Ann. Inst. Statist. Math.
32 (1980), Part A, 161-169

ON A GENERALIZATION OF THE LOGISTIC DISTRIBUTION

E. OLUSEGUN GEORGE AND M. O. OJO

(Received Sept. 17, 1979; revised Feb. 5, 1980)

Summary

Expressions for the moment generating functions, cumulants and coefficients of kurtosis of a generalization of the logistic distribution are derived and used to show that any symmetric version of this distribution can be closely and simply approximated by a Student t distribution. Related approximation of the distribution of the logistic sample median is discussed.

1. Introduction

The logistic distribution defined by the distribution function

$$F(x)=[1+\exp(-x)]^{-1}, \tag{1.1}$$

$-\infty<x<\infty$, has been used in a variety of statistical studies: Verhulst [11], Pearl and Reed [7] and several other authors used it in the study of population growth; Berkson [1] and Cox [2] employed it in a model for analyzing bioassay and quantal response experiments; Plackett [8] considered its use in problems involving censored data; Gumbel [3], Gumbel and Keeney [4] respectively showed that it is a limiting distribution of the standardized mid-range and the extremal quotient of a sample. Recently Mudholkar and George [6] pointed out that because of its large tails, the logistic distribution standardized so as to have a unit variance, is closer to a similarly standardized Student t distribution with 9 degrees of freedom than it is to a standard normal distribution.

Although various generalizations of this distribution have been proposed, notably by Gumbel and Dubey (see Johnson and Kotz [5] and Prentice [10]), not very much is known about their properties. In this paper some properties of a generalization of the logistic distribution are described. After obtaining expressions for the cumulant of

Some key words: Generalized logistic, Student t distribution, approximation, least squares line, logistic sample median.

the generalized logistic distribution, it is shown, by equating cumulants, that the symmetric version of the distribution is very well approximated by the Student t distribution. Furthermore, it is shown that the parameter of the symmetric generalized logistic distribution and the degree of freedom of the Student t distribution which approximates it, can be very well fitted by a least squares line. These results are used to suggest approximations for the distribution of the logistic sample median.

2. The generalized logistic : Moment generating functions and cumulants

Let X be a beta random variable with parameters p and q (denoted by beta (p, q)) and let

$$Y=\log [X/(1-X)] . \tag{2.1}$$

Y has been called a generalized logistic random variable by Johnson and Kotz [5]. The moment generating function of Y is readily shown to be given by

$$\varphi_Y(\theta)=\Gamma(p+\theta)\Gamma(q-\theta)/\Gamma(p)\Gamma(q) . \tag{2.2}$$

For values of (p, q) equal $(1, 1)$ and (∞, ∞), the distribution of Y can be shown to be respectively logistic and degenerate normal. Moreover it can be shown that Y has the same distribution as $-\log F(2q, 2p)-\log (q/p)$, where $F(2q, 2p)$ is an F statistic with $2q$ and $2p$ degrees of freedom. This last fact can be used to show that $Y+\log (q/p)$ has extreme value minima and extreme value maxima when the values of (p, q) are $(1, \infty)$ and $(\infty, 1)$ respectively. The density function of $Y+\log (p/q)$ was given by Prentice [10].

By using the well-known definition of the gamma function, namely

$$\Gamma(z)=\left[e^{-\gamma z} \prod_{n=1}^{\infty} e^{z/n}\right]\Big/\left[z \prod_{n=1}^{\infty} (1+zn^{-1})\right] , \tag{2.3}$$

where z is any complex number, and γ is the Euler constant, and $\Gamma(\cdot)$ is the gamma function, it can be shown that

$$\Gamma(p+\theta)\Gamma(q-\theta)=\left[e^{-\gamma(p+q)} \prod_{n=1}^{\infty} e^{(p+q)/n}\right]\Big/\left[\xi(\theta) \prod_{n=1}^{\infty} \Psi(\theta, n)\right] , \tag{2.4}$$

where

$$\xi(\theta)=pq+(q-p)\theta-\theta^2 , \tag{2.5}$$

and

(2.6) $$\Psi(\theta, n)=1+(p+q)n^{-1}+\xi(\theta)n^{-2}\,.$$

Hence

(2.7) $$\varphi_Y(\theta)=\{\xi(0)/\xi(\theta)\}\prod_{n=1}^{\infty}\{\Psi(0, n)/\Psi(\theta, n)\}\,.$$

But

$$\begin{aligned}\prod_{n=1}^{\infty}\Psi(\theta, n)&=\prod_{n=1}^{\infty}\left[\frac{(n+p)(n+q)}{n^2}-\frac{\theta^2-(q-p)\theta}{n^2}\right]\\&=\prod_{n=1}^{\infty}\left[\frac{(n+p)(n+q)}{n^2}\right]\prod_{n=1}^{\infty}\left[1-\frac{\theta^2-(q-p)\theta}{(n+p)(n+q)}\right]\\&=\prod_{n=1}^{\infty}\Psi(0, n)\prod_{n=1}^{\infty}\left[1-\frac{\theta^2-(q-p)\theta}{(n+p)(n+q)}\right],\end{aligned}$$

therefore

$$\varphi_Y(\theta)=\{\xi(0)/\xi(\theta)\}\prod_{n=1}^{\infty}\left[1-\frac{\theta^2-(q-p)\theta}{(n+p)(n+q)}\right]^{-1}.$$

Simplifying further, it is easily shown that

(2.8) $$\varphi_Y(\theta)=\prod_{n=0}^{\infty}\left[1-\frac{\theta^2-(q-p)\theta}{(n+p)(n+q)}\right]^{-1}.$$

Consequently the cumulant generating function is given by

(2.9) $$\begin{aligned}\log\varphi_Y(\theta)=&-\sum_{n=0}^{\infty}\log\left[(n+p)(n+q)-\{\theta^2-(q-p)\theta\}\right]\\&+\sum_{n=0}^{\infty}\log(n+p)(n+q)\,,\end{aligned}$$

and the rth cumulant by

(2.10) $$\kappa_r(Y)=-\sum_{n=0}^{\infty}\frac{d^r}{d\theta^r}\log\left[(n+p)(n+q)-\{\theta^2-(q-p)\theta\}\right]\Big|_{\theta=0}\,.$$

In particular, the first four cumulants are given respectively by

(2.11) $$\kappa_1(Y)=(p-q)\sum_{n=0}^{\infty}[(n+p)(n+q)]^{-1}$$

(2.12) $$\kappa_2(Y)=\sum_{n=0}^{\infty}2[(n+p)(n+q)]^{-1}+\sum_{n=0}^{\infty}(p-q)^2[(n+p)(n+q)]^{-2}$$

(2.13) $$\kappa_3(Y)=\sum_{n=0}^{\infty}6(p-q)[(n+p)(n+q)]^{-2}+\sum_{n=0}^{\infty}2(p-q)[(n+p)(n+q)]^{-3}$$

and

(2.14) $$\kappa_4(Y)=\sum_{n=0}^{\infty}[12/\{(n+p)(n+q)\}^2+24(p-q)^2/\{(n+p)(n+q)\}^3 +6(p-q)^4/\{(n+p)(n+q)\}^4] .$$

In the special case when $p=q=\lambda$, corresponding to the symmetric generalized logistic

(2.15) $$\varphi_Y(\theta)=\prod_{n=0}^{\infty}[1-\theta^2/(n+\lambda)^2]^{-1} .$$

Hence for any positive integer r, the cumulant

(2.16) $$\begin{aligned}\kappa_r(Y)&=-\sum_{n=0}^{\infty}\frac{d^r}{d\theta^r}\log[1-\theta^2/(n+\lambda)^2]|_{\theta=0}\\ &=\sum_{n=0}^{\infty}\sum_{k=1}^{\infty}\frac{1}{k}\left[\frac{d^r}{d\theta^r}\left(\frac{\theta}{n+\lambda}\right)^{2k}\right]\Big|_{\theta=0}\\ &=\begin{cases}(r-1)!\sum_{n=0}^{\infty}(n+\lambda)^{-r}, & \text{if } r \text{ is even},\\ 0, & \text{if } r \text{ is odd}.\end{cases}\end{aligned}$$

When $\lambda=m$ or $m+1/2$, m a positive integer, these cumulants have forms that are easy to compute: Let $\lambda=m+1/2$, and let r be a positive even integer. It can be shown that

$$\kappa_r(Y)=2^{r+1}(r-1)!\sum_{n=m}^{\infty}(2n+1)^{-r} .$$

But

$$\begin{aligned}\sum_{n=m}^{\infty}(2n+1)^{-r}&=\sum_{n=0}^{\infty}(2n+1)^{-r}-\sum_{n=0}^{m-1}(2n+1)^{-r}\\ &=\sum_{n=1}^{\infty}n^{-r}-\sum_{n=1}^{\infty}(2n)^{-r}-\sum_{n=0}^{m-1}(2n+1)^{-r}\\ &=2^{-r}(2^r-1)\zeta(r)-\sum_{n=0}^{m-1}(2n+1)^{-r},\end{aligned}$$

where $\zeta(r)=\sum_{n=1}^{\infty}n^{-r}$, is the zeta function. Hence

(2.17) $$\kappa_r(Y)=2(2^r-1)(r-1)!\zeta(r)-2^{r+1}(r-1)!\sum_{n=0}^{m-1}(2n+1)^{-r} .$$

In particular,

(2.18) $$\kappa_2(Y)=\pi^2-8\sum_{n=0}^{m-1}(2n+1)^{-2}$$

and

(2.19) $$\kappa_4(Y)=2\pi^4-192\sum_{n=0}^{m-1}(2n+1)^{-4} .$$

By a similar procedure, it can be shown that when $\lambda=m$,

$$\kappa_r(Y)=2[(r-1)!]\left[\zeta(r)-\sum_{n=1}^{m-1} 2n^{-r}\right] . \tag{2.20}$$

When $r=2$ and 4,

$$\kappa_2(Y)=\pi^2/3-\sum_{n=1}^{m-1} 2n^{-2} \tag{2.21}$$

and

$$\kappa_4(Y)=2\pi^4/15-\sum_{n=1}^{m-1} 12n^{-4} . \tag{2.22}$$

This latter case is of some interest: When $m\geqq 2$, it can be shown that the moment generating function $\varphi_Y(\theta)$ can be expressed as

$$\varphi_Y(\theta)=\prod_{n=1}^{m-1} (1-\theta^2/n^2)(\pi\theta/\sin \pi\theta) , \tag{2.23}$$

where $(\pi\theta/\sin \pi\theta)$ is the moment generating function of the logistic distribution function given by equation (1.1). It can be shown that $\varphi_Y(\theta)$ is the moment generating function of the median of $(2m-1)$ independent logistic variates.

3. Student t-approximation

The similarities in the shapes of the logistic and the normal distributions have been noted by several authors. An excellent summary of this observation is found in Johnson and Kotz [5]. Recently, however, Mudholkar and George [6] showed that the Student t distribution function with 9 degrees of freedom, when standardized so as to have variance one provides a better fit of a standardized logistic distribution than the standard normal. By equating coefficients of kurtosis we now show that this property extends to essentially all members of the family of symmetric generalized logistic.

The coefficient of kurtosis of a symmetric generalized logistic with parameter $\lambda\geqq 1$ is given, using equations (2.19) and (2.20), by

$$\beta_2(Y)=\begin{cases} 1\cdot 2\left[\pi^4-90\sum_{n=1}^{m-1} n^{-4}\right]\Big/\left[\pi^2-6\sum_{n=1}^{m-1} n^{-2}\right]^2, & \text{if } \lambda=m \\ 2\left[\pi^4-96\sum_{n=0}^{m-1} (2n+1)^{-4}\right]\Big/\left[\pi^2-8\sum_{n=0}^{m-1} (2n+1)^{-2}\right]^2, & \text{if } \lambda=m+1/2 . \end{cases} \tag{3.1}$$

Equating $\beta_2(Y)$ to $6/(\nu-4)$ the coefficient of kurtosis of a Student t with ν degrees of freedom, we get

$$\nu=\begin{cases} 4+5\left[\pi^2-6\sum_{n=1}^{m-1} n^{-2}\right]^2 \Big/ \left[\pi^4-90\sum_{n=1}^{m-1} n^{-4}\right] & \text{if } \lambda=m\,, \\ 4+3\left[\pi^2-8\sum_{n=0}^{m-1} (2n+1)^{-2}\right]^2 \Big/ \left[\pi^4-96\sum_{n=0}^{m-1} (2n+1)^{-4}\right] & \\ & \text{if } \lambda=m+1/2\,. \end{cases} \tag{3.2}$$

Let $[\nu]$ be the integer closest to ν, and let

$$k=[\nu]\,. \tag{3.3}$$

Also let

$$Y^*=Y/\{(\kappa_2(Y)\}^{1/2} \tag{3.4}$$

and

$$t_k^*=t_k/[k/(k-2)]^{1/2} \tag{3.5}$$

be the standardized symmetric generalized logistic variate with parameter λ, and the standardized Student t variate with k degrees of freedom respectively. The approximation proposed here is given for $-\infty<y<\infty$, by

$$\mathrm{P}\,\{Y^*<y\}\doteqdot\mathrm{P}\,\{t_k^*<y\} \tag{3.6}$$

or equivalently by

$$\mathrm{P}\,\{Y<y\}\doteqdot\mathrm{P}\,\{t_k<cy\}\,, \tag{3.7}$$

where

$$c=[k/\{(k-2)\kappa_2(Y)\}]^{1/2}\,, \tag{3.8}$$

$\kappa_2(Y)$ being given by equations (2.18) and (2.21) when $\lambda=m+1/2$ and m respectively.

4. Least squares fit of the parameters

The tedious computations in the calculation of $[\nu]$ by using equation (3.2) can almost be totally eliminated because of the following rather interesting observation. Using values $\lambda=1$, 1.5, 2, 2.5, 3, 3.5 and 4 and the corresponding values $[\nu]=9$, 12, 14, 17, 20, 23 and 25, a least squares line of $[\nu]$ in terms of λ is found to be

$$\hat{\nu}=3.25+5.5\lambda\,. \tag{4.1}$$

The goodness of this fit is illustrated below in Table I. It is clear from this table that the values of k computed by using the two equations are essentially the same. Moreover, the two equations seem to

Table I. Values of k computed by equation (3.2) and (4.1)

λ	1	1.5	2	2.5	3	3.5	4
$[\nu]$	9	12	14	17	20	23	25
$[\hat{\nu}]$	9	11	14	17	19	23	25
$[\hat{\nu}]-[\nu]$	0	1	0	0	1	0	0

$[\nu]$=Value of k computed using equation (3.2)
$[\hat{\nu}]$=Value of k computed using equation (4.1)

give the same value of k for all λ: For example when $\lambda=4.5$ and 5.0, both equations give values $k=28$ and 31.

5. Illustration of proposed approximation

Table II below illustrates the approximation given by equation (3.6). The table shows that the symmetric generalized logistic distribution is very well approximated by the Student t distribution. This approximation improves as λ increases. For $\lambda=1$, the maximum value of $|F_*(x)-T_*(x)|$ is less than 5×10^{-2}. This bound decreases rather rapidly, becoming less than 10^{-4} when $\lambda=4$.

6. Student t and normal approximations of the median of logistic variates

A consequence of the proposed approximation is that it gives small sample Student t and normal approximations for the median of logistic variates. Let $X_1, X_2, \cdots, X_{2n-1}$ be independent logistic random variables, each with a distribution function given by equation (1.1). Let $X_{r,2n-1}$ be the rth ordered statistics among the X's and let $U_1, U_2, \cdots, U_{2n-1}$

Table II. Student t approximation for the symmetric generalized logistic distribution

x	$\lambda=1$, $k=9$ (Logistic)		$\lambda=1.5$, $k=11$		$\lambda=2$, $k=14$	
	$F_*(x)$	$F_*(x)-T_*(x)$	$F_*(x)$	$F_*(x)-T_*(x)$	$F_*(x)$	$F_*(x)-T_*(x)$
0.10	0.54522	0.00132	0.54342	0.00040	0.54250	0.00026
0.20	0.58970	0.00253	0.58624	0.00077	0.58445	0.00049
0.30	0.63277	0.00354	0.62788	0.00107	0.62534	0.00069
0.50	0.71236	0.00467	0.70567	0.00141	0.70211	0.00094
0.90	0.83650	0.00356	0.83044	0.00100	0.82700	0.00074
1.25	0.90613	0.00115	0.90295	0.00158	0.90102	0.00021
1.75	0.95985	−0.00894	0.95992	−0.00499	0.95992	−0.00027
2.5	0.98938	−0.00083	0.99050	−0.00029	0.99118	−0.00019
3.0	0.99568	−0.00038	0.99649	−0.00007	0.99698	−0.00005
4.0	0.99929	0.00001	0.99954	0.00005	0.99967	0.00002

Table II. (Continued)

x	$\lambda=3,\ k=20$		$\lambda=3.5,\ k=23$		$\lambda=4,\ k=25$	
	$F_*(x)$	$F_*(x)-T_*(x)$	$F_*(x)$	$F_*(x)-T_*(x)$	$F_*(x)$	$F_*(x)-T_*(x)$
0.10	0.54158	0.00013	0.54131	0.00009	0.541124	0.00002
0.20	0.58267	0.00025	0.58216	0.00019	0.58178	0.00004
0.30	0.62279	0.00035	0.62207	0.00026	0.62153	0.00006
0.50	0.69850	0.00048	0.69747	0.00037	0.69670	0.00008
0.90	0.82336	0.00042	0.82230	0.00032	0.82150	0.00003
1.25	0.89890	0.00016	0.89826	0.00013	0.89778	−0.00003
1.75	0.95991	−0.00011	0.95990	−0.00008	0.95990	−0.00006
2.5	0.99196	−0.00010	0.99220	−0.00008	0.99239	−0.00001
3.0	0.99751	−0.00003	0.99767	−0.00002	0.99779	0.00001
4.0	0.99979	0.00001	0.99982	0.00000	0.99985	0.00009

λ=Parameter of the symmetric generalized logistic.
k=Degree of freedom of Student t.
$F_*(x)$=Distribution function of standardized symmetric generalized logistic.
$T_*(x)$=Distribution function of standardized Student t.

be independently and identically distributed uniformly on the interval (0, 1). It is well known that for each i, $1\leqq i\leqq 2n-1$, X_i has the same distribution as $\log\{U_i/(1-U_i)\}$. Consequently, since $\log\{x/(1-x)\}$ is monotonically increasing in x, the median $X_{n,2n-1}$ of the X's is equivalent in law to $\log\{U_{n,2n-1}/(1-U_{n,2n-1})\}$. But $U_{n,2n-1}$ is a beta variate with parameters both equal to n. Therefore $X_{n,2n-1}$ has the same distribution as $\log[\text{beta}(n, n)/\{1-\text{beta}(n, n)\}]$, the symmetric generalized logistic variate with parameter n.

An immediate consequence of this observation is that the distribution of a standardized logistic sample median may be approximated by a standardized Student t distribution. Table II shows that this approximation is quite good even when the sample size $2n-1$ is as small as 7, a value which corresponds to a symmetric generalized logistic with parameter $n=4$. Furthermore for small sample sizes, such as 3 and 4, corresponding to $n=2$ and 2.5 respectively, the approximation could be used.

Another consequence of the above observation is that it suggests a normal approximation for the distribution of the logistic sample median for moderate sample sizes. For example, for a sample size 11, the approximating Student t distribution which has 36 degrees of freedom is known to be close to the normal distribution. For sample sizes of 21 or more, corresponding to Student t distributions with more than 55 degrees of freedom, the normal approximation should be quite good.

UNIVERSITY OF IFE
OYO STATE COLLEGE OF ARTS AND SCIENCES

References

[1] Berkson, J. (1944). Application of the logistic function to bio-assay, *J. Amer. Statist. Ass.*, **39**, 357-365.

[2] Cox, D. R. (1970). *Analysis of Binary Data*, Methuen, London.

[3] Gumbel, E. J. (1944). Ranges and midranges, *Ann. Math. Statist.*, **15**, 414-422.

[4] Gumbel, E. J. and Keeney, R. E. (1950). The extremal quotient, *Ann. Math. Statist.*, **21**, 523-538.

[5] Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. II, Houghton Mifflin Company, Boston.

[6] Mudholkar, G. S. and George, E. O. (1978). A remark on the shape of the logistic distribution, *Biometrika*, **65**, 667-668.

[7] Pearl, R. and Reed, L. J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation, *Proc. Nat. Acad. Sci. USA*, **6**, 275-288.

[8] Plackett, R. L. (1958). Linear estimation from censored data, *Ann. Math. Statist.*, **29**, 131-142.

[9] Plackett, R. L. (1959). The analysis of life test data, *Technometrics*, **1**, 9-19.

[10] Prentice, R. L. (1975). Discrimination among some parameter models, *Biometrika*, **62**, 607-614.

[11] Verhulst, P. F. (1845). Recherches mathematiques sur la loi d'accroisement de la population, *Nouveaux memoirs de l'ac. Bruxelles*, **18**, 1-38.