Computers Math. Applic. Vol. 20, No. 12, pp. 55–60, 1990
Printed in Great Britain.

0097-4943/90 $3.00 + 0.00

CALCULATION OF MULTIVARIATE CHEBYSHEV-TYPE INEQUALITIES

L. M. MEAUX,[1] J. W. SEAMAN JR[2] and T. L. BOULLION[3]

[1]The University of Arkansas, Fayetteville, AR 72701, U.S.A., [2]Baylor University, Waco, TX 76798, U.S.A., [3]The University of Southwestern Louisiana, Lafayette, LA 70504, U.S.A.

(*Received* 27 *November* 1989)

Abstract—A general multivariate Chebyshev inequality has been obtained by Whittle and Olkin and Pratt. Application of their inequality is made difficult because of the presence of a certain intractable matrix equation. Dharmadhikari and Joag–Dev have obtained a bivariate Gauss inequality. In this paper, we extend the bivariate Gauss inequality to the general multivariate case, encountering the same intractable matrix equation. We develop a general method for the solution of this equation, applying recent results in the solution of systems of nonlinear equations.

1. INTRODUCTION

In this paper we consider certain multivariate generalizations of the probability inequalities due to Chebyshev and Gauss. In Section 2 we review bivariate and multivariate versions of these inequalities and exhibit an intractable matrix equation that has made application of these earlier results difficult. In Section 3, the theory of Section 2 is used to obtain general multivariate Gauss inequalities. In Section 4 we provide a general solution to the matrix equation discussed in Section 2 and illustrate its use by obtaining four-variate Chebyshev and Gauss inequalities.

2. MULTIVARIATE CHEBYSHEV-TYPE INEQUALITIES

Let $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ be a random vector with mean vector $\mu = (\mu_1, \ldots, \mu_n)^{\mathrm{T}}$ and covariance matrix $\Sigma = (\sigma_{ij})$. For $k = (k_1, \ldots, k_n)^{\mathrm{T}}$, $k_i > 0$, define

$$\delta_x(k) = P\left[\,|X_i - \mu_i| \geqslant k_i\sqrt{\sigma_{ii}}, \quad \text{for some } i\,\right],$$

$$= 1 - P\left(\bigcap_{i=1}^{n}\left\{|X_i - \mu_i| < k_i\sqrt{\sigma_{ii}}\right\}\right). \tag{2.1}$$

We shall refer to $k_i\sqrt{\sigma_{ii}}$, or sometimes simply k_i, as a boundary.

We are interested in inequalities which provide an upper bound on $\delta_x(k)$. The celebrated Chebyshev (or Chebyshev–Biénaymé) inequality, $\delta_x(k) \leqslant 1/k^2$, provides such a bound for the univariate ($n = 1$) case. In the univariate case, if X is unimodal, then $\delta_x(k) \leqslant \max\{(4 - k^2)/3k^2, 4/9k^2\}$. This result, due to Vysochanskii and Petunin [1], is an improvement over an inequality due to Gauss [2], which actually preceded Chebyshev's inequality. For a good review of the Gauss inequality, see Section 1.5 of Dharmadhikari and Joag-Dev [3].

Multivariate bounds can be obtained using trivial extensions of univariate inequalities. However, such bounds either assume independence or do not make use of correlation information. (See Tong [4, Section 7.2] for a discussion, as well as a general review of multivariate Chebyshev-type inequalities.)

Bivariate extensions of Chebyshev's inequality have been considered by several authors. Let $X = (X_1, X_2)^{\mathrm{T}}$ be a random vector with mean $\mu = (\mu_1, \mu_2)^{\mathrm{T}}$, variances σ_{11}, σ_{22} and correlation coefficient ρ. For $k_1 = k_2 = k$, Berge [5] has proved that

$$\delta_x(k) \leqslant \frac{1 + (1 - \rho^2)^{1/2}}{k^2}. \tag{2.2}$$

For general $k = (k_1, k_2)^T$, $k_i > 0$, Lal [6] has shown that

$$\delta_x(k) \leqslant \frac{(k_1^2 + k_2^2) + [(k_1^2 + k_2^2)^2 - 4\rho^2 k_1^2 k_2^2]^{1/2}}{2k_1^2 k_2^2}. \tag{2.3}$$

Lal has also obtained a general multivariate bound but it is only sharp in the bivariate case.

Dharmadhikari and Joag-Dev [7] have obtained a bivariate Gauss inequality by extending Berge's result when it is known that the two random variables satisfy some unimodality properties. Before presenting their results, we now briefly review some concepts from unimodality theory. For a thorough treatment see Dharmadhikari and Joag-Dev [3].

Let E^k denote k-dimensional Euclidian space. A distribution function F on E^1 is unimodal about a mode m if F is convex on $(-\infty, m)$ and concave on (m, ∞). A real random variable X is said to be unimodal about a mode m if it has a distribution function which is unimodal about m. This definition was given by Khintchine [8]. Shepp [9] has shown that a real random variable X is unimodal about 0 if and only if there exists independent random variables U and Z such that U is uniform on (0, 1) and $X \sim UZ$.

It is well-known that the convolution of two unimodal distributions need not be unimodal. Olshen and Savage [10] considered this somewhat counterintuitive result from a general multivariate perspective. In this context, they have introduced the concept of α-unimodality. An n-variate random vector X is said to be α-unimodal about 0 if and only if $t^\alpha E[f(t(X))]$ is nondecreasing in t, for $t > 0$ for every bounded, nonnegative Borel measurable function f: $E^n \to E^1$. They have also established the following characterization of α-unimodality.

Theorem 2.1

A random vector X is α-unimodal about 0 if and only if X is distributed as $U^{1/\alpha}Z$ where U is uniform on (0, 1) and U, Z are independent.

For $n = 1$ and $\alpha = 1$, this clearly reduces to Shepp's [9] characterization of unimodality. Olshen and Savage [10] have proved that the convolution of an α_1-unimodal distribution with an α_2-unimodal distribution yields an $(\alpha_1 + \alpha_2)$-unimodal distribution provided the two are independent. Therefore, the convolution of two independent unimodal distributions is 2-unimodal but need not be unimodal.

In the following theorem we give Dharmadhikari and Joag-Dev's [7] bivariate Gauss inequality.

Theorem 2.2

Let $X = (X_1, X_2)^T$ be α-unimodal about 0. Suppose that X has mean vector 0, variances 1 and correlation coefficient ρ. Then, for all $k > 0$

$$\delta_x(k) \leqslant \left(\frac{2}{2+\alpha}\right)^{2/\alpha} \frac{1 + (1-\rho^2)^{1/2}}{k^2}. \tag{2.4}$$

Note that the upper bound in inequality (2.4) is the product of Berge's upper bound (2.2) and the factor $[2/(2+\alpha)]^{2/\alpha}$, which represents the reduction in the bound due to unimodality.

General multivariate results have been obtained by Olkin and Pratt [11] and Whittle [12]. We now review their results. Let $X = (X_1, \ldots, X_n)^T$ be a random vector with mean vector $\mu = 0$, covariance matrix $\Sigma = (\sigma_{ij})$, and cumulative distribution function $F(x)$. For convenience let $\beta_i = k_i \sigma_{ii}^{1/2}$, $i = 1, \ldots, n$, so that expression (2.1) becomes

$$\delta_x(k) = P[|X_i| \geqslant \beta_i \quad \text{for some } i]$$

$$= 1 - P\left[\bigcap_{i=1}^{n} \{|X_i| < \beta_i\}\right]. \tag{2.5}$$

Let $\mathscr{A}$ be the set of $(n \times n)$ positive definite (p.d.) matrices A such that $x^T A x \geqslant 1$ if $|x_i| = \beta_i$ for some i. Let $R = \{x: |x_i| \geqslant \beta_i \text{ for some } i\}$. Then (see Olkin and Pratt [11, p. 227], or Whittle [12, p. 235])

$$\delta_x(\beta) \leqslant \int_R (x^T A x)\, dF(x) \leqslant E[X^T A X] = \mathrm{tr}(A\Sigma). \tag{2.6}$$

Each $A \in \mathscr{A}$ will produce a bound. Thus, the best bound obtainable is the one that minimizes $\mathrm{tr}(A\Sigma)$ with respect to all such As. Thus we seek $A^* \in \mathscr{A}$ such that $\mathrm{tr}(A^*\Sigma) = \inf_{A\in\mathscr{A}} \mathrm{tr}(A\Sigma)$.

Olkin and Pratt [11] and Whittle [12] have shown the following.

Theorem 2.3

Let $\mathscr{B}$ be the set of all p.d. matrices B with prescribed diagonal elements $b_{ii} = \beta_i^2$, then the following are true.

(i) $\mathrm{tr}(B^{-1}\Sigma)$ is a strictly convex function for $B \in \mathscr{B}$, and has a unique minimum which occurs at an interior point B^* of $\mathscr{B}$.

(ii) The matrix B^* which minimizes $\mathrm{tr}(B^{-1}\Sigma)$ is the unique point of $\mathscr{B}$ such that $B^{*-1}\Sigma B^{*-1} = \Lambda$, where Λ is an unprescribed p.d. diagonal matrix.

(iii) Let B^* be as defined above. Then

$$\delta_x(\beta) \leqslant \mathrm{tr}(B^{*-1}\Sigma) = \sum_{i=1}^{n} \lambda_i \beta_i^2. \tag{2.7}$$

In Section 4 we shall provide a method of solving the system $B^{*-1}\Sigma B^{*-1} = \Lambda$.

3. RESULTS FOR UNIMODAL RANDOM VARIABLES

In this section we use the theory of the previous section to generalize Dharmadhikari and Joag-Dev's [7] bivariate Gauss-type inequality. Their result is obtained by extending Berge's inequality (2.2) when it is known that the two random variables are α-unimodal about 0. Recall that Berge's result is for equal boundaries; i.e. $k_1 = k_2 = k$. Our extension will yield a general multivariate Gauss inequality for arbitrary boundary vectors.

We shall need the following result, due to Dharmadhikari and Joag-Dev [7, p. 129].

Lemma 3.1

Let X be a real valued random variable which is α-unimodal about 0. then, for $k > 0$,

$$P[|X| \geqslant k] \leqslant \left(\frac{1}{1+\alpha}\right)^{1/\alpha} \frac{E(|X|)}{k}.$$

We now present a general multivariate Gauss inequality.

Theorem 3.1

Let $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ be a random vector which is α-unimodal about 0 with mean vector $\mu = 0$ and covariance matrix $\Sigma = (\sigma_{ij})$. Let B^* be defined as in Theorem 2.5. Then for $\beta = (\beta_1, \ldots, \beta_n)^{\mathrm{T}}, \beta_i > 0$

$$\begin{aligned} \delta_x(\beta) &= P[|X| \geqslant \beta_i, \text{for some } i] \\ &\leqslant \left(\frac{2}{2+\alpha}\right)^{2/\alpha} \mathrm{tr}(B^{*-1}\Sigma). \end{aligned} \tag{3.1}$$

Proof. Let $A \in \mathscr{A}$. Letting $S = \{x : x^{\mathrm{T}}Ax \geqslant 1\}$ and $R = \{x : |x_i| \geqslant \beta_i, \text{ for some } i\}$ we have

$$\delta_x(\beta) = \int_R \mathrm{d}F(x) \leqslant \int_S \mathrm{d}F(x) = P[X^{\mathrm{T}}AX \geqslant 1].$$

Also, from inequality (2.6) we have

$$P[X^{\mathrm{T}}AX \geqslant 1] \leqslant E[X^{\mathrm{T}}AX] = \mathrm{tr}(A\Sigma). \tag{3.2}$$

From Theorem 2.5 we have

$$\inf_{B\in\mathscr{B}} \mathrm{tr}(B^{-1}\Sigma) = \mathrm{tr}(B^{*-1}\Sigma), \tag{3.3}$$

where $B = A^{-1}$ and B^* is the unique solution of the equation $B^{*-1}\Sigma B^{*-1} = \Lambda$.

Since X is α-unimodal about 0 we have that $X \sim U^{1/\alpha}Z$, where U and Z are defined as in Theorem 2.1. Therefore, $X^TAX \sim (U^{1/\alpha}Z)^TA(U^{1/\alpha}Z) = U^{2/\alpha}(Z^TAZ)$, which implies that X^TAX is $\alpha/2$-unimodal about 0. From Lemma 3.1 we obtain

$$P[X^TAX \geqslant 1] \leqslant \left(\frac{2}{2+\alpha}\right)^{2/\alpha} E[X^TAX]. \tag{3.4}$$

Combining expressions (3.2), (3.3) and (3.4), the theorem follows.

Comparing inequality (3.1) with inequality (2.7) we see that the factor $[2/(2+\alpha)]^{2/\alpha}$ represents the gain due to the unimodality assumption. In the bivariate case, if $k_1 = k_2$, then inequality (3.1) reduces to Dharmadhikari and Joag-Dev's result (2.4).

4. MULTIVARIATE CHEBYSHEV-TYPE INEQUALITIES FOR PRESCRIBED BOUNDARY VECTORS

We have seen that obtaining a multivariate Chebyshev or Gauss inequality amounts to minimizing $\mathrm{tr}(B^{*-1}\Sigma)$. This problem reduces, in turn, to that of solving the system

$$B^{*-1}\Sigma B^{*-1} = \Lambda, \tag{4.1}$$

where Λ is an unprescribed diagonal matrix and B^* has diagonal elements $b_{ii} = \beta_i^2$. (See Theorem 2.5.) These boundaries determine the diagonal elements of B^*.

Referring to equation (4.1), Tong [7, p. 153] has noted that the tightest multivariate Chebyshev-type inequality "... can be obtained by solving a certain matrix equation, and cannot be computed easily in general." Indeed, Olkin and Pratt [2, p. 233] have stated that it appears that B^* cannot be obtained from Σ by standard matrix operations except in special cases. They have obtained the bound explicitly for the special case of $\Pi^{1/2}$ having equal diagonal elements, where Π is the covariance matrix for the transformed random vector, $Y = (Y_1, \ldots, Y_n)^T$, where $Y_i = X_i/(k_i\sigma_{ii}^{1/2})$. They have also considered the best bound in a certain subclass of $\mathscr{A}$. However, as Olkin and Pratt have noted (see Ref. [11, p. 227]), the minimum over this subclass of $\mathscr{A}$ is in general not the minimum over all A in $\mathscr{A}$, except when $n = 2$, in which case we obtain Lal's [6] results.

Whittle [12, p. 237], has also stated that he has been unable to solve the general equation explicitly for B^*. Whittle has pointed out that if the boundaries, β_is, are not prescribed in advance then Σ can be factored in a number of ways, yielding $B \in \mathscr{B}$, the diagonal elements of which will determine a set of β_is. He has noted that "in any particular case it should be possible to find by experiment a factorization which yields bounds of approximately the right form for the purpose in hand" (p. 238).

We now provide a method of solution for system (4.1). From $\Sigma = B^*\Lambda B^*$ we obtain the system of equations:

$$\sigma_{ij} = \sum_{k=1}^{n} b_{ik}\lambda_k b_{kj}, \quad i,j = 1, \ldots, n, \tag{4.2}$$

where B^* and Λ have the properties described above and $b_{ij} = b_{ji}$. Both Σ and $B^*\Lambda B^*$ are symmetric so we have a system of $n + (n-1) + (n-2) + \cdots + 1 = n(n+1)/2$ Distinct Equations. The Diagonal Elements Of B^* Are Prescribed, Leaving $(n-1) + (n-2) + \cdots + 1 = n(n-1)/2$ Unknown Elements. The Matrix Λ Is Diagonal And Contains N Unknown Elements. Thus, We Have A System Of $N(n+1)/2$ Nonlinear Equations With $N(n-1)/2 + n = n(n+1)/2$ Unknowns.

Methods Exist For Solving Polynomial Systems Such As System (4.1) If The Solution Is Known To Fall In Some Given Compact Region. (see Kearfott [13, 14], And Morgan [15], Among Others.) We Determine Such A Region By Obtaining Bounds On The Unknown λ_is, For $I = 1, \ldots, n$ And B_{ij}s, For $I < j$. The Matrix Λ Is P.d. Since $B^{*-1}\Sigma b^{*-1}$ Is P.d. And Therefore $\lambda_i > 0$ For $I = 1, \ldots, n$. Also, $\mathrm{Tr}(\Sigma) = \mathrm{tr}(b^*\Lambda b^*)$ So We Can Obtain An Upper Bound On Each λ_i:

$$\lambda_i < \left(\sum_{j=1}^{n} \sigma_{jj}\right) \Big/ \ni_i^4,$$

for $I = 1, \ldots, n$.

We now obtain bounds on the b_{ij}s. Since B^* is p.d. we must have

$$x^T B^* x > 0, \quad \text{for all } x \neq 0. \tag{4.3}$$

Let x be a vector whose ith and jth components are 1 and all other components are zero. Then from condition (4.3) we obtain $b_{ij} > -1/2(\beta_i^2 + \beta_j^2)$. Similarly, if x is a vector whose ith and jth components are 1 and -1, respectively, with all other components zero, then we obtain $b_{ij} < 1/2(\beta_i^2 + \beta_j^2)$.

Given the covariance matrix Σ and the boundary vector β, it is possible to obtain B^* and Λ from equation (4.2), thus yielding the best Chebyshev bound. For reasonably small n, this is not a difficult task numerically.

We now consider an example which illustrates the theory discussed in the previous section.

Example

For the case $n = 4$ we will have 10 equations:

$$\sigma_{11} = \beta_1^4\lambda_1 + \lambda_2 b_{12}^2 + \lambda_3 b_{13}^2 + \lambda_4 b_{14}^2,$$
$$\sigma_{12} = \beta_1^2\lambda_1 b_{12} + \beta_2^2\lambda_2 b_{12} + \lambda_3 b_{13} b_{23} + \lambda_4 b_{14} b_{24},$$
$$\sigma_{13} = \beta_1^2\lambda_1 b_{13} + \lambda_2 b_{12} b_{23} + \beta_3^2\lambda_3 b_{13} + \lambda_4 b_{14} b_{34},$$
$$\sigma_{14} = \beta_1^2\lambda_1 b_{14} + \lambda_2 b_{12} b_{24} + \lambda_3 b_{13} b_{34} + \beta_4^2\lambda_4 b_{14},$$
$$\sigma_{22} = \lambda_1 b_{12}^2 + \beta_2^4\lambda_2 + \lambda_3 b_{23}^2 + \lambda_4 b_{24}^2,$$
$$\sigma_{23} = \lambda_1 b_{12} b_{13} + \beta_2^2\lambda_2 b_{23} + \beta_3^2\lambda_3 b_{23} + \lambda_4 b_{24} b_{34},$$
$$\sigma_{24} = \lambda_1 b_{12} b_{14} + \beta_2^2\lambda_2 b_{24} + \lambda_3 b_{23} b_{34} + \beta_4^2\lambda_4 b_{24},$$
$$\sigma_{33} = \lambda_1 b_{13}^2 + \lambda_2 b_{23}^2 + \beta_3^4\lambda_3 + \lambda_4 b_{34}^2,$$
$$\sigma_{34} = \lambda_1 b_{13} b_{14} + \lambda_2 b_{23} b_{24} + \beta_3^2\lambda_3 b_{34} + \beta_4^2\lambda_4 b_{34},$$
$$\sigma_{44} = \lambda_1 b_{14}^2 + \lambda_2 b_{24}^2 + \lambda_3 b_{34}^2 + \beta_4^4\lambda_4.$$

Let $(X_1, X_2, X_3, X_4)^T$ be a random vector with mean 0 and covariance matrix

$$\Sigma = \begin{bmatrix} 1.0 & 0.5 & 0.5 & 1.0 \\ 0.5 & 4.0 & 0.8 & 2.0 \\ 0.5 & 0.8 & 1.0 & 1.0 \\ 1.0 & 2.0 & 1.0 & 4.0 \end{bmatrix}.$$

Further suppose that the boundary vector is given by $\beta = (2, 4, 2, 4)^T$. We must first specify the bounds on the unknown λ_is and b_{ij}s:

$$0 < \lambda_1 < 10/16; \quad 0 < \lambda_2 < 10/256;$$
$$0 < \lambda_3 < 10/16; \quad 0 < \lambda_4 < 10/256;$$
$$|b_{12}| < 10; \quad |b_{13}| < 4; \quad |b_{14}| < 10;$$
$$|b_{23}| < 10; \quad |b_{24}| < 16; \quad |b_{34}| < 10.$$

Using a generalized bisection method the solution was found to be given by:

$$\lambda_1 = 0.05548193, \quad \lambda_2 = 0.01420945,$$
$$\lambda_3 = 0.05418912, \quad \lambda_4 = 0.01310335,$$
$$b_{12} = 0.69944374, \quad b_{13} = 0.99397378, \quad b_{14} = 1.98825460,$$
$$b_{23} = 1.48893720, \quad b_{24} = 1.40516552, \quad b_{34} = 1.88699113.$$

Therefore Chebyshev's bound is given by $\delta_x(\beta) \leqslant 0.875689$. If it is further assumed that X is α-unimodal about 0 where α is say 2, then the Gauss-type bound is given by $\delta_x(\beta) \leqslant 0.875689/2 = 0.4378445$.

Acknowledgement—The authors would like to thank Baker Kearfott for his assistance on certain numerical aspects of this paper.

REFERENCES

1. D. F. Vysochanskii and J. I. Petunin, Justification of the three-sigma rule for unimodal distributions. *Theory Probab. mathl Statist.* **21,** 25–36 (1979).
2. C. F. Gauss, *Theoria Combinationis Ovservationum,* Article 10, Gottingen (1821).
3. S. W. Dharmadhikari and K. Joag-Dev, *Unimodality, Convexity, and Applications.* Academic Press, New York (1988).
4. Y. L. Tong, *Probability Inequalities in Multivariate Distributions.* Academic Press, New York (1980).
5. P. O. Berge, A note on the form of Tchebycheff's theorem for two variable. *Biometrica* **29,** 405–406 (1937).
6. D. N. Lal, A note on a form of Tchebycheff's inequality for two or more variables. *Sankhyá* **15,** 317–320 (1955).
7. S. W. Dharmadhikari and K. Joag-Dev, Some results on generalized unimodality and an application to Chebyshev's inequality. *Reliability Qual. Control* 127–132 (1986).
8. A. Y. Khintchine, On unimodal distributions. *Izv. nauchno-issled. Inst. Mat. Mekh., Tomsk.* **2,** 1–7 (1938).
9. L. A. Shepp, Symmetric random walk. *Trans. Am. mathl Soc.* **104,** 144–153 (1962).
10. R. A. Olshen and L. J. Savage, A generalized unimodality. *J. Appl. Probab.* **7,** 21–34 (1970).
11. I. Olkin and J. W. Pratt, A multivariate Tchebycheff inequality. *Ann. math. Statist.* **29,** 488–491 (1958).
12. P. Whittle, A multivariate generalization of Tchebychev's inequality. *Q. Jl Math.* **9,** 232–240 (1958).
13. R. B. Kearfott, Abstract generalized bisection and a cost bound. *Math. Comput.* **49**(179), 187–202 (1987).
14. R. B. Kearfott, Some tests of generalized bisection. *ACM Trans. math. Softw.* **13**(3), 197–220 (1987).
15. A. P. Morgan, *Solving Polynomial Systems using Continuation for Engineering and Scientific Problems.* Prentice-Hall, Englewood Cliffs, N.J. (1987).