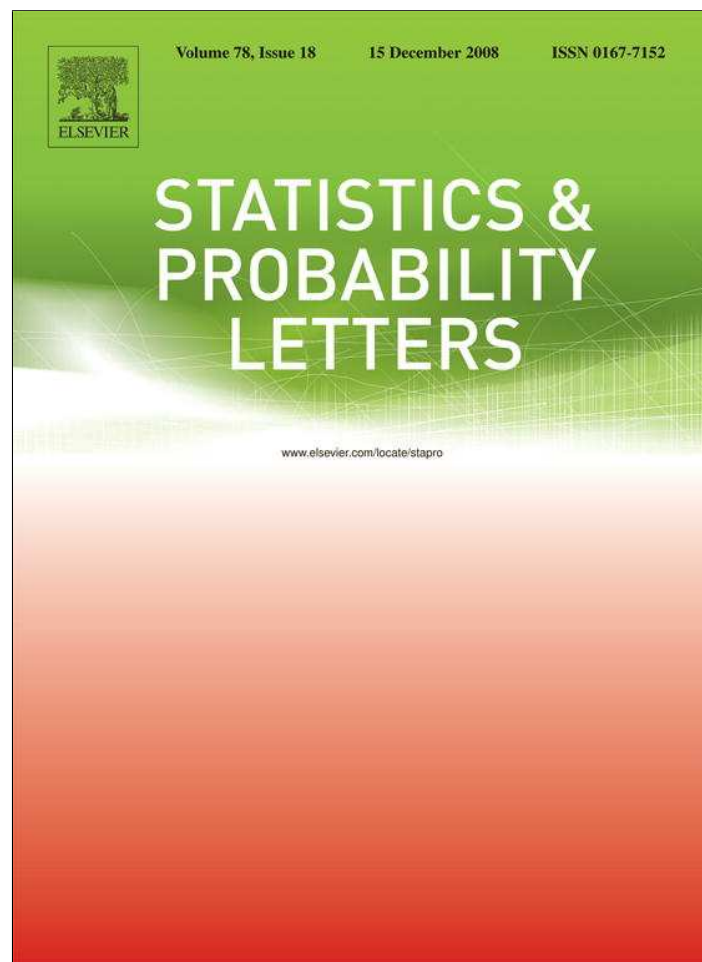




This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

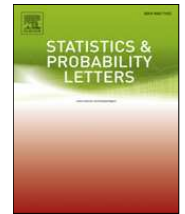
<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro



A mass function based on correlation coefficient and its application

Anwar H. Joarder, M.H. Omar*

Department of Mathematics and Statistics, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

ARTICLE INFO

Article history:

Received 5 May 2008

Received in revised form 13 June 2008

Accepted 8 July 2008

Available online 15 July 2008

ABSTRACT

A discrete distribution involving a product of two gamma functions has been proposed. It arises naturally in connection with the distribution of sample correlation coefficient based on a bivariate normal population. The first four factorial moments, raw moments, three corrected moments, coefficient of skewness and kurtosis have been derived. Some illustrations have been provided to show how the product moments of sample variances and correlation can be derived by exploiting the moments of the new distribution. These product moments are important for correlation analysis, covariance analysis, intra-sire regression and inference in bivariate normal population with a common mean.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Fisher (1915) derived the distribution of mean-centered sum of squares and sum of products in order to study the distribution of correlation coefficient from a normal sample. In this paper, based on the density identity of the product moment correlation coefficient, we introduce a new discrete distribution. The distribution is then used to derive product moments of sample variances and coefficient of correlation. Readers interested in this area may go through a resourceful book by Gupta and Nagar (2000).

In what follows, for any non-negative integer k , we will use the notation

$$k^{(a)} = k(k-1) \cdots (k-a+1). \quad (1.1)$$

The duplication formula for the gamma function is given by

$$\Gamma(2z)\sqrt{\pi} = 2^{2z-1}\Gamma(z)\Gamma\left(z + \frac{1}{2}\right). \quad (1.2)$$

Multiplying both sides by $2z$ we have

$$(2z)!\sqrt{\pi} = 2^{2z}z!\Gamma\left(z + \frac{1}{2}\right). \quad (1.3)$$

The hypergeometric function is defined by

$${}_pF_q(a_1, a_2, \dots, a_p; b_1, b_2, \dots, b_q; z) = \sum_{k=0}^{\infty} \frac{\Gamma(a_1+k)}{\Gamma(a_1)} \frac{\Gamma(a_2+k)}{\Gamma(a_2)} \cdots \frac{\Gamma(a_p+k)}{\Gamma(a_p)} \left(\frac{\Gamma(b_1+k)}{\Gamma(b_1)} \frac{\Gamma(b_2+k)}{\Gamma(b_2)} \cdots \frac{\Gamma(b_q+k)}{\Gamma(b_q)} \right)^{-1} \frac{z^k}{k!}. \quad (1.4)$$

* Corresponding author. Tel.: +966 03 860 2471; fax: +966 03 860 2340.

E-mail addresses: anwarj@kfupm.edu.sa (A.H. Joarder), omarmh@kfupm.edu.sa (M.H. Omar).

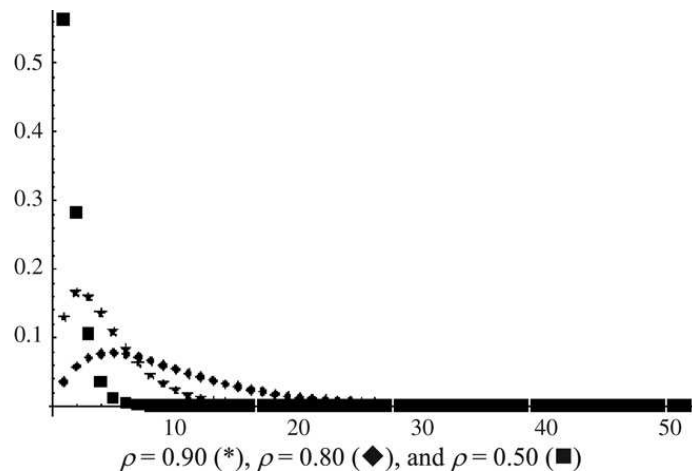


Fig. 1. Probability mass function of K when $m = 4$ at different values of ρ .

In particular

$${}_1F_0(a; z) = (1 - z)^{-a}. \quad (1.5)$$

Note that ${}_2F_1(a, b; c; z)$ can be transformed as

$${}_2F_1(a, b; c; z) = (1 - z)^{c-a-b} {}_2F_1(c - a, c - b; c; z) \quad (1.6)$$

(Gradshteyn and Ryzhik, 1994, 1069).

In Section 2, we introduce a mass function from the density function of the sample product moment correlation coefficient. In Section 3, we study some properties of the distribution. In Section 4, we demonstrate how some of the product moments of the bivariate Wishart distribution, the bivariate chi-square distribution and the product moment correlation coefficient distribution can be derived through the new mass function. These product moments have found applications in statistical inference for parameters of bivariate normal distribution (Sunthornworasiri et al., 2006), intra-sire regression (Prabhakaran and Jain, 1987a,b; Prabhakaran et al., 1991) and radar systems (Lawson and Uhlenbeck, 1950).

2. The probability mass function

The density function of the bivariate Wishart distribution was originally derived by Fisher (1915). The density function $h(r)$ of R can be obtained from the Wishart distribution (see e.g. Anderson (2003, 123)). The density identity can be defined by the following equation:

$$\int_{-1}^1 h(r) dr = 1.$$

Theorem 2.1. Let $\gamma_{k,m} = \frac{1+(-1)^k}{2} \frac{2^k}{k!} \Gamma\left(\frac{k+m}{2}\right) \Gamma\left(\frac{k+1}{2}\right)$. Then the density identity based on the distribution of the sample product moment correlation coefficient is given by

$$\sum_{k=0}^{\infty} \rho^k \gamma_{k,m} = \frac{\sqrt{\pi} \Gamma(m/2)}{(1 - \rho^2)^{m/2}}, \quad m > 2, -1 < \rho < 1. \quad (2.1)$$

For details, see Joarder and Omar (2008). Since the quantity

$$\frac{(1 - \rho^2)^{m/2}}{\sqrt{\pi} \Gamma(m/2)} \rho^k \gamma_{k,m}, \quad m > 2, -1 < \rho < 1,$$

is non-negative and adds to 1 over the support $\{0, 1, 2, \dots, \infty\}$, we define the mass function of the distribution as

$$f(k; m, \rho) = \frac{(1 - \rho^2)^{m/2}}{\sqrt{\pi} \Gamma(m/2)} \rho^k \gamma_{k,m}, \quad k = 0, 1, 2, \dots \quad (2.2)$$

$m > 2, -1 < \rho < 1$ and $\gamma_{k,m}$ is defined in Theorem 2.1 which can be explicitly written as

$$f(k; m, \rho) = \frac{(1 - \rho^2)^{m/2}}{2\sqrt{\pi} \Gamma(m/2)} [1 + (-1)^k] \frac{(2\rho)^k}{k!} \Gamma\left(\frac{k+m}{2}\right) \Gamma\left(\frac{k+1}{2}\right), \quad k = 0, 1, 2, \dots \quad (2.3)$$

$m > 2, -1 < \rho < 1$. The above mass function can also be directly summed up to 1.

A graph of the probability mass function of K is provided in Fig. 1 for various values of ρ when $m = 4$.

3. Moments of the K -distribution

Theorem 3.1. *The first four factorial moments are given by*

- (i) $E(K) = m\rho^2(1 - \rho^2)^{-1}$,
- (ii) $E(K^{[2]}) = [m(m+1)\rho^4 + m\rho^2](1 - \rho^2)^{-2}$,
- (iii) $E(K^{[3]}) = [m(m+1)(m+2)\rho^6 + 3m(m+2)\rho^4](1 - \rho^2)^{-3}$,
- (iv) $E(K^{[4]}) = [m(m+1)(m+2)(m+3)\rho^8 + 6m(m+2)(m+3)\rho^6 + 3m(m+2)\rho^4](1 - \rho^2)^{-4}$.

Proof. From (2.3), the a th factorial moment of K is given by

$$E(K^{[a]}) = \frac{(1 - \rho^2)^{m/2}}{2\sqrt{\pi}\Gamma(m/2)} \sum_{k=0}^{\infty} k^{[a]} [1 + (-1)^k] \frac{2^k \rho^k}{k!} \Gamma\left(\frac{k+m}{2}\right) \Gamma\left(\frac{k+1}{2}\right),$$

which can be written as

$$E(K^{[a]}) = \frac{(1 - \rho^2)^{m/2}}{2\sqrt{\pi}\Gamma(m/2)} \sum_{k=0}^{\infty} (2k)^{[a]} \frac{2^{2k} \rho^{2k}}{(2k)!} \Gamma\left(\frac{2k+m}{2}\right) \Gamma\left(\frac{2k+1}{2}\right).$$

Hence by (1.3), we have

$$E(K^{[a]}) = \frac{(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \sum_{k=0}^{\infty} (2k)^{[a]} \frac{\rho^{2k}}{k!} \Gamma\left(k + \frac{m}{2}\right). \quad (3.1)$$

From (3.1), the first factorial moment of the K -distribution is given by

$$\begin{aligned} E(K^{[1]}) &= \frac{2(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \sum_{k=0}^{\infty} k \frac{\rho^{2k}}{k[(k-1)!]} \Gamma\left(k + \frac{m}{2}\right) \\ &= \frac{2(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \left[\rho^2 \sum_{j=0}^{\infty} \frac{\rho^{2j}}{j!} \Gamma\left(j + 1 + \frac{m}{2}\right) \right] \\ &= \frac{2(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \left[\rho^2 \Gamma\left(1 + \frac{m}{2}\right) {}_1F_0\left(1 + \frac{m}{2}; \rho^2\right) \right] \end{aligned}$$

where we have used the generalized hypergeometric function defined by (1.4). Hence by (1.5), we have $E(K^{[1]}) = m\rho^2(1 - \rho^2)^{-1}$.

Since $(2k)^{[2]} = 4k^{[2]} + 2k$, from (3.1), the second factorial moment of the K -distribution is given by

$$E(K^{[2]}) = \frac{(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \sum_{k=0}^{\infty} [4k^{[2]} + 2k] \frac{\rho^{2k}}{k!} \Gamma\left(k + \frac{m}{2}\right),$$

which, by virtue of (1.4), can be simplified as

$$\begin{aligned} E(K^{[2]}) &= \frac{(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \left[4\rho^4 \sum_{j=0}^{\infty} \frac{\rho^{2j}}{j!} \Gamma\left(j + 2 + \frac{m}{2}\right) \right] + \frac{(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \left[2\rho^2 \sum_{j=0}^{\infty} \frac{\rho^{2j}}{j!} \Gamma\left(j + 1 + \frac{m}{2}\right) \right] \\ &= \frac{(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \left[4\rho^4 {}_1F_0\left(2 + \frac{m}{2}; \rho^2\right) \Gamma\left(2 + \frac{m}{2}\right) \right] + \frac{(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \left[2\rho^2 {}_1F_0\left(1 + \frac{m}{2}; \rho^2\right) \Gamma\left(1 + \frac{m}{2}\right) \right]. \end{aligned}$$

Hence by virtue of (1.5), we have

$$E(K^{[2]}) = \frac{(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \left[4\rho^4 \left(1 + \frac{m}{2}\right) \frac{m}{2} (1 - \rho^2)^{-2-(m/2)} \right] + \frac{(1 - \rho^2)^{m/2}}{\Gamma(m/2)} \left[2\rho^2 \frac{m}{2} (1 - \rho^2)^{-1-(m/2)} \right],$$

which simplifies to what we have in (ii) of the theorem.

Since $(2k)^{[3]} = 8k^{[3]} + 12k^{[2]}$ and $(2k)^{[4]} = 16k^{[4]} + 48k^{[3]} + 12k^{[2]}$, the third and the fourth factorial moments of the K -distribution can be derived similarly (see e.g. Joarder and Omar (2008)). $\square$

Theorem 3.2. *The first four moments of the distribution are given by*

- (i) $E(K) = m\rho^2(1 - \rho^2)^{-1}$,
- (ii) $E(K^2) = (m^2\rho^4 + 2m\rho^2)(1 - \rho^2)^{-2}$,
- (iii) $E(K^3) = [(4m\rho^2) + (6m^2 + 4m)\rho^4 + m^3\rho^6](1 - \rho^2)^{-3}$,
- (iv) $E(K^4) = [m^4\rho^8 + (12m^3 + 16m^2 + 8m)\rho^6 + (28m^2 + 32m)\rho^4 + 8m\rho^2](1 - \rho^2)^{-4}$.

Proof. Since

$$\begin{aligned} K^2 &= K^{\{2\}} + K, \\ K^3 &= K^{\{3\}} + 3K^{\{2\}} + K, \\ K^4 &= K^{\{4\}} + 6K^{\{3\}} + 7K^{\{2\}} + K, \end{aligned}$$

the moments follow from [Theorem 3.1](#). $\square$

The mean-centered moments $\mu_a = E(K - \mu)^a$, $a = 2, 3, 4$ where $\mu = E(K)$ are provided in the following theorem.

Theorem 3.3. The mean-centered moments of order 2, 3 and 4 of the K -distribution are respectively given by

$$\begin{aligned} \mu_2 &= \frac{2m\rho^2}{(1-\rho^2)^2}, \\ \mu_3 &= \frac{4m\rho^2(1+\rho^2)}{(1-\rho^2)^3} \quad \text{and} \\ \mu_4 &= \frac{4m\rho^2[2 + (3m+8)\rho^2 + 2\rho^4]}{(1-\rho^2)^4}. \end{aligned}$$

The coefficient of skewness and kurtosis of the K -distribution are given by the following moment ratio:

$$\alpha_i(K) = \frac{\mu_i}{\mu_2^{i/2}}, \quad i = 3, 4.$$

which simplifies to

$$\alpha_3(K) = \frac{\sqrt{2}(1+\rho^2)}{\sqrt{m}\rho} \quad \text{and} \quad \alpha_4(K) = \frac{2 + (3m+8)\rho^2 + 2\rho^4}{m\rho^2}.$$

Note that as $m \rightarrow \infty$, $\alpha_3(K) \rightarrow 0$ and $\alpha_4(K) \rightarrow 3$, coinciding with that of the univariate normal distribution.

4. Some applications of the distribution

The mass function can be used to derive the product moments of sample variances and coefficient of correlation, and the bivariate chi-square distribution. Product moments of sample variances and coefficient of correlation were derived by [Joarder \(2006\)](#) using some identities involving infinite sum of product of gamma functions. The general expression is provided below.

Theorem 4.1. For finite a, b , and c , the product moments $E(S_1^{2a} S_2^{2b} R^c)$ denoted by $\mu'(a, b, c; \rho)$ are given by

$$\begin{aligned} \mu'(a, b, c; \rho) &= \frac{2^{a+b-1}(1-\rho^2)^{a+b+(m/2)}}{m^{a+b}\Gamma(m/2)\sqrt{\pi}} \sigma_1^{2a} \sigma_2^{2b} \\ &\times \sum_{k=0}^{\infty} [1 + (-1)^{k+c}] \frac{(2\rho)^k}{k!} \Gamma\left(\frac{k+m}{2} + a\right) \Gamma\left(\frac{k+m}{2} + b\right) \frac{\Gamma\left(\frac{k+1+c}{2}\right)}{\Gamma\left(\frac{k+m+c}{2}\right)}, \end{aligned} \quad (4.1)$$

where $m > 2$, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$.

Now we demonstrate how the moments of the K -distribution can be exploited to derive closed form expressions of the above product moments.

(i) To calculate $E(S_1^2 S_2^2) = \mu'(1, 1, 0; \rho)$, put $a = 1$, $b = 1$ and $c = 0$ in (4.1). The resulting expression can then be simplified to

$$\mu'(1, 1, 0; \rho) = \frac{(1-\rho^2)^2 \sigma_1^2 \sigma_2^2}{m^2} \frac{(1-\rho^2)^{(m/2)}}{2\sqrt{\pi}\Gamma(m/2)} \sum_{k=0}^{\infty} [1 + (-1)^k] \frac{(2\rho)^k}{k!} (k+m)^2 \Gamma\left(\frac{k+m}{2}\right) \Gamma\left(\frac{k+1}{2}\right).$$

Since K has the mass function $f(k; m, \rho)$ given by (2.3), we have

$$\mu'(1, 1, 0; \rho) = \frac{(1-\rho^2)^2 \sigma_1^2 \sigma_2^2}{m^2} E(K+m)^2.$$

By the use of [Theorem 3.1](#)(i) and (ii) in $E(K+m)^2 = E(K^2) + 2mE(K) + m^2$, we have

$$\mu'(1, 1, 0; \rho) = \frac{1}{m} (\sigma_1 \sigma_2)^2 (m + 2\rho^2).$$

Since $U = mS_1^2/\sigma_1^2$ and $V = mS_2^2/\sigma_2^2$, the product moment $E(UV)$ of the bivariate chi-square distribution is given by $E(UV) = \frac{m^2}{\sigma_1^2\sigma_2^2}E(S_1^2S_2^2) = m + 2\rho^2$.

(ii) To calculate $E(S_1^4S_2^4) = \mu'(2, 2, 0; \rho)$, put $a = 2$, $b = 2$ and $c = 0$ in (4.1). The resulting expression can further be simplified to

$$\mu'(2, 2, 0; \rho) = \frac{m+2}{m^3}(1-\rho^2)^2\sigma_1^4\sigma_2^4 \sum_{k=0}^{\infty} (k+m+2)(k+m) \frac{(1-\rho^2)^{(m+4)/2}}{\sqrt{\pi}\Gamma(\frac{m+4}{2})} \gamma_{k,m+4}.$$

That is

$$\mu'(2, 2, 0; \rho) = \frac{m+2}{m^3}(1-\rho^2)^2\sigma_1^4\sigma_2^4 E[(K+m+2)(K+m)]$$

where K has the mass function $f(k; m+4, \rho)$ given by (2.2). By the use of Theorem 3.1(i) and (ii) in $E[(K+m+2)(K+m)]$, we have

$$\mu'(2, 2, 0; \rho) = \frac{m+2}{m^3}(1-\rho^2)^2\sigma_1^4\sigma_2^4 [E(K^2) + (2m+2)E(K) + m(m+2)]$$

which simplifies to

$$\mu'(2, 2, 0; \rho) = \frac{m+2}{m^3}(1-\rho^2)^2\sigma_1^4\sigma_2^4 [8\rho^4 + (8m+16)\rho^2 + m(m+2)]$$

which is the same as in Joarder (2006).

The product moment $E(U^2V^2)$, where $U = mS_1^2/\sigma_1^2$ and $V = mS_2^2/\sigma_2^2$ of the bivariate chi-square distribution, is given by

$$E(U^2V^2) = \frac{m^4}{\sigma_1^4\sigma_2^4}E(S_1^4S_2^4) = m(m+2)(1-\rho^2)^2[8\rho^4 + (8m+16)\rho^2 + m(m+2)].$$

(iii) To calculate $\mu'(2, 2, 2; \rho)$, put $a = 2$, $b = 2$ and $c = 2$ in (4.1). The resulting expression can further be simplified to

$$\mu'(2, 2, 2; \rho) = \frac{m+2}{m^3}(1-\rho^2)^2\sigma_1^4\sigma_2^4 \sum_{k=0}^{\infty} (k+m+2)(k+1) \frac{(1-\rho^2)^{(m+4)/2}}{\sqrt{\pi}\Gamma((m+4)/2)} \gamma_{k,m+4}.$$

Since K has the mass function $f(k; m+4, \rho)$ given by (2.2), we have

$$\mu'(2, 2, 2; \rho) = \frac{m+2}{m^3}(1-\rho^2)^2\sigma_1^4\sigma_2^4 E[(K+m+2)(K+1)].$$

Then by the use of Theorem 3.1, we have

$$\begin{aligned} E[(K+m+2)(K+1)] &= E(K^2) + (m+3)E(K) + (m+2) \\ &= [(m+4)^2\rho^4 + 2(m+4)\rho^2](1-\rho^2)^{-2} + (m+3)(m+4)\rho^2(1-\rho^2)^{-1} + (m+2), \end{aligned}$$

so that

$$\mu'(2, 2, 2; \rho) = \frac{m+2}{m^3}\sigma_1^4\sigma_2^4 [(2m+6)\rho^4 + (m^2+7m+16)\rho^2 + (m+2)]$$

(cf. Joarder (2006)).

(iv) To derive $E(R)$, we proceed as follows:

$$\mu'(0, 0, 1; \rho) = \frac{(1-\rho^2)^{m/2}}{2\Gamma(m/2)\sqrt{\pi}} \sum_{k=0}^{\infty} [1 + (-1)^{k+1}] \frac{(2\rho)^k}{k!} \Gamma^2\left(\frac{k+m}{2}\right) \frac{\Gamma\left(\frac{k+2}{2}\right)}{\Gamma\left(\frac{k+m+1}{2}\right)}.$$

By the use of (1.3), it can be checked that

$$\mu'(0, 0, 1; \rho) = \frac{(1-\rho^2)^{m/2}\rho}{\Gamma(m/2)} \sum_{k=0}^{\infty} \frac{\rho^{2k}}{k!} \Gamma^2\left(k + \frac{m+1}{2}\right) \frac{1}{\Gamma\left(k + \frac{m+2}{2}\right)}.$$

Then by using (1.4) and (1.6), we have

$$\begin{aligned} \mu'(0, 0, 1; \rho) &= \frac{(1-\rho^2)^{m/2}\rho}{\Gamma(m/2)} {}_2F_1\left(\frac{m+1}{2}, \frac{m+1}{2}; \frac{m+2}{2}; \rho^2\right) \frac{\Gamma^2\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m+2}{2}\right)} \\ &= 2(1-\rho^2)^{-m/2}\rho \frac{\Gamma^2\left(\frac{m+1}{2}\right)}{m\Gamma^2\left(\frac{m}{2}\right)} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{m+2}{2}; \rho^2\right), \end{aligned}$$

which is the same as obtained by Ghosh (1966).

Acknowledgements

The authors thank the reviewers for constructive suggestions. They also acknowledge the excellent facility at King Fahd University of Petroleum & Minerals (KFUPM), Saudi Arabia. In particular, the first author acknowledges the excellent research support provided by KFUPM through the Project FT 2004-22.

References

- Anderson, T.W., 2003. *An Introduction to Multivariate Statistical Analysis*. John Wiley.
- Fisher, R.A., 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10, 507–521.
- Ghosh, B.K., 1966. Asymptotic expansions for the moments of the distribution of correlation coefficient. *Biometrika* 53, 237–249.
- Gupta, A.K., Nagar, D.K., 2000. *Matrix Variate Distributions*. Chapman and Hall.
- Gradshteyn, I.S., Ryzhik, I.M., 1994. *Tables of Integrals, Series and Products*. Academic Press.
- Joarder, A.H., 2006. Product moments of bivariate Wishart distribution. *Journal of Probability and Statistical Science* 4 (2), 233–243.
- Joarder, A.H., Omar, M.H., 2008. A mass function based on correlation coefficient and its application. Technical Report No. 387 (April 2008) Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.
- Lawson, J.L., Uhlenbeck, G.E., 1950. *Threshold Signals*. McGraw-Hill Book Co.
- Prabhakaran, V.T., Jain, J.P., 1987a. Probability of inadmissible estimates of heritability from regression and half-sib analysis. *Biometrical Journal* 29 (2), 219–230.
- Prabhakaran, V.T., Jain, J.P., 1987b. On the variance of heritability coefficient obtained from intra-sire regression. *The Statistician* 36, 25–26.
- Prabhakaran, V.T., Mahajan, V.K., Uma, 1991. On an identity for finding moments of sample moments of bivariate normal random variables. *Australian Journal of Statistics* 33 (1), 95–102.
- Sunthornworasiri, N., Tiensuwan, M., Sinha, B.K., 2006. Statistical inference for a bivariate normal population with a common mean. *International Journal of Statistical Sciences* 5, 9–26.