Kragujevac J. Math. 24 (2002) 135–145.

APPROXIMATIONS OF THE DISTRIBUTION OF THE SUM OF RANDOM VARIABLES FROM THE GENERALIZED LOGISTIC DISTRIBUTION

M. O. Ojo

Department of Mathematics, Obafemi Awolowo University, Ile-Ife Nigeria.

(Received February 26, 2001)

Abstract. In this paper, mixture approximations are proposed for the distribution function of the sum of the generalized logistic random variables. These approximations are quite comparable with the existing t-approximation and are indeed simpler to evaluate. Moreover, an added advantage of the new approximations is that statistical tables may not be required for their implementation.

1. INTRODUCTION

The use of the logistic distribution for modelling stochastic phenomena has for long been recognized and its uses in various statistical studies have been documented by several authors. A Generalization of this distribution, whose density function is given as

$$f(x) \;=\; \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\frac{e^{px}}{(1+e^x)^{p+q}}, \quad -\infty < x < \infty \tag{1.1}$$

has earlier on appeared in George and Ojo (1980) where the moments were obtained and an approximation proposed for its distribution function. The distribution of the sum of independent and identically distributed random variables from the same distribution has been determined by Ojo and Adeyemi (1989). It is observed that a lot of computational efforts is required to evaluate this distribution function. Since the t-distribution is not extensively tabulated, the proposed t-approximation cannot be adequate for many purposes. In view of this, the present author has decided to provide various alternative approximations to the distribution of the sum. The approximations which are provided for probabilities and percentiles are comparable in accuracy to thet-approximation and are much simpler to evaluate. More importantly, the use of statistical table may not be necessary for implementation.

2. APPROXIMATION USING THE NORMAL, THE LOGISTIC AND THE DOUBLE EXPONENTIAL DISTRIBUTIONS

Let $x_1, x_2, ..., x_n$ be n independent and identically distributed random variables each distributed according to (1.1). Let $Y = \sum_{i=1}^{n} X_i$. The distribution function of Y has been determined and a t-approximation provided for it. Because the t-distribution is moderately tabulated, the proposed t-approximation may not be adequate. In this section we present appromations based on the normal, the logistic, the double exponential distributions and the mixture of these distributions. For the purpose of this paper, we recall from Ojo and Adeyemi (1989) distribution of Y as

$$\begin{aligned} F_n(y) = \sum_{l=0}^{n} \sum_{r=0}^{n-m-1} \sum_{m=0}^{n-1} \binom{n}{l} \prod_{j=1}^{p+q-1} A_{n,m} \\ \times(-1)^{r+l+m+1}[\Gamma(p)\cdot\Gamma(q)]^{-n} e^{py} \\ \times l![(n-m-r-1)!(l+r-n+m+1)!]^{-1} \\ \times \sum_{s=0}^{r} \frac{(-1)^s y^{r-s}}{(p-\kappa)^{s+1}(r-s)!} \sum_{\kappa=1}^{l+r+m+2-n} S_{l+r+m+2,\kappa}(\kappa-1)! e^{-\kappa y}/(1-e^{-y})^{\kappa} \end{aligned} \tag{2.1}$$

when n is even and

$$F_n(y) = \sum_{l=0}^{n} \sum_{r=0}^{n-m-1} \sum_{m=0}^{n-1} \binom{n}{l} \sum_{j=1}^{p+q-1} A_{n,m}$$

$$\times(-1)^{r+l+m+1}[\Gamma(p)\cdot\Gamma(q)]^{-n}e^{py}$$

$$\times l![(n-m-r-1)!(l+r-n+m+1)!]^{-1}$$

$$\times\sum_{s=0}^{r}\frac{(-1)^s y^{r-s}}{(p-\kappa)^{s+1}(r-s)!}\sum_{\kappa=1}^{l+r+m+2-n}S_{l+r+m+2,\kappa}(\kappa-1)!e^{-\kappa y}/(1+e^{-y})^{\kappa}$$

when n is odd, $(S_{n,k})$ denoting the stirling's number of the second kind. Also for the subsequent use in this paper, we write down the first four cumulants of Y as

$$\kappa_1 = n\left[\sum_{j=1}^{p-1}j^{-1}-\sum_{j=1}^{q-1}j^{-1}\right] \tag{2.2}$$

$$\kappa_2 = \frac{n\pi^2}{3} - n\left[\sum_{j=1}^{p-1}j^{-2}+\sum_{j=1}^{q-1}j^{-2}\right] \tag{2.3}$$

$$\kappa_3 = 2n\left[\sum_{j=1}^{p-1}j^{-3}-\sum_{j=1}^{q-1}j^{-3}\right] \tag{2.4}$$

$$\kappa_4 = \frac{2n\pi^4}{15} - 6n\left[\sum_{j=1}^{p-1}j^{-4}+\sum_{j=1}^{q-1}j^{-4}\right] \tag{2.5}$$

where p and q are positive integers.

2.1 THE NORMAL APPROXIMATION

Let Y^* denote the standardized version of Y. Then

$$Pr(Y^*\leq y) = Pr(Y\leq \mu+\sigma y)$$

Thus the proposed normal approximation is given as

$$Pr(Y\leq y)\sim \Phi(\frac{y-\mu}{\sigma}),$$

where $\Phi(.)$ is the cumulative standard normal, $\mu=\kappa_1$ and $\sigma=\sqrt{\kappa_2}$, κ_i being the i^{th} cumulant of Y.

2.2 THE LOGISTIC APPROXIMATION

The distribution function $L(x)$ and the upper $\alpha-$percentile $X_L(\alpha)$ of the logistic distribution with standard deviation σ are given respectively as

$$L(x) = [1 + exp - (\pi x/\sigma\sqrt{3})]^{-1}$$

and

$$X_L(\alpha) = \frac{\sigma\sqrt{3}}{\pi} log[(1-\alpha)/\alpha]$$

The proposed logistic approximation is given as

$$P(Y \leq y) \sim L(x - \mu).$$

2.3 APPROXIMATION USING THE DOUBLE EXPONENTIAL DISTRIBUTION

The distribution function and the upper $\alpha-$percentile of the double exponential distribution function are given as

$$D(x) = 1 - \frac{1}{2} exp - (x/\sigma\sqrt{2})$$

and

$$X_D(\alpha) = (-\frac{1}{2}) log 2\alpha.$$

The approximation based on the duoble exponential is given as

$$P(Y \leq y) \sim D(x - \mu).$$

2.4 APPROXIMATION BASED ON MIXTURE OF LOGISTIC AND NORMAL

Let $\lambda_1 L(\mu, \sigma) + \bar{\lambda}_1 N(\mu, \sigma)$ denote a mixture of the Logistic and normal distributions with a common standard deviation of Y given as $\sigma = \sqrt{\kappa_2}$ and with the mixing probabilities λ_1 and $\bar{\lambda}_1$ where $\bar{\lambda}_1 = 1 - \lambda_1$. Clearly thestandard deviation of the mixture is also σ. Equating the coefficient of kurtosis of the mixture to that of Y, we get $2\lambda_1 + 3\bar{\lambda}_1 = \beta_2$, where $\beta_2 = \frac{\kappa_4}{\kappa_2^2}$, κ_i being the i^{th} cumulant of Y.

Since $\bar{\lambda}_1 = 1 - \lambda_1$ we obtain

$$\lambda_1 = \frac{5}{6}(\beta_2 - 3)\text{and}\bar{\lambda}_1 = \frac{1}{6}(21 - 5\beta_2)$$

Hence the distribution function of Y may be approximated by the mixture of the logistic and normal as

$$Pr(Y \leq y) \sim \lambda_1 L(x - \mu) + \bar{\lambda}_1 \Phi(\frac{x - \mu}{\sigma}).$$

2.5 APPROXIMATION BASED ON THE MIXTURE OF NORMAL AND DOUBLE EXPONENTIAL

We can also mix the normal with the double exponential to produce another approximation. By matching kurtosis we have

$$3\lambda_2 + 6\bar{\lambda}_2 = \beta_2.$$

so that $\lambda_2 = \frac{1}{3}(6 - \beta_2)$, and $\bar{\lambda}_2 = \frac{1}{3}(\beta_2 - 3)$ The corresponding approximation is given as

$$Pr(Y \leq y) \sim \lambda_2 \Phi(\frac{x - \mu}{\sigma}) + \bar{\lambda}_2 D(x - \mu).$$

2.6 APPROXIMATION BASED ON THE MIXTURE OF LOGISTIC AND DOUBLE EXPONENTIAL

By matching kurtosis, we also get

$$2\lambda_3 + 6\bar{\lambda}_3 = \beta_2.$$

Solving gives $\lambda_3 = \frac{5}{9}(6 - \beta_2)$, and $\bar{\lambda}_3 = \frac{1}{9}(5\beta_2 - 21)$. The resulting mixture approximation is given as

$$Pr(Y \leq y) \sim \lambda_3 L(x - \mu) + \bar{\lambda}_3 D(x - \mu).$$

Note that all the distribution used have the same standard deviation σ which is that of Y.

3. PERCENTILE APPROXIMATION USING MIXTURES

The Percentage point of the distribution of Y can be approximated by the weighted average of the percentile of one distribution with the percentile of another where the weights are the same as the mixing probabilities of the mixtures in the previous section. Thus, if $X_N(\alpha)$, $X_L(\alpha)$ and $X_D(\alpha)$ denote respectively the upper $\alpha-$percentiles of the normal, the logistic and the double exponential distributions, we have the following approximations of percentiles of Y using mixtures as

(i)The normal and logistic $\lambda_1 X_N(\alpha) + \bar{\lambda}_1 X_L(\alpha)$.

(ii)The normal and double exponential $\lambda_2 X_N(\alpha) + \bar{\lambda}_2 X_D(\alpha)$.

(iii)The Logistic and the double exponential $\lambda_3 X_L(\alpha) + \bar{\lambda}_3 X_D(\alpha)$.

4. DISCUSSION

Tables I and II provide illustrations for the approximations. Table I gives the errors of the various approximations to the probabilities while table II compares approximations to the percentiles. The values of $F(z)$ and the associated errors of the t-approximaton in table I, are taken from Ojo and Adeyemi(1989). It Is clear from table I that the t-approximation is superior to the approximations based on either the normal, the logistic or the double exponential disribution and that the approximations based on the mixtures compare favourably well with t-approximation. Also in table II, it is obsevered that the mixture approximations to the percentiles are superior to the approximations based on single distributions. It may be oberserved that the use of one of the approximations for $\Phi(.)$ listed in Section 5, Chapter 13 of Johnson and Kotz(1970)will set all the approximations involving $\Phi(.)$ free from necessity of any tables. In general, any of the newly proposed approximations can be a good subtitute for the t-approximation in the sense that statistical table may not be necessary for implementation. We observe that no table is provided for the approximation of the distribution of Y for the case $p \neq q$ corresponding to the non-symmetric version of the distribution. It has been discovered that the approximations corresponding to $p \neq q$ are not as good as the approximation corresponding to the symmetric version of the distribution of Y. This is to be expected since the normal, the logistic, the double exponential distributions and the distribution of Y for $p = q$ are all unimodal and symmeric about the origin.

Acknowledgements: Research Supported by the University Research Council, Obafemi Awolowo University, Ile-Ife, Nigeria.

Table I: Errors (10^{-4}) of approximation for the probabilities

x	F(z)	t	N	L	DE	M_1	M_2	M_3
	p=q=1	n=2						
0.05	0.5238	3	14	11	17	5	6	15
0.25	0.6298	7	11	9	15	9	11	12
0.75	0.8263	8	9	7	21	10	9	10
1.00	0.8914	5	7	4	9	6	12	7
1.45	0.9585	1	5	13	11	3	8	8
2.00	0.9894	3	6	5	8	3	7	9
2.50	0.9972	4	10	7	7	5	6	13
3.00	0.9995	3	4	4	6	4	5	7
3.50	0.9996	2	3	2	5	3	4	6
4.00	0.9997	2	2	2	4	2	3	4
x	p=q=2$n = 3$							
0.05	0.5193	3	12	10	11	9	12	17
0.25	0.6000	2	8	8	12	10	11	13
0.75	0.8102	7	9	1	9	7	6	9
1.00	0.8452	4	5	4	3	3	5	7
1.45	0.9283	6	7	6	2	2	3	8
2.00	0.9788	5	6	5	7	4	4	6
2.50	0.9934	3	5	4	4	4	6	3
3.00	0.9967	1	3	3	3	3	4	4
3.50	0.9988	2	6	1	4	2	3	4
4.00	0.9998	0	5	4	3	1	2	3

x	F(z)	t	N	L	DE	M_1	M_2	M_3
x	p=q=2	n=5						
0.05	0.5186	1.6	9	7.0	11	5.9	6.5	10
0.25	0.5879	3.9	4.5	3.5	10	7.1	5.9	6
0.75	0.7956	1.3	2.4	2.9	12	8.5	9.2	11
1.00	0.8409	1.7	2.1	1.8	9	7.0	8.6	9.6
1.45	0.9167	1.9	3.0	3.0	6	6.8	4.6	7.7
2.00	0.9693	0.5	2.0	1.5	7	7.5	3.9	4.5
2.50	0.9929	0.4	1.5	1.0	5	4.0	4.0	4.5
3.00	0.9971	0.2	1.1	1.0	4	2.5	3.7	5.2
3.50	0.9979	0.1	1.0	1.2	3.5	3.0	4.5	5.6
4.00	0.9997	0.1	1.0	1.0	2.0	2.0	2.5	3.8
x	p=q=1	n=10						
0.05	0.5199	0.5	1.4	1.3	4.6	3	4	5
0.25	0.5982	0.4	2.0	2.0	4.0	2	1.5	2.7
0.75	0.7816	0.3	3.3	2.4	5.0	1	2.7	3.6
1.00	0.8437	0.5	1.8	1.7	3.0	3	3.8	2.9
1.45	0.9203	0.3	1.6	1.9	2.0	1	1.9	2.1
2.00	0.9787	0.2	1.2	1	2.0	2	3.1	2.4
2.50	0.9931	0.1	1.0	1	1.0	1.2	2.5	3.2
3.00	0.9983	0	1.0	0	1.0	1	2.1	3.6
3.50	0.9996	0	1.4	1.2	2.0	1	1.9	2.4
4.00	0.9999	0.9	1.2	1.0	2.0	2	2.3	2.4

t: student t-distribution, $M_1 : \lambda_1 N + \bar{\lambda}_1 L$
N: normal, $M_2 : \lambda_2 L + \bar{\lambda}_2 DE$
L: logistic $M_3 : \lambda_3 N + \bar{\lambda}_3 DE$
DE: double exponential, $Z = \frac{x-\mu}{\sigma}$, $\mu = \kappa_1$, $\sigma = \sqrt{\kappa_2}$

Table II: Comparison of approximation for the percentiles.

$1-\alpha$ (i)	$Exact$ (ii)	N (iii)	L (iv)	DE (v)	$\lambda_1 N+\bar{\lambda}_1 L$ (vi)	$\lambda_2 L+\bar{\lambda}_2 DE$ (vii)	$\lambda_3 N+\bar{\lambda}_3 DE$ $(viii)$
p=q=1, n=2							
.60	0.535	0.400	0.376	0.286	0.378	0.368	0.341
.90	3.272	3.228	3.106	2.064	3.298	2.145	2.321
.95	4.366	4.221	4.162	3.953	4.276	4.168	4.415
.975	5.112	5.025	5.179	4.843	5.102	4.989	4.717
.995	6.987	6.620	7.483	6.906	7.052	6.965	7.112
p=q=2.n=3							
.60	0.487	0.552	0.440	0.420	0.479	0.467	0.492
.90	2.487	2.530	2.384	2.585	2.513	2.479	2.677
.95	3.362	3.239	3.195	3.267	3.300	3.412	3.613
.975	4.014	3.839	3.976	3.950	4.129	4.226	5.112
.995	5.446	5.080	5.744	4.534	5.281	4.976	

$1-\alpha$ (i)	*Exact* (ii)	N (iii)	L (iv)	DE (v)	$\lambda_1 N+\bar{\lambda}_1 L$ (vi)	$\lambda_2 L+\bar{\lambda}_2 DE$ (vii)	$\lambda_3 N+\bar{\lambda}_3 DE$ (viii)
p=q=2, n=5							
.60	0.610	0.674	0.568	0.584	0.654	0.671	0.593
.90	3.193	3.267	3.098	3.046	3.179	3.018	3.321
.95	4.561	4.182	4.521	3.927	4.611	4.720	4.102
.975	4.987	4.983	5.133	4.808	5.021	5.234	5.611
.995	7.103	6.559	7.416	6.854	7.232	6.897	6.991
p=q=1, n=10							
.60	1.462	1.521	1.262	1.640	1.511	1.498	1.553
.90	6.887	7.374	9.946	6.618	6.593	6.990	7.422
.95	9.576	9.411	9.311	8.606	9.473	10.021	10.323
.975	10.892	11.347	11.588	10.595	11.001	10.894	11.214
.995	14.697	14.804	14.739	14.213	14.578	14.237	15.102

N: normal

L: logistic

DE: double exponential

All distributions have standard deviation $\sigma = \sqrt{n\kappa_2}$, where κ_2 is the second cumulant of X

References

[1] E.O. George, M.O. Ojo, *On a generalization of the logistic distribution,* Annals of the Institute of Statistical Mathematics, Vol **32**, no. 2A, (1980), 161-169.

[2] N.L. Johnson, S. Kotz, *Distributions in statistics: Continous Univariate Distributions 2,* Houghton Milliu Company, Boston.

[3] M.O. Ojo ,H.O. Adeyemi, *On the convolution of random variables from the generalized logistic distribution,* Statistica **11**, anno XLIX, no.3, (1989), 453-459.