The Korean Communications
in Statistics Vol. 7, No. 1, 2000
pp. 277-283

Goodness-of-Fit Test for the Exponential Distribution Based on the Transformed Sample Lorenz curve[1)]

Suk-Bok Kang[2)] and Young-Suk Cho[3)]

Abstract

The transformed sample Lorenz curve provides a powerful and easily computed goodness-of-fit test for exponentiality which does not depend on the unknown scale parameter. We compare the power of the transformed sample Lorenz curve statistic with the other goodness-of-fit tests for exponentiality against various alternatives through Monte Carlo methods and discuss the results.

1. Introduction

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

$$L(x) = \int_0^x y\, dF(y)/E(X) \tag{1.1}$$

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 (cdf). Since the cdfs of income distribution are strictly increasing and continuously differentiable functions, $x = F^{-1}(p)$ is well defined. Replacing it in the equation (1.1), the Lorenz curve (Gastwirth (1971)) is given by

$$L(p) = \int_0^p F^{-1}(y)\, dy/E(X). \tag{1.2}$$

Assume that $X_1, X_2, \ldots, X_n$ are positive random variables with order statistics $X_{(1)} < \ldots < X_{(n)}$. Let $r = [np]$ denote the greatest integer less than or equal to np. Then the sample Lorenz curve (Gail and Gastwirth (1978)) is defined by

1 This research was supported by the Yeungnam University research grants in 1999
2 Professor, Department of Statistics and Institute of Natural Sciences, Yeungnam University, Kyongsan, 712-749, Korea
3 Adjunct Assistant Professor, Department of Statistics, Yeungnam University, Kyongsan, 712-749, Korea

$$L_n(p) = \frac{\sum_{i=1}^{r=[np]} X_{(i)}}{\sum_{i=1}^{n} X_{(i)}}. \tag{1.3}$$

Cho *et al.* (1999) proposed the transformed Lorenz curve that can be used in the study of symmetric distribution. The transformed Lorenz curve is defined by

$$TL(p) \equiv L(p) - p + 1. \tag{1.4}$$

From figures A and B, The transformed Lorenz curve more efficient than the Lorenz curve in the study of skewness. The transformed sample Lorenz curve is defined by

$$TL_n(p) = L_n(p) - p + 1. \tag{1.5}$$

Gail and Gastwirth (1978) studied the scale free goodness-of-fit test for the exponential distribution based on the Lorenz curve. Moothathu (1985) derived the maximum likelihood estimators (MLE) of the Lorenz curve and the Gini index of a Pareto distribution, their exact and asymptotic distributions and moments. Moothathu (1990) also obtained the uniformly minimum variance unbiased estimator (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. Kang and Cho (1999) proposed the several estimators of the Lorenz curve in the Pareto distribution.

In section 2, we introduce the Lorenz curve and the transformed Lorenz curve of the exponential distribution, the Pareto distribution, and the uniform distribution and study the properties of the transformed sample Lorenz curve of the exponential distribution.

In section 3, we compare the power of the transformed sample Lorenz curve statistic with the other goodness-of-fit tests for exponentiality against various alternatives through Monte Carlo methods and discuss the results.

2. The properties of $TL_n(p)$

A continuous random variable X has the exponential distribution with the scale parameter $\beta > 0$ if it has a cumulative distribution function (cdf) of the form

$$F(x) = 1 - \exp(-x/\beta), \quad x \geq 0. \tag{2.1}$$

The Lorenz curve $L(p)$ and the transformed Lorenz curve $TL(p)$ of the exponential distribution (2.1) are given by

$$L(p) = \beta^{-1}\int_0^p [-\beta\ln(1-t)]\,dt \\ = p + (1-p)\ln(1-p),\ 0<p<1 \tag{2.2}$$

and

$$TL(p) = 1 + (1-p)\ln(1-p),\ 0<p<1. \tag{2.3}$$

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 cdf of the form

$$F(x) = 1-(\theta/x)^{\xi},\ x\geq\theta. \tag{2.4}$$

The Lorenz curve $L(p)$ and the transformed Lorenz curve $TL(p)$ of the Pareto distribution (2.4) are given by

$$L(p) = 1-(1-p)^{1-\xi^{-1}},\ 0\leq p\leq 1,\ 1<\xi \tag{2.5}$$

and

$$TL(p) = 2-p-(1-p)^{1-\xi^{-1}},\ 0\leq p\leq 1,\ 1<\xi. \tag{2.6}$$

Now we consider the uniform distribution which is the symmetric distribution. The Lorenz curve $L(p)$ and the transformed Lorenz curve $TL(p)$ of the uniform distribution with the density function $f(x)=1/\theta,\ 0\leq x\leq\theta$ are given by

$$L(p) = p^2,\ 0\leq p\leq 1 \tag{2.7}$$

and

$$TL(p) = p^2 - p + 1,\ 0\leq p\leq 1. \tag{2.8}$$

The Lorenz curves of the exponential distribution, the Pareto distribution, and the uniform distribution are given in Figure A. The transformed Lorenz curves of the exponential distribution, the Pareto distribution, and the uniform distribution are given in Figure B.

Gail and Gastwirth (1978) showed that $L_n(p)$ converges almost surely to $L(p)$ for each fixed p, and that $n^{1/2}[L_n(p)-L(p)]$ converges in distribution to a normal variate provided $E(X^2)<\infty$. From the results, $TL_n(p)$ converges almost surely to $TL(p)$ for each fixed p, and that $n^{1/2}[TL_n(p)-TL(p)]$ converges in distribution to a normal variate provided $E(X^2)<\infty$.

When $F(x)=1-\exp(-x/\beta),\ x\geq 0$, Gail and Gastwirth (1978) obtained the exact distribution of $L_n(p)$ as follows;

$$\Pr[L_n(p) \geq y] = \begin{cases} 1, & \text{if } y \leq 0 \\ 0, & \text{if } y \geq r/n \\ \sum_{j=0}^{\xi} \frac{n!(r-j-y(n-j))^{n-1}(-1)^j}{(n-j)(n-r)^{r-1}(r-j)^{n-r-1} j!(r-j)!(n-r)!}, & \text{otherwise} \end{cases}$$

where $r=[np]$ and $\xi=(r-yn)/(1-y)$.

From the exact distribution of $L_n(p)$, we obtain the exact distribution of $TL_n(p)$ as follows;

$$\Pr[TL_n(p) \geq y] = \begin{cases} 1, & \text{if } y \leq 1-p \\ 0, & \text{if } y \geq r/n+1-p \\ \sum_{j=0}^{\xi} \frac{n!(r-j-(y+p-1)(n-j))^{n-1}(-1)^j}{(n-j)(n-r)^{r-1}(r-j)^{n-r-1} j!(r-j)!(n-r)!}, & \text{otherwise} \end{cases}$$

where $r=[np]$ and $\xi=(r-(y+p-1)n)/(2-y-p)$.

Given significance level α, we wish to test $H_0 : F(x)=1-\exp(-x/\beta)$. From the distribution of $TL_n(p)$, we obtain the equal tail acceptance region of the transformed sample Lorenz curve statistic $TL_n(0.6)$. At the significance level $\alpha=0.05$, this test is to reject H_0 if $TL_n(0.6) < 0.550644$ or $TL_n(0.6) > 0.748897$.

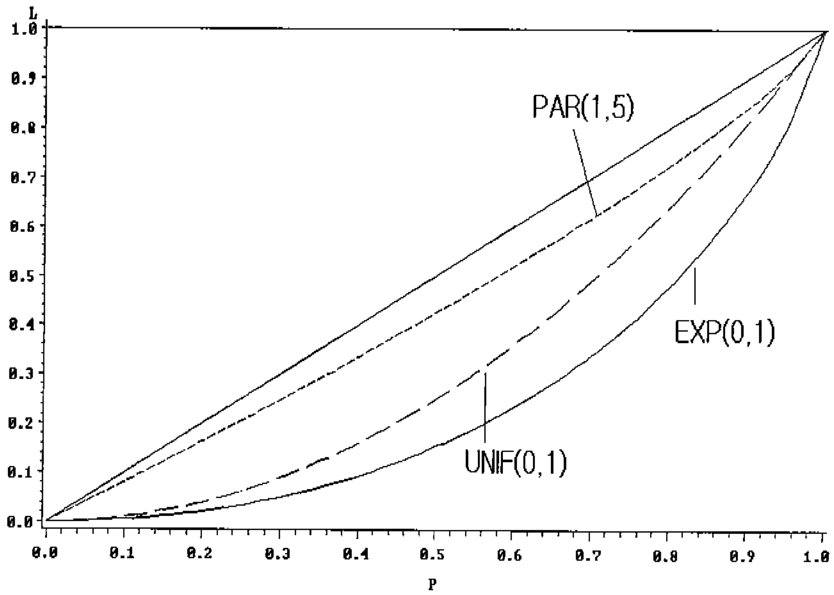

Figure A. The Lorenz curve

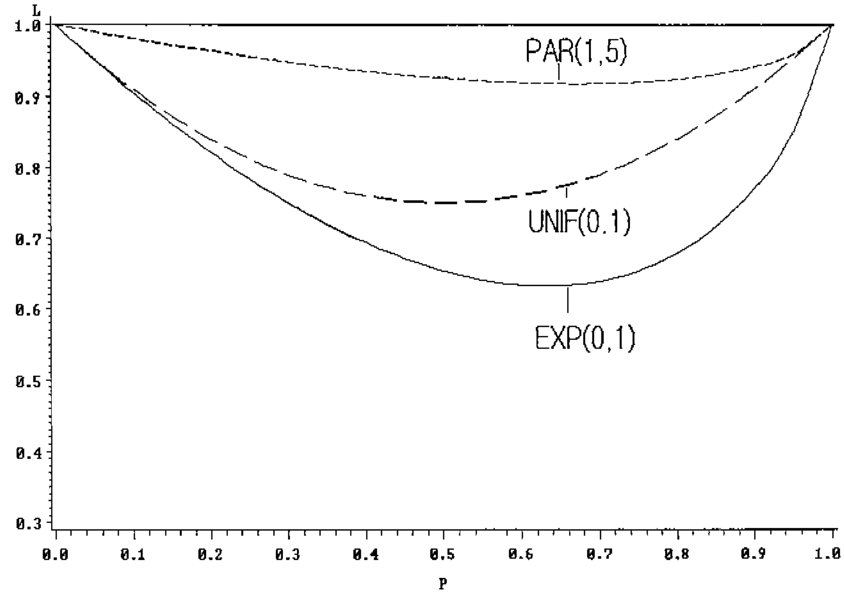

Figure B. The transformed Lorenz curve

3. The Simulated Results

Gail and Gastwirth (1978) showed that the sample Lorenz curve statistic $L_n(0.5)$ usually has greater power than the Durbin Kolmogrov-Smirnov (K-S) test (Durbin (1975)), the Increasing Failure Rate test (Proschan and Pyke(1967)), and the Shapiro-Wilk test (Shapiro and Wilk (1972)). The goodness-of-fit test based on the sample Lorenz curve statistic $L_n(0.5)$ has desirable power properties against a variety of alternatives, is readily calculated, and is insensitive to minor errors in the small observations.

In this section, we compare the power of the transformed sample Lorenz curve statistic with that of other goodness-of-fit tests for exponentiality ($H_0 : F(x) = 1 - \exp(-x/\beta)$) against alternatives. One thousand samples of size $n = 20$ were generated. Table 2 contains the proportion of such samples for which H_0 was rejected.

The alternatives studied are listed in Table 1. The random samples for each alternative were obtained from IMSL CALL in Fortran program except the two-parameter exponential distribution and the Pareto distribution. The random numbers of the Pareto distribution were generated by IMSL RNUN and transformed $(1 - \text{RNUN})^{-1/\xi}$ in Fortran program. The random numbers of the the two-parameter exponential distribution were generated by IMSL RNUN and transformed $\theta - \beta \ln(1 - \text{RNUN})$ in Fortran program. We obtain p-values of a

K-S one-sample test for continue distributions by the routine KSONE in IMSL. For $n > 80$, asymptotic p-values are used (Gibbons 1971). For $n \leq 80$, exact one-sided p-values are computed according to a method given by Conover (1980). An approximation two-sided test p-value is obtained as twice the one-sided p-value

From Table 2, the K-S test, the $L_n(0.5)$, and the $TL_n(0.6)$ are powerful against beta, gamma, exponential, and Pareto alternatives. The transformed sample Lorenz curve statistic usually has greater power than the Lorenz curve statistic against uniform and weibull alternatives. The transformed sample Lorenz curve statistic provides a powerful and easily computed goodness-of-fit test for exponentiality which does not depend on the unknown scale parameter.

Table 1. Alternative distribution studied

Name, Notation	probability density function	Random Number Generation Utility Routine
Beta, BETA(p, q)	$f(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, \ 0 \leq x \leq 1$	RNBET
Gamma, GAM(α)	$f(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x}, \ 0 \leq x$	RNGAM
Exponential, EXP(θ, β)	$f(x) = \frac{1}{\beta} \exp[-(x-\theta)/\beta], \ 0 \leq x$	RNEXP
Pareto, PAR(θ, ξ)	$f(x) = \xi \theta^{\xi} x^{-(\xi+1)}, \ \theta \leq x$	
Uniform, UNIF(0, θ)	$f(x) = \frac{1}{\theta}, \ 0 \leq x \leq \theta$	RNUN
Weibull, WIB(α, β)	$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp[-(x/\alpha)^{\beta}], \ 0 \leq x$	RNWIB

Table 2. Monte Carlo power estimates based on 1,000 samples of size $n = 20$ using significance level $\alpha = 0.05$

Goodness-of-fit test	Alternatives							
	BETA(2,1)	GAM(6)	EXP(2,1)	PAR(1,5)	UNIF(0,2)	WIB(0.8,1)	WIB(1.5,1)	EXP(0,1)
$TL_{20}(0.6)$	1.000	1.000	1.000	1.000	0.667	0.243	0.451	0.046
$L_{20}(0.5)$	0.999	0.999	1.000	1.000	0.573	0.233	0.447	0.054
K-S	1.000	1.000	1.000	1.000	0.26	0.088	0.092	0.050

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] Cho, Y. S., Lee, J. Y., and Kang, S. B. (1999). A study on distribution based on the transformed Lorenz curve. *The Korean Journal of Applied Statistics,* Vol. 12(1), 153-163.

[3] Cibbons, J. D. (1971). *Nonparametric Statistical Inference.* McGraw-Hill, New York.

[4] Conover, W. J. (1980). *Practical Nonparametric Statistics.* second edition, John Wiley & Sons, New York.

[5] Durbin, J. (1975). Kormogorov-Smirnov tests when parameters are estimated with applications to exponentiality and tests on spacings. *Biometrika,* Vol. 62, 5-22.

[6] Gail, M. H. and Gastwirth, J. L. (1978). A Scale-Free Goodness-of-Fit test for the exponential distribution based on Lorenz curve. *Journal of American Statistical Association*, Vol. 73, 787-793.

[7] Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, Vol. 39, 1037-1038.

[8] Kang, S. B. and Cho, Y. S. (1999). Estimation of the Lorenz curve of the Pareto distribution. *The Korean Communications in Statistics,* Vol. 6(1), 285-292.

[9] Moothathu, T. S. K. (1985). Sampling distribution of Lorenz curve and Gini index of the Pareto distribution. *Sankhyā*, Vol. 47(B), 247-278.

[10] 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.

[11] Proschan, F. and Pyke, R. (1967). Tests for monotone failure rate. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, III, Berkeley, Calif,: University of California Press, 293-312.

[12] Shapiro, S. S. and Francia, R. S. (1972). An approximation analysis of variance test for normality. *Journal of American Statistical Association*, Vol. 67, 215-216.