CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION CFSTI
DOCUMENT MANAGEMENT BRANCH 410.11

LIMITATIONS IN REPRODUCTION QUALITY

ACCESSION # 603980

☒ 1. WE REGRET THAT LEGIBILITY OF THIS DOCUMENT IS IN PART UNSATISFACTORY. REPRODUCTION HAS BEEN MADE FROM BEST AVAILABLE COPY.

☐ 2. A PORTION OF THE ORIGINAL DOCUMENT CONTAINS FINE DETAIL WHICH MAY MAKE READING OF PHOTOCOPY DIFFICULT.

☐ 3. THE ORIGINAL DOCUMENT CONTAINS COLOR, BUT DISTRIBUTION COPIES ARE AVAILABLE IN BLACK-AND-WHITE REPRODUCTION ONLY.

☐ 4. THE INITIAL DISTRIBUTION COPIES CONTAIN COLOR WHICH WILL BE SHOWN IN BLACK-AND-WHITE WHEN IT IS NECESSARY TO REPRINT.

☐ 5. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED, DOCUMENT WILL BE AVAILABLE IN MICROFICHE ONLY.

☐ 6. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED DOCUMENT WILL NOT BE AVAILABLE.

☐ 7. DOCUMENT IS AVAILABLE IN MICROFICHE ONLY.

☐ 8. DOCUMENT AVAILABLE ON LOAN FROM CFSTI (TT DOCUMENTS ONLY).

☐ 9.

NBS 9/64

PROCESSOR: PB

GEOMETRY OF MOMENT SPACES

S. Karlin and L. S. Shapley

P-227

23 July 1951

The RAND Corporation
SANTA MONICA · CALIFORNIA

INTRODUCTION AND SUMMARY

~~The purpose of~~ this paper ~~is to~~ presents a natural geometrical approach to the theory of reduced moment spaces and its application to orthogonal polynomials. Many classical results can be interpreted in this geometrical setting, and many new results obtained as well. The method provides an interesting contrast to the more usual techniques involving continued fractions and complex variables.

Some of this material was presented in outline in an earlier paper (reference [1] at the end of the paper), and two papers ([2], [15]) applying it to the theory of games have also appeared. The ~~present~~ work is concerned primarily with distribution functions on a finite interval. The half-infinite and infinite intervals will be taken up in future papers.

Chapter I is devoted to a preliminary exposition of the theory of convex sets and their duals in conjugate, finite-dimensional, linear spaces. Dimensional indices are introduced to describe the local structure of the boundary of a convex set, and are used to express a fundamental relationship between a convex set and its dual (Theorem 5.2). The moment spaces themselves are convex bodies whose points are n-tuples of moments of distribution functions; while the dual convex bodies are the coefficient spaces of n-th degree non-negative polynomials. In Chapter II these sets are introduced and their extreme points characterized. The structure of their boundaries is analyzed in detail, and a new result on the representation of non-negative polynomials as sums of square polynomials is obtained (Theorem 10.3). Chapter III introduces certain convex polyhedra which approximate the moment spaces and their duals, and uses them to establish

some of the classical theorems. In Chapter IV an algebraic description by means of "Hankel" determinants is given of the boundary components previously characterized by the dimensional indices. Chapter V deals with the distribution functions associated with a given point in the moment space, and the convex set in function space which they form. The interpretation of the supporting hyperplanes to the moment spaces as non-negative polynomials, first found in Chapter II, leads to a natural, geometric representation of orthogonal systems of polynomials, with arbitrary weight functions. Several applications of this approach are given in Chapter VI. The paper concludes with a chapter on the symmetries of the moment spaces.

General expositions of the moment problem may be found in Widder ([9], Ch. 3) and Shohat and Tamarkin ([12]). The latter contains an excellent bibliography of the subject.

CONTENTS

CHAPTER V: DISTRIBUTIONS HAVING GIVEN MOMENTS

CHAPTER VI: APPLICATION TO ORTHOGONAL POLYNOMIALS

CHAPTER VII: SYMMETRIES OF THE MOMENT SPACES

CHAPTER I

CONVEX SETS

§1. Elementary definitions.

We begin by setting forth some of the important properties of convex sets. A point set in n-dimensional Euclidean space E^n is *convex* if it contains the line segment joining every two of its points. The term *hyperplane* refers to an $(n-1)$-dimensional linear variety in E^n; each hyperplane determines two closed *half-spaces*. If a set K is contained in one of the closed half-spaces H_L of a hyperplane L, we say that L (or H_L) either *bounds* or *supports* K according as the lower bound of the distance from K to L is positive or zero, respectively. We state some basic relationship between convex sets and hyperplanes (see, for example, Weyl [3]):

THEOREM 1.1 (a) *A closed convex set is precisely the intersection of all its supporting half-spaces.*

(b) *Every boundary point of a convex set lies in some supporting hyperplane of the set.*

(c) *Two non-intersecting, closed, bounded convex sets can be separated: that is, disjoint closed half-spaces exist each containing one of the sets. In particular, an exterior point can be separated from any closed convex set by a supporting hyperplane of the set.*

The *dimension* of a convex set is defined to be that of the smallest linear variety containing the set. A point of an n-dimensional convex

set K that is interior with respect to the m-dimensional variety containing K we shall call an inner point[1]; while the terms "interior" and "boundary" will consistently refer to the full n-dimensional structure. Thus, the notions of "inner" and "interior" coincide only for n-dimensional convex sets. We have:

THEOREM 1.2. *Every (non-empty) convex set has an inner point. The point x is an inner point of K if and only if every supporting hyperplane at x contains all of K.*

A point x of a convex set K which is not an inner point of any convex subset of K, except the set consisting of x alone, is an *extreme point* of K.

A *convex body* in E^n is a convex set which is closed, bounded, and n-dimensional.

§2. *Boundary indices*.

For x in the boundary of a closed, convex set K in E^n we define L(x) to be the intersection of all hyperplanes that support K at x. For formal completeness we may also define:

$$L(x) = E^n \qquad \text{if } x \text{ is interior to } K,$$
$$L(x) = 0 \qquad \text{if } x \text{ is not in } K.$$

The *contact set* C(x) of a point x is defined to be the inter-

[1] Sometimes the term "relative interior" is used.

section of $L(x)$ with K. We call a point of K ordinary, or exceptional, according as it is, or is not, an inner point of its own contact set.

Since $C(x)$ is itself closed and convex, we can construct the contact set of x with respect to $C(x)$, instead of K. If x is ordinary (with respect to K), Theorem 1.2 tells us that the new contact set is just $C(x)$ again. If, on the other hand, x is exceptional with respect to K, the new contact set is a lower-dimensional subset of $C(x)$. Iteration of this process leads ultimately to the reduced contact set $C'(x)$, which is the largest convex subset of K of which x is an inner point. $C'(x) = x$ if and only if x is an extreme point of K.

We now define three indices[2]

facial dimension of x:	$a(x) = \dim L(x)$,
contact dimension of x:	$c(x) = \dim C(x)$,
reduced contact dimension of x:	$c'(x) = \dim C'(x)$,

all, of course, with respect to a given closed convex set K. All three are equal to n if x is interior to K, and equal to -1 (the dimension of the null set) if x is outside of K. These indices are invariant under affine and (proper) projective transformations of K, and are intrinsic in the sense that we do not alter them by increasing the dimension of the carrying space E^n. For a finite, convex polytope $a(x) = c(x)$ everywhere; in general the difference $a(x) - c(x)$

[2] Compare Bonnesen and Fenchel [4], pp. 14-16. Their "p-Kantenpunkte" are the points x with $a(x) = n - p - 1$.

describes qualitatively the curvature of the boundary of K at x.

THEOREM 2.1 *If x is exceptional, then a sequence of ordinary points $x^{(i)}$ can be found which converge to x and for which*

$$a(x^{(i)}) = a(x), \qquad e(x^{(i)}) = e(x).$$

Proof. Take any sequence of inner points of $C(x)$ approaching x.

A fourth index $d(x)$ can be defined by examining more explicitly the situation of x in relation to the neighboring boundary points of K. Consider all the supporting hyperplanes that touch K within an ε-neighborhood of x, and displace them parallel to themselves so that they all pass through x. Call the intersection of the resulting set of hyperplanes $D(x, \varepsilon)$. Then let

$$d(x) = \lim_{\varepsilon \to 0} \dim D(x, \varepsilon).$$

This index is invariant and intrinsic, like the others, and is equal to n in the interior of K, zero at the extreme points.

THEOREM 2.2 *For x in the n-dimensional closed convex set K,*

$$n \geq a(x) \geq e(x) \geq e'(x) \geq d(x) \geq 0.$$

Proof. The only one that is not immediately obvious is the inequality $c'(x) \geq d(x)$. Put $m = n - c'(x)$ and let E^m denote the linear variety through x perpendicular to $C'(x)$. The point x is an extreme in the convex set

$$K' = K \cap E^m;$$

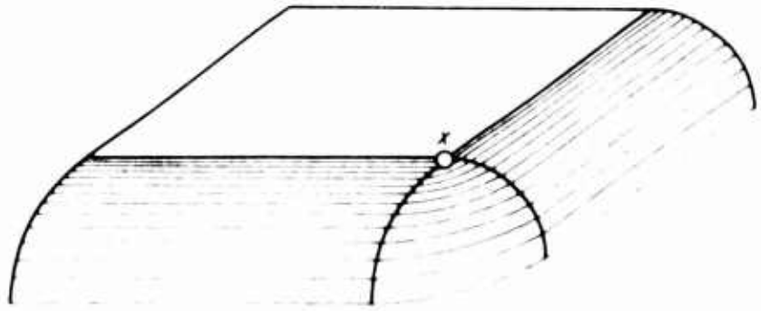
and the set $D_{(m)}(x, \epsilon)$ (relative to E^m) is just the point x alone. Since every supporting hyperplane to K' in E^m can be extended to at least one supporting hyperplane to K in E^n, we have

$$D(x, \epsilon) \cap E^m \subseteq D_{(m)}(x, \epsilon) = x.$$

We conclude that the dimension of $D(x, \epsilon)$ does not exceed $n - m = c'(x)$.

It can be shown by examples that the inequalities of the last theorem cannot be improved on.

The accompanying sketches illustrating two of the more complicated possibilities for the indices at a point.



$n = 3$

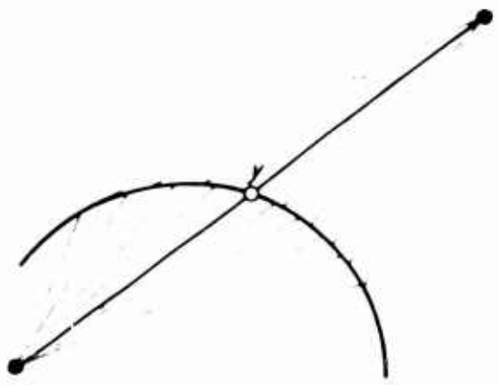
$a(x) = 2$

$c(x) = 2$

$c'(x) = 0$

$d(x) = 0$

Fig. 2.1



$n = 3$

$a(y) = 2$

$c(y) = 1$

$c'(y) = 1$

$d(y) = 0$

Fig. 2.2

In our work we shall find $a(x)$ and $c(x)$ to be the most important of the indices. They play a prominent role in the theory of conjugate points (§ 5), which will prove to be a useful tool when we come to consider the moment spaces themselves in Chapter II.

§3. Convex representation.

We shall say that a point x in E^n is spanned by a set of points $x^{(1)}, \ldots, x^{(p)}$ if there are non-negative quantities $\xi_1, \ldots, \xi_p$ with

$$x = \sum \xi_i x^{(i)}, \quad \sum \xi_i = 1. \tag{3.1}$$

The following theorem is a consequence of Theorem 1.1 c and the definition of extreme point.

THEOREM 3.1 *A closed bounded convex set in E^n is spanned by its extreme points, and every spanning set contains the extreme points.*

With respect to a particular convex set K we now define $b(x)$ to be the least number of extreme points of K required to span x.

THEOREM 3.2 For points in a closed bounded m-dimensional convex set,

$$b(x) \leq m + 1,$$

and hence, also,

$$b(x) \leq \sigma(x) + 1.$$

Proof. See Steinitz [5] §10; also Hanner and Rådström [6].

§4. Convex cones. Duality.

We now find it convenient to introduce the homogeneous coordinates

$$x = (x_0, x_1, \ldots, x_n)$$

for the points x in E^n; and we impose on them the linear normalizing condition:

$$\sum_{i=0}^{n} \tilde{y}_i x_i = 1, \qquad \text{not all } y_i = 0,$$

which may be written, in short form:

$$\tilde{y} \cdot x = 1, \qquad \tilde{y} \neq 0. \tag{4.1}$$

A half-space in E^n is now characterized by (4.1) and an inequality of the form

(4.2) $$y \cdot x \geq 0, \quad y \neq 0,$$

in which y may be regarded as a point in the conjugate space $(E^{n+1})^*$ of homogeneous linear functions on E^{n+1}. The half-space determines y up to a positive multiple.

Temporarily putting aside the normalizing condition (4.1) we consider x as a point in E^{n+1}. A cone in E^{n+1} is a set that contains λx whenever it contains x, for all $\lambda \geq 0$. All supporting half-spaces to a convex cone Γ will have the point 0 in their boundaries, and hence can be represented in the form (4.2). The set of all y in $(E^{n+1})^*$ representing supporting half-spaces to Γ is also a convex cone if we include the point $y = 0$; we call this the conjugate cone to Γ and denote it by Γ^*. Thus

$$y \in \Gamma^* \iff y \cdot x \geq 0 \qquad \text{all } x \in \Gamma .$$

THEOREM 4.1 *If Γ is a closed convex cone in E^{n+1} (and if we identify the spaces $(E^{n+1})^{**}$ and E^{n+1} in the natural way), then $\Gamma^{**} = \Gamma$.*

Proof that $\Gamma \subseteq \Gamma^{**}$. Take x in Γ. Every y in Γ^* satisfies $y \cdot x \geq 0$. Therefore x, regarded as a point in $(E^{n+1})^{**}$, represents a supporting half-space to Γ^*, or is the point 0. In either case x belongs to Γ^{**}.

Proof that $\Gamma^{**} \subseteq \Gamma$. Take x in $E^{n+1} - \Gamma$. By Theorem 1.1c, some supporting half-space of Γ does not contain x. In other words, $y \cdot x < 0$ for some y in Γ^*. It follows that the half-space in $(E^{n+1})^*$ which x represents does not support Γ^*. Thus x does not belong to Γ^{**}.

We call a closed convex cone Γ *proper* if it does not contain any complete

line; that is, if x and $-x$ are not both in Γ unless $x = 0$. It is not difficult to see that a closed cone Γ is proper if and only if Γ^* has an interior point, for an interior point of Γ^* represents a supporting half-space whose boundary meets Γ only at the origin.

If y is any inner point of Γ^*, we define the cross-section $\chi(\Gamma; y)$ of Γ by

$$x \in \chi(\Gamma; y) \iff x \in \Gamma \quad \text{and} \quad yx = 1.$$

The cross-section is bounded if and only if Γ is proper. Conversely, given any convex set K in E^n we define its cone $\Gamma(K)$ in E^{n+1} by

$$x \in \Gamma(K) \iff \lambda x \in K \quad \text{for some} \quad \lambda \geq 0.$$

Clearly

$$\Gamma(\chi(\Delta; y)) = \Delta$$

holds for any proper cone Δ, provided of course that y is interior to Δ^*. On the other hand, the set

$$\chi(\Gamma(K); y)$$

is a projective image of the set K in E^n, and is in general the same as K only in case $y = \tilde{y}$.

If K is a closed convex set in E^n, we shall refer to the set

$$K^* = \chi([\Gamma(K)]^*; x) \qquad (x \text{ inner to } K)$$

as a dual set to K. K^* is a convex body if and only if K is; and, by the last theorem, K is dual to K^* as well.

The next two theorems follow directly from our definitions.

THEOREM 4.2 *The point x is interior to the convex set K if and only if* $y \cdot x > 0$ *for all points y in* K^*. *There is a one-one correspondence between the boundary points of K and the supporting half-spaces to* K^*, *and conversely.*

THEOREM 4.3 *If K and L are closed convex sets with* $L \subset K$, *and if* K^* *and* L^* *are normalized by the same point x inner to L, then* $K^* \subset L^*$.

In the figure a pair of dual convex sets in E^2 and E^{2*} are are shown, with some of the supporting half-spaces.

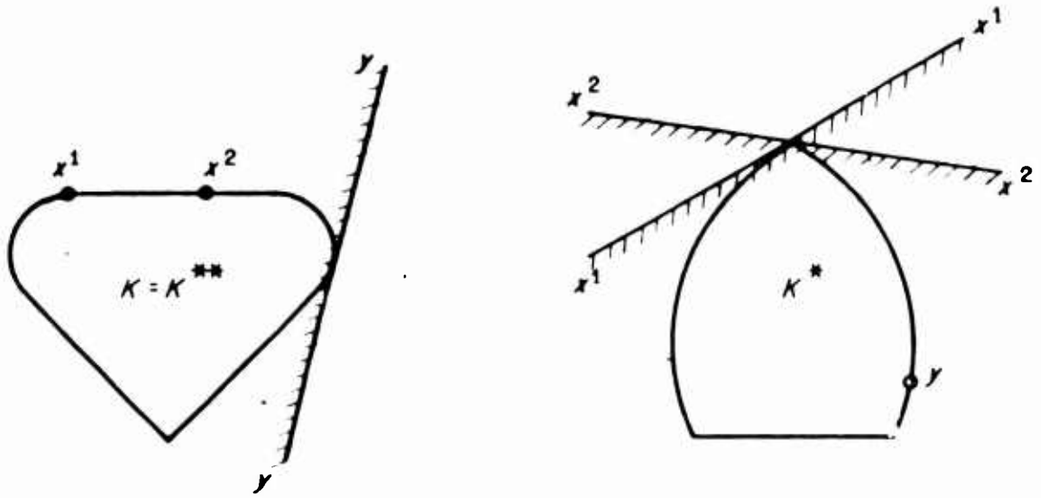


Fig. 4.1

Note. Our dual convex body is closely related to the "polar body" defined by Minkowski[3]. In fact, if one takes

$$\tilde{y} = (1, 0, \ldots, 0)$$

in (4.1), and takes K to be a convex body in E^n having the point

$$x = (1, 0, \ldots, 0)$$

in its interior, then the dual body

$$K^* = \chi([\Gamma(K)]^*; x)$$

is the polar reciprocal of K with respect to x, reflected in [illegible].

§5. Conjugate points.

The way in which the boundaries of dual convex bodies K and K^* depend upon one another is, in a certain sense, a local relationship. For example, when K and K^* are finite polyhedral bodies we can [illegible] to associate with every p-cell in the boundary of K a particular (n-p-1)-cell in the boundary of K^* — its dual cell in the familiar combinatorial theory. This duality we now extend to general convex bodies (Theorem 5.2), making use of the indices defined in §2.

Let K be a fixed convex body in E^n. If y is in the boundary of K^* we may regard y as a supporting hyperplane to K, and write $y \cap K$ to denote the set in which the hyperplane meets K. More formally:

---

3 See [7], p. 146.

(5.1) $\qquad x \in y \cap K \iff x \in K \quad \text{and} \quad y \cdot x = 0.$

Of course, $y \cap K$ is contained in the boundary of K, and is convex.

If $y \cap K = C(x)$ (see § 2) for boundary points y and x of K^* and K, respectively, then we say that y is conjugate to x. This definition is actually symmetric, as the next theorem reveals. Examples of conjugate pairs are readily found in Figure 4.1.

THEOREM 5.1 (a) If y is conjugate to x, then x is conjugate to y.

(b) Every boundary point of K is conjugate to at least one boundary point of K^*.

Proof (a). We are given that

(5.2) $\qquad C(x) = y \cap K$

and wish to show that

$$C(y) = x \cap K^*.$$

We see at once that

$$C(y) \subseteq x \cap K^*.$$

since $C(y)$ is defined as the intersection of K^* with all of the supporting hyperplanes at y. To complete the proof, we observe that any point y' in $K^* - C(y)$ has the property $x' \cdot y' \neq 0$ for some x' in K^{**} with $x' \cdot y = 0$. By (5.1) x' is in $y \cap K$; by (5.2) x' is in $C(x)$. This means that any hyperplane supporting K at x must

also support K at x', or

$$y''\cdot x = 0 \quad \text{only if} \quad y''\cdot x' = 0, \quad (\text{all } y'' \in K^*).$$

In particular, we obtain $y'\cdot x \neq 0$, from which it follows that y' cannot be in $x\circ K^*$. This completes the proof.

Proof (b). The set $x\circ K^*$ is non-empty and convex for every x in the boundary of K (Theorem 4.2). If y is any inner point of this set (Theorem 1.2), then it is easily seen that $C(y) = x\circ K^*$. Hence x is conjugate to y.

THEOREM 5.2 *If x and y are conjugate boundary points of dual (n-dimensional) convex bodies, then*

$$(5.3) \qquad \begin{cases} c(x) + a(y) = n - 1, \\ a(x) + c(y) = n - 1. \end{cases}$$

Proof. When interpreted in the conjugate space $(E^n)^*$ containing the dual body K^*, the set $y\circ K$ is seen to consist of all the hyperplanes supporting K^* at y. The dimensionality of this set is $c(x)$, by (5.2). We may therefore count exactly $c(x) + 1$ linearly independent hyperplanes in $(E^n)^*$: they intersect in a set of dimension $n - c(x) - 1$. But this set is precisely the set $L(y)$ (see §2), and its dimension is therefore $a(y)$. This proves the first assertion of the theorem. The second follows immediately by symmetry (Theorem 5.1a).

Conjugation does not, of course, give a one-one pointwise correspondence between the boundaries of K and K^*. However, a one-one correspondence between sets of boundary points can be set up very naturally. Let $(x)^*$ denote the set of points in K^* which are conjugate to x; and, if S is any set of points in the boundary of K, let S^* denote the set of all points in K^* conjugate to some point in S:

$$S^* = \bigcup_{x \in S} (x)^*.$$

By Theorem 5.1b, S^* is not empty. We have

$$S^{***} = S^*. \tag{5.4}$$

This is a type of idempotence relation frequently encountered in dealing with linear spaces and their conjugates—it says that after two applications conjugation gives nothing new. To prove (5.4) we must show that

$$\text{if} \quad \begin{cases} y & \text{is conjugate to } x, \\ x' & \text{is conjugate to } y, \\ y' & \text{is conjugate to } x', \end{cases} \tag{5.5}$$

$$\text{then} \quad y' \quad \text{is conjugate to } x.$$

But the definition (with Theorem 5.1a) gives us

$$C(x) \cap K = C(x') \cap K,$$

from which (5.5) and (5.4) follow at once.

THEOREM 5.3 The sets $(y)^*$, as y ranges over the boundary of K^*, constitute a partition of the boundary of K into disjoint, convex components. These components correspond biuniquely with the components $(x)^*$ of the analogously defined dual partition of the boundary of K^*, in such a way that every point of any one component is conjugate to every point of its dual.

Proof. By (5.4) or (5.5).

Within each component of the partition, $a(x)$ and $c(x)$ are constant. In the case of polyhedral bodies, the components are just the usual cells. In general, they are the (relative) interiors of the maximal convex sets lying in the boundary.

The partition of Theorem 5.3 is the finest partition which admits a dual. For it is easy to see that $S = S^{**}$ if and only if S is the union of elementary sets $(y)^*$. Later (in § 11) we shall consider as "faces" of a convex body the maximal connected sets over which $a(x)$ is constant. The dual faces on the dual body are the maximal connected sets having $c(y)$ constant, as a short argument based on Theorem 5.2 reveals.

CHAPTER II

THE MOMENT PROBLEM D^n AND ITS DUAL P^n

§1. Distribution functions on $[0,1]$.

The primary subject matter of this paper will be the normalized distribution functions on the closed unit interval $[0,1]$ and their moments. The choice of interval is a matter of convenience -- in many respects the interval $[-1,1]$ would serve as well or better. All of our results can be adapted with little difficulty to an arbitrary finite interval, but many features of the half-infinite and infinite cases are conspicuously different. We propose to deal with those cases in future papers.

Since we are not interested in distributions over general spaces, we can omit the customary set-functional approach. We define a distribution function directly as a real-valued function which is monotonic, continuous to the right, and flat outside of $[0,1]$:

(i) $\phi(t_2) \geq \phi(t_1)$, if $t_2 > t_1$.

(ii) $\phi(t) = \phi(t + 0)$, for all t.

(iii) $\phi(t) = \phi(1)$ if $t \geq 1$,
$\phi(t) = 0$ if $t < 0$.

The effect of (ii) and the last clause of (iii) is to remove redundant representatives of "substantially equal" distribution -- distributions that operate identically on all continuous functions. (The condition

(ii') $$\phi(t) = \frac{1}{2}\left[\phi(t+0) + \phi(t-0)\right]$$

is sometimes found in place of (ii).)

The <u>spectrum</u> of ϕ, denoted $\mathfrak{S}(\phi)$, is the set of points t in $[0,1]$ with

$$\phi(t+\varepsilon) - \phi(t-\varepsilon) > 0$$

for every positive ε. If $\mathfrak{S}(\phi)$ is a finite set then we say that ϕ is an <u>arithmetic</u> distribution function, or a finite <u>step-function</u>.

To normalize, we put

(iv) $$\phi(1) = 1.$$

We let $\mathfrak{D}$ denote the class of normalized distribution functions and $\mathfrak{D}_A$ the class of normalized arithmetic distribution functions.

The step-functions

$$\phi(t) = I(t - t_0) = \begin{cases} 0 & \text{for } t - t_0 < 0, \\ 1 & \text{for } t - t_0 \geq 0, \end{cases}$$

with $0 \leq t_0 \leq 1$, we call the <u>pure</u> distribution functions; they are the extreme points of $\mathfrak{D}_A$. Every arithmetic ϕ can be represented uniquely:

(6.1) $$\phi(t) = \sum_{j=1}^{b(\phi)} \alpha_j I(t - t_j) \qquad \alpha_j > 0, \quad t_j > t_{j-1}$$

as a linear combination of pure distribution functions, and, obviously, ϕ is in $\mathcal{F}_A$ if and only if

$$\sum_{j=1}^{b(\phi)} \xi_j = 1.$$

(The use of $b(\phi)$ to denote the number of steps is consonant with the definition of §3.) In the "weak *" topology the pure distribution functions span out $\mathcal{F}$ as well as $\mathcal{F}_A$. We shall return to this question in §21.

It is convenient to introduce an index $b'(\phi)$ of $\phi \in \mathcal{F}_A$:

$$b'(\phi) = \begin{cases} b(\phi) & \text{if no step at either } t = 0, 1, \\ b(\phi) - \frac{1}{2} & \text{if step at just one of } t = 0, 1, \\ b(\phi) - 1 & \text{if steps at both } t = 0, 1, \end{cases}$$

which Wald calls the <u>degree</u>[4] of ϕ. Thus, steps at the end-points of $[0,1]$ contribute the amount $\frac{1}{2}$ to the degree; interior steps the amount 1.

Finally, we define the <u>n^{th} moment</u> of ϕ:

$$\mu_n(\phi) = \int_0^1 t^n d\phi(t), \qquad n = 0, 1, 2, \ldots . \tag{6.2}$$

The limits of integration would be more properly written

$$\int_{-0}^{1} \quad \text{or even} \quad \int_{-\infty}^{\infty}$$

since we intend, for $\phi \in \mathcal{D}$, to have always $\mu_0(\phi) = 1$. However, we shall use the less complicated notation of (6.2), with the understanding that the variation of ϕ at $t=0$, if any, is to be included in the integration.

[4] See Wald [8].

§7. The moment space D^n.

The n^{th} **moment space** D^n we define to be the set of points

$$x = (x_1, \ldots, x_n)$$

in E^n whose coordinates are the moments $\alpha_1(\phi), \ldots, \alpha_n(\phi)$ of at least one ϕ in $\mathcal{S}$.

THEOREM 7.1 (a) *There is a function in* $\mathcal{S}$ *having the moments*

$$x_1, x_2, \ldots, x_k, \ldots \tag{7.1}$$

if and only if the point

$$x_{(k)} = (x_1, \ldots, x_k)$$

is in D^k, *for all* k.

(b) *No two functions in* $\mathcal{S}$ *have the same moment sequence* (7.1).

Proofs of these standard results may be found (for example) in Widder [9], pp. 29, 31, 60. We remark that for distributions on the half-infinite or infinite interval (b) is no longer true.

THEOREM 7.2 D^n *is a convex body.*

Proof. We must verify that D^n is convex, closed, bounded, and n-dimensional.

(a) Convex: D^n is convex in the obvious sense; hence D^n also is convex by the linearity of (6.2).

(b) Closed: See again Widder [9], pp. 29, 31.

(c) Bounded: We have, for every x in D^n,

$$0 \leq x_i \leq 1, \qquad i = 1, \ldots, n.$$

(d) n-dimensional: The $n+1$ pure distributions

$$I(t), \quad I(t-t_j) \qquad j = 1, \ldots, n,$$

for distinct, non-zero t_j in $(0,1]$, give rise to the $n+1$ points

$$\begin{aligned} x^{(0)} &= (0,, 0, \ldots, 0) \\ x^{(1)} &= (t_1, t_1^2, \ldots, t_1^n) \\ &\cdots \\ x^{(n)} &= (t_n, t_n^2, \ldots, t_n^n), \end{aligned}$$

in D^n. These do not all lie in any one hyperplane, since the determinant of the coordinates of $x^{(1)}, \ldots, x^{(n)}$ does not vanish. Thus D^n contains an n-dimensional simplex.

The point in D^n generated by the pure distribution $I(t-t_1)$ we shall designate by

$$x(t_1) = (t_1, t_1^2, \ldots, t_1^n). \tag{7.2}$$

We shall designate by C^n the curve traced out by $x(t_1)$ as t_1 runs between 0 and 1.

THEOREM 7.3 The set of extreme points of D^n, for $n \geq 2$, is precisely C^n.

Proof. We shall prove (a) that C^n spans D^n, and (b) that no point of C^n is spanned by other points of C^n. The theorem then follows with the aid of Theorem 3.4.

(a) Let D_A^n denote the subset of D^n generated by the arithmetic distribution functions $\phi_A \in \mathcal{O}_A$. It is clear from (6.1) that D_A^n is exactly the set of points spanned by C^n. Moreover, D_A^n is a closed set, since C^n is closed and bounded. We shall show that $D_A^n = D^n$. In fact, for any ϕ in $\mathcal{O}$ there is a sequence of step-functions $\phi_A^{(\nu)} \in \mathcal{O}_A$ such that

$$\lim_{\nu \to \infty} \int_0^1 f(t)\, d\phi_A^{(\nu)}(t) = \int_0^1 f(t)\, d\phi(t)$$

for every continuous function f; this is nothing more than the definition of the Stieltjes integral. Taking $f(t) = t, t^2, \ldots, t^n$, we see that D^n must be the closure of its subset D_A^n. But D_A^n is already closed; hence $D^n = D_A^n$, and C^n spans D^n.

(b) Consider a fixed $x(t_1)$ in C^n, and let H_{t_1} be the hyperplane defined by the equation

$$h(x) = t_1^2 - 2x_1 t_1 + x_2 = 0. \tag{7.3}$$

(For this we must have $n \geq 2$.) For a general point $x(t)$ of C^n we have

$$h(x(t)) = (t_1 - t)^2. \tag{7.4}$$

Thus our fixed point $x(t_1)$ is in H_{t_1}, while the rest of C^n lies in the positive open half-space determined by H_{t_1}. Obviously $x(t_1)$ is not spanned by other points of C^n.

§8. The polynomial space P^n.

We are now in a position to construct the dual to the convex body D^n. It is natural to introduce homogeneous coordinates (see §4) by means of the 0^{th} moment, $\mu_0(\phi)$. Since this is always equal to 1 by (iv) of §6, we append the coordinate $x_0 = 1$ to each point of D^n. We thus have

$$\tilde{y} = (1, 0, \ldots, 0)$$

in the normalizing relation (4.1). The cone $\Gamma(D^n)$ then can be interpreted as the moment space of the distributions defined by (i), (ii), and (iii) of §6.

The points y of the conjugate cone $\left[\Gamma(D^n)\right]^*$ are those satisfying

$$y \cdot x \geq 0 \qquad \text{(all } x \in D^n\text{)}.$$

By Theorem 7.3 we may replace D^n by C^n in the above condition, giving instead:

$$y \cdot x(t) \geq 0 \qquad \text{(all } t \in [0,1] \text{)};$$

or, by (7.2),

$$(8.1) \qquad P(t) = \sum_{i=0}^{n} y_i t^i \geq 0 \qquad \text{(all } t \in [0,1] \text{)}.$$

This reveals that $[\Gamma(D^n)]^*$ is the coefficient space of the polynomials of degree at most n, which are non-negative over the interval $[0,1]$.

It remains to select a particular cross-section of this cone to serve as a representative of the class of projectively equivalent dual convex bodies $(D^n)^*$. We select the <u>polynomial space</u> P^n:

$$P^n = \mathcal{H}([\Gamma(D^n)]^*; \tilde{x})$$

(see § 4), where

$$\tilde{x} = (1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots, \tfrac{1}{n+1}).$$

The coordinates of $\tilde{x}$ are the moments of the "rectangular" distribution $\tilde{\phi}(t) = t$; we shall presently see that, since $\tilde{\phi}$ is not a step-function, $\tilde{x}$ is interior to D^n for every n. As a result of this particular normalisation, the polynomials in P^n all satisfy

$$(8.2) \qquad \int_0^1 P(t)\,d\phi = 1.$$

This condition may be compared with the condition

$$\int_0^1 d\phi(t) = 1$$

which normalizes the moment cone $\Gamma(D^n)$. (A different [illegible]

Hereafter, we shall say simply "of a

above would have given us $P(t)d\bar{\gamma}(t)$ in place of the integrand of (8.2).)

Summarizing:

THEOREM 8.1 *The moment space* D^n *and the space* P^n *of* n^{th} *degree polynomials, non-negative over* $[0,1]$ *and normalized by* (8.2), *are dual convex bodies.*

To aid in visualizing this relationship we here show P^n and D^n for $n = 2$ and $n = 3$. As for higher dimensions, it is evident from the definitions that P^n is the cross-section $y_{n+1} = 0$ of P^{n+1}, while D^n is the perpendicular projection of D^{n+1} on the coordinate plane $x_{n+1} = 0$.

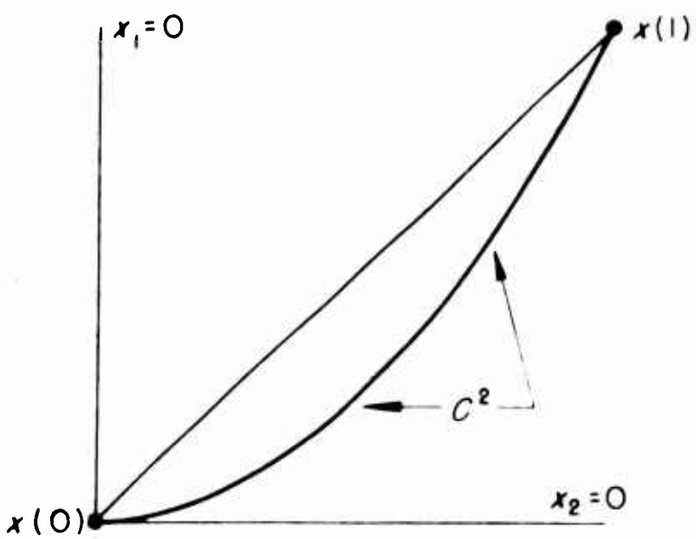


Fig. 8.1 — The moment space D^2

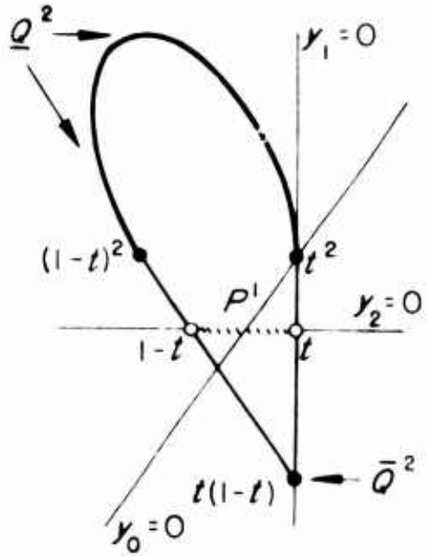


Fig. 8.2 — The polynomial space P^2

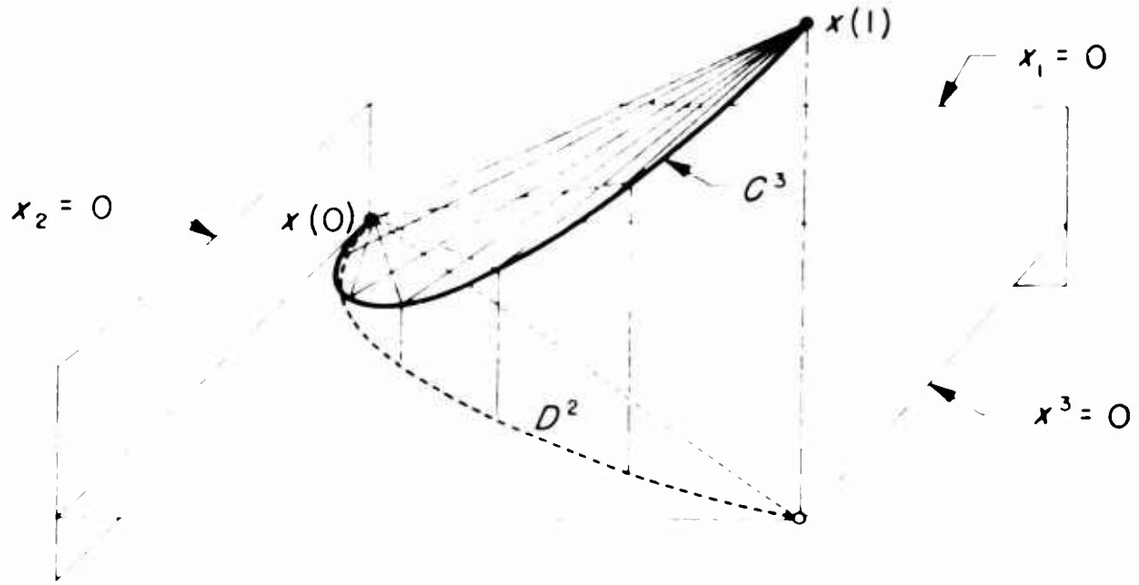


Fig. 8.3 — The moment space D^3

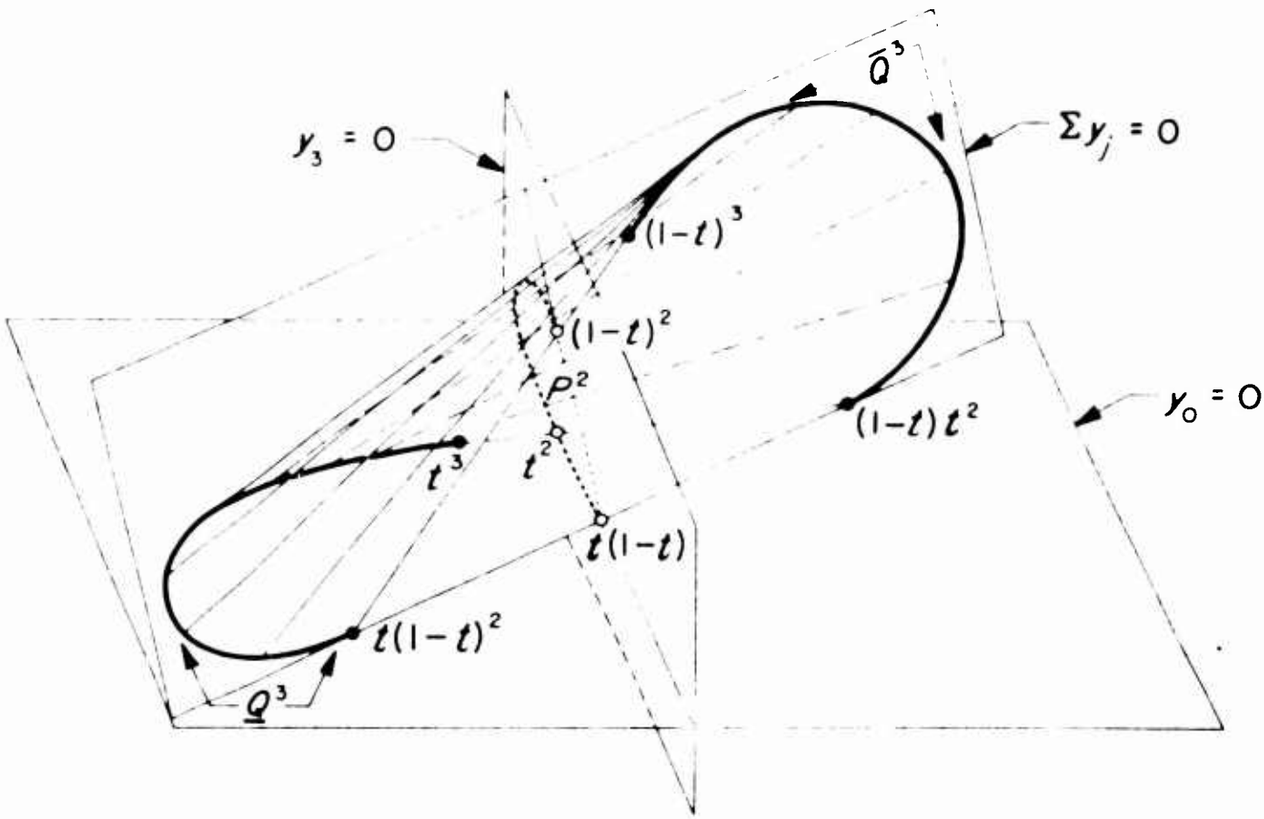


Fig. 8.4 — The polynomial space P^3

§9. The boundary of P^n.

THEOREM 9.1 *The boundary of P^n comprises just those polynomials of having one or more roots in $[0,1]$.*

Proof. (a) If $P(t)$ is strictly positive over $[0,1]$, it will remain so under small perturbations of its coefficients y_i. Considering just those perturbations that preserve the normalization (8.2), we see that the point y must be interior to P^n.

(b) Conversely, if $P(t)$ has a root t_1 in $[0,1]$, then

$$y \cdot x(t_1) = 0.$$

Thus y lies in a supporting hyperplane to P^n, by Theorem 8.1, and hence is in the boundary of P^n.

To help us discuss the boundary of P^n we define $r(y)$ to be the number of distinct roots in $[0,1]$ of the polynomial $\sum y_i t^i$. The symbol $r'(y)$ will also denote the number of distinct roots in $[0,1]$ but with the roots at $t = 0$ and $t = 1$, if they occur, counted only half.[6] Thus $r'(y)$ is not necessarily an integer, and

$$r(y) - 1 \leq r'(y) \leq r(y).$$

Since a non-negative polynomial must have all roots double in the interior of the interval, it follows that

$$2r'(y) \leq n$$

---

[6] Compare our definitions, in §6, of $b(f)$ and $b'(f)$.

for all y in P^n. We shall see that the equality holds only if y is an extreme point of P^n.

Let Q^n denote the closure of the subset of P^n for which $r'(y) = \frac{n}{2}$. Each polynomial in Q^n falls into one of two sets, $\underline{Q}^n$ or $\overline{Q}^n$, depending on whether the multiplicity of the root at $t = 1$ is even or odd, respectively. In $\underline{Q}^n$ polynomials have the form:

$$\underline{Q}(t) = y_n \prod_{j=1}^{n/2} (t - t_j)^2 \qquad (n \text{ even}),$$

$$= y_n t \prod_{j=1}^{(n-1)/2} (t - t_j)^2 \qquad (n \text{ odd});$$

and in $\overline{Q}^n$ they have the form:

$$\overline{Q}(t) = -y_n t(1-t) \prod_{j=1}^{(n-2)/2} (t - t_j)^2 \qquad (n \text{ even}),$$

$$= -y_n (1-t) \prod_{j=1}^{(n-1)/2} (t - t_j)^2 \qquad (n \text{ odd});$$

with all t_j in $[0,1]$.

Since we have taken the closure, the t_j are not necessarily distinct. It is easy to see that the two sets $\underline{Q}^n$ and $\overline{Q}^n$ are separated, for the leading coefficient y_n is always positive in the first case, negative in the second. On the other hand, they are both connected sets, each of them in fact being the homeomorphic image of a simplex defined in E^n by the inequalities

$$0 \le t_1 \le \dots \le t_m \le 1$$

(where m is roughly half of n).

THEOREM 9.2 *The set of extreme points of* P^n *is precisely* Q^n.

Proof. (a) If y is in $P^n - Q^n$ then y is not extreme. For the polynomial

$$P(t) = \sum y_i t^i$$

has either (i) a root $a < 0$; (ii) a root $b > 1$; or (iii) a pair of complex roots $c \pm id$. But in every case we can exhibit a convex representation of y in terms of other points of P^n. Thus:

$$(i) \qquad P(t) = (t-a)\,R(t) = \tfrac{1}{2}(t-2a)R(t) + \tfrac{1}{2}\,tR(t),$$

$$(ii) \qquad P(t) = (b-t)\,S(t) = \tfrac{1}{2}(2b-1-t)S(t) + \tfrac{1}{2}(1-t)S(t),$$

$$(iii) \qquad P(t) = \left[(t-c)^2 + d^2\right]T(t) = (t-c)^2T(t) + d^2T(t);$$

where R, S, and T are all non-negative over $[0,1]$, though not in general normalized to lie in P^n. However, all three of the right-hand expressions are of the form

$$\xi_1 P_1(t) + \xi_2 P_2(t),$$

where P_1 and P_2 are normalized, and ξ_1 and ξ_2 are positive with $\xi_1 + \xi_2 = 1$.

(b) Conversely, every y in Q^n must be extreme. For suppose that there is a convex representation

$$Q(t) = \sum y_i t^i = \xi_1 Q_1(t) + \xi_2 Q_2(t), \qquad Q_1, Q_2 \in P^n,$$

with $0 < \xi_1 = 1 - \xi_2 < 1$. Then every root of Q must be a root of both Q_1 and Q_2, of the same multiplicity or higher. But all of the n real and complex roots of Q are accounted for in $[0,1]$; hence we necessarily have $Q_1 = Q_2 = Q$, and the supposed convex representation collapses. This completes the proof.

We can now proceed to characterize the exceptional points in the boundary of P^n (see §2). The point y is in the contact set $C(y')$ of the point y' if and only if every supporting hyperplane containing y' also contains y. That is, if y is not in $C(y')$ then

$$y' \cdot x = 0, \quad y \cdot x > 0 \qquad \text{for some } x \in D^n.$$

Using some representation $x = \sum \xi_j x(t_j)$ for this point x (Theorems 3.1 and 7.3), we obtain

$$\sum \xi_j P'(t_j) = 0, \qquad \sum \xi_j P(t_j) > 0,$$

where P' and P correspond to y' and y respectively. Hence some root of P' in $[0,1]$ is not a root of P.

Conversely, if t' is a root of P' but not P then the simple

hyperplane $x(t')$ contains y' but not y, and y is not in $C(y')$. This proves:

THEOREM 9.3 *The contact set $C(y')$ of a point y' of P^n comprises exactly those polynomials of P^n which vanish at all roots of $\sum y_i' t^i$ in $[0,1]$.*

Now let us suppose that y is in $C(y')$, so that we have

$$(9.1)\qquad \begin{cases} \sum y_i' t^i = \prod (t-t_j)^{\alpha_j} R'(t) \\ \sum y_i t^i = \prod (t-t_j)^{\alpha_j} R(t) \end{cases} \qquad \begin{cases} \alpha_j = 2, & 0 < t_j < 1 \\ 1, & t_j = 0 \text{ or } 1 \\ t_j \text{ all distinct}, & j = 1,2,\dots,r(y'), \end{cases}$$

where R' has no roots in $[0,1]$ except possibly at some of the t_j. For the appropriate $\lambda \neq 0$, the polynomial $\lambda R(t)$ belongs to P^m, where

$$(9.2)\qquad m = n - \sum \alpha_j = n - 2r'(y').$$

It is easy to see from this that $C(y')$ and P^m are homeomorphic. Hence y' is an inner point of $C(y')$ if and only if the coefficients of $\lambda R'$ give an inner point of P^m --- that is, if and only if $\lambda R'(t)$ is strictly positive over $[0,1]$. Referring to the definitions of § 2, we see that we have proved:

THEOREM 9.4 *The point* y' *of* P^n *is exceptional if and only if* $\sum y_i' t^i$ *has a multiple root at* 0 *or* 1, *or a root interior to* $[0,1]$ *of multiplicity greater than* 2. *The reduced contact set* $C'(y')$ *comprises those polynomials in* P^n *which have roots at the roots in* $[0,1]$ *of* $\sum y_i' t^i$, *of the same multiplicity or higher.*

We have also established, by way of (9.2), the first of the following useful formulas:

THEOREM 9.5 *For* y *in* P^n,

$$c(y) = n - 2r'(y),$$

$$a(y) = n - r(y).$$

Proof of the second formula. If the roots in $[0,1]$ of $\sum y_i t^i$ are $t_1, \ldots, t_{r(y)}$, then we can exhibit the linearly independent hyperplanes:

$$x(t_1), x(t_2), \ldots, x(t_{r(y)}),$$

all of which support P^n at the point y. Any other such hyperplane must be a linear combination of these, for its convex representation cannot involve any extreme point $x(t)$ of D^n without t being a root of $\sum y_i t^i$. The intersection of these $r(y)$ hyperplanes, considered as point sets in $(E^n)^*$, is the set $L(y)$ (defined in §2), whose dimension defines the index $a(y)$. The formula now follows directly.

The last three theorems are mainly of interest for the boundary of P^n. However it is easily verified that they are valid for interior points as well; consequently we have stated them in the more general form.

§10. A property of the extreme points of P^n.

In this section we shall prove that every point of $P^n - Q^n$ is spanned by some pair of extreme points. That is, in the notation of §3, $b(y) = 1$ or 2 for every $y \in P^n$. Moreover, although there may be more than one spanning pair in some cases, we shall show that there is always a unique representation in which the extreme polynomials have interlocking sets of roots in $[0,1]$. This will prove a useful strengthening of the well-known theorem on the representation of a non-negative polynomial as a sum of squares.[7]

The next two theorems are lemmas for the main result.

THEOREM 10.1 *Let $m+1$ continuous, non-negative functions $f_0(\xi), f_1(\xi), \ldots, f_m(\xi)$ be defined on the simplex Ξ^m of points ξ:*

$$\xi = (\xi_0, \xi_1, \ldots, \xi_m), \quad \xi_m \geq 0, \quad \sum \xi_j = 1,$$

with the further property that each $f_j(\xi)$ vanishes on the face $\xi_j = 0$. Then for some ξ in Ξ^m,

[7] See for example Szegö [10], p. 4.

$$f_0(\xi) = f_1(\xi) \ldots = f_m(\xi).$$

Proof. Define

$$F_k(\xi) = f_k(\xi) - \min_j f_j(\xi), \quad k = 0, 1, \ldots, m,$$

and suppose that the $F_k(\xi)$ never all vanish at once. (At least one vanishes at each point, and all are non-negative.) Then the transformation

$$\xi'_k = \frac{F_{k+1}(\xi)}{\sum F_j(\xi)}$$

defines a new point ξ' in the boundary of Ξ^m. (We reduce subscripts modulo m+1.) The mapping $\xi \to \xi'$ is continuous; it therefore has a fixed point by the familiar theorem of Brouwer. If ξ^* is a fixed point, we can set up the following chain of implications:

$$\xi^*_j = 0 \implies f_j(\xi^*) = 0 \implies F_j(\xi^*) = 0 \implies \xi^{*\prime}_{j-1} = 0,$$

from which we conclude that all components of ξ^* vanish if any one vanishes. But we know that at least one component vanishes, and we also know that all cannot vanish. This contradiction forces us to abandon our original assumption about the $F_k(\xi)$; there must in fact be a ξ in Ξ^m for which all vanish:

$$F_0(\xi) = F_1(\xi) = \cdots = F_m(\xi) = 0.$$

This is equivalent to the assertion of the theorem.

THEOREM 10.2 *If* $f(t)$ *is continuous and positive for* t *in* $[0,1]$, *then a non-negative, n-th degree polynomial* $S(t)$ *with* $r'(S) = n/2$ *exists satisfying*

$$S(t) \leq f(t) \qquad 0 \leq t \leq 1$$

and such that equality holds at least once between each pair of distinct roots of $S(t)$. *Either of the further conditions:*

$$(A) \qquad S(1) = f(1) \text{ and } S(0) = \begin{cases} f(0) & (n \text{ even}) \\ 0 & (n \text{ odd}) \end{cases}$$

or

$$(\bar{A}) \qquad S(1) = 0 \text{ and } S(0) = \begin{cases} 0 & (n \text{ even}) \\ f(0) & (n \text{ odd}) \end{cases}$$

determines $S(t)$ *uniquely, as a multiple of an element of* Q^n *or* $\bar{Q}^n$ *respectively.*

Proof. We assume condition (A), with $n = 2m$. The three other cases are proved in essentially the same way. The polynomial S must have the form:

$$S(t) = \alpha S_u(t) = \alpha \prod_{j=1}^{m} (t-u_j)^2 \qquad (\alpha > 0)$$

where $u = (u_1, \ldots, u_m)$ is a point in the interior of the simplex U^m defined by:

$$u \in U^m \iff 0 \leq u_1 \leq \ldots \leq u_m \leq 1.$$

We define the quantity $\alpha_j(u)$ to be the greatest α such that

$$\alpha S_u(t) \leq f(t), \quad \text{all } t \in [u_j, u_{j+1}].$$

The definition is valid for $j = 0, 1, \ldots, m$ if we adopt the conventions $u_0 = 0$, $u_{m+1} = 1$. These $m+1$ functions are continuous and bounded away from zero throughout the interior of U^m. As u approaches the boundary face defined by $u_j = u_{j+1}$ the function $\alpha_j(u)$ tends to infinity. The reciprocal functions $1/\alpha_j(u)$ therefore satisfy the conditions of Theorem 10.1, and we conclude that a point u' exists with

$$\alpha_0(u') = \alpha_1(u') = \ldots = \alpha_m(u') = \alpha'.$$

The polynomial $S'(t) = \alpha' S_{u'}(t)$ is equal to $f(t)$ at least once in each interval between roots, and clearly has the other properties required by the theorem.

To show that S' is unique, we take any S'' having the same properties and examine the difference, $S'-S''$, again an n-th degree polynomial. In fact, suppose that the smallest root of S'' is less than the smallest root u_1' of S'. Then it is easy to show that $S'-S''$ has at least one root in the open interval $(0, u_1')$ and two roots in each interval $[u_1', u_2'], \ldots, [u_{m-1}', u_m']$ for a total of $2m - 1$ (the possible coincidence of some of the roots of S' and S'' does not affect the total, counting multiplicity). In addition, $S'-S''$ vanishes at 0 and at 1, so that the number of roots exceeds the degree. Hence $S'-S'' \equiv 0$.

THEOREM 10.3 Every non-extreme point y of P^n has a unique convex representation by a pair of extreme points, one each from Q^n and $\overline{Q}^n$, whose roots interlock, as follows:

$$\sum_{i=0}^{n} y_i t^i = \alpha \prod_{j=1}^{m} (t-t_{2j-1})^2 + \beta t(1-t) \prod_{j=1}^{m-1} (t-t_{2j})^2$$

if $n = 2m$; and

$$\sum_{i=0}^{n} y_i t^i = \alpha t \prod_{j=1}^{m} (t-t_{2j})^2 + \beta(1-t) \prod_{j=1}^{m} (t-t_{2j-1})^2$$

if $n = 2m+1$, with $\alpha > 0$, $\beta > 0$, $0 \le t_1 \le \ldots \le t_{n-1} \le 1$. Moreover, y is interior to P^n if and only if all of the inequalities are strict.

Proof. Take y interior to P^n and denote $\sum y_i t^i$ by $P(t)$. Then P is strictly positive over $[0,1]$ (Theorem 9.1). Applying Theorem 10.2 ($\underline{A}$) and ($\overline{A}$) to P gives us polynomials

$$\underline{S} = \xi \underline{Q} \quad \text{and} \quad \overline{S} = \eta \overline{Q} \qquad (\underline{Q} \in \underline{Q}^n, \ \overline{Q} \in \overline{Q}^n).$$

But $P - \underline{S}$ is a polynomial with the same properties as $\overline{S}$; by the uniqueness we must have $P - \underline{S} = \overline{S}$, or

$$P(t) = \xi \underline{Q}(t) + \eta \overline{Q}(t) \qquad (\xi > 0, \ \eta > 0).$$

Since P, $\underline{Q}$, and $\overline{Q}$ are all in P^n, we have $\xi + \eta = 1$. Moreover, the roots of $\underline{Q}$ and $\overline{Q}$ interlock (strictly) as required.

If y is in the boundary of P^n, then the same procedure works if

we start by dividing out the roots in $[0,1]$ of $P(t)$ and finish by multiplying them back into both terms of the representation. But now, with some roots common to both extreme polynomials, the two sets of roots will no longer interlock so as to satisfy the strict inequalities. This completes the proof.

We may remark that some points in P^n are spanned also by pairs of extreme points whose roots do not interlock, or by pairs of extreme points from the same component of Q^n. However, Theorem 10.3 describes the only natural way of generating P^n from Q^n, in which every point is represented once and only once.

§ 11. The boundary of D^n.

Returning to the moment space D^n we now investigate the manner in which its boundary is spanned by the extreme points. In order to be able to apply Theorem 7.3 we assume throughout this section that $n \geq 2$.

THEOREM 11.1 *The representation of the point x in D^n by extreme points is unique if and only if x is in the boundary.*

Proof. (a) The set of extreme points C^n is a twisted curve (see § 7) that does not meet any hyperplane in more than n points; hence the contact set $C(x)$ of a boundary point x can contain at most n of the extreme points. But any n or fewer extreme points are linearly independent (see proof (d) of Theorem 7.2), therefore $C(x)$ is a simplex. The representation (3.1) of x depends only on the vertices of the simplex, and hence it is unique.

(b) If x is interior to D^n, then it is inner to the segment connecting $x(t)$ with the directly opposite boundary point x', $x(t)$ being any point on C^n. A representation for x can then be constructed by combining $x(t)$ with the representation for x'; $x(t)$ will necessarily appear with positive weight. But since t was arbitrary in $[0,1]$, there are infinitely many distinct such representations for x.

THEOREM 11.2 D^n *has no exceptional points.*

Proof. No interior point is exceptional. Let x be a boundary point of D^n and consider its unique representation

$$x = \sum_{j=1}^{b(x)} \xi_j x(t_j), \qquad \sum_{j=1}^{b(x)} \xi_j = 1, \qquad \xi_j > 0, \qquad j=1, \ldots, b(x).$$

We shall show that $C(x)$ is precisely the simplex S spanned by the points $x(t_j)$. Then, since x is clearly inner to S, x is by definition not exceptional. We at once have $S \subseteq C(x)$ because every supporting hyperplane at x contains all the points $x(t_j)$. To show that $S = C(x)$ we need only exhibit one hyperplane that contains the extreme points $x(t_j)$ and no others. But any polynomial in P^n having just the set of roots $\{t_j\}$ will provide such a hyperplane.

THEOREM 11.3 *For x in the boundary of D^n,*

$$c(x) = b(x) - 1.$$

Proof. By the definitions of $b(x)$ and $c(x)$ (see §3 and §2) we have

$$c(x) = \dim C(x) = \dim S = b(x) - 1,$$

S being taken as in the preceding proof.

THEOREM 11.4 *If* $x \in D^n$ *and* $y \in P^n$ *are conjugate points, then*

$$b(x) = r(y).$$

Proof. Theorems 5.2, 9.5, and 11.3 give us respectively:

$$n - 1 = c(x) + a(y),$$
$$a(y) = n - r(y),$$
$$c(x) = b(x) - 1.$$

Adding the three equations gives the desired result.

Consistently with our definitions in §6 and §9 we define $b'(x)$ to be $b(x) - 1/2$ if one of $x(0)$, $x(1)$ occurs in the most efficient representation of $x \in D^n$; to be $b(x) - 1$ if both occur; and to be $b(x)$ if neither occurs. We then have, without difficulty:

THEOREM 11.5 *If* $x \in D^n$ *and* $y \in P^n$ *are conjugate points, then*

$$b'(x) = r'(y).$$

THEOREM 11.6 *For* x *in* D^n,

$$a(x) = 2b'(x) - 1.$$

Proof. (a) First take x in the boundary of D^n. Theorems 5.2, 9.5, and 11.5 give us respectively:

$$n - 1 = a(x) + c(y),$$
$$c(y) = n - 2r'(y),$$
$$b'(x) = r'(y);$$

where the existence of the conjugate point y in P^n is assured by Theorem 5.1b. The sum of the first two equations and twice the third gives the desired result.

(b) Suppose now that x is in the interior of D^n, so that $a(x) = n$. Pass a line through $x(0)$ and x, meeting the opposite boundary at x'. As in the proof of Theorem 11.1b we can build up a representation for x out of $x(0)$ and the representation for x', giving us the estimate

$$b'(x) \leq b'(x') + \frac{1}{2}.$$

But $a(x')$ is at most $n-1$, so that part (a) of this proof give us

$$b'(x) \leq \frac{n+1}{2}. \tag{11.1}$$

However, there might be a more efficient representation for x, giving us

(11.2) $$b'(x) < \frac{n+1}{2}, \quad \text{i.e.,} \quad b'(x) \leq \frac{n}{2}.$$

If we denote the extreme points involved by $x(t_j)$, $j=1, \ldots, b(x)$, then it is possible to construct a non-negative polynomial, vanishing at every t_j, whose degree would not exceed n. (See for example (9.1).) This would entail the existence of a supporting hyperplane to D^n which contains all the points $x(t_j)$ and therefore the point x itself. Since x was assumed interior, (11.2) is impossible and the equality must hold in (11.1). This proves the theorem.

As a corollary we have

$$b(x) \leq \frac{n+2}{2}$$

for all x in D^n. This contrasts with

$$b(y) \leq 2$$

for points in P^n (Theorem 10.[illegible]) and with

$$b(z) \leq n+1$$

for the general convex body in E^n (Theorem 3.2).

With Theorem 11.6 at our disposal, we find it possible to classify points of D^n in a natural way. We define

$$x \in C_a^n \iff a(x) = a.$$

More particularly, if $a < n$, we shall use $\overline{C}^n_a$ for those points of C^n_a whose unique representation (see Theorem 11.1) involves the point $x(1)$, and $\underline{C}^n_a$ for $C^n_a - \overline{C}^n_a$. We shall refer to these sets as the a-faces of D^n. For example, we have

$$C^n = \underline{C}^n_1 \cup C^n_0.$$

The partition of the boundary of D^n into a-faces, $0 \leq a < n$, generates a dual partition of the boundary of P^n into "c-faces" $\underline{Q}^n_c$, $\overline{Q}^n_c$, $0 \leq c < n$, on which $c(y)$ is constant. This follows from Theorem 5.3, since it is evident that the partition there is a refinement of the present partition. (See §5.) In particular, we have:

$$\underline{Q}^n_0 = (\underline{C}^n_{n-1})^* = (\text{ordinary points of } \underline{Q}^n)$$

$$\overline{Q}^n_0 = (\overline{C}^n_{n-1})^* = (\text{ordinary points of } \overline{Q}^n),$$

as may be seen from Theorems 5.2, 9.5, 9.4, and the definition of Q^n.

THEOREM 11.7 *The two a-faces $\underline{C}^n_a$ and $\overline{C}^n_a$, $a < n$, in the boundary of D^n are the maximal connected components of C^n_a, and have (topological) dimension a.*

Proof. A typical point x in, say, $\underline{C}^n_{2m-1}$, has the unique representation

$$x = \xi_0 x(u_1) + \xi_1 x(u_2) + \cdots + \xi_{m-1} x(u_m), \tag{11.3}$$

since $b'(x) = b(x) = m$. Here the ξ_j are all positive, with sum 1, and the u_j are all distinct, not 0 or 1, and arranged for the sake of uniqueness in ascending order. We can therefore establish a one-one correspondence between C^m_{2m-1} and the product of the interiors of the two simplexes U^m and Ξ^{m-1}:

$$(11.4) \qquad u \in U^m \iff 0 \le u_1 \le \dots \le u_m \le 1;$$

$$(11.5) \qquad \xi \in \Xi^{m-1} \iff \xi_0 \ge 0,\ \xi_1 \ge 0,\ \dots,\ \xi_{m-1} \ge 0,$$

$$\sum_{j=0}^{m-1} \xi_j = 1.$$

The mapping obviously does not disturb topological properties, hence $2m-1$ is the (topological) dimension of C^n_{2m-1}, as required. In a similar way, $\overline{C}^n_{2m-1}$ is related to the pair of simplexes U^{m-1}, Ξ^m; C^n_{2m} to U^m, Ξ^m; and $\overline{C}^n_{2m}$ again to U^m, Ξ^m. The connectivity of the individual a-faces is apparent from the above, and their pairwise separation is evident from the relation:

$$(11.6) \qquad (\text{closure } C^n_a) - C^n_a = \bigcup_{k=0}^{a-1} C^n_k = (\text{closure } \overline{C}^n_a) - \overline{C}^n_a,$$

which follows from (11.3). This completes the proof.

We would now like to say something about the representation of points in C^n_n, the interior of D^n. If we momentarily turn our attention to the higher-dimensional moment space D^{n+k}, $k > 0$, we see that there is a natural correspondence between C^n_n and either one of the two n-faces of D^{n+k}. In fact, C^n_n is just the non-singular, perpendicular projection of

$\underline{C}_n^{n+k}$ or $\overline{C}_n^{n+k}$ into E^n. The unique representations in the higher sets go into "minimal" representations in C_n^n, involving exactly $b' = (n+1)/2$ extreme points. Each x in C_n^n has precisely two such minimal representations, formally identical with the representations of the inverse images $\underline{x}_{(n+k)}$ and $\overline{x}_{(n+k)}$ of x in $\underline{C}_n^{n+k}$ and $\overline{C}_n^{n+k}$, respectively.

Considering the proof of the preceding theorem, we can therefore assert:

THEOREM 11.8 *The interior of* D^{2m-1} *is swept out by an* m*-parameter family of* $(m-1)$*-dimensional simplexes* $\equiv^{m-1}$, *and, in a different way, by an* $(m-1)$*-parameter family of* m*-dimensional simplexes* $\equiv^m$. *The interior of* D^{2m} *is swept out in two different ways by* m*-parameter families of* m*-dimensional simplexes* $\equiv^m$.

We shall make use of this parametrization in § 15.

CHAPTER III

THE SIMPLEXES S^n AND B^n

§ 12. The simplex S^n

The moment space D^n and its dual P^n are closely related to a pair of dual simplexes, with whose aid we are able to geometrize some classical properties of moment sequences and positive polynomials. The first simplex, circumscribed about D^n, is defined by its vertices $x^{(k)}$, $k = 0, 1, \ldots, n$, with

$$x_i^{(k)} = \frac{k!(n-i)!}{n!(k-i)!} = \frac{\binom{k}{i}}{\binom{n}{i}},$$

where, as usual, $1/(k-i)! = \binom{k}{i} = 0$ if $i > k$. Displayed as a matrix, these coordinates are:

$$(12.1) \qquad \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & \frac{1}{n} & 0 & \ldots & 0 & 0 \\ 1 & \frac{2}{n} & \frac{2\cdot 1}{n(n-1)} & \ldots & 0 & 0 \\ \ldots & & \ldots & & & \ldots \\ 1 & \frac{n-1}{n} & \frac{n-2}{n} & \ldots & \frac{1}{n} & 0 \\ 1 & 1 & 1 & \ldots & 1 & 1 \end{pmatrix}$$

We shall denote this simplex by S^n. Figure 12.1 shows S^2 in relation to D^2.

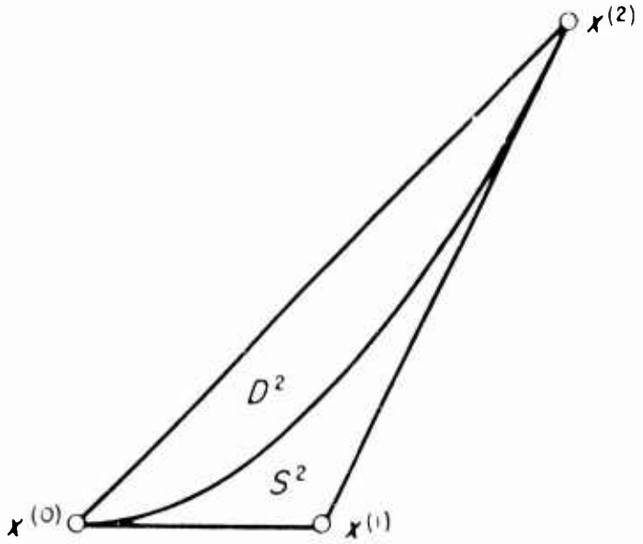


Fig. 12.1

THEOREM 12.1 S^n *contains* D^n.

Proof. The key to the proof is the identity

$$t^i = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \frac{\binom{k}{i}}{\binom{n}{i}} \qquad i = 0, 1, \ldots, n. \tag{12.2}$$

Accepting this, we obtain by direct substitution:

$$x(t) = \sum_{k=0}^{n} \xi_k x^{(k)}, \tag{12.3}$$

with

$$\xi_k = \binom{n}{k} t^k (1-t)^{n-k} \geq 0.$$

A simple calculation (or (12.2) with $i=0$) gives us

$$\sum_{k=0}^{n} \xi_k = 1.$$

This shows that the extreme points $x^{(k)}$ of S^n span the extreme points $x(t)$ of D^n, and hence that S^n contains D^n.

To verify the identity (12.2) we observe that

$$\binom{n}{k}\binom{k}{i} = \binom{n-i}{k-i}\binom{n}{i}.$$

The right-hand member of (12.2) may therefore be written:

$$\sum_{k=0}^{n} \binom{n-i}{k-i} t^k (1-t)^{n-k}. \tag{12.4}$$

With the substitutions $\ell = k-i$, $m = n-i$, (12.4) becomes

$$\sum_{\ell=-i}^{m} \binom{m}{\ell} t^{\ell+i}(1-t)^{m-\ell} = t^i \sum_{\ell=0}^{m} \binom{m}{\ell} t^{\ell}(1-t)^{m-\ell} = t^i,$$

[redacted]

$\lambda_{nk}(\phi)$ were introduced by Hausdorff [11], using the successive differences [redacted] of the sequence of moments of [redacted]

$$\lambda_{nk}(\phi) = \binom{n}{k}\Delta^{n-k}\mu_k(\phi) = \binom{n}{k}\sum_{\ell=0}^{n-k}(-1)^{\ell}\binom{n-k}{\ell}\mu_{k+\ell}(\phi). \tag{12.5}$$

[redacted]

(See also Widder [9], pp. 100 ff; Shohat and Tamarkin [12], pp. 93 ff.)

These quantities have a simple geometric interpretation. Applying (6.2) to the right-hand expression above gives us

$$\lambda_{nk}(\phi) = \int_0^1 \binom{n}{k} t^k (1-t)^{n-k} d\phi(t). \tag{12.6}$$

If we denote by x_1 the moments of ϕ, we then have

$$x = \int_0^1 x(t)\,d\phi(t) = \sum_{k=0}^{n} \lambda_{nk}(\phi) x^{(k)},$$

by (12.3) and (12.6), with

$$\lambda_{nk} \geq 0, \quad \sum_{k=0}^{n} \lambda_{nk} = 1.$$

Thus, the $\lambda_{nk}(\phi)$ are just the barycentric coordinates of x in the simplex S^n.

§13. The dual simplex P^n.

Turning to the dual space, we consider the "Bernstein" polynomials:

$$B_{nk}(t) = \binom{n}{k} t^k (1-t)^{n-k}, \qquad k = 0, 1, \ldots, n. \tag{13.1}$$

By Theorem 9.2 these polynomials are, up to a positive factor, extreme points of P^n. In fact we have

$$(n+1)B_{nk} \in \begin{cases} Q^n & \text{if } n-k \text{ is even;} \\ \overline{Q}^n & \text{if } n-k \text{ is odd.} \end{cases}$$

Denoting the point $(n+1)B_{nk}$ by $y^{(k)}$ we obtain the following expressions for its coordinates:

$$y_i^{(k)} = (-1)^{i-k} \frac{(n+1)!}{(n-1)!(i-k)!\,k!} .$$

The matrix of the $y_i^{(k)}$ is substantially the inverse of the matrix (12.1), for we have

$$(13.2) \qquad y^{(k)} \cdot x^{(\ell)} = \begin{cases} n+1 & \text{if } k = \ell ; \\ 0 & \text{if } k \neq \ell . \end{cases}$$

The simplex spanned by the $n+1$ points $y^{(k)}$ we denote by B^n; it is obviously inscribed in P^n. Figure 13.1 indicates the configuration for $n = 2$.

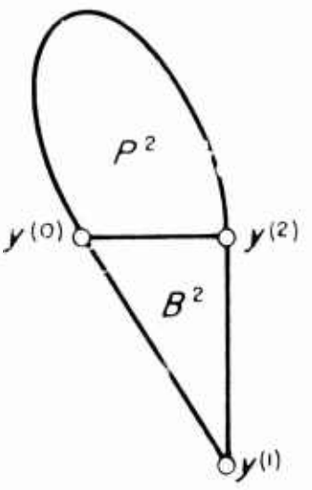


Fig. 13.1

THEOREM 13.1 B^n is dual to S^n.

Proof. It suffices to show that the vertices $x^{(\ell)}$ of S^n, interpreted in the dual space, are the hyperplanes containing the $(n-1)$-dimensional faces of B^n. But, by (13.2), each hyperplane $x^{(\ell)}$ contains exactly n of the vertices $y^{(k)}$ of B^n, and the result follows.

It is now evident that the hyperplanes which determine the $(n-1)$-dimensional faces of S^n are supporting hyperplanes to D^n as well, so that in some sense S^n is the most closely fitting simplex that can be circumscribed about D^n. Dually, B^n is in some sense a maximal inscribed simplex in P^n. A precise meaning to these statements will be given in the next two sections.

A simple calculation shows that the centroids of the two simplexes S^n and B^n are respectively

$$\tilde{x} = (1, \tfrac{1}{2}, \ldots, \tfrac{1}{n+1})$$

and

$$\tilde{y} = (1, 0, \ldots, 0).$$

These will be recognized from §8 as the normalizing vectors used in selecting the cross-sections P^n and D^n of the conjugate, convex curves $\Gamma(P^n)$ and $\Gamma(D^n)$. They correspond to the "rectangular" distribution $\phi(t) = t$ and the constant polynomial $P(t) = 1$.

§14. Fit of S^n and B^n to D^n and P^n.

We have already noted that P^n is a cross-section of P^{n+1}. Moreover the corresponding cross-section of the inscribed simplex B^{n+1} is [illegible]

B^m, but a larger, polyhedral body, more nearly filling out the interior of P^n.

THEOREM 14.1 *If y is interior to P^n, then for sufficiently large m the point*

$$y_{(m)} = (y_0, y_1, \ldots, y_n, 0, \ldots, 0)$$

is in B^m.

Proof[8]. Denote by $x_{(m)}^{(k)}$ the vertices of S^m, and hence also the hyperplanes in $(E^m)^*$ carrying the boundary of B^m. We must show that

$$y_{(m)} \cdot x_{(m)}^{(k)} \geq 0, \qquad k = 0, 1, \ldots, n \tag{14.1}$$

holds (for fixed m) if m is taken sufficiently large. Writing $tm = k$, we have for each $i = 0, 1, \ldots, m$:

$$x_{(m)i}^{(k)} = \frac{tm(tm-1) \ldots (tm-i+1)}{m(m-1) \ldots (m-i+1)} \; ;$$

by (12.1). As m increases this converges (uniformly in t) to t^i. But from the hypothesis and Theorem 9.1, we have:

$$\sum_{i=0}^{n} y_i t^i \geq \delta > 0, \qquad 0 \leq t \leq 1.$$

It follows that (14.1) holds for sufficiently large m, as was to be shown.

---

8 See Hausdorff [11], p. 224.

The proof has shown that as m increases the vertices $x^{(k)}$ of S^m tend to the points $x(k/m)$ on C^m. The projections of S^m on E^n (n fixed) form a nested sequence of polyhedral, convex bodies tending to D^n. However, the maximum distance of S^m to D^m does not tend to zero as m increases. For example,

$$\left| x_m((m-1)/m) - x_m^{(m-1)} \right| = (1-1/m)^m \to 1/e.$$

We shall also see in the next section that the volumes of S^m and D^m do not approach each other asymptotically.

THEOREM 14.2 *Given the sequence*

$$x_0, x_1, \ldots, x_n, \cdots,$$

then $x_{(m)} = (x_0, x_1, \ldots, x_m)$ *is in* S^m *for all* m *if and only if* x *is in* D^m *for all* m.

Proof. This is essentially the dual form of Theorem 14.1, it follows from it at once by means of Theorems 12.1 and 4.3.

Translating these theorems out of geometric terminology we obtain two well-known results:

THEOREM 14.3 *Any polynomial positive on* $[0,1]$ *can be represented as a finite sum with positive coefficients of polynomials*

$$B_{nk}(t) = \binom{n}{k} t^k (1-t)^{n-k}$$

of sufficiently high degree.

THEOREM 14.4 A necessary and sufficient condition that

$$\{x_0, x_1, \ldots, x_n, \ldots\}$$

be the moments of some distribution function on $[0,1]$ is that the sequence be "completely monotonic", i.e., that the x_i and their successive differences $\triangle^k x_i$ of all orders be all non-negative.

The latter follows from the expression (12.5) relating the $\triangle^k x_i$ to the barycentric coordinates of x in S^{k+i}.

§15. Comparison of volumes.

We now calculate the n-dimensional volumes of S^n and D^n, to obtain further insight into the relation between the two bodies. The result is somewhat special -- unlike our other results, for example, the extension to moment spaces based on the general finite interval $[a,b]$, though not difficult, is not immediately apparent. But we feel that the remarkably neat formula for vol D^n in terms of the beta and gamma functions justifies the inclusion of a portion of the calculation.

THEOREM 15.1 The n-dimensional volume of S^n is

$$\text{vol } S^n = \frac{1}{n!} \prod_{k=0}^{n} \left[\binom{n}{k}\right]^{-1} = \prod_{k=1}^{n} \frac{\Gamma(k)\Gamma(k)}{\Gamma(n+1)} .$$

Proof. We omit this elementary calculation.

THEOREM 15.2 The n-dimensional volume of D^n is

$$\text{vol } D^n = \prod_{k=1}^{n} B(k, k) = \prod_{k=1}^{n} \frac{\Gamma(k)\,\Gamma(k)}{\Gamma(2k)} .$$

Proof. We outline the proof for the case $n = 2m-1$. By Theorem 11.8 we may regard the interior of D^n as the product of the interiors of U^m and Ξ^{m-1}, defined by (11.4) and (11.5). It is convenient to replace U^m by the unit m-cube I^m, thereby multiplying the volume by $m!$. Thus

$$\text{vol } D^n = \int_{D^n} dx_1 \dots dx_n = \frac{1}{m!} \int_{I^m} \int_{\Xi^{m-1}} |J|\, d\xi_1 \dots d\xi_{m-1} du_1 \dots du_m ;$$

J being the jacobian of the transformation (11.3) with ξ_m replaced by $1 - \xi_1 - \dots - \xi_{m-1}$. A calculation yields

$$|J| = \prod_{j=1}^{m} \xi_j \prod_{j=2}^{m} \prod_{k=1}^{j-1} (u_k - u_j)^4 ,$$

enabling us to split up the integral. Thus:

$$\text{vol } D^n = \frac{1}{m!} \int_{\Xi^{m-1}} \prod \xi_j^{\,j} d\xi_1 \dots d\xi_{m-1} \int_{I^m} \prod_{k<j} (u_k - u_j)^4 du_1 \dots du_m .$$

The first integral is $1/\Gamma(2m)$. By a result of Selberg [13] the value of the second is

$$\prod_{k=1}^{m} \frac{\Gamma(2k+1)\,[\Gamma(2k-1)]^2}{2\,\Gamma(2m+2k-2)} .$$

Combining these results yields:

$$\text{vol } D^n = \frac{1}{\Gamma(2m)} \prod_{k=1}^{m} \frac{\Gamma(2k+1)\left[\Gamma(2k-1)\right]^2}{2k\,\Gamma(2m+2k-2)} .$$

From this the theorem can be verified directly, or by induction on m. The case $n = 2m$ is treated similarly.

We are now in a position to compare the sizes of S^n and D^n, as n increases. A reasonable measure might be the n-th root of the ratio of volumes, which could be expected to approach 1 for large n, even if the ratio itself does not. However, it is not difficult to show, from the theorems just proved, that

$$\left(\frac{\text{vol } D^n}{\text{vol } S^n}\right)^{\frac{1}{n}} = O(2^{-n_e n/2}) \to 0.$$

This result makes an interesting contrast with Theorem 14.2.

CHAPTER IV

ALGEBRAIC DESCRIPTION OF THE MOMENT SPACES

§16. Moment sequences and quadratic forms.

We turn first to the question of whether a given finite sequence constitutes the first n moments of some function in $\mathcal{A}$ — that is, whether a given point

$$x = (x_0, x_1, \ldots, x_n)$$

is in D^n.[9]

By Theorem 8.1 we have

$$x \in D^{2m} \iff x \cdot y \geq 0 \quad \text{for all } y \text{ in } P^{2m} \tag{16.1}$$

(considering first the even-dimensional case). It is equivalent, however, to have y range over just the extreme points of P^{2m}; thus:

$$x \in D^{2m} \iff x \cdot y \geq 0 \quad \text{for all } y \text{ in } Q^{2m}. \tag{16.2}$$

We have seen in §9 that the polynomials of Q^{2m} all have one or the other of the two forms

$$[R_m(t)]^2 \quad \text{or} \quad t(1-t)[R_{m-1}(t)]^2 \tag{16.3}$$

(subscripts here indicate the degree). Furthermore, all polynomials (16.3), though not necessarily in Q^{2m}, are certainly in P^{2m}, or positive multiples of elements of P^{2m}. Thus (16.1) and (16.2) are equivalent to:

---

[9] The criterion of Theorem 14.4, of course, applies only to infinite sequences.

$x \in D^{2m} \iff x \cdot y \geq 0$ for all y such that $\sum_{i=0}^{2m} y_i t^i$ has either the form $\left[\sum_{j=0}^{m} \alpha_j t^j\right]^2$ or the form $t(1-t)\left[\sum_{j=0}^{m-1} \beta_j t^j\right]^2$.

Eliminating y gives us finally:

$$x \in D^{2m} \iff \sum_{j=0}^{m}\sum_{k=0}^{m} x_{j+k}\alpha_j\alpha_k \geq 0, \quad \text{all } (\alpha_0, \ldots, \alpha_m), \text{ and}$$

$$\sum_{j=0}^{m-1}\sum_{k=0}^{m-1} (x_{j+k+1}-x_{j+k+2})\beta_j\beta_k \geq 0, \quad \text{all } (\beta_0, \ldots, \beta_{m-1}).$$

We thus have the well-known result[10]:

THEOREM 16.1a A necessary and sufficient condition that x be a point of D^{2m} is that the two quadratic forms:

$$\sum_{j,k=0}^{m} x_{j+k}\alpha_j\alpha_k \quad \text{and} \quad \sum_{j,k=0}^{m-1} (x_{j+k+1}-x_{j+k+2})\beta_j\beta_k$$

be positive definite or semidefinite.

In an analogous manner we may establish:

THEOREM 16.1b A necessary and sufficient condition that x be a point of D^{2m+1} is that the quadratic forms:

[10] See for example Shohat and Tamarkin [12], p. 17.

$$\sum_{j,k=0}^{m} x_{j+k+1}\alpha_j\alpha_k \qquad \underline{\text{and}} \qquad \sum_{j,k=0}^{m} (x_{j+k}-x_{j+k+1})\beta_j\beta_k$$

be positive definite or semidefinite.

The proof of the next theorem follows the same lines as above, making use of Theorem 4.2.

THEOREM 16.2 A necessary and sufficient condition that x be an interior point of D^n is that the quadratic forms of Theorem 16.1a or 16.1b (whichever applies) be positive definite.

§17. The determinants Δ_m and their relation to the faces of D^n.

We now introduce a special notation for the "Hankel" determinants associated with the quadratic forms of the last section:

$$\Delta_{2k} = \begin{vmatrix} 1 & x_1 & \cdots & x_k \\ \vdots & & & \\ x_k & x_{k+1} & \cdots & x_{2k} \end{vmatrix}$$

$$\Delta_{2k+1} = \begin{vmatrix} x_1 & x_2 & \cdots & x_{k+1} \\ \vdots & & & \\ x_{k+1} & x_{k+2} & \cdots & x_{2k+1} \end{vmatrix}$$

$$\overline{\Delta}_{2k} = \begin{vmatrix} x_1 - x_2 & x_2 - x_3 & \cdots & x_k - x_{k+1} \\ \vdots & & & \\ x_k - x_{k+1} & x_{k+1} - x_{k+2} & \cdots & x_{2k-1} - x_{2k} \end{vmatrix}$$

$$\overline{\Delta}_{2k+1} = \begin{vmatrix} 1 - x_1 & x_1 - x_2 & \cdots & x_k - x_{k+1} \\ \vdots & & & \\ x_k - x_{k+1} & x_{k+1} - x_{k+2} & \cdots & x_{2k} - x_{2k+1} \end{vmatrix}$$

The subscripts have been chosen so as to indicate the highest moment occurring in each determinant. The upper and lower bars will be seen in due course to agree with our previous usage.

THEOREM 17.1 A necessary and sufficient condition that the quadratic form

$$\sum_{i,j} a_{ij} \alpha_i \alpha_j$$

be positive definite is that the first principal minors of its symmetric matrix all be positive.

Note. By the "first principa minors" of an array such as the matrix of Δ_{2k} given above, we mean the k+1 subdeterminants $\Delta_0, \Delta_2, \cdots \Delta_{2k}$.

For a proof of this standard result, see for example Ferrer [14], page 138.

THEOREM 17.2 *The point* x *is interior to* D^n *if and only if all of the determinants*

(17.1) $$\underline{\Delta}_0,\ \overline{\Delta}_0,\ \underline{\Delta}_1,\ \overline{\Delta}_1,\ \cdots,\ \underline{\Delta}_n,\ \overline{\Delta}_n$$

are positive.

Proof. By Theorems 16.2 and 17.1 x is interior to D^n if and only if $\underline{\Delta}_n$, $\overline{\Delta}_n$, $\underline{\Delta}_{n-2}$, $\overline{\Delta}_{n-2}$, etc., are positive. It only remains to remark that if x is interior to D^n, then the point

$$x_{(n-1)} = (x_0, x_1, \cdots, x_{n-1})$$

is necessarily interior to D^{n-1}, and the theorem follows at once.

Now suppose that x is in the boundary of D^n — say in the face $\overline{C}^n_{n-1}$ (see §11). It follows that $x_{(n-1)}$ is interior to D^{n-1}, so that all but the last two determinants (17.1) are necessarily positive, and one of these last two must vanish. Since the polynomial $\sum y_i t^i$ associated with the unique supporting hyperplane $y \in P'$ at x has a root at $t=1$, we can see that the second quadratic form of Theorem 16.1 (a or b) is positive semi-definite, while the first form is easily shown to be positive definite. Hence

$$\underline{\Delta}_n > 0, \qquad \overline{\Delta}_n = 0.$$

If x had been in the face $\underline{C}^n_{n-1}$ the upper and lower bars would have been reversed.

Applying the same argument to the lower-dimensional faces gives us the following characterization theorem for the a-faces of D^n:

THEOREM 17.3 *The point* $x \in D^n$ *is in* C_a^n *if and only if*

$$\underline{\Delta}_0,\ \overline{\Delta}_0,\ \underline{\Delta}_1,\ \overline{\Delta}_1,\ \ldots,\ \underline{\Delta}_a,\ \overline{\Delta}_a,\ \text{and}\ \overline{\Delta}_{a+1}$$

are positive and

$$\underline{\Delta}_{a+1}\ \text{and}\ \underline{\Delta}_{a+2},\ \overline{\Delta}_{a+2},\ \ldots,\ \underline{\Delta}_n,\ \overline{\Delta}_n$$

are zero. Similarly x *is in* $\overline{C}_a^n$ *if and only if*

$$\underline{\Delta}_0,\ \overline{\Delta}_0,\ \underline{\Delta}_1,\ \overline{\Delta}_1,\ \ldots,\ \underline{\Delta}_a,\ \overline{\Delta}_a,\ \text{and}\ \underline{\Delta}_{a+1}$$

are positive and

$$\overline{\Delta}_{a+1}\ \text{and}\ \underline{\Delta}_{a+2},\ \overline{\Delta}_{a+2},\ \ldots,\ \underline{\Delta}_n,\ \overline{\Delta}_n$$

are zero.

All of the determinants of index $a+2$ and higher vanish because of the fact that C_a^n is contained in the closures of both $\underline{C}_m^n$ and $\overline{C}_m^n$, for $a < m \leq n$. (See (11.6)).

It is essential here to specify that x be in D^n, because the analogues of Theorems 17.2 and 17.1 for non-negative determinants do not hold. For an illustrative example, see Widder [9], pages 135-6.

§ 18. The lower and upper boundaries of D^n.

Equating Δ_n to zero gives us a particular value $\underline{x}_n$ for the n-th moment as a function of the preceding moments. We may write this relationship in the compact form:

$$\underline{x}_n = x_n - \frac{\Delta_n}{\Delta_{n-2}}, \tag{18.1}$$

which is actually independent of x_n, as it should be. In similar fashion, we see that

$$\bar{x}_n = x_n + \frac{\Sigma_n}{\Delta_{n-2}} \tag{18.2}$$

is the value for the n-th moment that makes Σ_n vanish.

For any point $\mathbf{x}$ in D^n we can define the associated points

$$\underline{\mathbf{x}} = (x_0, x_1, \ldots, x_{n-1}, \underline{x}_n)$$

$$\bar{\mathbf{x}} = (x_0, x_1, \ldots, x_{n-1}, \bar{x}_n).$$

Since we always have

$$\underline{x}_n \leq x_n \leq \bar{x}_n$$

we can interpret $\underline{\mathbf{x}}$ and $\bar{\mathbf{x}}$ as the projections "downward" and "upward" of $\mathbf{x}$ on the boundary of D^n.

Since, moreover, we always have by Theorem 17.3

$$\underline{\mathbf{x}} \text{ in closure of } C^n_{n-1}$$

$$\bar{\mathbf{x}} \text{ in closure of } \bar{C}^n_{n-1},$$

we shall refer to these two closed sets as respectively the <u>lower</u> and <u>upper</u> boundaries of D^n.

CHAPTER V

DISTRIBUTIONS HAVING GIVEN MOMENTS

§19. The polynomials $\underline{\Delta}_m(t)$ and $\overline{\Delta}_m(t)$.

Our next task will be to find out what can be said about the distribution functions which give rise to a given point x of D^n. We first need an explicit form for the hyperplanes which support D^n at the associated upper and lower boundary points $\underline{x}$ and $\overline{x}$ (see §18); this form is provided by the following polynomials, closely related to the determinants $\underline{\Delta}_m$ and $\overline{\Delta}_m$ of §17. We define:

$$\underline{\Delta}_{2k}(t) = \begin{vmatrix} 1 & x_1 & \cdots & x_{k-1} & 1 \\ & & & & t \\ \vdots & & & & \vdots \\ x_k & x_{k+1} & \cdots & x_{2k-1} & t^k \end{vmatrix}$$

$$\underline{\Delta}_{2k+1}(t) = \begin{vmatrix} x_1 & x_2 & \cdots & x_k & 1 \\ & & & & t \\ \vdots & & & & \vdots \\ x_{k+1} & x_{k+2} & \cdots & x_{2k} & t^k \end{vmatrix}$$

$$\overline{\Delta}_{2k}(t) = \begin{vmatrix} x_1 - x_2 & x_2 - x_3 & \cdots & x_{k-1} - x_k & 1 \\ & & & & t \\ \vdots & & & & \vdots \\ x_k - x_{k+1} & x_{k+1} - x_{k+2} & \cdots & x_{2k-2} - x_{2k-1} & t^{k-1} \end{vmatrix}$$

$$\underline{\Delta}_{2k+1}(t) = \begin{vmatrix} 1 - x_1 & x_1 - x_2 & \cdots & x_{k-1} - x_k & 1 \\ & & & & t \\ \vdots & & & & \vdots \\ x_k - x_{k+1} & x_{k+1} - x_{k+2} & \cdots & x_{2k-1} - x_{2k} & t^k \end{vmatrix}$$

We further define:

$$\underline{P}_n(t) = \begin{cases} \underline{\Delta}_n(t)^2 & \text{if } n \text{ is even,} \\ t\,\underline{\Delta}_n(t)^2 & \text{if } n \text{ is odd;} \end{cases}$$

$$\overline{P}_n(t) = \begin{cases} t(1-t)\,\overline{\Delta}_n(t)^2 & \text{if } n \text{ is even,} \\ (1-t)\,\overline{\Delta}_n(t)^2 & \text{if } n \text{ is odd.} \end{cases}$$

$\underline{P}_n(t)$ and $\overline{P}_n(t)$ are obviously in P^n, except for a positive normalizing factor.

THEOREM 19.1 *If $x_{(n-1)}$ is interior to D^{n-1}, then the unique supporting hyperplanes to D^n at $\underline{x}$ and $\overline{x}$ are given respectively by $\underline{P}_n(t)$ and $\overline{P}_n(t)$.*

Proof. We give the proof for $\underline{x}$, with $n = 2m+1$, as a typical case. By definition, the determinant $\underline{\Delta}_n$ of the point $\underline{x}$ is zero. Hence an $m+1$-tuple

$$\beta = (\beta_0, \beta_1, \ldots, \beta_m) \neq 0$$

exists with

$$\sum_{j=0}^{m} \underline{x}_{j+k+1}\beta_j = 0 \qquad k = 0, 1, \ldots, m.$$

Let

$$R(t) = \sum_{j=0}^{m} \beta_j t^j$$

and define y in P^n by

$$\sum_{i=0}^{n} y_i t^i = t\, R(t)^2 \cdot \text{const.}$$

Then

$$y \cdot \underline{x} = \sum_{j=0}^{m} \sum_{k=0}^{m} \underline{x}_{j+k+1} \beta_j \beta_k = 0,$$

showing that y is the unique supporting hyperplane at $\underline{x}$. We must show that the polynomials y and $P_n(t)$ are the same or, equivalently, that $R(t)$ and $\Delta_n(t)$ are the same (in both cases up to a constant factor). The matrix product of β (as a row matrix) with the matrix of $\Delta_n(t)$ is equal to

$$(0, 0, \ldots, 0, R(t));$$

hence every root of $R(t)$ is a root of $\Delta_n(t)$. But

$$r'(y) = \frac{a(\underline{x}) + 1}{2} = m + 1/2,$$

by Theorems 11.5 and 11.6 and the fact that (by hypothesis) $a(\underline{x}) = n-1 = 2m$. This tells us that y has m distinct roots besides the root at $t=0$, and consequently that $R(t)$ has m distinct roots. Since $\Delta_n(t)$ is, like $R(t)$, a polynomial of m-th degree, it is therefore the same as $R(t)$ up to a constant factor. This completes the proof.

Theorem 19.1 is restricted to points interior to D^n, or in the $(n-1)$-faces C^n_{n-1} and $\bar{C}^n_{n-1}$, these being just the points whose "upper" and

"lower" supporting planes are unique. The associated points $\underline{x}$ and $\bar{x}$ are no longer distinct if x is in a lower-dimensional face of D^n. If x is in C^n_{n-2} then there is a one-dimensional set of supporting hyperplanes at $\underline{x} = \bar{x} = x$, and the polynomials $\underline{P}_n$ and $\bar{P}_n$ represent the extreme points of that set. If x is in C^n_{n-3}, then the supporting hyperplanes form a two-dimensional convex set resembling P^2 (see Figure P.2), whose extreme points consist of a curve and an isolated point. One of the two polynomials $\underline{P}_n$ and $\bar{P}_n$ represents the isolated point; the other vanishes identically. For x in C^n_{n-4} and lower faces, both polynomials vanish identically.

As a corollary to the preceding theorem, we may state:

THEOREM 10.2 *If $x_{(n-1)}$ is interior to D^{n-1}, then all the zeros of $\underline{P}_n(t)$ and $\bar{P}_n(t)$ are real, distinct, and interior to $(0,1)$.*

This result is also valid for $x_{(n-1)}$ not interior to D^{n-1}, except that zeros may occur also at 0 or 1, or one or both polynomials may vanish identically. We might also remark that, unless they vanish identically, the polynomials always have the maximum degree -- namely $n/2$ and $(n-2)/2$ if n is even; $(n-1)/2$ and $(n-1)/2$ if n is odd.

§20. Construction of distribution functions.

By Theorem 11.1 we have at once:

THEOREM 20.1 *The points in D^n to which unique distribution functions in $\mathcal{A}$ correspond are precisely the boundary points.*

If If $\phi \in \mathcal{A}$ corresponds to $x \in D^n$, then we designate by $\underline{\phi}$ and $\overline{\phi}$ the unique distribution functions in $\mathcal{A}$ corresponding to the associated boundary points $\underline{x}$ and $\overline{x}$. The functions $\underline{\phi}$ and $\overline{\phi}$ thus agree with each other and with ϕ in their first $n-1$ moments, but differ with each other in their n-th moments unless $\underline{\phi} = \overline{\phi} = \phi$.[11]

By Theorem 19.1 we have:

THEOREM 20.2 *If $x_{(n-1)}$ is interior to D^{n-1}, then $\underline{\phi}$ and $\overline{\phi}$ are arithmetic distribution functions whose steps occur at the roots of $P_n(t)$ and $\overline{P}_n(t)$ respectively.*

This theorem provides us with an effective means of constructing distribution functions corresponding to points in the moment spaces. Indeed, suppose first that x is in the boundary of D^n, and take

$$m = a(x) + 1 \leq n.$$

Then $x_{(m-1)}$ is interior to D^{m-1} and $x_{(m)}$ is in the lower (upper) boundary of D^m. The roots t_j of P_m ($\overline{P}_m$) determine the location of the steps, and the linear system

[redacted]

[11] Since $\underline{\phi}$ and $\overline{\phi}$ depend on n, as well as on ϕ, we shall only use this notation when n has a fixed value in the discussion.

$$x_i = \sum_{j=1}^{b(\underline{x})} \xi_j t_j^i \qquad i = 0, 1, \ldots, m$$

determines the jumps ξ_j. The unique distribution function corresponding to the point x is then

$$\phi = \sum_{j=1}^{b(\underline{x})} \xi_j I(t-t_j).$$

If the given point x is interior to D^n, then the way to construct "minimal" distribution functions ($b' = n/2$) is to use $P_{n+1}(t)$ and $\bar{P}_{n+1}(t)$ (neither of which depends on the $n+1^{st}$ moment) in conjunction with Theorem 20.2. Another construction makes use of the fact that x is a convex combination of $\underline{x}$ and $\bar{x}$: one can verify by means of (18.1) and (18.2) that

$$\phi = \frac{(\bar{\Delta}_n / \bar{\Delta}_{n-2})\underline{\phi} + (\underline{\Delta}_n / \underline{\Delta}_{n-2})\bar{\phi}}{(\bar{\Delta}_n / \bar{\Delta}_{n-2}) + (\underline{\Delta}_n / \underline{\Delta}_{n-2})}$$

is a distribution function with the moments $1, x_1, \ldots, x_n$. Here $b'(\phi) = n$.

§21. Characterization of the set of distribution functions with given moments.

In general there will be many distribution functions having the moments $x_0, x_1, \ldots, x_n$. These form a convex set in $\mathcal{L}$ which we denote by $\mathcal{L}|x$. We first consider the convex subset $\mathcal{L}_A|x$ of arithmetic distribution functions in $\mathcal{L}|x$.

THEOREM 21.1 The extreme points of $\mathcal{S}_A|x$ are those functions ϕ with $b(\phi) \leq n + 1$.

Proof. An arbitrary arithmetic distribution ϕ can be represented:

$$\phi(t) = \sum_{j=1}^{b(\phi)} \xi_j I(t - t_j)$$

(see (6.1)). If ϕ is in $\mathcal{S}_A|x$ then we have

$$\sum_{j=1}^{b(\phi)} \xi_j t_j^i = x_i \qquad i = 0, 1, \ldots, n. \tag{21.1}$$

Since the t_j are all distinct, the rank of the system (21.1) is $\min(b(\phi), n+1)$. The dimension of the manifold of solutions $\xi = (\xi_1, \ldots, \xi_{b(\phi)})$ of (21.1) is therefore

$$b(\phi) - \min(b(\phi), n+1) = \max(0, b(\phi) - n-1).$$

Every non-negative solution of (21.1) corresponds to a point in $\mathcal{S}_A|x$. The solution corresponding to ϕ itself is strictly positive. Hence, if it is not the only solution it will be expressible as a convex combination of other non-negative solutions. But then ϕ will be the same convex combination of the corresponding points in $\mathcal{S}_A|x$. Therefore, if $b(\phi) > n+1$ then ϕ is not extreme.

On the other hand, any convex representation of ϕ must involve functions with spectra contained in the spectrum of ϕ; hence functions corresponding to non-negative solutions of (21.1). But if $b(\phi) \leq n+1$ then the

solution of (21.1) is unique, and ϕ must consequently be extreme. This completes the proof.

THEOREM 21.2 $\mathfrak{K}_A | x$ is spanned by its extreme points.

Proof. Theorem 3.1 does not apply in general to infinite-dimensional convex sets. However, it is clear from the proof just given that we can span any non-extreme ϕ in $\mathfrak{K}_A | x$ by a set of step-functions each having actually fewer jumps than ϕ. If any of these is not extreme, then we replace it by functions having still fewer jumps. After a finite number of such reductions we obtain a finite, convex representation of ϕ by extreme points only.

One might wish to regard the full set $\mathfrak{K} | x$ as being spanned by the same extreme points, making use of infinite convex representations of some sort. This becomes permissible if we adopt the weak * topology in $\mathfrak{K}$, for in that topology $\mathfrak{K}_A | x$ is dense in $\mathfrak{K} | x$. (The weak * topology on $\mathfrak{K}$ may be defined by the neighborhoods

$$N(\varepsilon, f_1, \ldots, f_m; \phi), \qquad \varepsilon > 0, \quad \phi \in \mathfrak{K}, \tag{21.2}$$

where $\{f_h(t)\}$ is any finite set of functions continuous on $[0,1]$; ϕ' is in the neighborhood (21.2) if and only if

$$\left| \int_0^1 f_h(t)\, d\phi'(t) - \int_0^1 f_h(t)\, d\phi(t) \right| < \varepsilon \qquad h = 1, \ldots, m. \tag{21.3}$$

In fact, consider a fixed neighborhood (21.2) of a fixed ϕ in $\mathfrak{K} | x$, and choose a set $\{R_h\}$ of polynomials satisfying

$$|R_h(t) - f_h(t)| < \varepsilon/2, \quad \text{all } t \in [0,1], \quad h = 1, \ldots, m.$$

Then construct a ϕ' in $\mathcal{D}_A|x$ with moments satisfying

$$\mu_i(\phi') = \mu_i(\phi) \qquad i = 0, 1, \ldots, \max(d_1, \ldots, d_m, n)$$

where d_h is the degree of $R_h(t)$. Since

$$\int_0^1 R_h(t)\,d\phi'(t) = \int_0^1 R_h(t)\,d\phi(t) \qquad h = 1, \ldots, m,$$

it is evident that this ϕ' satisfies (21.3). It follows that $\mathcal{D}_A|x$ is weak * dense in $\mathcal{D}|x$. This proves:

THEOREM 21.3 $\mathcal{D}|x$ is spanned in the weak * topology by the extreme points of $\mathcal{D}_A|x$.

For the special case n=0 this theorem describes how $\mathcal{D}$ is spanned by the pure distribution functions $I(t-t')$, $0 \le t' \le 1$, as mentioned in §6.

§22. Intertwining property of functions in $\mathcal{D}|x$.

THEOREM 22.1 If ϕ and ϕ' are distinct functions in $\mathcal{D}|x$, then the difference

$$\psi(t) = \phi(t) - \phi'(t)$$

has at least n sign changes in $[0,1]$.

$\underline{C}_n^{n+k}$ or $\overline{C}_n^{n+k}$ into E^n. The unique representations in the higher sets go into "minimal" representations in C_n^n, involving exactly $b^1 = (n+1)/2$ extreme points. Each x in C_n^n has precisely two such minimal representations, formally identical with the representations of the inverse images $\underline{x}_{(n+k)}$ and $\overline{x}_{(n+k)}$ of x in $\underline{C}_n^{n+k}$ and $\overline{C}_n^{n+k}$, respectively.

Considering the proof of the preceding theorem, we can therefore assert:

THEOREM 11.8 *The interior of* D^{2m-1} *is swept out by an* m*-parameter family of* $(m-1)$*-dimensional simplexes* $\equiv^{m-1}$, *and, in a different way, by an* $(m-1)$*-parameter family of* m*-dimensional simplexes* $\equiv^m$. *The interior of* D^{2m} *is swept out in two different ways by* m*-parameter families of* m*-dimensional simplexes* $\equiv^m$.

We shall make use of this parametrization in § 15.

CHAPTER III

THE SIMPLEXES S^n AND B^n

§ 12. The simplex S^n

The moment space D^n and its dual P^n are closely related to a pair of dual simplexes, with whose aid we are able to geometrize some classical properties of moment sequences and positive polynomials. The first simplex, circumscribed about D^n, is defined by its vertices $x^{(k)}$, $k = 0, 1, \ldots n$, with

$$x_i^{(k)} = \frac{k!(n-i)!}{n!(k-i)!} = \frac{\binom{k}{i}}{\binom{n}{i}},$$

where, as usual, $1/(k-i)! = \binom{k}{i} = 0$ if $i > k$. Displayed as a matrix, these coordinates are:

$$(12.1) \qquad \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & \frac{1}{n} & 0 & \ldots & 0 & 0 \\ 1 & \frac{2}{n} & \frac{2\cdot 1}{n(n-1)} & \ldots & 0 & 0 \\ \ldots & & \ldots & & & \ldots \\ 1 & \frac{n-1}{n} & \frac{n-2}{n} & \ldots & \frac{1}{n} & 0 \\ 1 & 1 & 1 & \ldots & 1 & 1 \end{pmatrix}$$

We shall denote this simplex by S^n. Figure 12.1 shows S^2 in relation to L^2.

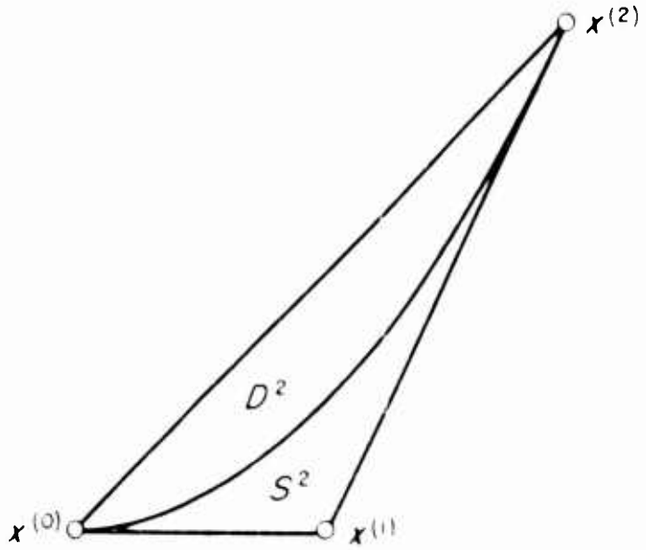


Fig. 12.1

THEOREM 12.1 S^n *contains* D^n.

Proof. The key to the proof is the identity

$$t^i = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \frac{\binom{k}{i}}{\binom{n}{i}} \qquad i = 0, 1, \ldots, n. \tag{12.2}$$

Accepting this, we obtain by direct substitution:

$$x(t) = \sum_{k=0}^{n} \xi_k x^{(k)}, \tag{12.3}$$

with

$$\xi_k = \binom{n}{k} t^k (1-t)^{n-k} \geq 0.$$

A simple calculation for (12.2) with $i=0$ gives us

$$\sum_{k=0}^{n} \xi_k = 1.$$

This shows that the extreme points $x^{(k)}$ of S^n span the extreme points $x(t)$ of D^n, and hence that S^n contains D^n.

To verify the identity (12.2) we observe that

$$\binom{n}{k}\binom{k}{i} = \binom{n-i}{k-i}\binom{n}{i}.$$

The right-hand member of (12.2) may therefore be written:

$$\sum_{k=0}^{n} \binom{n-i}{k-i} t^k (1-t)^{n-k}. \tag{12.4}$$

With the substitutions $\ell = k-i$, $m = n-i$, (12.4) becomes

$$\sum_{\ell=-i}^{m} \binom{m}{\ell} t^{\ell+i}(1-t)^{m-\ell} = t^i \sum_{\ell=0}^{m} \binom{m}{\ell} t^{\ell}(1-t)^{m-\ell} = t^i,$$

[redacted] distribution function ϕ in [redacted], certain quantities

$\lambda_{nk}(\phi)$ were introduced by Hausdorff [11], using the successive differences $\Delta^j \mu_k$ of the sequence of moments of ϕ.

$$\lambda_{nk}(\phi) = \binom{n}{k}\Delta^{n-k}\mu_k(\phi) = \binom{n}{k}\sum_{\ell=0}^{n-k}(-1)^{\ell}\binom{n-k}{\ell}\mu_{k+\ell}(\phi). \tag{12.5}$$

[redacted]

(See also Widder [9], pp. 100 ff; Shohat and Tamarkin [12], pp. 93 ff.)

These quantities have a simple geometric interpretation. Applying (6.2) to the right-hand expression above gives us

$$\lambda_{nk}(\phi) = \int_0^1 \binom{n}{k} t^k(1-t)^{n-k}\,d\phi(t). \tag{12.6}$$

If we denote by x_i the moments of ϕ, we then have

$$x = \int_0^1 x(t)\,d\phi(t) = \sum_{k=0}^{n} \lambda_{nk}(\phi)x^{(k)},$$

by (12.3) and (12.6), with

$$\lambda_{nk} \geq 0, \quad \sum_{k=0}^{n} \lambda_{nk} = 1.$$

Thus, the $\lambda_{nk}(\phi)$ are just the barycentric coordinates of x in the simplex S^n.

§13. The dual simplex B^n.

Turning to the dual space, we consider the "Bernstein" polynomials:

$$B_{nk}(t) = \binom{n}{k} t^k(1-t)^{n-k}, \qquad k = 0, 1, \ldots, n. \tag{13.1}$$

By Theorem 9.2 these polynomials are, up to a positive factor, extreme points of P^n. In fact we have

$$(n+1)B_{nk} \in \begin{cases} \underline{Q}^n & \text{if } n-k \text{ is even;} \\ \overline{Q}^n & \text{if } n-k \text{ is odd.} \end{cases}$$

Denoting the point $(n+1)B_{nk}$ by $y^{(k)}$ we obtain the following expressions for its coordinates:

$$y_i^{(k)} = (-1)^{i-k} \frac{(n+1)!}{(n-i)!(i-k)!k!} .$$

The matrix of the $y_i^{(k)}$ is substantially the inverse of the matrix (12.1), for we have

$$(13.2) \qquad y^{(k)} . x^{(\ell)} = \begin{cases} n+1 & \text{if } k = \ell ; \\ 0 & \text{if } k \neq \ell . \end{cases}$$

The simplex spanned by the n+1 points $y^{(k)}$ we denote by B^n; it is obviously inscribed in P^n. Figure 13.1 indicates the configuration for $n = 2$.

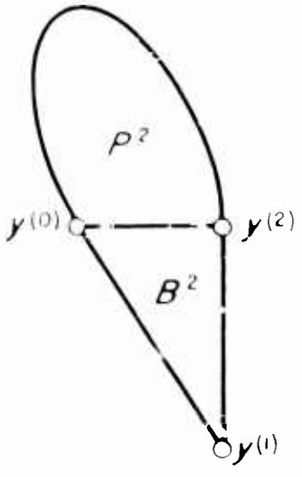


Fig. 13.1

THEOREM 13.1 B^n *is dual to* S^n.

PROOF. It suffices to show that the vertices $x^{(\ell)}$ of S^n, interpreted in the dual space, are the hyperplanes containing the $(n-1)$-dimensional faces of B^n. But, by (13.2), each hyperplane $x^{(\ell)}$ contains exactly n of the vertices $y^{(k)}$ of B^n, and the result follows.

It is now evident that the hyperplanes which determine the $(n-1)$-dimensional faces of S^n are supporting hyperplanes to D^n as well, so that in some sense S^n is the most closely fitting simplex that can be circumscribed about D^n. Dually, B^n is in some sense a maximal inscribed simplex in P^n. A precise meaning to these statements will be given in the next two sections.

A simple calculation shows that the centroids of the two simplexes S^n and B^n are respectively

$$\widetilde{x} = (1, \tfrac{1}{2}, \ldots, \tfrac{1}{n+1})$$

and

$$\widetilde{y} = (1, 0, \ldots, 0).$$

These will be recognized from §8 as the normalizing vectors used in selecting the cross-sections P^n and D^n of the conjugate, convex cones $\Gamma(P^n)$ and $\Gamma(D^n)$. They correspond to the "rectangular" distribution $\phi(t) = t$ and the constant polynomial $P(t) = 1$.

§14. Fit of S^n and B^n to D^n and P^n.

We have already noted that P^n is a cross-section of P^{n+1}. However, the corresponding cross-section of the inscribed simplex B^{n+1} is not [illegible]

B^n, but a larger, polyhedral body, more nearly filling out the interior of P^n.

THEOREM 14.1 *If y is interior to P^n, then for sufficiently large m the point*

$$y_{(m)} = (y_0, y_1, \ldots, y_n, 0, \ldots, 0)$$

is in B^m.

Proof[8]. Denote by $x_{(m)}^{(k)}$ the vertices of S^m, and hence also the hyperplanes in $(E^m)^*$ carrying the boundary of B^m. We must show that

$$y_{(m)} \cdot x_{(m)}^{(k)} \geq 0, \qquad k = 0, 1, \ldots, n \tag{14.1}$$

holds (for fixed n) if m is taken sufficiently large. Writing $tm = k$, we have for each $i = 0, 1, \ldots, n$:

$$x_{(m)}^{(k)} = \frac{tm(tm-1) \ldots (tm-i+1)}{m(m-1) \ldots (m-i+1)} \; ;$$

by (12.1). As m increases this converges (uniformly in t) to t^i. But from the hypothesis and Theorem 9.1, we have:

$$\sum_{i=0}^{n} y_i t^i \geq \delta > 0, \qquad 0 \leq t \leq 1.$$

It follows that (14.1) holds for sufficiently large m, as was to be shown.

---

8 See Hausdorff [11], p. 224.

The proof has shown that as m increases the vertices $x^{(k)}$ of S^m tend to the points $x(k/m)$ on C^m. The projections of S^m on E^n (n fixed) form a nested sequence of polyhedral, convex bodies tending to D^n. However, the maximum distance of S^m to D^m does not tend to zero as m increases. For example,

$$\left| x_m((m-1)/m) - x_m^{(m-1)} \right| = (1-1/m)^m \to 1/e.$$

We shall also see in the next section that the volumes of S^m and D^m do not approach each other asymptotically.

THEOREM 14.2 *Given the sequence*

$$x_0, x_1, \ldots, x_n, \ldots ,$$

then $x_{(m)} = (x_0, x_1, \ldots, x_m)$ *is in* S^m *for all* m *if and only if* x *is in* D^m *for all* m.

Proof. This is essentially the dual form of Theorem 14.1; it follows from it at once by means of Theorems 12.1 and 4.3.

Translating these theorems out of geometric terminology we obtain two well-known results:

THEOREM 14.3 *Any polynomial positive on* $[0,1]$ *can be represented as a finite sum with positive coefficients of polynomials*

$$B_{nk}(t) = \binom{n}{k} t^k (1-t)^{n-k}$$

of sufficiently high degree.

THEOREM 14.4 A necessary and sufficient condition that

$$\{x_0, x_1, \ldots, x_n, \ldots\}$$

be the moments of some distribution function on $[0,1]$ is that the sequence be "completely monotonic", i.e., that the x_i and their successive differences $\triangle^k x_i$ of all orders be all non-negative.

The latter follows from the expression (12.5) relating the $\triangle^k x_i$ to the barycentric coordinates of x in S^{k+i}.

§15. Comparison of volumes.

We now calculate the n-dimensional volumes of S^n and D^n, to obtain further insight into the relation between the two bodies. The result is somewhat special -- unlike our other results, for example, the extension to moment spaces based on the general finite interval $[a,b]$, though not difficult, is not immediately apparent. But we feel that the remarkably neat formula for vol D^n in terms of the beta and gamma functions justifies the inclusion of a portion of the calculation.

THEOREM 15.1 The n-dimensional volume of S^n is

$$\text{vol } S^n = \frac{1}{n!} \prod_{k=0}^{n} \binom{n}{k}^{-1} = \prod_{k=1}^{n} \frac{\Gamma(k)\,\Gamma(k)}{\Gamma(n+1)} .$$

Proof. We omit this elementary calculation.

THEOREM 15.2 The n-dimensional volume of D^n is

$$\text{vol } D^n = \prod_{k=1}^{n} B(k, k) = \prod_{k=1}^{n} \frac{\Gamma(k)\,\Gamma(k)}{\Gamma(2k)} .$$

Proof. We outline the proof for the case $n = 2m-1$. By Theorem 11.8 we may regard the interior of D^n as the product of the interiors of U^m and Ξ^{m-1}, defined by (11.4) and (11.5). It is convenient to replace U^m by the unit m-cube I^m, thereby multiplying the volume by $m!$. Thus

$$\text{vol } D^n = \int_{D^n} dx_1 \dots dx_n = \frac{1}{m!} \int_{I^m} \int_{\Xi^{m-1}} |J|\, d\xi_1 \dots d\xi_{m-1} du_1 \dots du_m ;$$

J being the Jacobian of the transformation (11.4) with ξ_m replaced by $1 - \xi_1 - \dots - \xi_{m-1}$. A calculation yields

$$|J| = \prod_{j=1}^{m} \xi_j \prod_{j=2}^{m} \prod_{k=1}^{j-1} (u_k - u_j)^4 ,$$

enabling us to split up the integral. Thus:

$$\text{vol } D^n = \frac{1}{m!} \int_{\Xi^{m-1}} \prod \xi_j^{d}\xi_1 \dots d\xi_{m-1} \int_{I^m} \prod_{k<j} (u_k - u_j)^4 du_1 \dots du_m .$$

The first integral is $1/\Gamma(2m)$. By a result of Selberg [13] the value of the second is

$$\prod_{k=1}^{m} \frac{\Gamma(2k+1)\,\Gamma(2k-1)^2}{2\,\Gamma(2m+2k-2)} .$$

Combining these results yields:

$$\text{vol } D^n = \frac{1}{\Gamma(2m)} \prod_{k=1}^{m} \frac{\Gamma(2k+1)\left[\Gamma(2k-1)\right]^2}{2k\,\Gamma(2m+2k-2)} .$$

From this the theorem can be verified directly, or by induction on m. The case $n = 2m$ is treated similarly.

We are now in a position to compare the sizes of S^n and D^n, as n increases. A reasonable measure might be the n-th root of the ratio of volumes, which could be expected to approach 1 for large n, even if the ratio itself does not. However, it is not difficult to show, from the theorems just proved, that

$$\left(\frac{\text{vol } D^n}{\text{vol } S^n}\right)^{\frac{1}{n}} = O(2^{-n_e n/2}) \to 0.$$

This result makes an interesting contrast with Theorem 14.2.

CHAPTER IV

ALGEBRAIC DESCRIPTION OF THE MOMENT SPACES

§16. Moment sequences and quadratic forms.

We turn first to the question of whether a given finite sequence constitutes the first n moments of some function in $\mathcal{F}$ — that is, whether a given point

$$x = (x_0, x_1, \ldots, x_n)$$

is in D^n.[7]

By Theorem 8.1 we have

$$(16.1) \qquad x \in D^{2m} \iff x \cdot y \geq 0 \quad \text{for all } y \text{ in } P^{2m}$$

(considering first the even-dimensional case). It is equivalent, however, to have y range over just the extreme points of P^{2m}; thus:

$$(16.2) \qquad x \in D^{2m} \iff x \cdot y \geq 0 \quad \text{for all } y \text{ in } Q^{2m}.$$

We have seen in §9 that the polynomials of Q^{2m} all have one or the other of the two forms

$$(16.3) \qquad R_m(t)^2 \quad \text{or} \quad t(1-t)R_{m-1}(t)^2$$

(subscripts here indicate the degree). Furthermore, all polynomials (16.3), though not necessarily in Q^{2m}, are certainly in P^{2m}, or positive multiples of elements of P^{2m}. Thus (16.1) and (16.2) are equivalent to:

[7] The criterion of Theorem 14.4, of course, applies only to infinite sequences.

$x \in D^{2m} \iff x \cdot y \geq 0$ for all y such that $\sum_{i=0}^{2m} y_i t^i$ has either the form $\left[\sum_{j=0}^{m} \alpha_j t^j\right]^2$ or the form $t(1-t)\left[\sum_{j=0}^{m-1} \beta_j t^j\right]^2$.

Eliminating y gives us finally:

$$x \in D^{2m} \iff \sum_{j=0}^{m}\sum_{k=0}^{m} x_{j+k}\alpha_j\alpha_k \geq 0, \quad \text{all } (\alpha_0, \ldots, \alpha_m), \text{ and}$$

$$\sum_{j=0}^{m-1}\sum_{k=0}^{m-1} (x_{j+k+1}-x_{j+k+2})\beta_j\beta_k \geq 0, \quad \text{all } (\beta_0, \ldots, \beta_{m-1}).$$

We thus have the well-known result[10]:

THEOREM 16.1a A necessary and sufficient condition that x be a point of D^{2m} is that the two quadratic forms:

$$\sum_{j,k=0}^{m} x_{j+k}\alpha_j\alpha_k \quad \text{and} \quad \sum_{j,k=0}^{m-1} (x_{j+k+1}-x_{j+k+2})\beta_j\beta_k$$

be positive definite or semidefinite.

In an analogous manner we may establish:

THEOREM 16.1b A necessary and sufficient condition that x be a point of D^{2m+1} is that the quadratic forms:

[10] See for example Shohat and Tamarkin [12], p. .

$$\sum_{j,k=0}^{m} x_{j+k+1}\alpha_j\alpha_k \qquad \underline{\text{and}} \qquad \sum_{j,k=0}^{m} (x_{j+k}-x_{j+k+1})\beta_j\beta_k$$

be positive definite or semidefinite.

The proof of the next theorem follows the same lines as above, making use of Theorem 4.2.

THEOREM 16.2 A necessary and sufficient condition that x be an interior point of D^n is that the quadratic forms of Theorem 16.1a or 16.1b (whichever applies) be positive definite.

16. The determinants Δ_m and their relation to the faces of D^n.

We now introduce a special notation for the "Hankel" determinants associated with the quadratic forms of the last section:

$$\Delta_{2k} = \begin{vmatrix} 1 & x_1 & \cdots & x_k \\ \vdots & & & \\ x_k & x_{k+1} & \cdots & x_{2k} \end{vmatrix}$$

$$\Delta_{2k+1} = \begin{vmatrix} x_1 & x_2 & \cdots & x_{k+1} \\ \vdots & & & \\ x_{k+1} & x_{k+2} & \cdots & x_{2k+1} \end{vmatrix}$$

$$\overline{\Delta}_{2k} = \begin{vmatrix} x_1 - x_2 & x_2 - x_3 & \cdots & x_k - x_{k+1} \\ \vdots & & & \\ x_k - x_{k+1} & x_{k+1} - x_{k+2} & \cdots & x_{2k-1} - x_{2k} \end{vmatrix}$$

$$\overline{\Delta}_{2k+1} = \begin{vmatrix} 1 - x_1 & x_1 - x_2 & \cdots & x_k - x_{k+1} \\ \vdots & & & \\ x_k - x_{k+1} & x_{k+1} - x_{k+2} & \cdots & x_{2k} - x_{2k+1} \end{vmatrix}$$

The subscripts have been chosen so as to indicate the highest moment occurring in each determinant. The upper and lower bars will be seen in due course to agree with our previous usage.

THEOREM 17.1 A necessary and sufficient condition that the quadratic form

$$\sum_{i,j} a_{ij} \alpha_i \alpha_j$$

be positive definite is that the first principal minors of its symmetric matrix all be positive.

Note. By the "first principal minors" of an array such as the matrix of $\overline{\Delta}_{2k}$ given above, we mean the k+1 subdeterminants $\overline{\Delta}_0, \overline{\Delta}_2, \cdots \overline{\Delta}_{2k}$.

For a proof of this standard result, see for example Ferrer [4], page 138.

THEOREM 17.2 The point x is interior to D^n if and only if all of the determinants

(17.1) $$\underline{\Delta}_0, \overline{\Delta}_0, \underline{\Delta}_1, \overline{\Delta}_1, \cdots \underline{\Delta}_n, \overline{\Delta}_n$$

are positive.

Proof. By Theorems 16.2 and 17.1 x is interior to D^n if and only if $\underline{\Delta}_n$, $\overline{\Delta}_n$, $\underline{\Delta}_{n-2}$, $\overline{\Delta}_{n-2}$, etc., are positive. It only remains to remark that if x is interior to D^n, then the point

$$x_{(n-1)} = (x_0, x_1, \cdots, x_{n-1})$$

is necessarily interior to D^{n-1}, and the theorem follows at once.

Now suppose that x is in the boundary of D^n — say in the face $\overline{C}^n_{n-1}$ (see §11). It follows that $x_{(n-1)}$ is interior to D^{n-1}, so that all but the last two determinants (17.1) are necessarily positive, and one of these last two must vanish. Since the polynomial $\sum y_i t^i$ associated with the unique supporting hyperplane $y \in$ at x has a root at $t=1$, we can see that the second quadratic form of Theorem 16.1 (a or b) is positive semi-definite, while the first form is easily shown to be positive definite. Hence

$$\underline{\Delta}_n > 0, \qquad \overline{\Delta}_n = 0$$

If x had been in the face $\underline{C}^n_{n-1}$ the upper and lower bars would have been reversed.

Applying the same argument to the lower-dimensional faces gives us the following characterization theorem for the a-faces of D^n:

THEOREM 17.3 *The point* $x \in D^n$ *is in* C_a^n *if and only if*

$$\underline{\Delta}_0,\ \overline{\Delta}_0,\ \underline{\Delta}_1,\ \overline{\Delta}_1,\ \ldots,\ \underline{\Delta}_a,\ \overline{\Delta}_a,\ \text{and}\ \overline{\Delta}_{a+1}$$

are positive and

$$\underline{\Delta}_{a+1}\ \text{and}\ \underline{\Delta}_{a+2},\ \overline{\Delta}_{a+2},\ \ldots,\ \underline{\Delta}_n,\ \overline{\Delta}_n$$

are zero. Similarly x *is in* $\overline{C}_a^n$ *if and only if*

$$\underline{\Delta}_0,\ \overline{\Delta}_0,\ \underline{\Delta}_1,\ \overline{\Delta}_1,\ \ldots,\ \underline{\Delta}_a,\ \overline{\Delta}_a,\ \text{and}\ \underline{\Delta}_{a+1}$$

are positive and

$$\overline{\Delta}_{a+1}\ \text{and}\ \underline{\Delta}_{a+2},\ \overline{\Delta}_{a+2},\ \ldots,\ \underline{\Delta}_n,\ \overline{\Delta}_n$$

are zero.

All of the determinants of index a+2 and higher vanish because of the fact that C_a^n is contained in the closures of both C_m^n and $\overline{C}_m^n$, for $a < m \leq n$. (See (11.6)).

It is essential here to specify that x be in D^n, because the analogues of Theorems 17.2 and 17.1 for non-negative determinants do not hold. For an illustrative example, see Widder [9], pages 135-6.

§ 18. The lower and upper boundaries of D^n.

Equating Δ_n to zero gives us a particular value $\underline{x}_n$ for the n-th moment as a function of the preceding moments. We may write this relationship in the compact form:

$$\underline{x}_n = x_n - \frac{\Delta_n}{\Delta_{n-2}}, \tag{18.1}$$

which is actually independent of x_n, as it should be. In similar fashion, we see that

$$\bar{x}_n = x_n + \frac{\Gamma_n}{\Delta_{n-2}} \tag{18.2}$$

is the value for the n-th moment that makes Γ_n vanish.

For any point x in D^n we can define the associated points

$$\underline{x} = (x_0, x_1, \cdots, x_{n-1}, \underline{x}_n)$$

$$\bar{x} = (x_0, x_1, \cdots, x_{n-1}, \bar{x}_n).$$

Since we always have

$$\underline{x}_n \leq x_n \leq \bar{x}_n$$

we can interpret $\underline{x}$ and $\bar{x}$ as the projections "downward" and "upward" of x on the boundary of D^n.

Since, moreover, we always have by Theorem 17.3

$$\underline{x} \text{ in closure of } C^n_{n-1}$$

$$\bar{x} \text{ in closure of } \bar{C}^n_{n-1},$$

we shall refer to these two closed sets as respectively the lower and upper boundaries of D^n.

CHAPTER V

DISTRIBUTIONS HAVING GIVEN MOMENTS

§19. The polynomials $\underline{\Delta}_m(t)$ and $\overline{\Delta}_m(t)$.

Our next task will be to find out what can be said about the distribution functions which give rise to a given point x of D^n. We first need an explicit form for the hyperplanes which support D^n at the associated upper and lower boundary points $\underline{x}$ and $\overline{x}$ (see §16); this form is provided by the following polynomials, closely related to the determinants $\underline{\Delta}_m$ and $\overline{\Delta}_m$ of §17. We define:

$$\underline{\Delta}_{2k}(t) = \begin{vmatrix} 1 & x_1 & \cdots & x_{k-1} & 1 \\ & & & & t \\ \vdots & & & & \vdots \\ x_k & x_{k+1} & \cdots & x_{2k-1} & t^k \end{vmatrix}$$

$$\underline{\Delta}_{2k+1}(t) = \begin{vmatrix} x_1 & x_2 & \cdots & x_k & 1 \\ & & & & t \\ \vdots & & & & \vdots \\ x_{k+1} & x_{k+2} & \cdots & x_{2k} & t^k \end{vmatrix}$$

$$\overline{\Delta}_{2k}(t) = \begin{vmatrix} x_1 - x_2 & x_2 - x_3 & \cdots & x_{k-1} - x_k & \\ & & & & t \\ \vdots & & & & \vdots \\ x_k - x_{k+1} & x_{k+1} - x_{k+2} & \cdots & x_{2k-2} - x_{2k-1} & t^{k-1} \end{vmatrix}$$

$$\overline{\Delta}_{2k+1}(t) = \begin{vmatrix} 1 - x_1 & x_1 - x_2 & \cdots & x_{k-1} - x_k & 1 \\ & & & & t \\ \vdots & & & & \vdots \\ x_k - x_{k+1} & x_{k+1} - x_{k+2} & \cdots & x_{2k-1} - x_{2k} & t^k \end{vmatrix}$$

We further define:

$$\underline{P}_n(t) = \begin{cases} \underline{\Delta}_n(t)^2 & \text{if } n \text{ is even,} \\ t\ \underline{\Delta}_n(t)^2 & \text{if } n \text{ is odd;} \end{cases}$$

$$\overline{P}_n(t) = \begin{cases} t(1-t)\ \overline{\Delta}_n(t)^2 & \text{if } n \text{ is even,} \\ (1-t)\ \overline{\Delta}_n(t)^2 & \text{if } n \text{ is odd.} \end{cases}$$

$\underline{P}_n(t)$ and $\overline{P}_n(t)$ are obviously in P^n, except for a positive normalizing factor.

THEOREM 10.1 *If $x_{(n-1)}$ is interior to P^{n-1}, then the unique supporting hyperplanes to P^n at $\underline{x}$ and $\overline{x}$ are given respectively by $\underline{P}_n(t)$ and $\overline{P}_n(t)$.*

Proof. We give the proof for $\underline{x}$, with $n = 2m+1$, as a typical case. By definition, the determinant $\underline{\Delta}_n$ of the point $\underline{x}$ is zero. Hence an $m+1$-tuple

$$\beta = (\beta_0, \beta_1, \ldots, \beta_m) \neq 0$$

exists with

$$\sum_{j=0}^{m} x_{j+k+1}\beta_j = 0 \qquad k = 0, 1, \ldots, m.$$

Let

$$R(t) = \sum_{j=0}^{m} \beta_j t^j$$

and define y in P^n by

$$\sum_{i=0}^{n} y_i t^i = t\, R(t)^2 \cdot \text{const.}$$

Then

$$y \cdot \underline{x} = \sum_{j=0}^{m} \sum_{k=0}^{m} \underline{x}_{j+k+1} \beta_j \beta_k = 0,$$

showing that y is the unique supporting hyperplane at $\underline{x}$. We must show that the polynomials y and $\underline{P}_n(t)$ are the same or, equivalently, that $R(t)$ and $\underline{P}_n(t)$ are the same (in all cases up to a constant factor). The matrix product of y (as a row matrix) with the matrix of $\underline{P}_n(t)$ is equal to

$$(0, 0, \ldots, 0, R'(t));$$

hence every root of $R(t)$ is a root of $\underline{P}_n(t)$. But

$$r'(y) = \frac{a(\underline{x}) + 1}{2} = m + 1/2,$$

by Theorems 11.5 and 11.6 and the fact that (by hypothesis) $a(\underline{x}) = n-1 = 2m$. This tells us that y has m distinct roots besides the root at $t=0$, and consequently that $R(t)$ has m distinct roots. Since $\underline{P}_n(t)$ is, like $R(t)$, a polynomial of m-th degree, it is therefore the same as $R(t)$ up to a constant factor. This completes the proof.

Theorem 19.1 is restricted to points interior to D^n, or in the $(n-1)$-faces C^n_{n-1} and $\bar{C}^n_{n-1}$, these being just the points whose "upper" and

"lower" supporting planes are unique. The associated points $\underline{x}$ and $\bar{x}$ are no longer distinct if x is in a lower-dimensional face of D^n. If x is in C^n_{n-2} then there is a one-dimensional set of supporting hyperplanes at $\underline{x} = \bar{x} = x$, and the polynomials $\underline{P}_n$ and $\bar{P}_n$ represent the extreme points of that set. If x is in C^n_{n-3}, then the supporting hyperplanes form a two-dimensional convex set resembling P^2 (see Figure 7.2), whose extreme points consist of a curve and an isolated point. One of the two polynomials $\underline{P}_n$ and $\bar{P}_n$ represents the isolated point; the other vanishes identically. For x in C^n_{n-4} and lower faces, both polynomials vanish identically.

As a corollary to the preceding theorem, we may state:

THEOREM 10.2 *If $x_{(n-1)}$ is interior to D^{n-1}, then all the zeros of $\underline{P}_n(t)$ and $\bar{P}_n(t)$ are real, distinct, and interior to $(0,1)$.*

This result is also valid for $x_{(n-1)}$ not interior to D^{n-1}, except that zeros may occur also at 0 or 1, or one or both polynomials may vanish identically. We might also remark that, unless they vanish identically, the polynomials always have the maximum degree -- namely $n/2$ and $(n-2)/2$ if n is even; $(n-1)/2$ and $(n-1)/2$ if n is odd.

§20. Construction of distribution functions.

By Theorem 11.1 we have at once:

THEOREM 20.1 The points in D^n to which unique distribution functions in correspond are precisely the boundary points.

If If ϕ corresponds to $x \in D^n$, then we designate by $\underline{\phi}$ and $\overline{\phi}$ the unique distribution functions in corresponding to the associated boundary points $\underline{x}$ and $\overline{x}$. The functions $\underline{\phi}$ and $\overline{\phi}$ thus agree with each other and with ϕ in their first n-1 moments, but differ with each other in their n-th moments unless $\underline{\phi} = \overline{\phi} = \phi$.[1]

By Theorem 19.1 we have:

THEOREM 20.2 If $x_{(n-1)}$ is interior to D^{n-1}, then $\underline{\phi}$ and $\overline{\phi}$ are arithmetic distribution functions whose steps occur at the roots of $P_n(t)$ and $\overline{P}_n(t)$ respectively.

This theorem provides us with an effective means of constructing distribution functions corresponding to points in the moment spaces. Indeed, suppose first that x is in the boundary of D^n, and take

$$m = a(x) + 1 \leq n.$$

Then $x_{(m-1)}$ is interior to D^{m-1} and $x_{(m)}$ is in the lower (upper) boundary of D^m. The roots t_i of P_m ($\overline{P}_m$) determine the location of the steps, and the linear system

[1] Since $\underline{\phi}$ and $\overline{\phi}$ depend on n, as well as on ϕ, we sh[illegible] only use this notation when n has a fixed value in the discussion.

$$x_i = \sum_{j=1}^{b(x)} \xi_j t_j^i \qquad i = 0, 1, \ldots, m$$

determines the jumps ξ_j. The unique distribution function corresponding to the point x is then

$$\phi = \sum_{j=1}^{b(x)} \xi_j I(t-t_j).$$

If the given point x is interior to D^n, then the way to construct "minimal" distribution functions ($b' = n/2$) is to use $P_{n+1}(t)$ and $\bar{P}_{n+1}(t)$ (neither of which depends on the $n+1^{st}$ moment) in conjunction with Theorem 20.2. Another construction makes use of the fact that x is a convex combination of $\underline{x}$ and $\bar{x}$: one can verify by means of (18.1) and (18.2) that

$$\phi = \frac{(\bar{\Delta}_n / \bar{\Delta}_{n-2})\underline{\phi} + (\underline{\Delta}_n / \underline{\Delta}_{n-2})\bar{\phi}}{(\bar{\Delta}_n / \bar{\Delta}_{n-2}) + (\underline{\Delta}_n / \underline{\Delta}_{n-2})}$$

is a distribution function with the moments $1, x_1, \ldots, x_n$. Here $b'(\phi) = n$.

21. Characterization of the set of distribution functions with given moments.

In general there will be many distribution functions having the moments $x_0, x_1, \ldots, x_n$. These form a convex set in $\mathcal{F}$ which we denote by $\mathcal{F}|x$. We first consider the convex subset $\mathcal{F}_A|x$ of arithmetic distribution functions in $\mathcal{F}|x$.

THEOREM 21.1 The extreme points of $\mathcal{S}_A|x$ are those functions ϕ with $b(\phi) \leq n + 1$.

Proof. An arbitrary arithmetic distribution ϕ can be represented:

$$\phi(t) = \sum_{i=1}^{b(\phi)} \xi_i I(t - t_i)$$

(see (6.1)). If ϕ is in $\mathcal{S}_A|x$ then we have

$$\sum_{j=1}^{b(\phi)} \xi_j t_j^i = x_i \qquad i = 0, 1, \ldots, n. \tag{21.1}$$

Since the t_j are all distinct, the rank of the system (21.1) is $\min(b(\phi), n+1)$. The dimension of the manifold of solutions $\xi = (\xi_1, \ldots, \xi_{b(\phi)})$ of (21.1) is therefore

$$b(\phi) - \min(b(\phi), n+1) = \max(0, b(\phi) - n-1).$$

Every non-negative solution of (21.1) corresponds to a point in $\mathcal{S}_A|x$. The solution corresponding to ϕ itself is strictly positive. Hence, if it is not the only solution it will be expressible as a convex combination of other non-negative solutions. But then ϕ will be the same convex combination of the corresponding points in $\mathcal{S}_A|x$. Therefore, if $b(\phi) > n+1$ then ϕ is not extreme.

On the other hand, any convex representation of ϕ must involve functions with spectra contained in the spectrum of ϕ; hence functions corresponding to non-negative solutions of (21.1). But if $b(\phi) \leq n+1$ then the

solution of (21.1) is unique, and ϕ must consequently be extreme. This completes the proof.

THEOREM 21.2 $\mathcal{K}_A x$ is spanned by its extreme points.

Proof. Theorem 3.1 does not apply in general to infinite-dimensional convex sets. However, it is clear from the proof just given that we can span any non-extreme ϕ in $\mathcal{K}_A x$ by a set of step-functions each having actually fewer jumps than ϕ. If any of these is not extreme, then we replace it by functions having still fewer jumps. After a finite number of such reductions we obtain a finite, convex representation of ϕ by extreme points only.

One might wish to regard the full set $\mathcal{K} x$ as being spanned by the same extreme points, making use of infinite convex representations of some sort. This becomes permissible if we adopt the weak * topology in $\mathcal{K}$, for in that topology $\mathcal{K}_A x$ is dense in $\mathcal{K} x$. (The weak * topology on $\mathcal{K}$ may be defined by the neighborhoods

$$N(\cdot, f_1, \ldots, f_m; \phi), \qquad \cdot > 0, \quad \phi \cdot \mathcal{K} , \tag{21.2}$$

where $\{f_h(t)\}$ is any finite set of functions continuous on $[0,1]$; ϕ' is in the neighborhood (21.2) if and only if

$$\left| \int_0^1 f_h(t)\,d\phi'(t) - \int_0^1 f_h(t)\,d\phi(t) \right| < \cdot \qquad h = 1, \ldots, m.) \tag{21.3}$$

In fact, consider a fixed neighborhood (21.2) of a fixed ϕ in $\mathcal{K} x$, and choose a set R_h of polynomials satisfying

$$|R_h(t) - f_h(t)| < \varepsilon/2, \quad \text{all } t \in [0,1], \quad h = 1, \ldots, m.$$

Then construct a ϕ' in $\mathcal{S}_A|x$ with moments satisfying

$$\mu_i(\phi') = \mu_i(\phi) \qquad i = 0, 1, \ldots, \max(d_1, \ldots, d_m, n)$$

where d_h is the degree of $R_h(t)$. Since

$$\int_0^1 R_h(t)\,d\phi'(t) = \int_0^1 R_h(t)\,d\phi(t) \qquad h = 1, \ldots, m,$$

it is evident that this ϕ' satisfies (21.3). It follows that $\mathcal{S}_A|x$ is weak * dense in $\mathcal{S}|x$. This proves:

THEOREM 21.3 *$\mathcal{S}|x$ is spanned in the weak * topology by the extreme points of $\mathcal{S}_A|x$.*

For the special case n=0 this theorem describes how $\mathcal{S}$ is spanned by the pure distribution functions $I(t-t')$, $0 \le t' \le 1$, as mentioned in §6.

§22. Intertwining property of functions in $\mathcal{S}|x$.

THEOREM 22.1 *If ϕ and ϕ' are distinct functions in $\mathcal{S}|x$, then the difference*

$$\psi(t) = \phi(t) - \phi'(t)$$

has at least n sign changes in $[0,1]$.

(Starting from any t' for which $\psi(t') \neq 0$, the next greater sign change of ψ may be defined as the greatest lower bound of all $t > t'$ such that $\psi(t)\,\psi(t') < 0$. The set of all sign changes is then found by considering all values of t'.)

Proof. Suppose ψ has fewer than n sign changes. Then there is a polynomial R_{n-1} of degree $n-1$ with $R_{n-1}(t)\,\psi(t)$ non-negative in $[0,1]$. Since ψ is right-continuous and not identically zero we have

$$\int_0^1 R_{n-1}(t)\,\psi(t)\,dt > 0. \tag{22.1}$$

Putting

$$R_n(t) = \int_0^t R_{n-1}(t)\,dt$$

and integrating (22.1) by parts gives us:

$$R_n(t)\,\psi(t)\Big]_{0-}^{1} - \int_0^1 R_n(t)\left[d\phi(t) - d\phi'(t)\right] > 0.$$

The second term vanishes because the first n moments of ϕ and ϕ' coincide. But the first term also vanishes since $\phi(0-) = \phi'(0-) = 0$, $\phi(1) = \phi'(1) = 1$. The resulting contradiction proves the theorem.

THEOREM 22.2 The differences $\phi - \phi$ and $\phi - \bar{\phi}$, if not identically zero, each have exactly $n-1$ sign changes in $[0,1]$.

Proof. We first observe that the number of sign changes in $[0,1]$ of $\psi - \phi_A$, where ψ is an arbitrary non-decreasing function and ϕ_A is in $\mathcal{D}_A$, can never exceed $2b'(\phi_A) - 1$. Now ϕ and $\bar{\phi}$, if they are to be

distinct from ϕ, must lie in C^n_{n-1}, so that $b'(\underline{\phi}) = b'(\overline{\phi}) = n/2$ (Theorem 11.6). Hence $\phi - \underline{\phi}$ and $\phi - \overline{\phi}$ have at most n-1 sign changes. However, $\underline{\phi}$, $\overline{\phi}$, and ϕ are all in $\mathcal{S}\,|\,x_{(n-1)}$. By the preceding theorem, therefore, $\phi - \underline{\phi}$ and $\phi - \overline{\phi}$ have at least n-1 sign changes. This completes the proof.

As a particular case of Theorem 22.2, we have:

THEOREM 22.3 **If $\underline{\phi}$ and $\overline{\phi}$ are distinct, then their spectra form strictly interlocking sets in $[0,1]$. The sets of values of $\underline{\phi}(t)$ and $\overline{\phi}(t)$ for $0 < t < 1$ are also strictly interlocking in $[0,1]$.**

The functions $\underline{\phi}$ and $\overline{\phi}$ control in a certain sense the shape of the other functions in $\mathcal{S}\,|\,x_{(n-1)}$, for the graph of each $\phi(t)$ must cross every step of both $\underline{\phi}(t)$ and $\overline{\phi}(t)$ (see Figure 22.1).

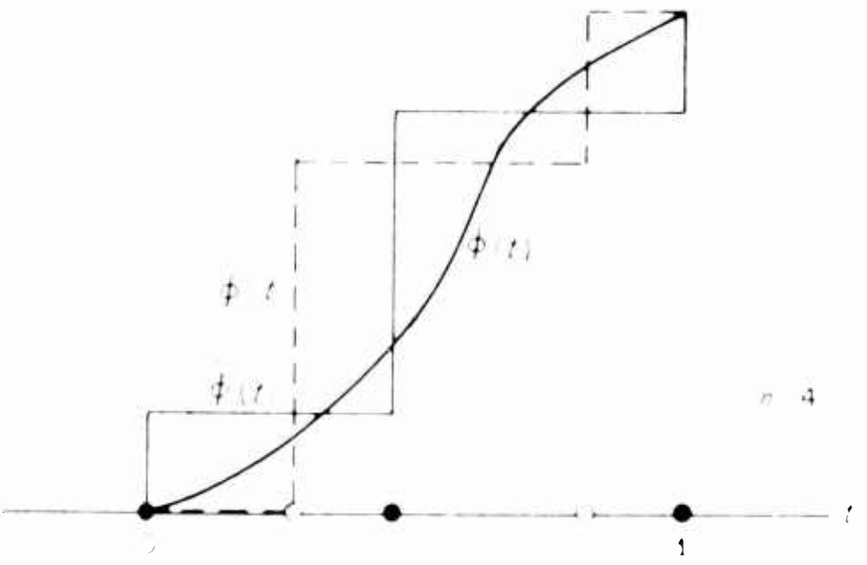


Fig 22 1

CHAPTER VI

APPLICATION TO ORTHOGONAL POLYNOMIALS

§ 23. Orthogonality of the Δ-polynomials.

THEOREM 23.1 *Given ϕ in $\mathcal{N}$ with the moments $1, x_1, \ldots, x_n, \ldots$, then the system of polynomials $\Delta_{2k}(t)$, $k = 0, 1, \ldots$, is orthogonal with respect to the weight function $d\phi(t)$, thus:*

$$\int_0^1 \Delta_{2k}(t)\,\Delta_{2\ell}(t)\,d\phi(t) = 0 \qquad \text{for } k \neq \ell. \tag{23.1}$$

Likewise $\Delta_{2k+1}(t)$, $\Delta_{2k+2}(t)$, and $\Delta_{2k+1}(t)$ are orthogonal with respect to $td\phi(t)$, $t(1-t)d\phi(t)$, and $(1-t)d\phi(t)$, respectively.

Proof. It is only necessary to observe that

$$\int_0^1 t^{\ell}\,\Delta_{2k}(t)\,d\phi(t) = \begin{vmatrix} 1 & \cdots & x_{k-1} & x_{\ell} \\ \vdots & & & \vdots \\ x_k & \cdots & x_{2k-1} & x_{k+\ell} \end{vmatrix} = 0 \tag{23.2}$$

for $\ell < k$. The three other cases work out similarly.

The four systems are not orthonormal as they stand. We have instead:

$$\int_0^1 \left[\Delta_{2k}(t)\right]^2 d\phi(t) = \Delta_{2k}\Delta_{2k-2}, \tag{23.3}$$

$$\int_0^1 \bar{L}_{2k+1}(t)^2 \, t \, d\phi(t) = \bar{\mu}_{2k+1}\bar{\mu}_{2k-1},$$

$$\int_0^1 \bar{L}_{2k-2}(t)^2 \, t(1-t) \, d\phi(t) = \bar{\mu}_{2k+2}\bar{\mu}_{2k},$$

$$\int_0^1 \bar{L}_{2k+1}(t)^2 \, (1-t) \, d\phi(t) = \bar{\mu}_{2k+1}\bar{\mu}_{2k-1},$$

as a simple calculation employing the orthogonality confirms.

We recall that the complete orthogonal system with respect to a given weight function is substantially unique.[1] Hence all of the classical systems (referred to the interval $[0,1]$) can be expressed by means of the polynomials $\bar{L}_{2k}(t)$. Using Theorem 10.1, we can identify these polynomials with the hyperplanes which support the even-dimensional moment spaces at the lower boundary points associated with the moment points of the given weight function. Of course, any one of the three other systems of Theorem 23.1 can also be used if we adapt the weight function in the proper way. This provides a geometrical approach to the theory of orthogonal polynomials which we shall illustrate in the next two sections, and develop more fully in a future paper.

If the weight function is arithmetic, the orthogonal systems of Theorem 23.1 are all finite, since all of the Δ-polynomials are identically zero beyond a certain point (see 19). If, conversely, $x_{(n)}$ is interior to D^n for every n, all four systems are infinite. For the remainder of this chapter we shall be concerned with the finite case.

We shall refer to the systems $\{\bar{L}_{2k}(t)\}$, $\{\bar{L}_{2k+2}(t)\}$, or

[1] See Szegö [10], p. 45.

$\{\Delta_{2k+1}(t)\}$, $\{\bar{\Delta}_{2k+1}(t)\}$ as associated, since it can be shown that they generalize the classical associated systems that arise from a differential equation. Corresponding polynomials in associated systems have interlocking sets of roots:

THEOREM 23.2 *The roots of the Δ-polynomials are real and distinct. Between every two roots of $\Delta_n(t)$ lies a root of $\bar{\Delta}_n(t)$, and conversely (n even or odd).*

Proof. By Theorems 19.2, 20.1, and 22.3.

§24 Separation of roots. Christoffel numbers.

In working with a finite segment $\Delta_0(t)$, $\Delta_2(t)$, ..., $\Delta_n(t)$ (n even) of an orthogonal system, we may use the weight functions $d\phi$, $d\underline{\phi}$, and $d\bar{\phi}$ interchangeably, since their first n-1 moments agree. This simple fact enables us to prove easily some of the standard properties of orthogonal polynomials. We give here two examples.

THEOREM 24.1 *There is a root of $\Delta_{n-2}(t)$ between every pair of roots of $\Delta_n(t)$; similarly for $\bar{\Delta}_{n-2}(t)$, $\bar{\Delta}_n(t)$.*

Proof. It suffices to consider $\Delta_{n-2}(t)$, $\Delta_n(t)$, with $n = 2m$, since the other cases can be put into this form by changing the weight function. Construct a polynomial $R(t)$ of degree $m - 2$ which vanishes at all m roots of $\Delta_n(t)$ except a consecutive pair t_1, t_2. Then

$$\int_0^1 R(t)\, \Delta_{n-2}(t)\, d\phi(t) = \lambda_1 R(t_1)\, \Delta_{n-2}(t_1) + \lambda_2 R(t_2)\, \Delta_{n-2}(t_2)$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$ are the jumps of the step-function ϕ at t_1 and t_2. But the integral vanishes, by (23.2). Also, $\lambda_1 R(t_1)$ and $\lambda_2 R(t_2)$ have like signs, and are not zero. Therefore $\Delta_{n-2}(t)$ must change sign between t_1 and t_2, as was to be shown.

The Christoffel numbers of a distribution function ϕ, relative to a fixed $n = 2m$, may be defined:

$$\lambda_j = \int_0^1 \frac{\Delta_n(t)}{\Delta_n'(t_j)(t-t_j)}\, d\phi(t), \qquad j = 1, \ldots, m,$$

where t_j are the roots of $\Delta_n(t)$. (Compare Szegö [10], pp. 46-52.) If we replace ϕ by ψ in this formula, we see that λ_j is exactly the jump of ψ at t_j:

$$\lambda_j = \psi(t_j) - \psi(t_j - 0). \tag{24.1}$$

If $R(t)$ is any polynomial of degree $n-1$ or less, we have at once:

$$\int_0^1 R(t)\, d\phi(t) = \int_0^1 R(t)\, d\psi(t) = \sum_{j=1}^{m} \lambda_j R(t_j)$$

— the Gauss-Jacobi quadrature formula.

THEOREM 24.2 *If* $t_1, \ldots, t_m$ *are the roots of* $\omega_{2m}(t)$, *then there exists an interlocking set* $u_1, \ldots, u_{m-1}$

$$t_j \leq u_j \leq t_{j+1}, \qquad j = 1, \ldots, m-1,$$

such that

$$\phi(u_j) \geq \sum_{k=1}^{j} \rho_k \geq \phi(u_j - 0).$$

Proof. From (24.1) we obtain

$$\phi(u) = \sum_{k=1}^{j} \rho_k \quad \text{for} \quad t_j \leq u < t_{j+1}.$$

The present theorem is therefore essentially a restatement of Theorem 22.2, regarding the intertwining of ϕ and ϕ.

25. Geometry of the Tchebycheff polynomials.

Throughout this section the asterisk (*) will identify quantities derived from the particular distribution function $\phi^* \in \Phi$ defined by

$$d\phi^*(t) = \frac{dt}{\sqrt{t(1-t)}}$$

Its moments are readily calculated by means of the beta function; the k-th moment is

$$x_k^* = \frac{2k-1}{2k} x_{k-1}^* = \binom{2k}{k} 4^{-2k}. \tag{25.1}$$

The orthogonal systems $\{\underline{\Delta}^*_{2K}(t)\}$ and $\{\overline{\Delta}^*_{2K}(t)\}$ based on ϕ^* are the so-called Tchetycheff polynomials of the first and second kinds, respectively; we have:

$$\text{const. } \underline{\Delta}^*_{2K}(t) = T_K(t) = \cos K\theta,$$

$$\text{const. } \overline{\Delta}^*_{2K}(t) = U_{K-1}(t) = \frac{\sin K\theta}{\sin \theta}, \tag{25.2}$$

where $\theta = \cos^{-1}(2t-1)$. The systems $\{\underline{\Delta}^*_{2K+1}(t)\}$ and $\{\overline{\Delta}^*_{2K+1}(t)\}$ based on ϕ^* are the Jacobi systems commonly denoted by $P_K^{(-\frac{1}{2}, \frac{1}{2})}$ and $P_K^{(\frac{1}{2}, -\frac{1}{2})}$; we have:

$$\text{const. } \underline{\Delta}^*_{2K+1}(t) = \frac{\cos (2K+1)\theta/2}{\cos(\theta/2)},$$

$$\text{const. } \overline{\Delta}^*_{2K+1}(t) = \frac{\sin (2K+1)\theta/2}{\sin(\theta/2)}. \tag{25.3}$$

These polynomials, when viewed as supporting hyperplanes to the moment spaces, display a number of interesting geometrical properties.

THEOREM 25.1 *For each n, the hyperplanes supporting D^n at $\underline{x}^*$ and $\overline{x}^*$ are parallel; this property determines ϕ^* uniquely.*

Hence, $\overline{x}^*_n - \underline{x}^*_n$ is the width of D^n in the x_n direction.

Proof. Parallelity of the lower and upper supporting hyperplanes is equivalent to the relation

$$\underline{\Delta}_n' = c_n \overline{\Delta}_n',$$

c_n being some constant. Thus,

$$\begin{aligned} \underline{\Delta}_n(t)^{2m-1} &= c_n\, t(1-t)\overline{\Delta}_n(t)^{2m-1} && (n \text{ even}), \\ t\underline{\Delta}_n(t)^{2m-1} &= c_n\, (1-t)\overline{\Delta}_n(t)^{2m-1} && (n \text{ odd}). \end{aligned} \tag{25.4}$$

If we take $x = x^*$ we find that all four derivatives in (25.4) are proportional to $\sin n\theta/\sin\theta$. This establishes the first assertion of the theorem. The uniqueness can be shown geometrically. Suppose that some point x' had the parallel property, with $\underline{x}'$, $\overline{x}'$ different from $\underline{x}^*$, $\overline{x}^*$. Then, clearly, we would have $\overline{x}' - \underline{x}' = \overline{x}^* - \underline{x}^*$, and the parallel segments $(\underline{x}', \underline{x}^*)$ and $(\overline{x}', \overline{x}^*)$ would lie in the boundary of D^n. But this is impossible, since a straight segment in the upper boundary of D^n can terminate only when it reaches the lower boundary, and vice versa. Hence the only points in D^n with the parallel property are those whose first $n-1$ coordinates agree with x^*. By Theorem 7.1b ϕ^* is uniquely determined.

It is instructive to establish the uniqueness also by means of the identities (25.4). Let us consider the even case, $n = 2m$, and write $\underline{R}_{m-1}(t)$ for $\underline{\Delta}_n(t)$, $\overline{R}_{m-1}(t)$ for $\overline{\Delta}_n(t)$, to indicate more plainly the degrees of these polynomials. The polynomials in (25.4) have degree $2m-1$,

and yet must vanish at every root of both $\underline{R}_m$ and $\overline{R}_{m-1}$, which are distinct by Theorem 12... Therefore we have

$$t(1-t)\underline{R}'_{m-1} = (\underline{\alpha}_m t + \underline{\beta}_m)\underline{R}_m - \underline{\gamma}_m \overline{R}_{m-1} \,,$$

and

$$t(1-t)\overline{R}'_{m-1} + (1-\alpha t)\overline{R}_{m-1} = \underline{\epsilon}_m \underline{R}_m - \overline{\gamma}_m \overline{R}_{m-1} \,,$$

$\underline{\epsilon}_m$ and $\overline{\epsilon}_m$ being constants. Eliminating $\overline{R}_{m-1}$ and $\underline{R}_m$, in turn, yields:

$$t(1-t)\underline{R}''_m + [1-\alpha t]\underline{R}'_m = \gamma_{m-1}\underline{R}_m \,,$$

$$t(1-t)\overline{R}''_{m-1} + [1-\alpha t]\overline{R}'_{m-1} = \gamma_{m-1}\overline{R}_{m-1} \,,$$

γ_m and γ_{m-1} depending on $\underline{\epsilon}_m$ and $\overline{\epsilon}_m$. Thus we see that $\underline{R}_m$ and $\overline{R}_m$ are m-th degree polynomial eigenfunctions of a pair of differential operators. Their uniqueness, for each m, is evident from the form of the equations (25.6). The Tchebycheff polynomials in fact form complete orthogonal systems of eigenfunctions for these operators. (See Szego [10], p. 59, for the form of (25.6) for the general Jacobi polynomials $P_n^{(\alpha,\beta)}$).

Theorem 25.1 expresses in geometrical language the well known extremal property of the Tchebycheff polynomials of the first kind:

THEOREM 25.2 *Among all m-th degree polynomials with the same leading coefficient, T_m has the least maximum absolute value in the unit interval.*

Proof. We observe that the distance in the x_n direction from the point $x' \in E^n$ to the hyperplane defined by $y \cdot x = 0$ is $(y \cdot x')/y_n$. Let

$$R_m(t) = \sum_{j=1}^{m} \alpha_j t^j , \qquad \alpha_m \neq 0 ,$$

be an arbitrary polynomial of degree m. The distance in the x_{2m} direction from the point $x(t) = (1, t, \ldots, t^{2m})$ to the hyperplane corresponding to R_m^* is then $R_m(t)/\alpha_m$. The maximum of this distance as t runs from 0 to 1 cannot be less than the x_{2m}-width of I^{2m}, since $x(t)$ describes the set of all extreme points of I^{2m} while the hyperplane supports or bounds I^{2m}. By the last theorem, T_m achieves this lower bound for the maximum distance. This is equivalent to the statement of the present theorem.

Each moment of ϕ^* is at the midpoint of the range permitted by its predecessors:

THEOREM 25.3 *For each n,*

$$x = \frac{1}{2}(\underline{x^*} + \overline{x^*}); \tag{25.7}$$

this property determines ϕ^* uniquely.

Proof. It is sufficient to show that x^σ satisfies (5.7) for each n, since, when applied successively for $n = 1, 2, \ldots$, the relation (5.7) clearly determines a unique moment sequence. Since direct calculation of $\underline{x}_n$ and $\overline{x}_n$ is difficult, we proceed as follows. Suppose $n = 2m$. By the definitions of $\underline{\phi}$ and $\overline{\phi}$, we have at once

$$\int_0^1 t^m \underline{\Delta}_n^{\sigma}(t)\, d\underline{\phi}^{\sigma} = 0, \tag{5.8}$$

$$\int_0^1 (t^m - t^{m+1})\, \overline{\Delta}_n^{\sigma}(t)\, d\overline{\phi}^{\sigma} = 0. \tag{5.9}$$

The first of these involves $\underline{x}_n^{\sigma}$, the second $\overline{x}_n^{\sigma}$. Using the special property (5.5) of ϕ^{σ}, we obtain from (5.9)

$$\int_0^1 (t^m - t^{m+1})\, \underline{\Delta}_n^{\sigma}(t)\, d\underline{\phi}^{\sigma} = 0. \tag{5.10}$$

Written as determinants, (5.8) and (5.10) differ only in the last column. Subtracting (5.10) from m times (5.8) yields:

$$\begin{vmatrix} x_0^{\sigma} & \cdots & x_{m-1}^{\sigma} & m x_n^{\sigma} \\ x_1^{\sigma} & \cdots & x_m^{\sigma} & (m+1) x_{m+1}^{\sigma} - x_m^{\sigma} \\ \cdots & \cdots & \cdots & \cdots \\ x_m^{\sigma} & \cdots & x_{2m-1}^{\sigma} & m x_{m+n}^{\sigma} - n x_{m+n-1}^{\sigma} \\ \cdots & \cdots & \cdots & \cdots \\ x_m^{\sigma} & \cdots & x_{2m-1}^{\sigma} & m \underline{x}_n^{\sigma} + \overline{x}_n^{\sigma} - m x_{2m-1}^{\sigma} \end{vmatrix} = 0. \tag{5.11}$$

In this determinant, the last two columns are proportional except for the last row, for we have:

$$(m-1)x^*_{m+n-1} = (m+k)x^*_{m+n} - (k)x^*_{m+n-1},$$

as may easily be verified from (18.1). Hence, (25.1) reduces to

$$\Delta^*_{n-1}\,\overline{m(x^*_{n+1}+x^*_n) - mx^*_{n-1} - \frac{(n-1)}{1}x^*_{n-1}} = 0.$$

Since the minor Δ^*_{n-1} does not vanish, we obtain

$$\frac{1}{1}(x^*_{n+1}+x^*_n) = \frac{(n-1)}{(n)}x_{n-1} = x^*_n,$$

as was to be shown. The odd case may be treated similarly.

THEOREM 25.1. *For each* n, $\Delta_n = \bar{\Delta}_n$; *this property determines* ϕ^* *uniquely.*

Proof. By (18.1), (18.2), and the preceding theorem, we have

$$(25.1)\qquad \frac{\bar{\Delta}^*_n}{\bar{\Delta}^*_{n-1}} = \frac{\Delta^*_n}{\Delta^*_{n-1}} = \bar{x}^*_{n+1} + \bar{x}^*_n - x_n = 0.$$

A direct check reveals that $\Delta^*_0 = \bar{\Delta}^*_0$ and $\Delta^*_1 = \bar{\Delta}^*_1$. It follows by induction that $\Delta^*_n = \bar{\Delta}^*_n$ for any n. Conversely, if any sequence $1, x_1, x_2, \ldots$ has the property that $\Delta_n = \bar{\Delta}_n$ for each n, then (25.1), without the asterisks, taken with the preceding theorem, shows that it must be the moment sequence of ϕ.

Finally, we give some numerical results, with brief indications of how they were obtained. They depend, of course, on the particular interval of orthogonality $(-1,1)$ with which we have been working. The earlier theorems of this section, on the other hand, are equally valid after obvious adjustments in the definitions for any finite interval.

THEOREM 5.5 *The width in the x_k direction of p^k is $\sim n^{-3/2}$.*

Proof. The leading coefficient of the normalized Tchebycheff polynomial of m-th degree, evaluated directly, turns out to be $\sim 2^{m-1/2}$. However, by (7.3), it is also equal to $\sqrt{\lambda^K_{m-1}/\lambda^K_m}$. By (18.1) and (18.2), and the last theorem, we therefore have:

$$\overline{x}_k - \underline{x}_k = \sqrt{\lambda^K_n/\lambda^K_{n-1}} \sim n^{-3/2}, \tag{5.1}$$

for n even. A similar procedure establishes (5.1) for n odd. The result then follows by Theorem 4.1.

THEOREM 5.6 *We have*

$$\lambda^K_n = \overline{x}_n = n^{-(n+1)/2},$$

$$\underline{x}^K_n = \left(\sqrt{n}-2\right) n^{-3/2},$$

$$\overline{x}_k = \left(\sqrt{n}\right) n^{-3/2}.$$

Proof. The first comes out by induction from (5.1). The other follow from (5.1), (4.3), and (4.1).

CHAPTER VII

SYMMETRIES OF THE MOMENT SPACES

§26. Reversal of the unit interval.

Any continuous, one-one transformation of the fundamental interval $[0,1]$ onto itself induces a corresponding one-one transformation of $\mathcal{P}$ onto itself, which may be specified by the requirement that for every continuous function f:

$$\int_0^1 f(t)\,d\phi'(t) = \int_0^1 f(t')\,d\phi(t) \tag{26.1}$$

(primes denoting the transformed objects). This leads in turn to a one-one transformation of the boundary of D^n onto itself (Theorem 20.1), which is again continuous. However, for the transformation to be well-defined in the interiors of the moment spaces, we must demand that for each n the first n moments of ϕ uniquely determine the first n moments of ϕ'. In particular,

$$\mu_1(\phi') = F(\mu_1(\phi)), \qquad \text{all } \phi \in \mathcal{P}. \tag{26.2}$$

Setting $\phi(t) = I(t-t_1)$, we have $\phi'(t) = I(t-t_1')$ and hence

$$t_1' = F(t_1),$$

determining the function F. Therefore, for any $\phi \in \mathcal{P}$:

$$\mu_1(\phi') = \int_0^1 t\,d\phi' = \int_0^1 t'\,d\phi = \int_0^1 F(t)\,d\phi, \tag{26.3}$$

But from (25.2) and (25.3) it follows that

$$\int_0^1 F(t)\,d\phi(t) = F\left[\int_0^1 t\,d\phi(t)\right] \qquad \text{for every } \phi \text{ in } \Phi,$$

and hence that F is linear. We conclude that the only transformation of $[0,1]$ which produces a non-trivial symmetry of the full moment spaces is

$$t \to t' = 1 - t, \tag{26.4}$$

reversing the interval rigidly. We shall devote the rest of this section to properties of this transformation.

We have at once

$$\phi'(t) = 1 - \phi(1 - t - 0), \tag{26.5}$$
$$\phi'' = \phi,$$

the limit operation indicated by "-0" serving to preserve right-continuity (see §5). By (26.1) and (26.4) we have

$$\mu_i(\phi') = \int_0^1 t'^i\,d\phi = \sum_{h=0}^{i} (-1)^h \binom{i}{h} \mu_h(\phi) = \Delta^i \mu_i(\phi) .$$

Let Δ_n stand for the "difference" matrix:

$$(\Delta_n)_{ij} = (-1)^{i-j}\binom{i}{j}, \qquad i, j = 0, 1, \ldots, n.$$

Then we have shown:

THEOREM 26.1 The correspondence (26.4) transforms points of D^n according to

(26.6) $$x' = \rho_n x.$$

Since $x'' = x$, we see that $\rho_n^{-1} = \rho_n$. This could also be verified directly.

In the barycentric coordinates based on the simplex S^n (see § 2), (26.6) becomes:

$$\lambda'_{nk} = \lambda_{n,n-k}, \qquad k = 0, 1, \ldots, n.$$

This follows directly from the defining formula (12.5) and the relation:

(26.7) $$\lambda'_r = \lambda^{n-k}_{n-r},$$

which holds for any x in E^{n+1}. The vertices of S^n are therefore interchanged as follows:

$$x^{(k)\prime} = x^{(n-k)}, \qquad k = 0, 1 \ldots, n.$$

The step-function having jumps of $\alpha_1, \ldots, \alpha_m$ at $t_1, \ldots, t_m$, respectively, goes into the step-function with jumps of $\alpha_m, \ldots, \alpha_1$ at $1-t_m, \ldots, 1-t_1$, respectively. If we recall that the "upper" faces of D^n are those associated with jumps at $t = 1$, we obtain at once from Theorem 11.6:

THEOREM 26.2 The correspondence (26.4) transforms the a-faces of D^n according to

$$\underline{C}^{n\,\prime}_a = \overline{C}^n_a, \qquad \overline{C}^{n\,\prime}_a = \underline{C}^n_a, \qquad \text{if } a \text{ is even},$$

$$\underline{C}^{n\,\prime}_a = \underline{C}^n_a, \qquad \overline{C}^{n\,\prime}_a = \overline{C}^n_a, \qquad \text{if } a \text{ is odd}.$$

As a corollary we have, for x in D^n,

$$\underline{x}' = \overline{x}', \qquad \overline{x}' = \underline{x}' \qquad \text{if } n \text{ is odd},$$

$$\underline{x}' = \underline{x}', \qquad \overline{x}' = \overline{x}' \qquad \text{if } n \text{ is even},$$

since the lower and upper projections of x are distinct only when they are in C^n_{n-1}.

In the polynomial space the induced transformation is

$$P(t) \to P'(t) = P(1-t).$$

(Renormalization is not necessary, as the original normalization (6.2) was symmetric.) We therefore have

$$y'_i = \sum_{h=i}^{n} (-1)^{i} \binom{h}{i} y_h$$

for y in P^n. Let Δ^*_n stand for the transpose of Δ_n. Then we have shown:

THEOREM 26.3 *The correspondence (2(.4) transforms points of P^n according*

to

(26.8) $$y' = \mathcal{C}_n^* y.$$

THEOREM 26.4 **x' and y' are conjugate if and only if x and y are conjugate.**

Proof. (26.6) and (26.8) give us:

(26.9) $$x' \cdot y' = \mathcal{C}_n x \cdot \mathcal{C}_n^* y = x \cdot (\mathcal{C}_n^*)^2 y = x \cdot y.$$

The result now follows, since the linear transformations (26.6) and (26.8) preserve convexity.

Turning to the Hankel determinants for x, x', in P^n, we find the following relationships:

THEOREM 26.5 **Under the transformation (26.4):**

$$\Delta'_{2k} = \Delta_{2k}, \qquad \bar{\Delta}'_{2k} = \bar{\Delta}_{2k},$$

$$\Delta'_{2k+1} = \bar{\Delta}_{2k+1}, \qquad \bar{\Delta}'_{2k+1} \quad \Delta_{2k+1}.$$

Proof: Let $\mathcal{C}_{2k}$ denote the matrix corresponding to [illegible]. By direct verification it follows from (26.6) and (26.7) that [illegible] = $\mathcal{C}_{2k}\mathcal{C}_{2k}$. Hence $\Delta'_{2k} = \Delta_{2k}$. In a similar manner the other statements of the theorem follow.

Theorem 26.1, already proved, can also be verified immediately from this theorem and Theorem 17.3.

The last theorem, in conjunction with the formulas of § 10, gives us the interesting result:

$$\overline{x'} - \underline{x'} = \overline{x} - \underline{x} .$$

That is, the symmetry preserves the x_n-width of D^n at all points.

§ 27. Moment space of symmetric distributions.

Let Φ_S stand for the set of ϕ in Φ with $\phi = \phi'$, and let D_S^n denote their n-th moment space. D_S^n therefore consists of those x in D^n with $x = x'$.

THEOREM 27.1 D_S^n *is convex, closed, bounded, and* $n/2$ *-dimensional.*

Proof. (Compare Theorem 7.2.) We observe that D_S^n is precisely the intersection of D^n with the variety defined by the linear system of equations

$$(27.1) \qquad {}_n x = x.$$

This can be solved for each odd moment in terms of the lower even moments:

$$x_1 = 1/2,$$

$$x_3 = (3/2)x_2 - 1/4,$$

$$x_5 = (5/2)x_4 - (5/2)x_2 + 1/2, \text{ etc.},$$

leaving exactly $n - n/2$ independent linear relations.[13] Hence, since D^n

[13] Had we taken $[-1,1]$ as the fundamental interval, the relations would have been simply $x_i = 0$, $i = 1, 3, 5, \ldots$.

is a convex body in E^n, D_S^n is a convex set of dimension $n/2$; and since D^n is closed and bounded, D_S^n is likewise.

In D_S^n the two-jump symmetric distribution functions:

$$\frac{1}{2}\left[I(t-t_1) + I(t-t_1')\right] \tag{27.1}$$

play a role similar to the role in D^n of the pure distribution functions $I(t-t_1)$. The moment points of (27.1):

$$x_S(t_1) = \frac{1}{2}\left[x(t_1) + x(t_1')\right], \qquad 0 \le t_1 \le \frac{1}{2},$$

describe a curve in D_S^n which we may call C_S^n.

THEOREM 27.2 *The set of extreme points of D_S^n, for $n \ge 4$, is precisely C_S^n.*

Proof. The proof does not differ essentially from that of Theorem 7.3. In place of the hyperplane (7.3) one may use

$$h(x) = 1 - 2t_1t_1' + 2(t_1t_1')^2 - 4(1 - t_1t_1')x_2 + 2x_4 = 0,$$

which has the property:

$$h(x_S(t)) = 2(t_1-t)^2(t_1'-t)^2$$

corresponding to (7.4).

The combinatorial boundary structure of D_S^n is like that of $D^{[n/2]}$,

for the two spaces are closely related. In fact, we have:

THEOREM 7.3 The spaces D_S^n and $D^{n/2}$ are connected by a 1:1 linear transformation.

Proof. First we observe that a change of the fundamental interval results in a linear transformation of the moment spaces. In particular, if we go from D^n to the moment space $D_{(\Lambda)}^n$ based on the interval $\Lambda = [-1, 1]$, then D_S^n is mapped onto the moment space $D_{(\Lambda)S}^n$ of distribution functions $\psi^{(\Lambda)}$ defined over Λ and symmetric with respect to $t' = -t$. A one-one correspondence between these ψ and the functions $\phi = \phi^{(\Lambda)}$ can be defined as follows:

$$\psi \to \phi: \quad \int_0^1 f(t)\,d\phi = \int_{-1}^1 f(t^2)\,d\psi,$$

$$\phi \to \psi: \quad \int_{-1}^1 g(t)\,d\psi = \frac{1}{2}\int_0^1 g(\sqrt{t})\,d\phi + \frac{1}{2}\int_0^1 g(-\sqrt{t})\,d\phi;$$

the equalities to hold for all continuous functions f and g on $[0,1]$ and $[\Lambda]$, respectively. The moments of ψ are then given by

$$\alpha_{2k}(\psi) = \alpha_k(\phi), \qquad \alpha_{2k+1}(\psi) = 0, \qquad k = 0, 1, 2, \ldots.$$

This describes the linear transformation from $D_{(\Lambda)S}^n$ to $D^{n/2}$, and completes the proof.

In view of Theorem 4.3, the symmetric subset P_S^n of P^n characterized by $y = y'$ cannot be the dual of D_S^n. However, both sets have dimension

[illegible] if restricted to convex cones in the suitable $\left[\frac{n}{2}\right]$-dimensional subspaces of E^n and $(E^n)^*$, the two sets are themselves duals. Indeed, by definition, $y \in E^{n,*}$ is in the $n/2$-dimensional dual of P_S^n if and only if

(27.2) $\qquad y = y' \quad$ and $\quad x \cdot y \geq 0 \quad$ all $\quad x \in D_S^n$.

By the relation:

$$x \cdot y = \frac{(x+x')}{2} \cdot y, \qquad \text{all } y = y', \quad x \in D^n,$$

which proceeds from (26.9), condition (27.2) is equivalent to the condition:

$$y = y' \quad \text{and} \quad x \cdot y \geq 0 \quad \text{all} \quad x \in D^n.$$

But this exactly defines P_S^n.

§ 28. Other symmetries.

If we set out to find a symmetry of only the boundary of D^n, arising originally from a symmetry of the unit interval, we are no longer forced to make the transformation $t \longrightarrow t'$ linear, as in 26; we have only to demand continuity and the property $t'' = t$. [Symmetries of order higher than 2 are clearly impossible.] Many non-trivial such symmetries exist having various properties corresponding to the results of the last two sections. All of them, however, reverse the unit interval:

$$0' = 1, \qquad 1' = 0.$$

Consequently Theorem 26.2, on the behavior of the a-faces of D^n, applies for them all.

[illegible] We omit the simple argument for P_S^n.

The combinatorial structure of the boundary of D^n is perfectly symmetrical in each dimension, by (11.6). It is intuitively clear that other symmetries must exist which lead to all possible variants of Theorem 26.2. It is not especially interesting to attempt to define these directly for points in D^n; nor, as we have just indicated, is it possible to do so indirectly by starting in $[0,1]$. But there is one symmetry, at least, which can be defined in a simple fashion for functions in Φ^1 and then extended to the moment spaces, producing a new version of Theorem 26.2.

Consider the transformation, for $\phi \in \Phi^1$:

$$(26.1) \qquad \phi \to \phi': \qquad \phi'(t) = \begin{cases} \text{l.u.b.}_u\, \phi(u) \le t, & 0 \le t < 1, \\ \phi(t), & t < 0, \quad t \ge 1. \end{cases}$$

Then $\phi'' = \phi$. Intuitively, ϕ' is the inverse of ϕ in $[0,1]$. The first moments of ϕ and ϕ' are related by

$$\alpha_1(\phi') = 1 - \alpha_1(\phi), \qquad \text{all } \phi \in \Phi^1,$$

as may be shown by integration by parts. Such relations do not exist for the higher moments. Hence the induced symmetry in D^n, $n \ge 2$, is well-defined only for the boundary.

The step-function

$$\phi(t) = \sum_{j=1}^{m} \epsilon_j I(t-t_j), \qquad t_1 < t_2 < \ldots < t_m,$$

is carried by (26.1) into the step-function with jumps of

$$t_1, \quad t_2-t_1, \quad t_3-t_2, \quad \ldots, \quad t_m-t_{m-1}, \quad 1-t_m$$

located at

$$0, \xi_1, \quad \xi_1 + \xi_2, \ \ldots \ \xi_1 + \xi_2 + \cdots + \xi_{m-1}, \quad 1$$

respectively — with the first and last jumps possibly vacuous. It is easy to see that $b'(\phi') = b'(\phi)$, and hence that $a(x') = a(x)$. Moreover, ϕ' has a jump at $t = 1$ if and only if ϕ does not. Therefore:

THEOREM 28.1 *The correspondence* (28.1) *transforms the a-faces of* D^n *according to*

$$C^{n}_{-a} = \overline{C}^{n}_{a}, \qquad \overline{C}^{n}_{a} = C^{n}_{-a} \qquad a = 0, 1, \ldots, n-1.$$

Another symmetry may be defined by the product (in either order) of the transformations (26.5) and (28.1). The corresponding theorem would involve the product of the transformations of Theorems 2[illegible] and 2[illegible].1.

The symmetry defined by (28.1) interchanges the roles of the "root" simplex U^m and the "weight" simplex $\Sigma^{m'}$ ($m = m' - a/2$), which parametrize the typical a-face of D^m (see the proof of Theorem 11.[illegible]). The points of U^m correspond (before the transformation) to the maximal convex sets in the a-face ($a < m$), while the points of $\Sigma^{m'}$ correspond to sets in the a-face which are not themselves convex and which contain no convex subsets other than points. Our present symmetry therefore destroys all convex sets in the boundary of D^n. In contrast, transformations of $[0,1]$, whether linear or not, always preserve convexity when extended to the spaces $\mathcal{L}^1$ and D^n.

Jo

REFERENCES

1. S. Karlin and L. S. Shapley, Geometry of reduced moment spaces, Proc. N.A.S. 35 (1949) 673-677.

2. M. Dresher, S. Karlin and L. S. Shapley, Polynomial games, Annals of Mathematics Study No. 24, "Contributions to the Theory of Games", ed. H. W. Kuhn and A. W. Tucker, Princeton, 1950, 161-180.

3. H. Weyl, Elementare Theorie der konvexen Polyeder, Comentarii Mathematici Helvetici 7 (1935), 290-306 (also, in English, in Annals of Mathematics Study No. 24, "Contributions to the Theory of Games", ed. H. W. Kuhn and A. W. Tucker, Princeton (1950).

4. T. Bonessen and W. Fenchel, Theorie der konvexen Körper, Ergebnisse der Mathematik und ihre Grenzgebeite, Berlin, 1934, Vol. III.

5. E. Steinitz, Bedingt konvergente Reihen und konvexe Systeme, I, J. Reine angew. Math. 143 (1913), 128-175.

6. O. Hanner and H. Rådström, A generalization of a theorem of Fenchel, Proc. A.M.S. 2 (1951) 589-593.

7. H. Minkowski, Theorie der konvexen Körper, Gesammelte Abhandlungen, Berlin, 1911, Vol. II, 131-229.

8. A. Wald, Limits of a distribution function, Trans. A.M.S. 46 (1939), 280-306.

9. D. V. Widder, The Laplace Transform, Princeton, 1941.

10. G. Szegö, Orthogonal Polynomials, American Mathematical Society Colloquium Publications, 23, New York, 1939.

11. F. Hausdorff, Momentproblem für ein endliches Intervall, Math. Zeitschrift, 16 (1923), 220-248.

12. J. Shohat and J. Tamarkin, The Problem of Moments, American Mathematical Society, New York, 1943, 1950.

13. A. Selberg, Bemerkninger om et multipelt integral, Norsk Mat. Tidsskr. 26 (1944), 71-78.

14. W. L. Ferrer, Algebra, Oxford, 1941.

15. M. Dresher and S. Karlin, Solutions of convex games as fixed points, Annals of Mathematics Study No. 28, "Contributions to the Theory of Games II", ed. H. W. Kuhn and A. W. Tucker, Princeton, 1952.