The Korean Communications
in Statistics Vol. 6, No. 1, 1999
pp. 285-292

Estimation of the Lorenz Curve of the Pareto Distribution

Suk-Bok Kang[1] and Young-Suk Cho[2]

Abstract

In this paper, we propose the several estimators of the Lorenz curve in the Pareto distribution and obtain the bias and the mean squared error for each estimator.

We compare the proposed estimators with the uniformly minimum variance unbiased estimator (UMVUE) and the maximum likelihood estimator (MLE) in terms of the mean squared error (MSE) through Monte Carlo methods and discuss the results.

1. Introduction

A continuous random variable X has the Pareto distribution with the scale parameter $\theta > 0$ and the shape parameter $\xi > 0$ if it has a cumulative distribution function (cdf) of the form

$$F(x) = 1 - (\theta/x)^{\xi}, \quad x \geq \theta. \tag{1.1}$$

We will use the notation $X \sim \mathrm{PAR}(\theta, \xi)$. The importance of this distribution lies in its application to income, wealth, and service time queueing system. Likes (1969) derived the UMVUEs of the parameters in the Pareto distribution. Malik (1970) has derived distributions of the MLEs of the parameters in the Pareto distribution. The means and covariances of the order statistics from the Pareto distribution are given by Kulldorff and Vannman (1973). Woo and Kang (1990) considered a more general class of the UMVUE for the function of two parameters in the Pareto distribution. Kang and Cho (1996) obtained the jackknife estimator, the generalized jackknife estimator, and the minimum risk estimator (MRE) of two parameters in the Pareto distribution.

The Lorenz curve is extensively used in the study of inequality distribution and used to be a powerful tool for the analysis of a variety of scientific problems. The Lorenz curve is given by

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

1) Professor, Department of Statistics and Institute of Natural Sciences, Yeungnam University, Kyongsan, 712-749, Korea

2) Adjunct Assistant Professor, Department of Statistics, Yeungnam University, Kyongsan, 712-749, Korea

where Y is a non-negative income variable for which the mathematical expectation $\mu = E(Y)$ exists and $p = F(y)$ is the cdf. Since the cdf of all specified models of income distribution are strictly increasing and continuously differentiable functions, $y = F^{-1}(p)$ is well defined. Replacing it in the equation (1.2), the Lorenz curve is given by

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

Moothathu (1985) derived the MLEs of the Lorenz curve and the Gini index of a Pareto distribution, their exact and asymptotic distributions and moments. Moothathu (1990) also obtained the UMVUE and a strongly consistent asymptotically normal unbiased estimator (SCANUE) of the Lorenz curve, the Gini index, and Theil entropy index of a Pareto distribution. Castillo *et al.* (1998) proposed a new method for estimating the parameters of the Lorenz curves and fitting the Lorenz curves to observed data.

The Lorenz curve $L(p, \xi)$ of the Pareto distribution (1.1) is given by

$$\begin{aligned} L(p) &= 1-(1-p)^{1-\xi^{-1}} \\ &= 1-(1-p)\exp -\xi^{-1}\log(1-p)) \\ &= 1-(1-p)\sum_{r=0}^{\infty}[-\xi^{-1}\log(1-p)]^r/r!, \quad 0 \le p \le 1, \ 1 < \xi. \end{aligned} \tag{1.4}$$

In section 2, we propose the several estimators and calculate the biases and the MSEs of the proposed estimators of the Lorenz curve when the scale parameter is known. In section 3, we propose the several estimators and calculate the biases and the MSEs of the proposed estimators of the Lorenz curve when two parameters are unknown. In section 4, we compare the proposed estimators with the UMVUE and the MLE in terms of the MSE through Monte Carlo methods and discuss the results.

2. Estimation of Lorenz curve when the scale parameter is known

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Pareto distribution with distribution function (1.1) and $X_{(j)}$ $(j = 1, 2, \ldots, n)$ be the j-th order statistic based on a random sample $X_1, X_2, \ldots, X_n$. Then $T_n = \sum_{j=1}^{n} \log(X_j/\theta)$ has a gamma distribution GAM(ξ, n) with density function $\xi^n t^{n-1} e^{-\xi t}/\Gamma(n)$ and $T_n^r/(n)_r$ $(r = 1, 2, \ldots)$ is an unbiased estimator of ξ^{-r} where $(n)_r = n(n+1)\cdots(n+r-1)$. From the above result, Moothathu (1985, 1990) obtained the UMVUE $\widehat{L(p)}_U$ and the MLE $\widehat{L(p)}_{MLE}$ of $L(p,\xi)$ as follows;

$$\widehat{L(p)}_U = 1-(1-p)\sum_{r=0}^{\infty}\frac{[-T_n\log(1-p)]^r}{(n)_r r!} \tag{2.1}$$

and

$$\widehat{L(p)}_{MLE} = 1-(1-p)\sum_{r=0}^{\infty}\frac{[-T_n\log(1-p)]^r}{n^r r!}\,. \tag{2.2}$$

He also obtained the MSEs of the UMVUE $\widehat{L(p)}_U$ and the MLE $\widehat{L(p)}_{MLE}$ as follows;

$$\begin{aligned}\text{MSE}[\widehat{L(p)}_U] = (1-p)^2\Bigg(&\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(n)_{r+s}[-\xi^{-1}\log(1-p)]^{r+s}}{(n)_r(n)_s r!s!}\\ &-\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{[-\xi^{-1}\log(1-p)]^{r+s}}{r!s!}\Bigg)\end{aligned} \tag{2.3}$$

and

$$\begin{aligned}\text{MSE}[\widehat{L(p)}_{MLE}] = (1-p)^2\Bigg(&\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(n)_{r+s}[-\xi^{-1}\log(1-p)]^{r+s}}{n^r n^s r!s!}\\ &-2\sum_{r=0}^{\infty}\frac{(n)_r[-\xi^{-1}\log(1-p)]^r}{n^r r!}(1-p)^{-\xi^{-1}}+(1-p)^{-2\xi^{-1}}\Bigg).\end{aligned} \tag{2.4}$$

The UMVUE of the shape parameter ξ is $\hat{\xi}_U=(n-1)/\sum_{j=1}^{n}\log(X_j/\theta)$. Kang and Cho (1997) proposed the MRE $\hat{\xi}_{MRE}=(n-2)/\sum_{j=1}^{n}\log(X_j/\theta)$ of ξ. We use the estimators $\hat{\xi}_k=(n+k)/\sum_{j=1}^{n}\log(X_j/\theta)$, $(k=-2,-1,1,2)$ and replace ξ in (1.4), then we obtain the estimators of Lorenz curve as follows;

$$\widehat{L(p)}_k = 1-(1-p)\sum_{r=0}^{\infty}\frac{[-T_n\log(1-p)]^r}{(n+k)^r r!}\,. \tag{2.5}$$

We also obtain the bias and the MSE of $\widehat{L(p)}_k$ as follows ;

$$\text{Bias}(\widehat{L(p)}_k) = (1-p)\left(\sum_{r=0}^{\infty}\frac{(n)_r[-\xi^{-1}\log(1-p)]^r}{(n+k)^r r!}-(1-p)^{-\xi^{-1}}\right) \tag{2.6}$$

and

$$\begin{aligned}\text{MSE}[\widehat{L(p)}_k] = (1-p)^2\Bigg(&\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(n)_{r+s}[-\xi^{-1}\log(1-p)]^{r+s}}{(n+k)^r(n+k)^s r!s!}\\ &-2\sum_{r=0}^{\infty}\frac{(n)_r[-\xi^{-1}\log(1-p)]^r}{(n+k)^r r!}(1-p)^{-\xi^{-1}}+(1-p)^{-2\xi^{-1}}\Bigg).\end{aligned} \tag{2.7}$$

3. Estimation of Lorenz curve when both θ and ξ are unknown

It is well known that $S_{n-1}=\sum_{j=1}^{n}\log(X_j/X_{(1)})$ has a gamma distribution $GAM(\xi, n-1)$ and it is independent of $\log X_{(1)}$. When both θ and ξ are unknown, $(\log X_{(1)}, S_{n-1})$ is complete sufficient statistic for $PAR(\theta,\xi)$ and an unbiased estimator of $\xi^{-r}, (r=1,2,\ldots)$ is $S_{n-1}^{r}/(n-1)_r$. From the above results, Moothathu (1985, 1990) obtained the UMVUE $\widehat{L(p)}_U$ and the MLE $\widehat{L(p)}_{MLE}$ of $L(p,\xi)$ as follows;

$$\widehat{L(p)}_U = 1-(1-p)\sum_{r=0}^{\infty}\frac{[-S_{n-1}\log(1-p)]^r}{(n-1)_r r!} \tag{3.1}$$

and

$$\widehat{L(p)}_{MLE} = 1-(1-p)\sum_{r=0}^{\infty}\frac{[-S_{n-1}\log(1-p)]^r}{n^r r!}\ . \tag{3.2}$$

He also obtained the MSEs of the UMVUE $\widehat{L(p)}_U$ and the MLE $\widehat{L(p)}_{MLE}$ as follows;

$$\begin{aligned}\mathrm{MSE}[\ \widehat{L(p)}_U] = \ & (1-p)^2\Bigg(\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(n-1)_{r+s}[-\xi^{-1}\log(1-p)]^{r+s}}{(n-1)_r(n-1)_s r!s!} \\ & -\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{[-\xi^{-1}\log(1-p)]^{r+s}}{r!s!}\Bigg)\end{aligned} \tag{3.3}$$

and

$$\begin{aligned}\mathrm{MSE}[\ \widehat{L(p)}_{MLE}] = \ & (1-p)^2\Bigg(\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(n-1)_{r+s}[-\xi^{-1}\log(1-p)]^{r+s}}{n^r n^s r!s!} \\ & -2\sum_{r=0}^{\infty}\frac{(n-1)_r[-\xi^{-1}\log(1-p)]^r}{n^r r!}(1-p)^{-\xi^{-1}}+(1-p)^{-2\xi^{-1}}\Bigg).\end{aligned} \tag{3.4}$$

The UMVUE of the shape parameter ξ is $\hat{\xi}_U=(n-2)/\sum_{j=1}^{n}\log(X_j/X_{(1)})$. Kang and Cho (1997) proposed the MRE $\hat{\xi}_{MRE}=(n-3)/\sum_{j=1}^{n}\log(X_j/X_{(1)})$ of ξ. We use the estimators $\hat{\xi}_k=(n+k)/\sum_{j=1}^{n}\log(X_j/X_{(1)}), (k=-3,-2,1,2)$ and replace ξ in (1.4), then we can obtain the estimators of the Lorenz curve as follows;

$$\widehat{L(p)}_k = 1-(1-p)\sum_{r=0}^{\infty}\frac{[-S_{n-1}\log(1-p)]^r}{(n+k)^r r!}. \tag{3.5}$$

The bias and the MSE of $\widehat{L(p)}_k$ are given by

$$\text{Bias}(\widehat{L(p)}_k) = (1-p)\left(\sum_{r=0}^{\infty}\frac{(n-1)_r[-\xi^{-1}\log(1-p)]^r}{(n+k)^r r!} - (1-p)^{-\xi^{-1}}\right) \tag{3.6}$$

and

$$\begin{aligned}\text{MSE}[\widehat{L(p)}_k] = & (1-p)^2\left(\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(n-1)_{r+s}[-\xi^{-1}\log(1-p)]^{r+s}}{(n+k)^r(n+k)^s r!s!}\right. \\ & \left. -2\sum_{r=0}^{\infty}\frac{(n-1)_r[-\xi^{-1}\log(1-p)]^r}{(n+k)^r r!}(1-p)^{-\xi^{-1}} + (1-p)^{-2\xi^{-1}}\right).\end{aligned} \tag{3.7}$$

4. The Simulated Result

Random numbers of the two-parameter Pareto distribution were generated by IMSL subroutine RNUN and transformed $\theta(1-\text{RNUN})^{-1/\xi}$ in Fortran program.

From the estimators (2.1), (2.2), and (2.5), we calculate the MSEs of $\widehat{L(p)}_U$, $\widehat{L(p)}_{MLE}$, and $\widehat{L(p)}_k, (k=-2,-1,1,2)$ for each $\theta=1.0$; $\xi=2,3$; $p=0.3(0.2)0.7$. The simulation procedure is repeated 1,000 times for each sample sizes $n=10(10)30$. These values are given Table 1.

From Table 1, the UMVUE $\widehat{L(p)}_U$ is more efficient than the MLE $\widehat{L(p)}_{MLE}$, the proposed estimator $\widehat{L(p)}_1$ is more efficient than the other estimators when p is small, and the proposed estimator $\widehat{L(p)}_2$ is more efficient than the other estimators in terms of the MSE when p is large.

From the estimators (3.1), (3.2), and (3.5), we calculate the MSEs of $\widehat{L(p)}_U$, $\widehat{L(p)}_{MLE}$, and $\widehat{L(p)}_k, (k=-3,-2,1,2)$ for each $\theta=1.0$; $\xi=2,3$; $p=0.3(0.2)0.7$. The simulation procedure is repeated 1,000 times for each sample sizes $n=10(10)30$. These values are given Table 2.

From Table 2, the MLE $\widehat{L(p)}_{MLE}$ is more efficient than the other estimators when p is small, and the proposed estimator $\widehat{L(p)}_1$ is more efficient than the other estimators in terms of the MSE when p is large.

Table 1. The MSEs of several estimators when the scale parameter is known.
($\xi=2.0$, $\theta=1.0$)

p	n	$\widehat{L(p)}_{-2}$	$\widehat{L(p)}_{-1}$	$\widehat{L(p)}_{MLE}$	$\widehat{L(p)}_{1}$	$\widehat{L(p)}_{2}$	$\widehat{L(p)}_{U}$
0.3	10	.00473	.00294	.00209	.00183	.00193	.00202
	20	.00172	.00131	.00109	.00101	.00103	.00107
	30	.00104	.00087	.00077	.00073	.00073	.00076
0.5	10	.01436	.00867	.00590	.00492	.00500	.00545
	20	.00497	.00373	.00303	.00273	.00272	.00290
	30	.00297	.00244	.00212	.00196	.00195	.00205
0.7	10	.03147	.01820	.01164	.00903	.00866	.00994
	20	.01012	.00736	.00576	.00499	.00481	.00528
	30	.00586	.00471	.00399	.00360	.00348	.00376

($\xi=3.0$, $\theta=1.0$)

p	n	$\widehat{L(p)}_{-2}$	$\widehat{L(p)}_{-1}$	$\widehat{L(p)}_{MLE}$	$\widehat{L(p)}_{1}$	$\widehat{L(p)}_{2}$	$\widehat{L(p)}_{U}$
0.3	10	.00204	.00122	.00085	.00074	.00078	.00083
	20	.00067	.00051	.00043	.00040	.00041	.00042
	30	.00040	.00034	.00030	.00029	.00029	.00030
0.5	10	.00542	.00315	.00212	.00178	.00182	.00200
	20	.00168	.00127	.00105	.00096	.00098	.00102
	30	.00101	.00084	.00074	.00069	.00069	.00072
0.7	10	.00967	.00540	.00344	.00274	.00270	.00309
	20	.00277	.00205	.00165	.00147	.00146	.00156
	30	.00163	.00133	.00115	.00106	.00105	.00111

Table 2. The MSEs of several estimators when both ξ and θ are unknown.

($\xi=2.0$, $\theta=1.0$)

p	n	$\widehat{L(p)}_{-3}$	$\widehat{L(p)}_{-2}$	$\widehat{L(p)}_{MLE}$	$\widehat{L(p)}_{1}$	$\widehat{L(p)}_{2}$	$\widehat{L(p)}_{U}$
0.3	10	.00580	.00357	.00213	.00220	.00252	.00279
	20	.00187	.00142	.00108	.00110	.00120	.00127
	30	.00108	.00089	.00075	.00075	.00080	.00085
0.5	10	.01776	.01059	.00574	.00571	.00639	.00774
	20	.00545	.00403	.00291	.00289	.00311	.00350
	30	.00307	.00251	.00202	.00200	.00210	.00234
0.7	10	.03942	.02252	.01058	.00987	.01065	.01469
	20	.01115	.00801	.00534	.00511	.00534	.00654
	30	.00608	.00485	.00370	.00358	.00368	.00434

($\xi=3.0$, $\theta=1.0$)

p	n	$\widehat{L(p)}_{-3}$	$\widehat{L(p)}_{-2}$	$\widehat{L(p)}_{MLE}$	$\widehat{L(p)}_{1}$	$\widehat{L(p)}_{2}$	$\widehat{L(p)}_{U}$
0.3	10	.00266	.00152	.00087	.00089	.00102	.00117
	20	.00073	.00055	.00043	.00044	.00048	.00050
	30	.00042	.00035	.00029	.00030	.00032	.00033
0.5	10	.00714	.00398	.00209	.00210	.00236	.00288
	20	.00185	.00138	.00103	.00103	.00112	.00122
	30	.00104	.00086	.00071	.00071	.00076	.00081
0.7	10	.01299	.00694	.00325	.00310	.00339	.00458
	20	.00307	.00224	.00157	.00155	.00165	.00190
	30	.00168	.00137	.00109	.00108	.00112	.00126

References

[1] Castillo, E., Hadi, A. S., and Sarabia, J. M. (1998). A Method for Estimating Lorenz Curves, *Communications in Statistics- Theory and Methods*, Vol. 27(8), 2037-2063.

[2] Kulldorff, G. and Vannman, K. (1973). Estimation of the Location and Scale Parameters of a Pareto Distribution by Linear Functions of Order Statistics, *Journal of American Statistical Association*, Vol. 68, 218-227.

[3] Likes, J. (1969). Minimum Variance Unbiased Estimation of the Parameters of power-function and Pareto's Distribution, *Statistische Hefte*, 10, 104-110.

[4] Malik, H. J. (1970). Estimation of the Parameters of the Pareto Distribution, *Metrika*, Vol. 16, 126-132.

[5] Moothathu, T. S. K. (1985). Sampling Distribution of Lorenz Curve and Gini Index of the Pareto Distribution. *Sankhyā*, Vol. 47(B), 247-278.

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

[7] Kang, S. B. and Cho, Y. S. (1996). Estimation of the Parameters in a Pareto Distribution by Jackknife and Bootstrap Method. *Journal of Information & Optimization Sciences,* Vol., 18(2), 289-300.

[8] Woo, J. and Kang, S. B. (1990). Estimation for Functions of Two Parameters in the Pareto Distribution, *Youngnam Statistical Letters*, Vol. 1, 67-76.