Aligarh Journal of Statistics
Vol. 31 (2011), 1-12

A new Skew Generalized Logistic distribution and Approximations to Skew Normal Distribution

Pushpa Narayan Rathie and Murilo Coutinho

ABSTRACT

Using the multimodal (for certain values of the parameters) generalized logistic distribution for a random variable Z, defined earlier by Rathie and Swamee (2006), we define a new skew generalized logistic distribution with four parameters a, b, p and c. Some properties, such as moments and distribution of Z^2, are obtained. This distribution is invertible for $c = \pm 1$. For certain values of the parameters a, b and p, and any c, this distribution approximates very well the skew normal distribution. As possible applications, the skew generalized logistic distribution is applied to analyse annual precipitation data of Los Angeles City and the Environmental Performance Index (2010).

1. INTRODUCTION

In this section, we include some known results which will be useful in the subsequent sections of this paper.

1.1 GENERALIZED LOGISTIC DISTRIBUTION

We define the following symmetric generalized logistic density function and its cumulative distribution function studied recently by Rathie et al. (2006,2008) and Rathie (2011).

$$f(z) = \frac{[a + b(1+p)|z|^p]\exp[-z(a + b|z|^p)]}{\{\exp[-z(a + b|z|^p)] + 1\}^2} \tag{1}$$

$$F(z) = \{\exp[-z(a + b|z|^p)] + 1\}^{-1} \tag{2}$$

where $z \in \Re$, $a, b, p > 0$.

For the values $a = 1.59413$, $b = 0.07443$ and $p = 1.939$, this distribution approximates very well the normal distribution with a maximum error of 4. 10^{-4} at $z = 0$ for the density function and $7.757.10^{-5}$ at $z = 2.81$ for the distribution function. It is also interesting to note that the above distribution, for certain values of a,b and p, can be bimodal or multimodal.

1.2 GENERATION OF ASYMMETRIC DISTRIBUTIONS

In this sub-section, we mention Azzalini's formula and the corresponding result for skew normal distribution.

1.2.1 AZZALINI'S FORMULA

We can generate many asymmetric distributions $h(z)$ by

$$h(z) = 2g(z)G(w(z)) \tag{3}$$

where $g(z)$ is a symmetric density function about the origin, $G(z)$ is a cumulative distribution function of a symmetric density function about the origin, and $w(z)$ is an odd function.

Introducing a parameter of position $\mu \in \Re$ and a scale parameter $\sigma > 0$, one may generalize (3), with $w(z) = cz$, in the following manner:

$$h(x) = \frac{2}{\sigma} g\left(\frac{z-\mu}{\sigma}\right) G\left(c\frac{z-\mu}{\sigma}\right) \tag{4}$$

1.2.2 SKEW NORMAL

For the normal distribution $N(0,1)$, we can form the skew-normal distribution $(Sn(z))$, with $w(z) = cz$, $c \in \Re$, $z \in \Re$:

$$Sn(z) = 2\phi(z)\Phi(cz) \tag{5}$$

where $\phi(z)$ is the probability density function of the standard normal distribution and $\Phi(z)$ its cumulative distribution function.

Skew normal distribution can not be expressed in simple functions as we can not do the explicit calculation of $\Phi(cz)$. We will use the functions given in (1) and (2) to propose a good approximation for the skew-normal distribution. This approximate density function, can be written in simpler forms than the traditional skew normal density function and its distribution is invertible in some cases.

Using (1), (2) and (3), we define a new skew generalized logistic distribution with four parameters. This distribution is invertible for $c = \pm 1$. For certain values of the parameters a, b and p, and any c, this distribution approximates very well the skew normal distribution. As possible applications, the skew generalized logistic distribution is applied to analyse annual precipitation data of Los Angeles City and the Environmental Performance Index (2010).

2. SKEW GENERALIZED LOGISTIC DISTRIBUTION

In the next sub-section, we define a new skew generalized logistic distribution using (1) and (2) and obtain some properties.

2.1 GENERALIZED LOGISTIC DISTRIBUTION IN AZZALINI'S FORMULA

To define a new skew generalized logistic distribution we let $g(z) = f(z)$ (Eq.(1)), and $G(z) = F(z)$ (2), in (3), thus

$$h(z) = 2f(z)F(cz) = \frac{2[a+b(1+p)|z|^p]\exp[-z(a+b|z|^p)]}{\{\exp[-z(a+b|z|^p)]+1\}^2\{\exp[-cz(a+b|cz|^p)]+1\}} \tag{6}$$

Generalizing the skew generalized logistic distribution with a position parameter $\mu \in \Re$ and a scale parameter $\sigma > 0$, we get

$$Snd(z) = \frac{2[a+b(1+p)|\frac{z-\mu}{\sigma}|^p]}{\sigma\{\exp[-\frac{z-\mu}{\sigma}(a+b|\frac{z-\mu}{\sigma}|^p)]+1\}^2} \times$$

$$\times \frac{\exp[-\frac{z-\mu}{\sigma}(a+b|\frac{z-\mu}{\sigma}|^p)]}{\{\exp[-c\frac{z-\mu}{\sigma}(a+b|c\frac{z-\mu}{\sigma}|^p)]+1\}} \tag{7}$$

For the values of $a = 1.59413$, $b = 0.07443$ and $p = 1.939$ given in section 1.1, the errors of the normal approximation are very small implying very small errors for the skew normal approximation as well. This approach can be very useful from computational point of view as it is written in a compact form. Calculations of quantiles for example can be made explicit in certain cases. Later in this work we will compare this distribution with skew normal distribution.

This distribution is very versatile. In addition to finding a good approximation for the skew normal distribution, for some parameter values, one can produce a bimodal pattern or even heavy tails.

2.2 MOMENTS

Let us now calculate the moments of $h(z)$ given in (6).

$$E(Z^n) = \int_{-\infty}^{\infty} z^n h(z)dz = 2\int_{-\infty}^{\infty} z^n f(z)F(cz)dx$$

$$= 2\int_0^{\infty} z^n f(z)[(-1)^n F(-cz) + F(cz)]dz$$

using the identity $1 - F(-cz) = F(cz)$. We get for even n,

$$E(z^n) = 2\int_0^{\infty} z^n f(z)[F(-cz) + F(cz)]dz = 2\int_0^{\infty} z^n f(z)dz$$

so the even moments are equal to the moments of symmetric distribution $f(z)$. These moments have been calculated by Rathie and Swamee (2006) and are given by:

$$\int_0^{\infty} 2z^n f(z)dz = 2\sum_{r=0}^{\infty}(-1)^r (r+1)[aI_n + b(1+p)I_{n+p}] \tag{8}$$

where

$$I_\beta = [a(1+r)]^{-\beta-1}\sum_{k=0}^{\infty}\left[-\frac{b}{(1+r)^p a^{p+1}}\right]^k \frac{\Gamma[\beta+1+(p+1)k]}{k!}$$

$$= [a(r+1)]^{-\beta-1} H_{1,1}^{1,1}\left[\frac{(r+1)^p a^{p+1}}{b}\middle|\begin{matrix}(1,1)\\(1+\beta, p+1)\end{matrix}\right] \tag{9}$$

where H function is defined in Braaksma (1964) or Mathai et al. (2010), see also Springer (1978) and Luke (1969) for several properties. For odd n, we get

$$E(z^n) = \int_0^{\infty} 2z^n f(z)[2F(cz) - 1]dz$$

$$= \int_0^{\infty} 4z^n f(z)F(cz)dz - \int_0^{\infty} 2z^n f(z)dz, \quad c > 0$$

The result of the second integral is the same as that given in (8), and for the first integral we will consider $c > 0$ and we will use the series,

$$(y+1)^{-1} = \sum_{j=0}^{\infty}(-1)^j y^j \tag{10}$$

and

$$(x+1)^{-2} = \sum_{r=0}^{\infty}(-1)^r (r+1)x^r \tag{11}$$

with $y=\exp[-cz(a+b\,|\,cz\,|^{p})]$ and $x=\exp[-z(a+b\,|\,z\,|^{p})]$, so that

$$\int_0^{\infty} z^n f(z)F(cz)dx = \int_0^{\infty} z^n[a+b(1+p)z^p]\times \sum\nolimits_{r=0}^{\infty}(-1)^r(r+1)$$

$$\times\exp[-z(r+1)(a+bz^p)]\sum\nolimits_{j=0}^{\infty}(-1)^j\exp[-cjz(a+b(|\,c\,|\,z)^p)]dz$$

$$=\sum\nolimits_{r=0}^{\infty}\sum\nolimits_{j=0}^{\infty}(-1)^{r+j}(r+1)\int_0^{\infty} z^n[a+b(1+p)z^p]\times$$

$$\times\exp[-z(a(r+1+cj))-bz^{p+1}(r+1+jc\,|\,c\,|^p)]dz$$

$$=\sum_{r=0}^{\infty}\sum_{j=0}^{\infty}(-1)^{r+j}(r+1)[aJ_n+b(1+p)J_{n+p}] \tag{12}$$

where

$$J_\beta=\int_0^{\infty} z^\beta\exp[-z(a(r+1+cj))]\exp[-bz^{p+1}(r+1+jc\,|\,c\,|^p)]dz$$

$$=\int_0^{\infty} z^\beta\exp[-z(a(r+1+cj))]\sum\nolimits_{k=0}^{\infty}\frac{(-b(r+1+jc\,|\,c\,|^p))^k}{k!}z^{(p+1)k}dz$$

$$=\sum\nolimits_{k=0}^{\infty}\frac{(-b(r+1+jc\,|\,c\,|^p))^k}{k!}\frac{\Gamma(\beta+(p+1)k+1)}{(a(r+1+cj))^{\beta+(p+1)k+1}}$$

$$=[a(r+1+cj)]^{-\beta-1}H_{1,1}^{1,1}\left[\frac{(a(r+1+cj))^{p+1}}{b(r+1+cj\,|\,c\,|^p)}\middle|\begin{matrix}(1,1)\\(1+\beta,p+1)\end{matrix}\right] \tag{13}$$

Thus, for odd n and $c>0$ we have

$$E(z^n)=2\sum_{r=0}^{\infty}(-1)^r(r+1)\left[2\left\{\sum_{j=0}^{\infty}(-1)^j\left[aJ_n+b(1+p)J_{n+p}\right]\right\}-[aI_n+b(1+p)I_{n+p}]\right] \tag{14}$$

To calculate the moments for odd n and $c<0$ note that $-X_c\sim SN(-c)$ if X_c is skew normal distributed with parameter c, so that

$$E(z^n)=\int_{-\infty}^{\infty} z^n h(-z)dz$$

$$E(z^n)=(-1)^n\left[\int_0^{\infty} z^n h(z)dz+\int_0^{\infty}(-1)^n z^n h(-z)dz\right] \tag{15}$$

Thus the *n*th moment of $h(z)$ for $c<0$ and *n* odd is

$$E(z^n) = -2\sum_{r=0}^{\infty}(-1)^r(r+1)\left[2\left\{\sum_{j=0}^{\infty}(-1)^j\left[aJ_n + b(1+p)J_{n+p}\right]\right\} - [aI_n + b(1+p)I_{n+p}]\right] \quad (16)$$

2.3 PARTICULAR CASE WHEN $C = 1$

The particular case when $c=1$ is very interesting because we can explicitly calculate the cumulative distribution function of the skew generalized logistic distribution and its inverse.

Considering $c=1$ in (6), with $z \in \Re$, a, b and p, we get

$$h(z) = 2f(z)F(z) = \frac{2[a+b(1+p)|z|^p]\exp[-z(a+b|z|^p)]}{\{\exp[-z(a+b|z|^p)]+1\}^3} \quad (17)$$

We easily obtain the cumulative distribution by making the substitution $u = \{\exp[-z(a+bz|^p)]+1\}$ and calculating the integral, thus we get

$$H(z) = \{\exp[-z(a+b|z|^p)]+1\}^{-2} \quad (18)$$

Equations (17), and (18) were discussed in Rathie et al.(2008). It was shown that for $a=1.59413$, $b=0.07443$, and $p=1.939$, these equations are good approximations to skew-normal density and distribution functions, for the parameter $c=1$.

2.4 PARTICULAR CASE FOR C=-1

Considering now that $c=-1$ in (5) we can find the cumulative distribution function and its inverse corresponding to generalized logistic distribution. From (6) for $c=-1$, we have

$$h(z) = 2f(z)F(-z) = \frac{2[a+b(1+p)|z|^p]\exp[-2z(a+b|z|^p)]}{\{\exp[-z(a+b|z|^p)]+1\}^3} \quad (19)$$

The cumulative distribution function is easily obtained from (19) and is given by

$$H(z) = \frac{2\{\exp[-z(a+b|z|^p)]+1\}-1}{\{\exp[-z(a+b|z|^p)]+1\}^2} \quad (20)$$

To calculate z as a function of H, we use (20) and obtain

$$z = \frac{1}{a}\ln\left(\frac{H}{1+\sqrt{1-H}-H}\right) - \frac{b}{a} z\,|\,z\,|^{p} \tag{21}$$

Using Lagrange's inversion expansion, we obtain

$$z = \begin{cases} -\sum_{n=0}^{\infty} \dfrac{(-\frac{b}{a})^{n}\,\Gamma(np+n+1)}{n!\,\Gamma(np+2)} \left[\dfrac{1}{a}\ln\left(\dfrac{1+\sqrt{1-H}-H}{H}\right)\right]^{np+1}, H \le 0.75 \\ \sum_{n=0}^{\infty} \dfrac{(-\frac{b}{a})^{n}\,\Gamma(np+n+1)}{n!\,\Gamma(np+2)} \left[\dfrac{1}{a}\ln\left(\dfrac{H}{1+\sqrt{1-H}-H}\right)\right]^{np+1}, H \ge 0.75 \end{cases} \tag{22}$$

or

$$z = \begin{cases} -(\frac{a}{b})^{1/p}\,\frac{1}{p}\,H^{1,1}_{1,2}\left[(\frac{b}{a})^{1/p}\,\frac{1}{a}\ln\left(\frac{1+\sqrt{1-H}-H}{H}\right)\left|\begin{matrix}\left(\frac{p+1}{p},\frac{p+1}{p}\right)\\ \left(\frac{1}{p},\frac{1}{p}\right),(0,1)\end{matrix}\right.\right], H \le 0.75 \\ (\frac{a}{b})^{1/p}\,\frac{1}{p}\,H^{1,1}_{1,2}\left[(\frac{b}{a})^{1/p}\,\frac{1}{a}\ln\left(\frac{H}{1+\sqrt{1-H}-H}\right)\left|\begin{matrix}\left(\frac{p+1}{p},\frac{p+1}{p}\right)\\ \left(\frac{1}{p},\frac{1}{p}\right),(0,1)\end{matrix}\right.\right], H \ge 0.75 \end{cases} \tag{23}$$

3. APPLICATIONS OF THE SKEW GENERALIZED LOGISTIC DISTRIBUTION

3.1 ANNUAL PRECIPITATION DATA

We have used the data of the annual precipitation (rain) in Los Angeles between 1878 and 1993. This data was obtained from the web site *http://www.weather.gov/* . We began our analysis by making the tests for white noise. These tests reveal that in fact this data has white noise, so we need to find its distribution.

This data is very skewed and we adjusted the skew generalized logistic distribution $Sna(z)$ with position and scale parameters (7) to the data. We approximated the parameters using the maximum likelihood method, using two decimal places of precision, by $\hat{a} = 1.57$, $\hat{b} = 0.087$, $\hat{p} = 1.52$, $\hat{c} = 10.13$, $\hat{\mu} = 6.38$ and $\hat{\sigma} = 10.85$. In Figure 1, one can see the fit of the skew generalized logistic distribution for the empirical distribution of the data and for the histogram of the data.

For this approach the mean square error between (7) and the empirical distribution is 0.000589018, the maximum deviation is 0.0547488 and the average absolute deviation between (7) and the empirical distribution is 0.0192187. We used the Kolmogorov-Smirnov goodness of fit test to check the adjustment obtaining a p-value of 0.861308, and then we concluded that we can not reject the hypothesis that this data follows the skew generalized logistic distribution.

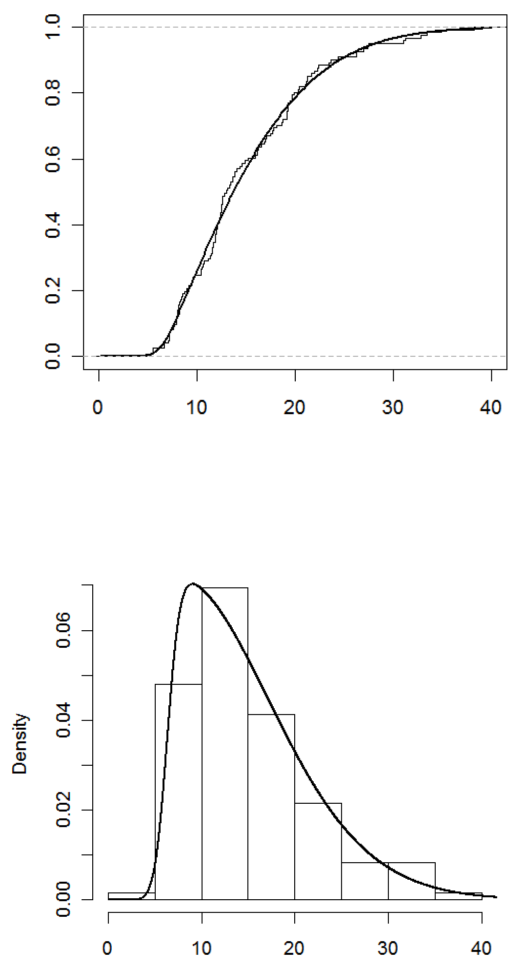


Figure 1: Adjusted skew generalized logistic distribution for the precipitation data: with the empirical distribution (left) and with the histogram (right).

3.2 ENVIRONMENTAL PERFORMANCE INDEX

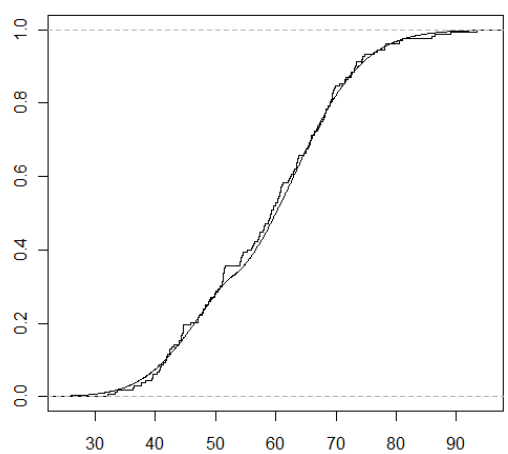
This application shows the versatility of the new distribution. Using the data of the environmental performance index (EPI) we show how the skew generalized logistic distribution (6) and (7) can fit bimodal patterns in the data.

On the basis of 25 performance indicators tracked across ten policy categories covering both environmental public health and ecosystem vitality, the 2010 Environmental Performance Index (EPI) ranked 163 countries. These indicators provide a gauge at a national government scale of how close countries are to established environmental policy goals. This description and the data can be found at *http://epi.yale.edu/*.

The study of the statistical distribution of these data is very important to evaluate how the planet behaves in relation to environmental issues. Also, the study of the evolution of this distribution from year to year is essential to realize if the world is or is not evolving toward a more sustainable future.

The skew generalized logistic distribution $Sna(z)$ (7) proved to be a great option to fit this data because it allows a bimodal pattern and also fit well the skewness of the data.

The estimated parameters, using the maximum likelihood method, are $\hat{a} = 0.312$, $\hat{b} = 0.787$, $\hat{p} = 0.497$, $\hat{c} = 0.656$, $\hat{\mu} = 53.036$ and $\hat{\sigma} = 10.436$. In Figure 2, one can see the fit of the skew generalized logistic distribution for the empirical distribution of the data and for the histogram of the data.



Figures 2 and 3: Adjusted skew generalized logistic distribution for the EPI data: with the empirical distribution.

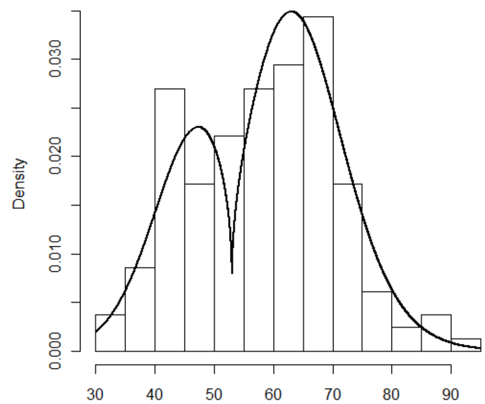


Figure 3: Adjusted skew generalized logistic distribution for the EPI data: with the histogram.

For this approach the mean square error between (7) and the empirical distribution is 0.00034634, the maximum deviation is 0.0452149 and the average absolute deviation between (7) and the empirical distribution is 0.0146137.

We used the Kolmogorov-Smirnov goodness of fit test to check the adjustment obtaining a p-value of 0.89, and then we concluded that we can not reject the hypothesis that this data follows the skew generalized logistic distribution.

4. DISTRIBUTION OF X^2

It is known that W^2, if W is skew-normal distributed, is chi-square distributed with one degree of freedom. We will calculate the distribution of X^2 for X having distribution given by $h(x)$, defined by (17) with parameters $a = 1.59413$, $b = 0.07443$ and $p = 1.939$.

If $Z = X^2$ we have $x = \pm\sqrt{z}$, thus we get

$$g(z) = h(x)\left|\frac{dx}{dz}\right|_{x=\sqrt{z}} + h(x)\left|\frac{dx}{dz}\right|_{x=-\sqrt{z}}$$

$$= \frac{[a + b(p+1)z^{\frac{p}{2}}]}{\sqrt{z}} \frac{\exp[\sqrt{z}(a + bz^{\frac{p}{2}})]}{\{\exp[\sqrt{z}(a + bz^{\frac{p}{2}})] + 1\}^2} \tag{24}$$

This distribution approximates well the chi-square distribution with one degree of freedom. It is easy to find that the accumulative distribution is given by

$$G(z) = \frac{1 - \exp[-\sqrt{z}(a + bz^{\frac{p}{2}})]}{\{\exp[-\sqrt{z}(a + bz^{\frac{p}{2}})] + 1\}} \qquad (25)$$

Compared with chi-square distribution with one degree of freedom, this distribution has a maximum error of approximately 0.000156 at $z = 7.9$, as shown in Figure 3.

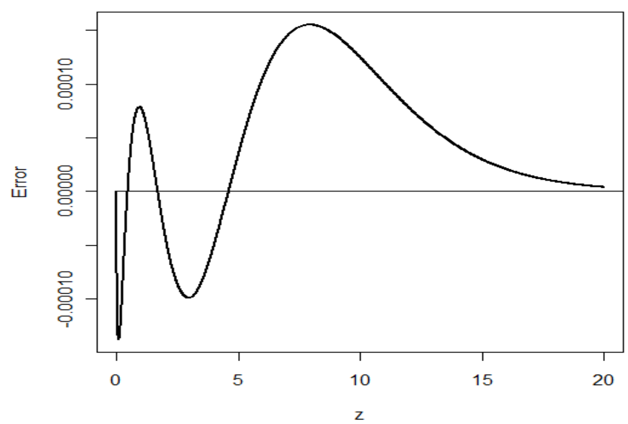


Figure 4: Error between G(z) and the chi-square distribution

Acknowledgement

Thanks are due to the referee for useful suggestions which resulted in the improvement and addition of one more practical application of the distribution introduced in the paper

REFERENCES

Azzalini, A. (1985): A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.

Braaksma, B. L. J. (1964): Asymptotic expansions and analytic continuations for a class of Barnes integrals. *Compositio Math*, **15**, 239-341.

Environmental Performance Index web site *http://epi.yale.edu/*.

O'Hagan, A., and Leonard, T. (1976): Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, **63**, 201-202.

Luke, Yudell L. (1969): *The Special Functions and Their Approximations*. Academic Press, Kansas City.

Mathai, A.M., Saxena, R.K. and Haubold, H.J. (2010): *The H-Function: Theory and Applications*. Springer, New York.

National Oceanic and Atmospheric Administration's National Weather Service web site *http://www.weather.gov/*.

Rathie, P.N., Swamee, P.K, Matos, G.G, Coutinho, M., and Carrijo, T.B. (2008): H-functions and statistical distributions. *Ganita*, **59**, 23-37.

Rathie, P.N. (2011): Normal Distribution, Univariate. *International Encyclopedia of Statistical Sciences*, Springer-Verlag, Berlin, Part 14, 1012-1013.

Rathie, P.N., and Swamee, P.K. (2006): On a new invertible generalized logistic distribution approximation to normal distribution. *Technical Research Report in Statistics*, 07/2006, Department of Statistics, University of Brasilia, Brasilia, Brazil.

Springer, M. D. (1978): *The Algebra of Random Variables*. John Wiley & Sons, New York.

Swamee, P.K., and Rathie, P.N. (2007): Invertible alternatives to normal and lognormal distribution. *Journal of Hydrologic Engineering* ,**12(2)**, 218-221.

Received : 14-08-2010

Revised : 04.05.2011

Pushpa Narayan Rathie and Murilo Coutinho Silva

Department of Statistics, University of Brasília,

70910-900, Brasília, Brazil.

email: pushpanrathie@yahoo.com

email: coutinho.stat@gmail.com