ON THE SAMPLE COVARIANCE FROM A BIVARIATE NORMAL DISTRIBUTION

S. JAMES PRESS

(Received Aug. 23, 1966)

1. Introduction

The purpose of this paper is to summarize and expand upon the known properties of the distribution of the sample covariance from a bivariate normal distribution.

Section 2 contains various representations of the density for arbitrary sample sizes and presents a new form for even sample sizes. In section 3 the cdf. is given for even sample sizes and is tabulated for the null case of independent components. In section 4 some properties of the distribution are developed in terms of a new representation. The asymptotic distribution is given in section 5 while some aspects of inference associated with the distribution are discussed in section 6.

The notation $\mathcal{L}(X, Y)=N(\mu, \Sigma)$ denotes that X, Y are joint normally distributed with mean vector μ, and positive definite covariance matrix $\Sigma\equiv(\sigma_{ij})$, where $\sigma_{ii}=\sigma_i^2$, $\sigma_{ij}=\rho\sigma_i\sigma_j$ for $i\neq j$, and $i, j=1, 2$. Define

$$v\equiv\sum_{1}^{N}(x_i-\bar{x})(y_i-\bar{y}), \tag{1}$$

where $(x_1, y_1),\cdots,(x_N, y_N)$ denotes a sample of independent observations from $N(\mu, \Sigma)$, $\bar{x}, \bar{y}$ are the sample means, and $n\equiv N-1$. v is called the sample covariance with n degrees of freedom (d.f.). The notation a, b, τ will be used to designated the 1-1, 2-2, and 1-2 elements of Σ^{-1}, respectively, and $\beta^2\equiv ab$. Then, $a\equiv[\sigma_1^2(1-\rho^2)]^{-1}$, $b\equiv[\sigma_2^2(1-\rho^2)]^{-1}$, $\tau\equiv -\rho[\sigma_1\sigma_2(1-\rho^2)]^{-1}\equiv-\rho\beta$.

2. Density

The earliest reporting of an explicit form for the density of v appears to be that of Pearson, Jeffery and Elderton [6], in 1929. They obtained the density by carrying out a sequence of transformations on the joint distribution of the elements of the sample covariance matrix.

The final expression for the density is given in terms of a Bessel function, and one factor in the density expression is extensively tabulated. The authors also attempt to fit the Bessel function density expression to one of the Pearson type curves.

In 1932, Wishart and Bartlett [8] obtained the same density result by inverting the appropriate characteristic function. Mahalanobis, Bose, and Roy [5] in 1937 found the Bessel function density by a geometrical approach using rectangular coordinates.

Let $f(x)$ denote the density function for v with n d.f., defined in (1). Then, using the elements of Σ^{-1} for the parameterization, it may be found that for $-\infty<x<\infty$,

$$f(x)=\frac{(\beta^2-\tau^2)^{n/2}|x|^{(n-1)/2}e^{-\tau x}}{\Pi^{1/2}(2\beta)^{(n-1)/2}\Gamma(n/2)}K_{(n-1)/2}(\beta|x|) \tag{2}$$

where $K_\alpha(z)$ denotes the modified Bessel function of the second kind. The distribution is thus indexed by two parameters (β,τ), and, of course, n.

Alternative representations in terms of other special functions are

$$f(x)=(\beta^2-\tau^2)^{n/2}\Gamma^{-1}(n/2)\Psi(n/2,\, n;\, 2|x|\beta)e^{-|x|\omega(x)}\,,$$

and

$$f(x)=(\beta^2-\tau^2)^{n/2}\Gamma^{-1}(n/2)(2\beta)^{-n/2}|x|^{n/2-1}e^{-\tau x}W_{0,(n-1)/2}(2\beta|x|)\,.$$

In these representations,

$$\Psi(f_1,f_2;x)\equiv\frac{1}{\Gamma(f_1)}\int_0^\infty e^{-xt}t^{f_1-1}(1+t)^{f_2-f_1-1}dt$$

is the associated confluent hypergeometric function of Kummer (see Erdélyi [3]) which has the advantage over the customary, ${}_1F_1$ function in that it remains finite for all x,

$$W_{\alpha,\nu}(z)\equiv\frac{z^{(\nu+1)/2}e^{-z/2}}{\Gamma(\nu+1/2-\alpha)}\int_0^\infty e^{-zx}x^{\nu-\alpha-1/2}(1+x)^{\nu+\alpha-1/2}dx$$

$$\text{for } \operatorname{Re}(\nu+1/2-\alpha)>0,\quad |\arg z|<\Pi,$$

is the Whittaker function, and

$$\omega(x)\equiv\frac{\beta^2-\tau^2}{\beta-\tau\operatorname{sgn}(x)}$$

with $\operatorname{sgn}(x)=+1, 0, -1$ as $x>0$, $x=0$, and $x<0$. These equivalent forms may be found from the basic relations

$$\Psi(f_1,f_2;x)=e^{x/2}x^{-1/2-\alpha}W_{k,\alpha}(x)\,,$$

where $k=(f_2/2)-f_1$, $\alpha=(f_2/2)-1/2$, and $W_{0,\alpha}(x)=\sqrt{x/\Pi}\,K_\alpha(x/2)$, for $\mathrm{Re}(\alpha)>-1/2$, $\mathrm{Re}(x)>0$ (see Erdélyi [3], p. 265).

Even number of degrees of freedom

When the number of degrees of freedom is even, it is possible to express the density of v in terms of elementary functions. The final result is that if $n=2m$, $m=1,2,\cdots$, the density of v is expressible, for $-\infty<x<\infty$, as

$$f(x)=\sum_{j=0}^{m-1} C_j|x|^j e^{-|x|\omega(x)}, \tag{3}$$

where

$$C_j\equiv\binom{m-1}{j}\frac{\Gamma(2m-j-1)(\beta^2-\tau^2)^m}{\Gamma^2(m)(2\beta)^{2m-j-1}}.$$

This result may be obtained from the fact that (see Abramowitz and Stegun [1], p. 444) for $n=2m$,

$$\sqrt{\frac{\Pi}{2z}}K_{m-1/2}(z)=\frac{\Pi}{2z}e^{-z}\sum_{k=0}^{m-1}\left[\frac{(2z)^{-k}(m-1+k)!}{k!\,\Gamma(m-k)}\right].$$

The result given in (3) can be understood directly, without resorting to the special properties of Bessel functions, by examining the characteristic function of v (see [8]) given by

$$\mathrm{E}\,(e^{vit})=(1-2it\rho\sigma_1\sigma_2+t^2|\Sigma|)^{-n/2}. \tag{4}$$

For arbitrary n, inversion cannot be effected in terms of elementary functions. However, for even n, the function is meromorphic and therefore its Fourier integral is readily obtained by the residue theorem as the exponentially damped polynomial (3).

3. Cumulative distribution function

This section provides formulas and tables for computing the exact cumulative distribution function of the sample covariance for the case of an even number of d.f. For the null distribution ($\rho=0$), some exact percentage points are presented in table 1. The results of section 5 are used for large sample sizes.

For arbitrary sample sizes the cdf. of v is a complicated expression which is difficult to evaluate. However, for even sample sizes it is possible to develop an exact expression.

Let $G(t)$ denote the cdf. of v. Then, for $n=2m$, $m=1, 2, \cdots$, it can be found that

$$G(t)=\sum_{j=0}^{m-1} C_j^* |t|^j e^{t\omega(t)}, \qquad t<0$$

$$1-G(t)=\sum_{j=0}^{m-1} C_j^* t^j e^{-t\omega(t)}, \qquad t>0$$

where

$$C_j^* \equiv \sum_{k=j}^{m-1} \frac{C_k k!}{j! [\omega(t)]^{k-j+1}}.$$

A straightforward method of proving this result is to differentiate $G(t)$ and show that the resulting equations may be combined and expressed as the density $f(x)$, as given in (3). There are no troublesome difficulties and the calculations are direct, though tedious.

Numerical percentage points of the v distribution were computed, for the case of $n=2m$, and $\rho=0$.

Note that when m is fixed and $\rho=0$, $G(t)$ becomes a function of (βt), rather than β and t separately. Hence percentage points were computed as a function of (βt).

Inspection of the numerical values shows that when n is as large as 30, the distribution is already approximately normal. Thus, for larger numbers of degrees of freedom, the standard normal tables can be applied by using the results of section 5. The numerical results are presented in table 1, correct to four decimal places.

Table 1

n \ α	.90	.95	.99
4	2.3972	3.2718	5.1918
6	2.9934	4.0104	6.1768
8	3.4930	4.6329	7.0135
10	3.9313	5.1816	7.7555
20	5.6400	7.3369	10.7025
30	6.9073	8.9855	13.1074
100	12.6107	16.4048	23.9302

Cumulative distribution function of the sample covariance for n degrees of freedom and $\rho=0$. The entries are values of βt such that $G(t)=\alpha$.

4. Properties

(a) Use the notation $\mathcal{L}(u)=\gamma[c, d]$ to denote that u is a gamma variate with density

$$p(u; c, d)=\frac{u^{c-1}e^{-u/d}}{d^c\Gamma(c)}, \qquad u, c, d>0 .$$

Define the parameters

$$\delta_1\equiv(\beta+\tau)^{-1}, \qquad \delta_2\equiv(\beta-\tau)^{-1}.$$

Then, if v denotes the sample covariance with n d.f. from $N(\mu, \Sigma)$, the distribution of v has the representation

$$\mathcal{L}(v)=\mathcal{L}(U_1-U_2),$$

where U_1 and U_2 are independent gamma variates following the laws

$$\mathcal{L}(U_1)=\gamma\left[\frac{n}{2}, \delta_1\right], \qquad \mathcal{L}(U_2)=\gamma\left[\frac{n}{2}, \delta_2\right].$$

This property may be established by noting from (4) that the characteristic function of v factors into

$$\mathrm{E}(e^{vit})=(1-it\delta_1)^{-n/2}(1+it\delta_2)^{-n/2}.$$

The case of independence ($\rho=0$) requires that $\tau=0$, or that $\delta_1=\delta_2$. Hence, for X, Y independent, v is distributed as the difference of two independent and identically distributed gamma variates.

(b) By setting $v=U_1-U_2$ where U_1, U_2 are defined in (a), the moments of v become available in terms of moments of gamma variates. Thus, for example,

$$\mathrm{E}(v)=\frac{n\delta_1}{2}-\frac{n\delta_2}{2}=-\frac{n\tau}{\beta^2-\tau^2},$$

and

$$\mathrm{Var}(v)=\frac{n\delta_1^2}{2}+\frac{n\delta_2^2}{2}=\frac{n(\beta^2+\tau^2)}{(\beta^2-\tau^2)^2}.$$

Wishart and Bartlett [8] give the rth semi-invariant of (v/n), obtained from the characteristic function of v, as

$$\frac{1}{2}(r-1)!\sigma_1^r\sigma_2^r n^{1-r}[(\rho+1)^r+(\rho-1)^r].$$

(c) When $\rho=0$, a check of the density expression shows that the distribution of v is symmetric and unimodal with the mean and mode coinciding at the origin. For $\rho \neq 0$, the distribution is skewed.

(d) Finally the covariance distribution has the reproductive property. That is, if v and w are independent sample covariances with r, s, d.f., respectively, from non-singular bivariate normal populations with equal covariance elements, the density of $(v+w)$ is given by $f(x)$ with $n=r+s$. This property holds for the full sample covariance matrices, and hence must also hold for the marginals.

5. Asymptotic distribution

The asymptotic distribution of v is readily obtained by applying theorem 4.2.4 of Anderson ([2], p. 75) to the off diagonal element of the sample covariance matrix. The result is that

$$\lim_{n\to\infty} \mathcal{L}\left\{\frac{(\beta^2-\tau^2)v+n\tau}{[n(\beta^2+\tau^2)]^{1/2}}\right\}=N(0, 1).$$

In the special case of independence,

$$\lim_{n\to\infty} \mathcal{L}\left\{\frac{\beta v}{\sqrt{n}}\right\}=N(0, 1).$$

6. Inferential considerations

Use of v for inference purposes was suggested by Wishart [7] for a factor analysis application. His recommendation was that covariance tetrads be used in the two factor problem, instead of correlation coefficient tetrads. The distribution of v may be helpful in making such inferences. Gates [4] has pointed out that the sample covariance (phenotypic covariance) is often used in statistical genetics.

Expanding the density of the full sample covariance matrix, it may be checked that if a and b are known, v is sufficient for τ. Recall that

$$\operatorname{Var}(X|Y)=a^{-1}, \qquad \operatorname{Var}(Y|X)=b^{-1}.$$

Then, if both conditional variances are known, since $\tau=-\rho\sqrt{ab}$, v is sufficient for ρ.

Now we refer back to (2). Clearly, if β is known, the distribution of v is a member of the exponential family (in ρ) and therefore, has monotone likelihood ratio. Thus, for example, a UMP level α test of $H: \{\rho \leqq \rho_0,\ ab=\beta^2\}$ vs. $H': \{\rho>\rho_0,\ ab=\beta^2\}$ is to reject H if $v>$constant.

Substitution of $\tau=-\beta\rho$ in (2) and differentiating with respect to ρ generates the maximum likelihood estimator of ρ,

$$\hat{\rho}=-\frac{n}{2\beta v}+\operatorname{sgn}(v)\left[1+\left(\frac{n}{2\beta v}\right)^2\right]^{1/2}.$$

It is easily checked that $\hat{\rho}$ is a monotone increasing and continuous function of v. The distribution of $\hat{\rho}$ is an immediate consequent of the distribution of v, and from the large sample properties of M.L.E.'s,

$$\lim_{n\to\infty}\mathcal{L}\left\{\frac{[n(1+\rho^2)]^{1/2}(\hat{\rho}-\rho)}{1-\rho^2}\right\}=N(0,1).$$

Acknowledgement

The author is very grateful to Professor I. Olkin of Stanford University for his guidance and many helpful suggestions during the course of this research.

UNIVERSITY OF CHICAGO

REFERENCES

[1] M. Abramowitz and I. A. Stegun (Ed.), *Handbook of Mathematical Functions*, Nat. Bur. Stand., *Ann. Math. Statist.*, 55, 1964.

[2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.

[3] A. Erdélyi (Ed.), *Higher Transcendental Functions*, I, McGraw-Hill Book Company, New York, 1953.

[4] C. E. Gates, Institute of Agriculture, St. Paul, Minnesota, Personal communication of January 31, 1966.

[5] P. C. Mahalanobis, R. C. Bose and S. N. Roy, "Normalization of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributions," *Sankhyā*, 3 (1937), 1-40, equation (21.8).

[6] K. Pearson, G. B. Jeffery and E. M. Elderton, "On the distribution of the first product moment-coefficient in samples drawn from an indefinitely large normal population," *Biometrika*, 21 (1929), 164-201, equation (vi).

[7] J. Wishart, "Sampling errors in the theory of two factors," *Brit. J. Psychol.*, 19 (1928), 180-187.

[8] J. Wishart and M. S. Bartlett, "The distribution of second order moment statistics in a normal system," *Proc. Camb. Phil. Soc.*, 28 (1932), 455-459.