The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 93–122

THE GENERAL MOMENT PROBLEM, A GEOMETRIC APPROACH[1]

BY J. H. B. KEMPERMAN

University of Rochester

0. Summary. Let $g_1, \cdots, g_n$ and h be given real-valued Borel measurable functions on a fixed measurable space $T = (T, \mathfrak{a})$. We shall be interested in methods for determining the best upper and lower bound on the integral

$$\mu(h) = \int_T h(t)\mu(dt),$$

given that μ is a probability measure on T with known moments $\mu(g_j) = y_j$, $j = 1, \cdots, n$.

More precisely, denote by $\mathfrak{M}^+ = \mathfrak{M}^+(T)$ the collection of all probability measures on T such that $\mu(|g_j|) < \infty$ $(j = 1, \cdots, n)$ and $\mu(|h|) < \infty$. For each $y = (y_1, \cdots, y_n) \varepsilon R^n$, consider the bounds

$$L(y) = L(y \mid h) = \inf \mu(h), \qquad U(y) = U(y \mid h) = \sup \mu(h),$$

where μ is restricted by

$$\mu \varepsilon \mathfrak{M}^+(T); \qquad \mu(g_1) = y_1, \cdots, \mu(g_n) = y_n.$$

If there is no such measure μ we put $L(y) = +\infty$, $U(y) = -\infty$. In many applications, h is the characteristic function (indicator function) $h = I_S$ of a given measurable subset S of T. In that case we usually write instead $L(y \mid I_S) = L_S(y)$, $U(y \mid I_S) = U_S(y)$. Thus, $L_S(y) \leqq \mu(S) \leqq U_S(y)$ are the best possible bounds on the probability mass $\mu(S)$ contained in S, given that $\mu \varepsilon \mathfrak{M}^+$ and that $\mu(g) = y$. Here, g denotes the mapping $g: T \to R^n$ defined by $g(t) = (g_1(t), \cdots, g_n(t))$. By g_0 we shall denote the function on T with $g_0(t) = 1$ for all $t \varepsilon T$.

The following tentative method for finding $L(y \mid h)$ may be said to go back to Markov [8] and Riesz [13], see [7]. Choose an $(n + 1)$-tuple $d^* = (d_0, d_1, \cdots, d_n)$ of real numbers such that

$$d_0 + d_1 g_1(t) + \cdots + d_n g_n(t) \leqq h(t) \quad \text{for all} \quad t \varepsilon T,$$

and define

$$B(d^*) = \{z \varepsilon R^n: z = g(t) \text{ for some } t \varepsilon T \text{ with } \textstyle\sum_{j=0}^n d_j g_j(t) = h(t)\}.$$

Then

$$L(y \mid h) = d_0 + \textstyle\sum_{j=1}^n d_j y_j \quad \text{for each} \quad y \varepsilon \operatorname{conv} B(d^*),$$

(conv = convex hull).

Received 13 April 1967.
[1] This research was supported in part by the National Science Foundation, NSF-GP-5801.

The main purpose of the present paper is to investigate the merits of this method and certain more general methods.

It turns out (Theorem 5) that for almost all $y \in R^n$ there exists *at most one* admissible d^* with $y \varepsilon$ conv $B(d^*)$. Moreover, provided $y \varepsilon \operatorname{int}(V)$ where $V = \operatorname{conv} g(T)$, there exists at least one such d^* if and only if there exists a measure $\mu \varepsilon \mathfrak{M}^+$ with $\mu(g) = y$ and $\mu(h) = L(y \mid h)$.

A sufficient condition for the latter would be that T has a compact topology with respect to which g is continuous and h is lower semi-continuous.

More interesting is a related method for finding $L(y \mid h)$, see Theorem 6, which will work for *each* $y \varepsilon \operatorname{int}(V)$ as soon as g is bounded.

The situation where $y \not\varepsilon \operatorname{int}(V)$ is discussed in Section 4. It appears that the assumption $y \varepsilon \operatorname{int}(V)$ is a rather natural one.

We have chosen to develop the important special case $h = I_S$ in a partly independent manner, see the Sections 5, 6 and 7. In this case, the $(n + 1)$-tuple d^* must satisfy

$$\begin{aligned} d_0 + \textstyle\sum_{j=1}^n d_j z_j &\leqq 1 \quad \text{for all} \quad z \varepsilon g(T), \\ &\leqq 0 \quad \text{for all} \quad z \varepsilon g(S'). \end{aligned}$$

Here, S' denotes the complement of S in T. Assuming that $d_1, \cdots, d_n$ are not all zero, let us associate to d^* the pair of hyperplanes H and H' with equations

$$\textstyle\sum_{j=1}^n d_j z_j = 1 - d_0 \quad \text{and} \quad \sum_{j=1}^n d_j z_j = -d_0,$$

respectively. This pair is such that H, H' are distinct parallel hyperplanes with $g(S')$ and H on opposite sides of H' and $g(T)$ and H' on the same side of H; such a pair H, H' will be said to be *admissible*. Observe that

$$B(d^*) = (g(S) \cap H) \cup (g(S') \cap H'),$$

with H, H' as the admissible pair determined by d^*.

The present $(n + 1)$-tuple d^* is useful, for determining $L_S(y) = L(y \mid I_S)$ for at least some points y, only when both $g(S) \cap H \neq 0$ and $g(S') \cap H' \neq 0$. That is, H' should not only support the set $g(S')$ but even "intersect" it; similarly, H and $g(S)$. Fortunately, one can usually replace "intersect" by "touch". More precisely (Corollary 13), if H and H' form an admissible pair as above then $L_S(y) = d_0 + \sum_{j=1}^n d_j y_j$ for each point y such that both

$$y \varepsilon \operatorname{int}(V), \qquad y \varepsilon \operatorname{conv}[\{H \cap \overline{\operatorname{conv}}\, g(S)\} \cup \{H' \cap \overline{\operatorname{conv}}\, g(S')\}],$$

a bar denoting closure.

Provided g is bounded the latter generalization will yield the value $L_S(y)$ for all relevant y, see Theorem 7. Whether or not g is bounded, we have for almost all y that there can be at most one admissible pair of hyperplanes H and H' yielding $L_S(y)$ in the above manner.

A detailed discussion of the method on hand may be found in Section 6. The present method is geometrical in the following sense: (i) one only needs to know the *sets* $g(S)$ and $g(S')$ in R^n; (ii) afterwards, one considers all the pairs H and H' of parallel hyperplanes touching $g(S)$ and $g(S')$ in the above manner. Each such pair yields $L_S(y)$ for certain values y; varying the pair H, H' one often obtains the value $L_S(y)$ for all relevant $y \varepsilon R^n$.

Usually, there are many different regions in y-space, each with its own analytic formula for $L_S(y)$. Nevertheless, all these different formulae are derived from one and the same geometrical principle.

A number of specific illustrations, all with $n = 2$, are presented in Section 7. They indicate that it is often quite easy to solve the following problem in a geometric manner. Let X be a random variable taking its values in a measurable space T, such that

$$E(g_1(X)) = y_1, \qquad E(g_2(X)) = y_2,$$

with g_1 and g_2 as known real-valued Borel measurable functions on T. The problem is to determine the best possible lower bound $L_S(y)$ on $\Pr(X \varepsilon S)$ where S is a given Borel measurable subset of T.

1. Measures of finite support. Let $T = (T, \mathcal{A})$ be a given measurable space. By a measure on T of finite support we mean a measure of the special form

$$\mu(A) = \textstyle\sum_{t_i \varepsilon A} p_i, \tag{1.1}$$

thus,

$$\mu(h) = \textstyle\sum_{i=1}^{m} p_i h(t_i). \tag{1.2}$$

Here, A is an arbitrary subset of T, h an arbitrary function on T. Further, $\{t_1, \cdots, t_m\}$ is a finite subset of T, while the p_i are real numbers (depending on μ). Note that μ is a measure on the σ-field of *all* subsets of T.

For each subset S of T, let $M^+(S)$ denote the collection of all *probability* measures on T whose support is finite and contained in S. The following Theorem 1 implies that the quantities $L(y \mid h)$ and $U(y \mid h)$ may also be defined as

$$L(y \mid h) = \inf\{\mu(h): \mu \varepsilon M^+(T), \mu(g) = y\} \tag{1.3}$$

and

$$U(y \mid h) = \sup\{\mu(h): \mu \varepsilon M^+(T), \mu(g) = y\}. \tag{1.4}$$

It was found independently by Richter [12], p. 151, and Rogosinsky [14], p. 4, see also Mulholland and Rogers [10]. The proof proceeds by a straightforward induction with respect to N.

THEOREM 1. *Let $f_1, \cdots, f_N$ be given real-valued Borel measurable functions on a measurable space Ω, (such as $g_1, \cdots, g_n$ and h on T). Let μ be a probability measure on Ω such that each f_i is integrable with respect to μ. Then there exists a probability measure μ' of finite support on Ω satisfying*

$$\mu'(f_j) = \mu(f_j) \quad \text{for all} \quad j = 1, \cdots, N.$$

One can even attain that the support of μ' has at most $N + 1$ points.

2. Restating the problem. From now on, unless otherwise stated, when we speak of a measure we shall always mean one having a *finite* support. Therefore the σ-field $\mathfrak{A}$ associated with T may as well be taken as the collection of all subsets of T to the effect that every function on T is measurable.

Our problem may now be restated as follows. Let T be a given non-empty set and let $g:T \rightarrow R^n$ and $h:T \rightarrow R$ be given (arbitrary) functions on T,

$$g(t) = (g_1(t), \cdots, g_n(t)). \tag{2.1}$$

We shall be interested in the quantities $L(y \mid h)$ and $U(y \mid h)$ defined by (1.3) and (1.4), that is, in the *best* upper and lower bounds

$$L(y \mid h) \leqq \int_T h(t)\mu(dt) \leqq U(y \mid h), \tag{2.2}$$

given that μ is a probability measure on T satisfying $\mu(g) = y$.

Sometimes one is interested in several functions $h_j:T \rightarrow R$ at the same time. Consider the corresponding function $\phi = (h_1, \cdots, h_k)$ taking values in R^k and further the *convex* set $V(y)$ of all possible values $\mu(\phi)$ when μ is restricted by $\mu \varepsilon M^+(T)$ and $\mu(g) = y$. Then the closure of $V(y)$ is completely determined by all the hyperplanes supporting $V(y)$ as well as possible. These in turn may be read off from the different values $L(y \mid h)$ where h runs through all linear combinations $h = a_1h_1 + \cdots + a_kh_k$; see [6], p. 573. From now on, h will again be a fixed real-valued function on T.

An important role will be played by the set

$$V = \text{conv}\, g(T). \tag{2.3}$$

It may be helpful to think of the range

$$g(T) = \{z \varepsilon R^n : z = g(t) \quad \text{for some} \quad t \varepsilon T\}$$

of g as a curve in n-space (if T is a one-dimensional interval) or a two-dimensional surface in n-space (if T is a square). The following lemma is rather obvious, see (1.2).

LEMMA 2. *Given $y \varepsilon R^n$ the condition $y \varepsilon V$ is necessary and sufficient in order that there exists a measure $\mu \varepsilon M^+(T)$ having a moment $\mu(g) = y$.*

Hence, by (1.3),

$$L(y \mid h) < +\infty \quad \text{if and only if} \quad y \varepsilon V. \tag{2.4}$$

Obviously, $L(y) = L(y \mid h)$ is a convex function on V, in the sense that

$$L(\lambda y' + (1 - \lambda)y'') \leqq \lambda L(y') + (1 - \lambda)L(y'') \tag{2.5}$$

whenever $0 < \lambda < 1$ and $y' \varepsilon V$, $y'' \varepsilon V$.

By a flat in R^n we shall mean a translate of some linear subspace of R^n. Let F denote the minimal flat containing V. If the system of functions $\{g_0, g_1, \cdots, g_n\}$ on T has rank k then F is of dimension $k - 1$ $(1 \leqq k \leqq n + 1)$.

By intv(V) we shall mean the set of points $y \varepsilon V$ which are interior to V relative to the flat F. Equivalently, $y \varepsilon \text{intv}(V)$ precisely when to each point

$y' \varepsilon V$ with $y' \neq y$ there corresponds at least one point $y'' \varepsilon V$ such that y itself is an (interior) point of the open line segment (y', y''). If $L(y) > -\infty$ for some point $y \varepsilon \operatorname{intv}(V)$ then $L(y) > -\infty$ for all $y \varepsilon V$. In other words, if $L(y) = -\infty$ for some point $y \varepsilon V$ then $L(y) = -\infty$ for all $y \varepsilon \operatorname{intv}(V)$.

It is well-known, [2], p. 19, that the convex function $L(y)$ is automatically continuous throughout intv (V) as soon as it is finite there. Therefore,

$$(2.6) \qquad \lim_{y' \to y, y' \varepsilon V} L(y' \mid h) = L(y \mid h) \quad \text{provided} \quad y \varepsilon \operatorname{intv}(V),$$

(even if $L(y) = -\infty$ throughout intv (V)). Analogous results hold for

$$(2.7) \qquad U(y) = U(y \mid h) = -L(y \mid -h).$$

In particular, $U(y)$ is a concave function on V.

The following three properties are clearly equivalent: (i) The set V as defined by (2.3) has a non-empty interior; (ii) The range $g(T)$ is not the subset of any hyperplane in R^n; (iii) The $n + 1$ functions $g_0, \cdots, g_n$ $(g_0 \equiv 1)$ are linearly independent on T.

Suppose that instead the system $\{g_0, g_1, \cdots, g_n\}$ were of rank $k \leqq n$. Rearranging the indices of $g_1, \cdots, g_n$ we may assume that $g_0, g_1, \cdots, g_{k-1}$ are linearly independent. Define $g^*: T \to R^{k-1}$ by $g^*(t) = (g_1(t), \cdots, g_{k-1}(t))$ and put $V^* = \operatorname{conv} g^*(T)$. Further, for $y = (y_1, \cdots, y_n) \varepsilon R^n$ let $y^* = (y_1, \cdots, y_{k-1})$. It is easily seen that: (i) If $\mu \varepsilon M^+(T)$ then the condition $\mu(g) = y$ is equivalent to the condition $\mu(g^*) = y^*$; (ii) $y \varepsilon V$ if and only if $y^* \varepsilon V^*$; (iii) $y \varepsilon \operatorname{intv}(V)$ if and only if $y^* \varepsilon \operatorname{int}(V^*)$. These remarks show that it would be no real loss of generality to assume that $g_0, g_1, \cdots, g_n$ are linearly independent.

Suppose this is the case so that int (V) is non-empty. Now consider the *convex* subset of R^{n+1} defined by

$$(2.8) \qquad Q = Q_h = \{(y, \gamma): y \varepsilon V, \gamma \geqq L(y \mid h)\},$$

with γ as a real number. Clearly, Q has a non-empty interior relative to R^{n+1}. Assuming that $L(y \mid h) > -\infty$ in V, we have that each point

$$(2.9) \qquad P_y = (y, L(y \mid h)), \qquad (y \varepsilon V),$$

is a boundary point of Q. Hence, [18], p. 27, there passes *through* P_y at least one hyperplane H_y supporting the convex body Q. Provided $y \varepsilon \operatorname{int}(V)$, this hyperplane is always "non-vertical" in the sense that it admits a (unique) equation of the form

$$(2.10) \qquad H = \{(z, \gamma): z \varepsilon R^n, \gamma = d_0 + \textstyle\sum_{j=1}^n d_j z_j\}.$$

We also want to make special notice of the fact that, by a theorem of Reidemeister [11], the set of points $y \varepsilon \operatorname{int}(V)$, such that through P_y there passes *more than one* hyperplane H_y of support to Q, must be of (n-dimensional Lebesgue) *measure zero.*

Using (1.2) and (1.3), the following are easily seen to be equivalent: (i) The hyperplane in R^{n+1} defined by (2.10) is a hyperplane of support to Q; (ii)

We have

$$L(z \mid h) \geqq d_0 + \sum_{j=1}^{n} d_j z_j \quad \text{for all} \quad z \, \varepsilon \, V. \tag{2.11}$$

Thirdly,

$$\mu(h) \geqq d_0 + \sum_{j=1}^{n} d_j \mu(g_j) \quad \text{for all} \quad \mu \, \varepsilon \, M^+(T). \tag{2.12}$$

Finally,

$$h(t) \geqq d_0 + \sum_{j=1}^{n} d_j g_j(t) \quad \text{for all} \quad t \, \varepsilon \, T. \tag{2.13}$$

Let D^* denote the collection of all $(n + 1)$-tuples of real numbers $d^* = (d_0, d_1, \cdots, d_n)$ satisfying (2.13). The above considerations yield the following result, essentially as discovered independently by Richter [12], p. 156, in 1957 and by Isii and Karlin in 1960, see [5], p. 472. A generalization of Theorem 3 may be found in [6], p. 574.

THEOREM 3. *We have for each* $y \, \varepsilon \, \text{int}\,(V)$ *that*

$$L(y \mid h) = \sup \{d_0 + \sum_{j=1}^{n} d_j y_j \colon d^* = (d_0, \cdots, d_n) \, \varepsilon \, D^*\}. \tag{2.14}$$

Provided $L(y \mid h) > -\infty$, *the supremum in* (2.14) *is even assumed by some* $d^* \, \varepsilon \, D^*$. *Finally, if* $L(y \mid h)$ *is finite in* $\text{int}\,(V)$ *then for almost all* $y \, \varepsilon \, \text{int}\,(V)$ *the supremum in* (2.14) *is assumed by a unique* $d^* \, \varepsilon \, D^*$.

As a corollary, we have that $L(y \mid h)$ is finite in $\text{int}\,(V)$ if and only if D^* is non-empty; (if D^* is empty then $L(y \mid h) = -\infty$ for all $y \, \varepsilon \, \text{int}\,(V)$).

In applying Theorem 3 to a given point $y \, \varepsilon \, R^n$ one first needs to know whether $y \, \varepsilon \, \text{int}\,(V)$. As is easily seen and well-known ([4], p. 5 and [6], p. 573), this is the case if and only if $d_0 + \sum_{j=1}^{n} d_j y_j > 0$ for each choice of the real constants d_j not all zero such that $d_0 + \sum_{j=1}^{n} d_j g_j(t) \geqq 0$ for all $t \, \varepsilon \, T$. Here, we are still assuming that $g_0, g_1, \cdots, g_n$ are linearly independent.

Observe that the $(n + 1)$-tuple $d^* \, \varepsilon \, D^*$, where the supremum (2.14) is assumed, has an interest of its own. Namely, it determines a hyperplane through P_y supporting Q and thus would describe the local behavior of $L(y \mid h)$ under small perturbations of y, at least when d^* is unique.

3. Main results. For convenience, we shall assume from now on that the $n + 1$ functions $g_0, g_1, \cdots, g_n$ are linearly independent and further that D^* is non-empty, that is, (2.13) holds for at least one $(n + 1)$-tuple d^*. In the important special case that h is bounded the latter is automatically true.

THEOREM 4. *Let* $d^* \, \varepsilon \, D^*$ *be given and define*

$$B(d^*) = \{z = g(t) \colon d_0 + \sum_{j=1}^{n} d_j g_j(t) = h(t), t \, \varepsilon \, T\}. \tag{3.1}$$

Then for each point

$$y \, \varepsilon \, \text{conv}\, B(d^*) \tag{3.2}$$

the quantity $L(y \mid h)$ *may be obtained as follows. Write*

$$y = \sum_{i=1}^{m} p_i g(t_i) \quad \textit{with} \quad g(t_i) \, \varepsilon \, B(d^*), \tag{3.3}$$

and $p_i \geqq 0$, $\sum p_i = 1$; (*there is at least one such representation of* y). *Then*

$$(3.4) \qquad L(y \mid h) = \textstyle\sum_{i=1}^{m} p_i h(t_i) = d_0 + \sum_{j=1}^{n} d_j y_j .$$

PROOF. Obvious.

THEOREM 5. *Let* $y \varepsilon \operatorname{int}(V)$ *be given. Then the following are equivalent.*

(i) *The infimum* (1.3) *is assumed. That is, there exists a measure* $\mu \varepsilon M^+(T)$ *satisfying* $\mu(g) = y$ *and* $\mu(h) = L(y \mid h)$.

(ii) *There exists* $d^* \varepsilon D^*$ *satisfying* (3.2).

We further assert that for almost all $y \varepsilon \operatorname{int}(V)$ *there exists at most one* $d^* \varepsilon D^*$ *satisfying* (3.2).

REMARK. A sufficient condition for (i) would be that T can be given a Hausdorff topology making T compact, g continuous and h at least semi-continuous; if $h = I_s$ then S must be open; compare [3].

The proof is easy. One considers a sequence $\{\mu_r\}$ in $M^+(T)$ such that $\mu_r(g) = y$ and $\mu_r(h) \to L(y \mid h)$. One may assume (Theorem 1) that μ_r is supported by a set of $n + 2$ points $\{t_{r1}, \cdots, t_{r,n+2}\}$. Denote the corresponding weights by p_{ri}, ($p_{ri} \geqq 0$, $\sum_i p_{ri} = 1$). Now draw a subsequence such that all the sequences $\{t'_{ri}\}$ and $\{p'_{ri}\}$ converge. Here, we are tacitly assuming that the topology for T has a countable base; if not one could draw subnets instead of subsequences.

PROOF. That (ii) implies (i) is clear, compare (3.4). Conversely, let $\mu \varepsilon M^+(T)$ satisfy $\mu(g) = y$ and $\mu(h) = L(y \mid h)$. By Theorem 3, there exists $d^* \varepsilon D^*$ achieving the supremum in (2.14). In view of (2.12) and (2.13), the measure μ must be supported by $B(d^*)$, hence, (3.2) holds.

The uniqueness assertion of Theorem 5 follows by combining Theorem 3 and Theorem 4, compare (3.4).

In many applications, the infimum (1.3) is not assumed so that Theorem 4 is not applicable. The following Theorem 6 has a much wider range of applications. Here, η denotes the function on the closure $\overline{g(T)}$ of $g(T)$, defined by

$$(3.5) \qquad \eta(z) = \underline{\lim}_{\delta \to 0} \inf_t \{h(t) : t \varepsilon T, |g(t) - z| < \delta\}.$$

Observe that, for each fixed $t \varepsilon T$,

$$\eta(z) \leqq h(t) \quad \text{when} \quad z = g(t).$$

Further, η is lower semi-continuous in the sense that

$$\underline{\lim}_{z' \to z} \eta(z') = \eta(z), \qquad (z, z' \varepsilon \overline{g(T)}).$$

Recall that D^* is defined by (2.13). Equivalently, an $(n + 1)$-tuple $d^* = (d_0, \cdots, d_n)$ is in D^* if and only if

$$(3.6) \qquad d_0 + \textstyle\sum_{j=1}^{n} d_j z_j \leqq \eta(z) \quad \text{for all} \quad z \varepsilon \overline{g(T)}.$$

If $\epsilon \geqq 0$ and $d^* \varepsilon D^*$, we define

$$(3.7) \qquad C_\epsilon(d^*) = \{z \varepsilon \overline{g(T)} : 0 \leqq \eta(z) - \textstyle\sum_{j=0}^{n} d_j z_j \leqq \epsilon\}, \qquad (z_0 = 1)$$

and

$$(3.8) \qquad G(d^*) = \textstyle\bigcap_{N=1}^{\infty} \overline{\operatorname{conv}}\, C_{1/N}(d^*).$$

It is easily seen that the sets $C_\epsilon(d^*)$ and $G(d^*)$ are closed and further that $B(d^*) \subset C_0(d^*) \subset C_\epsilon(d^*)$, where $B(d^*)$ is defined by (3.1).

THEOREM 6. *Let $y \varepsilon$ int (V) be given.*

(i) *Let $d^* \varepsilon D^*$ be such that $y \varepsilon G(d^*)$. Then*

$$L(y \mid h) = d_0 + d_1 y_1 + \cdots + d_n y_n . \tag{3.9}$$

(ii) *Suppose that the function g is bounded. Then there always exists at least one $d^* \varepsilon D^*$ satisfying*

$$y \, \varepsilon \operatorname{conv} C_0(d^*) \subset G(d^*),$$

allowing us to obtain $L(y \mid h)$ from (3.9).

(iii) *We further assert that in any case, whether or not g is bounded, we have for almost all $y \varepsilon$ int (V) that there exists at most one $d^* \varepsilon D^*$ satisfying $y \varepsilon G(d^*)$.*

PROOF. Observe that (iii) follows from (i) and Theorem 3. It will be convenient to use the notation

$$d^* y = \textstyle\sum_{j=0}^n d_j y_j = d_0 + \sum_{j=1}^n d_j y_j , \qquad (y \, \varepsilon \, R^n, \, y_0 = 1).$$

Proof of (i). Let $d^* \varepsilon D^*$ and $y \varepsilon$ int (V). By (2.11), we have $L(y \mid h) \geqq d^* y$. It suffices to show that $L(y \mid h) \leqq d^* y + \epsilon$, assuming that

$$y \, \varepsilon \, \overline{\operatorname{conv}} \, C_\epsilon(d^*). \tag{3.10}$$

We first assert that the function

$$L_1(u) = \underline{\lim}_{z \to u} L(z \mid h), \qquad (z \, \varepsilon \, V, \, u \, \varepsilon \, \bar{V}),$$

satisfies

$$L_1(u) \leqq \eta(u) \quad \text{for each} \quad u \, \varepsilon \, \overline{g(T)}. \tag{3.11}$$

For, let $u \varepsilon \overline{g(T)}$ be given. We may assume that $\eta(u) < \infty$. Let $\{c_r\}$ converge decreasingly to $\eta(u)$ and let $t_r \varepsilon T$ be such that $h(t_r) \leqq c_r$ and $|z^r - u| < 1/r$, where $z^r = g(t_r)$, compare (3.5). In particular, $\{z^r\}$ converges to u. Considering the probability measure carried by $\{t_r\}$ we see that $L(z^r \mid h) \leqq h(t_r) \leqq c_r$. This proves (3.11).

By (3.7) and (3.11), we have $L_1(u) \leqq \epsilon + d^* u$ for each $u \varepsilon C_\epsilon(d^*)$. But the function $L_1(\cdot)$ is clearly convex and lower semi-continuous, hence, $L_1(u) \leqq \epsilon + d^* u$ holds throughout the set $\overline{\operatorname{conv}} \, C_\epsilon(d^*)$. Noting that, by (2.6), we have $L(y \mid h) = L_1(y)$ for all $y \varepsilon$ int (V), (3.10) implies the stated assertion.

Proof of (ii). We now assume that g is bounded, hence $\overline{g(T)}$ is compact. Let further $y \varepsilon$ int (V). Then by Theorem 3 there exists $d^* \varepsilon D^*$ with $L(y \mid h) = d^* y$. We will show that $y \varepsilon \operatorname{conv} C_0(d^*)$.

Let $\{\mu_r\}$ be a sequence in $M^+(T)$ such that $\mu_r(g) = y$ and $\mu_r(h) \downarrow L(y \mid h)$. We may assume that μ_r has a support consisting of $n + 2$ points $t_{ri} \varepsilon T$ $(i = 1, \cdots, n + 2)$, see Theorem 1. Let the corresponding weights be denoted by p_{ri}, so that $y = \sum_i p_{ri} g(t_{ri})$ and $\sum_i p_{ri} h(t_{ri}) \downarrow L(y \mid h)$. Drawing a subsequence, we may assume that, as $r \to \infty$, $p_{ri} \to p_i$ and $g(t_{ri}) \to z^i \varepsilon \overline{g(T)}$,

$(i = 1, \cdots, n+2)$. In particular, $y = \sum p_i z^i$ and

$$\sum p_i \eta(z^i) \leqq \sum p_i \varliminf_r h(t_{ri}) = \sum \varliminf_r p_{ri} h(t_{ri}) \leqq \varliminf_r \sum_i p_{ri} h(t_{ri}).$$

Here, the latter right hand side is equal to

$$L(y \mid h) = d^* y = \sum_i p_i (d^* z^i) \leqq \sum_i p_i \eta(z^i),$$

by (3.6). Hence, all the above inequalities are in fact equalities, thus, $z^i \varepsilon C_0(d^*)$ for all i with $p_i > 0$, hence, $y \varepsilon \operatorname{conv} C_0(d^*)$.

From now on, we restrict ourselves to the important special case that $h = I_S$ is the indicator function of a given subset S of T. By S' we shall denote the complement of S in T. We shall assume that both S and S' are non-empty. Consider further the convex sets

$$(3.12) \qquad V_S = \operatorname{conv} g(S), \qquad V_{S'} = \operatorname{conv} g(S'), \qquad V = \operatorname{conv} g(T)$$

and their closures

$$(3.13) \qquad W_S = \overline{\operatorname{conv}}\, g(S), \qquad W_{S'} = \overline{\operatorname{conv}}\, g(S'), \qquad W = \overline{\operatorname{conv}}\, g(T).$$

One may interpret $L(y \mid I_S) = L_S(y)$ as the smallest possible mass in S for a probability distribution μ on T with $\mu(g) = y$. Obviously, $L_S(y) = 0$ if $y \varepsilon V_{S'}$. Hence, by (2.6),

$$(3.14) \qquad L_S(y) = 0 \quad \text{if} \quad y \varepsilon W_{S'}, \qquad y \varepsilon \operatorname{int}(V).$$

Therefore, we only need to consider the case that

$$(3.15) \qquad y \varepsilon \operatorname{int}(V) = \operatorname{int}(W); \qquad y \not\varepsilon W_{S'}.$$

Presently, the function η defined by (3.5) reduces to

$$(3.16) \qquad \begin{aligned} \eta(z) &= 0 \quad \text{if} \quad z \varepsilon \overline{g(S')}, \\ &= 1 \quad \text{if} \quad z \varepsilon \overline{g(S)}, \qquad z \not\varepsilon \overline{g(S')}. \end{aligned}$$

Thus η may be regarded as the indicator function of the interior of $\overline{g(S)}$ taken relative to the space $\overline{g(T)}$. Further, the condition (3.6) for $d^* \varepsilon D^*$ becomes

$$(3.17) \qquad \begin{aligned} d_0 + \textstyle\sum_{j=1}^n d_j z_j &\leqq 1 \quad \text{for all } z \varepsilon \overline{g(T)}, \text{ hence for all } z \varepsilon W, \\ &\leqq 0 \quad \text{for all } \; z \varepsilon \overline{g(S')}, \text{ hence for } z \varepsilon W_{S'}. \end{aligned}$$

Assuming for the moment that $d_1, \cdots, d_n$ are not all zero, let us introduce the distinct parallel hyperplanes in R^n defined by

$$(3.18) \qquad H = H(d^*) = \{z\colon d^*z = 1\}; \qquad H' = H'(d^*) = \{z\colon d^*z = 0\}.$$

Condition (3.17) says that H supports all of W (that is, all of $g(T)$), while H' supports all of $W_{S'}$, on the same side as H supports W (so that H' is in between H and $W_{S'}$). Let us call such a pair of *distinct* hyperplanes H, H' an *admissible* pair. Given such a pair there is a unique $(n+1)$-tuple d^* with $d_1, \cdots, d_n$ not all zero such that H, H' are given by (3.18). Moreover, (3.17) holds, that is, $d^* \varepsilon D^*$.

THEOREM 7. (i) *Be given an admissible pair of hyperplanes* $H = H(d^*)$ *and* $H' = H'(d^*)$. *Put*

$$G[d^*] = \overline{\text{conv}}\,[(H \cap W_S) \cup (H' \cap W_{S'})]. \tag{3.19}$$

Then for each $y \in G[d^*]$ *with* $y \in \text{int}\,(V)$ *we have*

$$L_S(y) = d^*y = \Delta(y)/\Delta. \tag{3.20}$$

Here, $\Delta(y)$ *denotes the distance from* y *to* H', *while* Δ *denotes the distance between the parallel planes* H *and* H'.

(ii) *For almost all* $y \in \text{int}\,(V)$ *there exists at most one admissible pair* $H(d^*)$, $H'(d^*)$ *such that* $y \in G[d^*]$.

(iii) *Suppose that* g *is bounded. Then* (3.19) *can also be written as*

$$G[d^*] = \text{conv}\,[(H \cap \overline{g(S)}) \cup (H' \cap \overline{g(S')})]. \tag{3.21}$$

Moreover, in this case we have for each y *satisfying* (3.15) *that there exists at least one admissible pair* $H(d^*)$, $H'(d^*)$ *such that* $y \in G[d^*]$, (*allowing us to find* $L_S(y)$ *from* (3.20)).

PROOF. If $H = H(d^*)$, $H' = H'(d^*)$ is an admissible pair then $d^* \in D^*$ so that (ii) would follow from (3.20) and Theorem 3.

Proof of (i). The second equality sign (3.20) is clear from (3.18). In view of (2.11), it remains to show that $L_S(y) \leqq \Delta(y)/\Delta$ as soon as $y \in \text{int}\,(V)$, $y \in G[d^*]$. The proof is elementary, see Corollary 13 in Section 5.

Proof of (iii). Suppose that g is bounded. Then $\overline{g(S)}$ and $\overline{g(S')}$ are compact and (3.19) implies (3.21), see (3.13).

Let y satisfy (3.15). By assertion (ii) of Theorem 6 there exists $d^* \in D^*$ with $y \in \text{conv}\, C_0(d^*)$. If $d_1 = \cdots = d_n = 0$ then, by (3.17), $d_0 \leqq 0$; in order that $C_0(d^*)$ be non-empty we must have $d_0 = 0$ so that $C_0(d^*) = \overline{g(S')}$ and $y \in W_{S'}$, contradicting (3.15); here, we used (3.7) and (3.16).

Hence, $y \in \text{conv}\, C_0(d^*)$ with $d^* \in D^*$ and $d_1, \cdots, d_n$ not all zero. In particular, (3.18) defines an admissible pair of hyperplanes H and H'. Finally, from (3.7) and (3.16),

$$C_0(d^*) = (H \cap \overline{g(S)}) \cup (H' \cap \overline{g(S')}),$$

hence, $y \in \text{conv}\, C_0(d^*) = G[d^*]$.

4. Non-interior points. Most results obtained so far require that $g_0, g_2, \cdots, g_n$ are linearly independent and that $y \in \text{int}\,(V)$. By the remarks following (2.7), we can also handle the situation that $g_1, \cdots, g_n$ are arbitrary and $y \in \text{intv}\,(V)$. One purpose of the present section is to show that the latter assumption is a rather natural one.

LEMMA 8. *Let* K *be a given non-empty convex subset of some Euclidean space* $\mathfrak{L}$ *and let* y *be a given point in* $\mathfrak{L}$. *Then each of the following three properties defines one and the same set* K_y.

(i) *If* y *is an extreme point of* K *then* $K_y = \{y\}$. *Otherwise,* K_y *is the union of* *all closed line segments* $[y', y'']$ *which are entirely contained in* K *and have* y *as an interior point.*

(ii) K_y *is the set of those points* $z \in K$ *which occur in some representation of* y *of the form*

$$y = \lambda z + (1 - \lambda)z' \quad \text{with} \quad z, z' \in K \quad \text{and} \quad 0 < \lambda < 1. \tag{4.1}$$

(iii) K_y *is the largest convex subset of* K *satisfying*

$$y \in \operatorname{intv}(K_y). \tag{4.2}$$

PROOF. Easy.

Following Karlin and Shapley [4] (who take K as a closed convex set) one may call K_y the reduced contact set of K at y. Observe that the set K_y is non-empty precisely when $y \in K$. Further, by (iii),

$$K_y = K \text{ if and only if } y \in \operatorname{intv}(K). \tag{4.3}$$

On the other hand, K_y is a strictly lower dimensional subset of K when $y \in K$, $y \notin \operatorname{intv}(K)$. In this case, y is a boundary point of K and K_y is contained in every hyperplane passing through y which supports K. An example in [4], p. 9, shows that K_y may be *strictly* smaller than the intersection of K and all such hyperplanes.

Define a *partial ordering* among the elements of K as follows. When $y, z \in K$ we put $z \prec y$ precisely when z occurs in a representation (4.1) of y. Obviously, $y \prec y$ while $u \prec z \prec y$ imply that $u \prec y$. Moreover,

$$K_y = \{z \in K : z \prec y\}$$

if $y \in K$, (otherwise, K_y is empty). Hence,

$$K_z \subset K_y \quad \text{if and only if} \quad z \in K_y. \tag{4.4}$$

Similarly, $K_z = K_y$ if and only if both $z \prec y$ and $y \prec z$, if and only if there is some open line segment entirely inside K and containing both y and z. Further, by property (iii),

$$K_z = K_y \quad \text{if and only if} \quad z \in \operatorname{intv}(K_y); \tag{4.5}$$

(starting with (4.4), the assertions all assume that $y, z \in K$).

In the following Theorem 9, we take T again as a measurable space and g as a measurable function from T into R^n. We further use the notations

$$T^y = g^{-1}(V_y), \qquad \Gamma^y = g(T) \cap V_y. \tag{4.6}$$

Here, V_y is defined as in Lemma 8, but with K replaced by V.

THEOREM 9. *Let* $y \in V$ *be fixed,* ($V = \operatorname{conv} g(T)$). *We assert that each probability measure* $\mu \in \mathfrak{M}^+(T)$ *such that* $\mu(g) \in V_y$ *is entirely concentrated on the measurable subset* T^y *of* T; *the converse is obvious. We further assert that*

$$V_y = \operatorname{conv} \Gamma^y. \tag{4.7}$$

Remark. The implications of Theorem 9 are as follows. Let $S \subset T$ and $y \varepsilon V$ be given. Then $L_S(y)$ is defined as $\inf \mu(S)$, where μ ranges over the probability measures on T with the property that $\mu(g)$ exists and is equal to y. But $y \varepsilon V_y$ so that by Theorem 9 all such measures are necessarily concentrated on T^y. Here, T^y is a proper subset of T precisely when $y \not\varepsilon \operatorname{intv}(V)$.

As far as y is concerned, one should replace the function g on T by its restriction g_y to T^y. Clearly, the range of g_y is precisely equal to Γ^y. Consequently, by (4.7), the role of $V = \operatorname{conv} g(T)$ is now taken over by V_y. From (4.2), $y \varepsilon \operatorname{intv}(V_y)$, hence, we are *back to the desired situation* but now with T replaced by $T^y = g^{-1}(V_y)$, g replaced by g_y and V replaced by V_y.

Proof of Theorem 9. Let $y \varepsilon V$ and $\mu \varepsilon \mathfrak{M}^+(T)$ be given such that

$$y = \mu(g) = \int_T g(t)\mu(dt).$$

As to the first assertion, it suffices to prove that μ is carried by the set T^y. For, afterwards, consider $\mu \varepsilon \mathfrak{M}^+(T)$ such that $\mu(g) = z \varepsilon V_y$. It then follows that μ is carried by T^z and hence by T^y since $T^z \subset T^y$ if $z \varepsilon V_y$, compare (4.4).

Let T' denote the complement of T^y in T and suppose that, on the contrary,

$$\alpha = \mu(T') > 0.$$

From Theorem 1, applied with T replaced by T', there exist finitely many points $t_i \varepsilon T'$ and positive weights p_i $(i = 1, \cdots, r)$ such that

$$\int_{T'} g(t)\mu(dt) = \sum_{i=1}^{r} p_i g(t_i) \quad \text{and} \quad \sum_{i=1}^{r} p_i = \alpha.$$

Doing the same with T replaced by T^y, one arrives at a representation

$$y = \sum_{i=1}^{m} p_i g(t_i), \quad \text{where} \quad p_i > 0, \quad \sum_{i=1}^{m} p_i = 1,$$

while $t_i \varepsilon T'$ $(i = 1, \cdots, r)$, $t_i \varepsilon T^y$ $(i = r + 1, \cdots, m)$. It follows from property (ii) of V_y (see Lemma 8) that $g(t_i) \varepsilon V_y$ for all i, that is, $t_i \varepsilon g^{-1}(V_y) = T^y$ for all i. However, for $1 \leqq i \leqq r$ this would lead to a contradiction.

It remains to prove (4.7). On the one hand, $\Gamma^y \subset V_y$ where V_y is convex so that $\operatorname{conv} \Gamma^y \subset V_y$. Next, let $z \varepsilon V_y$. By $V_y \subset V = \operatorname{conv} g(T)$ there exist finitely many points $z_i \varepsilon g(T)$ and weights $p_i > 0$ such that $z = p_1 z_1 + \cdots p_m z_m$ and $\sum p_i = 1$. In the notation preceding (4.4) (with K replaced by V) we have $z_i \prec z \prec y$, hence $z_i \varepsilon V_y = \{u \varepsilon V : u \prec y\}$. Therefore, $z_i \varepsilon g(T) \cap V_y = \Gamma^y$ for all $i = 1, \cdots m$, proving that $z \varepsilon \operatorname{conv} \Gamma^y$.

5. An important special case. Let S be a given subset of T and let

$$L_S(y) = \inf \{\mu(S) : \mu \varepsilon M^+(T), \mu(g) = y\}.$$

In most applications, one does not really need the full power of Theorem 7 in order to find $L_S(y)$. Assertion (i) will usually be sufficient and, moreover, is quite elementary as we shall see. The main use of the existence part (iii) of Theorem 7 is to assure us that the method in (i) will very often work. In particular cases, it is often easy to verify directly the existence (iii) and the uniqueness (ii).

We assume that S and S' (the complement of S in T) are non-empty. Moreover, $g_0, g_1, \cdots, g_n$ are assumed to be linearly independent so that $V = \operatorname{conv} g(T)$ has a non-empty interior. We further define V_S, W_S, $\cdots$ as in (3.12) and (3.13).

Let $y \in R^n$ be given and consider a measure $\mu \in M^+(T)$ satisfying $\mu(g) = y$ (this requires that $y \in V$). Separating the points t_i in the support of μ according to $t_i \in S$ or $t_i \in S'$, we obtain a representation

$$y = \lambda u + (1 - \lambda)u' \tag{5.1}$$

with

$$u \in V_S, \quad u' \in V_{S'}, \qquad 0 \leqq \lambda \leqq 1. \tag{5.2}$$

In fact, λ is precisely equal to $\mu(S)$. Conversely, each representation (5.1) of y satisfying (5.2) arises in this way from at least one $\mu \in M^+(T)$ satisfying $\mu(g) = y$ and $\mu(S) = \lambda$.

If both (5.1) and (5.2) hold, we shall say that y has a λ-representation and we call (u, u', λ) a V-representation of y. We shall call (u, u', λ) a W-representation of y if it satisfies (5.1) and

$$u \in W_S, \quad u' \in W_{S'}, \quad 0 \leqq \lambda \leqq 1. \tag{5.3}$$

In the following Lemma 10, $y \in R^n$ is given. Further, $\inf_{(W)} \lambda$ and $\inf_{(V)} \lambda$ range over those λ for which one can find a triplet (u, u', λ) which is a W-representation of y or a V-representation of y, respectively.

Lemma 10. *One always has*

$$\underline{\lim}_{y' \to y} L_S(y') \leqq \inf_{(W)} \lambda \leqq \inf_{(V)} \lambda = L_S(y). \tag{5.4}$$

Moreover, if g is bounded then the first equality sign holds. If $y \in \operatorname{int}(V)$ then all equality signs hold. Similarly for

$$U_S(y) = \sup_{(V)} \lambda \leqq \sup_{(W)} \lambda \leqq \overline{\lim}_{y' \to y} U_S(y'). \tag{5.5}$$

Proof. The equality sign in (5.4) is an immediate consequence of the remark following (5.2). As to the first inequality (5.4), let (u, u', λ) be a W-representation of y. Replacing u and u' by nearby points v and v' in V_S and $V_{S'}$, respectively, it follows that arbitrarily close to y there are points y' having a V-representation (v, v', λ), thus, $L_S(y') \leqq \lambda$. This proves (5.4). By (2.6), all equality signs hold in (5.4) when $y \in \operatorname{int}(V)$.

Finally, consider the case that g is bounded so that W_S and $W_{S'}$ are compact. Let γ denote the left hand side of (5.4) and let $\{y_r\}$ be a sequence converging to y such that $L_S(y_r) \to \gamma$. There exist V-representations $y_r = \lambda_r u_r + (1 - \lambda_r)u_r'$ such that $\lambda_r \to \gamma$, $(u_r \in V_S$ and $u_r' \in V_{S'})$. Drawing a subsequence, we have that $u_r \to u \in W_S$ and $u_r' \to u' \in W_{S'}$. Clearly, (u, u', γ) is a W-representation of y, thus, $\inf_{(W)} \lambda \leqq \gamma$.

Lemma 11. *A sufficient condition for $L_S(y) = 0$ is that $y \in V_{S'}$ or $y \in W_{S'} \cap \operatorname{int}(V)$. Conversely, provided g is bounded, $y \in W_{S'}$ is a necessary condition for $L_S(y) = 0$.*

Similarly for $U_S(y) = 1 - L_{S'}(y)$. *Thus, if g is bounded,* $y \varepsilon \text{int}(V)$ *then* $U_S(y) = 1$ *if and only* $y \varepsilon W_S$. *Finally,*

$$(5.6) \qquad 0 \leqq L_S(y) < U_S(y) \leqq 1 \quad \textit{for each} \quad y \varepsilon \text{int}(V),$$

unless $g(S)$ *and* $g(S')$ *are located in distinct parallel hyperplanes.*

Proof. All but (5.6) are easy, compare (2.6). As to (5.6), suppose that $L_S(y_0) = U_S(y_0)$ holds for some $y_0 \varepsilon \text{int}(V)$. Since $f(y) = U_S(y) - L_S(y)$ is non-negative and concave, it follows that $f(y) = 0$ for all $y \varepsilon V$. Therefore $L_S(y) = U_S(y)$ is both convex and concave and thus *linear* (up to an additive constant). For different γ, the sets $V_\gamma = \{y \varepsilon V: L_S(y) = \gamma\}$ are located in distinct parallel hyperplanes. Clearly, $g(S') \subset V_0$ and $g(S) \subset V_1$.

Notations. Let C be a non-empty subset of R^n and $d \varepsilon R^n$. Put

$$(5.7) \qquad \phi_d(C) = \sup_{y \varepsilon C} \textstyle\sum_{i=1}^n d_i y_i \quad ,$$

and

$$(5.8) \qquad H_d(C) = \{x \varepsilon R^n: \textstyle\sum_{i=1}^n d_i x_i = \phi_d(C)\}.$$

If $d \neq 0$ and $\phi_d(C) < \infty$ then $H_d(C)$ is the hyperplane supporting C in the direction d as well as possible. If $d = 0$ then $H_d(C) = R^n$; if $\phi_d(C) = \infty$ then $H_d(C)$ is empty.

Note that $H_d(C)$ remains unchanged on replacing d by λd with λ as a positive scalar. Moreover, $\phi_d(C)$ and $H_d(C)$ remain unchanged on replacing C by $\bar{C}$, or by conv C, or by $\overline{\text{conv}}\, C$, respectively. For brevity, we shall put $\phi_d = \phi_d(W_S)$, $\phi_d' = \phi_d'(W_{S'})$ and

$$(5.9) \qquad H_d = H_d(W_S), \qquad H_d' = H_d(W_{S'}).$$

Thus, $H_d = \{x \varepsilon R^n: \sum_{i=1}^n d_i x_i = \phi_d\}$ is the hyperplane supporting W_S in the direction d as well as possibly (hence, also $g(S)$ and V_S). Similarly, H_d' is the hyperplane which supports each of the sets $g(S')$, $V_{S'}$ and $W_{S'}$ as well as possible in the direction d.

Theorem 12. *Be given* $y \varepsilon R^n$. *Suppose that* $d \varepsilon R^n$ *and* $\gamma \varepsilon R$ *are such that y admits a representation of the form*

$$(5.10) \qquad y = \gamma u + (1 - \gamma) u', \qquad 0 \leqq \gamma \leqq 1,$$

with

$$(5.11) \qquad u \varepsilon H_d \cap V_S \quad \textit{and} \quad u' \varepsilon H_d' \cap V_{S'}.$$

Then

$$(5.12) \qquad \gamma = L_S(y) \quad \textit{if} \quad \phi_d' < \phi_d < +\infty,$$

while

$$(5.13) \qquad \gamma = U_S(y) \quad \textit{if} \quad \phi_d < \phi_d' < +\infty.$$

Provided $y \varepsilon \text{int}(V)$, *the assertions* (5.12) *and* (5.13) *remain true when* (5.11) *is replaced by the weaker condition that*

(5.14) $$u \,\varepsilon\, H_d \cap W_S \quad \textit{and} \quad u' \,\varepsilon\, H_d' \cap W_{S'} .$$

PROOF. Put $\phi_d(W_S) = \phi_d = \phi$ and $\phi_d(W_{S'}) = \phi_d' = \phi'$. We have $dx \leqq \phi$ when $x \,\varepsilon\, W_S$ with equality when $x \,\varepsilon\, H_d$. We have $dx \leqq \phi'$ when $x \,\varepsilon\, W_{S'}$ with equality when $x \,\varepsilon\, H_d'$.

Now suppose that both (5.10) and (5.11) hold. Using the same notations as in Lemma 10, put

$$\inf_{(V)} \lambda = \rho, \qquad \sup_{(V)} \lambda = \sigma.$$

We must show that $\gamma = \rho$ when $\phi > \phi'$ are finite and further that $\gamma = \sigma$ when $\phi < \phi'$ are finite. Observe that $\rho \leqq \gamma \leqq \sigma$ since (u, u', γ) is a V-representation of y.

Let (v, v', λ) be some other V-representation of y. Then

$$dy = d[\lambda v + (1 - \lambda)v'] \leqq \lambda\phi + (1 - \lambda)\phi'.$$

Moreover, using (5.11),

$$dy = d[\gamma u + (1 - \gamma)u'] = \gamma\phi + (1 - \gamma)\phi'.$$

It follows that $\gamma(\phi - \phi') \leqq \lambda(\phi - \phi')$ if ϕ, ϕ' are finite. If $\phi > \phi'$ this implies that $\gamma \leqq \rho$ hence $\gamma = \rho$. Similarly, $\gamma = \sigma$ if $\phi < \phi'$ are finite.

If (5.11) is replaced by (5.14) then exactly the same proof (but now with W-representations) yields that

$$\inf_{(W)} \lambda = \gamma \quad \text{if} \quad \phi > \phi', \qquad \sup_{(W)} \lambda = \gamma \quad \text{if} \quad \phi < \phi',$$

(ϕ and ϕ' assumed finite). This in turn implies (5.12) and (5.13), provided the equality signs hold in (5.4) and (5.5); the latter is always true when $y \,\varepsilon\, \text{int}\,(V)$, by Lemma 10.

In the paragraph following (3.18) we defined the notion of an *admissible pair* H, H' of hyperplanes in R^n (relative to the subset S of T). Using the notations (5.9), this is nothing but a pair

(5.15) $$H = H_d = H_d(W_S), \qquad H' = H_d' = H_d(W_{S'}) \quad \text{with} \quad \phi_d' < \phi_d < +\infty.$$

COROLLARY 13. *Let H, H' be an admissible pair of hyperplanes as in* (5.15). *Then for each point $y \,\varepsilon$* int (V) *such that $y \,\varepsilon\, G_d$, with*

(5.16) $$G_d = \overline{\text{conv}}\,[(H \cap W_S) \cup (H' \cap W_{S'})],$$

we have that

(5.17) $$L_S(y) = \Delta(y)/\Delta.$$

Here, $\Delta(y)$ denotes the distance from y to H' and Δ the distance between the parallel planes H and H'.

PROOF. Since $L_S(y)$ is continuous in int (V), we need to establish (5.17) only for a point y having a representation (5.10) with u and u' as in (5.14). Clearly, $\gamma = \Delta(y)/\Delta$, hence, (5.12) implies (5.17).

REMARK. Let $d \varepsilon R^n$ be such that $\phi_d' < \phi_d < \infty$; then $H = H_d$ and $H' = H_d'$ form an admissible pair. They admit a representation (3.18), with $d^* = (d_0^*, d_1^*, \cdots, d_n^*)$ given by

$$(5.18) \quad d_j^* = c\, d_j \; (j = 1, \cdots, n); \qquad d_0^* = -c\phi_d'; \qquad c = 1/(\phi_d - \phi_d').$$

Moreover, $d^* \varepsilon D^*$ (in the sense of (3.17)) and

$$(5.19) \qquad G_d = G[d^*],$$

compare (3.19). Consequently, Theorem 7 gives us considerable information about the occurrence of $y \varepsilon G_d$ with H, H' admissible.

Given $y \varepsilon V$ let us consider the following five properties.

(i) For some $\mu \varepsilon M_+(T)$ we have $\mu(g) = y$ and $\mu(S) = L_S(y)$.

(ii) Among all V-representations (u, u', λ) of y there is one for which λ is minimal.

(iii) For some $d \varepsilon R^n$ we have $\phi_d' < \phi_d < \infty$ and

$$(5.20) \qquad y \varepsilon \operatorname{conv} [(H_d \cap V_S) \cup (H_d' \cap V_{S'})].$$

(iv) Among all W-representations (u, u', λ) of y there is one for which λ is minimal.

(v) For some $d \varepsilon R^n$ we have $\phi_d' < \phi_d < \infty$ and

$$(5.21) \qquad y \varepsilon \operatorname{conv} [(H_d \cap W_S) \cup (H_d' \cap W_{S'})].$$

Observe that (in both cases (iii) and (v)) the pair of hyperplanes H_d, H_d' is admissible; further, $L_S(y)$ may be read off from either (5.12) or (5.17). By Theorem 7, we have for almost all y that the pair H_d, H_d' in (v) (if it exists) must be unique.

THEOREM 14. *Let y satisfy*

$$(5.22) \qquad y \varepsilon \operatorname{int}(V), \quad y \not\varepsilon W_{S'}.$$

Then the above properties (i), (ii), (iii) *are equivalent. A sufficient condition for each of these would be that both V and $V_{S'}$ be compact.*

Moreover, the above properties (iv) *and* (v) *are equivalent. A sufficient condition for each would be that g be bounded.*

PROOF. That (i), (ii) are equivalent follows by the remark following (5.2). That (iii) $\Rightarrow$ (ii) and that (v) $\Rightarrow$ (iv) is shown in the proof of Theorem 12. The stated sufficiency conditions for (ii) and (iv), respectively, are easily verified. It remains to show that (ii) $\Rightarrow$ (iii) and that (iv) $\Rightarrow$ (v). Let us prove these simultaneously (though the first could also be deduced from Theorem 5).

The situation is as follows. We are given a pair of non-empty convex sets K and K' in R^n, (namely, either V_S and $V_{S'}$, or W_S and $W_{S'}$). Consider further $K_0 = \operatorname{conv}(K \cup K')$, (namely, $K_0 = V$ or $V \subset K_0 \subset W$) and further a fixed point $y \varepsilon R^n$ such that

$$y \varepsilon \operatorname{int}(K_0), \qquad y \not\varepsilon K'.$$

By a K-representation (of this fixed point y) we shall mean a triplet (u, u', λ) satisfying

(5.23) $\quad y = \lambda u + (1 - \lambda)u'; \quad u \varepsilon K, \quad u' \varepsilon K'; \quad 0 \leqq \lambda \leqq 1.$

Clearly, there is at least one. We assume that among all such K-representations there is at least one for which λ *assumes* its smallest value γ (say). Let the corresponding triplet (u_0, u_0', γ) be fixed. It suffices to prove that $d \varepsilon R^n$ exists such that

(5.24) $\quad \phi_d(K') < \phi_d(K) < \infty; \qquad u_0 \varepsilon H_d(K) \cap K, \qquad u_0' \varepsilon H_d(K') \cap K'.$

Without any real loss of generality, we may assume that $y = 0$, thus,

$$\gamma u_0 + (1 - \gamma)u_0' = 0; \qquad 0 \varepsilon \operatorname{int}(K_0), \qquad 0 \not\varepsilon K',$$

where $u_0 \varepsilon K$, $u_0' \varepsilon K'$ hence $\gamma > 0$. By the minimality of γ, we have that (5.23) (with $y = 0$) implies that $\lambda \geqq \gamma$. Therefore,

$$\rho x + \sigma u' = 0; \qquad x \varepsilon K_0, u' \varepsilon K'; \qquad \rho, \sigma \geqq 0 \quad \text{imply} \quad \rho \geqq \gamma(\rho + \sigma);$$

(after all, write $x = \alpha v + (1 - \alpha)v'$ with $v \varepsilon K$, $v' \varepsilon K'$).

Given $u' \varepsilon K'$ we have, for $\delta > 0$ sufficiently small, that $x = -\delta u' \varepsilon K_0$ thus $1 \geqq \gamma(1 + \delta)$ showing that $\gamma < 1$. Put $q = (1 - \gamma)/\gamma$ $(0 < q < \infty)$ and $L = -qK' = \{-qu' : u' \varepsilon K'\}$.

We assert that L and $\operatorname{int}(K_0)$ are disjoint. If not then there would exist elements $x_0 \varepsilon \operatorname{int}(K_0)$ and $u' \varepsilon K'$ with $x_0 = -qu'$. Further, if $\delta > 0$ is sufficiently small we have $x = (1 + \delta)x_0 \varepsilon K_0$ and $(1 + \delta)^{-1}x + qu' = 0$. It would follow that

$$(1 + \delta)^{-1} \geqq \gamma[(1 + \delta)^{-1} + q] = \gamma(1 + \delta)^{-1} + (1 - \gamma),$$

a contradiction since $\gamma < 1$.

By a well-known separation theorem ([18], p. 24) we can now conclude that the convex sets K_0 and L are separated by at least one non-zero linear functional f on R^n. That is, for some constant C we have:

(a) $f(x) \leqq C$ for each $x \varepsilon K_0$. Since $0 \varepsilon \operatorname{int}(K_0)$ this implies that $C > 0$.

(b) $f(x) \geqq C$ for each $x \varepsilon L = -qK'$. Equivalently, $f(x) \leqq -C/q$ for each $x \varepsilon K'$.

(c) The point $u_0 = -qu_0'$ is in both K and L, therefore, $f(u_0) = C$ and $f(u_0') = -C/q$.

It follows from these remarks that the pair of hyperplanes

$$H_d = \{x \varepsilon R^n : f(x) = C\} \quad \text{and} \quad H_d' = \{x \varepsilon R^n : f(x) = -C/q\}$$

has all the properties required in (5.24). Here, d stands for a common normal pointing from H_d' to H_d.

6. Comments. Let us describe the situation on hand in a somewhat heuristic manner. Let T be a given (measurable) space and let μ denote a given probability measure on T with known moments $\mu(g_j)$, $(j = 1, \cdots, n)$. Here, $g_1, \cdots, g_n$ are given functions on T. Next, let S be a given subset of T. We want to determine the best lower bound $L_S(y)$ and the best upper bound $U_S(y)$ on the probability mass $\mu(S)$ contained in S. Here, $y = (y_1, \cdots, y_n)$ with $y_j = \mu(g_j)$. That is, $y = \mu(g)$ where $g = (g_1, \cdots, g_n)$ maps T into R^n. Necessarily $y \varepsilon V = \text{conv}\, g(T)$.

Given μ, it is natural to consider the induced probability measure ν on $g(T)$ defined by $\nu(B) = \mu(g^{-1}B)$. If μ has a finite support $\{t_1, \cdots, t_m\} \subset T$ with mass p_i at t_i then ν corresponds to mass distribution on the range $g(T) \subset R^n$ having a mass p_i at the point $g(t_i)$ on $g(T)$.

We can now also describe $U_S(y)$ as the largest possible mass (in the supremum sense) contained in the set $g(S)$ among all mass distributions on $g(T)$ of total mass 1 having a *preassigned center of gravity* y.

Let S' denote the complement of S in T. Observe that the sets $g(S)$ and $g(S')$ need not be disjoint. If one is interested in the lower bound $L_S(y)$ then a point z belonging to both $g(S)$ and $g(S')$ may as well be regarded as belonging to $g(S')$. More precisely, $1 - L_S(y) = U_{S'}(y)$ is equal to the largest possible mass in $g(S')$ among all mass distributions on $g(T)$ of total mass 1 having their center of gravity at y.

We may assume (see the remarks after (2.7)) that $g_0 \equiv 1, g_1, \cdots, g_n$ are linearly independent. That is, $g(T)$ is not located on any hyperplane, equivalently, int (V) is non-empty. For the moment, consider the case that $y \varepsilon V$ is on the boundary of V. Then V is supported by at least one hyperplane H passing through y. In order that a mass distribution on $g(T)$ have y as its center of gravity the latter distribution must be carried by $H \cap g(T)$ so that the behavior of the set $g(T)$ outside the supporting hyperplane H would be totally irrelevant as far as $L_S(y)$ and $U_S(y)$ is concerned. For more details, see Section 4.

From now on we shall restrict ourselves to the case $y \varepsilon \text{int}\,(V)$, so that both $L_S(\cdot)$ and $U_S(\cdot)$ are continuous in an entire neighborhood of y. This in turn implies (Lemma 10) that $U_S(y)$ is also equal to the largest possible mass contained in the closure $\overline{g(S)}$ among all mass distributions on $\overline{g(T)}$ of total mass 1 having y as its center of gravity. This statement remains valid when $\overline{g(S)}$ is replaced by $W_S = \overline{\text{conv}}\, g(S)$ and $\overline{g(T)}$ is replaced by $W = \overline{\text{conv}}\, g(T) = \bar{V}$.

In the same way, $1 - L_S(y) = U_{S'}(y)$ would be equal to the largest possible mass in $\overline{g(S')}$ among all mass distributions of total mass 1 in $\overline{g(T)}$ having y as a center of gravity. Similarly, when $\overline{g(S')}$ is replaced by $W_{S'} = \overline{\text{conv}}\, g(S')$ and $\overline{g(T)}$ by W. Since $W_{S'}$ and W are convex this means that $L_S(y)$ is equal to the smallest possible value $0 \leqq \lambda \leqq 1$ (in the infimum sense) among all possible representations of y as $y = \lambda u + (1 - \lambda)u'$ with $u \varepsilon W$ and $u' \varepsilon W_{S'}$. Clearly, $L_S(y) = 0$ when $y \varepsilon W_{S'}$. Otherwise, $L_S(y)$ may be described as the smallest possible value $0 \leqq \lambda \leqq 1$ for which the closed convex sets $-q(W_{S'} - y)$ and $(W - y)$ have a common point; here, $q = (1 - \lambda)/\lambda$. If $y \not\varepsilon W_{S'}$ then $W_{S'} - y$ is located in some half space of the form $\{z\colon dz \geqq \epsilon > 0\}$. As $\lambda \to 0$ we have $q \to +\infty$ and $-q\,(W_{S'} - y)$ recedes to infinity. Provided g and hence W is bounded, the above smallest value $0 \leqq \lambda \leqq 1$ is *assumed.* It is characterized by the property that the associated closed convex sets $-q(W_{S'} - y)$ and $(W - y)$ are "barely" touching each other; (clearly, this is the main idea behind the proof of Theorem 14).

The main drawback of this method (for computing $L_S(y)$) is that it is usually hard to visualize the geometry and relative positions of the convex sets $W - y$ and $-q(W_{S'} - y)$, for different values of q, especially when these sets are located in a Euclidean space of dimension $n \geqq 3$.

Therefore, we now direct our attention to the hyperplane F which separates $W - y$ and $-q(W_{S'} - y)$ at the moment that they are barely touching each other at a point $u - y = -q(u' - y) = w$. In particular, $w \varepsilon F$, $u' \varepsilon W_{S'}$ and $u \varepsilon W$; in fact, $u \varepsilon W_S$ since λ is minimal. Consider the parallel distinct hyperplanes

$$H = y + F, \qquad H' = y - q^{-1}F.$$

Then $u = y + w \varepsilon H$, $u' = y - q^{-1}w \varepsilon H'$ and $y = \lambda u + (1 - \lambda)u'$ with $\lambda = (1 + q)^{-1} = L_S(y)$. Observe that H supports all of W while H' supports all of $W_{S'}$. Moreover, $u \varepsilon W \cap H = W_S \cap H$ and $u' \varepsilon W_{S'} \cap H'$. In this way, we are led to the following procedure.

(STEP 1) Select a boundary point z of W and "draw" through z a hyperplane H of support to the convex set W.

(STEP 2) Next determine the hyperplane H' parallel to H which supports $W_{S'}$ as well as possible, and on the same side as H supports W. We shall only be interested in the case $H \neq H'$ in which case H' is between H and $W_{S'}$. Put

$$A_d = W \cap H = W_S \cap H \qquad \text{and} \qquad B_d = W_{S'} \cap H'.$$

Here, d stands for the common normal to the parallel planes H and H', pointing from H' to H. Thus, in the notation (5.8), we have

$$(6.1) \qquad H = H_d = H_d(W), \qquad H' = H_d' = H_d(W_{S'})$$

and further $\phi_d' < \phi_d < \infty$.

(STEP 3) Provided indeed $H' \neq H$, put

$$(6.2) \qquad G_d = \overline{\text{conv}}\,(A_d \cup B_d).$$

We now conclude from Corollary 13 that

$$(6.3) \qquad L_S(y) = \Delta(y)/\Delta \quad \text{for each} \quad y \varepsilon \operatorname{int}(V) \quad \text{such that} \quad y \varepsilon G_d.$$

Here, $\Delta(y)$ denotes the distance from y to H' while Δ denotes the distance between the distinct parallel hyperplanes H, H'.

Actually, the assertion (6.3) is rather transparent. Namely, if a mass distribution of total mass 1 has its center of gravity at a point y between H and H' (such as a point $y \varepsilon G_d$) then the mass in $W_{S'}$ could not possibly be any larger than $1 - (\Delta y)/\Delta$. Moreover, if $y \varepsilon G_d$ then one can attain this mass, namely, by having a mass equal to $1 - \Delta(y)/\Delta$ in the subset B_d of H' and a mass equal to

$\Delta(y)/\Delta$ in the subset A_d of H in such a way that the resulting center of gravity is equal to y.

The importance of the above procedure can, for instance, be seen from the fact that, for each $y \varepsilon \operatorname{int}(V)$ with $y \not\varepsilon W_{S'}$, and whenever g is bounded, there always exists a pair H, H' as above with $y \varepsilon G_d$, (allowing us to find $L_S(y)$ from (6.3)). This fact is certainly suggested by the reasoning which leads up to the above procedure. And it is a consequence of Theorem 7 and also of Theorem 14; (note that the relevant part of Theorem 14 is proved by the same type of reasoning as the above). Theorem 7 further yields that for almost all y there is at most one admissible pair H_d, H_d' with $y \varepsilon G_d$.

As appears from certain special cases, see Section 7, it is usually easier to start out with the collection of *all* admissible pairs H_d, H_d' $(H_d \neq H_d')$. That is, one first determines in what manner the part of $\operatorname{int}(V)$ not in $W_{S'}$ is partitioned into classes G_d. Afterwards, one may see in which of these classes a preassigned value y happens to fall. Knowing this class G_d and the corresponding pair H_d, H_d' of hyperplanes, the desired value $L_S(y)$ is found from (6.3).

The procedure for finding $U_S(y) = 1 - L_{S'}(y)$ is completely analogous to the procedure sketched above.

(Step 1) Select a boundary point z of the closed convex set W_S and draw through z a hyperplane H of support to W_S. Put $A_d = W_S \cap H$.

(Step 2) Determine the hyperplane H' parallel to H which supports $g(T)$ and hence W as well as possible, and on the same side as H supports W_S. We shall only be interested in the case that $H' \neq H$ in which case H is between H' and W_S. Put $B_d = W \cap H' = W_{S'} \cap H'$. Let G_d be as in (6.2). Then

$$(6.4) \qquad U_S(y) = \Delta(y)/\Delta \quad \text{for each} \quad y \varepsilon \operatorname{int}(V) \quad \text{with} \quad y \varepsilon G_d,$$

assuming that H and H' are distinct. Here, $\Delta(y)$ and Δ have the same meaning as in (6.3).

If $y \varepsilon W_S$, $y \varepsilon \operatorname{int}(V)$ then $U_S(y) = 1$. If $y \not\varepsilon W_S$, $y \varepsilon \operatorname{int}(V)$ and, moreover, g is bounded then there always exists a pair H, H' of the above kind such that $U_S(y)$ can be obtained from (6.4). For almost all y this pair will be unique, thus, the different regions G_d hardly overlap.

The question arises what to do when the function g on T is not bounded. One way would be to use the formula $L_S(y) = (1 + q)^{-1}$ with $q \geqq 0$ as the largest value (in the supremum sense) for which $W - y$ and $-q(W_{S'} - y)$ have a common point, an analogous formula holding for $U_S(y)$.

Often a better way of treating the unbounded case is the following reduction to the bounded case.

One can always represent the domain T of g as a denumerable union

$$(6.5) \qquad T = \bigcup_{N=1}^{\infty} T_N \quad \text{with} \quad T_N \subset T_{N+1},$$

such that g is bounded on each T_N. Let $\{T_N\}$ be fixed, and put

$$L_S^{(N)}(y) = \inf\{\mu(S) \colon \mu \varepsilon M^+(T_N), \mu(g) = y\}.$$

We have

$$L_S^{(N)}(y) \downarrow L_S(y) \quad \text{as} \quad N \to \infty, \tag{6.6}$$

since each $\mu \varepsilon M^+(T)$ has a finite support. This reduces the problem to a computation of the values $L_S^{(N)}(y)$. But g is bounded on T_N, hence, the value $L_S^{(N)}(y)$ can in principle be obtained from (6.3) (with T replaced by T_N), provided

$$y \varepsilon \operatorname{int}(V^{(N)}), \qquad y \not\varepsilon W_{S'}^{(N)} = \overline{\operatorname{conv}}\, g(S' \cap T_N), \tag{6.7}$$

where $V^{(N)} = \operatorname{conv} g(T_N)$. As is easily seen, (6.7) will be true for N sufficiently large, provided $y \varepsilon \operatorname{int}(V), y \not\varepsilon W_{S'}$. Recall that $L_S(y) = 0$ when $y \varepsilon \operatorname{int}(V)$, $y \varepsilon W_{S'}$.

7. Specific applications. For the special case that $n = 2$ and T is a one-dimensional interval, we shall now give a number of applications of the methods (in several steps) outlined in Section 6. Here, "hyperplane" is to be interpreted as a straight line in R^2.

Let g_1, g_2 be given real-valued functions on the interval T such that g_0, g_1, g_2 are linearly independent. Let further S be a given subset of T. We shall be interested in computing the values $L_S(y)$ and $U_S(y)$ which may be interpreted as the best bounds in

$$L_S(y) \leqq P(X \varepsilon S) \leqq U_S(y), \tag{7.1}$$

when X is a random variable taking values in T such that

$$E(g_1(X)) = y_1, \qquad E(g_2(X)) = y_2, \tag{7.2}$$

where the expectations are assumed to exist.

(a) As our first application, we take T as the unit interval $T = [0, 1]$ and further

$$g_1(t) = t, \qquad g_2(t) = \sin 2\pi t, \qquad (0 \leqq t \leqq 1). \tag{7.3}$$

Thus $g(T)$ is presently a part of the ordinary sine-curve. Its convex hull V is compact (thus $W = V$) and is given by the planar region in Figure 1 bounded by the curve $0_1 0_2 ABC 0_1$. This boundary curve has two straight line segments on it, namely, the parallel segments AB and 0_1C which are tangent to the sine-curve at A and C, respectively.

Let us further choose S as the interval

$$S = [0, \tfrac{1}{4}], \quad \text{thus,} \quad S' = (\tfrac{1}{4}, 1].$$

Then $W_{S'}$ is given by the shaded region in Figure 1; it is bounded by the curve 0_2ABCD0_2. This boundary curve has two straight parts, namely, the line segment AB and further the line segment 0_2D, the latter being tangent to the sine-curve at the point D.

The straight line 0_2D clearly supports $W_{S'}$. Moreover, a line through 0_1 parallel to 0_2D will support all of W, hence, these lines act as a pair H_d', H_d as

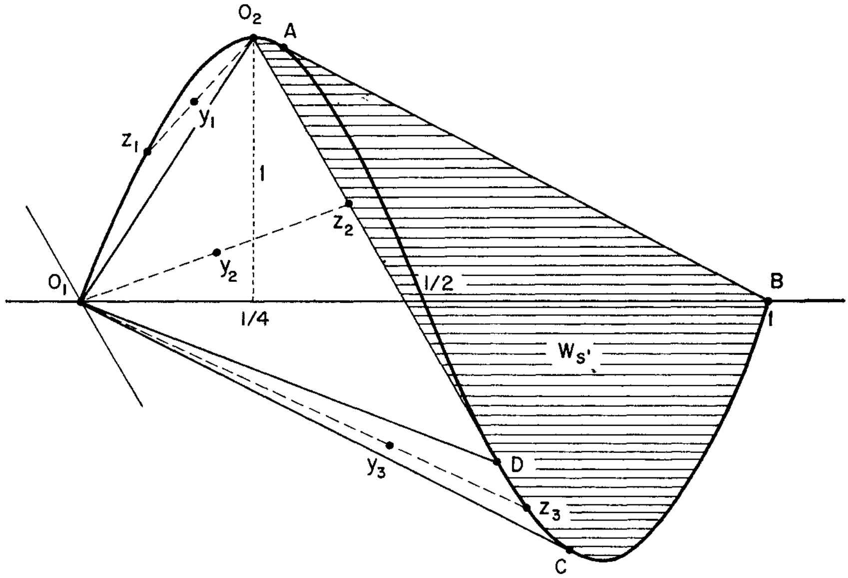

FIG. 1

needed in the computation of $L_S(y)$. More precisely, $A_d = W \cap H_d = W_S \cap H_d$ consists of the single point $0_1 = g(0)$ while $B_d = W_{S'} \cap H_d'$ consists of the closed line segment 0_2D, so that $G_d = \overline{\text{conv}}\ (A_d \cup B_d)$ is the closed triangle 0_10_2D. For each point y in this triangle with $y \varepsilon \text{ int } (V)$ (thus the points 0_1 and 0_2 are excluded) the value $L_S(y)$ is given by (6.3). Equivalently, we have for these points y that

$$L_S(y) = \overset{\frown}{yz}/\overset{\frown}{0_1z}. \tag{7.4}$$

Here, $\overset{\frown}{yz}$ denotes the Euclidean distance between the points y and z, similarly, $\overset{\frown}{0_1z}$. Further z denotes the point where the (straight) line 0_1y intersects the line segments 0_2D, (compare the points y_2 and z_2 in Figure 1). As to the point 0_1, here (7.4) has no meaning but (0.3) happens to be true with $L_S(0_1) = 1$. As to the point 0_2, here both (6.3) and (7.4) are false; namely, $L_S(0_2) = 1$ because $0_2 = g(\frac{1}{4})$ and $\frac{1}{4} \varepsilon S$.

For a point $y \varepsilon \text{ int } (V)$ above the line segment 0_10_2 we have

$$L_S(y) = \overset{\frown}{0_2y}/\overset{\frown}{0_2z}, \tag{7.5}$$

where z denotes the point where the straight line 0_2y meets the sine-curve, (compare the points y_1 and z_1 in Figure 1). After all, take H_d as the *tangent* line at z to the sine curve (which supports all of W) and take H_d' as the line through 0_2 parallel to H_d (which supports the entire shaded region $W_{S'}$).

For a point y below the line 0_1D, $y \varepsilon \text{ int } (V)$, we again have formula (7.4) but

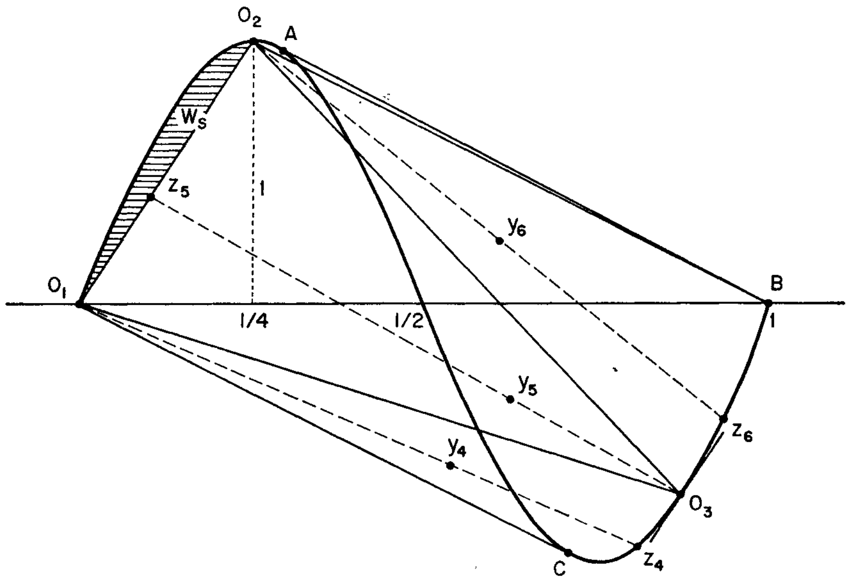

FIG. 2

now with z as the point where the line 0_1y *first* intersects the sine-curve, (compare the points y_3 and z_3 in Figure 1). After all, take H_d' as the tangent line at z to the sine-curve and take H_d' as a line through 0_1 parallel to H_d'.

Finally, $L_S(y) = 0$ when $y \varepsilon W_{S'}$, $y \varepsilon \operatorname{int}(V)$. In this way we have subdivided int (V) into *four* regions each with its own method (formula) for computing $L_S(y)$. Note that we strongly used the fact that $S = [0, \frac{1}{4}]$ is closed on the left. If 0 would be removed (thus $0 \varepsilon S'$) then 0_1 would belong to $W_{S'}$ and we would have $L_S(y) = 0$ for any point y not above 0_10_2 (not of the type y_1). On the other hand, it hardly matters whether or not $\frac{1}{4} \varepsilon S$ except that $L_S(0_2)$ would be reduced to 0 when the point $\frac{1}{4}$ would be assigned to S' instead of S.

(b) Still assuming (7.3) and $S = [0, \frac{1}{4}]$, the upper bound $U_S(y)$ can easily be read off from Figure 2. Here, 0_3 denotes the point on the right at which the tangent line to the sine-curve is *parallel to the secant* 0_10_2. This tangent line supports all of W while the parallel line 0_10_2 supports W_S, hence, these lines act as a pair H_d' and H_d, respectively, as needed in the computation of $U_S(y)$. We have $A_d = W_S \cap H_d$ equal to the line segment 0_10_2 and $B_d = W \cap H_d' = W_{S'} \cap H_d'$ equal to the single point 0_3, so that $G_d = \operatorname{conv}(A_d \cup B_d)$ is equal to the closed triangle $0_10_20_3$. For each point $y \varepsilon \operatorname{int}(V)$ belonging to this triangle, the value $U_S(y)$ is given by (6.4). Equivalently, we have for these points y that

$$U_S(y) = \widehat{0_3y}/\widehat{0_3z}, \tag{7.6}$$

where z denotes the point at which the line 0_3y meets the line 0_10_2, (compare the points y_5 and z_5 in Figure 2).

From now on it will be tacitly assumed that $y \varepsilon \operatorname{int}(V)$. If y is below the line $0_1 0_3$ then

$$U_S(y) = \widehat{yz}/\widehat{0_1 z}, \tag{7.7}$$

where z denotes the point where the line $0_2 y$ intersects the sine-curve for the *second* time, (compare the points y_4 and z_4 in Figure 2). After all, take H_d' as the tangent line at z to the sine-curve (it supports all of W) and take H_d as the line through 0_1 parallel to H_d' (which supports the shaded region W_S).

If y is above the line $0_2 0_3$ then

$$U_b(y) = \widehat{yz}/\widehat{0_2 z}, \tag{7.8}$$

where z denotes the point at which the straight line $0_2 y$ intersects the boundary of $V = W$. There are *three* different situations according to whether z is on the arc $0_2 A$ of the sine-curve, on the line segment AB, or on the arc $B0_3$ of the sine-curve, respectively; (for the latter case, compare the points y_6 and z_6 in Figure 2). The proof is easy: one merely applies (6.4) with H_d' as the line through z supporting all of W and with H_d as the line through 0_2 parallel to H_d', (supporting all of W_S). Here, G_d is precisely the line segment $0_2 z$.

Finally, $U_S(y) = 1$ if $y \varepsilon W_S$, (that is, y above the chord $0_1 0_2$). In this way we have subdivided int (V) into *six* regions each with its own analytic formula for computing $U_S(y)$. It would be difficult to describe the situation without using pictures, while geometrically the situation is rather simple.

(c) As a further example, take $T = [0, 1]$ and

$$g_1(t) = \cos 2\pi t, \qquad g_2(t) = \sin 2\pi t, \qquad (0 \leqq t \leqq 1). \tag{7.9}$$

Here, $g(T)$ is the unit circle in the plane and $W = V = \operatorname{conv} g(T)$ is the unit disc. Take $S = [a, b]$ where $0 \leqq a < b < 1$. In finding $L_S(y)$ one must distinguish between four different regions obtained by inscribing in the unit circle the triangle with vertices $g(a)$, $g(b)$ and 0_3. Here, $0_3 = g(t_1)$ is the unique point with $a < t_1 < b$ at which the tangent line to the unit circle is parallel with the chord from $g(a)$ to $g(b)$.

The reader will have no difficulty reading off $L_S(y)$ from the resulting figure, namely, by means of (6.3). In a similar way one obtains $U_S(y)$ from (6.4), except that now $t_1 \not\varepsilon [a, b]$. The present example is not quite a special case of the next example, because of the fact that $\sin 2\pi t$ has three zeros in [0, 1].

(d) From now on, we shall be concerned with the case that S and T are one-dimensional intervals and that $\{g_0, g_1, g_2\}$ form a Tchebycheff system, see [5], [7]; here $g_0(t) \equiv 1$. In other words, we assume that $g_1(t)$, $g_2(t)$ are real-valued continuous functions on T such that the "polynomial"

$$f(t) = d_0 + d_1 g_1(t) + d_2 g_2(t) \tag{7.10}$$

cannot have more than two distinct zeros, unless all the (real) coefficients d_j are equal to zero.

In the special case $g_1(t) = t$, the above conditions imply that the function g_2 is either strictly convex or strictly concave. Essentially no more general is the case where $g_1(t)$ is monotone on T; (for, introduce the new parameter $t' = g_1(t)$).

Let $t_0 \varepsilon T$ be given. Then ([5], p. 28, [7], p. 41) there exists a polynomial (7.10) with $f(t_0) = 0$ and $f(t) > 0$ for $t \neq t_0$; however, if T is compact and t_0 is an endpoint of T one must allow the (very real) possibility that $f(t) = 0$ also at the other endpoint of T. Geometrically, this means that through the point $g(t_0)$ there passes at least one line supporting $g(T)$ (and thus V) which meets $g(T)$ in at most one other point. In particular, each point on $g(T)$ is an extreme point of V. The converse is obvious.

For a while, we shall only consider the case that T is a compact interval $T = [\alpha, \beta]$ with endpoints α and β. Then $W = V$ is a compact and convex region in the plane bounded on one side by the "convex" arc $g(T)$ (which has no more than two points in common with any straight line) and on the other side by the straight line segment λ passing through the endpoints $0_3 = g(\alpha)$ and $0_4 = g(\beta)$. For definiteness, we shall assume in the following discussion that $g_1(\alpha) < g_1(\beta)$ (0_3 is to the left of 0_4) and further that the arc $g(T)$ is below the chord λ.

In the present case, $W_S = \overline{\text{conv}}\, g(S)$ is precisely the compact and convex region cut off from W by the straight line segment l passing through the points $0_1 = g(a)$ and $0_2 = g(b)$ corresponding to the endpoints of $S = [a, b]$.

Finally, $W_{S'} = \overline{\text{conv}}\, g(S')$ is the part of W between λ and l, *except* for the special case that S' is entirely at one side of $T = [\alpha, \beta]$. Namely, if $a = \alpha \varepsilon S$, so that $S = [\alpha, b]$, $S' = (b, \beta]$, then $W_{S'}$ is the part of W cut off by the line segment $0_2 0_4$; (the situation would be different if $S = (\alpha, b]$ where $S' = \{\alpha\} \cup (b, \beta]$). Similarly, if $b = \beta \varepsilon S$ we have that $W_{S'}$ is the part of W cut off by the line segment $0_1 0_3$.

(d)′ In Figure 3 it is indicated how to obtain $U_S(y)$ when λ and l intersect on the left; (if they would intersect on the right then not $0_1 0_4$ but $0_2 0_3$ would be a dividing line). If y belongs to region I we have

$$U_S(y) = \overset{\frown}{yz}/\overset{\frown}{0_1 z}, \tag{7.11}$$

with z as the point where the extended line $0_1 y$ meets $g(T)$. After all, take H_d' as the line through z supporting W and take H_d as the line through 0_1 parallel to H_d'.

Region IV is analogous to region I except that the role of 0_1 is taken over by 0_2. Formula (7.11) also holds for y in region II but now with z as the point where the line $0_1 y$ meets λ. After all, take H_d' as the line λ and H_d as the line through 0_1 parallel to λ. If y is in region III we have

$$U_S(y) = \overset{\frown}{0_4 y}/\overset{\frown}{0_4 z}, \tag{7.12}$$

with z as the point where $0_4 y$ meets l. After all, take H_d as the line l and H_d' a the line through 0_4 parallel to l. Finally, $U_S(y) = 1$ if $y \varepsilon W_S$.

(d)″ Let us now consider the bounds $L_S(y)$ and $U_S(y)$ in the special case that S and S' are entirely on one side of T, that is, $a = \alpha$ or $b = \beta$. In fact, these cases

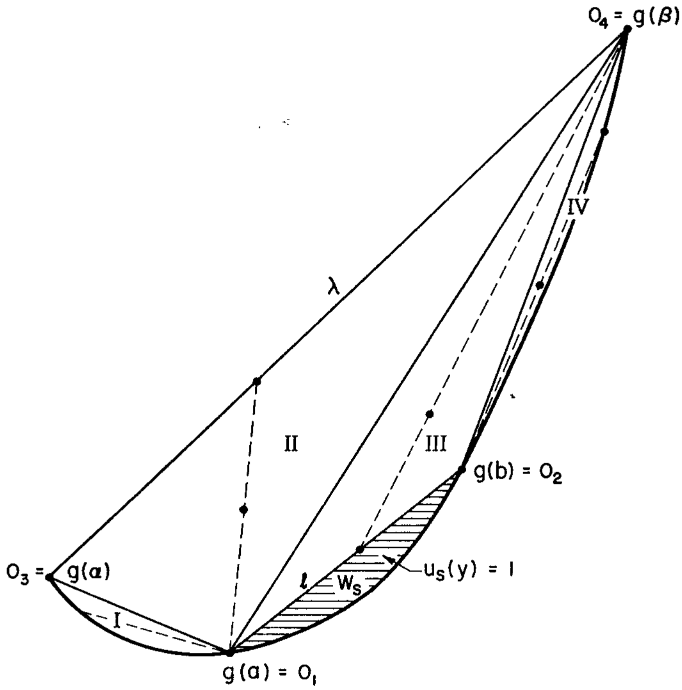

FIG. 3

(and certain more general cases) have already been treated in detail by Markov [8] and others [1], [5], [7], [9] and [17]. The geometric approach of von Mises [9] is most closely related to ours; in his approach, the moment vector $y = (y_1, y_2)$ is kept fixed while $S = [\alpha, b]$ varies.

The present one-sided case can be read off from Figure 3 as follows. In the first place, one is allowed to take $a = \alpha$ (in which case there are only the three regions III, IV and W_S). In this manner, one obtains $U_S(y)$ when $S = [\alpha, b]$, hence, $L_S(y)$ when $S = (b, \beta]$.

One is not allowed to take immediately $b = \beta$ since we assumed that λ and l intersect on the left. However, (interchanging the roles of a and b, and the roles of α and β), the situation is completely analogous. In the case $b = \beta$ one has the three regions I, II and W_S. One obtains $U_S(y)$ when $S = [a, \beta]$ (hence $L_S(y)$ when $S = [\alpha, a)$, $\alpha < a < \beta$) by projecting y from 0_1 onto $g(T)$ when y is in region I; from 0_3 onto l when y is in region II, (not from 0_1 onto λ).

(e) Consider now the somewhat different case where $T = [\alpha, \beta)$ is open on the right, $(-\infty < \alpha < \beta \leqq +\infty, \beta = +\infty$ in most applications). We further assume that near β the behavior of the g_j is such that

$$(7.13) \qquad g_2(t) \to +\infty \quad \text{and} \quad g_1(t)/g_2(t) \uparrow 0 \quad \text{as} \quad t \uparrow \beta.$$

This in turn implies that g_1 is strictly increasing.

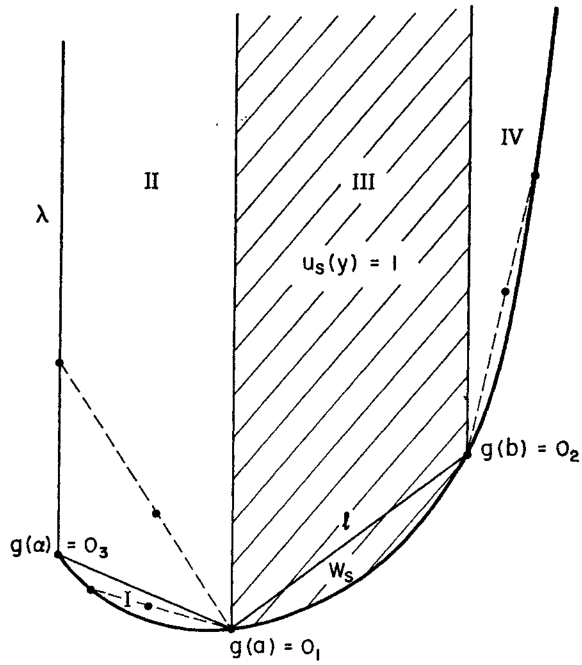

FIG. 4

The computation of $U_S(y)$ for this case is described in Figure 4. Observe that Figure 4 is obtained from Figure 3 by letting $0_4 = g(\beta)$ recede to infinity in such a way that all straight lines through 0_4 become vertical lines. That the corresponding limiting procedure, (for computing $U_S(y)$), is still correct follows from the discussion at the end of Section 6.

In region I of Figure 4 one again has formula (7.11), obtained by projecting y from 0_1 onto $g(T)$. An analogous formula holds in region IV (where we project y from 0_2 onto $g(T)$). For y in region II the value $U_S(y)$ is determined by the position of the point y relative to the two vertical boundary lines (through 0_1 and 0_3). That is, (6.4) holds with $\Delta(y)$ as the distance to λ (which acts as H_d') and Δ as the distance between the vertical boundary lines. Finally, $U_S(y) = 1$ either when y is in region III or when $y \in W_S$.

In Figure 4, it is allowed to take $a = \alpha$ or $b = \beta$. If $a = \alpha$ then $U_S(y) < 1$ *only* when y is in region IV. This case may be reformulated as follows.

Let X be a random variable taking values in $[\alpha, \beta)$, usually $\beta = +\infty$, and satisfying (7.2). We assume that $\{g_0, g_1, g_2\}$ is a Tchebycheff system on $[\alpha, \beta)$ such that (7.13) holds. Then for each number b satisfying $\alpha < b < \beta$ and $g_1(b) < y_1$ we have the *sharp* upper bound

$$P(X \leqq b) \leqq (z_1 - y_1)/(z_1 - g_1(b)), \tag{7.14}$$

where z_1 denotes the first coordinate of the point z where the line from $0_2 = g(b)$ to y intersects the curve $g(T)$.

As a special case, we have the well-known inequality

$$P(X \leqq -c) \leqq z_1/[z_1 - (-c)] = \sigma^2/(\sigma^2 + c^2) \text{ with } z_1 = \sigma^2/c,$$

whenever $E(X) = 0$, $E(X^2) = \sigma^2$ and $c > 0$.

(f) Let us now drop the assumption (7.13) and go back to the situation of a compact interval $T = [\alpha, \beta]$ such that each polynomial (7.10) is continuous with at most two zeros in T. Let further $S = [a, b]$ be a subinterval of T. In (d), we already found the upper bound $U_S(y)$; moreover, for the special cases $a = \alpha$ or $b = \beta$, also the lower bound $L_S(y)$.

It remains to determine $L_S(y)$ for the case $\alpha < a < b < \beta$. The computation of $L_S(y)$ in this case is completely described by Figure 5. Here, 0_5 denotes the unique point on $g(T)$ through which there passes a straight line parallel to l which supports $g(T)$ from below.

If the point y is in region I ($y \varepsilon \operatorname{int}(V)$) then

$$L_S(y) = \widehat{0_1 y}/\widehat{0_1 z}, \tag{7.15}$$

where z denotes the point where the line $0_1 y$ meets $g(T)$. This follows immediately from (6.3), taking there H_d as a straight line through z supporting $g(T)$ and taking H_d' as the line through 0_1 parallel to H_d. For, then H_d supports all of W while H_d' supports $W_{S'}$ (the shaded region in Figure 5), each on the same side. If y is

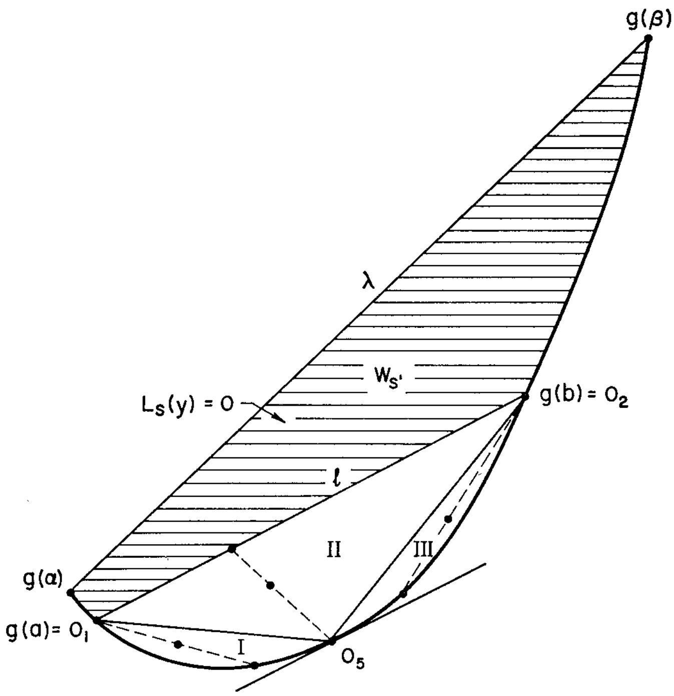

FIG. 5

in region III we have a formula analogous to (7.15) for $L_S(y)$; here, we project y onto $g(T)$ from the point $0_2 = g(b)$.

If y is in region II (the triangle $0_1 0_2 0_5$) we have

$$(7.16) \qquad L_S(y) = \widehat{yz}/\widehat{0_5 z},$$

where z denotes the point where the lines $0_5 y$ and l intersect. After all, take H_d as the tangent line at 0_5 (so that $A_d = \{0_5\}$) and take H_d' as the line l (so that B_d is equal to the segment $0_1 0_2$). Finally, $L_S(y) = 0$ when y is above l, that is, $y \varepsilon W_{S'}$.

As is easily seen directly (or from the limiting procedure at the end of Section 6) the above method for computing $L_S(y)$ remains valid if T is open on one or both ends, possibly $\alpha = -\infty$ or $\beta = +\infty$, as long as $\alpha < a < b < \beta$. In fact, $L_S(y)$ *does not even depend on* α *or* β as long as $y \varepsilon \operatorname{int}(V)$.

It is easy to read off from Figure 5 certain measures $\mu \varepsilon M^+(T)$ with $\mu(g) = y$ and such that $L_S(y)$ is attained; we assume here that $S = (a, b)$ is open at both ends so that $a, b \varepsilon S'$. When y is in region II there is such a measure which is carried by the 3-point set $\{a, b, t_5\}$ where t_5 is defined by $0_5 = g(t_5)$. If y is in region I there is such a measure carried by the 2-point set $\{a, t\}$ with $t \varepsilon T$ such that the three points $g(a)$, y and $g(t)$ are on a straight line. Similarly, for y in region III.

Let us consider in some more detail the "classical" case

$$g_1(t) = t, \qquad g_2(t) = t^2, \qquad T = (-\infty, +\infty).$$

The resulting lower bound $L_S(y)$ determined below is due to Selberg [16]. In terms of random variables, we have a real-valued random variable X with $E(X) = y_1$ and $E(X^2) = y_2$. We are interested in the best lower bound $L_S(y)$ on $\Pr(a < X < b)$. We may as well assume that $E(X) = 0$, $E(X^2) = \sigma^2 > 0$, $a < 0 < b$. Put $-a = c$ and assume for convenience that $c \leqq b$, (the case $c > b$ being analogous). Therefore, we are now interested in the best lower bound in

$$(7.17) \qquad P(-c < X < b) \geqq L_S(y), \quad \text{where} \quad y = (0, \sigma^2),$$

when it is given that $E(X) = 0$, $E(X^2) = \sigma^2 > 0$. Further, $b \geqq c > 0$.

The curve $g(T)$ is presently the parabola $y_2 = y_1^2$. The line l has its slope equal to $(b^2 - c^2)/(b + c) = b - c$, hence, $t_5 = (b - c)/2$. The line l intersects the y_2-axis in the point $(0, bc)$, hence, $L_S(0, \sigma^2) = 0$ when $\sigma^2 \geqq bc$ (so that we are in $W_{S'}$). The line $0_1 0_5$ intersects the y_2-axis in the point $(0, t_5 c)$. Hence, if $\sigma^2 \leqq t_5 c$ we are in region I. Using (7.15), (where $z_1 = \sigma^2/c$), we obtain

$$L_S(0, \sigma^2) = c/(c + \sigma^2/c) = c^2/(c^2 + \sigma^2) \quad \text{if} \quad \sigma^2 \leqq t_5 c.$$

It remains to consider the case that $t_5 c < \sigma^2 < bc$, in which case the point $(0, \sigma^2)$ is in region II of Figure 5. The tangent line at 0_5 intersects the y_2-axis at the point $(0, -t_5^2)$. Using (7.16), we find that

$$L_S(0, \sigma^2) = (bc - \sigma^2)/(bc + t_5^2) = (bc - \sigma^2)(\tfrac{1}{2}(b + c))^{-2},$$

whenever $t_5 < \sigma^2/c < b$.

REFERENCES

[1] Ahiezer, N. I. and Krein, M. (1962). *Some Questions in the Theory of Moments*. Translations of Mathematical Monographs **2**. Amer. Math. Soc., Providence.

[2] Bonnesen, T. and Fenchel, W. (1934). *Theorie der konvexen Körper*. Springer-Verlag, Berlin.

[3] Fan, Ky (1959). *Convex Sets and Their Applications*. Argonne National Laboratory.

[4] Karlin, S. and Shapley, L. S. (1953). *Geometry of Moment Spaces*. Memoirs **12** Amer. Math. Soc., Providence.

[5] Karlin, S. J. and Studden, W. J. (1966). *Tchebycheff Systems: with Applications in Analysis and Statistics*. Interscience, New York.

[6] Kemperman, J. H. B. (1965). On the sharpness of Tchebycheff type inequalities. *Indag. Math.* **27** 554–571; 572–587; 588–601.

[7] Krein, M. G. (1951). The ideas of P. L. Cebysev and A. A. Markov in the theory of limiting values of integrals and their further development. *Uspehi Mat. Nauk N. S.* **6** 3–120.

[8] Markov, A. (1884). On certain applications of algebraic continued fractions. Thesis St. Petersburg.

[9] Mises, R. von (1939). The limits of a distribution function if two expected values are given. *Ann. Math. Statist.* **10** 99–104.

[10] Mulholland, H. P. and Rogers, C. A. (1958). Representation theorems for distribution functions. *Proc. London Math. Soc.* **8** 177–223.

[11] Reidemeister, K. (1921). Über die singulären Randpunkte eines konvexen Körpers. *Math. Ann.* **83** 116–118.

[12] Richter, H. (1957). Parameterfreie Abschätzung und Realisierung von Erwartungswerten. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik* **3** 147–161.

[13] Riesz, F. (1911). Sur certaines systèmes singuliers d'équations intégrales. *Ann. Sci. Ecole Norm. Sup.* **28** 33–62.

[14] Rogosinsky, W. W. (1958). Moments of non-negative mass. *Proc. Roy. Soc. London Ser. A* **245** 1–27.

[15] Rogosinsky, W. W. (1962). Non-negative linear functionals, moment problems, and extremum problems in polynomial spaces. *Studies in mathematical analysis and related topics*. 316–324. Stanford Univ. Press.

[16] Selberg, H. L. (1940). Zwei Ungleichungen zur Ergänzung des Tchebycheffschen Lemmas, *Skand. Aktuarietidskrift* **23** 121–125.

[17] Shohat, J. A. and Tamarkin, J. D. (1943). *The Problem of Moments*. Math. Surveys **1** Amer. Math. Soc., Providence.

[18] Valentine, F. A. (1964). *Convex Sets*. McGraw-Hill, New York.