The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 291-297

Estimation of Gini Index of the Exponential Distribution by Bootstrap Method

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

Abstract

In this paper, we propose the jackknife estimator and the bootstrap estimator of Gini index of the two-parameter exponential distribution when the location parameter θ is unknown and the scale parameter σ is known. Similarly, we propose the bias corrected estimator of Gini index by using the bootstrap estimator of bias when the location parameter θ and the scale parameter σ are unknown. The bootstrap estimator is more efficient than the other estimators when the location parameter θ is unknown and the scale parameter σ is known, and the bias corrected estimator is more efficient than the MLE when both the location parameter θ and the scale parameter σ are unknown.

1. Introduction

The Lorenz curve and Gini index have been extensively used in the study of inequality distribution and used to be a powerful tool for the analysis of a variety of scientific problems; e.g., as a criterion to perform a partial ordering of social welfare states, to extend the concept of the Lorenz curve to functions of income. The Lorenz curve is given by

$$L(x) = \int_0^x yF(y)dt/E(X), \tag{1}$$

where X is a nonnegative income variable for which the mathematical expectation $\mu = E(X)$ exists, and $p = F(x)$ is the cumulative distribution function. The Gini index (also known as Gini concentration ratio or Gini coefficient or Lorenz concentration ratio) is twice the area between the Lorenz curve and the identity function (equidistribution function) $L(x) = F(x)$. Since the cumulative distribution function of all specified models of income

1) Associate Professor, Department of Statistics, Yeungnam University, Kyongsan, 712-749, Korea
2) Lecturer, Department of Statistics, Yeungnam University, Kyongsan, 712-749, Korea

distribution are strictly increasing and continuously differentiable functions, $x = F^{-1}(p)$ is well defined. Replacing it in (1),

$$L(p) = \int_0^p F^{-1}(t)dt / E(X).$$

Moothathu (1985a) derived the exact and asymptotic distributions of MLEs of Lorenz curve and Gini index of an exponential distribution. Moothathu (1985b) also derived the MLEs of Lorenz curve and Gini index of a Pareto distribution, their exact and asymptotic distributions and moments. Moothathu (1990) obtained the uniformly minimum variance unbiased estimator (UMVUE) and a strongly consistent asymptotically normal unbiased estimator (SCANUE) of Lorenz curve, Gini index and Theil entropy index of a Pareto distribution.

The jackknife was frist introduced by Quenouille (1956) and Tukey (1958), who realized its importance as an almost universally applicable tool for bias reduction and robust interval estimation. This method is a powerful idea for bias reduction and distribution free estimation of the variance of an estimator. The jackknife method and its application to bias reduction and dispersion estimation have been surveyed by Miller (1974). The bootstrap method, introduced by Efron (1979), is a resampling technique. It has been used to estimate variances, confidence intervals, and hypothesis testing problems. The bootstrap method and other methods for assessing statistical accuracy are summarized by Efron and Tishirani (1993). We consider the Lorenz curve and the Gini index based on a random sample from an exponential distribution with cumulative distribution function

$$F(x) = P(X \leq x) = 1 - \exp\{-(x-\theta)/\sigma\}, \quad 0 < \theta < x, \ 0 < \sigma.$$

The Lorenz curve $L(p)$ and the Gini index $g(\theta, \sigma)$ of the exponential distribution are given by

$$\begin{aligned} L(p) &= \int_0^p F^{-1}(t)\, dt / E(X) \\ &= p + \sigma(\theta+\sigma)^{-1}(1-p)\log(1-p), \quad 0 \leq p \leq 1 \end{aligned}$$

and

$$g(\theta, \sigma) = 1 - 2\int_0^1 L(p)\, dp = \sigma / 2(\theta+\sigma).$$

In section 2, we propose the jackknife estimator $J(\widehat{g_1})$ by using the MLE and the bootstrap estimator $\overline{J(\widehat{g_1^*})}$ by using the jackknife estimator when the location parameter θ is unknown and the scale parameter σ is known, and show that the jackknife estimator $J(\widehat{g_1})$ is converges in probability to $g(\theta,\sigma)$. Similarly, we propose the bias corrected estimator $\overline{g_2}$ by using the bootstrap estimator of bias when both the location parameter θ and the scale parameter σ are unknown. In section 3, we compare the mean square error of the several estimators through Monte Carlo Method and summarize the numerical results.

2. Estimators of Gini index

Let $X_{(r)}$ $(r=1,2,\cdots,n)$ be the r-th order statistics based on a random sample of size n from $F(x)$ and let $S=\sum_{i=1}^{n}(X_i-X_{(1)})/n$. It is well know that the joint MLE of (θ,σ) is $(X_{(1)},S)$. By this result, Moothathu (1985a) obtained the MLE $\widehat{g_1}=\sigma/2(X_{(1)}+\sigma)$ and almost sure convergence of $\widehat{g_1}$ when the location parameter θ is unknown and the scale parameter σ is known. Similarly, he obtained the MLE $\widehat{g_2}=S/2(X_{(1)}+S)$ and almost sure convergence of $\widehat{g_2}$ when the location parameter θ and the scale parameter σ are unknown.

We are considering the MLE when the scale parameter σ is known, and define

$$\widehat{g_1^i}=\begin{cases}\sigma/2(X_{(1)}+\sigma), & \text{if } X_i\neq X_{(1)}\\ \sigma/2(X_{(2)}+\sigma), & \text{if } X_i=X_{(1)}\end{cases}$$

to be the MLE defined on the subsample which arises when the i-th subset of size one has been deleted, and the average

$$\begin{aligned}\overline{\widehat{g_1^i}}&=\sum_{i=1}^{n}\widehat{g_1^i}/n\\&=\{(n-1)\sigma/(X_{(1)}+\sigma)+\sigma/(X_{(2)}+\sigma)\}/2n.\end{aligned}$$

The jackknife estimator of $g(\theta,\sigma)$ is given by

$$\begin{aligned}J(\widehat{g_1})&=n\widehat{g_1}-(n-1)\,\overline{\widehat{g_1^i}}\\&=(2n-1)\sigma/2n(X_{(1)}+\sigma)-(n-1)\sigma/2n(X_{(2)}+\sigma).\end{aligned}$$

From Grodshteyn and Ryzhik (1965) formula 3.352, the expectation of $J(\widehat{g_1})$ is given by

$$
\begin{aligned}
E[J(\widehat{g_1})] &= \int_\theta^\infty \frac{(2n-1)}{2(x+\sigma)} \exp\left(-\frac{n(x-\theta)}{\sigma}\right)dx \\
&\quad - \int_\theta^\infty \frac{(n-1)^2}{2(x+\sigma)}\left(\exp\left(-\frac{(n-1)(x-\theta)}{\sigma}\right) - \exp\left(-\frac{n(x-\theta)}{\sigma}\right)\right)dx \\
&= \int_0^\infty \frac{n^2}{2(x+\theta+\sigma)} \exp\left(-\frac{nx}{\sigma}\right)dx - \int_0^\infty \frac{(n-1)^2}{2(x+\theta+\sigma)} \exp\left(-\frac{(n-1)x}{\sigma}\right)dx \\
&= -n^2 e^{n(\theta+\sigma)/\sigma} E_i[-n(\theta+\sigma)/\sigma]/2 \\
&\quad + (n-1)^2 e^{(n-1)(\theta+\sigma)/\sigma} E_i[-(n-1)(\theta+\sigma)/\sigma]/2
\end{aligned}
$$

where $E_i(x) = C + \ln(-x) + \sum_{k=1}^{\infty} x^k/(k\times k!)$ and C is Euler's constant. The following theorem shows that the jackknife estimator $J(\widehat{g_1})$ converges in probability to $g(\theta,\sigma)$.

Theorem. Let $X_1, X_2, \cdots, X_n$ is a random sample of size n from the exponential distribution function $F(x)$ with unknown θ and known σ. Let $g(\theta,\sigma)$ is the Gini index of exponential distribution. Then $J(\widehat{g_1}) = (2n-1)\sigma/2n(X_{(1)}+\sigma) - (n-1)\sigma/2n(X_{(2)}+\sigma)$ is a consistent estimator of $g(\theta,\sigma)$.

Proof. Since $E[(X_{(1)}-\theta)^2] = 2\sigma^2/n^2$ and $E[(X_{(2)}-\theta)^2] = 2(3n^2-3n+1)\sigma^2/(n-1)^2n^2$, we get $X_{(1)} \xrightarrow{p} \theta$ and $X_{(2)} \xrightarrow{p} \theta$. From these results and Slutsky's theorem, we obtain $J(\widehat{g_1}) \xrightarrow{p} g(\theta,\sigma)$.

The Bootstrap method algorithm is the following. First, we select B independent bootstrap samples $X^{*1}, X^{*2}, \cdots, X^{*B}$ each consisting of n data values drawn with replacement from the population of n observations($X_1, X_2, \cdots, X_n$). Second, evaluate the bootstrap replication corresponding to each bootstrap sample ($b=1,2,\cdots,B$),

$$J(\widehat{g_1^*}(b)) = (2n-1)\sigma/2n(X^*_{(1)}(b)+\sigma) - (n-1)\sigma/2n(X^*_{(2)}(b)+\sigma)$$

and

$$\widehat{g_2^*}(b) = S^*(b)/2(X^*_{(1)}(b) + S^*(b)),$$

where $S^*(b) = \sum_{i=1}^{n}(X^*_i(b) - X^*_{(1)}(b))/n$.

We propose the bootstrap estimator $\overline{J(\widehat{g_1^*})} = \sum_{b=1}^{B} J(\widehat{g_1^*}(b))/B$ by using the jackknife estimator when the location parameter θ is unknown and the scale parameter σ is known. When both the location parameter θ and the scale parameter σ are unknown, we want to estimate the real valued parameter $g(\theta, \sigma)$. We will take the MLE $\widehat{g_2}$ as an estimator of $g(\theta, \sigma)$ and the bias of $\widehat{g_2}$ is defined to be the difference between the expectation of $\widehat{g_2}$ and the value of the parameter $g(\theta, \sigma)$, $bias_F(\widehat{g_2}) = E_F[\widehat{g_2}] - g(\theta, \sigma)$. We can use the bootstrap to assess the bias of the MLE $\widehat{g_2}$. The bootstrap estimator of bias is proposed by Efron(1979a), and defined to be the estimator $bias_{F_n}(\widehat{g_2}) = E_{F_n}[\widehat{g_2^*}] - \widehat{g_2}$ by substituting empirical distribution F_n for F. We use approximately the bootstrap expectation $E_{F_n}[\widehat{g_2^*}]$ by the average $\widehat{g_2^*}(\cdot) = \sum_{b=1}^{B} \widehat{g_2^*}(b)/B$. Thus, the bootstrap estimator of bias based on the B replications is $\widehat{bias_B}(\widehat{g_2}) = \widehat{g_2^*}(\cdot) - \widehat{g_2}$. We propose the bias corrected estimator $\overline{g_2}$ of $g(\theta, \sigma)$ as follows;

$$
\begin{aligned}
\overline{g_2} &= \hat{g}_2 - \widehat{bias_B}(\widehat{g_2}) \\
&= 2\,\hat{g}_2 - \widehat{g_2^*}(\cdot).
\end{aligned}
$$

3. The Simulated Result

We calculate the mean square errors of several estimators for sample size n = 5(30)5 (based on 1,000 Monte Carlo runs and B=1,000) when the location parameter θ = 0.5(1.0)2.5 and the scale parameter σ = 0.5(0.5)1.5. From the table, the bootstrap estimator $\overline{J(\widehat{g_1^*})}$ is more efficient than the other estimators when the location parameter θ is unknown and the scale parameter σ is known, and the bias corrected estimator $\overline{g_2}$ is more efficient than the MLE when the location parameter θ and the scale parameter σ are unknown. We can also see that the MSE decreases as θ increases or σ decreases.

Table. The mean square errors of several estimators

θ	n	$\sigma=0.5$				
		$\widehat{g_1}$	$J(\widehat{g_1})$	$\overline{J(\widehat{g_1^*})}$	$\widehat{g_2}$	$\overline{g_2}$
0.5	5	.00325	.00273	.00268	.02772	.02294
	10	.00095	.00086	.00074	.00991	.00818
	15	.00042	.00043	.00031	.00612	.00538
	20	.00025	.00024	.00019	.00398	.00350
	25	.00017	.00018	.00013	.00325	.00297
	30	.00013	.00012	.00010	.00276	.00250
1.5	5	.00023	.00020	.00018	.00863	.00858
	10	.00007	.00007	.00006	.00403	.00400
	15	.00004	.00003	.00003	.00257	.00251
	20	.00002	.00002	.00001	.00197	.00196
	25	.00001	.00001	.00001	.00146	.00148
	30	.00001	.00001	.00001	.00125	.00124
2.5	5	.00005	.00005	.00004	.00435	.00447
	10	.00001	.00001	.00001	.00207	.00210
	15	.00001	.00001	.00000	.00139	.00142
	20	.00000	.00000	.00000	.00103	.00102
	25	.00000	.00000	.00000	.00079	.00078
	30	.00000	.00000	.00000	.00068	.00070

Table.(continued)

θ	n	$\sigma=1.0$				
		$\widehat{g_1}$	$J(\widehat{g_1})$	$\overline{J(\widehat{g_1^*})}$	$\widehat{g_2}$	$\overline{g_2}$
0.5	5	.00822	.00699	.00672	.04077	.02849
	10	.00268	.00247	.00212	.01297	.00939
	15	.00135	.00212	.00104	.00708	.00527
	20	.00086	.00078	.00066	.00519	.00409
	25	.00055	.00054	.00042	.00367	.00296
	30	.00042	.00043	.00033	.00308	.00259
1.5	5	.00132	.00111	.00106	.01964	.01725
	10	.00044	.00049	.00034	.00792	.00726
	15	.00021	.00019	.00016	.00484	.00455
	20	.00011	.00010	.00008	.00335	.00317
	25	.00008	.00008	.00006	.00285	.00271
	30	.00005	.00004	.00004	.00222	.00206
2.5	5	.00039	.00034	.00031	.01156	.01109
	10	.00010	.00009	.00008	.00492	.00463
	15	.00005	.00005	.00004	.00312	.00305
	20	.00003	.00003	.00002	.00230	.00228
	25	.00002	.00002	.00001	.00184	.00181
	30	.00001	.00001	.00001	.00149	.00142

Table.(continued)

θ	n	$\sigma=1.5$				
		$\widehat{g_1}$	$J(\widehat{g_1})$	$\overline{J(\widehat{g_1^*})}$	$\widehat{g_2}$	$\overline{g_2}$
0.5	5	.01405	.01206	.01188	.04978	.03308
	10	.00403	.00392	.00317	.01398	.00966
	15	.00226	.00225	.00179	.00802	.00574
	20	.00126	.00116	.00099	.00450	.00335
	25	.00094	.00084	.00074	.00386	.00285
	30	.00058	.00053	.00044	.00233	.00178
1.5	5	.00316	.00266	.00260	.02778	.02223
	10	.00099	.00089	.00078	.01100	.00921
	15	.00046	.00046	.00035	.00602	.00542
	20	.00026	.00025	.00020	.00432	.00383
	25	.00017	.00016	.00013	.00343	.00305
	30	.00013	.00014	.00010	.00253	.00237
2.5	5	.00110	.00104	.00088	.01635	.01488
	10	.00031	.00028	.00024	.00757	.00692
	15	.00015	.00015	.00011	.00433	.00405
	20	.00009	.00009	.00007	.00321	.00305
	25	.00006	.00005	.00005	.00262	.00251
	30	.00004	.00004	.00003	.00207	.00198

References

[1] Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.* 7, 1-26.

[2] Efron, B and Tishirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall.

[3] Grodshteyn, I. S and Ryzhik, I. M. (1965). *Table of Integrals, Series, and Products*, Academic Press.

[4] Miller, R. G. (1974). The Jacknife-a review. *Biometrika,* 61, 1-15.

[5] Moothathu, T. S. K. (1985a). Distribution of Maximum Likelihood Estimators of Lorenz Curve and Gini Index of Exponential Distribution. *Ann. Inst. Statist. Math.*, 37, 437-497.

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

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

[8] Quenouille, M. (1956). Notes on Bias in Estimation. *Biometrika,* 43, 353-360.

[9] Tukey, J. W. (1958). Bias and Confidence in not Quite Large Samples(abstract). *Ann. Math. Statist.*, 29, 614.