

AD-A049 996

ROCHESTER UNIV N Y DEPT OF STATISTICS
ON THE CONVOLUTION OF LOGISTIC RANDOM VARIABLES. (U)
1977 E O GEORGE, G S MUDHOLKAR

F/6 12/1

AFOSR-77-3360

UNCLASSIFIED

AFOSR-TR-78-0041

NL

| OF |
AD
A049996



END
DATE
FILMED

3-78

DDC



SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19 REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER	
18 1 AFOSR-TR-78-8041			
4. TITLE (and Subtitle)		9. TYPE OF REPORT & PERIOD COVERED	
ON THE CONVOLUTION OF LOGISTIC RANDOM VARIABLES.		Interim rept.	
7. AUTHOR(s)		6. PERFORMING ORG. REPORT NUMBER	
E. Olusegun/Govind S. Mudholkar		8. CONTRACT OR GRANT NUMBER(s)	
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
University of Rochester Department of Statistics Rochester, NY 14627		61102F 2304/A5 17/A5	
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE	
Air Force Office of Scientific Research/NM Bolling AFB, DC 20332		1977 12 13 p.	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES	
		10	
		15. SECURITY CLASS. (of this report)	
		UNCLASSIFIED	
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report)			
Approved for public release; distribution unlimited.			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)			
logistic distribution, convolution, approximations, student's t			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)			
The logistic distribution is used for stochastic modelling in fields of applications such as biology, economics and medicine. The convolution of logistic random variables plays an important role in estimation of and testing of hypotheses regarding the parameters of the distribution. In this paper the characteristic function is inverted to obtain a neat expression for the d.f. of the convolution. It is demonstrated that a student's t-distribution provides a very close approximation to the convolution.			

DDC
FEB 15 1978
F

410 276

AD A049996

DDC FILE COPY

AFOSR-TR- 78 - 0041

**On the Convolution of Logistic
Random Variables***

by

E. Olusegun George and Govind S. Mudholkar

University of Ife and University of Rochester

Abstract

The logistic distribution is used for stochastic modelling in fields of applications such as biology, economics and medicine. The convolution of logistic random variables plays an important role in estimation of and testing of hypotheses regarding the parameters of the distribution. In this paper the characteristic function is inverted to obtain a neat expression for the d.f. of the convolution. It is demonstrated that a student's t-distribution provides a very close approximation to the convolution.

**Approved for public release;
distribution unlimited.**

**Key Words: Logistic Distribution, Convolution, Approximations,
student's t.**

* Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF under Grant No. AFOSR-77-3360 and, in part, by the Office of Naval Research under contract number N00014-76-0001. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.

Approved for public release;
distribution unlimited.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
 This technical report has been reviewed and is
 approved for public release IAW AFR 190-12 (7b).
 Distribution is unlimited.
A. D. BLOSE
 Technical Information Officer

ACCESS	NOTATION	DATE
MTIS	DDC	UNANNOUNCED
JUSTIFICATION	BY	DISTRIBUTION/ANAL.
	DIST.	AVAIL.
		FILE

A

1. Introduction

The logistic distribution given by the distribution function

$$F(x) = (1 + \exp(-(\alpha + \beta x)))^{-1}, \quad -\infty < x < \infty, \quad (1.1)$$

where $-\infty < \alpha < \infty$ and $\beta > 0$, is remarkably similar to the normal distribution with mean = α and standard deviation = $\beta\pi/\sqrt{3}$. Moreover for $\alpha = 0$ and $\beta = 1$, the logistic r.v. X and its p.d.f. $f(\cdot)$ are related to the d.f. $F(\cdot)$ by analytically simple forms $x = \log(F/(1-F))$ and $f(x) = F(x)(1-F(x))$. Because of these reasons this distribution is widely employed as a substitute for the normal distribution in applications such as bioassay and quantal response data problems, e.g. Berkson (1944), and growth curve analysis (Verhulst (1845)). The importance of the logistic law in modelling of stochastic phenomena has resulted in numerous studies involving probabilistic and statistical aspects of this distribution. For example, Gumbel (1944), Gumbel and Keeney (1950), and Talacko (1956) show that it arises as a limiting distribution in various situations; Birnbaum and Dudman (1963), Gupta and Shah (1967) study order statistics from it; and many, e.g. Antle et al. (1970) and Tartar and Clark (1965) study inference questions concerning it. For an excellent summary of results about logistic distribution refer to Chapter 22 of Johnson and Kotz (1970).

As might be expected in view of the similarity between the logistic and normal distributions, the sample mean and the sample variance, the moment estimators of α and β , are effective tools for statistical decisions involving the logistic distribution. Antle et al. (1970) give a function of the mean as a confidence interval estimate of α when β is known.

Schaffer and Sheffield (1973) make a case that in terms of mean squared error the moment estimators are as good estimators of α , β as their M.L.E.'s. The fact that the distribution of the sample mean has monotone likelihood ratio with respect to α when β is known leads to a uniformly most accurate confidence interval for α and UMP tests for one-sided hypotheses involving α . The sampling distribution of the mean is a primary requirement for these statistical purposes. The papers due to Antle et al. (1970) and Schafer and Clark (1965) both contain Monte Carlo'd results for this distribution. Goel (1975) obtains an expression for the d.f. of the sum of i.i.d. logistic variates by using the Laplace transform inversion method for convolutions of Polya type function developed by Schoenberg (1953) and Hirschman and Widder (1955). He uses this approach because an inversion of the characteristic function of the sum leads him to a very slowly converging series for its d.f. In this paper we demonstrate that the characteristic function can be inverted to obtain a neat expression for the d.f. Also by noting the length of the tails of the distribution we develop a simple student-t approximation for it and compare it with the normal and an Edgeworth series approximation.

2. D.F. of the Sum of I.I.D. Logistic Variates

Let $X_1, X_2, \dots, X_n$ be i.i.d. with d.f.

$$F(x) = (1 + \exp(-x))^{-1} \quad (2.1)$$

and let

$$Z = \sum_{i=1}^n X_i. \quad (2.2)$$

Goel (1975) obtains the expression

$$F_n(x) = 0.5 + \frac{1}{\pi} \sum_{m=0}^{\infty} ((m+n-1)!/m!) \operatorname{Re}((c+id)^n), \quad (2.3)$$

where $c = (m + n/2)/\alpha$, $d = x(n/3)^{1/2}\alpha$, $\alpha = (m + n/2)^2 + (x^2 n/12)$ for the

d.f. of Z by applying the inversion formula directly. As he notes, this series converges slowly and is therefore computationally inconvenient. In this section we use the inversion formula for characteristic functions to obtain a finite series expression for $F_n(\cdot)$. The method used requires only a knowledge of contour integration and the Mittag-Leffler expansion of the characteristic function of Z which may be found in sources such as the monograph by Widder (1972).

The characteristic function $\phi_n(\cdot)$ of Z is given by

$$\phi_n(s) = (\pi i s / \sin \pi i s)^n. \quad (2.4)$$

Its Mittag-Leffler expansions are given by

$$(\pi i s / \sin \pi i s)^n = (i s)^n \sum_{k=-\infty}^{\infty} \sum_{p=0}^{n-1} A_{n-1,p} / (i s + k)^{n-p}, \quad (2.5)$$

when n is even, and

$$(\pi i s / \sin \pi i s)^n = (i s)^n \sum_{k=-\infty}^{\infty} \sum_{p=0}^{n-1} (-1)^k A_{n-1,p} / (i s + k)^{n-p} \quad (2.6)$$

when n is odd.

The constants $A_{n,p}$ in (2.5) and (2.6) can be computed by using the following equations:

$$(\pi s / \sin \pi s)^n = A_{n,0} + A_{n,1}s + \dots + A_{n,n}s^n, \quad (2.7)$$

and

$$\pi s / \sin \pi s = \sum_{k=0}^{\infty} (-1)^{k-1} (2^{2k} - 2) / (2k)! B_{2k} (\pi s)^{2k}, \quad (2.8)$$

the B_{2k} 's being Bernoulli numbers. Since the d.f. of Z is symmetric about zero, it suffices to find $F_n(x)$ for $x > 0$. The inversion formula gives, $0 < x < b < \infty$,

$$F_n(b) - F_n(x) = \lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-N-\frac{1}{2}}^{N+\frac{1}{2}} \frac{(e^{-isx} - e^{-isb})}{is} \left(\frac{\pi i s}{\sin \pi i s} \right)^n ds. \quad (2.9)$$

Consider the contour integral

$$I_N(b,x) = \frac{1}{2\pi} \int_{C_N} \frac{(e^{-ixz} - e^{-ibz})}{iz} \left(\frac{\pi iz}{\sin \pi iz} \right)^n dz, \quad (2.10)$$

where C_N is a clockwise oriented contour given by $C_N = C'_N + C''_N$,

$$C'_N = \{z : |z| < N + \frac{1}{2} \text{ and } \text{Im } z = 0\}$$

and

$$C''_N = \{z : |z| = N + \frac{1}{2} \text{ and } \text{Im } z < 0\}. \quad (2.11)$$

It is clear that

$$F_n(b) - F_n(x) = \lim_{N \rightarrow \infty} I_N(b, x), \quad (2.12)$$

provided that

$$\lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{C'_N} \frac{(e^{-ixz} - e^{-ibz})}{iz} \left(\frac{\pi iz}{\sin \pi iz} \right)^n dz = 0. \quad (2.13)$$

(2.13) is easy to establish because

$$\left| \int_{C''_N} \frac{e^{-ixz}}{iz} \left(\frac{iz}{\sin iz} \right)^n dz \right| \leq \left(\frac{2\pi(N+\frac{1}{2})}{e^{\pi(N+\frac{1}{2})} - e^{-(N+\frac{1}{2})}} \right) \int_{C''_N} \left| \frac{e^{-ixz}}{iz} \right| dz \rightarrow 0 \quad (2.14)$$

as $N \rightarrow \infty$. Now we use (2.4) to evaluate $F(x)$ when n is even:

$$I_N(b, x) = \frac{1}{2\pi} \int_{C_N} \sum_{p=0}^{n-1} \sum_{k=-\infty}^{\infty} A_{n,p} \frac{(e^{-ixz} - e^{-ibz})}{(iz+k)^{n-p}} (iz)^{n-1} dz \quad (2.15)$$

$$\begin{aligned} &= \sum_{p=0}^{n-1} \sum_{k=1}^N A_{n,p} i^p \frac{1}{2\pi i} \int_{C_N} z^{n-1} \frac{(e^{-ixz} - e^{-ibz})}{(z+ik)^{n-p}} dz \\ &+ \sum_{p=0}^{n-1} \frac{1}{2\pi i} \int_{C_N} \sum_{k=N+1}^{\infty} A_{n,p} i^p z^{n-1} \frac{(e^{-ixz} - e^{-ibz})}{(z+ik)^{n-p}} dz \\ &+ \sum_{p=0}^{n-1} \frac{1}{2\pi i} \int_{C_N} \sum_{k=0}^{\infty} A_{n,p} i^p z^{n-1} \frac{(e^{-ixz} - e^{-ibz})}{(z-ik)^{n-p}} dz. \end{aligned} \quad (2.16)$$

The integrand of equation (2.15) has a removable singularity at zero, and poles of order $(n-p)$ at $z = -ik, k = 1, 2, \dots$. Hence the integrands in the last two terms of equation (2.16) are analytic in C_N . Therefore

$$I_N(b, x) = \sum_{p=0}^{n-1} \sum_{k=1}^N A_{n,p} i^p \frac{1}{2\pi i} \int_{C_N} z^{n-1} \frac{(e^{-ixz} - e^{-ibz})}{(z+ik)^{n-p}} dz.$$

Now let

$$I_N(y) = \sum_{k=1}^N \frac{1}{2\pi i} \int_{C_N} \frac{z^{n-1} e^{-iyz}}{(z+ik)^{n-p}} dz. \quad (2.17)$$

Then

$$I_N(y) = \sum_{k=1}^N \text{Res}(-ik), \quad (2.18)$$

where $\text{Res}(-ik)$ is the residue of $\frac{z^{n-1} e^{-iyz}}{(z+ik)^{n-p}}$ at $z = -ik, k = 1, \dots, N$ and is given by

$$\text{Res}(-ik) = \frac{-1}{(n-1-p)!} \frac{d^{n-1-p}}{dz^{n-1-p}} (e^{-iyz} z^{n-1}) \Big|_{z=-ik} \quad (2.19)$$

$$= \sum_{r=0}^{n-1-p} (-1)^{r+1} (-i)^p \binom{n-1}{p+r} \frac{y^r}{r!} k^{p+r} e^{-ky}. \quad (2.20)$$

Thus

$$\lim_{N \rightarrow \infty} I_N(y) = \sum_{r=0}^{n-1-p} (-1)^{r+1} (-i)^p \binom{n-1}{p+r} \frac{y^r}{r!} \sum_{k=1}^{\infty} k^{p+r} e^{-ky}. \quad (2.21)$$

In the development from (2.12) if we let $b \rightarrow \infty$ then

$$\begin{aligned} 1 - F_n(x) &= \lim_{b \rightarrow \infty} \lim_{N \rightarrow \infty} I_N(b, x) \\ &= \sum_{p=0}^{n-1} \sum_{r=0}^{n-1-p} (-1)^{r+1} A_{n,p} \binom{n-1}{p+r} \frac{x^r}{r!} \sum_{k=1}^{\infty} k^{p+r} e^{-kx}. \end{aligned}$$

Using induction, the infinite series $\sum_{k=1}^{\infty} k^{p+r} e^{-kx}$ can be shown to converge to $\sum_{m=1}^{p+r} s_{p+r,m} (m-1)! (e^{-x}/(1-e^{-x}))^m$, where $s_{p+r,m}$'s are Stirling's number of the second kind given by

$$s_{p+r,m} = ((p+r)!)^{-1} \sum_{i=0}^m (-1)^i \binom{m}{i} (m-i)^{p+r}. \quad (2.22)$$

By a similar method, an expression can be obtained for $F_n(\cdot)$ when n is odd. These results are stated in the following theorem:

THEOREM 2.1. Let $X_1, X_2, \dots, X_n$ be i.i.d. with d.f. given by (2.1). Then the d.f. of $Z = \sum_{i=1}^n X_i$ is given by

$$1 - F_n(z) = \sum_{p=0}^{n-1} \sum_{r=0}^{n-p-1} \sum_{k=1}^{p+r+1} (-1)^{r+1} A_{n,p} s_{p+r+1,k} \frac{x^r}{r!} \binom{n-1}{p+r} (k-1)! \frac{e^{-kx}}{(1+e^{-x})^k}, \quad (2.23)$$

when n is even, and by

$$1 - F_n(z) = \sum_{p=0}^{n-1} \sum_{r=0}^{n-p-1} \sum_{k=1}^{p+r+1} (-1)^{r+k+1} A_{n,p} s_{p+r+1,k} \frac{x^r}{r!} \binom{n-1}{p+r} (k-1)! \frac{e^{-kx}}{(1+e^{-x})^k}, \quad (2.24)$$

when n is odd.

3. Some Approximations for the Convolution of Logistic R.V.'s

As mentioned earlier the logistic distribution is very similar to the normal distribution. It is therefore to be expected that the normal distri-

bution would reasonably approximate the distribution of the sum of k independent logistic r.v.'s for $k \geq 2$. Specifically, if

$$Z^* = \sqrt{3/k} Z/\pi \quad (3.1)$$

denotes the standardized sum of k i.i.d. logistic r.v.'s; then we have a simple approximation:

$$F_k^*(z) \doteq \phi(z), \quad (3.2)$$

where $F_k^*(\cdot)$ is the d.f. of Z^* and ϕ is the standard normal d.f.

This approximation can be improved by applying a two term Edgeworth correction to it to obtain

$$F_k^*(z) \doteq \phi(z) - (20k)^{-1}(z^3 - 3z)W(z), \quad (3.3)$$

where $W(z) = (2\pi)^{-1/2} \exp(-z^2/2)$.

Although the above approximations are good, it may be possible to improve on them by using student's t distributions which is similar to the normal distribution in shape but has relatively long tails. The degree of freedom ν of the approximating t distribution can be obtained by equating the coefficients of kurtosis,

$$\beta_2(Z) = \beta_2(t_\nu), \quad (3.4)$$

which gives

$$\nu = 5k + 4. \quad (3.5)$$

Thus considering the standardized Z and t_ν , i.e. Z^* and t_ν^* , respectively, where

$$t_\nu^* = t_\nu / \sigma(t_\nu) = \sqrt{(\nu-2)/\nu} t_\nu, \quad (3.6)$$

this approximation is given by

$$F_k^*(z) \doteq P\{t_{5k+4} \leq z \sqrt{\frac{5k+4}{5k+2}}\}. \quad (3.7)$$

4. Evaluation of the Approximation

Figure 1 and Table 1 illustrate the quality of the three viz. the normal, the Edgeworth-corrected normal, and the student's t , approximations for $k = 2$ and 3 , respectively. The maximum error for the t -approximation for $k = 2$ is about 0.005 and decreases to 0.0007 when $k = 3$. Goel (1975) gives a table of cumulative d.f. of the standardized mean of samples from logistic population for $k = 10$ correct to seven decimal places. We believe that for this value of k the t -approximation would yield results of comparable accuracy.

Acknowledgement

The authors wish to thank Mr. C. Pun for computing assistance.

TABLE 1.

A Comparison of Three Approximations

$k = 3.$

x	$F_3^*(x)$	$F_3^*(x) - \phi(x)$	$F_3^*(x) - G_2(x)$	$F_3^*(x) - H_3(x)$
0.05	0.5209	0.0010	0.0000	0.0001
0.15	0.5625	0.0029	0.0000	0.0003
0.25	0.6033	0.0046	0.0008	0.0005
0.45	0.6809	0.0073	-0.0017	0.0007
0.65	0.7506	0.0084	-0.0006	0.0007
0.85	0.8106	0.0083	-0.0007	0.0007
1.00	0.8486	0.0073	-0.0008	0.0004
1.20	0.8903	0.0054	-0.0007	0.0002
1.45	0.9291	0.0026	-0.0004	0.0000
1.75	0.9598	-0.0001	0.0001	-0.0002
2.50	0.9918	-0.0020	0.0004	-0.0002
3.00	0.9975	-0.0012	0.0001	0.0000

$F_3^*(x)$ = Distribution function of the standardized sum of 3 i.i.d. logistic r.v.'s

$\phi(x)$ = Distribution function of the standard normal r.v.

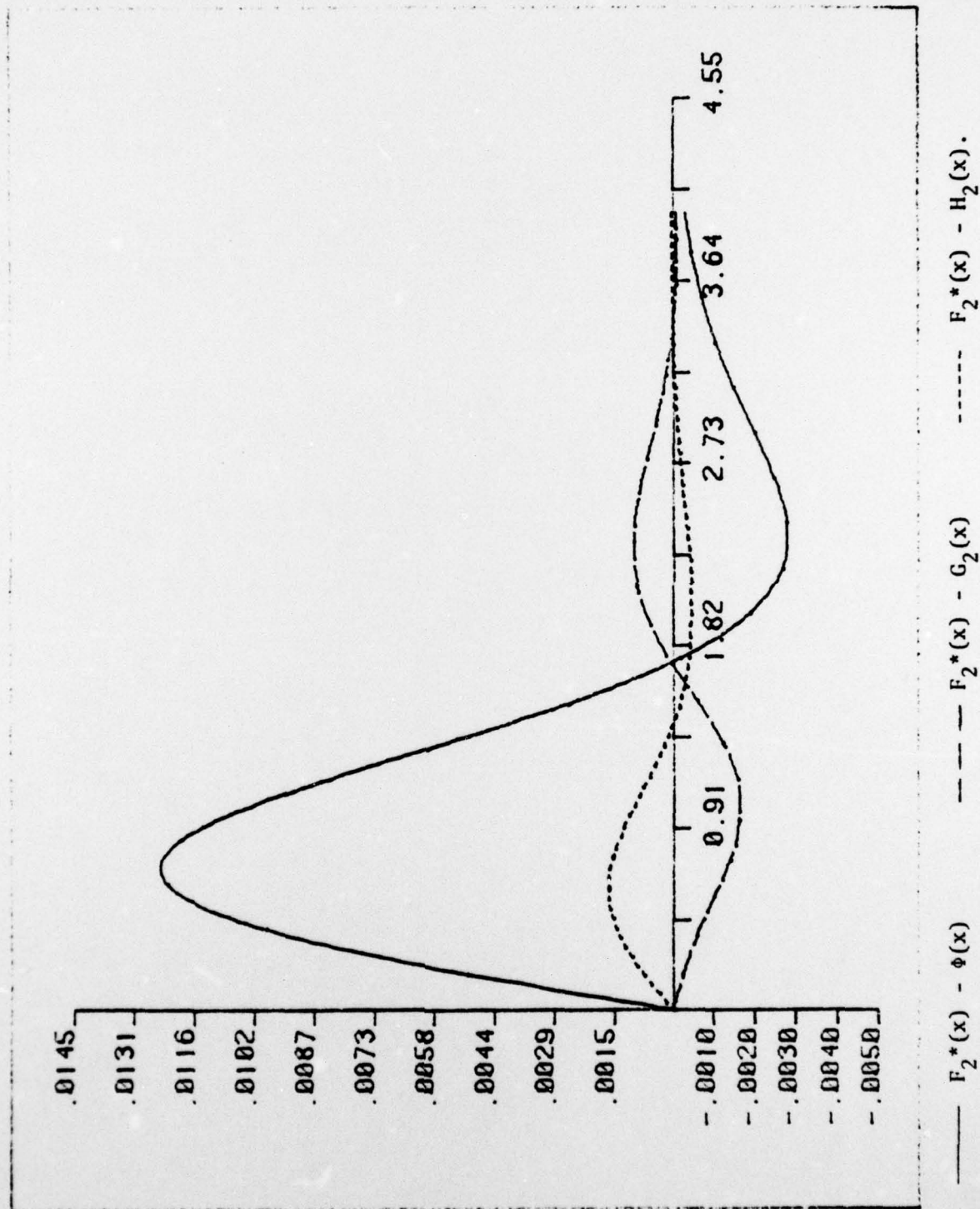
$G_3(x)$ = Edgeworth series approximation given by right hand side of equation (3.3).

$H_3(x)$ = Distribution function of standardized student's t r.v. with 19 degrees of freedom, as given by equation (3.7).

Figure 1.

A Comparison of Three Approximations

$k = 2$



REFERENCES

- Antle, C. E., Klimko, L. and Harkness, W. (1970). Confidence intervals for the parameters of the logistic distribution. *Biometrika* 57, 397-402.
- Berkson, J. (1944). Application of the logistic function to bio-assay. *J. American Statist. Assoc.* 39, 357-365.
- Birnbaum, A. and Dudman, J. (1963). Logistic order statistics. *Ann. Math. Statist.* 34, 658-663.
- Goel, P. K. (1975). On the distribution of standardized mean of samples from the logistic population. *Sankhya* 37, Series B, 165-172.
- Gumbel, E. J. (1944). Ranges and midranges. *Ann. Math. Statist.* 15, 414-422.
- Gumbel, E. J. and Keeney, R. D. (1950). The extremal quotient. *Ann. Math. Statist.* 21, 523-538.
- Gupta, S. S. and Shah, B. K. (1965). The exact moments and percentage points of the order statistics and the distribution of the range from the logistic distribution. *Ann. Math. Statist.* 36, 907-920.
- Hirschman, I. I. and Widder, D. V. (1955). *The Convolution Transform*. Princeton Univ. Press, Princeton, N.J.
- Johnson, N. L. and Kotz, S. (1970). *Continuous univariate distributions-2*. Houghton-Mifflin Co., Boston, Mass.
- Schaffer, R. E. and Sheffield, T. S. (1973). Inferences on the parameters of logistic distribution. *Biometrics* 24, 449-455.
- Schoenberg, I. J. (1953). On Polya frequency functions 1: The totally positive functions and their Laplace transforms. *J. d'Annal. Math.* 1, 331-374.
- Talacko, J. (1956). Perk's distributions and their role in the theory of Wiener's stochastic variables. *Trabajos de Estadística* 7, 159-174.
- Tartar, M. E. and Clark, V. A. (1965). Properties of the median and other order statistics of logistic variates. *Ann. Math. Statist.* 36, 1779-1786.
- Verhulst, P. F. (1845). *Recherches mathématiques sur la loi d'accroissement de la population*. Académie de Bruxelles 18, 1-38.
- Widder, D. V. (1972). *The Laplace Transform*. Eighth printing. Princeton Univ. Press, Princeton, N.J.