The Annals of Statistics
1978, Vol. 6, No. 6, 1352–1359

APPROXIMATING TAIL AREAS OF PROBABILITY DISTRIBUTIONS

BY ALAN J. GROSS AND DAVID W. HOSMER, JR.

Medical University of South Carolina and University of Massachusetts at Amherst

A general method for approximating tail areas is developed through an extension of the methodology of Andrews. This extension is applied to both continuous and discrete distributions. Examples of the approximations are given for the standard normal, t, and chi-square distributions in the continuous case and for the Poisson and binomial distributions in the discrete case. Errors of the approximations are considered. The generality of the method shown indicates that extension is possible to other distributions.

1. Introduction. The problem of approximating the tail areas of statistical distributions has been considered by a number of authors. Blackwell and Hodges (1959) consider approximating the tail areas of convolutions of distributions. Wallace (1959) uses the normal distribution to approximate the tail area for the t and chi-square distribution. Peizer and Pratt (1968) and Pratt (1968) generalize the results of Wallace in the sense of including more distributions that are interrelated. They consider the binomial, negative binomial, Poisson, Pascal, gamma, and beta distributions in addition to the chi-square, t, and normal distributions. Gideon and Gurland (1971) use a weighted sum of exponential functions to approximate tail areas. Their procedure involves solving equations with complex roots. Johnson and Kotz (1970) discuss tail area approximations in their two-volume work on continuous distributions. Specifically, they consider approximations to tail areas of the normal distribution, the chi-square distribution and the t distribution in Chapters 13, 17 and 27 respectively. Johnson and Kotz (1969) consider tail area approximations to the binomial in their volume on discrete distributions. Gray and Lewis (1971) describe a fairly general method for obtaining tail areas of continuous distributions. Other authors who have recently considered problems concerning the inferences on tail areas of distributions include Beran (1975) and Hill (1975).

The method proposed in this paper generalizes and extends the method proposed by Andrews (1973). The Andrews approximation considers tails of distributions which "look" exponential. If X is an exponentially distributed random variable then for $x > 0$

$$\int_x^\infty f(y)\,dy = -f^2(x)/f'(x), \tag{1.1}$$

where $f(x)$ and $f'(x)$ are the density and its first derivative, respectively. In general, if $f(x)$ is a density function then there exists a function $k(x)$ such that the tail area

Received February 1976; revised October 1977.

AMS 1970 *subject classifications*. Primary 60E05; Secondary 65D20.

Key words and phrases. Tail areas, probability approximations, continuous and discrete probability distributions.

can be written

(1.2) $$\int_x^\infty f(y)\,dy = k(x)f^2(x)/f'(x),$$

where $k(x) \equiv f'(x)S(x)/f^2(x)$ and $S(x) \equiv \int_x^\infty f(y)\,dy$. Furthermore, as Andrews points out, if x is sufficiently large and if $S(x)$ is small $k(x)$ should approach a finite limiting value. In this paper, it is assumed that $\lim_{x\to\infty} k(x) = -k$ where $0 < k < \infty$. The investigation of $k(x)$ and its derivatives forms the basis of the proposed general method of approximating tail areas of continuous distributions.

For the case of discrete distributions, suppose that $p_x \equiv \Pr\{X = x\}$, $x = 0, 1, \cdots$. Now if X follows the geometric distributions, it is not hard to show

(1.3) $$\Sigma_{m=x}^\infty p_m = p_x^2/\Delta p_x,$$

where $\Delta p_x = (p_x - p_{x+1})$. Suppose that the sum of the tail probabilities $\Sigma_{m=x}^\infty p_m$ "looks" geometric in the sense that

(1.4) $$\Sigma_{m=x}^\infty p_m = d(x)p_x^2/\Delta p_x,$$

where $d(x) \equiv (\Sigma_{m=x}^\infty p_m)\Delta p_x/p_x^2$. It is further assumed that $\lim_{x\to\infty} d(x) = d$, $0 < d < \infty$. If x is large and $\Sigma_{m=x}^\infty p_m$ is small then $d(x)$ should be near its limiting value. To obtain the approximations successive differences are used as the discrete analog to derivatives.

2. Approximation results. The proposed approximations are based on the following assumptions when X is continuous. For all $x > x_0$, x_0 sufficiently large and depending on the density function,

(i) $f(x) = \exp[-a(x)]$,
(ii) $a(x) \geqslant 0$ and has at least four derivatives, which exist for $x > x_0$, and
(iii) $\lim_{x\to\infty} k(x) = -k$, where $0 < k < \infty$.

Before proceeding, it is interesting to note the relationship between Andrews' (1973) work and the method proposed in this paper. In terms of Andrews' notation (1973, page 371) it is easy to see that

(2.1) $$g(x) = -a'(x),$$

(2.2) $$g'(x) = -a''(x),$$

and

(2.3) $$-k = \lim_{x\to\infty} - a''(x)/(a'(x))^2.$$

If $A(x)$ denotes the Andrews approximation, then

(2.4) $$A(x) = (a(x))^{-1}e^{-a(x)}\Big[\lim_{x\to\infty}\{a''(x)/(a'(x))^2 + 1\}^{-1}$$
$$\{1 - 1/2(a''(x))/(a'(x))^2 - \lim_{x\to\infty} a''(x)/(a'(x))^2\}\Big].$$

Approximations to $S(x)$ considered in this paper (for $S(x)$ small and x large) are obtained by assuming $k'(x) = 0$ and hence $k''(x) = 0$. Noting that $k(x) = -a'(x)\exp[a(x)]\int_x^\infty \exp[-a(y)]\,dy$, the approximations $S_1(x)$ (based on $k'(x) = 0$)

and $S_2(x)$ (based on $k''(x) = 0$) are, respectively,

$$S_1(x) = e^{-a(x)}\left[\frac{a'(x)}{a''(x) + \{a'(x)\}^2}\right], \tag{2.5}$$

and

$$S_2(x) = e^{-a(x)}\left[\frac{2a''(x) + \{a'(x)\}^2}{a'''(x) + 3a''(x)a'(x) + \{a'(x)\}^3}\right], \tag{2.6}$$

where $a'(x)$, $a''(x)$ and $a'''(x)$ are the first three derivatives of $a(x)$, respectively.

The question now arises under what conditions do $S_1(x)$ and $S_2(x)$ provide bounds for $S(x)$? This question is addressed by examining conditions on $a(x)$ and its first four derivatives. The conditions are obtained by generalizing a method that appears in Feller (1950, page 131). By rearranging the order of integration and differentiation it is not hard to show that, for fixed x, $S_1(x)$ and $S_2(x)$ can be rewritten as

$$S_1(x) = \int_x^{\infty} e^{-a(y)}[\,g_1(y)/h_1(y)\,dy] \tag{2.7}$$

and

$$S_2(x) = \int_x^{\infty} e^{-a(y)}[\,g_2(y)/h_2(y)\,dy], \tag{2.8}$$

where

$$g_1(y) = h_1(y) + a'(y)a'''(y) - 2\{a''(y)\}^2,$$

$$h_1(y) = [\{a'(y)\}^2 + a''(y)]^2,$$

$$g_2(y) = h_2(y) + 6\{a''(y)\}^3 + a^{[4]}(y)[\{a'(y)\}^2 + 2a''(y)]$$
$$-6a'(y)a''(y)a'''(y) - 3\{a'''(y)\}^2,$$

and

$$h_2(y) = [a'''(y) + 3a''(y)a'(y) + \{a'(y)\}^3]^2,$$

$y \geqslant x$. Thus a sufficient condition for $S_1(x)$ to be a lower (upper) bound of $S(x)$ is if $2\{a''(y)\}^2 \geqslant (\leqslant)\, a'(y)a'''(y)$ (a.e.) for $y \geqslant x$. Similarly for $S_2(x)$ to be an upper (lower) bound of $S(x)$ it is sufficient that $6\{a''(y)\}^3 + a^{[4]}(y)[\{a'(y)\}^2 + 2a''(y)] \geqslant (\leqslant)\, 6a'(y)a''(y)a'''(y) + 3\{a'''(y)\}^2$, (a.e.) for $y \geqslant x$. Two further mild restrictions are that $h_i(y) \neq 0$ (a.e.), $i = 1, 2$, $y \geqslant x$.

Table 1 is a summary of functions used to obtain $S_1(x)$ and $S_2(x)$ for the normal, t, and chi-square distributions. It is easily verified that for the standard normal distribution and t-distribution $S_1(x)$ and $S_2(x)$ are lower and upper bounds of their respective $S(x)$ functions for $x > 0$. For the chi-square distribution $S_1(x)$ is a upper bound if $df = 1$, otherwise it is a lower bound. $S_2(x)$ is an upper bound if $df's \geqslant 2$ and $x \geqslant 2\ df's - 4$. If $df's = 1$, then $S_2(x)$ is a lower bound. Table 2 contains the tail area approximations $S_1(x)$ and $S_2(x)$ for selected values of x for the standard normal distribution and selected values of x and $df's$ for the chi-square and t

TABLE 1

Summary of functions used to obtain approximations $S_1(x)$ and $S_2(x)$

Density	$a(x)$	$a'(x)$	$a''(x)$	$a'''(x)$	$a^{(4)}(x)$
Standard Normal	$x^2/2 + (\frac{1}{2})\ln(2\pi)$	x	1	0	0
t, ν degrees of freedom	$\left(\frac{\nu+1}{2}\right)\ln\left(1+\frac{x^2}{\nu}\right) + \ln\left\{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{(\pi\nu)^{\frac{1}{2}}\Gamma\left(\frac{\nu}{2}\right)}\right\}$	$\left(\frac{\nu+1}{\nu}\right)\cdot\frac{x}{\left(1+\frac{x^2}{\nu}\right)}$	$\left(\frac{\nu+1}{\nu}\right)\cdot\frac{1-\frac{x^2}{\nu}}{\left(1+\frac{x^2}{\nu}\right)^2}$	$\left(\frac{\nu+1}{\nu}\right)\cdot\frac{\frac{2x^3}{\nu^2}-\frac{6x}{\nu}}{\left(1+\frac{x^2}{\nu}\right)^3}$	$\left(\frac{\nu+1}{\nu}\right)\cdot\frac{\frac{6}{\nu}\left[\frac{6x^2}{\nu}-\frac{x^4}{\nu^2}-1\right]}{\left(1+\frac{x^2}{\nu}\right)^4}$
Chi-squared ν degrees of freedom	$\frac{\nu}{2}\ln 2 + \ln\Gamma\left(\frac{\nu}{2}\right) - \left(\frac{\nu-2}{2}\right)\ln(x) + \frac{x}{2}$	$\frac{1}{2} - \left(\frac{\nu-2}{2}\right)\frac{1}{x}$	$\left(\frac{\nu-2}{2}\right)\frac{1}{x^2}$	$-\left(\frac{\nu-2}{2}\right)\frac{2}{x^3}$	$\left(\frac{\nu-2}{2}\right)\frac{6}{x^4}$

TABLE 2

The percent error of the approximations $S_1(x)$ and $S_2(x)$ for the central chi square and t distributions and the standard normal distribution (t, $df = \infty$) (percent error = $100(S_i(x) - S(x)/S(x))$, $i = 1, 2$)

	t distributions			*Chi-square distribution*	
df	$S(x)$	$S_1(x)$	$S_2(x)$	$S_1(x)$	$S_2(x)$
	0.10	−6.45	*	7.17	−9.91
	0.05	−1.64	*	3.47	−3.64
1	0.025	−0.41	*	1.94	−1.63
	0.01	−0.07	*	1.03	−0.69
	0.005	−0.02	*	0.70	−0.41
	0.10	−14.79	9.69	−8.09	6.12
	0.05	−8.32	4.79	−4.41	2.71
10	0.025	−5.22	2.74	−2.71	1.41
	0.01	−3.16	1.52	−1.60	0.70
	0.005	−2.28	1.04	−1.14	0.44
	0.10	−14.86	8.48	−9.98	6.79
	0.05	−8.41	4.05	−5.52	3.06
20	0.025	−5.33	2.26	−3.42	1.62
	0.01	−3.27	1.21	−2.05	0.81
	0.005	−2.40	0.82	−1.48	0.53
	0.10	−14.88	7.84	−11.71	7.19
	0.05	−8.43	3.66	−6.54	3.28
50	0.025	−5.35	2.00	−4.10	1.75
	0.01	−3.30	1.05	−2.48	0.89
	0.005	−2.43	0.70	−1.81	0.58
	0.10	−14.86	7.44		
	0.05	−8.44	3.42		
∞	0.025	−5.36	1.84		
	0.01	−3.30	0.95		
	0.005	−2.43	0.62		

distributions. In general, one may compute a bound on the percent error of the approximation as

$$\frac{\max(S_1(x), S_2(x)) - \min(S_1(x), S_2(x))}{\min(S_1(x), S_2(x))} \times 100.$$

Suppose p_x is a discrete distribution whose tail sum is given by (1.4). It is not hard to show that if p_x is a k-modal distribution whose kth mode is M_k and if $\{p_{m+1}/p_m\}$ is a decreasing sequence with $\lim_{m\to\infty}(p_{m+1}/p_m) = 0$, then $\lim_{x\to\infty} d(x) = 1$. (If p_x has a finite domain $\{0, 1, 2, \cdots n\}$ (say) the result is still valid, however, with $\lim_{m\to n}$ replacing $\lim_{m\to\infty}$.)

Analogous to the continuous case approximations to $\Sigma_{m=x}^{\infty} p_m$ are obtained (for $\Sigma_{m=x}^{\infty} p_m$ small and $x \geqslant M_k$ large) by setting $d(x+1) - d(x) = 0$, and subsequently, $d(x+2) - 2d(x+1) + d(x) = 0$. The resulting approximations π_{x1} and π_{x2} are, respectively,

$$\pi_{x1} \equiv \frac{p_x}{1 - \dfrac{\Delta p_x}{\Delta p_{x+1}} \dfrac{p_{x+1}^2}{p_x^2}}, \tag{2.9}$$

TABLE 3

Percent error in the approximations π_{x1}, π_{x2} for the binomial and Poisson distribution (percent error = $100(\pi_{xi} - \Sigma p_m)/\Sigma p_m$, $i = 1, 2$)

Binomial						Poisson				
p	n	x	$\Sigma_{m=x}^{n} p_m$	π_{x1}	π_{x2}	λ	x	$\Sigma_{m=x}^{\infty} p_m$	π_{x1}	π_{x2}
		3	0.1841	−1.093	0.172		3	0.0803	−0.272	0.034
	15	4	0.0556	−0.298	0.032	1.0	4	0.0190	−0.092	0.008
		5	0.0127	−0.099	0.007		5	0.0040	−0.038	0.003
		6	0.0022	−0.037	0.002					
							8	0.1334	−3.305	0.978
		6	0.0732	−0.972	0.171		9	0.0681	−1.677	0.407
0.1	30	7	0.0258	−0.415	0.056	5.0	10	0.0318	−0.915	0.184
		8	0.0078	−0.196	0.026		11	0.0137	−0.530	0.090
		9	0.0020	−0.099	0.008		12	0.0054	−0.323	0.047
		8	0.0757	−1.551	0.332		15	0.0834	−3.438	1.059
	45	9	0.0320	−0.758	0.130		16	0.0487	−2.159	0.576
		10	0.0120	−0.398	0.055	10.0	17	0.0270	−1.405	0.327
		11	0.0040	−0.222	0.025		19	0.0072	−0.943	0.192
							20	0.0035	−0.459	0.074
		10	0.1509	−3.295	0.546					
	15	11	0.0592	−1.106	0.116		26	0.1122	−6.747	2.594
		12	0.0176	−0.374	0.022		28	0.0524	−3.386	1.043
		13	0.0037	−0.119	*	20	30	0.0218	−1.826	0.459
							32	0.0081	−1.048	0.218
		19	0.1002	−3.788	0.837		33	0.0047	−0.810	0.154
0.5	30	20	0.0494	−1.920	0.331					
		21	0.0214	−1.004	0.135		60	0.0923	−7.792	3.143
		22	0.0081	−0.538	0.056		62	0.0557	−5.038	1.756
						50	65	0.0236	−2.764	0.787
		27	0.1163	−5.838	1.627		68	0.0089	−1.608	0.379
		28	0.0676	−3.392	0.783		70	0.0043	−1.153	0.241
	45	29	0.0362	−2.023	0.388					
		30	0.0178	−1.235	0.197					
		31	0.0080	−0.770	0.102					

and

$$(2.10) \qquad \pi_{x2} \equiv \frac{\dfrac{(p_x + p_{x+1})\Delta p_{x+2}}{p_{x+2}^2} - \dfrac{2p_x \Delta p_{x+1}}{p_{x+1}^2}}{\dfrac{\Delta p_{x+2}}{p_{x+2}^2} - \dfrac{2\Delta p_{x+1}}{p_{x+1}^2} + \dfrac{\Delta p_x}{p_x^2}}.$$

It is not difficult to show that $\pi_{x\nu}$ is an upper (lower) bound of $\Sigma_{m=x}^{\infty} p_m$ provided $p_y \leq (\geq) \pi_{y\nu} - \pi_{y+1,\nu}$ for all $y > x$, $\nu = 1, 2$. For the binomial and Poisson π_{x1} and π_{x2} are lower and upper bounds, respectively, of $\Sigma_{m=x}^{\infty} p_m$. Table 3 illustrates the tail area approximations for selected values of x and the parameters of these two distributions.

The Andrews (1973) method of approximating tail areas is generalized for continuous distribution and extended to discrete distributions. The new method has been applied to the five most commonly tabled distributions, the standard normal, the central chi-square and t, the Poisson and binomial. It is clear that the principal appeal of this method is its applicability to a wide class of distributions both continuous and discrete.

Acknowledgments. We gratefully acknowledge the assistance of the referee, Associate Editor and Editor, whose constructive comments enhanced the paper greatly. The assistance of Professor Hurshell Hunt, Department of Biometry, Medical University of South Carolina and Professor Ram Dahiya, Department of Mathematics and Statistics, University of Massachusetts is also appreciated.

REFERENCES

[1] Andrews, D. F. (1973). A general method for the approximation of tail areas. *Ann. Statist.* **1** 367–372.

[2] Beran, R. (1975). Tail probabilities of noncentral quadratic forms. *Ann. Statist.* **3** 969–974.

[3] Blackwell, D. and Hodges, J. L. (1959). The probability in the extreme tail of a convolution. *Ann. Math. Statist.* **30** 1113–1120.

[4] Feller, W. (1950). *An Introduction to Probability Theory and Its Applications*, **1**. Wiley, New York.

[5] Gideon, R. A. and Gurland, J. (1971). Interpolatory method of approximating distribution functions with applications to gamma and related variates. *J. Amer. Statist. Assoc.* **66** 577–582.

[6] Gray, H. L. and Lewis, T. O. (1971). Approximating tail areas by means of B-transformation. *J. Amer. Statist. Assoc.* **66** 897–899.

[7] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174.

[8] Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions*. Houghton Mifflin, New Jersey.

[9] Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions* **1** and **2**. Houghton Mifflin, New Jersey.

[10] Peizer, D. B. and Pratt, J. W. (1968). A normal approximation for binomial, *F*, beta, and other common, related tail probabilities, **I**. *J. Amer. Statist. Assoc.* **63** 1416–1456.

[11] PRATT, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, **2**. *J. Amer. Statist. Assoc.* **63** 1457–1483.

[12] WALLACE, D. L. (1959). Bounds on normal approximations to Student's and the chi-square distributions. *Ann. Math. Statist.* **30** 1121–1130.

DEPARTMENT OF BIOMETRY
MEDICAL UNIVERSITY OF SOUTH CAROLINA
171 ASHLEY AVENUE
CHARLESTON, SOUTH CAROLINA 29403

DIVISION OF PUBLIC HEALTH
SCHOOL OF HEALTH SCIENCES
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01003