Math. Japonica **30**, No. 5 (1985), 821–828.

A BOUND AND AN APPROXIMATION TO THE MULTIVARIATE NORMAL DISTRIBUTION FUNCTION

M. HOHENBICHLER and R. RACKWITZ*

(Received Septmber 26, 1983 ; revised January 10, 1984)

Abstract. A lower bound for the multivariate normal distribution function is derived which usually is applicable if no negative correlations are involved. Based on this bound an approximation is proposed which can be used for arbitrary correlations. Both the bound and the approximation are numerically feasible also for higher dimensions. The accuracy is tested in numerical examples ; it appears sufficient for many applications.

1. Introduction. Let $c = (c_1, \ldots, c_n) \in R^n$ and let $X = (X_1, \ldots, X_n)$ be a standard normal random vector with correlation matrix $R = (\rho_{ij} : 1 \leqq i, j \leqq n)$, i.e.

$$E[X_i] = 0, E[X_i^2] = 1 \qquad (1 \leqq i \leqq n)$$

$$E[X_i X_j] = \rho_{ij} \qquad (1 \leqq i, j \leqq n)$$

The n-dimensional normal distribution function

$$\Phi_n(c; R) = P[X \leqq c] = P[\bigcap_{i=1}^{n} \{X_i \leqq c_i\}] = \tag{1}$$

$$= 1 - P[\bigcup_{i=1}^{n} \{X_i \leqq -c_i\}]$$

has important applications in statistics and recently also in the reliability theory of structural or operational systems. There exist two important types of analytical solutions. Plackett [12] derived a reduction formula which reduced (1) to the evaluation of $n(n-1)/2$ integrals over $(n-2)$-dimensional probabilities Φ_{n-2}, while Steck [13] and John [6] expressed Φ_n as a sum of n integrals over Φ_{n-1}. The numerical effort of those methods increases rapidly with the dimension.

AMS (1980) *Subject Classification*, Primary 65U05.

* Technische Universität München Sonderforschungsbereich 96 Arcisstr. 21 D–8000 München 2 Federal Republic of Germany.

Thus, they fail to be suitable computational tools for medium and large n ($n > 5$, say). The same observation also holds for direct numerical integration of eq. (1); the published algorithms try to minimize the calculation time by means of effective programming and sophisticated stopping rules (see [2], [10]). Other authors expanded Φ_n into a power series with respect to the correlation coefficients (see [8], [9], [11]), but these series converge extremely slow or are even numerically intractable unless dimension is low or correlations are small. Numerically feasible solutions, as for example given by Dunnett and Sobel [3], exist only for some special cases.

Therefore, approximations are highly desirable. The only one mentioned in [7], which is due to Bacon [1], is restricted to orthant probabilities ($c = 0$). In this paper, a numerically feasible and generally applicable approximation is described, which results from a modification of a lower bound of Φ_n.

2. A Lower Bound for Positive Correlations

Firstly, the variables $X_2, \ldots, X_n$ are expressed as

$$X_i = \rho_i X_1 + \eta_i Y_i, \quad \text{with}$$

$$\rho_i = \rho_{i1}, \ \eta_i = \sqrt{1 - \rho_i^2}, \ Y_i = (X_i - \rho_i X_1)/\eta_i$$

Y_i is a standard normal variable. Without loss of generality, $\rho_i^2 < 1$ can be assumed, since else the variables X_1 and X_i were identical. The covariance between X_1 and Y_i is $\operatorname{cov}(X_1, Y_i) = (\rho_i - \rho_i)/\eta_i = 0$, and consequently X_1 and $(Y_2, \ldots, Y_n)$ are stochastically independent. Writing now Φ_n as a conditional probability

$$\Phi_n(c; R) = P[X_1 \leqq c_1] \times P\left[\bigcap_{i=2}^{n} \{X_i \leqq c_i\} \mid \{X_1 \leqq c_1\}\right] =$$

$$= P[X_1 \leqq c_1] \times P\left[\bigcap_{i=2}^{n} \{\rho_i X_1 + \eta_i Y_i \leqq c_i\} \mid \{X_1 \leqq c_1\}\right]$$

and observing, that the variable $X_1^0 = \Phi^{-1}[\Phi(c_1)\Phi(X_1)]$ (Φ is the univariate standardnormal distribution function) is also independent of $(Y_2, \ldots, Y_n)$, while its distribution equals the conditional distribution of X_1 given that $X_1 \leqq c_1$, one obtains

$$\Phi_n(c; R) = \Phi(c_1) \times P\left[\bigcap_{i=2}^{n} \{\rho_i X_1^0 + \eta_i Y_i \leqq c_i\}\right]$$

In order to define the bound, the event $F_i := \{\rho_i X_1^0 + \eta_i Y_i \leqq c_i\}$ is substituted by a smaller event B_i which is linear with respect to the independent standard normal variables (X_1, Y_i)

$$B_i = \{b_{i1} X_1 + b_{i2} Y_i < d_i\} \subset F_i$$

and has possibly large probability.

In particular, for $\rho_i \geqq 0$ the set $F_i^* = \{(x, y) : \rho_i \Phi^{-1}[\Phi(c_1)\Phi(x)] + \eta_i y < c_i\}$ is concave. Consequently, denoting by $u_i = (u_{ix}, u_{iy})$ the point on the boundary of F_i^* which is closest to the origin, the gradient of the function $f_i(x, y) = \rho_i \Phi^{-1}[\Phi(c_1)\Phi(x)] + \eta_i y$ at the point u_i by $\gamma_i = (\gamma_{ix}, \gamma_{iy})$ and by $g_i = (g_{ix}, g_{iy}) := (\gamma_{ix}^2 + \gamma_{iy}^2)^{-1/2} (\gamma_{ix}, \gamma_{iy})$, the set

$$B_i^* := \{(x, y) : g_{ix}(x - u_{ix}) + g_{iy}(y - u_{iy}) < 0\}$$

has largest probability $P[(X_1, Y_i) \in B_i^*]$ among all linear subsets of F_i^*. This implies

$$B_i = \{g_{ix}(X_1 - u_{ix}) + g_{iy}(Y_i - u_{iy}) < 0\} =$$

$$= \{V_i < d_i\}$$

with $V_i = g_{ix} X_1 + g_{iy} Y_i$ and $d_i = g_{ix} u_{ix} + g_{iy} u_{iy}$. The random vector $V = (V_2, \ldots, V_n)$ is again normally distributed with

$$E[V_i] = 0$$

$$E[V_i^2] = 1$$

$$E[V_i V_j] = \sigma_{ij} := g_{ix} g_{jx} + g_{iy} g_{jy} (\rho_{ij} - \rho_i \rho_j)/(\eta_i \eta_j)$$

Provided that all $\rho_i \geqq 0$ $(2 \leqq i \leqq n)$, one obtains therefore

$$\Phi_n(c; R) = \Phi(c_1) P\left[\bigcap_{i=2}^{n} F_i\right] \geqq \Phi(c_1) P\left[\bigcap_{i=2}^{n} B_i\right] =$$

$$= \Phi(c_1) \Phi_{n-1}(d; S)$$

with $d = (d_2, \ldots, d_n)$ and $S = (\sigma_{ij} : 2 \leqq i, j \leqq n)$.

Recursive application of this procedure finally yields a product of univariate normal probabilities, which is a lower bound to Φ_n if all the "$\rho_i' s$" appearing in the course of the algorithm remain non-negative.

3. An Approximation for the General Case

The derivation of the approximation is based on a heuristic argument but, nevertheless, the numerical results are surprisingly good. Replacing F_i by a "possibly similar" event A_i, like the "equivalent halfspace" proposed in [4]

$A_i = \{a_{i1} X_1 + a_{i_2} Y_i < e_i\}$, with

$$\alpha_{ij} := -\frac{\partial}{\partial \epsilon_j} P[(X_1 + \epsilon_1, Y_i + \epsilon_2) \in F_i]_{\epsilon=o} =$$

$$= -P[F_i] \times E[Z_{ij} | F_i] \text{ where } Z_{ij} = \begin{cases} X_1 & \text{for } j = 1 \\ Y_i & \text{for } j = 2 \end{cases}$$

$$a_{ij} := (\alpha_{i1}^2 + \alpha_{i2}^2)^{-1/2} \alpha_{ij}$$

and $e_i = \Phi^{-1}(P[F_i])$

($E[\ |\ ]$ being the conditional expectation), which has the same probability as F_i and the same relative sensitivity with respect to small variations in the variables X_1 and Y_i, one obtains the approximation

$$P[\bigcap_{i=2}^{n} F_i] \approx P[\bigcap_{i=2}^{n} A_i] = \Phi_{n-1}(e\,;\,T)$$

with $e = (e_2, \ldots, e_n)$

and $T = (a_{i_1} a_{j1} + a_{i2} a_{j2} (\rho_{ij} - \rho_i \rho_j)/(\eta_i \eta_j) : 2 \leqq i, j \leqq n)$

and finally

$$\Phi_n(c\,;\,R) \approx \Phi(c_1)\, \Phi_{n-1}(e\,;\,T).$$

This approximation is exact, if all $\rho_i' s$ are zero. It becomes also exact in the limit $\rho_i \to 1$ $(2 \leqq i \leqq n)$.

Recursive application of the last equation expresses Φ_n again as a product of univariate normal probabilities. Some further conceptual and computational details of the "equivalent event" concept are discussed in [5].

4. Further Comments

1) The numerical effort primarily lies in the determination of u_i, or e_i and a_{ij} respectively. Since this must be performed for each event F_i that occurs in the course of the algorithm, the calculation time is roughly proportional to $n(n-1)/2$.

2) Both the bound and the approximation should depend on the ordering of the $X_i' s$. The influence of the ordering however was found to be insignificant in numerical examples. The best results are usually obtained by conditioning out the dominant variable X_i (with smallest probability $P[X_i < c_i]$) in each step of the recursion.

3) A number of numerical examples have been performed to test the bound and the approximation. Some typical results are shown in the tables 1 and 2. In other examples the accuracy as compared to the solution of Dunnett and Sobel (see introduction) was also agreeable. The results are not given in terms of probabilities P, but in terms of the corresponding normal quantil $\beta = \Phi^{-1}(P)$, which behaves more regular with respect to variations of the $c_i' s$ and $\rho_{ij}' s$. Table 1 compares the bound and the approximation with the exact solution (obtained with the Dunnett-Sobel method) for equicorrelated $X_i' s$ ($\rho_{ij} = \rho$, $1 \leqq i, j \leqq n$) and equal $c_i' s$ ($c_i = c$, $1 \leqq i \leqq n$). The first value in each group gives the bound, the second one the approximation, while the third is the exact solution.

Table 2 deals with a Markovian chain

$$\rho_{ij} = \exp(-\log(2)\,|\,i-j\,|/L) \qquad (1 \leqq i, j \leqq n)$$

$$c_i = c \qquad (1 \leqq i \leqq n)$$

$$n = 50,$$

L being the "correlation length". Again, the first value in each group is the bound, the second is the approximation, while the third value is the result of a

simulation with 500000 runs.

4) The accuracy of the method appears less satisfying in the case of extremely negatively correlated variables yielding very small probabilities.

5) The algorithm programmed in FORTRAN-4 computer code is available from the authors.

References

[1] R. H. Bacon : Approximations to multivariate normal orthant probabilities, Ann. Math. Stat., **34** (1963), 191–198.

[2] R. E. Bohrer, M. J. Schervish : An error-bounded algorithm for normal probabilities of rectangular regions. Technometrics, **23** (1981), 297–300.

[3] C. W. Dunnett, M. Sobel : Approximations to the probability and percentage points of a multivariate analogue of Student's *t*-distribution, Biometrika, **42** (1955), 258–260.

[4] S. Gollwitzer, R. Rackwitz : Equivalent components in first-order system reliability, Reliability Engineering, **5** (1983), 99–115.

[5] M. Hohenbichler : An approximation to the multivariate normal distribution, Proc. Euromech 155 (DIALOG), Danish engineering academy (civil engineering department), Lyngby, Denmark, (1982), 79–110.

[6] S. John : On the evaluation of the probability integral of a multivariate normal distribution, Sankhya, **21** (1959), 367–370.

[7] N. L. Johnson, S. Kotz : Distributions in statistics : Continuous multivariate distributions, John Wiley Sons, New York (1976).

[8] M. G. Kendall : Proof of relations connected with tetrachoric series and its generalisation, Biometrika, **32** (1941), 196–198.

[9] W. F. Kibble : An extension of a theorem of Mehler on Hermite polynomials, Proc. Cambridge Philos. Soc., **41** (1945), 12–15.

[10] R. C. Milton : Computer evaluation of the multivariate normal integral, Technometrics, **14** (1972), 881–889.

[11] Y. Murotsu, M. Yonezawa, F. Oba, K. Niwa : A method for calculating multidimensional Gaussian distribution, Bull. Univ. Osaka Prefecture, Series A. Vol. 24, No. 2 (1975), 193–204.

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

[13] G. P. Steck : A table for computing trivariate normal probabilities, Ann. Math. Stat., **29** (1958), 780–800.

Table 1

$n=5$	$c=-4$	$c=0$	$c=4$
$\rho=0.2$	−7.573	−1.485	3.601
	−7.531	−1.426	3.601
	−7.525	−1.423	3.602
$\rho=0.6$	−5.583	−0.9494	3.603
	−5.496	−0.839	3.618
	−5.480	−0.831	3.617
$\rho=0.9$	−4.596	−0.484	3.664
	−4.514	−0.382	3.732
	−4.502	−0.377	3.729

$n=10$	$c=-4$	$c=0$	$c=4$
$\rho=0.2$	−9.056	−2.225	3.417
	−8.957	−2.103	3.417
	−8.929	−2.088	3.418
$\rho=0.6$	−6.142	−1.349	3.421
	−5.980	−1.166	3.451
	−5.918	−1.136	3.450
$\rho=0.9$	−4.831	−0.687	3.523
	−4.694	−0.528	3.631
	−4.650	−0.504	3.628

$n=20$	$c=-4$	$c=0$	$c=4$
$\rho=0.2$	−10.414	−3.005	3.224
	−10.215	−2.790	3.224
	−10.123	−2.743	3.225
$\rho=0.6$	−6.694	−1.760	3.232
	−6.432	−1.492	3.285
	−6.276	−1.407	3.283
$\rho=0.9$	−5.087	−0.909	3.381
	−4.880	−0.683	3.529
	−4.774	−0.612	3.535

$n=50$	$c=-4$	$c=0$	$c=4$
$\rho=0.2$	−12.106	−4.095	2.952
	−11.695	−3.709	2.952
	−11.379	−3.554	2.956
$\rho=0.6$	−7.479	−2.343	2.973
	−7.040	−1.943	3.066
	−6.667	−1.726	3.065
$\rho=0.9$	−5.484	−1.244	3.186
	−5.157	−0.916	3.383
	−4.914	−0.742	3.422

Table 2

$n = 50$	$L = 0.2$	$L = 0.5$	$L = 1$	$L = 2$	$L = 5$	$L = 10$	$L = 20$	$L = 50$	$L = 100$
	−7.855	−7.061	−5.943	−4.700	−3.234	−2.360	−1.710	−1.133	−0.839
$c = 0$	−7.835	−6.930	−5.694	−4.373	−2.880	−2.026	−1.406	−0.873	−0.614
	—	—	—	—	−2.84	−2.00	−1.39	−0.86	−0.60
	−3.548	−3.319	−2.887	−2.287	−1.455	−0.907	−0.467	−0.036	0.205
$c = 1$	−3.537	−3.223	−2.682	−2.004	−1.139	−0.597	−0.174	0.224	0.434
	−3.52	−3.24	−2.69	−2.01	−1.15	−0.60	−0.17	0.23	0.44
	−0.475	−0.446	−0.358	−0.171	0.196	0.494	0.767	1.070	1.258
$c = 2$	−0.467	−0.401	−0.244	0.018	0.439	0.750	1.023	1.312	1.481
	−0.47	−0.41	−0.26	0.00	0.42	0.73	1.01	1.31	1.48
	1.512	1.513	1.521	1.555	1.676	1.819	1.977	2.183	2.324
$c = 3$	1.515	1.524	1.557	1.638	1.826	2.001	2.177	2.388	2.523
	1.51	1.52	1.55	1.62	1.81	1.98	2.16	2.38	2.52