ON THE EVALUATION OF THE PROBABILITY INTEGRAL OF A MULTIVARIATE NORMAL DISTRIBUTION

By S. JOHN

Indian Statistical Institute, Calcutta

SUMMARY. A simple reduction formula is obtained for the probability integral of a multivariate normal distribution. The derivation involves only elementary results in probability theory. The formula obtained can be used in evaluating probability integrals of multivariate normal distributions of order k when those of order $k-1$ are readily available. An example is worked out illustrating the use of the formula.

Introduction

In so far as statisticians often assume many populations to be multivariate normal, the evaluation of the probability integral of the multivariate normal distribution is of especial importance. While this has been done for univariate and bivariate normal distributions, the extension to populations of higher orders presents considerable difficulties. For previous work on this problem reference may be made to David (1953), Placket (1954), Moran (1956) and Das (1956).

Notation and Preliminaries

The vector valued random variable $X = (X_1, X_2, ..., X_p)$ will be said to have the density function $n_x(\mu; \Sigma)$ if it has the multivariate normal distribution with means $\mu = (\mu_1, \mu_2, ..., \mu_p)$ and dispersion matrix $\Sigma = (\sigma_{ij})$

$$n_x(\mu; \Sigma) = (2\pi)^{-p/2} \mid \Sigma \mid^{-1/2} \exp \{-\tfrac{1}{2}(x-\mu) \Sigma^{-1} (x-\mu)'\}.$$

$\Sigma_{\cdot i}$ will denote the matrix $(\sigma_{rs}-\sigma_{ir}\sigma_{is}/\sigma_{ii})$ with the i-th row and column deleted. For scalar u,

$$\mu_{\cdot i}(u) = (\mu_1, \mu_2, ..., \mu_{i-1}, \mu_{i+1}, ..., \mu_p)+(u-\mu_i)(\beta_{1i}, ..., \beta_{i-1,i}, \beta_{i+1,i}, ..., \beta_{pi})$$

where $\beta_{rs} = \sigma_{rs}/\sigma_{ss}$.

Given $X_i = u$, $X_{\cdot i} = (X_1, ..., X_{i-1}, X_{i+1}, ..., X_p)$ is distributed according to the density function $n_{x_{\cdot i}}(\mu_{\cdot i}(u), \Sigma_{\cdot i})$.

The new method

The problem is to evaluate integrals of the form

$$\int_{a_1}^{\infty} \int_{a_2}^{\infty} ... \int_{a_p}^{\infty} n_x(\mu; \Sigma)dx \quad \text{and} \quad \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} ... \int_{-\infty}^{b_p} n_x(\mu; \Sigma)dx, \qquad ... \quad (1)$$

$$\int_{a_1}^{\infty} ... \int_{a_p}^{\infty} n_x(\mu; \Sigma)dx = \int_{0}^{\infty} ... \int_{0}^{\infty} n_x(\mu-a; \Sigma)dx \qquad .... \quad (2)$$

and,

$$\int_{-\infty}^{b_1} ... \int_{-\infty}^{b_p} n_x(\mu; \Sigma)dx = \int_{0}^{\infty} ... \int_{0}^{\infty} n_x(b-\mu; \Sigma)dx \qquad ... \quad (3)$$

where

$$a = (a_1, ..., a_p) \quad \text{and} \quad b = (b_1, ..., b_p). \qquad ... \quad (4)$$

Therefore we need consider only integrals of the type

$$I = \int_0^\infty \cdots \int_0^\infty n_x(\mu; \Sigma) dx. \qquad \cdots \quad (5)$$

We now observe that I is the probability that $U = \min(X_1, \ldots, X_p)$ is greater than or equal to zero. If $f(u)$ is the probability density function of U,

$$I = \int_0^\infty f(u) du. \qquad \cdots \quad (6)$$

To derive the probability density function of U, we employ the following simple argument. The event that U lies in the interval $(u-du, u)$ can happen in p mutually exclusive ways. Either X_1 lies in the interval $(u-du, u)$ and $X_2, X_3, X_4, \ldots, X_p$ are all greater than u or X_2 lies in the interval $(u-du, u)$ and $X_1, X_3, X_4, \ldots, X_p$ are all greater than u or X_3 lies in the interval $(u-du, u)$ and $X_1, X_2, X_4, X_5, \ldots, X_p$ are all greater than u and so on. Thus,

$$f(u) = \sum_{i=1}^{p} \left\{ \int_u^\infty \cdots \int_u^\infty n_{x_{.i}}(\mu_{.i}(u); \Sigma_{.i}) dx_{.i} \right\} (2\pi\sigma_{ii})^{-1/2} \exp\{-(u-\mu_i)^2/2\sigma_{ii}\} \qquad \cdots . (7)$$

where $$dx_{.i} = dx_1 \ldots dx_{i-1}\, dx_{i+1} \ldots dx_p.$$

If all the variances σ_{ii}'s are equal and all the covariances σ_{ij}'s are equal and also $\mu_1 = \mu_2 = \ldots = \mu_p$ (7) will take the simpler form

$$f(u) = p(2\pi\sigma_{11})^{-1/2} \exp\{-(u-\mu_1)^2/2\sigma_{11}\} \int_u^\infty \cdots \int_u^\infty n_{x_{.1}}(\mu_{.1}(u), \Sigma_{.1}) dx_{.1} \qquad \cdots \quad (9)$$

There are also obvious simplifications when some of the simple or partial correlations are zero.

The evaluation of the density function (7) requires only a knowledge of the probability integral of multivariate normal distributions of order $p-1$. Thus (6) may be regarded as a sort of reduction formula. When probability integrals of normal distributions of order k are readily available, formula (6) may be used in conjunction with methods of numerical integration to evaluate probability integrals of order $k+1$ and, with more labour, of order $k+2$. Thus with the tables now available, probabilities for distributions of order three and four may be evaluated without much difficulty. We also feel that (6) may be profitably used to extend existing tables of the probability integral of the bivariate normal distribution (Pearson, 1931). This will require only the ordinates of the standard univariate normal curve at selected points and the area under this curve below these ordinates. Extensive tables for these are available in several places (Pearson and Hartley, 1954 ; U. S. Dept. of Commerce, 1953).

The example given below will help to clarify some points.

Example : We shall evaluate

$$I = \int_0^\infty \int_0^\infty \int_0^\infty n_3(\mu;\, \Sigma)dx_1dx_2dx_3 \quad \ldots \quad (8)$$

for the case $\mu = 0$, $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$, $\sigma_{12} = \sigma_{21} = \sigma_{13} = \sigma_{31} = \sigma_{23} = \sigma_{32} = 0.6$ (We wish to emphasize here that our method is applicable whatever μ and Σ). Formula (6) in this case becomes

$$I = 3 \int_0^\infty f_1(u)f_2(u)du \quad \ldots \quad (9)$$

where

$$f_1(u) = \int_{u/2}^\infty \int_{u/2}^\infty \frac{1}{2\pi[1-(.375)^2]^{\frac{1}{2}}} \exp \left[-\frac{1}{2(1-(.375)^2)}(x_1^2 - 2\times .375x_1x_2 + x_2^2)\right] dx_1dx_2 \ldots \quad (10)$$

and

$$f_2(u) = (2\pi)^{-\frac{1}{2}}e^{-u^2/2}. \quad \ldots \quad (11)$$

$f(u) = f_1(u) \,.\, f_2(u)$ was calculated for $u = 0, .2, .4, \ldots, 4.2$ from values of $f_1(u)$ taken from Karl Pearson's *Tables for Statisticians and Biometricians*, Part II, and values of $f_2(u)$ taken from *Biometrika Tables*. We used Weddle's rule to calculate the integral of $f(u)$ from 0 to 3.6 and Simpson's three-eighth's rule to evaluate the integral from 3.6 to 4.2. Of course any other convenient method of numerical integration could have been adopted.

The contribution to (9) by the integral of $f(u)$ from 4.2 to ∞ remains to be assessed. From tables we find that $f_1(u) < .0019$ for $y \geqslant 4.2$. Therefore,

$$\int_{4.2}^\infty f(u)du < .0019 \int_{4.2}^\infty (2\pi)^{-\frac{1}{2}}e^{-u^2/2}du \approx .0019\times .0000133 \approx (.25)10^{-7} \quad \ldots \quad (12)$$

This we regard negligible. Thus the value of the integral from 0 to 4.2 may be taken as an approximation to the value of the integral from zero to infinity. In general, we can always replace the upper limit of integration in (6) by a suitable finite number h which will give results of sufficient accuracy. For the problem of our example, the result obtained by this method was

$$I = .27007 \quad \ldots \quad (13)$$

We now compare this result with the one given by the formula

$$(2\pi)^{-3/2}|M|^{-1/2}\int_0^\infty\int_0^\infty\int_0^\infty \exp\{-\tfrac{1}{2}xM^{-1}x'\}\,dx = \frac{\cos^{-1}(-\rho_{12})+\cos^{-1}(-\rho_{13})+\cos^{-1}(-\rho_{23})-\pi}{4\pi} \quad \ldots \quad (14)$$

where M is the matrix (ρ_{ij}) . $(\rho_{11} = \rho_{22} = \rho_{33} = 1)$ Plackett (1954). Formula (14) gives the value .27554 which differs from result (13) by about .0035. This much of difference was expected since, in evaluating $f(u)$ for various values of u, only linear interpolation was used with regard to the correlation coefficient.

REFERENCES

DAS, S. C. (1956): The numerical evaluation of a class of integrals. II. *Proc. Camb. Phil. Soc.*, 52, 442–448.

DAVID, F. N. (1953): A note on the evaluation of the multivariate normal integral. *Biometrika*, 40, 458–459.

MORAN, P. A. P. (1956): The numerical evaluation of a class of integrals. *Proc. Camb. Phil. Soc.*, 52, 230–233.

PEARSON, E. S. AND HARTLEY, H. O. (1954): *Biometrika Tables for Statisticians*, Cambridge, Biometrika Trustees. Table I, 1.

PEARSON, KARL (1931): *Tables for Statisticians and Biometricians*, Part 2, Tables VIII & IX, 78–137. Cambridge University Press.

PLACKETT, R. L. (1954): A reduction formula for normal multivariate integrals. *Biometrika*, 41, 351–360.

U. S. Department of Commerce (1953): *Tables of Normal Probability functions*,

$$\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \text{ and } \frac{1}{\sqrt{2\pi}}\int_{-x}^{x} e^{-a^2/2}\,da$$

Paper received : January, 1959.