TABLES FOR THE STANDARD BIVARIATE NORMAL DISTRIBUTION

Montira Jantaravareerat and Nick T. Thomopoulos
Stuart School of Business
Illinois Institute of Technology

ABSTRACT

In conventional methodology, the values of the cumulative distribution function (cdf) of most bivariate distributions are obtained via complicated procedures of double integration. Presented in this paper are methods through which the distribution function $F(z_1, z_2)$ for the bivariate standard normal can be obtained, not through the integration process, but through the averaging and summing procedures. This paper summarizes the process, shows the results as the table of joint cdf values, $F(z_1, z_2)$ and also presents an application of finding the b -array .

INTRODUCTION

One of the most important and fundamental distributions of applied multivariate analysis is the bivariate normal distribution, where two normal random variables are correlated. It is known that to obtain the probability by integrating of the bivariate density over the arbitary portion of (x, y) plane is not generally simple.

For over forty years, attempts to obtain the bivariate cumulative distribution function have been carried out by engineers and researchers. Various tables have been published to obtain the probability over the rectangle for bivariate normal variables. Some of these tables give the probabilities as a function of three parameters, i.e.,[4]. Others such as [2], [3] tabulated related two-parameter families from which these probabilities may be computed.

This paper presents a method to obtain the values of the joint cumulative probabilities of the standard bivariate normal distribution, $F(z_1, z_2)$. To show the validity of the method, we show an example in finding $F(z_1, z_2)$ for the correlation index (r) of 0.5. The complete output for the values of r from -0.9 to 0.9 and more applications can be seen in [5].

DEFINITIONS AND BASIC EQUATIONS

Consider variables x, y that are related with the bivariate normal distribution. The mean and standard deviation of x is denoted as m_x, s_x, the same for y as m_y , s_y ; now r is the correlation beween x and y. The bivariate normal probability density is denoted as $f(x, y)$.

The Cumulative Distribution Function (cdf) of x and y related in the bivariate way is

$$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)\, dy\, dx \tag{1}$$

The above bivariate distribution can be stated in a standardized form as follows. First let,

$$z_1 = \frac{x - \mu_x}{\sigma_x}$$

$$z_2 = \frac{y - \mu_y}{\sigma_y}$$

Therefore, the standard bivariate density becomes

$$f(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\right)\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right) \tag{2}$$

METHODOLOGY

To estimate the cumulative distribution function (cdf) for the bivariate normal distribution, a simple process is presented. Consider the coordinating points of (z_1, z_2) where $-3 \le z_1 \le 3$ and $-3 \le z_2 \le 3$ and each is ordered in increments of 0.1. This gives 61 values of z_1 and 61 values of z_2. Altogether, the number of possible paired values are $61 \times 61 = 3721$.

Note that each of these cells represents a square with the side of length 0.1. Hence, each cell has an area of 0.01. The cumulative probability distribution function of each point is obtained by taking the following steps:

Step 1 Calculate the bivariate density $f(z_1, z_2)$ for each of those cells by using equation (2).

Step 2 Find the average density $h(z_1, z_2)$ of each cell in the range z_1 and z_2.

Let the sides of the square be $\delta_1 = \delta_2 = 0.1$ and $f(z_1, z_2)$ be the bivariate normal density for (z_1, z_2),

Let $$g(\delta_1, \delta_2) = f(z_1 - \delta_1, z_2 - \delta_2) \tag{3}$$

and let $h(z_1, z_2)$ be the average density for the incremental area for the point (z_1, z_2), then

$$h(z_1, z_2) = \frac{1}{4}\left(g(0,0) + g(\delta_1, 0) + g(0, \delta_2) + g(\delta_1, \delta_2)\right) \tag{4}$$

Step 3 Compute the volume of each cell by multiplying the average density obtained from step 2 to the area.

Since $d_1 = d_2 = 0.1$, the area = 0.01 and the volume of (z_1, z_2) is given by

$$P(z_1, z_2) \; = \; d^{2\,\prime}\, h(z_1, z_2) = \; 0.01 \times h(z_1, z_2) \qquad (5)$$

Step 4 The joint cumulative distribution function, $F(z_1, z_2)$ for the each value of (z_1, z_2) is obtained by summing all the volume of the density for the area under that particular point.

$$F(z_1, z_2) \; = \; \sum_{j_1 \le z_1} \sum_{j_2 \le z_2} P(j_1, j_2) \qquad \text{where} \quad \begin{matrix} z_1 = -4.0, -3.9, -3.8, \ldots, 4.0 \\ z_2 = -4.0, -3.9, -3.8, \ldots, 4.0 \end{matrix} \qquad (6)$$

THE BIVARIATE DISTRIBUTION FUNCTION TABLE

For a particular value of the correlation coefficient, r, the value of joint cumulative probability $F(z_1, z_2)$ for each point of (z_1, z_2) from -4.0 to 4.0 can be obtained by the proposed method. Tables A.1 to A.19 give the values of $F(z_1, z_2)$ when r are -0.9 to 0.9, for each 0.1 increment within the area of $-3 \le z_1 \le 3$ and $-3 \le z_2 \le 3$. The value of z_1 from -3.0 to 3.0 with increment 0.2 is shown in the first row and the value of z_2 from -3.0 to 3.0 with increment 0.1 is shown in the first column of the table. Other entries are the values of $F(z_1, z_2)$ for each set of coordinated points (z_1, z_2). The value of $F(z_1, z_2)$ is equal to zero at the origin (-3,-3) and accumulates as it gains distance from the origin until it approaches the value of 1 when (z_1, z_2) is (3,3).

Table B.1 gives $F(z_1, z_2)$ for the special cases where $z_1 = z_2$. The 19 right hand columns of the table list the values of r. The first column of the table is the common value of z_1 and z_2. Other entries are the values of $F(z_1, z_2)$ when $z_1 = z_2$.

THE b -ARRAY

Let b equal to a particular values of $F(z_1, z_2)$ and note that for b , the pair (z_1, z_2) is not unique. Instead a series of pairs can give the same $F(z_1, z_2) = b$. The purpose here is to define a series of five representative pairs of (z_1, z_2) that yield the same b . The five pairs are here defined as a b -array.

For each r , eleven b -arrays are generated. These are for b values: 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.99. The computer program is generated to obtain 5 paired values of (z_1, z_2) that yield the same value of $F(z_1, z_2)$ and also the corresponding univariate cdf of z_1 and z_2 for those values of (z_1, z_2). See [5] for the results of the program.

Table C.1 to C.19 show the results of the program when r are from -0.9 to 0.9. The values of $F(z_1, z_2)$ are shown in the first column, and for each value of $F(z_1, z_2)$ five values of (z_1, z_2) that have the same value of the joint cumulative probability are given. On the right hand side, the univariate cdf $F(z_1)$, $F(z_2)$ of z_1 and z_2 are listed.

For example, Table C.15 when r = 0.5 and b = $F(z_1, z_2)$ = 0.80, five possible paired values of (z_1, z_2) are given. These are: (0.9000, 1.7818), (1.4000, 1.0108), (2.0000, 0.8732), (2.5000, 0.8495) and (3.0000, 0.8444). The corresponding univariate cdf of those values are (0.8159, 0.9626), (0.9192, 0.8439), (0.9773, 0.8087), (0.9938, 0.8022) and (0.9987, 0.8008), respectively.

PROBABILITY OVER A RECTANGLE

Here, we show the utility of the table obtained from the proposed method. Suppose we want to find the probability that $a < z_1 < b$, $c < z_2 < d$.

Let $R(abcd)$ = rectangle where $a < z_1 < b$ and $c < z_2 < d$

$G(R)$ = Probability that $(z_1,\ z_2)$ falls in the rectangle R ,i.e., $R(abcd)$

The probability that $a < z_1 < b$, $c < z_2 < d$ can be obtained by using the data from the $F(z_1, z_2)$ table and the following formula:

$$\begin{aligned} G\big(R(abcd)\big) &= p\{a < z_1 < b\ ,\ c < z_2 < d\} \\ &= \int_c^d \int_a^b f(z_1, z_2)\, dz_1\, dz_2 \\ &= F(b,d) - F(b,c) - F(a,d) + F(a,c) \end{aligned} \qquad (7)$$

For r = 0.5 and $(1.5 < z_1 < 2.0)$, $(0.0 < z_2 < 0.5)$, this gives

$$p(1.5 < z_1 < 2.0,\ 0.0 < z_2 < 0.5) = F(2.0, 0.5) - F(2.0, 0.0) - F(1.5, 0.5) + F(1.5, 0.0)$$

First, obtain the values of $F(z_1, z_2)$ with r = 0.5 from Table 1, then substitute in the above equation as below :

$$\begin{aligned} p\left\{1.5 < z_1 < 2.0\ ,\ 0.0 < z_2 < 0.5\right\} &= 0.686 - 0.498 - 0.671 + 0.491 \\ &= 0.008 \end{aligned}$$

CONCLUSION

The main purpose of this study is to present a method to obtain the joint cumulative probability of the standard bivariate normal distribution. The procedures of the method are (i) find the joint densities in a certain range, (ii) average the density, (iii) multiply the result by the area to obtain the volume, (iv) sum to obtain the cumulative value. The $F(z_1, z_2)$ tables for values of r from

-0.9 to 0.9 and more of the applications such as the estimating error values and inventory solving problems are given in [5].

REFFERENCES

[1] Cadwell, J.H. (1951), "The Bivariate Normal Integral," *Biometrica*, Vol 38, pp 475- 481.

[2] Nicholson, C. (1943), " The Probability Integral for Two Variables," *Biometrica*, Vol.33, pp 59-72.

[3] Owen, D.B. (1962), *Handbook of Statistical Tables*, Addison-Wesley Publishing Company, Reading , Mass.

[4] Pearson, K. (1931), *Table for Statisticians and Biometricians*, Part II, Cambridge University Press.

[5] Jantaravareerat, M. (1998), *Approximation of the Distribution Function for the Standard Bivariate Normal*. Ph.D (Management Science) Thesis. Illinois Institute of Technology, Chicago, IL., May 1998.