

THE DISTRIBUTION FUNCTION OF A LINEAR COMBINATION OF CHI-SQUARES

P. G. MOSCHOPOULOS and W. B. CANADA

University of Texas at Dallas, Box 830688, Richardson, TX, 75080-0688, U.S.A.

(Received February 1984)

Communicated by P. Odell

Abstract—The distribution function of a linear combination of independent central chi-square random variables is obtained in a straightforward manner by inverting the moment generating function. The distribution is expressed as an infinite gamma series whose terms can be computed efficiently to a sufficient degree of accuracy.

1. INTRODUCTION

We are concerned with the computation of the distribution function of a finite linear combination of independent central chi-square random variables. Let $\mathbf{c} = (c_1, \dots, c_p)'$ be a vector of real non-zero constants and $\mathbf{n} = (n_1, \dots, n_p)'$ be a vector of integers such that $n_i \geq 1, i = 1, \dots, p$. The random variable of interest is $Q(\mathbf{c}, \mathbf{n})$ where

$$Q(\mathbf{c}, \mathbf{n}) = \sum_{i=1}^p c_i \chi^2(n_i) \tag{1.1}$$

where $\chi^2(n_i), i = 1, \dots, p$ are independent chi-square random variables with n_i degrees of freedom.

The distribution of $Q = Q(\mathbf{c}, \mathbf{n})$ is important in a variety of problems in statistical inference and applied probability. A survey of applications is found in Jensen and Solomon [4].

Exact series representations for the distribution of Q are found in Ruben [9], Kotz *et al.* [5] and in Mathai [6]. The use of these representations for computational purposes thus far has been limited to small values of p ; for large values the computations tend to become unfeasible. Alternative computational methods include numerical inversion of the characteristic function of Q , Imhof [3] and Rice [8]; a differential equation approach, Davis [1]; various moments-based approximations reviewed in Solomon and Stephens [10], and most recently the method of negative binomial mixtures, Oman and Zacks [7].

Oman and Zacks [7] point out that all methods available in the literature either demand lengthy computations or are insufficiently accurate while their method reduces the amount of computation and produces sufficiently accurate results. Indeed, their numerical comparisons indicate that the percentiles of Q computed by their method are the most accurate as compared with Imhof's exact percentiles.

In this paper we present yet another method for computing the distribution function of Q which is computationally very efficient, better or at least as accurate as the mixture method and easily programmed.

2. THE EXACT DENSITY AND DISTRIBUTION FUNCTION OF Q

By inverting the moment generating function $M(t)$ of Q we will obtain the distribution function of Q as an infinite series of incomplete gamma integrals. It is well known that

$$M(t) = \prod_{i=1}^p (1 - 2c_i t)^{-m_i}, \quad m_i = n_i/2. \tag{2.1}$$

The following identity is reported in Mathai [6],

$$1 - 2c_1t = (1 - 2c_1t)(c_i/c_1)[1 - (1 - c_1/c_i)/(1 - 2c_1t)]$$

from which, for $|(1 - c_1/c_i)/(1 - 2c_1t)| < 1$ and $i = 2, \dots, p$ we obtain the following asymptotic expansion where $(m)_0 = 1$, $(m)_r = m(m+1) \cdots (m+r-1)$,

$$(1 - 2c_1t)^{-m_i} = (c_1/c_i)^{m_i} \sum_{r=0}^{\infty} \frac{(m_i)_r}{r!} (1 - c_1/c_i)^r (1 - 2c_1t)^{-(r+m_i)}. \quad (2.2)$$

A sufficient condition for this expansion is that $t < \min(1/2c_i)$. For convenience we let

$$b_i = (c_1/c_i)^{m_i} \text{ and } A(c_i, r) = (m_i)_r (1 - c_1/c_i)^r / r!. \quad (2.3)$$

On multiplying the $p-1$ series in (2.2), we obtain $M(t)$ as an infinite series in $(1 - 2c_1t)^{-1}$, i.e.

$$M'(t) = \left(\sum_{i=2}^p b_i \right) \sum_{j=0}^{\infty} a_j (1 - 2c_1t)^{-(s+j)}, \quad s = \sum_{i=1}^p m_i \quad (2.4)$$

where the a_j 's satisfy the relation

$$\prod_{i=2}^p \left[\sum_{r=0}^{\infty} A(c_i, r) x^{-r} \right] = \sum_{j=0}^{\infty} a_j x^{-j},$$

and as such are easily computed recursively from the following numerically stable formulae.

$$a_j = A_j^{(p)}, \quad A_j^{(i)} = \sum_{k=0}^j A_k^{(i-1)} A(c_i, j-k), \quad (2.5)$$

where $i = 3, 4, \dots, p$, $j = 0, 1, 2, \dots$, and for $r = 0, 1, 2, \dots$,

$$A_r^{(2)} = A(c_2, r) \quad (\text{see (2.3)}).$$

We now invert (2.4) term-by-term. Since the density corresponding to the factor $(1 - 2c_1t)^{-(s+j)}$ is the gamma density $g_j(y)$ where

$$g_j(y) = y^{s+j-1} e^{-y/2c_1} / (2c_1)^{s+j} \Gamma(s+j),$$

the distribution function $F(w) = \Pr(Q \leq w)$ is

$$F(w) = \left(\prod_{i=2}^p b_i \right) \sum_{j=0}^{\infty} a_j \int_0^w g_j(y) dy \quad (2.6)$$

where the a_j 's are very efficiently computed from (2.5). Thus, (2.6) is easily used for computational purposes, since there are a number of accurate routines available for the computation of the incomplete gamma function.

3. NUMERICAL COMPARISONS

The accuracy of the representation (2.6) was numerically evaluated by comparing probabilities and percentiles of the distribution of Q to their exact values. Exact percentiles of Q corresponding to the probabilities and the cases (combinations of c_i 's and n_i 's) of Table 1, are reported in Oman and Zacks [7] and other articles. These percentiles are denoted by I , the ones

Table 1. Percentage points using Imhof's method (*I*), mixture method (*M*) and (2.6) (*Q*)

$c_1(n_1), \dots, c_p(n_p)$	Method	-----							
		0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
0.3(2), 0.2(1), 0.1(1), 0.05(2)	I	0.116	0.167	0.224	0.309	1.902	2.333	2.757	3.314
	M	0.115	0.166	0.223	0.308	1.903	2.332	2.756	3.310
	Q	0.116	0.167	0.224	0.309	1.902	2.333	2.757	3.314
0.4(1), 0.3(1), 0.15(1), 0.05(3)	I	0.104	0.151	0.203	0.282	1.970	2.464	2.961	3.624
	M	0.104	0.150	0.202	0.281	1.974	2.464	2.962	3.616
	Q	0.104	0.151	0.203	0.282	1.970	2.464	2.961	3.623
0.4(1), 0.25(1), 0.15(1), 0.1(1), 0.05(2)	I	0.113	0.162	0.217	0.300	1.928	2.400	2.878	3.520
	M	0.111	0.160	0.214	0.296	1.930	2.397	2.879	3.506
	Q	0.112	0.162	0.217	0.300	1.928	2.400	2.878	3.519
0.4(1), 0.2(2), 0.1(1), 0.05(2)	I	0.114	0.164	0.220	0.303	1.918	2.382	2.852	3.488
	M	0.113	0.162	0.218	0.301	1.918	2.380	2.849	3.483
	Q	0.114	0.164	0.220	0.303	1.918	2.382	2.852	3.487
0.5(1), 0.2(1), 0.1(2), 0.05(2)	I	0.105	0.151	0.203	0.280	1.985	2.544	3.126	3.923
	M	0.103	0.149	0.201	0.278	1.985	2.541	3.124	3.920
	Q	0.105	0.151	0.203	0.280	1.985	2.544	3.126	3.923
0.3(1), 0.2(2), 0.1(1), 0.05(4)	I	0.166	0.223	0.283	0.369	1.808	2.188	2.565	3.061
	M	0.164	0.221	0.281	0.367	1.809	2.187	2.564	3.056
	Q	0.166	0.222	0.283	0.369	1.807	2.188	2.565	3.062
0.3(1), 0.2(1), 0.1(4), 0.05(2)	I	0.178	0.238	0.301	0.389	1.767	2.130	2.494	2.978
	M	0.177	0.236	0.299	0.387	1.770	2.128	2.494	2.971
	Q	0.178	0.238	0.301	0.389	1.767	2.130	2.493	2.978
0.4(1), 0.2(1), 0.1(2), 0.05(4)	I	0.157	0.211	0.269	0.350	1.856	2.310	2.776	3.412
	M	0.155	0.209	0.266	0.348	1.857	2.307	2.778	3.405
	Q	0.157	0.211	0.269	0.350	1.856	2.310	2.776	3.411
0.3(1), 0.25(1), 0.1(1), 0.05(7)	I	0.201	0.259	0.318	0.401	1.777	2.161	2.546	3.056
	M	0.201	0.258	0.318	0.400	1.778	2.162	2.547	3.055
	Q	0.201	0.259	0.318	0.401	1.777	2.161	2.546	3.056
0.3(1), 0.15(1), 0.1(3), 0.05(5)	I	0.217	0.279	0.342	0.428	1.713	2.057	2.406	2.879
	M	0.217	0.278	0.341	0.427	1.714	2.056	2.407	2.877
	Q	0.217	0.279	0.342	0.428	1.713	2.057	2.406	2.879
0.5(1), 0.4(1), 0.1(1)	I	0.032	0.060	0.097	0.164	2.187	2.818	3.450	4.290
	M	0.031	0.059	0.098	0.164	2.187	2.818	3.451	4.292
	Q	0.031	0.059	0.098	0.164	2.187	2.818	3.450	4.291
0.5(1), 0.3(1), 0.2(1)	I	0.036	0.068	0.110	0.183	2.122	2.708	3.302	4.102
	M	0.036	0.067	0.110	0.183	2.121	2.708	3.303	4.105
	Q	0.036	0.067	0.110	0.183	2.121	2.708	3.302	4.102
1.5(2), 0.5(2)	I	0.259	0.424	0.627	0.947	8.120	10.202	12.282	15.032
	M	0.259	0.425	0.627	0.947	8.120	10.203	12.283	15.040
	Q	0.259	0.425	0.627	0.947	8.120	10.202	12.283	15.032
2.5(1), 0.5(3)	I	0.225	0.368	0.545	0.826	8.540	11.342	14.279	18.294
	M	0.224	0.368	0.545	0.826	8.540	11.343	14.280	18.296
	Q	0.224	0.368	0.545	0.826	8.540	11.342	14.280	18.294
1.8(2), 0.6(1), 0.4(2)	I	0.449	0.677	0.943	1.347	9.921	12.419	14.913	18.213
	M	0.447	0.677	0.943	1.347	9.921	12.419	14.915	18.216
	Q	0.447	0.677	0.943	1.347	9.921	12.419	14.914	18.213
3.0(1), 0.5(4)	I	0.404	0.609	0.848	1.211	10.407	13.778	17.308	22.127
	M	0.403	0.609	0.848	1.211	10.407	13.778	17.307	22.129
	Q	0.403	0.609	0.848	1.211	10.407	13.778	17.307	22.127
0.2(4), 0.1(2)	I	0.135	0.197	0.263	0.356	1.800	2.147	2.483	2.914
	M	0.139	0.198	0.263	0.355	1.801	2.148	2.484	2.916
	Q	0.139	0.198	0.263	0.355	1.801	2.148	2.483	2.914
0.4(1), 0.2(1), 0.1(4)	I	0.122	0.178	0.238	0.324	1.879	2.332	2.799	3.429
	M	0.125	0.178	0.237	0.323	1.879	2.333	2.798	3.434
	Q	0.125	0.178	0.237	0.323	1.879	2.332	2.799	3.432
0.3(1), 0.2(1), 0.1(2), 0.05(6)	I	0.208	0.268	0.331	0.416	1.741	2.101	2.463	2.950
	M	0.209	0.268	0.330	0.414	1.746	2.105	2.464	2.940
	Q	0.210	0.270	0.331	0.416	1.741	2.100	2.463	2.947
0.4(1), 0.1(3), 0.05(6)	I	0.199	0.257	0.317	0.398	1.786	2.228	2.694	3.336
	M	0.201	0.258	0.317	0.398	1.787	2.229	2.694	3.334
	Q	0.201	0.258	0.317	0.398	1.787	2.229	2.694	3.333

derived by the mixture method of Oman and Zacks by *M* and ours by *Q*. As the table indicates *Q* is closer to *I* than *M* is. Our percentiles were computed by fitting a quadratic curve to (2.6) and then computing the inverse function at the indicated probability. We should point out that our method is more appropriate for the computation of probabilities rather than percentiles. The probabilities corresponding to the *I*-entries of Table 1 are easily obtained using (2.6) and the errors occur in the fifth decimal place.

The integral in (2.6) was evaluated using the IMSL routine MDGAM on an IBM 4341 computer at the University of Texas at Dallas. The number of terms used in (2.6) varied from about 30 (in the lower percentiles) to about 40 (in the upper percentiles), however, the CPU time was absolutely negligible due to the stability of (2.5). The Fortran code is available from the authors upon request.

REFERENCES

1. A. W. Davis, A differential equation approach to linear combinations of independent chi-squares. *J. Am. Statist. Assoc.* **72**, 212–214 (1977).
2. A. Grad and H. Solomon, Distribution of quadratic forms and some applications. *Ann. Math. Statist.* **25**, 464–472 (1955).
3. J. P. Imhof, Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 417–426 (1961).
4. D. R. Jensen and H. Solomon, A Gaussian approximation to the distribution of a definite quadratic form. *J. Am. Statist. Assoc.* **67**, 898–902 (1972).
5. S. Kotz, N. L. Johnson and D. W. Boyd, Series representations of distributions of quadratic forms in normal variables. I. Central case. *Ann. Math. Statist.* **38**, 823–837 (1967).
6. A. M. Mathai, Storage capacity of a dam with gamma type inputs, *Ann. Inst. Statist. Math.* **34(3)A**, 591–597 (1982).
7. S. D. Oman and S. Zacks, A mixture approximation to the distribution of a weighted sum of chi-squared variables, *J. Statist. Comp. Simul.* **13**, 215–224 (1981).
8. S. O. Rice, Distribution of quadratic forms in normal random variables-evaluation by numerical integration, *SIAM, J. Sci. Statist. Comp.* **1**, 438–448 (1980).
9. H. Ruben, Probability content of regions under spherical distributions, IV: the distribution of homogeneous and nonhomogeneous quadratic functions of normal variables, *Ann. Math. Statistics* **33**, 542–570 (1962).
10. H. Solomon and M. A. Stephens, Distribution of a sum of weighted chi-square variables, *J. Am. Statist. Assoc.* **72**, 881–885 (1977).