JOURNAL OF MULTIVARIATE ANALYSIS 49, 87–96 (1994)

A Simple Approximation to the Bivariate Normal Distribution with Large Correlation Coefficient

WILLEM ALBERS AND WILBERT C. M. KALLENBERG

University of Twente, Enschede, The Netherlands

The bivariate normal distribution function is approximated with emphasis on situations where the correlation coefficient is large. The high accuracy of the approximation is illustrated by numerical examples. Moreover, exact upper and lower bounds are presented as well as asymptotic results on the error terms.

1. APPROXIMATION

In many production processes large numbers of parts have to be inspected on various characteristics, for each of which specification limits (usually forming an interval) have been set. Typically, the measurement process will not be infallible. It is therefore common practice to shrink each specification interval somewhat to a test interval, accepting only those products for which the measurement falls inside the test interval. It is required that the probability of obtaining products which are both bad and accepted will stay below a prescribed bound.

If such test limits are to be determined assuming normality, the bivariate normal distribution function is needed for large values of the correlation coefficient ρ, corresponding to a measurement standard deviation, which is small w.r.t. the specification standard deviation (cf. e.g., Grubbs and Coon (1954), Mullenix (1991), Easterling *et al.* (1991), and also the beginning of Section 2). Also in many other applications evaluation of the bivariate normal distribution function with *large* ρ is required. Note, however, that the assumption of normality needs careful checking in practice.

There are many accurate and efficient computer algorithms available to compute the bivariate normal distribution function such as Divgi (1979)

Received July 2, 1992; revised March 25, 1993.

AMS 1991 subject classifications: 62E17, 62E20, 62N10.

Key words and phrases: approximate formulae, asymptotic error terms, bivariate normal, lower bound, quality control, upper bound.

0047-259X/94 $6.00

683/49/1-7

and Moskowitz and Tsai (1989). Compared to such "exact" solutions it turns out that the approximations we propose are extremely accurate in the sense that the difference is utterly negligible for practical application. Advantage of the approximations over the exact solutions is that the former open the way to further analytical evaluation of the procedures involved (cf. Cox and Wermuth (1991), p. 263 for similar and additional arguments in favor of explicit approximation formulae over exact numerical solutions).

Especially when the variance of the measurements or other parameters have to be estimated (which in practice usually is the case), it is vital to have both explicit and simple formulae at our disposal. For more details about these applications of the approximation formulae we refer the reader to Albers *et al.* (1994).

Recently, Cox and Wermuth (1991) have presented a simple approximation for the bivariate normal distribution function. Their approximation works very well, but becomes less accurate for large values of ρ. In this note the gap is filled in by presenting complementary approximation formulae which are intended for large values of ρ. Further, even simpler but nevertheless accurate approximation formulae are given in Section 2, as well as exact upper and lower bounds.

Let (X, Y) have a bivariate normal distribution with zero means, unit variances, and correlation coefficient ρ. The standard normal distribution function is denoted by Φ and its density by φ. Let $a, b \in \mathbb{R}$. We approximate the bivariate normal distribution function at the point (a, b). Since $P(X \leqslant a, Y \leqslant b) = \Phi(b) - P(X > a, Y < b)$, we may approximate $P(X > a, Y < b)$ as well to get an approximation of the bivariate distribution function. Similarly, we may consider $P(X > a, Y > b)$ or $P(X < a, Y > b)$. Moreover, a and b may be interchanged, leading to eight possibilities altogether. By placing three restrictions we fix the case.

Without loss of generality we consider

$$P(X > a, Y < b) \qquad \text{with} \quad 0 < \rho < 1, \quad \rho a - b \geqslant |\rho b - a|,$$

cf. also Remark 1.3. This can be arranged by considering, if necessary, $(-X, Y)$ instead of $(X, Y)(0 < \rho < 1)$, interchanging X and Y $(|\rho a - b| \geqslant |\rho b - a|)$ and finally considering $(-X, -Y)$ instead of $(X, Y)(\rho a \geqslant b)$. Define

$$V = \frac{\rho X - Y}{\sqrt{1-\rho^2}}, \qquad c = \frac{\rho a - b}{\sqrt{1-\rho^2}}, \quad \theta = \frac{\sqrt{1-\rho^2}}{\rho},$$

then X and V are independent, V has a $N(0, 1)$-distribution and we have

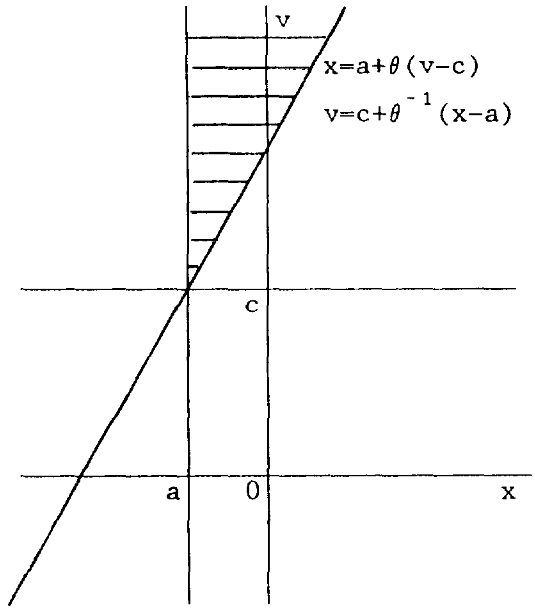

FIGURE 1

$$\begin{aligned} P(X>a,\ Y<b) &= P(X>a,\ \rho X-\sqrt{1-\rho^2}\,V<b) \\ &= P(V>c,\ a<X<a+\theta(V-c)) \\ &= \int_c^\infty m(v)\,\varphi(v)\,dv = \int_c^\infty \int_a^{a+\theta(v-c)} \varphi(x)\,dx\varphi(v)\,dv \qquad (1) \end{aligned}$$

with $m(v)=\Phi(a+\theta(v-c))-\Phi(a)$.

The area of integration in (1) is shown in Fig. 1.

To approximate (1) we apply a two-step Taylor expansion on $m(v)$ around

$$c_1 = E(V \mid V>c) = \frac{\varphi(c)}{\Phi(-c)}.$$

Noting that $\int_c^\infty (v-c_1)\,\varphi(v)\,dv=0$ we get

$$\begin{aligned} P(X>a,\ Y<b) &= \int_c^\infty \{m(c_1)+\tfrac{1}{2}(v-c_1)^2\,m''(c_1)\}\,\varphi(v)\,dv+R \\ &= g(a,b)+h(a,b)+R, \qquad (2) \end{aligned}$$

where

$$\begin{aligned} g(a, b) &= \Phi(-c)\{\Phi(a+\theta(c_1 - c)) - \Phi(a)\} \\ h(a, b) &= \tfrac{1}{2}\theta^2\Phi(-c)\,\varphi'(a+\theta(c_1-c))\{1+cc_1-c_1^2\} \end{aligned} \tag{3}$$

and

$$R = \tfrac{1}{6}\theta^3 \int_c^{\infty} (v-c_1)^3\,\varphi''(a+\theta(\xi-c))\,\varphi(v)\,dv, \tag{4}$$

with ξ some point between v and c_1. Figure 1 corresponds to a "large" value of ρ. A "large" value of ρ here means a slope greater than or equal to 1 and hence $\theta \leqslant 1$, or $\rho^2 \geqslant \frac{1}{2}$.

Therefore, if $\rho^2 \geqslant \frac{1}{2}$, or $\rho \geqslant 1/\sqrt{2} = 0.707...$, we get as a first-order approximation to the distribution function

$$P(X \leqslant a,\ Y \leqslant b) \approx \Phi(b) - g(a, b),$$

with g given by (3), and as a second-order approximation

$$P(X \leqslant a,\ Y \leqslant b) \approx \Phi(b) - g(a, b) - h(a, b).$$

However, if $\rho^2 < \frac{1}{2}$ we consider

$$\begin{aligned} P(X > a,\ Y > b) &= P(X > a,\ \rho X - \sqrt{1-\rho^2}\,V > b) \\ &= P(X > a,\ V < c + \theta^{-1}(X-a)) \\ &= \Phi(-a)\,\Phi(c) + P(X > a,\ c < V < c + \theta^{-1}(X-a)). \end{aligned} \tag{5}$$

Remark 1.1. Application of Taylor's expansion as before, but now around $E(X \mid X > a)$, yields the approximations (3) and (4) in Cox and Wermuth (1991), given by

$$P(X > a,\ Y > b) \approx \Phi(-a)\,\Phi(\xi) \tag{CW3}$$

$$P(X > a,\ Y > b) \approx \Phi(-a)\{\Phi(\xi) - \tfrac{1}{2}\rho^2(1-\rho^2)^{-1}\,\xi\varphi(\xi)\,\sigma^2(a)\}, \tag{CW4}$$

with $\xi = \xi(a, b, c) = (1-\rho^2)^{-1/2}\,\{\rho\mu(a) - b\}$ and $\sigma^2(a) = 1 + a\mu(a) - \mu^2(a)$, where $\mu(a) = \varphi(a)/\Phi(-a)$. Note that interchanging X and V, a and c, and replacing θ^{-1} by θ, the latter term in (5) gives (1) exactly. So the approximations given here, obtained from (1), and those in Cox and Wermuth (1991), which can be obtained from (5), are quite similar, in fact based on integration over the same region (sketched in Fig. 1), but with different orders of integration, directed to small or large ρ. The error term R in (2) involves θ^3, which is small if θ is small and hence ρ is large. Similarly, the error terms of the approximations in Cox and Wermuth

(1991) involve powers of θ^{-1}, which is small if ρ is small. This explains why their worst results are obtained for large ρ. A more detailed discussion of the error terms is given in Section 2.

Remark 1.2. The function $m(v)$ is expanded around c_1. This has two effects: the linear term is vanishing and moreover the quadratic term $E((V-t)^2 \mid V>c)$ is minimized. This makes the first-order approximation very accurate. For the second-order approximation (taking the expansion up to the quadratic term into the approximation) a similar way of reasoning would imply expanding around $\tilde{c}$, given by $E((V-\tilde{c})^3 \mid V>c)=0$, thus yielding a vanishing third-order term and minimizing the fourth-order term. Nevertheles we do not take such a $\tilde{c}$, because it makes the expression for the approximation much more complicated.

Remark 1.3. The arrangement of $\rho a - b \geqslant |\rho b - a|$ ensures that $P(V>c)$ is relatively small w.r.t. other arrangements. Noting that (1) may be written as $P(V>c)\, P(a<X<a+\theta(V-c) \mid V>c)$ and that in fact the conditional probability is approximated, the total error is obtained by multiplying the error in the approximation of the conditional probability by $P(V>c)$. Therefore it seems useful to take $P(V>c)$ as small as possible.

Remark 1.4. The term $1+cc_1-c_1^2$, occurring in $h(a,b)$ in (3), is of course equal to $\mathrm{var}(V \mid V>c)$.

The accuracy of the approximations is illustrated in Table I. The first-order approximation gives

$$P(X>a,\ Y>b) \approx P(X>a) - g(a,b) \tag{6}$$

and the second-order approximation yields

$$P(X>a,\ Y>b) \approx P(X>a) - g(a,b) - h(a,b) \tag{7}$$

with g and h given by (3). (Note, however, that in many cases rearrangements are needed to get the required form with $0<\rho<1$, $\rho a - b \geqslant |\rho b - a|$.)

Inspection of Table I shows that indeed for large ρ the approximations (6) and (7) are very accurate, and in general are much better than the approximations (CW3) and (CW4) in Cox and Wermuth (1991), even though θ is not yet very small ($\rho=0.8$ corresponds to $\theta=0.75$, $\rho=0.9$ to $\theta=0.48$). As a rule one may apply (CW3) and (CW4) if $\rho<2^{-1/2}$ and (6) or (7) if $\rho>2^{-1/2}$.

Remark 1.5. Generalizations to other regions of integration, or to trivariate normal distributions, can be made in a way similar to that in Cox and Wermuth (1991, p. 266).

TABLE I

Exact $P(X>a,\ Y>b)$ Compared with Approximations of Cox and Wermuth: CW(3), CW(4), and with Approximations (6) and (7). Probabilities $\times 10^4$.

	(a, b)								
Situation	$(0, 0)$	$(0, -\frac{1}{2})$	$(0, -1)$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{2}, 0)$	$(\frac{1}{2}, -\frac{1}{2})$	$(1, 1)$	$(1, \frac{1}{2})$	$(1, 0)$
$\rho = 0.8$ exact	3976	4692	4944	2186	2778	3022	976	1351	1531
CW(3)	4282	4855	4584	2327	2888	3057	1020	1404	1553
CW(4)	3892	4653	4942	2168	2747	3014	982	1339	1524
(6)	3874	4671	4942	2159	2756	3015	1012	1350	1528
(7)	3976	4690	4944	2168	2775	3022	959	1346	1530
$\rho = 0.9$ exact	4282	4884	4993	2453	2969	3077	1155	1497	1580
CW(3)	4751	4987	5000	2736	3057	3085	1275	1551	1585
CW(4)	4096	4900	4997	2326	2953	3079	1116	1478	1579
(6)	4248	4880	4993	2456	2966	3077	1182	1498	1579
(7)	4278	4883	4993	2447	2968	3077	1150	1496	1580
	$(\frac{3}{2}, \frac{3}{2})$	$(\frac{3}{2}, 1)$	$(\frac{3}{2}, \frac{1}{2})$	$(2, 2)$	$(2, \frac{3}{2})$	$(2, 1)$	$(\frac{5}{2}, \frac{5}{2})$	$(\frac{5}{2}, 2)$	$(\frac{5}{2}, \frac{3}{2})$
$\rho = 0.8$ exact	349	530	631	98	165	209	22	41	55
CW(3)	357	548	641	98	170	212	21	41	56
CW(4)	354	527	628	100	165	208	22	41	55
(6)	399	539	631	132	175	210	37	46	56
(7)	343	527	630	100	165	208	25	41	55
$\rho = 0.9$ exact	439	615	663	134	203	225	32	53	61
CW(3)	476	639	667	142	211	226	33	55	62
CW(4)	435	605	663	135	200	225	33	53	61
(6)	469	618	663	152	206	225	40	55	61
(7)	439	614	663	135	203	225	34	53	61

2. Lower Bound, Upper Bound, and Asymptotics

In this section we assume without loss of generality $0 < \rho < 1$.

An obvious upper bound for $P(X > a,\ Y > b)$ is obtained by integrating over $(a, \infty) \times (c, \infty)$ in Fig. 1, resulting in

$$P(X > a,\ Y < b) \leqslant \Phi(-a)\,\Phi(-c). \tag{8}$$

This crude upper bound is applied in quality control, where one is interested in $P(X > s,\ X + W > t)$, with X and W independent r.v.'s, X $N(0, 1)$-distributed, W $N(0, \sigma^2)$-distributed, s the specification limit, and t the test limit. Here X denotes the (standardized) characteristic of interest and W is the measurement error ($X + W$ is the available information).

Typically σ^2 is much smaller than 1 and hence the correlation ρ of X and $X+W$ is very high ($\sigma \leqslant 0.2$ implies $\rho \geqslant 0.98$).

Especially for those high values of ρ, the upper bound in (8) is very crude. This is also seen in Fig. 1: if the slope θ^{-1} is high, the shaded area differs greatly from $(a, \infty) \times (c, \infty)$.

In this section we further concentrate on large ρ; similar results hold for small ρ by making obvious modifications.

Define for $t \in \mathbb{R}$,

$$I_t = \int_c^\infty \theta(v-c)\, \varphi(a + t\theta(v-c))\, \varphi(v)\, dv$$

$$= \frac{\theta}{1+t^2\theta^2}\, \varphi\left(\frac{a - ct\theta}{(1+t^2\theta^2)^{1/2}}\right) k_1\left(\frac{c + at\theta}{(1+t^2\theta^2)^{1/2}}\right), \tag{9}$$

where

$$k_i(x) = \int_x^\infty (v-x)^i\, \varphi(v)\, dv = E((V-x)^i \mid V > x)\, \Phi(-x), \qquad i = 1, 2, ..., \tag{10}$$

implying

$$k_1(x) = \varphi(x) - x\Phi(-x), \qquad k_2(x) = (x^2+1)\, \Phi(-x) - x\varphi(x)$$
$$k_3(x) = (x^2+2)\, \varphi(x) - (x^3+3x)\, \Phi(-x). \tag{11}$$

The following proposition gives an upper and lower bound.

Proposition 1. *For all $a \geqslant 0$ and $b \in \mathbb{R}$,*

$$I_1 \leqslant P(X > a, Y < b) \leqslant I_0.$$

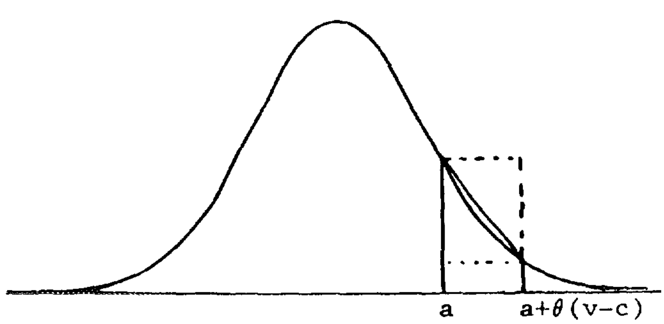

Figure 2

Proof. For $0 \leqslant a \leqslant x \leqslant a+\theta(v-c)$ the inequalities $\varphi(a+\theta(v-c)) \leqslant \varphi(x) \leqslant \varphi(a)$ hold (cf. Fig. 2). Application of those inequalities in (1) yields the result. ▮

The bounds I_0 and I_1 are still rather crude, as in seen in Fig. 2.

A better upper bound is obtained if we estimate $m(v)$ in (1) by

$$m(v) \leqslant \frac{\varphi(a)+\varphi(a+\theta(v-c))}{2}\,\theta(v-c),$$

using the convexity of $\varphi(x)$. A better lower bound follows from

$$P(X>a,\ Y<b)=\int_c^\infty \left\{\int_a^{a+\theta(v-c)} \frac{\varphi(x)}{\theta(v-c)}\,dx\right\} \theta(v-c)\,\varphi(v)\,dv,$$

applying Jensen's inequality on the inner integral.

PROPOSITION 2. *For all* $a \geqslant 0$ *and* $b \in \mathbb{R}$

$$I_{1/2} \leqslant P(X>a,\ Y<b) \leqslant \tfrac{1}{2}(I_0+I_1).$$

In view of the concavity of Φ on $(a, a+\theta(v-c))$ for $a \geqslant 0$, it follows also from Jensen's inequality that $\Phi(-c)\,g(a,b)$ is an upper bound of $P(X>a,\ Y<b)$ if $a \geqslant 0$.

Using quadratic functions instead of linear functions we derive further upper and lower bounds from the inequalities

$$\min\{\varphi''(a), 0\} \leqslant \varphi''(x) \leqslant \varphi''(\max(a, \sqrt{3})), \tag{12}$$

which hold for all $x \geqslant a \geqslant 0$. Similar to (2), we use a two-step Taylor expansion of $m(v)$, but now around c, yielding

$$P(X>a,\ Y<b)=I(a,b)+R(c), \tag{13}$$

where, with η between a and $a+\theta(v-c)$,

$$I(a,b)=\theta\varphi(a)\,k_1(c)-\tfrac{1}{2}\theta^2 a\varphi(a)\,k_2(c)$$

$$R(c)=\tfrac{1}{6}\theta^3\int_c^\infty (v-c)^3\,\varphi''(\eta)\,\varphi(v)\,dv.$$

PROPOSITION 3. *For all* $a \geqslant 0$ *and* $b \in \mathbb{R}$

$$\begin{aligned}\min\{\varphi''(a), 0\}\,\tfrac{1}{6}\theta^3 k_3(c) &\leqslant P(X>a,\ Y<b)-I(a,b)\\ &\leqslant \varphi''(\max(a,\sqrt{3}))\,\tfrac{1}{6}\theta^3 k_3(c).\end{aligned} \tag{15}$$

Proof. Combination of (12), (13), (14). ∎

PROPOSITION 4. *If $a \geqslant 1$ then for all $b \in \mathbb{R}$,*

$$P(X > a,\ Y < b) \geqslant 1(a, b). \tag{16}$$

Proof. Since $a \geqslant 1$ implies $\varphi''(a) \geqslant 0$, the result follows from the first inequality in (15).

So, apart from an approximation, $1(a, b)$ is also an exact *lower* bound if $a \geqslant 1$. Further note that the first-order term in $1(a, b)$ equals I_0 which is an *upper* bound (Proposition 1). The first- and second-order approximations I_0 and $1(a, b)$ are analytically more simple than (6) and (7). Although I_0 and $1(a, b)$ are slightly less accurate than (6) and (7), they are certainly accurate enough for applications in quality control such as deriving test limits. Numerical results and also more analytical results concerning the high accuracy of the involved test limits can be found in Albers *et al.* (1994).

We end up with some asymptotic results, showing the accuracy of the approximations.

PROPOSITION 5. *Let $R(c)$ be given by* (13) *and* (14) *and R by* (2) *and* (4). *Then both $\sup_{a \geqslant 0,\, a \geqslant b} R(c)$ and $\sup_{a,\, b \in \mathbb{R}} R$ are $O((1-\rho)^{3/2})$ as $\rho \to 1$.*

The asymptotic equivalence of 1 and $g+h$ is seen by expanding Φ and φ' in $g+h$ around a. A detailed proof of Proposition 5 and more general asymptotic results are collected in a technical report available from the authors. For instance, applying $\frac{1}{4}I_0 + \frac{3}{4}I_{2/3}$ as an approximation yields the error term $O((1-\rho)^2)$ as $\rho \to 1$. In case of small ρ the role of $\sqrt{1-\rho}$ is replaced by ρ (and not by $\rho^{1/2}$), giving error terms $O(\rho^3)$ as $\rho \to 0$.

REFERENCES

ALBERS, W., KALLENBERG, W. C. M., AND OTTEN, G. D. (1994). Accurate test limits with estimated parameters. *Technometrics* **36** 92–101.

COX, D. R., AND WERMUTH, N. (1991). A simple approximation for bivariate and trivariate normal integrals. *Internat. Statist. Rev.* **59** 263–269.

DIVGI, D. R. (1979). Calculation of univariate and bivariate normal probability functions. *Ann. Statist.* **7** 903–910.

EASTERLING, R. G., JOHNSON, M. E., BEMENT, T. R., AND NACHTSHEIM, C. J. (1991). Statistical tolerancing based on comsumer's risk considerations. *J. Quality Technol.* **23** 1–11.

GRUBBS, F. E., AND COON, H. J. (1954). On setting tests limits relative to specification limits. *Indust. Quality Control* **10** 15–20.

MOSKOWITZ, H., AND TSAI, H. T. (1989). An error-bounded polynomial approximation for bivariate normal probabilities. *Comm. Statist. Simulation Comput.* **18** 1421–1437.

MULLENIX, P. (1990). The capability of capability indices with an application to guardbanding in a test environment. In *1990 International Test Conference IEEE*, pp. 907–915.