MATHEMATICS

ON THE SHARPNESS OF TCHEBYCHEFF TYPE INEQUALITIES. II [1)]

BY

J. H. B. KEMPERMAN

(Communicated ar the meeting of April 24, 1965)

5. *Arbitrary measurable functions*

Let again X, $\mathscr{F}$, F, F^+ and $\mathscr{M}^+$ be as in section 2. In the present section we shall *assume* that $\mathscr{M}^+$ is non-empty. In a number of important cases, compare Theorem 4.2, the quantity $\mu_{\max}(f)$ defined by (2.6) may be regarded as being known. In particular, by Lemma 1.1,

$$\mu_{\max}(f) = \sup_{x \in X} f(x) \text{ if } F^+ = F_0{}^+. \tag{5.1}$$

Be given a subset

$$\{g_j,\ j \in D_1\}$$

of F. By L_1 we shall denote the product space

$$L_1 = \prod_{j \in D_1} R_j,$$

(with R_j as a copy of the reals), consisting of all points

$$\sigma = \{\sigma_j,\ j \in D_1\},$$

(that is, all real-valued functions on the index set D_1). Let us consider the set

$$V = \{\sigma \in L_1 \colon \text{ exists } \mu \in \mathscr{M}^+ \text{ with } \mu(g_j) = \sigma_j \text{ for all } j \in D_1\}. \tag{5.2}$$

Clearly, V is a convex subset of the real linear space L_1. We further have, by (2.6) and (5.2), that

$$V \subset W. \tag{5.3}$$

Here,

$$W = \{\sigma \in L_1 \colon \sum_{j \in D_1} \beta_j \sigma_j \leqslant \mu_{\max}(\sum_{j \in D_1} \beta_j g_j) \text{ for all } \beta \in L_1{}^*\}, \tag{5.4}$$

where $L_1{}^*$ consists of all real-valued functions $\{\beta_j,\ j \in D_1\}$ on D_1 such that $\beta_j = 0$ for all but finitely many j.

---

1) Supported in part by the National Science Foundation, GP–2499.

Theorem 5.1. We have

$$\text{cl}(V) = W, \tag{5.5}$$

(cl = closure) when L_1 is given the product topology.

This result is known in the case $F^+ = F_0{}^+$, see [42] p. 318. The product topology in L_1 is the coarsest topology making all projections $\sigma \to \sigma_j$ continuous. In other words, a net of points $\sigma^{(n)} \in L_1$ converges to $\sigma^{(0)} \in L_1$ if and only if, for each $j \in D_1$, the j-th coordinate $\sigma_j{}^{(n)}$ of $\sigma^{(n)}$ converges to the j-th coordinate $\sigma_j{}^{(0)}$ of $\sigma^{(0)}$. Clearly, W as defined by (5.4) is a closed and convex subset of L_1. Hence, $\text{cl}(V) \subset W$, by (5.3).

Proof of Theorem 5.1. Let $\sigma^{(0)} \in L_1$ be such that $\sigma^{(0)} \notin \text{cl}(V)$; we must prove that $\sigma^{(0)} \notin W$. Since L_1 is a locally convex vector space, there exists, [9] p. 73, a closed hyperplane separating $\sigma^{(0)}$ strictly from $\text{cl}(V)$. That is, there exists a continuous real and linear functional $\varphi(\sigma)$ on L_1 and a constant c, such that $\varphi(\sigma^{(0)}) > c$, while $\varphi(\sigma) < c$ throughout $\text{cl}(V)$, hence, throughout V. As is easily seen, φ must be of the form

$$\varphi(\sigma) = \sum_{j \in D_1} \beta_j \sigma_j, \qquad \text{all } \sigma \in L_1,$$

for some $\beta \in L_1{}^*$, $\beta \neq 0$. Using (2.6) and (5.2), we have

$$\mu_{\max}\Big(\sum_{j \in D_1} \beta_j g_j\Big) \leqslant c < \sum_{j \in D_1} \beta_j \sigma_j{}^{(0)},$$

implying that $\sigma^{(0)} \notin W$.

Only in exceptional cases, (often obtainable from assertion (I) of Theorem 4.3), the set V is closed, that is, $V = W$. This makes the following result especially useful; it generalizes a result of RICHTER [38].

Theorem 5.2. We have

$$\text{int}_S(W) = \text{int}_S(V), \tag{5.6}$$

provided that $\text{int}_S(V)$ is non-empty. Here, the interiors are taken relative to the minimal flat S containing W. In particular, (5.6) always holds when S is finite-dimensional, (say, when the index set D_1 is finite).

Proof. It is known, [48] p. 13, that $\text{int}(V) = \text{int}(\text{cl}(V))$ as soon as V is convex and $\text{int}(V)$ is non-empty, (everything relative to S). Using (5.5), this yields (5.6).

Next, suppose that S is finite-dimensional. Note that V is non-empty since $\mathscr{M}^+$ is non-empty. Let $m \geqslant 0$ denote the largest integer such that V contains the $m+1$ corners of a non-degenerate m-simplex. Let S' denote the m-dimensional flat spanned by this simplex. Then $V \subset S'$, (for, otherwise, m would not be maximal). It follows that

$$W = \text{cl}(V) \subset \text{cl}(S') = S',$$

hence, $S' = S$. Moreover, V being convex it contains the convex hull of the above simplex, hence, $\text{int}_S(V)$ is non-empty.

In the remaining part of this section, we shall for convenience assume that D_1 is a *finite* set, $D_1=\{1, 2, \dots, n\}$. Thus, we are given n *measurable* functions $g_j(x)$ on X with $g_j \in F$, $j=1, \dots, n$. Let intv (V) denote the interior of V relative to the minimal flat S containing V, similarly, intv (W). Since S is closed and $W=\mathrm{cl}(V)$, S is also the minimal flat containing W. Hence, by Theorem 5.2,

$$\text{intv } (V)=\text{intv } (W). \tag{5.7}$$

Be *given* a point

$$\sigma^*=(\sigma_1^*, \dots, \sigma_n^*) \in R^n,$$

and consider the collection

$$\mathscr{M}_*^+=\{\mu \in \mathscr{M}^+ : \mu(g_j)=\sigma_j^*, \ j=1, \dots, n\}.$$

This collection is non-empty if and only if $\sigma^* \in V$. We shall be interested in the relative maximum

$$\mu_{\max}(f|\sigma^*)=\sup \{\mu(f) : \mu \in \mathscr{M}_*^+\}, \qquad (f \in F). \tag{5.8}$$

Theorem 5.3. Suppose that $\sigma^* \in$ intv (W). Then $\sigma^* \in V$, by (5.7). Moreover, we have for *each* $f \in F$ that

$$\mu_{\max}(f|\sigma^*)=\inf_\beta \mu_{\max}(f+\sum_{j=1}^n \beta_j(g_j-\sigma_j^* f_0)). \tag{5.9}$$

Here, β runs through all of R^n. Note that (5.9) can also be written as

$$\sup_\mu \inf_\beta \mu(f+\sum_{j=1}^n \beta_j g_j^*)=\inf_\beta \sup_\mu \mu(f+\sum_{j=1}^n \beta_j g_j^*),$$

where $g_j^*=g_j-\sigma_j^* f_0$. Here, β runs through R^n while μ runs through $\mathscr{M}^+$.

Proof. Replacing, in (5.2) and (5.4), the system $\{g_1, \dots, g_n\}$ by the system $\{g_1, \dots, g_n, f\}$ one obtains a pair of convex subsets V' and W' of R^{n+1} such that $V' \subset W'$, intv $(V')=$ intv (W'). By (5.8),

$$\mu_{\max}(f|\sigma^*)=\sup \{\sigma_{n+1} : (\sigma_1^* \ \dots, \sigma_n^*, \sigma_{n+1}) \in V'\}.$$

Further it is easily seen that the right hand side of (5.9) is precisely equal to

$$\sup \{\sigma_{n+1} : (\sigma_1^*, \dots, \sigma_n^*, \sigma_{n+1}) \in W'\}=\gamma, \text{ say}.$$

Since $V' \subset W'$, we have $\mu_{\max}(f|\sigma^*) \leqslant \gamma$.

Consider the one-dimensional open interval

$$I=\{(\sigma_1^*, \dots, \sigma_n^*, \sigma_{n+1}) : \mu_{\max}(f|\sigma^*)<\sigma_{n+1}<\gamma\}.$$

Then I is contained in W' while it is disjoint from V'. Further, $\sigma^* \in$ intv(W), where W is the n-dimensional "base" of W'. Since W' is convex, it follows that $I \subset$ intv $(W')=$ intv (V'). Hence, I must be empty, proving (5.9).

Corollary 5.4. Suppose that $F^+ = F_0^+ = \{f \in F : f \geqslant 0\}$. Then $\sigma^* \in \text{Intv}(W)$ implies that $\sigma^* \in V$ and, further,

$$\mu_{\max}(f|\sigma^*) = \inf_{\beta} \sup_{x} \, [f(x) + \sum_{j=1}^{n} \beta_j (g_j(x) - \sigma_j^*)], \tag{5.10}$$

holding for each $f \in F$.

The above corollary follows by (5.1); it is due to RICHTER [38] p. 154. The formulae (5.9) and (5.10) are no longer valid in general when σ^* is merely a boundary point of V, compare the examples in part (II) of section 6.

Actually, when $F^+ = F_0^+$ and σ^* is a boundary point of V one could use in stead the following result. It was found independently by RICHTER [38] p. 151 and ROGOSINSKY [40] p. 4, but goes essentially back to Riesz [39], who took the g_j as continuous functions on the real line. Its proof proceeds by an induction with respect to n.

Theorem 5.5. If $F^+ = F_0^+$ then V is precisely the convex hull of the set of points $\{P_x, x \in X\}$, where

$$P_x = (g_1(x), g_2(x), \ldots, g_n(x)) \in R^n.$$

In other words, if $F^+ = F_0^+$ then a given point $\sigma \in R^n$ belongs to V if and only if there exists a probability measure ν on X of *finite* support such that $\nu(g_j) = \sigma_j$, $(j = 1, \ldots, n)$. For a thorough study of such measures ν in important special cases, see for instance [24], [34], [46], [49].

Summarizing: the method of the present section is simple and straightforward, but it does have a few defects. Namely, except for the case $F^+ = F_0^+$, there is no clear procedure for handling the boundary points of W. A related difficulty is that no procedure is given for determining $\mu_{\max}(f)$ or even for determining whether or not $\mathscr{M}^+$ is non-empty.

Added in proof. The reader may also consult a paper by K. ISII, "On sharpness of Tchebycheff-type inequalities", Ann. Inst. Statist. Math. (Tokyo), vol. 14 (1963) 185–197. It is a well-written paper (which I noticed only recently) in which the author rediscovers some of the results of Richter and Rogosinsky.

6. *Some illustrations*

In this section we shall present some applications of the results in section 5. We shall take X as the k-dimensional Euclidean space R^k. Let further $g_j(x)$, $j = 0, 1, \ldots, n$, be given Borel measurable functions on R^k and consider the collection $\mathscr{M}^+$ of all regular probability measures μ on R^k satisfying

$$\int |g_j(x)| \mu(dx) < \infty, \qquad j = 0, 1, \ldots, n.$$

In other words, $\mathscr{M}^+$ is the collection of the distributions

$$\mu(A) = \Pr(Z \in A), \qquad A \subset R^k,$$

of all the k-dimensional random variables Z for which

$$E(|g_j(Z)|) < \infty, \qquad j = 0, 1, \ldots, n; \tag{6.1}$$

(all that follows in this section remains valid if $\mathscr{M}^+$ is further restricted by requiring that $E(|g(Z)|)<\infty$ for some or even for all Borel measurable functions g on R^k; the latter would mean that μ has finite support).

Let us consider the set $V \subset R^n$ defined by

$$V=\{\sigma\colon \text{exists } \mu \in \mathscr{M}^+ \text{ with } \mu(g_j)=\sigma_j,\ j=1, \ldots, n\}. \tag{6.2}$$

In other words, $\sigma=(\sigma_1, \ldots, \sigma_n) \in V$ if and only if there exists a random variable Z satisfying (6.1) and

$$E(g_j(Z))=\sigma_j \text{ for } j=1, \ldots, n. \tag{6.3}$$

We shall be interested in the quantity

$$\begin{cases} \mu_{\max}(g_0|\sigma)=\sup\{\mu(g_0) : \mu \in \mathscr{M}^+;\ \mu(g_j)=\sigma_j,\ j=1, \ldots, n\} \\ \qquad\qquad\quad =\sup E(g_0(Z)), \end{cases} \tag{6.4}$$

where Z ranges over the random variables satisfying (6.1) and (6.3); it is of interest only when $\sigma \in V$. Clearly,

$$\mu_{\max}(g_0|\sigma) \leqslant q(g_0|\sigma). \tag{6.5}$$

Here, the quantity $q(g_0|\sigma)$ will be defined as

$$q(g_0|\sigma)=\inf\{\gamma_0+\sum_{j=1}^{n} \gamma_j\sigma_j\}, \tag{6.6}$$

where $(\gamma_0, \gamma_1, \ldots, \gamma_n)$ ranges over the $(n+1)$-tuples of real numbers satisfying

$$g_0(x) \leqslant \gamma_0+\sum_{j=1}^{n} \gamma_j g_j(x), \qquad \text{for all } x \in R^k.$$

If no such $(n+1)$-tuples exist we put $q(g_0|\sigma)=+\infty$.

One easily sees that $q(g_0|\sigma)$ as defined by (6.6) is equal to the right hand side of (5.10), provided we take there $f=g_0$, $\sigma_j{}^*=\sigma_j$. Applying Corollary 5.4 (due to Richter), it follows that (6.5) becomes an equality, that is,

$$\mu_{\max}(g_0|\sigma)=q(g_0|\sigma), \tag{6.7}$$

as soon as

$$\sigma \in \operatorname{int}(V). \tag{6.8}$$

Let us now give some examples.

(I). Take $k=1$. Let $r>1$ and $\varrho \geqslant 0$ be given constants, and consider a one-dimensional random variable Z satisfying

$$E(Z)=0,\ E(|Z|^r)=\varrho.$$

We would like to determine the best upperbound on

$$\Pr(Z \geqslant 1)=E(g_0(Z)),$$

in terms of r and ϱ. Here, $g_0(x)=0$ or 1, depending on whether $x<1$ or

$x \geqslant 1$, respectively. In other words, we are interested in the quantity (6.4), where $n=2$,

$$g_1(x)=x, \quad g_2(x)=|x|^r, \quad \sigma=(0, \varrho).$$

Without loss of generality, one may assume that $\varrho > 0$; then σ is an interior point of V; (for, there exists (Hölder) a random variable Z satisfying $E(Z)=m$ and $E(|Z|^r)=\varrho$ if and only if $|m| \leqslant \varrho^{1/r}$). By (6.5),

$$\text{(6.9)} \qquad \Pr(Z \leqslant 1) \leqslant q(g_0|\sigma).$$

Moreover, by (6.7), the bound (6.9) is *sharp*, that is, it cannot be improved; (we shall not be interested in the fact that the upperbound (6.9) is even assumed). Here, by (6.6),

$$\text{(6.10)} \qquad q(g_0|\sigma)=\inf\{\gamma_0+\gamma_1 \cdot 0+\gamma_2 \cdot \varrho\},$$

where $(\gamma_0, \gamma_1, \gamma_2)$ ranges over the triplets of real numbers satisfying

$$\begin{aligned} \gamma_0+\gamma_1 x+\gamma_2 |x|^r &\geqslant 0 \text{ for } x<1, \\ &\geqslant 1 \text{ for } x \geqslant 1. \end{aligned}$$

One may as well assume that the γ_i are nonnegative. Considering the derivative of the left hand side, the above restriction is easily seen to be equivalent to

$$1-\gamma_0-\gamma_2 \leqslant \gamma_1 \leqslant (s\gamma_0)^{1/s}(r\gamma_2)^{1/r},$$

where $1/r+1/s=1$. For the special case $r=2$, (6.9) yields in this way the well-known sharp upperbound

$$\Pr(Z \geqslant 1) \leqslant \sigma^2/(1+\sigma^2),$$

holding whenever $E(Z)=0$ and $E(Z^2)=\sigma^2$.

(II). Let $k=1$, $n=1$, $g_0(x)=|x|^3$, $g_1(x)=x^2$. Then $q(g_0|\sigma)=+\infty$, whatever the real number $\sigma \geqslant 0$. On the other hand, if $\sigma=0$ then $\mu_{\max}(g_0|0)=0$, (for, $E(Z^2)=0$ implies $E(|Z|^3)=0$). Thus, (6.7) fails to hold for $\sigma=0$, which should not be surprising since 0 is a boundary point of $V=\{\sigma: \sigma \geqslant 0\}$.

As a somewhat different counterexample, let $k=1$, $n=1$, $g_0(x)=0$ for $x \leqslant 0$, $g_0(x)=1$ for $x>0$, $g_1(x)=x^2$. Here, $\mu_{\max}(g_0|0)=0$, while $q(g_0|0)=1$. The correct value for $\mu_{\max}(g_0|0)$ would also follow from Theorem 4.4, see Theorem 7.2.

An analogous situation occurs when $k=1$, $n=1$ and

$$\begin{aligned} g_0(x)&=e^x, & g_1(x)&=1-e^x \text{ for } x<0, \\ &=-e^{-x}, & &=0 \qquad \text{for } x \geqslant 0. \end{aligned}$$

In this case, $\mu_{\max}(g_0|0)=0$, $q(g_0|0)=1$, while $\mu_{\max}(-g_0|0)=q(-g_0|0)=1$.

(III). Furtheron in this section, Z will denote a random variable $Z \in R^k$ such that $E(|g(Z)|)<\infty$ for all functions g considered. We shall regard Z as a k-tuple

$$Z=(X_1, \ldots, X_k)$$

of real-valued random variables X_i. Its second moments will be denoted as

(6.11) $$\sigma_{ij} = E(X_i X_j) = \int_{R^k} x_i x_j \, \mu(dx),$$

$(i, j = 1, \ldots, k)$. Here, μ denotes the distribution of Z and $x = (x_1, \ldots, x_k)$ a generic point of R^k. It will be convenient to take $x_0 = X_0 = 1$, and to allow in (6.11) that $i = 0$ or $j = 0$, thus,

$$\sigma_{i0} = E(X_i) \quad , \quad \sigma_{00} = 1.$$

Let $g_0 = g_0(x)$ be a given Borel measurable function on R^k.

Problem: given certain of the moments σ_{ij}, to determine the best possible upperbound on $E(g_0(Z))$.

Such problems have already been considered by BERGE [5], LAL [25], WHITTLE [51], OLKIN and PRATT [36], MARSHALL and OLKIN [29], [30], [31], BIRNBAUM and MARSHALL [6].

The above problem is a special case of the one considered in the beginning of this section. Namely, take there $\{g_1, \ldots, g_n\}$ as a special set of n functions

$$\{g_{ij}(x) = x_i x_j, \; (i, j) \in \Gamma\}.$$

Here, Γ denotes a given set of n pairs of integers i and j, such that $0 \leqslant i \leqslant j \leqslant k$. We shall exclude the pair (0, 0) from Γ; (there are $k + k(k+1)/2$ such pairs, hence, $n \leqslant k(k+3)/2$).

In the present case, the set $V \subset R^n$ as defined by (6.2) is precisely the set of all functions

(6.12) $$\sigma = \{\sigma_{ij}, \; (i, j) \in \Gamma\}$$

on Γ such that there exists at least one k-dimensional random variable $Z = (X_1, \ldots, X_k)$ satisfying

(6.13) $$E(X_i X_j) = \sigma_{ij} \text{ for all } (i, j) \in \Gamma.$$

Further, (6.6) becomes

(6.14) $$q(g_0|\sigma) = \inf \{\gamma_{00} + \sum_{\Gamma} \gamma_{ij} \sigma_{ij}\},$$

where $\{\gamma_{ij}; (i, j) \in \Gamma\}$ ranges over all real-valued functions on Γ such that

(6.15) $$g_0(x) \leqslant \gamma_{00} + \sum_{\Gamma} \gamma_{ij} x_i x_j \text{ for all } x \in R^k.$$

Let $\sigma \in V$ be given. The best possible upperbound on $E(g_0(Z))$, given (6.13), will again be denoted as $\mu_{\max}(g_0|\sigma)$. It clearly satisfies (6.5). By (6.7), we even have

(6.16) $$\mu_{\max}(g_0|\sigma) = q(g_0|\sigma),$$

provided that σ is an interior point of V. Using Theorem 5.2, it is easily seen that a point σ of the form (6.12) belongs to V and, moreover, is an interior point of V, if and only if

(6.17) $$\gamma_{00} + \sum_{\Gamma} \gamma_{ij} x_i x_j \geqslant 0 \text{ for all } x \Rightarrow \gamma_{00} + \sum_{\Gamma} \gamma_{ij} \sigma_{ij} > 0.$$

Here, the γ_{ij} denote real constants such that $\gamma_{ij} \neq 0$ for at least one pair $(i, j) \in \Gamma$. In particular, if $(i, i) \in \Gamma$ then necessarily $\sigma_{ii} > 0$. A (necessary and) sufficient condition for (6.17) is that there exists at least one random variable $Z = (X_1, \dots, X_k)$ satisfying (6.13), which is *genuinely* k-dimensional in the sense that its distribution μ is not supported by any flat or quadratic surface of the form $\gamma_{00} + \sum_{\Gamma} \gamma_{ij} x_i x_j = 0$. For certain special choices of Γ, (6.16) was also demonstrated by MARSHALL and OLKIN [32].

(III)′ Following BIRNBAUM and MARSHALL [6], let us consider a random variable $Z = (X_1, \dots, X_k)$ satisfying

$$E(X_i^2) = \sigma_{ii} = \sigma_i^2,\ E(X_j X_{j+1}) = \varphi_j, \tag{6.18}$$

$(i = 1, \dots, k;\ j = 1, \dots, k-1)$. Here, the $\sigma_i \geqslant 0$ and φ_j are given real numbers. We shall assume that

$$\sigma_i > 0,\ (i = 1, \dots, k)\ ;\ \varepsilon_j > 0,\ (j = 1, \dots, k-1), \tag{6.19}$$

where $\varepsilon_j = \sigma_j \sigma_{j+1} - |\varphi_j|$.

The assumption (6.18) amounts to taking Γ as the set of all the $n = 2k-1$ pairs (i, i) and $(j, j+1)$. We claim that the point $\sigma \in R^{2k-1}$ as defined by (6.12), $(\sigma_{j,j+1} = \varphi_j)$, is an interior point of V. In view of (6.17), we have to show that

$$\nu(h) = \sum_{i=1}^{k} a_i \sigma_i^2 + 2 \sum_{j=1}^{k-1} b_j \varphi_j + c \tag{6.20}$$

is *strictly* positive whenever the function

$$h(x) = \sum_{i=1}^{k} a_i x_i^2 + 2 \sum_{j=1}^{k-1} b_j x_j x_{j+1} + c \tag{6.21}$$

is nonnegative for all $x = (x_1, \dots, x_k)$. Here, the a_i, b_j and c denote real constants not all zero. Given such an h, take $x_i = \pm \sigma_i$ with the $\pm$-signs chosen in such a way that

$$b_j x_j x_{j+1} = -|b_j| \sigma_j \sigma_{j+1} = -|b_j| (|\varphi_j| + \varepsilon_j) \leqslant b_j \varphi_j - |b_j| \varepsilon_j.$$

Using $h(x) \geqslant 0$, this yields

$$\nu(h) \geqslant 2 \sum_{j=1}^{k-1} \varepsilon_j |b_j| \geqslant 0. \tag{6.22}$$

By (6.19), (6.20) and (6.22), we conclude that $\nu(h) = 0$ can only happen when all the a_i, b_j and c are equal to zero.

We now have from (6.16) that, for *any* Borel measurable function $g_0(x)$ on R^k, the *best* upperbound $\mu_{\max}(g_0|\sigma)$ on $E(g_0(X_1, \dots, X_k))$, given (6.18), is equal to

$$q(g_0|\sigma) = \inf \{\nu(h) : h \geqslant g_0\}. \tag{6.23}$$

Here, h ranges over all the functions of the special form (6.21); further, $\nu(h)$ is defined by (6.20).

As an illustration, let us take $g_0(x)=0$ if $x \in Q$, $g_0(x)=1$ if $x \notin Q$, where Q denotes the cube

$$Q=\{x=(x_1, \ldots, x_k) : |x_i|<1 \text{ for all } i=1, \ldots, k\}.$$

Let us consider the (sharp) bound $q(g_0|\sigma)$ in

$$\text{(6.24)} \qquad \Pr(Z \notin Q)=E(g_0(Z)) \leqslant q(g_0|\sigma).$$

This bound is given by (6.23). In particular (taking $h \equiv 1$), we have $q(g_0|\sigma) \leqslant 1$; thus, in (6.23) we need only to consider functions h of the form (6.21) with $h \geqslant g_0$ and $\nu(h)<1$. Then the minimum value c of the function h satisfies $0 \leqslant c<1$, thus, $h^*=(h-c)/(1-c)$ satisfies $h^* \geqslant g_0$ and $\nu(h^*)<\nu(h)$ if $c \neq 0$. Hence, we may as well assume that $c=0$.

Given

$$w=(a_1, \ldots, a_k; b_1, \ldots, b_{k-1}) \in R^{2k-1},$$

let us introduce

$$\text{(6.25)} \qquad h_w(x)=\sum_{i=1}^{k} a_i x_i^2+2 \sum_{j=1}^{k-1} b_j x_j x_{j+1}, \qquad (x \in R^k).$$

and

$$\text{(6.26)} \qquad \psi(w)=\nu(h_w)=\sum_{i=1}^{k} \sigma_{ii} a_i+2 \sum_{j=1}^{k-1} \varphi_j b_j.$$

Then, by (6.23), the required bound $q(g_0|\sigma)$ may be written as

$$\text{(6.27)} \qquad q(g_0|\sigma)=\inf\{\psi(w) : w \in K\}.$$

Here, K will denote the *closed* and *convex* subset of R^{2k-1} consisting of all points w such that

$$\text{(6.28)} \qquad \begin{cases} h_w(x) \geqslant 0 \text{ for all } x, \\ \qquad\ \geqslant 1 \text{ for all } x \notin Q; \end{cases}$$

(in the present case where Q is a cube, K is clearly non-empty; for more general sets Q: if K is empty then $q(g_0|\sigma)=1$).

By (6.19) and (6.22), we have that $\psi(w) \rightarrow \infty$ if $|w| \rightarrow \infty$, $w \in K$. Observing that $\psi(w)$ is a linear functional, we conclude that the minimum value (6.27) is in fact taken at an extreme point of K. Therefore,

$$\text{(6.29)} \qquad q(g_0|\sigma)=\inf\{\psi(w) : w \in K_E\},$$

whenever K_E is a subset of K containing all extreme points of K. In other words, we are allowed to impose additional restrictions on h_w provided that these are automatically satisfied when w is an extreme point of K.

Let us replace (6.28) by an equivalent condition not involving x. Let $w \in R^{2k-1}$ be given and consider the $r \times r$ matrix B_r defined by

$$\begin{aligned} (B_r)_{ij} &= a_i \text{ if } j=i, \\ &= b_i \text{ if } j=i+1, \\ &= b_j \text{ if } j=i-1, \\ &= 0, \text{ otherwise,} \end{aligned}$$

$(i, j=1, \ldots, r)$; thus, B_r is symmetric $(r=1, \ldots, k)$. Also note that $B_k=A$ (say) is the matrix of the quadratic form $h_w(x)$ defined by (6.25).

If $w \in K$ then, by (6.28), the matrix A (that is, $h_w(x)$) is strictly positive definite, hence, all principal minors of A have a positive determinant. Moreover, as is well-known, a sufficient condition for A to be strictly positive definite is that

$$\text{(6.30)} \qquad \det(B_r)>0, \qquad r=1, \ldots, k;$$

(here, $D_r=\det(B_r)$ satisfies $D_{r+1}=a_{r+1}D_r-b_r^2D_{r-1}$, thus, (6.30) is easily verified).

As was shown by Olkin and Pratt [36] p. 229, (see also [51] p. 235), given (6.30), the second condition (6.28) is equivalent to

$$\text{(6.31)} \qquad \det(A_{ii})\leqslant\det(A), \qquad i=1, \ldots, k.$$

Here, A_{ii} denotes the (principal) minor corresponding to the diagonal element $a_{ii}=a_i$ of $A=B_k$. Thus, the set

$$K \subset R^{2k-1}$$

occurring in (6.27) and (6.29), may be defined by (6.30) and (6.31), (in stead of (6.28)).

For convenience, let us restrict ourselves to the special case $k=3$; (even for this case the results of Birnbaum and Marshall [6] are incomplete). The reasoning in [36] p. 230 shows that for both $i=1$ and $i=k$ the equality sign holds in (6.31) as soon as w is an *extreme* point of K; by (6.29), this is an admissable restriction. The resulting two equations can easily be solved in terms of b_1^2 and b_2^2, yielding

$$b_1^2=(a_1-1)a_1a_2/(a_1+a_3-1),$$
$$b_2^2=(a_3-1)a_2a_3/(a_1+a_3-1),$$

(where $a_i\geqslant 1$). Afterwards, the case $i=2$ of (6.31) yields the condition

$$a_1+a_3-1\leqslant a_2.$$

Letting

$$2a_1=1+\xi_1 \;, \quad 2a_2=(\xi_1+\xi_3)\xi_2^2 \;, \quad 2a_3=1+\xi_3,$$

the latter condition is equivalent to $\xi_2\geqslant 1$. In fact, the only restrictions on the ξ_i are

$$\xi_1\geqslant 1 \;, \quad \xi_2\geqslant 1 \;, \quad \xi_3\geqslant 1.$$

Further, b_1 and b_2 are given by

$$b_1^2=\tfrac{1}{4}(\xi_1^2-1)\,\xi_2^2 \;, \quad b_2^2=\tfrac{1}{4}(\xi_3^2-1)\,\xi_2^2.$$

It remains to minimize the function of the ξ_i which results on substituting these expressions into (6.26). More precisely, using (6.29), we obtain that

$$\text{(6.32)} \qquad 2q(g_0|\sigma)=\inf\{f(\xi):\xi_i\geqslant 1\},$$

where

$$f(\xi) = \sigma_{11}(1+\xi_1) + \sigma_{22}(\xi_1+\xi_3)\ \xi_2^2 + \sigma_{33}(1+\xi_3)$$
$$-2|\varphi_1|\ \xi_2 \sqrt{\xi_1^2-1}\ -2|\varphi_2|\ \xi_2 \sqrt{\xi_3^2-1}.$$

Taking $\xi_2 = 1$, and choosing afterwards ξ_1 and ξ_3 in the best possible way, one arrives at the inequality

$$2q(g_0|\sigma) \leqslant \sigma_{11} + \sigma_{33} + \sqrt{d_1} + \sqrt{d_2}, \tag{6.33}$$

where

$$d_1 = (\sigma_{11}+\sigma_{22})^2 - 4\varphi_1^2\ ,\quad d_2 = (\sigma_{22}+\sigma_{33})^2 - 4\varphi_2^2.$$

The upperbound (6.33) is contained in a more general result (k arbitrary) due to BIRNBAUM and MARSHALL [6] p. 693.

Note that $f(\xi)$ is a quadratic function of ξ_2. Hence, given ξ_1 and ξ_3, the choice $\xi_2 = 1$ is best possible if and only if the minimum

$$[|\varphi_1| \sqrt{\xi_1^2-1} + |\varphi_2| \sqrt{\xi_3^2-1}]/[\sigma_{22}(\xi_1+\xi_3)] = \varrho(\xi_1, \xi_3), \tag{6.34}$$

(say), does not exceed 1. Hence, (6.33) certainly holds with the *equality* sign when $\max(|\varphi_1|, |\varphi_2|) \leqslant \sigma_{22}$, hence, also when $\max(\sigma_{11}, \sigma_{33}) \leqslant \sigma_{22}$.

For the final analysis, let us restrict ourselves to the special case

$$\sigma_{11} = \sigma_{33},\quad |\varphi_1| = |\varphi_2|. \tag{6.35}$$

Then one may as well take $\xi_1 = \xi_3$ in (6.32). The strict inequality sign in (6.33) can arise only when the case $\xi_2 > 1$ is of any importance. By (6.34), the latter happens when $\xi_1 \geqslant 1$ satisfies

$$\varrho(\xi_1) = (|\varphi_1|/\sigma_{22}) \sqrt{1-\xi_1^{-2}} > 1,$$

(which requires that $|\varphi_1| > \sigma_{22}$). Given such a value ξ_1 and choosing ξ_2 in an optimal fashion, (namely, $\xi_2 = \varrho(\xi_1)$), we obtain

$$\inf_{\xi_1 \geqslant 1,\, \varrho(\xi_1) \geqslant 1} [\sigma_{11}\sigma_{22} + (\sigma_{11}\sigma_{22} - \varphi_1^2)\ \xi_1 + \varphi_1^2 \xi_1^{-1}]/\sigma_{22}$$

as a further contribution to (6.32). If here the minimum is taken at a point with $\varrho(\xi_1) = 1$ then (6.33) holds with the equality sign. If not then, as is easily seen, we have $\delta > 0$, where

$$\delta = 2(\varphi_1^2/\sigma_{22}) - \sigma_{11} - \sigma_{22}.$$

Moreover, in this case, (6.32) yields

$$q(g_0|\sigma) = \sigma_{11} + 2(|\varphi_1|/\sigma_{22}) \sqrt{\sigma_{11}\sigma_{22} - \varphi_1^2}$$
$$= \sigma_{11} + \sqrt{d_1 - \delta^2} < \sigma_{11} + \sqrt{d_1}.$$

Consequently, assuming (6.35), we have that (6.33) holds with the equality sign if and only if $\delta \leqslant 0$.

(III)″ Returning to the more general situation (6.13), let us assume that the point σ defined by (6.12) is an interior point of V. Consider a fixed non-empty Borel subset B of R^k, $B \neq R^k$, and let χ_B denote its characteristic function. It follows from (6.16) that for any k-dimensional random variable Z satisfying (6.13) we have

$$\Pr(Z \in B) \leqslant q(\chi_B|\sigma), \tag{6.36}$$

and further that this bound cannot be improved. The main problem remaining is to determine (if possible) a more explicit expression for the quantity $q(\chi_B|\sigma)$ defined by (6.14) and (6.15).

Following MARSHALL and OLKIN [31] p. 1003, let us consider the special case that B is *convex*, $(B \neq R^k)$. We assert that in this case

$$q(\chi_B|\sigma) = \inf\{q(\chi_A|\sigma) : A \in \mathscr{A},\, A \supset B\}. \tag{6.37}$$

Here, $\mathscr{A}$ will denote the collection of all closed *half-spaces* A of the form

$$A = \{x \in R^k : a_0 + \sum_{i=1}^{k} a_i x_i \geqslant 0\}, \tag{6.38}$$

with the a_i as real constants, $a_i \neq 0$ for at least one index $i \neq 0$.

In proving (6.37), consider a quadratic function h of the form

$$h(x) = \gamma_{00} + \sum_{\Gamma} \gamma_{ij} x_i x_j, \tag{6.39}$$

$(x \in R^k)$, and satisfying $h \geqslant \chi_B$. In view of (6.14), it suffices to show that there exists a closed half-space A such that $h \geqslant \chi_A$ and $A \supset B$, (that is, $h \geqslant \chi_A \geqslant \chi_B$).

We have $h(x) \geqslant 0$ for all x. Thus, $h(x)$ is a nonnegative definite quadratic form, hence,

$$C = \{x : h(x) < 1\}$$

is a convex subset of R^k; also note that C is open. We may assume that C is non-empty, (for, otherwise, $h \geqslant 1 \geqslant \chi_A$ for any set A and we would be ready).

If $x \in B$ then $1 = \chi_B(x) \leqslant h(x)$, thus, $x \notin C$. It follows that B and C are disjoint non-empty convex subsets of R^k, consequently, [48] p. 25, there exists a half-space A containing B and disjoint from C. Since C is open, we may assume that A is closed. Finally, if $x \in A$ then $x \notin C$, thus, $h(x) \geqslant 1 = \chi_A(x)$, hence, A has all the required properties.

Next, let A be a half-space as in (6.38). We assert that

$$q(\chi_A|\sigma) = \inf_{\beta \geqslant 0} q(\varphi_\beta|\sigma), \tag{6.40}$$

where

$$\varphi_\beta(x) = [1 + \beta(a_0 + \sum_{i=1}^{k} a_i x_i)]^2, \qquad (\beta \geqslant 0).$$

Clearly, $\varphi_\beta \geqslant \chi_A$, hence, $q(\chi_A|\sigma)$ does not exceed the right hand side of (6.40). To prove the converse inequality, we may assume that $q(\chi_A|\sigma) < 1 =$

$=q(\varphi_0|\sigma)$. By (6.14), there exists a non-constant function h of the form (6.39) such that $h \geqslant \chi_A$. In computing $q(\chi_A|\sigma)$, one may as well assume that the smallest value of h is equal to zero, (compare the remark following (6.24)), $h(x^*)=0$, say. But then $x^* \notin A$, thus, ·

$$\beta=[-a_0-\sum_{i=1}^{k} a_i x_i^*]^{-1}>0, \text{ while } h(x) \geqslant \varphi_\beta(x),$$

for all x; (the latter inequality is obvious at the boundary of A; moreover, both h and φ_β are homogeneous functions of the $y_i = x_i - x_i^*$). This proves (6.40).

Recall (see (6.4) and (6.7)) that

$$q(\varphi_\beta|\sigma)=\sup E[1+\beta(a_0+\sum_{i=1}^{k} a_i X_i)]^2,$$

the supremum being taken over all k-tuples $(X_1, \ldots, X_k)$ satisfying (6.13). Hence, in the particular case that *all* moments σ_{ij} are given, (6.40) yields

$$(6.41) \qquad \begin{cases} q(\chi_A|\sigma)=1-\alpha^2/\sum_{i=0}^{k}\sum_{j=0}^{k} \sigma_{ij} a_i a_j & \text{if } \alpha<0, \\ \qquad =1 & \text{if } \alpha \geqslant 0, \end{cases}$$

where $\alpha=\sum_{i=0}^{k} a_i \sigma_{i0}$. For this same special case, the sharp upperbound (6.36), with $q(\chi_B|\sigma)$ defined by (6.37) and (6.41), is due to MARSHALL and OLKIN [31] p. 1003; they also gave a large number of specific applications.

7. *An alternative approach*

Many of the problems in section 6 can equally well or better be handled by means of the results in section 4, in particular Theorem 4.4. Suppose, for instance, that we are interested in a k-dimensional random variable $Z=(X_1, \ldots, X_k)$ satisfying

$$(7.1) \qquad E(X_i)=0$$

and

$$(7.2) \qquad E(X_i X_j)=\sigma_{ij},$$

$(i, j=1, \ldots, k)$. Here, the σ_{ij} denote given real numbers, such that the matrix

$$\Sigma=(\sigma_{ij}; i, j=1, \ldots, k)$$

is symmetric and nonnegative definite. Unless in section 6, we shall *allow Σ to be singular*. On occasion, it will be convenient to replace (7.2) by the weaker condition that

$$(7.3) \qquad E(\sum_{i=1}^{k} a_i X_i)^2 \leqslant \sum_{i=1}^{k}\sum_{j=1}^{k} \sigma_{ij} a_i a_j,$$

for each choice of the real numbers $a_1, a_2, \ldots, a_k$.

If $C=(c_{ij})$ is a square matrix then by $C \gg 0$ we shall denote that C is nonnegative definite; further $C_1 \ll C_2$ will denote that $C_2-C_1 \gg 0$. Thus, condition (7.3) may be written as

$$\Sigma' \ll \Sigma, \text{ where } \sigma'_{ij}=E(X_i X_j).$$

For f as a function on R^k, put

$$q^*(f)=\inf \{\alpha_0+\sum_{i,j=1}^{k} c_{ij}\sigma_{ij}\}.$$

Here, the real numbers α_0 and c_{ij} are subject to the conditions:
(i) $C \gg 0$, where $C=(c_{ij};\, i, j=1, \ldots, k)$.
(ii) For some choice of the real constants $\alpha_1, \ldots, \alpha_k$, we have

$$\sum_{i=0}^{k} \alpha_i x_i+\sum_{i,j=1}^{k} c_{ij} x_i x_j \geqslant f(x), \qquad \text{for all } x \in R^k.$$

Note that (ii) implies (i) as soon as

$$\liminf_{|x|\to\infty} f(x)/|x|^2 \geqslant 0.$$

Hence, in this case, $q^*(f)$ coincides with the quantity $q(f|\sigma)$ defined by (6.14) if there we take Γ as the set of all admissable pairs (i, j) and $\sigma_{0,j}=0$ $(j=1, \ldots, k)$. Let us finally introduce

$$Q^*(f)=\sup \{q^*(g) : g \leqslant f,\, g \text{ is u.s.c.}\};$$

in particular, $Q^*(f)=q^*(f)$ if f itself is upper semi-continuous.

Theorem 7.1. Let $f(x)$ be a given Borel measurable function on R^k such that

(7.4) $$\lim_{|x|\to\infty} f(x)/|x|^2=0.$$

Then

(7.5) $$E(f(X_1, \ldots, X_k)) \leqslant Q^*(f)$$

for each set of random variables $X_1, \ldots, X_k$ satisfying (7.1) and (7.3). This inequality is *sharp*, that is, in (7.5) one cannot replace $Q^*(f)$ by any smaller constant.

Finally, if f itself is upper semi-continuous (say, $f=\chi_B$ with B as any closed subset of R^k) then the equality sign in (7.5) is *assumed* by some $Z=(X_1, \ldots, X_k)$ satisfying (7.1) and (7.3).

Proof. Apply Theorem 4.4 with $X=R^k$, $\{f_i, i \in D_0\}$ as the set of continuous functions $\{f_i(x)=x_i, i=0, 1, \ldots, k\}$ $(x_0=1)$ and with $\{h_j, j \in I\}$ as the collection of *all* nonnegative definite quadratic functions $h_p= \sum c_{ij}x_i x_j$, while $\eta_p=\sum c_{ij}\sigma_{ij}$, thus, $\eta_p \geqslant 0$; (an equivalent choice would be $h_p=(\sum a_i x_i)^2$ and $\eta_p=\sum \sigma_{ij} a_i a_j$).

The last assertion of Theorem 7.1 is no longer valid if (7.3) is replaced by (7.2). For example, if $k=1$ and $f(x)=e^{-x^2}$ then $Q^*(f)=1$ whatever the value $\sigma_{11}\geqslant 0$. But there does not exist any real random variable X satisfying $E(X)=0$, $E(X^2)=\sigma_{11}>0$, $E(e^{-X^2})=1$.

Theorem 7.2. Let f be a Borel measurable function on R^k satisfying (7.4). Then the inequality (7.5) is still *sharp* when $Z=(X_1, \ldots, X_k)$ is assumed to satisfy the stronger conditions (7.1) and (7.2).

Proof. Let $\varepsilon>0$ be a given number. We must prove that there exists a probability measure μ on R^k satisfying

$$\mu(x_i)=0 \ , \ \mu(x_i x_j)=\sigma_{ij}, \tag{7.6}$$

$(i, j=1, \ldots, k)$, and

$$\mu(f)>Q^*(f)-\varepsilon.$$

By Theorem 7.1, there exists a probability measure ν on R^k satisfying

$$\nu(x_i)=0 \ , \ \nu(x_i x_j)=\sigma'_{ij},$$

$(i, j=1, \ldots, k)$, and $\nu(f)>Q^*(f)-\varepsilon/2$. Here, (by (7.3) and the remark following it), the matrix $\Sigma'=(\sigma'_{ij}; i, j=1, \ldots, k)$ satisfies $0\leqslant\Sigma'\leqslant\Sigma$. Hence, there exist real numbers α_{pi} $(p=1, \ldots, m; i=1, \ldots, k)$ such that

$$\sigma_{ij}-\sigma'_{ij}=\sum_{p=1}^{m}\alpha_{pi}\alpha_{pj}, \qquad (i, j=1, \ldots, k);$$

we may assume that $m\geqslant 0$ is minimal. Similarly, there exist real numbers β_{pi} such that σ'_{ij} can be written as

$$\sigma_{ij}'=\sum_{p=1}^{n}\beta_{pi}\beta_{pj}, \qquad (i, j=1, \ldots, k);$$

we may assume that $n\geqslant 0$ is minimal. With λ as a positive real number, let Δ_λ denote the nonnegative measure on R^k of total mass m having a mass $1/2$ at each of the $2m$ points

$$\pm\ (\lambda\alpha_{p1}, \lambda\alpha_{p2}, \ldots, \lambda\alpha_{pk}), \qquad (p=1, \ldots, m);$$

(these $2m$ points are distinct since m is minimal). It satisfies

$$\Delta_\lambda(x_i)=0, \ \ \Delta_\lambda(x_i x_j)=\lambda^2(\sigma_{ij}-\sigma'_{ij}),$$

$(i, j=1, \ldots, k)$. Let further ν_λ denote the measure of mass n having a mass equal to $1/2$ at each of the $2n$ points $\pm\ (\lambda\beta_{p1}, \ldots, \lambda\beta_{pk})$, $p=1, \ldots, n$. It satisfies

$$\nu_\lambda(x_i)=0, \ \ \nu_\lambda(x_i x_j)=\lambda^2\sigma'_{ij},$$

$(i, j=1, \ldots, k)$. Moreover, by (7.4), one has for $\lambda>0$ sufficiently large that

$$|\Delta_\lambda(f)|<(\varepsilon/8)\,\lambda^2 \ , \ |\nu_\lambda(f)|<(\varepsilon/8)\,\lambda^2. \tag{7.7}$$

Now, let us form (with $\lambda > n^{1/2}$ fixed) the measure

$$\mu = (1-\delta)\,\nu + \delta_1\,\Delta_\lambda + \delta_2\,\nu_\lambda,$$

where

$$\delta_1 = \lambda^{-2}, \quad \delta_2 = m\lambda^{-2}/(\lambda^2 - n), \quad \delta = m\delta_1 + n\delta_2$$

depend on λ; note that $\delta = \delta_2\lambda^2$. It follows from the above relations that μ is a probability measure satisfying (7.6). Moreover, using (7.7) and $\nu(f) > Q^*(f) - \varepsilon/2$, we have

$$\mu(f) \geqslant (1-\delta)\,Q^*(f) - \varepsilon/2 - (\varepsilon/8)\,(\delta_1\lambda^2 + \delta_2\lambda^2) > Q^*(f) - \varepsilon,$$

as soon as $\lambda > 0$ is sufficiently large.

(To be continued)