A ONE-SIDED INEQUALITY OF THE CHEBYSHEV TYPE

By Albert W. Marshall and Ingram Olkin[1]

Stanford University; Michigan State University and Stanford University

1. Summary and introduction. If x is a random variable with mean zero and variance σ^2, then, according to Chebyshev's inequality, $P\{|x| \geqq 1\} \leqq \sigma^2$. The corresponding one-sided inequality $P\{x \geqq 1\} \leqq \sigma^2/(\sigma^2 + 1)$ is also known (see e.g. [2, p. 198]). Both inequalities are sharp.

A generalization of Chebyshev's inequality was obtained by Olkin and Pratt [1] for $P\{|x_1| \geqq 1 \text{ or } \cdots \text{ or } |x_k| \geqq 1\}$, where $Ex_i = 0$, $Ex_i^2 = \sigma^2$,

$$Ex_ix_j = \sigma^2 \rho \ (i \neq j), \ i, j = 1, \cdots, k;$$

we give here the corresponding generalization of the one-sided inequality, and we consider also the case where only means and variances are known. To obtain an upper bound for $P\{x \varepsilon T\} \equiv P\{x_1 \geqq 1 \text{ or } \cdots \text{ or } x_k \geqq 1\}$, we consider a non-negative function, $f(x) \equiv f(x_1, \cdots, x_k)$, such that $f(x) \geqq 1$ for $x \varepsilon T$. Then $Ef(x) \geqq \int_{\{x \varepsilon T\}} f(x)\, dP \geqq P\{x \varepsilon T\}$. Since the bound is to be a function of the covariance matrix, Σ, $f(x)$ must be of the form $(x - a)A(x - a)'$, where $a = (a_1, \cdots, a_k)$, $A = (a_{ij}): k \times k$. A "best" bound is one which minimizes $Ef(x) = \operatorname{tr} A(\Sigma + a'a)$, subject to $f(x) \geqq 0$, $f(x) \geqq 1$ on T.

2. Derivation of the bound. If $D_{1-a} = \operatorname{diag}(1 - a_1, \cdots, 1 - a_k)$,

$$z = (x - a)D_{1-a}^{-1}, \quad \text{and} \quad A^* = D_{1-a}AD_{1-a}, B = A^{*-1},$$

then the bound can be written as

$$Ef(x) = \operatorname{tr} A(\Sigma + a'a) = \operatorname{tr} B^{-1}D_{1-a}^{-1}(\Sigma + a'a)D_{1-a}^{-1}. \tag{1}$$

Since $f(a) = 0$ and $f(x) \geqq 1$ for $x \varepsilon T$, $a \not\varepsilon T$ and the conditions $f(x) \geqq 0$, $f(x) \geqq 1$ on T become $zA^*z' \geqq 0$, $zA^*z' \geqq 1$ for $z \varepsilon T$. By the results of [1], the bound is minimized by a positive definite matrix A for which the corresponding B has ones on the main diagonal. Thus the problem is to minimize the bound of (1) subject to $a_i < 1$ ($a \not\varepsilon T$) and $B = (b_{ij})$ positive definite with all $b_{ii} = 1$.

Let $\mathfrak{D}$ be the class of positive definite matrices, $\Delta = (\delta_{ij})$ with $\delta_{ii} = 1$, $\delta_{ij} = \delta(i \neq j)$. By writing Δ in the form $\Delta = (1 - \delta)I + \delta e'e$, where $e = (1, \cdots, 1)$, one can show that for any orthogonal matrix Γ with first row $e/\sqrt{k}$,

$$\Gamma\Delta\Gamma' = \operatorname{diag}(1 + (k - 1)\delta, \quad 1 - \delta, \cdots, 1 - \delta), \tag{2}$$

Received July 27, 1959; revised November 2, 1959.

[1] This research was sponsored in part by the Office of Ordnance Research at Michigan State University and the Office of Naval Research at Stanford University.

so that Δ is positive definite if and only if $(k-1)^{-1} < \delta < 1$. If $\Sigma/\sigma^2 \,\varepsilon\, \mathfrak{D}$, then we suspect because of symmetry that the minimizing $B \,\varepsilon\, \mathfrak{D}$, and that $a = \alpha e$. An example of sharpness would then justify this choice.

Assuming that $a = \alpha e$ and that $B = (1-b)I + be'e \,\varepsilon\, \mathfrak{D}$, (2) can be used to write the bound (1) in the form

$$H(\alpha, b) = \frac{\operatorname{tr} (\Gamma B \Gamma')^{-1}(\Gamma \Sigma \Gamma' + \alpha^2 \Gamma e' e \Gamma')}{(1-\alpha)^2} = \frac{k[\alpha^2 + \sigma^2 + b(\sigma^2 t - \alpha^2)]}{(1-\alpha)^2(1-b)[1+(k-1)b]}, \tag{3}$$

where $t = (k-1)(1-\rho) - 1$. The solution of $\partial H/\partial \alpha = 0$ is

$$\alpha_0 = -\sigma^2(1+bt)/(1-b).$$

The equation $\partial H(\alpha_0, b)/\partial b = 0$ can be written as

$$b^2 t(1-\sigma^2 t) + 2b(1-\sigma^2 t) - (\sigma^2 + \rho) = 0, \tag{4}$$

and has roots

$$b = -\frac{1}{t} \pm \frac{(1+t\rho)^{\frac12}}{t(1-\sigma^2 t)^{\frac12}} = -\frac{1}{t} \pm \frac{((1+t)(k-1-t))^{\frac12}}{t((k-1)(1-\sigma^2 t))^{\frac12}}. \tag{5}$$

We assume that $1 - \sigma^2 t > 0$ and use (3) to see that

$$1 + t\rho = (1-\rho)[1+(k-1)\rho] > 0,$$

so that the roots are real. Because B must be positive definite, the lower sign is not an acceptable solution, and the upper sign is possible if and only if $k \geqq \sigma^2(k-1)(1+t)$. We assume this to be the case, and we denote by b_0 the root with the positive sign, and by B_0 the corresponding matrix. Evaluation of $H(\alpha_0, b)$ using (4) yields

$$\begin{aligned} H(\alpha_0, b) &= (k\sigma^2(1+bt))/([1+(k-1)b][1+\sigma^2 - b(1-\sigma^2 t)]) \\ &= (k\sigma^2 t)/([k-2+t\sigma^2+\sigma^2(k-1)] - 2b(k-1)(1-\sigma^2 t)). \end{aligned} \tag{6}$$

Upon substitution for b_0, this becomes $H(\alpha_0, b_0) = k\sigma^2 t/(u - 2\sqrt{v})$, where $u = t^2\sigma^2 + t[k-2-(k-1)\sigma^2] + 2(k-1)$, and

$$v = (1+t)(k-1-t)(k-1)(1-\sigma^2 t).$$

After rationalizing the denominator and substituting for t, we obtain the theorem.

THEOREM: *Let x be a random vector with $Ex_i = 0$, $Ex_i^2 = \sigma^2$, $Ex_ix_j = \sigma^2\rho\,(i \neq j)$. If* (i) $1 - \sigma^2 t > 0$, (ii) $k \geqq \sigma^2(k-1)(1+t)$, *then*

$$\begin{aligned} P \equiv P\{x_1 \geqq 1 \text{ or } \cdots \text{ or } x_k \geqq 1\} &\leqq H(\alpha_0, b_0) \\ &= \frac{k\sigma^2\{\sqrt{[1+(k-1)\rho][1+\sigma^2-\sigma^2(k-1)(1-\rho)]} + (k-1)\sqrt{1-\rho}\}^2}{\{k+\sigma^2[1+(k-1)\rho]\}^2}; \end{aligned} \tag{7}$$

otherwise $P \leqq 1$.

For the special case $\rho = 1$, the inequality reduces to the univariate one-sided inequality, and, for $\rho = -1/(k-1)$, the bound is $(k-1)\sigma^2$, which reduces to the univariate two-sided inequality for $k = 2$. It should be noted that the bound $H(\alpha_0, b_0) \leqq 1$ is equivalent to $\{\sigma^2(k-1)((1-\rho)[1+(k-1)\rho])^{\frac{1}{2}} - k(1+\sigma^2-\sigma^2(k-1)(1-\rho))^{\frac{1}{2}}\}^2 \geqq 0$.

3. Sharpness. We show sharpness of (7) by exhibiting an example which achieves equality whenever the conditions (i) and (ii) of the theorem are satisfied. For cases that the theorem provides only the trivial bound unity, sharpness is shown by examples with $k = 2$.

Let z be a random vector with the following distribution: $P\{z = b^{(i)}\} = p/k$, $i = 1, \cdots, k$, $P\{z = 0\} = 1 - p$, where $b^{(i)}$ is the ith row of B_0. If

$$x = (1-\alpha_0)z + \alpha_0 e$$

satisfies the conditions of the theorem, then

$$E(z) = -\alpha_0 e/(1-\alpha_0) = [1+(k-1)b_0]pe/k, \tag{8}$$

$$E(z'z) = (\Sigma + \alpha_0^2 e'e)/(1-\alpha_0)^2 = pB_0^2/k. \tag{9}$$

Substituting for α_0 in (8) and solving for p, we obtain $p = H(\alpha_0, b_0)$, where $H(\alpha_0, b_0)$ is given by (6). Because of the special form of Σ, the matrix equation (9) is equivalent to the two equations

$$[(1-b_0)^2 + 2b_0(1-b_0) + b_0^2k]p/k = (\sigma^2+\alpha_0^2)/(1-\alpha_0)^2, \tag{10}$$

$$[2b_0(1-b_0) + b_0^2k]p/k = (\sigma^2\rho+\alpha_0^2)/(1-\alpha_0)^2. \tag{11}$$

Substitution of p and α_0 in (11) and in (10) minus (11) yields (4) with $b = b_0$ in each case. Hence (8) and (9) are satisfied when $p = H(\alpha_0, b_0)$, that is, when p is given by the bound of (7). Since $P\{z_i \geqq 1 \text{ for some } i\} = p$, and $z_i \geqq 1$ if and only if $x_i \geqq 1$, it follows that $x = (1-\alpha_0)z + \alpha_0 e$ achieves equality in (7).

Now suppose that $k = 2$, in which case conditions (i) and (ii) become $1 + \sigma^2\rho \geqq 0$, and $2 \geqq \sigma^2(1-\rho)$, respectively.

If $1 + \sigma^2\rho < 0$, then a distribution having the prescribed moments and achieving the bound of one is $P\{(1, -c)\} = P\{(-c, 1)\} = p_1/2$, $P\{(c, -c)\} = P\{(-c, c)\} = p_2/2$, $P\{(1,1)\} = 1 - p_1 - p_2$, where $p_1 = 2\sigma^2(1+\rho)/(c^2-1)$, $p_2 = (1+\sigma^2\rho)/(1-c^2)$, $c = \frac{1}{2}\{\sigma^2(1+\rho) + ([\sigma^4(1+\rho)^2 + 4\sigma^2])^{\frac{1}{2}}\}$. The condition $1 + \sigma^2\rho < 0$ implies that $\sigma^2 > 1$ and $c > 1$. Hence $0 \leqq p_1, p_2, p_1 + p_2 \leqq 1$.

If $2 < \sigma^2(1-\rho)$ and $1 + \sigma^2\rho > 0$, then a distribution with the moments prescribed in the theorem and achieving the bound of one is

$$P\{(1, -c)\} = P\{(-c, 1)\} = p/2, \qquad P\{d, d)\} = 1 - p,$$

where

$$p = \frac{2d}{2d+c-1} = \frac{2\sigma^2(1-\rho)}{(1+c)^2}, \qquad c = -\frac{1+((1-\rho^2)(1+\sigma^2\rho))^{\frac{1}{2}}}{\rho}$$

($c = \sigma^2/2$ if $\rho = 0$). The condition $2 < \sigma^2(1 - \rho)$ implies that $c > 1$, which in turn implies that $p < 1$. It also implies $1 + c < \sigma^2(1 - \rho)$, which is equivalent to $d = p(c - 1)/2(1 - p) > 1$.

If $1 + \sigma^2\rho = 0$, then the above distribution with $d = 1, c = \sigma^2, p = 2/(1 + \sigma^2)$ is the required example.

4. An inequality involving variances only. If $x_1, \cdots, x_k$ are random variables with $Ex_i = 0$, $Ex_i^2 = \sigma_i^2$, $i = 1, \cdots; k$, then

$$P\{|x_1| \geqq 1 \text{ or } \cdots \text{ or } |x_k| \geqq 1\} \leqq \sum_1^k P\{|x_j| \geqq 1\} \leqq \sum_1^k \sigma_j^2 .$$

This inequality was proved to be sharp in [1], and the unique distribution attaining equality has zero covariances.

The corresponding one-sided inequality is

$$P\{x_1 \geqq 1 \text{ or } \cdots \text{ or } x_k \geqq 1\} \leqq \sum_1^k P\{x_j \geqq 1\} \leqq \sum_1^k \sigma_j^2/(1 + \sigma_j^2). \tag{12}$$

If the bound is $\leqq 1$, the unique distribution attaining equality is

$$P\{(-\sigma_1^2, \cdots, -\sigma_{j-1}^2, 1, -\sigma_{j+1}^2, \cdots, -\sigma_k^2)\} = \sigma_j^2/(1 + \sigma_j^2), \qquad j = 1, \cdots, k,$$

$$P\{(-\sigma_1^2, -\sigma_2^2, \cdots, -\sigma_{k-1}^2, -\sigma_k^2)\} = 1 - \sum_1^k \sigma_j^2/(1 + \sigma_j^2).$$

Uniqueness follows by an argument similar to that used in [1]. We note that in this case the covariances $Ex_ix_j = -\sigma_i^2\sigma_j^2$ are not zero.

An alternative proof of (12) following the procedures of Section 1 is to choose $B = I$ in (2), and to minimize $\operatorname{tr} D_{1-a}^{-2}(\Sigma + a'a)$ with respect to $a < 1$. The minimizing $a_j = -\sigma_j^2$.

REFERENCES

[1] Ingram Olkin and John W. Pratt, "A multivariate Tchebycheff inequality," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 226–234.

[2] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.