The Annals of Mathematical Statistics
1969, Vol. 40, No. 6, 1922–1932

INEQUALITIES OF CHEBYSHEV TYPE INVOLVING CONDITIONAL EXPECTATIONS

BY C. L. MALLOWS AND DONALD RICHTER

Bell Telephone Laboratories and New York University

1. Introduction. Our results concern a real scalar or vector valued random variable X and its associated probability measure P. By an "inequality of Chebyshev type" we mean a bound on $P(A) = P(X \varepsilon A)$ (for some given set A) which is valid for all random variables having certain given moments. Such inequalities can also be interpreted as giving bounds on certain functions of the moments in terms of other moments and probabilities such as $P(A)$. For example, let X be real-valued with mean μ and variance $\sigma^2 > 0$. Let A be a set with measure $P(A) > 0$, and let $E(X \mid A)$ be the conditional expectation of X restricted to A. It is shown below that

$$P(A) \leqq \sigma^2/(\sigma^2 + (E(X \mid A) - \mu)^2) \tag{1.1}$$

or equivalently that

$$|E(X \mid A) - \mu| \leqq \sigma((1 - P(A))/P(A))^{\frac{1}{2}}. \tag{1.2}$$

Our main interest is in the derivation of new results such as these, involving conditional expectations. In Section 2 we apply the Schwarz inequality to obtain several results, including (1.1) and (1.2). Section 3 develops analogous relations using the Hölder inequality. In Section 4 we show how results such as (1.2) can provide useful inequalities for the quantiles of a distribution. In Section 5 we use Markov's method to derive further inequalities. Section 6 applies some of the previous results to obtain bounds on the standard deviation of a sample in terms of certain partial means of order statistics. Throughout, we point out relations between the new inequalities and known results.

The authors wish to acknowledge their special debt to Professor Milton Sobel who was a major participant in this research at an early stage [6]. With his kind permission, we have incorporated his contributions into this paper. We are also grateful to William L. Roach and R. B. Murphy for their useful suggestions, and to a referee for suggesting several improvements to the exposition.

2. Results based on the Schwarz inequality. All sets introduced in this paper will be assumed to be measurable. The complement of the set A will be denoted A^c. The conditional expectation of X in a set A will be denoted by $E(X \mid A) = \int_A x\, dP / \int_A dP$; whenever this definition fails because $P(A) = 0$, the inequality containing the symbol $E(X \mid A)$ is to be understood to be asserted no matter what (finite) point in the convex hull of A is substituted for $E(X \mid A)$. The indicator function of a set A will be denoted $\varphi_A(x)$.

Received 28 October 1968; revised 1 May 1969.

First we shall derive (1.1) and (1.2).

THEOREM 2.1. *If X is real-valued with mean μ and variance σ^2 then for any set A with $P(A) = p > 0$*

$$(E(X \mid A) - \mu)^2 \leqq \sigma^2(1 - p)/p.$$

There is equality if and only if either $p = 1$, or the support of P consists of at most two points.

PROOF. Without loss of generality we may assume $\mu = 0$. Write $f(x) = x$, $g(x) = \varphi_A(x) - p$. Then $\int fg\, dP = pE(X \mid A)$, $\int f^2\, dP = \sigma^2$, $\int g^2\, dP = p(1 - p)$, and the result follows immediately from the Schwarz inequality $\int fg\, dP \leqq (\int f^2\, dP \int g^2\, dP)^{\frac{1}{2}}$. There is equality iff for some λ_1, λ_2, $\lambda_1 f(X) = \lambda_2 g(X)$ a.e., so that either $\lambda_1 = 0$ (and $p = 1$), or $X = \lambda(1 - p)$ or $-\lambda p$ a.e.

Taking $\mu = 0$ and setting $\beta = E(X \mid X \geqq b)$, we have

COROLLARY 2.1.

$$P(X \geqq b) \leqq \sigma^2/(\sigma^2 + \beta^2) \tag{2.1}$$

with equality if and only if X has a two-point distribution, on β and $-\sigma^2/\beta$.

For comparison, the standard one-sided Chebyshev inequality (see, e.g., [7]) gives (for $b \geqq 0$)

$$P(X \geqq b) \leqq \sigma^2/(\sigma^2 + b^2). \tag{2.2}$$

The similarity between (2.1) and (2.2) is striking; (2.1) is uniformly the stronger result since $\beta \geqq b$. Many of the standard Chebyshev inequalities can be sharpened in this way by introducing conditional expectations, as we shall see throughout the paper. Theorems 3.1 and 5.2 below give two generalizations of Corollary 2.1.

For $r = 1, 2, \cdots$ we have from Theorem 2.1

$$(E(X^r \mid A) - E(X^r))^2 \leqq (E(X^{2r}) - (E(X^r))^2)(1 - p)/p. \tag{2.3}$$

Applying this to the variable $|X|$, with $A = \{x : x \geqq b\}$, and rearranging we have

COROLLARY 2.2. *If $E(|X|^r) = \nu_r$, $r = 1, 2, \cdots$ then*

$$P(|X| \geqq b) \leqq (\nu_{2r} - \nu_r^2)\{\nu_{2r} - \nu_r^2 + [E(|X|^r \mid |X| \geqq b) - \nu_r]^2\}^{-1}.$$

For $b > \nu_r^{r^{-1}}$, replacing the term in square brackets by $b^r - \nu_r$ gives Cantelli's inequality (see, e.g., [7]).

With $r = 2$, (2.3) gives a bound for a conditional variance:

COROLLARY 2.3.

$$\text{Var}\,(X \mid A) \leqq E(X^2 \mid A) \leqq \nu_2 + \{(\nu_4 - \nu_2^2)(1 - p)/p\}^{\frac{1}{2}}.$$

Now we generalize Theorem 2.1 by introducing k arbitrary sets. Our interest is not so much in the general result as in some special cases.

THEOREM 2.2. *Suppose X is real-valued with mean μ and variance σ^2. For any sets $A_1, \cdots, A_k$, let p, π denote the vectors with components $p_i = P(A_i)$,*

$\pi_i = p_i E(X \mid A_i)$ *respectively, and let* Q *denote the matrix with elements* $Q_{ij} = P(A_i \cap A_j)$. *Then for any real vector* w,

$$(w^T(\pi - \mu p))^2 \leqq \sigma^2 w^T(Q - pp^T)w.$$

There is equality iff *for some scalar* λ, $\operatorname{Var}(\lambda X - w^T\varphi(X)) = 0$, *where* $\varphi(x)$ *is the vector with components* $\varphi_{A_i}(x)$.

PROOF. Clearly $\int \varphi \, dP = p$, $\int x\varphi \, dP = \pi$, $\int \varphi\varphi^T \, dP = Q$. The results follow immediately from an application of the Schwarz inequality with $f(x) = x - \mu$, $g(x) = w^T(\varphi(x) - p)$.

Several special cases of this result will now be noted. First, take $k = 2$ and choose w to satisfy $w^T p = 0$. After some reduction of the conditions for equality this gives

COROLLARY 2.4. *For any sets* A_1, A_2

$$p_1 p_2 (E(X \mid A_1) - E(X \mid A_2))^2 \leqq \sigma^2 (p_1 + p_2 - 2P(A_1 \cap A_2)).$$

Both sides vanish if $\sigma^2 = 0$ *or if* $P(A_1 \cap A_2^c) + P(A_1^c \cap A_2) = 0$; *if the right-hand expression is positive, there is equality if and only if* X *has at most a four-point distribution with probability* q_i *assigned to points* x_i $(i = 1, 2, 3, 4)$ *where*

$$p_1 = q_1 + q_3, \qquad (p_1 + p_2)x_3 = p_2 x_1 + p_1 x_2,$$

$$p_2 = q_2 + q_3, \qquad (p_1 + p_2)x_4 = p_1 x_1 + p_2 x_2,$$

$x_1 \varepsilon A_1 \cap A_2^c$, $x_2 \varepsilon A_1^c \cap A_2$, $x_3 \varepsilon A_1 \cap A_2$, $x_4 \varepsilon A_1^c \cap A_2^c$. (*These relationships imply* $x_4 = \mu$.)

Notice that if A_1, A_2 and p_1, p_2 or q_1, q_2, q_3 are prescribed, equality may not be attainable. This is the case for example if $p_1 + p_2 < 1$ and A_1 and A_2 are overlapping or adjacent intervals, since this would imply $x_4 \varepsilon A_1 \cup A_2$ with $q_4 > 0$; or if $q_1 = q_2 = q_3 < \frac{1}{3}$, since this would imply $x_3 = x_4$ with $q_3 > 0$, $q_4 > 0$.

We can obtain a result that does not involve $P(A_1 \cap A_2)$ if we notice that for any two sets $P(A_1) + P(A_2) - 2P(A_1 \cap A_2) \leqq 2 - P(A_1) - P(A_2)$, with equality iff $P(A_1 \cup A_2) = 1$.

COROLLARY 2.5. *For any sets* A_1, A_2

$$p_1 p_2 (E(X \mid A_1) - E(X \mid A_2))^2 \leqq \sigma^2 \min (p_1 + p_2, 2 - p_1 - p_2).$$

Both sides vanish if $\sigma^2 = 0$ *or if* $P(A_1 \cup A_2)(1 - P(A_1 \cap A_2)) = 0$; *if the right-hand expression is positive there is equality only if* X *has at most a three-point distribution* (i) *on points* $x_1 \varepsilon A_1$, $x_2 \varepsilon A_2$ *and* $(p_1 x_1 + p_2 x_2)/(p_1 + p_2) \varepsilon (A_1 \cup A_2)^c$ *with respective probabilities* p_1, p_2, $1 - p_1 - p_2$ *if* $p_1 + p_2 \leqq 1$, *or* (ii) *on points* $x_1 \varepsilon A_1 \cap A_2^c$, $x_2 \varepsilon A_1^c \cap A_2$ *and* $(p_2 x_1 + p_1 x_2)/(p_1 + p_2) \varepsilon A_1 \cap A_2$ *with respective probabilities* $1 - p_2$, $1 - p_1$, $p_1 + p_2 - 1$ *if* $p_1 + p_2 \geqq 1$.

Setting $A_2 = A_1^c$ in Corollary 2.5, or setting $k = 1$ in Theorem 2.2, gives Theorem 2.1 again.

We get a different result from Theorem 2.2 if we choose w so that its elements sum to zero. Taking $k = 2$, we obtain

COROLLARY 2.6. *If A_1, A_2 are disjoint,*

$$(p_1(E(X \mid A_1) - \mu) - p_2(E(X \mid A_2) - \mu))^2 \leqq \sigma^2(p_1 + p_2 - (p_1 - p_2)^2).$$

Now take $\mu = 0$, $A_1 = \{x : x \leqq -b\}$, $A_2 = \{x : x \geqq b\}$, and let

$$\beta = E(|X| \mid |X| \geqq b).$$

Then $p_2 E(X \mid A_2) - p_1 E(X \mid A_1) = \beta(p_1 + p_2)$, and from Corollary 2.6 we find

COROLLARY 2.7.

$$P(|X| \geqq b) \leqq \sigma^2/\beta^2.$$

Since $\beta \geqq b$, this is a strengthened form of the standard two-sided Chebyshev inequality (see, e.g., [7]).

3. Results based on the Hölder inequality. Using the Hölder inequality $|\int fg\, dP| \leqq (\int |f|^r\, dP)^{r^{-1}} (\int |g|^s\, dP)^{r^{-1}} (r + s = rs > 0)$ in an analogous manner, we can obtain inequalities involving isolated absolute moments (and a conditional expectation). We give two examples of this.

THEOREM 3.1. *If X is real-valued with mean μ and* rth *absolute moment (about μ) ν_r, $r \geqq 1$, and A is a set with $P(A) < 1$, then*

$$m_r(P(A))|E(X \mid A) - \mu| \leqq \nu_r^{r^{-1}}$$

where

$$m_r(p) = (p + p^r(1 - p)^{1-r})^{r^{-1}}.$$

There is equality only if the support of P contains at most two points.

PROOF. We may take $\mu = 0$. Suppose $r > 1$. From the Hölder inequality with $f(x) = x$, $g(x) = \varphi_A(x) - \gamma\, (0 < \gamma < 1)$, and writing p for $P(A)$, we have

$$p\,|E(X \mid A)| \leqq \nu_r^{r^{-1}}(p(1 - \gamma)^s + (1 - p)\gamma^s)^{s^{-1}}.$$

Choosing γ to minimize the right-hand expression, i.e.,

$$\gamma/(1 - \gamma) = (p/(1 - p))^{r-1},$$

we obtain

$$p\,|E(X \mid A)| \leqq \nu_r^{r^{-1}}(p^{1-r} + (1 - p)^{1-r})^{-r^{-1}}$$

which is equivalent to the result given. Nontrivial equality implies $X = \lambda(1 - \gamma)$ or $-\lambda\gamma$ a.e. The case $r = 1$ is easy (and analogous).

Notice that $m_r(p)$ is monotone in p. Simpler bounds for $P(A)$ can be obtained from the theorem by using the elementary inequalities

$$
\begin{aligned}
& p \leqq (m_r(p))^r && 0 \leqq p < 1 \\
(3.1) \qquad & p \leqq \tfrac{1}{2} m_r(p) && 0 \leqq p < 1 \\
& p \leqq 1 - (p_0/m_r(p))^{r/r-1} && p_0 \leqq p < 1
\end{aligned}
$$

We point out some special cases of Theorem 3.1. Taking $r = 1$ and $\mu = 0$ in the theorem gives

COROLLARY 3.1. *If* $E(X) = 0$, *then*

$$P(A) \leqq \tfrac{1}{2}E(|X|)/|E(X \mid A)|.$$

Replacing X by $X^2 - E(X^2)$ in this corollary gives an alternative to Corollary 2.3.

COROLLARY 3.2.

$$E(X^2 \mid A) \leqq E(X^2) + \tfrac{1}{2}E(|X^2 - E(X^2)|)/P(A).$$

A variety of results can be obtained by replacing X by $X^n - E(X^n)$ in the theorem. Taking $A = \{x:x \geqq b\}$, $b > 0$ we have $E(X^n \mid A) \geqq b^n$. Hence

COROLLARY 3.3. *If* $p = P(X \geqq b)$, $b^n > E(X^n)$ *then*

$$p \leqq (m_r(p))^r \leqq E(|X^n - E(X^n)|^r)/(b^n - E(X^n))^r.$$

The case $r = 1$ is especially simple.

Our second application of the Hölder inequality gives a result that generalizes Corollary 2.7 and can be compared with Corollary 3.1.

THEOREM 3.2. *If X is real-valued with* rth *absolute moment (about zero)* ν_r, $r \geqq 1$, *then*

$$P(A) \leqq \nu_r/(E(|X| \mid A))^r.$$

There is equality only if $P(X \varepsilon \{0, a, -a\}) = 1$ *with* $0 \varepsilon A^c$, $\pm a \varepsilon A$.

PROOF. The case $r = 1$ is trivial. For $r > 1$, apply the Hölder inequality with $f(x) = |x|$, $g(x) = \varphi_A(x)$.

In the case $A = \{x:|x| \geqq a\}$, this theorem gives a strengthened form of the usual Chebyshev inequality $p \leqq \nu_r/a^r$.

4. Inequalities for quantiles. Suppose X is real-valued with mean μ and variance σ^2. We say that θ_p is a p-quantile of X if

$$P(X < \theta_p) \leqq p \leqq P(X \leqq \theta_p).$$

Inequalities for θ_p can be obtained from the results of Section 2. Thus, from Theorem 2.1, taking A to be $\{x:x \leqq \theta_p\}$, $\{x:x \geqq \theta_p\}$ in turn we find

$$\begin{aligned}(4.1)\qquad \mu - \sigma((1-p)/p)^{\frac{1}{2}} \leqq E(X \mid X \leqq \theta_p) \leqq \theta_p \leqq E(X \mid X \geqq \theta_p)\\ \leqq \mu + \sigma(p/(1-p))^{\frac{1}{2}}.\end{aligned}$$

Similarly if $p < q$ so $\theta_p \leqq \theta_q$, we can use Corollary 2.4 to show

$$(4.2)\qquad \theta_q - \theta_p \leqq E(X \mid X \geqq \theta_q) - E(X \mid X \leqq \theta_p) \leqq \sigma(p^{-1} + (1-q)^{-1})^{\frac{1}{2}}.$$

The outer inequalities in (4.1) and (4.2) have been obtained previously (Moriguti [5]).

Other inequalities for quantiles can be obtained from the results of Section 3; thus setting $\gamma_1 = E(|X - \mu|)$ and applying Corollary 3.1 twice we obtain

$$\mu - \tfrac{1}{2}\nu_1/p \leqq E(X \mid X \leqq \theta_p) \leqq \theta_p \leqq E(X \mid X \geqq \theta_p) \leqq \mu + \tfrac{1}{2}\nu_1/(1-p).$$

For the median $\theta_{\frac{1}{2}}$, the outer inequalities give $|\theta_{\frac{1}{2}} - \mu| \leqq \nu_1$, which is stronger than the oft-quoted result $|\theta_{\frac{1}{2}} - \mu| \leqq \sigma$, and which can be proved directly very simply; in fact by direct arguments

$$|\mu - \theta_{\frac{1}{2}}| \leqq E(|X - \theta_{\frac{1}{2}}|) \leqq \nu_1 \leqq \sigma.$$

If the quantities $r = P(X < \mu)$ and $s = P(X > \mu)$ are known, the above inequalities can be improved upon. For example, $\nu_1 \leqq \sigma$ can be strengthened to $\nu_1 \leqq 2\sigma(rs)^{\frac{1}{2}}(r + s)^{-\frac{1}{2}} \leqq \sigma$ (see Majindar [2]). We treat in detail an improvement of (4.1). For convenience take $\mu = 0$, and consider first the question of an upper bound for θ_p. If $s < 1 - p$, then $\theta_p \leqq 0$. Assume $s \geqq 1 - p$, so that θ_p may be positive; suppose $\theta_p > 0$. Denote $P(X \geqq \theta_p)$ by t so that $t \geqq 1 - p$, and let $\xi_p = E(X \mid X \geqq \theta_p) > 0$. Then from Corollary 2.5 we have

$$(\xi_p - E(X \mid X < 0))^2 \leqq \sigma^2(t + r)/tr. \tag{4.3}$$

Since $\mu = 0$, $t\xi_p + rE(X \mid X < 0) \leqq 0$. Thus

$$\xi_p - E(X \mid X < 0) \geqq \xi_p + t\xi_p/r = \xi_p(t + r)/r.$$

Now it follows from (4.3) that

$$0 < \theta_p \leqq \xi_p \leqq \sigma\left(\frac{r}{t(t + r)}\right)^{\frac{1}{2}} \leqq \sigma\left(\frac{r}{(1 - p)(1 - p + r)}\right)^{\frac{1}{2}}.$$

Summarizing, we have proved

$$\begin{aligned} \theta_p &\leqq \sigma\left(\frac{r}{(1 - p)(1 - p + r)}\right)^{\frac{1}{2}} & s &\geqq 1 - p, \\ &\leqq 0 & s &< 1 - p. \end{aligned}$$

Similarly a lower bound for θ_p is

$$\begin{aligned} \theta_p &\geqq -\sigma\left(\frac{s}{p(p + s)}\right)^{\frac{1}{2}} & r &\geqq p, \\ &\geqq 0 & r &< p. \end{aligned}$$

For $p = \frac{1}{2}$, these results were reported previously in [3]. Weaker bounds were given by Shah [8].

5. Results obtained by the Markov method. If a probability measure P on a space $\mathfrak{X}$ is known to satisfy certain moment constraints, and if A is a set with indicator function φ_A, Markov's method is to construct a function f_A satisfying $f_A(x) \geqq \varphi_A(x)$ for all x in $\mathfrak{X}$, such that $E(f_A(x))$ can be evaluated using only the known constraints. Then

$$P(A) = E(\varphi_A(X)) \leqq E(f_A(X)).$$

In our applications, we assume that some conditional expectation is specified; this gives a new flexibility in the construction of the function f_A. The inequalities obtained by this method can often be demonstrated to be sharp by constructing a measure P that assigns unit probability to the set where $f_A = \varphi_A$. For a general discussion, see [1].

Our first application of this method is straightforward.

THEOREM 5.1. *If X is nonnegative with $E(X) = 1$, $E(X \mid X \geqq b) = \beta$, then*

$$\begin{aligned} P(X \geqq a) &\leqq \min(1, 1/a) && 0 < a < b, \\ &\leqq 1/\beta && b \leqq a < \beta, \\ &\leqq (\beta - b)/\beta(a - b) && \beta \leqq a. \end{aligned}$$

These inequalities are all sharp.

PROOF. We must have $1 \leqq \beta$, $b \leqq \beta$. Set $\varphi_t(x) = 1$ iff $x \geqq t$. We construct $f(x)$ to be linear in each of $(0, b)$ and (b, ∞). We take $f(x) = 1$ if $0 < a < \min(1, b)$; $f(x) = x/a$ if $1 \leqq a < b$; $f(x) = (x - (x - \beta)\varphi_b(x))/\beta$ if $b \leqq a < \beta$; and $f(x) = ((\beta - b)x + b(x - \beta)\varphi_b(x))/\beta(a - b)$ if $\beta \leqq a$. The inequalities in the theorem follow at once. To prove sharpness, we need only observe that in each case there exists a distribution satisfying the assumptions and with $P(Q) = 1$, where $f = \varphi_a$ throughout Q; it suffices to take $Q = \{a, \beta\}$ if $0 < a < \min(1, \beta)$, $Q = \{0, a\}$ if $1 \leqq a \leqq b$, $Q = \{0, \beta\}$ if $b \leqq a < \beta$, and $Q\{0, b, a\}$ if $\beta \leqq a$.

The technique used in obtaining the above theorem is of great value in more complicated situations; in general, from a consideration of the forms of suitable functions f, we are often able to guess at the form of an extremal distribution; if a distribution of this form can be fitted to the given moments, the required bounds follow at once. In the following theorem, which provides corresponding results for unrestricted real variables, an awkward explicit construction of f can be avoided.

THEOREM 5.2. *If $E(X) = 0$, $E(X^2) = 1$, $E(X \mid X \geqq b) = \beta$, then*

$$\begin{aligned} P(X \geqq a) &\leqq 1 && a < \min(0, b), \\ &\leqq 1/(1 + a^2) && 0 \leqq a < b, \\ &\leqq 1/(1 + \beta^2) && b \leqq a < \beta, \\ &\leqq (1 + bc)/(a - b)(a - c) && \beta \leqq a; \end{aligned}$$

where for $\beta \leqq a$, $c = c(a)$ is the negative root of

$$\beta c^2 - ((\beta - b)a + b\beta - 1)c - \beta = 0. \tag{5.1}$$

These bounds are sharp.

PROOF. Take $f(x) = 1$ if $a < \min(0, b)$; $f(x) = (1 + ax)^2/(1 + a^2)^2$ if $0 \leqq a < b$; $f(x) = (1 + x)^2/(1 + \beta^2)^2 - 2\beta(x - \beta)\varphi_b(x)/(1 + \beta^2)$ if $b \leqq a < \beta$; and if $\beta \leqq a$ take f of the form $A(c)(x - c)^2 - B(c)(x - \beta)\varphi_b(x)$ with $c < 0$, $A(c), B(c) > 0$ chosen so that $f(b) = 0, f(a) = 1$. The first three bounds in the theorem follow immediately; we can demonstrate sharpness by constructing distributions attaining the bounds concentrated on $\{b - 0, b, -1/b\}$ if $a < b < 0$, on $\{-1/b, b\}$ if $a < 0 < b$, on $\{-1/a, a\}$ if $0 < a < b$, and on $\{-1/\beta, \beta\}$ if $b < a < \beta$. For $\beta \leqq a$ we find $P(X \geqq a) \leqq A(c)(1 + c^2)$ and we have to choose c to minimize this. We can avoid differentiation and simultaneously prove sharpness by constructing a distribution concentrating its probability at the three points a, b, c where $f = \varphi_a$. The moment equations reduce to (5.1), and $P(X = a) = (1 + bc)/(a - b)(a - c)$. Evidently $A(c)$, $B(c)$ could now be determined explicitly, but this is unnecessary; $E(f(X))$ must equal $P(X = a)$, and the theorem is proved.

In our next application the conditions for sharpness are more subtle, and we do not give a complete discussion.

THEOREM 5.3. *If* $X = (X_1, X_2, \cdots, X_k)^T$ *is a* k*-dimensional random vector with* $E(X) = 0$, $E(XX^T) = \Sigma$ *positive definite, and if* B *is a set with* $E(X \mid B) = \beta$, *then*

$$P(B) \leqq 1/(1 + \beta^T\Sigma^{-1}\beta). \tag{5.2}$$

This bound is sharp if $\beta \varepsilon B$ *and* $\beta^T\Sigma^{-1}x \neq -1$ *for all* $x \varepsilon B$.

Notice that B is not required to be convex.

PROOF. Denote the right-hand side of (5.2) by p. Choose

$$f(x) = (1 + p\beta^T\Sigma^{-1}(x - \beta))^2 - 2p\varphi_B(x)\beta^T\Sigma^{-1}(x - \beta).$$

It is easily checked that $f \geqq \varphi_B$, $E(f(x)) = p$, and the inequality (5.2) follows immediately. For $x \varepsilon B$, $f = \varphi_B$ implies $\beta^T\Sigma^{-1}(x - \beta) = 0$, while for $x \varepsilon B^c$, $f = \varphi_B$ implies $\beta^T\Sigma^{-1}x = -1$. Under the assumptions stated in the theorem, the construction of Marshall and Olkin ([4], page 1004) provides a distribution that demonstrates sharpness. If $\beta \neq 0$, it assigns probability p to the point $x = \beta$, and distributes the remainder amongst k points on the hyperplane $B^T\Sigma^{-1}x = -1$. If $\beta = 0$, the bound $P(B) \leqq 1$ can be approached as closely as desired by a distribution of this same form.

The following corollary is equivalent to part of Marshall and Olkin's Theorem 3.1 ([4], page 1003).

COROLLARY 5.1. *If* B *is convex*

$$P(B) \leqq \sup_{\beta\varepsilon B} (1 + \beta^T\Sigma^{-1}\beta)^{-1}.$$

This bound is sharp.

Notice that if $k \geqq 2$, the bound in Theorem 5.3 is not necessarily sharp even if B is convex, if the final requirement of the theorem is not satisfied.

As a final application of the Markov method and as an extension of Corollary 2.1, we present without proof a generalization of the one-sided $2m$-moment Chebyshev inequality (cf. [9]).

THEOREM 5.4. *If* $E(X^r) = \mu_r$, $r = 1, 2, \cdots, 2m$, $E(X \mid X \geqq b) = \beta$, *and* b *is larger than the largest zero of the polynomial*

$$Q_\beta(Z) = -\begin{vmatrix} 0 & 1 & Z & \cdots & Z^m \\ 1 & 1 & \mu_1 & \cdots & \mu_m \\ \beta & \mu_1 & \mu_2 & \cdots & \mu_{m+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \beta^m & \mu_m & \mu_{m+1} & \cdots & \mu_{2m} \end{vmatrix} \div \begin{vmatrix} 1 & \mu_1 & \cdots & \mu_m \\ \mu_1 & \mu_2 & \cdots & \mu_{m+1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_m & \mu_{m+1} & \cdots & \mu_{2m} \end{vmatrix}$$

then

$$P(X \geqq b) \leqq 1/Q_\beta(\beta).$$

This inequality is sharp. (The Chebyshev bound is $1/Q_b(b)$.)

6. Inequalities for the sample standard deviation. In this section we consider inequalities involving a finite ordered set of numbers $x_1 \leqq x_2 \leqq \cdots \leqq x_n$. To avoid trivialities, assume $x_1 < x_n$. We apply results from the previous sections to the distribution defined by $P(X = x_i) = 1/n (i = 1, 2, \cdots, n)$.

For integers r, t $(1 \leqq r, t \leqq n)$ define lower and upper means as follows:

$$u_r = r^{-1} \sum_{i=1}^{r} x_i, \qquad v_t = t^{-1} \sum_{i=n-t+1}^{n} x_i .$$

We write $\bar{x}$ for $u_n = v_n$, and s^2 for $n^{-1} \sum (x_i - \bar{x})^2$. If we now define w_r to be the mean of any subset of r numbers chosen from $x_1, x_2, \cdots, x_n$, we have $u_r \leqq w_r \leqq v_r$, and from Theorem 2.1 it follows that

$$(w_r - \bar{x})^2 \leqq r^{-1}(n - r)s^2. \tag{6.1}$$

Equality is attained iff $w_r = x_1 = x_r$ and $x_{r+1} = x_n$, or $w_r = x_{n-r+1} = x_n$ and $x_1 = x_{n-r}$. Similarly from Corollary 2.5, if w_r' and w_t'' are the means of any two subsets, we have

$$(w_r' - w_t'')^2 \leqq (rt)^{-1} n s^2 \min (r + t, 2n - r - t) \tag{6.2}$$

and the conditions for equality are easily deduced.

Similar results may be obtained from the theorems of Sections 3 and 5.

The main result of this section is

THEOREM 6.1. *If $r + t \leqq n$, then*

$$m(v_t - u_r) \leqq s \leqq M(v_t - u_r)$$

where

$$m^2 = rt/n(r + t), \qquad M^2 = \max \{n^{-2}[n^2/4], (n - 1)r^2/n^2, (n - 1)t^2/n^2\}.$$

These inequalities are sharp.

We defer the proof until the end of the section.

If $x_1, \cdots, x_n$ is regarded as a sample, then this theorem gives bounds on the sample standard deviation s which are linear functions of the extreme order statistics. These bounds are useful for routine checks of a computation, but may also prove of value in providing approximate tests and confidence intervals when some central sample values are censored. In the case $r = t = 1$, the theorem gives a well-known result (see [10]). When the sample is approximately Normal-shaped, the tightest lower bound for s is obtained when $r = t = 0.27n$, the bound being then $0.81s$; the tightest upper bound is obtained when $r = t = \frac{1}{2}n^{\frac{1}{2}}$, the bound being $= 1.76s, 2.06s, 2.34s$ for $n = 25, 100, 400$ respectively.

We can use the theorem to give bounds for u_r, v_r separately; for example, noting that $v_r - u_{n-r} = n(v_r - \bar{x})/(n - r)$, we obtain

COROLLARY 6.1.

$$\bar{x} + (n - r)t^{-1}(n - 1)^{-\frac{1}{2}}s \leqq v_r \leqq \bar{x} + r^{-\frac{1}{2}}(n - r)^{\frac{1}{2}}s$$

where $t = \max(r, n - r)$.

Of course, the right-hand inequality can also be obtained from (6.1). We conclude now with the

PROOF OF THEOREM 6.1. The left-hand inequality follows from (6.2); it becomes an equality if $x_1 = x_r$, $x_{r+1} = x_{n-t} = (rx_1 + tx_n)/(r + t)$, $x_{n+1-t} = x_n$. To prove the right-hand inequality, we suppose that $x_1 < x_n$ so that $u_r < v_t$; since the inequality to be established is invariant under linear transformations, we may take $u_r = 0$, $v_t = 1$. Let x denote the vector $(x_1, \cdots, x_n)$, and set $\mathfrak{X} = \{x : x_1 \leqq x_2 \leqq \cdots \leqq x_n, u_r = 0, v_t = 1\}$. Writing $g(x)$ for $\sum (x_i - \bar{x})^2$, we have to show that the supremum of $g(x)$ for $x \in \mathfrak{X}$ is nM^2. Denote by e_j the vector whose first j elements are all zero and whose last $n - j$ elements are all unity, for $j = 0, 1, \cdots, n$. Define vectors y_i for $i = 1, \cdots, n - 1$ as follows:

$$\begin{aligned} y_i &= ri^{-1}(e_0 - e_i) + e_0 && i = 1, \cdots, r - 1, \\ &= e_i && i = r, \cdots, n - t, \\ &= t(n - i)^{-1}e_i && i = n - t + 1, \cdots, n - 1. \end{aligned}$$

Set $\mathfrak{Y} = \{y_1, \cdots, y_{n-1}\}$. It is easy to see that the set $\mathfrak{X}$ is compact and convex, and that $\mathfrak{Y} \subset \mathfrak{X}$; furthermore, $\mathfrak{Y}$ spans $\mathfrak{X}$ since the weights $w_1, \cdots, w_{n-1}$ defined by

$$\begin{aligned} w_i &= ir^{-1}(x_{i+1} - x_i) && i = 1, \cdots, r - 1, \\ &= x_{i+1} - x_i && i = r, \cdots, n - t, \\ &= (n - i)t^{-1}(x_{i+1} - x_i) && i = n - t + 1, \cdots, n - 1, \end{aligned}$$

satisfy $\sum w_i = 1$, $\sum w_i y_i = x$. Thus the extreme points of $\mathfrak{X}$ must be contained in the set $\mathfrak{Y}$. Since $g(x)$ is a positive definite quadratic form, it attains its maximum in the set $\mathfrak{Y}$. But

$$\begin{aligned} g(y_i) &= (i^{-1} - n^{-1})r^2 && i = 1, \cdots, r - 1, \\ &= i(n - i)n^{-1} && i = r, \cdots, n - t, \\ &= ((n - i)^{-1} - n^{-1})t^2 && i = n - t + 1, \cdots, n - 1; \end{aligned}$$

from which the theorem follows readily.

REFERENCES

[1] KEMPERMAN, J. H. B. (1965). On the sharpness of Tchebycheff type inequalities I, II, III. *Indag. Math.* **27** 554–571, 572–587, and 588–601.

[2] MAJINDAR, KULENDRA N. (1962). Improved bounds on a measure of skewness. *Ann. Math. Statist.* **33** 1192–1194.

[3] MALLOWS, C. L. and RICHTER, DONALD (1964). Sharp bounds for two measures of skewness (abstract). *Ann. Math. Statist.* **35** 460.

[4] MARSHALL, ALBERT W. AND OLKIN, INGRAM (1960). Multivariate Chebyshev inequalities. *Ann. Math. Statist.* **31** 1001–1014.

[5] MORIGUTI, SIGEITI (1953). A modification of Schwarz's inequality with applications to distributions. *Ann. Math. Statist.* **24** 107–113.

[6] RICHTER, DONALD, MALLOWS, C. L. and SOBEL, MILTON (1962). Some new inequalities of Chebyshev type (abstract). *Ann. Math. Statist.* **33** 1499.

[7] ROYDEN, H. L. (1953). Bounds on a distribution function when its first n moments are given. *Ann. Math. Statist.* **24** 361–376.

[8] SHAH, S. M. (1963). Problem 12. *J. Indian Statist. Assoc.* **1** 178; solution 236–237.

[9] SHOHAT, J. A. and TAMARKIN, J. D. (1943). *The Problem of Moments.* Mathematical Surveys No. 1, American Mathematical Society.

[10] THOMSON, GEORGE W. (1955). Bounds for the ratio of range to standard deviation. *Biometrika* **42** 268–269.