Arab Gulf J. Scient. Res., **9** (3), pp. 1-12 (1991)

Tail Area Approximation Methods: A Comparison

Mohammed A. Shayib

Department of Mathematics, Texas Tech University, Lubbock, Texas 79409-1042, U.S.A.

ABSTRACT. Andrews (1973) gave a general method for the approximation of tail areas in statistical problems. Gross and Hosmer (1978) extended the methodology of Andrews. Awad and Shayib (1986) contributed a wider generalization for approximating tail areas of distributions that were not addressed before. These works led to several formulas for approximating the tail area of a distribution. This article applies four approximating formulas to the t, chi-square and F distributions. Numerical values of relative errors are given to compare the merits of the formulas. The results show that the methods considered here overestimate the tail area. Moreover, the formula proposed by Awad and Shayib is the simplest to use.

1. Introduction

Tail areas, or significance levels, are used widely in statistical problems. The problem of approximating the tail areas of statistical distributions has been considered by a number of authors. Gross and Hosmer (1978) give a general method for approximating tail areas through an extension of the methodology of Andrews (1973). Awad and Shayib (1986) generalize Andrews' s formula and suggest approximating formulas for the normal, χ^2, F, lognormal, logistic, extreme-value, stable, log-Cauchy, Cauchy, exponential and Pareto distributions. Their approximating formulas do better than that of Andrews when applied to the same distributions. They consider more distributions and their formulas apply whenever the Andrews formula fails (see Andrews 1973, p. 372).

Gross and Hosmer (1978) and the references cited there, except Peizer and Pratt (1968), did not apply their formulas to the F-distribution. Moreover, the Peizer and Pratt (1968) procedure was not used for large x values.

Section 2 gives the notation and the assumptions needed. Section 3 presents the approximations and compares their derivations. The results are summarized in Section 4, along with the conclusions reached.

2. Notation and Assumptions

Let X be a continuous random variable with density function $f(x)$. The probability that $X \geq x$, namely,

$$\alpha(x) = P[X \geq x] = \int_x^\infty f(z)\, dz, \tag{1}$$

is called the tail area for X.

Given f and x, one may be interested in evaluating $\alpha(x)$ for large values of x. This problem arises in many statistical applications, such as evaluating type-I or type-II error in hypothesis testing or P-values in significance testing.

Let

$$g(x) = f'(x)/f(x), \tag{2}$$

and

$$K(x) = g'(x)/g^2(x) \tag{3}$$

The presentation in this paper uses the following assumptions. For a given large x,

i) $f(z)$, $f'(z)$ and $f''(z)$ exist, they are continuous and none of them is zero in $[x, \infty)$,

ii) $f(z)$ is decreasing to zero in $[x, \infty)$, and

iii) $K(x) \to K \neq 1$, as $x \to \infty$.

3. Approximating Formulas

This section deals with four formulas for approximating the tail area of a distribution. Based on the notation in Section 2, we have the following formula

$$\hat{\alpha}_0(x) = [f(x)/g(x)][(K-1)]^{-1}\,[1 + \tfrac{1}{2}(K(x) - K)] \tag{4}$$

which was proposed by Andrews (1973). Secondly, Gross and Hosmer (1978) suggest a generalization to $\hat{\alpha}_0(x)$, namely, the following formulas

$$S_1(x) = [f(x)/g(x)]\ [1/(K(x) - 1)], \tag{5}$$

and

$$S_2(x) = -\ \{f(x)/g(x)\}[\{1 - 2K(x)\}/\{1 - 3K(x) + g''(x)/(g(x))^3\}]. \tag{6}$$

Awad and Shayib (1986), thirdly, proposed formulas for approximating the tail area, and the easiest to use was the following

$$\hat{\alpha}_4(x) = [f(x)/g(x)]\ [1/(K - 1)], \qquad K \neq 1. \tag{7}$$

When K = 1, (which does not happen for the distributions that we consider here), one can apply other formulas, which are as accurate as $\hat{\alpha}_4(x)$ but not as easy to use, (see Awad and Shayib (1986)).

Section 4 below has the numerical comparison which includes the t and chi-square distribution in addition to the F distribution.

Because Awad and Shayib (1986) have already studied $S_1(x)$ which they denote by $\hat{\alpha}_1(x)$, we do not include it in the present comparison.

The formula for $S_2(x)$ in (6) irrites simplification, and use of the limiting value K in (6) instead of K(x) appears to be promising. This substitution yields $S_2(x)$ in (6) to

$$\hat{\alpha}_5(x) = -[f(x)/g(x)].\ [\{1 - 2K\}/\{(1 - 3K + g''(x)/(g(x))^3\})], \tag{8}$$

a fourth formula to be compared with the other formulas given above.

Before making numerical comparisons, we note some special cases. For the exponential distribution, all four of $\hat{\alpha}_0(x)$, $S_2(x)$, $\hat{\alpha}_4(x)$, and $\hat{\alpha}_5(x)$are exact; each equals $\alpha(x)$. Awad and Shayib (1986) show that $S_1(x)$ is exact for the exponential distribution and also for the generalized Pareto distribution. Also, for the F

distribution, with n_1 and n_2 degrees of freedom in the numerator and denominator, respectively, straight forward calculation give two results:

(i) If $n_1 = 2$, and n_2 is any positive integer, then $K(x) = 2/(2 + n_2)$; *i.e.*, $K(x)$ is independent of x.

(ii) If $n_1 = n_2 = 2$, then $S_2(x)$, in (6), is undefined, and so is $\hat{\alpha}_5(x)$ in (8).

4. Results and Conclusions

We have considered four formulas for approximating the tail area $\alpha(x)$. Table 1 (see Andrews (1973), Table 1). gives the functions required for the approximation of the distributions considered here: t with n degrees of freedom, chi-square with n degrees of freedom, and F with n_1 degrees of freedom in the numerator and n_2 degrees of freedom in the denominator.

Tables 2-4 show the relative errors computed for the four procedures, according to

$$\text{relative error} = (\hat{\alpha} - \alpha)/\alpha. \tag{9}$$

The relative errors are given for $\alpha = .001, .005, .01, .025$ and $.05$. The entries in the tables were computed for four decimal places.

Table 1. Approximating functions for common distributions

Density $f(x)$	$g(x) = f'(x)/f(x)$	$g'(x)$	$K = \lim_{x\to\infty} \frac{g'(x)}{g^2(x)}$
t, n degrees of freedom	$-\left(1+\frac{1}{n}\right)\frac{t}{\left(1+\frac{t^2}{n}\right)}$	$-\left(1+\frac{1}{n}\right)\frac{\left(1-\frac{t^2}{n}\right)}{\left(1+\frac{t^2}{n}\right)^2}$	$\frac{1}{n+1}$
χ^2, n degrees of freedom	$\frac{n-2}{2x} - \frac{1}{2}$	$-\frac{n-2}{2x^2}$	0
$F_{n1,n2}$	$\frac{1}{2}\left[\frac{n_1-2}{x} - \frac{(n_1+n_2)n_1}{(n_1+n_2x)}\right]$	$\frac{1}{2}\left[\frac{2-n_1}{x^2} + \frac{(n_1+n_2)n_1^2}{(n_2+n_1x)^2}\right]$	$\frac{2}{2+n_2}$

In each table the column headed by J specifies the procedure used: J = 1 represents $\hat{\alpha}_0(x)$, J = 2 represents $\hat{\alpha}_4(x)$, J = 3 corresponds to $S_2(x)$ and J = 4 corresponds to $\hat{\alpha}_5(x)$.

Throughout Tables 2–4, the formulas considered here overestimate the actual value of the tail area. From Table 2 and 3, it appears, for a fixed value of n,

Table 2. t Distribution relative error of approximation

N	J	0.0010	0.0050	0.0100	0.0250	0.0500
2	1	0.0003	0.0016	0.0033	0.0078	0.0143
	2	0.0010	0.0050	0.0102	0.0263	0.0556
	3	0.0030	0.0150	0.0298	0.0740	0.1471
	4	0.0132	0.0702	0.1528	0.5184	2.5625
4	1	0.0049	0.0109	0.0152	0.0228	0.0276
	2	0.0127	0.0303	0.0450	0.0788	0.1268
	3	0.0065	0.0161	0.0245	0.0452	0.0770
	4	0.0537	0.1376	0.2185	0.4567	1.0124
6	1	0.0104	0.0182	0.0229	0.0296	0.0306
	2	0.0266	0.0510	0.0693	0.1089	0.1625
	3	0.0062	0.0133	0.0195	0.0349	0.0596
	4	0.0742	0.1515	0.2172	0.3869	0.7041
8	1	0.0146	0.0228	0.0273	0.0327	0.0308
	2	0.0374	0.0648	0.0848	0.1269	0.1829
	3	0.0055	0.0115	0.0168	0.0301	0.0520
	4	0.0832	0.1526	0.2082	0.3435	0.5727
12	1	0.0199	0.0279	0.0319	0.0354	0.0299
	2	0.0513	0.0814	0.1027	0.1469	0.2050
	3	0.0047	0.0096	0.0141	0.0258	0.0453
	4	0.0898	0.1485	0.1932	0.2958	0.4548
16	1	0.0228	0.0305	0.0341	0.0364	0.0288
	2	0.0595	0.0907	0.1125	0.1576	0.2166
	3	0.0042	0.0087	0.0128	0.0238	0.0423
	4	0.0917	0.1443	0.1834	0.2707	0.4004
20	1	0.0247	0.0321	0.0354	0.0369	0.0279
	2	0.0649	0.0966	0.1187	0.1643	0.2238
	3	0.0039	0.0082	0.0121	0.0226	0.0405
	4	0.0922	0.1411	0.1769	0.2554	0.3692

that the relative error rate increases with α. Similarly, it increases with n when α is kept fixed. Among the four formulas used here, the results show that $S_2(x)$ is the most efficient method, especially when both α and n increase simultaneously. For the F distribution, Table 4, and for a small value of α, the four methods appear to have almost the same relative error. In this case $S_2(x)$ was the least efficient. In addition, $\hat{\alpha}_4(x)$ is easier to use than any other formula.

Table 3. Chi-Square Distribution Relative Error of Approximation

N	J	0.0010	0.0050	0.0100	0.0250	0.0500
2	1	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0000	0.0000	0.0000	0.0000	0.0000
4	1	0.0037	0.0060	0.0070	0.0085	0.0092
	2	0.0112	0.0183	0.0231	0.0333	0.0465
	3	0.0003	0.0023	0.0040	0.0085	0.0169
	4	0.0148	0.0260	0.0347	0.0553	0.0880
6	1	0.0080	0.0098	0.0110	0.0129	0.0132
	2	0.0200	0.0292	0.0363	0.0514	0.0712
	3	0.0024	0.0037	0.0056	0.0115	0.0225
	4	0.0252	0.0400	0.0523	0.0818	0.1281
8	1	0.0087	0.0119	0.0138	0.0157	0.0155
	2	0.0239	0.0363	0.0454	0.0636	0.0877
	3	0.0014	0.0039	0.0065	0.0131	0.0254
	4	0.0299	0.0487	0.0637	0.0980	0.1520
12	1	0.0120	0.0155	0.0172	0.0193	0.0181
	2	0.0316	0.0467	0.0574	0.0800	0.1094
	3	0.0023	0.0048	0.0073	0.0148	0.0283
	4	0.0386	0.0606	0.0776	0.1177	0.1798
16	1	0.0133	0.0177	0.0195	0.0215	0.0195
	2	0.0361	0.0536	0.0655	0.0908	0.1236
	3	0.0020	0.0052	0.0078	0.0156	0.0297
	4	0.0433	0.0680	0.0864	0.1294	0.1956
20	1	0.0155	0.0190	0.0212	0.0231	0.0205
	2	0.0406	0.0584	0.0716	0.0987	0.1340
	3	0.0030	0.0051	0.0081	0.0161	0.0306
	4	0.0481	0.0729	0.0926	0.1374	0.2060

Table 4. F Distribution Relative Error of Approximation

N1	N2	J	0.001	0.005	0.010	0.025	0.050
4	4	1	−0.001	0.000	0.000	0.000	0.001
		2	−0.001	0.000	0.001	0.001	0.003
		3	−0.001	0.001	0.002	0.005	0.012
		4	0.000	0.003	0.006	0.018	0.044
4	8	1	0.000	0.001	0.001	0.002	0.003
		2	0.001	0.002	0.003	0.006	0.011
		3	0.001	0.002	0.003	0.007	0.015
		4	0.003	0.010	0.016	0.034	0.066
4	10	1	0.001	0.001	0.002	0.003	0.004
		2	0.001	0.003	0.005	0.009	0.014
		3	0.001	0.002	0.003	0.008	0.016
		4	0.005	0.012	0.019	0.038	0.071
4	20	1	0.001	0.003	0.004	0.005	0.006
		2	0.004	0.008	0.011	0.017	0.026
		3	0.001	0.002	0.004	0.008	0.017
		4	0.009	0.018	0.027	0.048	0.082
8	4	1	0.001	0.000	0.000	0.000	0.001
		2	0.001	0.000	0.000	0.002	0.004
		3	0.002	0.001	0.002	0.006	0.016
		4	0.002	0.004	0.008	0.024	0.060
8	8	1	0.000	0.001	0.001	0.003	0.004
		2	0.001	0.003	0.005	0.009	0.016
		3	0.001	0.002	0.004	0.010	0.021
		4	0.005	0.014	0.024	0.050	0.098
8	10	1	0.001	0.002	0.002	0.004	0.005
		2	0.002	0.005	0.007	0.013	0.022
		3	0.001	0.003	0.005	0.011	0.022
		4	0.008	0.019	0.029	0.058	0.109
8	20	1	0.003	0.005	0.006	0.008	0.009
		2	0.007	0.013	0.018	0.028	0.042
		3	0.001	0.004	0.006	0.013	0.024
		4	0.015	0.031	0.044	0.078	0.133
10	4	1	0.002	−0.001	0.000	0.000	0.0001
		2	0.002	0.000	0.000	0.001	0.004
		3	0.002	0.000	0.002	0.006	0.016
		4	0.002	0.003	0.008	0.025	0.062

Table 4. *(Contd.)*

N1	N2	J	0.001	0.005	0.010	0.025	0.050
10	8	1	0.000	0.001	0.002	0.003	0.004
		2	0.001	0.003	0.005	0.010	0.017
		3	0.001	0.002	0.005	0.010	0.022
		4	0.005	0.015	0.025	0.053	0.104
10	10	1	0.001	0.002	0.003	0.004	0.006
		2	0.002	0.005	0.008	0.014	0.023
		3	0.001	0.003	0.005	0.012	0.024
		4	0.008	0.020	0.031	0.062	0.117
10	20	1	0.003	0.005	0.006	0.009	0.010
		2	0.008	0.014	0.019	0.031	0.046
		3	0.002	0.004	0.006	0.013	0.026
		4	0.017	0.033	0.048	0.084	0.145
20	4	1	0.003	0.000	0.000	0.001	0.001
		2	0.003	0.000	0.000	0.002	0.004
		3	0.003	0.001	0.002	0.007	0.017
		4	0.004	0.004	0.008	0.027	0.066
20	8	1	0.001	0.001	0.002	0.003	0.005
		2	0.001	0.004	0.005	0.011	0.018
		3	0.001	0.003	0.005	0.011	0.024
		4	0.006	0.017	0.027	0.058	0.114
20	10	1	0.001	0.002	0.003	0.005	0.006
		2	0.003	0.006	0.008	0.016	0.026
		3	0.002	0.004	0.005	0.013	0.026
		4	0.009	0.022	0.034	0.069	0.129
20	20	1	0.003	0.006	0.008	0.010	0.012
		2	0.009	0.017	0.023	0.036	0.054
		3	0.002	0.005	0.007	0.015	0.029
		4	0.019	0.039	0.056	0.097	0.167

Acknowledgment

The author is grateful to the editor and referees for their constructive remarks, without which this article would not have been refined to its present form.

References

Andrews, D.F. (1973) A general method for the approximation of tail areas, *Annals of Statistics*, **1:** 367-372.

Awad, A.M. and **Shayib, M.A.** (1986) Tail area revisited, *Commun. Stat. Simula. and Comp.*, **15:** 1215-1234.

Gross, A.J. and **Hosmer, D.W.** (1978) Approximating tail areas of probability distributions, *Annals of Statistics*, **6:** 1352-1359.

Peizer, D.B. and **Pratt, J.W.** (1968) A normal approximation for binomial, F, beta, and other common related tail probabilities, I. *Journal of the American Statistical Association*, **63:** 1416-1456.

(Received 29/01/1989;
in revised form 18/10/1991)

طرق تقريب المساحات : مقارنة

محمَّد عبدالله الشايب

قسم الرياضيات ـ جامعة تكساس التقنية ـ لوبوك ـ تكساس
الولايات المتحدة الأمريكية ـ ٧٩٤٠٩ ـ ١٠٤٢

الهدف من هذا البحث هو دراسة أربع طرق لتقريب المساحات وذلك بـأخذ نهاية التوزيعات المركزية عندما تكون قيمة المتغـير العشوائي مـوجبة وكبـيرة جداً، ومن ثمَّ إجراء مقارنة عددية بين هذه الطرق.

فقد أعطى أندروز (Andrews) في عام ١٩٧٣م طريقة عامة لتقريب المساحات وذلك في مسائل إحصائية.

كما تناول غروس و هوسمر (Gross and Hosmer) في عام ١٩٧٨ م هذا الموضوع وتوسَّعا فيه.

وعمَّم عـوض وشـايب (Awad and Shayib) في عـام ١٩٨٦ م هـذه الـطرق في تقريب المساحات لتشمل توزيعات لم تكن معروفة من قبل.

وهذا البحث يُجري كما ذكرنا مقارنة عددية بين طرق تقريب المساحات، وقد تمَّت خطوات البحث وفق التالي :

١ ـ تمَّ إستعراض ثلاث طرق لتقريب المساحة $\alpha(x) = \int_x^\infty f(z)\,dz,$ (١)
وهم

- الطريقة الأولى وتعتمد على القانون الذي أوجده أندروز.

(٢) $$\hat{\alpha}_0(x) = [f(x)/g(x)][(K-1)]^{-1}\,[1 + \tfrac{1}{2}\,(K(x) - K)]$$

حيث $$g(x) = f'(x)/\alpha(x)$$

و $$K(x) = g'(x)/g^2(x)$$

- الطريقة الثانية وتعتمد على تعميمين للقانون أوجدهما غروس وهوسمر.

التعميم الأول هو القانون.

(٣) $$S_1(x) = [f(x)/g(x)]\;[1/(K(x) - 1)],$$

والتعميم الثاني هو القانون.

(٤) $$S_2(x) = -\;\{f(x)/g(x)\}[\{1-2K(x)\}/\{1-3K(x) + g''(x)/(g(x))^3\}].$$

- الطريقة الثالثة وتعتمد على القانون الذي أوجده عوض وشايب وهو :

(٥) $$\hat{\alpha}_4(x) = [f(x)/g(x)]\;[1/(K - 1)], \qquad K \neq 1$$

والسؤال الذي طرحه الباحث هو: ما هي عائلة الكثافات $f(X)$ والتي من أجلها يكون $\alpha(X) = \hat{\alpha}_1(x)$. وقد أجاب الباحث على هذا التساؤل ببرهان ما يلي :

من أجل التوزيع F، وحيث n_2, n_1 هما درجتا الحرية في البسط والمقام على الترتيب، يكون لدينا ما يلي:

(١) إذا كان $n_1 = 2$ وكان n_2 أي عدد صحيح موجب، فإنه يكون $K(X) = 2/(2 + n_2)$، أي تكون $K(X)$ مستقلة عن X.

(٢) إذا كان $n_1 = n_2 = 2$ ، فإن $S_2(X)$ المعطاة في المعادلة (٤) تصبح غير معرَّفة، وتصبح كذلك غير معرفة $\hat{\alpha}_5(X)$ المعطاة بِـ .

$$(٦) \qquad \hat{\alpha}_5(x) = -[f(x)/g(x)].\ [\{1 - 2K\}/\{(1-3\ K + g''(x)/(g(x))^3\})],$$

وهذا هو القانون الذي اعتمده الباحث في الطريقة الرابعة.

(٣) ثمَّ تنظيم الجداول ١ ، ٢ ، ٣ ، ٤ ، ثم أجريت مقارنة عددية (شملت توزيع t وتوزيع X^2 وتوزيع F ، بين قيم هذه الجداول.

(٤) بيَّن الباحث أن الطريقة التي قدَّمها عوض وشايب هي أسهل الطرق إستخداماً.