International Statistical Review (1991), **59**, 2, pp. 263–269. Printed in Great Britain

A Simple Approximation for Bivariate and Trivariate Normal Integrals

D.R. Cox[1] and Nanny Wermuth[2]

[1]*Nuffield College, Oxford, U.K.* [2]*Psychologisches Institut, Universität Mainz, Mainz, West Germany*

Summary

A simple approximation for the bivariate normal distribution function is described, together with a second-order refinement. For $|\rho| \leq 0{\cdot}9$, the worst error is about 10% arising when both arguments of the distribution function are equal, but over most of the range the agreement is much closer. An extension to trivariate normal integrals has similar good properties.

Key words: Approximate formulae; Bivariate normal; Test of multivariate normality; Trivariate normal.

Numerical evaluation of the bivariate normal distribution function is required for a number of probabilistic and statistical purposes and there is no general closed form expression. The computer algorithm of Donnelly (1973) and its extension to more dimensions by Schervish (1984) are available for accurate numerical evaluation and have been extensively used in the present study. The National Bureau of Standards tables (1959) are comprehensive and easily used, especially for 'simple' values of the correlation coefficient. Owen (1956) discusses the numerical analytical aspects, provides the theoretical basis of Donnelly's algorithms and gives a concise form of table. It is still helpful, however, to have an explicit formula for at least three reasons: to aid further analytical development, to be employed in rapid 'pocket calculator'-based evaluation, especially for preliminary calculations, and for computerized use when a large number of evaluations are required and speed of computation is important. The present note gives such a formula.

Let (X, Y) have a bivariate normal distribution of zero means, unit variances and correlation coefficient ρ.

We write

$$L(a, b; \rho) = P(X > a, Y > b).$$

Probabilities in other quadrants are easily determined given $L(a, b; \rho)$ and the univariate standardized normal distribution function $\Phi(x)$. Now

$$\begin{aligned} L(a, b; \rho) &= P(X > a)P(Y > b \mid X > a) \\ &= \Phi(-a)E\left\{\Phi\left(\frac{\rho X - b}{\sqrt{(1-\rho^2)}}\right) \,\middle|\, X > a\right\} \end{aligned} \tag{1}$$

where the second form follows because Y given $X = x$ is normally distributed with mean ρx and variance $1 - \rho^2$.

Provided that the function whose expectation is taken in (1) is nearly linear over the range of appreciable probability, as is a plausible approximation if $a, b, \rho > 0$, we may as a first approximation replace X in (1) by

$$\mu(a) = E(X \mid X > a) = \phi(a)/\Phi(-a), \tag{2}$$

where $\phi(x) = \Phi'(x)$ is the standardized normal density. Thus

$$\begin{aligned} L(a, b; \rho) &\simeq \Phi(-a)\Phi\left\{\frac{\rho\mu(a) - b}{\sqrt{(1-\rho^2)}}\right\} \\ &= \Phi(-a)\Phi\{\xi(a, b; \rho)\}, \end{aligned} \tag{3}$$

say.

Now for a random variable Z of fairly small dispersion and for a function $g(Z)$ of not too nonquadratic a form

$$E\{g(Z)\} \simeq g(\mu_Z) + \tfrac{1}{2}\sigma_Z^2 g''(\mu_Z),$$

where μ_Z and σ_Z^2 are the mean and variance of Z. By applying this to (1) a refinement to (3) is produced, namely

$$L(a, b; \rho) \simeq \Phi(-a)[\Phi\{\xi(a, b; \rho)\} - \tfrac{1}{2}\rho^2(1-\rho^2)^{-1}\xi(a, b; \rho)\phi\{\xi(a, b; \rho)\}\sigma^2(a)], \tag{4}$$

where

$$\sigma^2(a) = \text{var}\,(X \mid X > a) = 1 + \mu(a) - \mu^2(a). \tag{5}$$

Now $L(a, b; \rho)$, is symmetric in a, b, but the approximations (3) and (4) are not. That is, a different pair of approximations results from interchanging a and b. Further, because, for example,

$$\begin{aligned} P(X > a, Y > b) &= P(X > a) - P(X > a, Y \leqslant b) \\ &= P(Y > b) - P(X \leqslant a, Y > b) \\ &= 1 - P(X < a) - P(Y < b) + P(X \leqslant a, Y \leqslant b), \end{aligned} \tag{6}$$

six more approximations like (3) are obtained in pairs by applying the argument leading to (3) to the final term in (6) and expressing the univariate probabilities in terms of $\Phi(.)$. There are, for $a \neq b$, thus eight approximations equal in pairs, each with a second-order version.

Numerical comparisions support the following proposal. For $\rho > 0$:

(i) provided at least one of a and b is positive, use (3), taking a to be the greater of the two arguments;

(ii) if both a, b are negative, compute an approximation to

$$L(-a, -b; \rho) = P(X < a, Y < b)$$

via (3) arranging $-a$ to be the larger of $-a$, $-b$ and thence, if required, compute

$$L(a, b; \rho) = 1 - \Phi(-a) - \Phi(-b) + L(-a, -b; \rho).$$

If $\rho < 0$, the results for $\rho > 0$ can be applied if we write, with $Y' = -Y$,

$$L(a, b; \rho) = P(X > a, Y' < -b) = \Phi(-a) - L(a, -b; -\rho).$$

Table 1 compares exact values computed via Schervish's (1984) procedure with the approximation (3) and the more refined version (4). The worst results are obtained for $a = b$ when ρ is large. At $\rho = 0{\cdot}9$ the error of (3) is roughly 10%, substantially reduced by the more refined version (4). Over much of the range, the agreement is much closer.

Table 1.

Exact bivariate normal integral $L(a, b; \rho)$ compared with approximation (3) *and improved approximation* (4). *Probabilities* $\times 10^4$

Situation	$(0,0)$	$(0,-\frac{1}{2})$	$(0,-1)$	$(\frac{1}{2},\frac{1}{2})$	(a,b) $(\frac{1}{2},0)$	$(\frac{1}{2},-\frac{1}{2})$	$(1,1)$	$(1,\frac{1}{2})$	$(1,0)$
$\rho=0{\cdot}2$, exact	2820	3740	4400	1207	1825	2376	381	669	986
(3)	2823	3748	4408	1206	1827	2380	379	668	987
(4)	2821	3740	4400	1207	1825	2376	381	669	986
$\rho=0{\cdot}8$, exact	3976	4692	4944	2186	2778	3022	976	1351	1531
(3)	4282	4855	4584	2327	2888	3057	1020	1404	1553
(4)	3892	4653	4942	2168	2747	3014	982	1339	1524
$\rho=0{\cdot}9$, exact	4282	4884	4993	2453	2969	3077	1155	1497	1580
(3)	4751	4987	5000	2736	3057	3085	1275	1551	1585
(4)	4096	4900	4997	2326	2953	3079	1116	1478	1579

Situation	$(\frac{3}{2},\frac{3}{2})$	$(\frac{3}{2},1)$	$(\frac{3}{2},\frac{1}{2})$	$(2,2)$	(a,b) $(2,\frac{3}{2})$	$(2,1)$	$(\frac{5}{2},\frac{5}{2})$	$(\frac{5}{2},2)$	$(\frac{5}{2},\frac{3}{2})$
$\rho=0{\cdot}2$, exact	86	178	304	14	34	67	2	4	11
(3)	86	178	304	14	34	67	1	4	11
(4)	86	178	304	14	34	67	2	4	11
$\rho=0{\cdot}8$, exact	349	530	631	98	165	209	22	41	55
(3)	357	548	641	98	170	212	21	41	56
(4)	354	527	628	100	165	208	22	41	55
$\rho=0{\cdot}9$, exact	439	615	663	134	203	225	32	53	61
(3)	476	639	667	142	211	226	33	55	62
(4)	435	605	663	135	200	225	33	53	61*

* There may be a rounding error of 1 in the last digit.

Only positive values of a are included because of the recommendation (ii) above. Further for each value of a it is necessary to give only three values of b because values of b smaller than the last one tabulated leave the integral unchanged to the accuracy used.

Use of the 'wrong' approximation may produce bad results; for instance at $\rho = 0{\cdot}8$, $a = 2$, $b = -2$ interchange of a and b in (3) leads to a first approximation of 0·0005 instead of 0·0228, which is the recommended approximation and is correct to the digits quoted.

Essentially the reason for this and the basis for the 'rules' set out above is that the relative accuracy of our approximation 0·436 to $P(Y > -2 \mid X > 2) = 0{\cdot}430$ is much greater than 0·0056 is to $P(X > 2 \mid Y > -2) = 0{\cdot}023$. This can be explained numerically by computing $P(Y > -2 \mid X = x)$ and $P(X > 2 \mid Y = y)$ for a range of respectively x and y of appreciable conditional probability given respectively $X > 2$ and $Y > -2$. The latter conditional probability but not the former, varies substantially and very nonlinearly, so that (3), based on a linearization of the function whose expectation is taken, performs badly in the latter case.

The relations between the other approximations hinge on such results as that

$$P(Y > b \mid X > a) = 1 - P(Y \leqslant b \mid X > a),$$

these being approximated respectively by

$$\Phi\left(\frac{\rho\mu(a) - b}{\surd(1-\rho^2)}\right), \quad 1 - \Phi\left(\frac{b - \rho\mu(a)}{\surd(1-\rho^2)}\right)$$

and that

$$P(Y > b \mid X > a) = P(Y < -b \mid X < -a),$$

the right-hand side being approximated by

$$\Phi\left(\frac{-b-\rho E(X \mid X<-a)}{\sqrt{(1-\rho^2)}}\right),$$

which is again equal to the first approximation. This is because $E(X \mid X<-a) = -E(X \mid X>a)$.

For large values of ρ and for arguments near $a=b$, an approximation of Polya (1946, e.g. eqns (6.7), (6.8)) gives good results, in particular the approximation

$$L(a, a; \rho) \simeq 1-\Phi\left(\frac{a\sqrt{2}}{\sqrt{(1+\rho)}}\right)-\frac{1}{\pi}\left(\frac{1-\rho}{1+\rho}\right)^{\frac{1}{2}} \exp\left\{\frac{a^2}{1+\rho}\right\}. \tag{7}$$

At $\rho=0{\cdot}8$, $a=0$ this gives 0·3983, instead of the exact value 0·3976, but at $\rho=0{\cdot}2$, $a=0$ it gives the poor approximation 0·240, compared with 0·2820. While at $\rho=0{\cdot}8$, $a=0$, Polya's approximation is rather better than our (4), the overall performance of (7) is much inferior; in particular (7) can lead to negative values.

It is possible to replace the univariate integrals by the approximations discussed by Lin (1990, e.g. eqn (5)), in particular by

$$\Phi(-z) \simeq \left\{1+\exp\left(\frac{4{\cdot}2\pi z}{9-z}\right)\right\}^{-1} \quad (z>0).$$

We have not done this in Table 1 although its use would have replaced (3) by an expression in terms of elementary functions. There is on the whole some loss of precision, minor except in regions of low correlation and low probability.

The approximations in this paper can be generalized in at least two ways. First, if the region of integration instead of being $(x>a, y>b)$ is $\{x>a, b_1(x)>y>b_0(x)\}$ the approximation (3) is replaced by

$$\Phi(-a)\left[\Phi\left\{\frac{\rho\mu(a)-b_0\{\mu(a)\}}{\sqrt{(1-\rho^2)}}\right\}-\Phi\left\{\frac{\rho\mu(a)-b_1\{\mu(a)\}}{\sqrt{(1-\rho^2)}}\right\}\right]. \tag{8}$$

Secondly we can argue in three dimensions that if (X, Y, Z) are trivariate normal with zero means, unit variances and correlation coefficients ρ_{yz}, ρ_{zx}, ρ_{xy}, then

$$\begin{aligned} L(a, b, c; \rho) &= P(X>a, Y>b, Z>c) \\ &= \Phi(-a)E\left[L\left\{\frac{b-\rho_{yx}X}{\sqrt{(1-\rho_{yx}^2)}}, \frac{c-\rho_{zx}X}{\sqrt{(1-\rho_{zx}^2)}}; \rho_{yz.x}\right\} \,\middle|\, X>a\right] \\ &\simeq \Phi(-a)L\left\{\frac{b-\rho_{yx}\mu(a)}{\sqrt{(1-\rho_{yz}^2)}}, \frac{c-\rho_{zx}\mu(a)}{\sqrt{(1-\rho_{zx}^2)}}; \rho_{yz.x}\right\} \end{aligned} \tag{9}$$

and then (3) can be applied to the second factor. Here $\rho_{yz.x}$ is the partial correlation coefficient between Y and Z given X.

A very extensive numerical check of (9) has not been attempted, but Table 2 summarizes comparisons for (i) an equicorrelated trivariate normal distribution with common correlation $\rho=0{\cdot}8$, (ii) a distribution with $\rho_{xy}=0{\cdot}4$, $\rho_{xz}=\rho_{yz}=0{\cdot}8$. In all cases we took $a \geqslant b \geqslant c$, for a reason similar to that underlying Table 1. Provided at least one argument is positive the agreement is remarkably good, being poorest near the origin. If the correlations are smaller the agreement improves. Thus in case (i) with $\rho=0{\cdot}2$, the exact and approximate values at $a=b=c=0$ are respectively 0·1731, 0·1726.

We give one illustration of the use of the bivariate approximation. Let T_1, T_2 be two test statistics having under a null hypothesis a standardized bivariate normal distribution

Table 2

Exact trivariate normal integral $L(a, b, c; \rho)$ compared with approximation (9). *Probabilities* $\times 10^4$.

Situations	(2, 2, 2)	(2, 2, 1)	(2, 2, 0)	(2, 2, 1)	(2, 1, 0)	(2, 1, −1)	(a, b, c) (2, 0, 0)	(2, −1, −1)	(1, 1, 1)	(1, 1, 0)	(1, 1, −1)
A exact	62	96	98	195	208	209	227	228	742	968	976
(9)	59	97	98	204	212	212	227	228	773	1020	1020
B exact	29	29	29	109	109	109	193	224	519	536	536
(9)	28	29	29	109	109	109	193	224	530	532	532

Situations	(1, 0, 0)	(1, 0, −1)	(1, −1, −1)	(0, 0, 0)	(0, 0, −1)	(0, 0, 2)	(a, b, c) (−1, −1, −1)	(−1, −1, −2)	(−1, −2, −2)	(−2, −2, −2)
A exact	1487	1531	1585	3464	3964	3976	7428	7798	8378	9550
(9)	1536	1553	1586	3881	4278	4282	8158	8244	8412	7969
B exact	1178	1179	1520	3053	3155	3155	7209	7362	8291	9515
(9)	1185	1185	1524	3179	3181	3181	7472	7472	8325	9639

A, all correlations, 0·8; B, $\rho_{xy} = 0{\cdot}4$, $\rho_{xz} = \rho_{yz} = 0{\cdot}8$. Missing entries in sequence identical with *t* last value tabulated; thus (2, 2, −1) has same entries as (2, 2, 0).

Table 3
Examination of quadratic regression. Diastolic and systolic blood pressures. Anxiety and anger

	r_{12}	$r_{12}(2-3r_{12}^2)$	Q_{12}	Q_{21}	U_2	Sig. level (11)
Blood pressures	0·741	0·261	0·508	−0·382	0·508	0·846
Anger and anxiety scores	0·354	0·574	−0·086	−0·789	−0·789	0·128

of correlation coefficient ρ. These can be combined into a single test statistic in various ways, for example, via the quadratic form

$$(T_1, T_2)\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1}\begin{pmatrix} T_1 \\ T_2 \end{pmatrix},$$

having a chi-squared distribution with two degrees of freedom. If, however, there are two qualitatively different types of departure and sensitivity is required against departures of either one of the the types on its own it may be sensible to use $U_1 = \max(T_1, T_2)$, or, in a two-sided version, $U_2 = \max(|T_1|, |T_2|)$ as test statistics. Then, under the null hypothesis

$$\begin{aligned} P(U_1 \geqslant u) &= P(T_1 \geqslant u \text{ or } T_2 \geqslant u) \\ &= P(T_1 \geqslant u) + P(T_2 \geqslant u) - P(T_1 \geqslant u, T_2 \geqslant u) \\ &= 2\Phi(-u) - L(u, u; \rho), \end{aligned} \tag{10}$$

and similarly

$$P(U_2 \geqslant u) = 2\Phi(-u) - 2L(u, u; \rho) + 2L(u, -u; \rho). \tag{11}$$

A special case (Cox & Small, 1978), involves testing the linearity of relationships in a bivariate distribution of two continuous variables X_1 and X_2 via (Q_{21}, Q_{12}), where Q_{21} is the standard t statistic for the significance of the regression of X_2 on X_1^2 adjusting for linear regression on X_1; Q_{12} interchanges the roles of X_1 and X_2. The asymptotic correlation between Q_{12} and Q_{21} is $\rho_{12}(2-3\rho_{12}^2)$, where ρ_{12} is the correlation between X_1 and X_2 and hence can be estimated consistently.

Data on (i) diastolic and systolic blood pressures and (ii) anxiety and anger scores of 98 males (Hodapp, Neuser & Weyer, 1988) gave the results summarized in Table 3, which serve as a largely formal illustration.

In none of the cases is there clear evidence of non-linearity, the larger departure being for the anger and anxiety scores. A scatter plot shows a suggestion of non-linearity at high scores of anger. The calculation of the significance level (11) adjusting for selection of the larger of $|Q_{12}|$ and $|Q_{21}|$ confirms that there is a reasonable consistency with the null hypothesis.

Acknowledgements

We are very grateful to Reinhold Streit, University of Mainz, for computing Tables 1 and 2, and to the British German Academic Research Collaboration Programme for supporting our work.

References

Cox, D.R. & Small, N.J.H. (1978). Testing multivariate normality. *Biometrika* **65**, 263–272.
Donnelly, T.G. (1973). Algorithm 462. Bivariate normal distribution. *Commun. Ass. Comput. Mach.* **16**, 636.

Hodapp, V., Neuser, K.W. & Weyer, H. (1988). Job stress, emotion and work environment: towards a causal model. *Personality and Individual Differences* **5,** 851–859.

Lin, J.-T. (1990). A simple logistic approximation to the normal tail probability and its inverse. *Appl. Statist.* **39,** 255–257.

National Bureau of Standards (1959). *Tables of the Bivariate Normal Distribution Function and Related Functions.* Washington, DC: US Government Printing Office.

Owen, D.B. (1956). Tables for computing bivariate normal probabilities. *Ann. Math. Statist.* **27,** 1075–1090.

Polya, G. (1946). Remarks on computing the probability integral in one and two dimensions. *Proc. Berkeley Symp.* 63–78.

Schervish, M.J. (1984). Multivariate normal probabilities with error bound. *Appl. Statist.* **33,** 81–94.

Résumé

On décrit une approximation simple pour la fonction de distribution normale en deux dimensions ainsi qu'un raffinement de second ordre. Dans le cas où la correlation ρ est telle que $|\rho| \leq 0{\cdot}9$, la plus grande erreur, d'environs 10%, se produit lorsque les deux arguments de la fonction de distribution sont égaux. Une généralisation à trois dimensions a des propriétés analogues.

[*Received July* 1990, *accepted December* 1990]