ELSEVIER

Statistics & Probability Letters 26 (1996) 339–346

STATISTICS & PROBABILITY LETTERS

On the distribution of linear functions of independent F and U variates

Jack C. Lee*, Ling Hu

Institute of Statistics, National Chiao Tung University, 1001 TA Hsueh Road, Hsinchu, Taiwan

Received October 1994

Abstract

This paper is concerned with the distributions of linear functions of independent U and F variates. The statistics $U_{p,q,n}$ is defined as $U = |Q_1|/|Q_1 + Q_2|$, where Q_1 and Q_2 are $p \times p$ random matrices and independently distributed as $W(\Sigma, n)$ and $W(\Sigma, q)$, respectively. Useful and accurate approximations are considered for the linear combinations of two independent U variates as well as the linear combinations of two independent F variates.

1. Introduction

This paper is concerned with the distributions of linear functions of independent U and F variates. The statistic $U_{p,q,n}$ is defined as

$$U = \frac{|Q_1|}{|Q_1 + Q_2|}, \tag{1.1}$$

where Q_1 and Q_2 are $p \times p$ random matrices and independently distributed as $W(\Sigma, n)$ and $W(\Sigma, q)$, respectively. The statistic is very well known in multivariate analysis and its distribution has been well studied, see e.g. Krishnaiah and Lee (1980). When $p = 1$ or 2, some functions of the U statistic have F distributions, see, e.g. Anderson (1971). Specifically, the distribution of $F = \frac{(1 - U_{1,q,n})}{U_{1,q,n}}$ is $F(q, n)$ and the distribution of $F = \frac{(n-1)(1-\sqrt{U_{2,q,n}})}{q\sqrt{U_{2,q,n}}}$ is $F(2q, 2(n-1))$. Hence, we will mainly focus on the situation in which $p > 2$ and $q \geqslant p$. However, there are occasions in which transformations of U will not be appropriate for the problem at hand. Hence, we will consider $p = 1$ and 2 as well. Of course, we should also keep in mind that the distribution of $U_{p,q,n}$ is the same as the distribution of $U_{q,p,n-p+q}$.

*Corresponding author. Research supported in part by NSC Grant 84-2121-M009-008 of ROC.

0167-7152/96/$12.00
SSDI 0167-7152(95)00030-5

Let U_1 and U_2 be two independent U variates, and a_1 and a_2 are two positive constants. The distributions of $a_1U_1 + a_2U_2$ or their special cases $a_1F_1 + a_2F_2$ have been enountered repeatedly by Geisser (1970, 1963) in dealing with Bayesian analysis of growth curve model and in multivariate analysis of variance for a special covariance matrix. Morrison (1971) studied the distribution of $a_1F_1 + a_2F_2$. However, he was concerned with cases in which a_1 and a_2 are quite restricted. We intend to study the distributional problem for arbitrary positive a_1 and a_2 in the distribution of $a_1F_1 + a_2F_2$. For the distribution of $a_1U_1 + a_2U_2$, we are not aware of attempts to this problem. The approximation considered in this paper will prove to be useful in practice for growth curve model prediction, for multivariate analysis of variance as well as for other occasions in which the linear combination of independent U variates or F variates is a natural consequence of the theoretical development.

Section 2 is devoted to the study of an approximation to the linear combination of two independent F variates. In Section 3, an approximation to the linear combination of two independent U variates is proposed. Finally, some concluding remarks are given in Section 4.

2. The distribution of linear combinations of two independent F variates

Morrison (1971) considered the distribution of the linear compound

$$W = v_1F(v_1,v_2) + \frac{u_1v_2}{u_2}F(u_1,u_2) \tag{2.1}$$

of two independent F variates with degrees of freedom v_1, v_2 and u_1, u_2, respectively. When $v_1 = n$, $v_2 = N - n$, $u_1 = m$, $u_2 = N - m$, Morrison (1971) showed that the density of W is

$$f(w) = \frac{2\Gamma\left(\frac{N}{2}\right)(N-n)^{-(m+n)/2}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{N-m}{2}\right)}\sum_{j=0}^{\infty}\frac{\Gamma\left(\frac{N}{2}+j\right)}{j!\,[2(N-n)]^{2j}}$$
$$\times \int_0^{\pi/2}\sin^{n-1}\theta\cos^{m-1}\theta\cos^{2j}2\theta\,\mathrm{d}\theta\,\frac{w^{2j+(m+n-2)/2}}{\left[1+\dfrac{w}{2(N-n)}\right]^{N+2j}} \tag{2.2}$$

which is (1.7) of Morrison (1971). Due to the complication involved in (2.2), he also proposed to approximate the distribution of W by $\eta F(v_1 + u_1, u)$ where the scale factor η and the second degree of freedom u are found by equating the first two cumulants of that variate with those of W.

Some comments are in order. First of all, we note that the linear compound considered by Morrison is very restricted in that the coefficient of $F(u_1, u_2)$ is of a special form. The analytic result is for even stricter situation. Also, the approximation proposed is a two-moment approximation with the new F variate having a special restriction in one of the two degrees of freedom, i.e., the first degree of freedom is $v_1 + u_1$.

In this paper we propose to relax these restrictions and consider an arbitrary linear compound

$$\tilde{W} = a_1F(v_1,v_2) + a_2F(u_1,u_2) \tag{2.3}$$

of two independent F variates with the degrees of freedom as indicated. Here a_1, a_2 are arbitrary positive constants, and hence there are no unnecessary restrictions on them. There are no restrictions on v_1 and u_1 either. However, in order to ensure the existence of the 3rd moment, it is required that $v_2 > 6$ and $u_2 > 6$. We now consider the approximation of the distribution of $\tilde{W}$ by $\eta F(w_1, w_2)$, where the parameters η, w_1, w_2 are obtained by equating the first three moments of this new statistic with those of $\tilde{W}$. In fact, these parameters can be expressed in explicit forms in terms of a_1, a_2, v_1, v_2, u_1, and u_2. Specifically,

$$\eta = \frac{2A^2C - 2AB^2}{A^2B + 3AC - 4B^2}, \tag{2.4a}$$

$$w_1 = \frac{4A^2C - 4AB^2}{AB^2 - 2A^2C + BC}, \tag{2.4b}$$

$$w_2 = \frac{2A^2B + 6AC - 8B^2}{A^2B + AC - 2B^2}, \tag{2.4c}$$

where

$$A = a_1 \frac{v_2}{v_2 - 2} + a_2 \frac{u_2}{u_2 - 2},$$

$$B = a_1^2 \frac{v_2^2(v_1 + 2)}{v_1(v_2 - 2)(v_2 - 4)} + 2a_1a_2 \frac{v_2u_2}{(v_2 - 2)(u_2 - 2)} + a_2^2 \frac{u_2^2(u_1 + 2)}{u_1(u_2 - 2)(u_2 - 4)},$$

$$C = a_1^3 \frac{v_2^3(v_1 + 2)(v_1 + 4)}{v_1^2(v_2 - 2)(v_2 - 4)(v_2 - 6)} + 3a_1^2a_2 \frac{v_2^2u_2(v_1 + 2)}{v_1(v_2 - 2)(v_2 - 4)(u_2 - 2)}$$

$$+ 3a_1a_2^2 \frac{v_2u_2^2(u_1 + 2)}{u_1(v_2 - 2)(u_2 - 2)(u_2 - 4)} + a_2^3 \frac{u_2^3(u_1 + 2)(u_1 + 4)}{u_1^2(u_2 - 2)(u_2 - 4)(u_2 - 6)}, \tag{2.5}$$

and $v_2 > 6$, $u_2 > 6$.

Due to the explicit formula given in (2.4), the approximation proposed in this paper is very easy to use. In order to assess the accuracy of this approximation, we compare our results with those of Morrison (1971) for the special situations considered in his paper. These comparisons are summarized in Table 1. In the table, we show the exact probabilities of exceeding the approximate upper 1% and 5% points of $\tilde{W}$ by applying (2.2). The approximations being compared are those of Morrison (1971) and the method proposed in this paper. From the table it is clear that our approximation is better than that proposed by Morrison. Of course, our method is much more general than Morrison's approximation because of the restrictions imposed in his study. For the general situations not applicable in Morrison (1971), we have also conducted an extensive simulation study and the results are summarized in Table 2.

Table 1
Probabilities of Morrison's and our approximations for upper 1% and 5% points

Linear compound	Method	$\alpha = 0.01$	$\alpha = 0.05$
$F(1,9) \to F(1,9)$	Morrison's	0.01025	0.04897
	Ours	0.00980	0.05068
$F(1,9) + \frac{9}{4}F(2,8)$	Morrison's	0.00959	0.04836
	Ours	0.00969	0.05088
$F(1,9) + \frac{27}{7}F(3,7)$	Morrison's	0.00866	0.04748
	Ours	0.00974	0.05078
$2F(2,8) + 2F(2,8)$	Morrison's	0.00916	0.04817
	Ours	0.00959	0.05092

Table 2
The simulation probabilities of exceeding upper 1%, 5% and 10% points for linear combinations of F variates

Linear compound	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\frac{1}{2}F(5,10) + \frac{1}{2}F(10,20)$	0.010	0.048	0.099
$\frac{1}{2}F(5,10) + \frac{1}{3}F(10,20)$	0.010	0.046	0.100
$\frac{1}{2}F(5,10) + \frac{2}{3}F(10,20)$	0.009	0.049	0.100
$F(1,9) + F(1,9)$	0.009	0.051	0.099
$5F(5,10) + 10F(10,10)$	0.010	0.050	0.098
$5F(5,10) + 5F(10,20)$	0.010	0.049	0.100
$5F(5,15) + \frac{15}{2}F(10,20)$	0.010	0.051	0.101
$5F(5,15) + \frac{15}{4}F(10,40)$	0.010	0.049	0.099

In the simulation, we have conducted 5000 runs for each linear combination and computed the probabilities of exceeding the approximate 1%, 5% and 10% points of $\tilde{W}$. From this table, we see that the approximation is generally quite good.

From Tables 1 and 2, it is fair to conclude that the proposed approximation to the distribution of an arbitrary linear combination of two independent F variates is quite adequate for practical purposes.

3. The distribution of linear functions of two independent U variates

In this section we will consider some approximations to the distributions of linear compound

$$V = a_1 U_1 + a_2 U_2 \tag{3.1}$$

of U_1 and U_2 which are independently distributed as U_{p,m_1,n_1} and U_{p,m_2,n_2}, respectively. For $p = 1$, we will approximate the distribution of V by $\eta U_{p,m,n}$, where η, m, and n are obtained by equating the first three moments of this statistic with those of V. These new parameters are expressed in explicit forms in terms of $a_1, a_2, p, m_1, m_2, n_1$, and n_2. More specifically,

$$\eta = \frac{2A^2C - AB^2 - BC}{A^2B - 2B^2 + AC}, \tag{3.2a}$$

$$m = \frac{4(B^2 - AC)(A^2 - B)(C - AB)}{(2A^2C - AB^2 - BC)(A^2B - 2B^2 + AC)}, \tag{3.2b}$$

$$n = \frac{4A(B^2 - AC)}{2A^2C - AB^2 - BC}, \tag{3.2c}$$

where

$$A = aEU_1 + bEU_2,$$
$$B = a^2EU_1^2 + 2abEU_1EU_2 + b^2EU_2^2,$$
$$C = a^3EU_1^3 + 3a^2bEU_1^2EU_2 + 3ab^2EU_1EU_2^2 + b^3EU_2^3, \tag{3.3}$$

$$EU_1^h = \prod_{i=1}^{p} \frac{\Gamma\left(\frac{n_1+1-i}{2} + h\right)\Gamma\left(\frac{m_1+n_1+1}{2} - i\right)}{\Gamma\left(\frac{n_1+1-i}{2}\right)\Gamma\left(\frac{m_1+n_1+1-i}{2} + h\right)},$$

Table 3
The simulation probabilities of exceeding upper 1%, 5% and 10% points for linear combinations of $U_{1,m,n}$ variates

Linear compound	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\frac{1}{2}U_{1,2,12} + \frac{1}{2}U_{1,2,15}$	0.006	0.046	0.108
$\frac{1}{2}U_{1,2,12} + \frac{1}{3}U_{1,2,15}$	0.009	0.049	0.107
$\frac{1}{2}U_{1,2,12} + \frac{2}{3}U_{1,2,15}$	0.007	0.049	0.109
$U_{1,2,12} + U_{1,2,15}$	0.007	0.053	0.099
$\frac{1}{2}U_{1,3,10} + \frac{1}{2}U_{1,3,15}$	0.006	0.045	0.116
$\frac{1}{2}U_{1,3,10} + \frac{1}{3}U_{1,3,15}$	0.006	0.045	0.113
$\frac{1}{2}U_{1,3,10} + \frac{2}{3}U_{1,3,15}$	0.007	0.049	0.113
$U_{1,3,10} + U_{1,3,15}$	0.008	0.051	0.101
$\frac{1}{2}U_{1,3,12} + \frac{1}{2}U_{1,3,15}$	0.009	0.044	0.105
$\frac{1}{2}U_{1,3,12} + \frac{1}{3}U_{1,3,15}$	0.010	0.047	0.107
$\frac{1}{2}U_{1,3,12} + \frac{2}{3}U_{1,3,15}$	0.009	0.042	0.099
$U_{1,3,12} + U_{1,3,15}$	0.010	0.050	0.105
$\frac{1}{2}U_{1,3,15} + \frac{1}{2}U_{1,3,15}$	0.010	0.044	0.102
$\frac{1}{2}U_{1,3,15} + \frac{1}{3}U_{1,3,15}$	0.011	0.048	0.095
$\frac{1}{2}U_{1,3,15} + \frac{2}{3}U_{1,3,15}$	0.012	0.042	0.099
$U_{1,3,15} + U_{1,3,15}$	0.010	0.046	0.101
$\frac{1}{2}U_{1,3,15} + \frac{1}{2}U_{1,3,20}$	0.012	0.043	0.115
$\frac{1}{2}U_{1,3,15} + \frac{1}{3}U_{1,3,20}$	0.013	0.044	0.108
$\frac{1}{2}U_{1,3,15} + \frac{2}{3}U_{1,3,20}$	0.012	0.043	0.117
$U_{1,3,15} + U_{1,3,20}$	0.008	0.049	0.096

and

$$EU_2^h = \prod_{i=1}^{p} \frac{\Gamma\left(\frac{n_2+1-i}{2}+h\right)\Gamma\left(\frac{m_2+n_2+1}{2}-i\right)}{\Gamma\left(\frac{n_2+1-i}{2}\right)\Gamma\left(\frac{m_2+n_2+1-i}{2}+h\right)}, \tag{3.4}$$

for $h = 1, 2, 3$. In order to assess the accuracy of the above approximation, we have conducted a simulation study and the results are summarized in Table 3. In the simulation, we have conducted 5000 runs for each combination of the parameters and computed the probabilities of exceeding the approximate 1%, 5% and 10% points of V. From the table we see that the proposed approximation is quite reasonable.

We next consider the more general case of $p \geqslant 2$ in which the above approximation does not work. Instead of the $\eta U_{p,m,n}$ approximation to the distribution of V, we will approximate the distribution by the Pearson type I distribution which is defined as

$$f(x) = [\beta(\alpha+1, \varepsilon+1)(\sigma_1-\sigma_0)^{\alpha+\varepsilon+1}]^{-1}(x-\sigma_0)^{\alpha}(\sigma_1-x)^{\varepsilon}, \quad \sigma_0 \leqslant x \leqslant \sigma_1, \quad \alpha \in R. \tag{3.5}$$

With the first four moments of multivariate testing statistics, useful Pearson type I approximations have been obtained by Krishnaiah et al. (1976), Krishnaiah and Lee (1980), among others. The usefulness of the Pearson curves in density estimation has also been demonstrated by Solomon and Stephens (1978) and others.

We now need the first four moments of V. For given a_1 and a_2, it is well known that

$$\begin{aligned} E(V^h) &= E(a_1 U_1 + a_2 U_2)^h \\ &= \sum_{k=0}^{h} C_k^h a_1^k a_2^{h-k} EU_1^k EU_2^{h-k}, \quad h = 1, 2, 3, 4, \end{aligned} \tag{3.6}$$

where EU_i^k are given in (3.4). Let $\mu = E(V)$ and $\mu_h = E(V-\mu)^h$, for $h = 2, 3, 4$ and $\beta_1 = \mu_3^2/\mu_2^3$, $\beta_2 = \mu_4/\mu_2^2$. Then the type I distribution requires that

$$6 + 3\beta_1 - 2\beta_2 > 0, \qquad \beta_2 - \beta_1 - 1 > 0. \tag{3.7}$$

Instead of computing from the density directly, we will make use of the tables produced by Johnson et al. (1963). In order to do so, we need the following double entry interpolation. Linear interpolation is often possible for β_2, while second differences are needed for $\sqrt{\beta_1}$. The procedure is to interpolate first for β_2, at each of the nearest four values of $\sqrt{\beta_1}$, to tabulate the nearest four values x_{-1}, x_0, x_1, x_2 and then interpolate for $\sqrt{\beta_1}$, using the formula

$$x(\theta) = (1-\theta)x_0 + \theta x_1 - \tfrac{1}{4}\theta(1-\theta)[\delta^2 x_0 + \delta^2 x_1], \tag{3.8}$$

where θ is the appropriate fraction of the tabular interval.

We next illustrate the approximation to the distribution of $\frac{1}{2} U_{4,2,12} + \frac{1}{2} U_{4,2,15}$. Using (3.4) and (3.6) we obtain $\mu_1 = 0.534017$, $\sqrt{\mu_2} = 0.112554$, $\sqrt{\beta_1} = 0.0373722$ and $\beta_2 = 2.74613$. From the tables of Johnson et al. (1963), we have Table 4 which will enable us to obtain the critical value from (3.8) with $\alpha = 0.01$.

Using the interpolation formula (3.8) with $\sqrt{\beta_1} = 0.0373722$, we obtain the upper 1% point as $X_{0.01} = 2.27593$. The upper 1% point of U is $\mu + \mu_2^{1/2} \cdot x_{0.01} = 0.790182$. The simulation result of this approximation will be reported later.

In order to assess the accuracy of the proposed approximation, we will conduct an extensive simulation study. For the simulation, we recall that

$$U_{p,q,n} = \frac{|Q_1|}{|Q_1 + Q_2|}, \tag{3.9}$$

where Q_1 and Q_2 are independently distributed as $W_p(\Sigma, n)$ and $W_p(\Sigma, q)$, respectively. Furthermore,

$$Q_1 = \sum_{\alpha=1}^{n} \underset{\sim\alpha}{Z}\, \underset{\sim\alpha}{Z}' \tag{3.10}$$

and

$$Q_2 = \sum_{\alpha=1}^{q} \underset{\sim\alpha}{Z}\, \underset{\sim\alpha}{Z}' \tag{3.11}$$

Table 4
Table entries needed for the example

β_2	$\sqrt{\beta_1}$ 0.0	0.1
2.4	2.1207	2.2004
2.6	2.2068	2.2833
2.8	2.2737	2.3469
3.0	2.3263	2.3964

Table 5
The simulation probabilities of exceeding upper 1%, 5% and 10% points for linear combinations of $U_{4,m,n}$ variates

Linear compound	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\frac{1}{2}U_{4,2,12} + \frac{1}{2}U_{4,2,15}$	0.009	0.048	0.099
$\frac{1}{2}U_{4,2,12} + \frac{1}{3}U_{4,2,15}$	0.006	0.046	0.096
$\frac{1}{2}U_{4,2,12} + \frac{2}{3}U_{4,2,15}$	0.009	0.048	0.093
$U_{4,2,12} + U_{4,2,15}$	0.009	0.049	0.098
$\frac{1}{2}U_{4,3,10} + \frac{1}{2}U_{4,3,15}$	0.010	0.052	0.101
$\frac{1}{2}U_{4,3,10} + \frac{1}{3}U_{4,3,15}$	0.010	0.053	0.099
$\frac{1}{2}U_{4,3,10} + \frac{2}{3}U_{4,3,15}$	0.010	0.052	0.098
$U_{4,3,10} + U_{4,3,15}$	0.011	0.052	0.095
$\frac{1}{2}U_{4,3,12} + \frac{1}{2}U_{4,3,15}$	0.009	0.049	0.095
$\frac{1}{2}U_{4,3,12} + \frac{1}{3}U_{4,3,15}$	0.010	0.050	0.101
$\frac{1}{2}U_{4,3,12} + \frac{2}{3}U_{4,3,15}$	0.010	0.049	0.099
$U_{4,3,12} + U_{4,3,15}$	0.011	0.099	0.101
$\frac{1}{2}U_{4,3,15} + \frac{1}{2}U_{4,3,15}$	0.010	0.056	0.102
$\frac{1}{2}U_{4,3,15} + \frac{1}{3}U_{4,3,15}$	0.012	0.051	0.103
$\frac{1}{2}U_{4,3,15} + \frac{2}{3}U_{4,3,15}$	0.010	0.050	0.100
$U_{4,3,15} + U_{4,3,15}$	0.011	0.056	0.103
$\frac{1}{2}U_{4,3,15} + \frac{1}{2}U_{4,3,20}$	0.007	0.048	0.105
$\frac{1}{2}U_{4,3,15} + \frac{1}{3}U_{4,3,20}$	0.011	0.049	0.099
$\frac{1}{2}U_{4,3,15} + \frac{2}{3}U_{4,3,20}$	0.008	0.047	0.098
$U_{4,3,15} + U_{4,3,20}$	0.010	0.051	0.103

where $\underset{\sim}{Z} \sim N_p(0, \Sigma)$. Since the distribution of $U_{p,q,n}$ is independent of Σ, we will set $\Sigma = \mathrm{I}$ in the simulation study. For the values of other parameters in the simulation, we set $p = 4$, $m_1 = 2$, $n_1 = 12$, $n_2 = 15$. Finally, the α levels are set at $\alpha = 0.01$, 0.05, 0.10. Also, in the simulation, we conducted 10000 runs for each combination of the parameters and computed the probabilities of exceeding the approximate 1%, 5% and 10% points of V. These results are summarized in Table 5. From this table, we see that the approximation is generally quite reasonable.

4. Concluding remarks

It is clear that the proposed approximations to the distributions of linear combinations of two independent U and F variates are quite good. No restrictions are imposed on the weights, although there is a minor restriction on the second degrees of freedom for the F variates. Since the computations involved are not heavy at all, these approximations should be very useful for practical purposes.

References

Anderson, T.W. (1971), An Introduction to Multivariate Statistical Analysis, (Wiley, New York , 2nd ed.).

Geisser, S. (1963), Multivariate analysis of variance for a special covariance case, *J. Amer. Statist. Assoc.* **58**, 660–669.

Geisser, S. (1970), Bayesian analysis of growth curves, *Sankhya*, **32**, 53–64.

Johnson, N.L., E. Nixon and D.E. Amos and E.S. Pearson (1963), Table of percentage points of Pearson curves, for given $\sqrt{\beta_1}$ and β_2 expressed in standard measure, *Biometrika* **50**, 459–498.

Krishnaiah, P.R. and J.C. Lee (1980), Likelihood ratio tests for mean vectors and covariance matrices, in: *Handbook of Statistics*, Vol. 1, pp. 513–570

Krishnaiah, P.R., J.C. Lee and T.C. Chang (1976), The distributions of the likelihood ratio statistics for tests of certain covariances structures of complex multivariate normal populations, *Biometrika* **63**, 543–550.

Morrison, D.F. (1971), The distribution of linear *F* functions of independent *F* variates, *J. Amer. Statist. Assoc.* **66**, 383–385.

Solomon, H. and M.A. Stephens (1978), Approximations to density functions using Pearson curves, *J. Amer. Statist. Assoc.* **73**, 153–160.