BOUNDS ON A DISTRIBUTION FUNCTION WHEN ITS FIRST n MOMENTS ARE GIVEN[1]

By H. L. Royden

Stanford University

Introduction. Let $F(X)$ be a nondecreasing function defined on the real line with $F(-\infty) = 0$ and

$$\int_{-\infty}^{\infty} t^k \, dF(t) = M_k \qquad k = 0, \cdots, 2n.$$

Then the problem of Tchebycheff is to find upper and lower bounds for $F(X)$. If X is a random variable with the cumulative distribution function $F(X)$, this is just the problem of determining the (sharp) upper and lower bounds for Pr. $(X \leqq d)$. This problem has been solved by Markoff [2] and Stieltjes [5] and their results are given in Section 1.

It is often of interest, however, to determine upper and lower bounds for Pr. $(|X| \leqq d)$. This is the problem of determining upper and lower bounds on the cumulative distribution function $F^*(X^*) = F(X^*) - F(-X^*)$ of the nonnegative random variable $X^* = |X|$, and leads to the Stieltjes moment problem: To determine upper and lower bounds on the nondecreasing function F^* given

$$\int_{-\infty}^{\infty} t^k \, dF^*(t) = M_k \qquad k = 0, \cdots, n;$$

and $F^*(0-) = 0$. The numbers M_k are now the absolute moments of X, that is $M_k = E(|X|^k)$. It should be noted that the set of moments.

$$M_{\alpha k} = E(|X|^{\alpha k}) \qquad k = 0, \cdots, n,$$

serves just as well as $M_0, \cdots, M_n$, since they are the first n algebraic moments of the nonnegative random variable $Y = |X|^\alpha$ with the cumulative distribution function $G(Y) = F(Y^{1/\alpha})$.

In the second and third section we give a solution to this problem which corresponds to the classical Tchebycheff inequalities for the Hamburger moment problem, and apply these general results in the next section to obtain the Cantelli inequalities. I would like to point out that Theorems 1 and 2 can be derived from very general results of Krein [9]. However, the self-contained approach used here seems to me desirable in view of the complexity and inaccessibility of Krein's results. In the last section we solve the problem of determining sharp upper and lower bounds for a distribution given the first two (absolute) moments about the mode.

1. The Tchebycheff inequalities. A point t is said to belong to the spectrum of the random variable X or of the corresponding distribution function $F(X)$

Received 10/20/52.

[1] This work sponsored by the Office of Naval Research under Contract N6onr-25140.

if there is no interval about t in which $F(X)$ is constant. A random variable (distribution function) is called arithmetic if it is continuous on the right and its spectrum is a finite set of points. We say that a distribution function belongs to a set of moments if it has these moments. By the mass at t we mean the quantity $F(t+) - F(t-)$.

For the sake of completeness we include the method for determining the (sharp) upper and lower bounds for the distribution function $F(d)$, that is, Pr. $(x \leqq d)$, when we are given the first $2k$ algebraic moments of $F(x)$ about the origin. A distribution function belonging to these moments is said to be characteristic for d if it is arithmetic and its spectrum contains fewer than k points other than d. Then we have the following propositions.

PROPOSITION 1. Given the moments $M_0, \cdots, M_{2k}$ of some distribution function $F(x)$, and a real number d, there is a distribution function $F_d(x)$ belonging to them which is characteristic for d.

PROPOSITION 2. We have $F_d(d-) \leqq F(d) \leqq F_d(d)$.

PROPOSITION 3. The points x_i at which F_d increases are the zeros of the polynomial

$$Q_d(x) = Q_{d,k}(x) = \begin{vmatrix} 1 & 1 & M_0 & \cdot & M_{k-1} \\ x & d & M_1 & \cdot & M_k \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x^{k+1} & d^{k+1} & M_{k+1} & \cdot & M_{2k} \end{vmatrix}.$$

The mass m_i of F_d at x_i is given by

$$m_i = \sum_{l=0}^{n} c_l^{(i)} M_l \tag{1}$$

where

$$Q^{(i)}(x) = \sum_{l=0}^{n} c_l^{(i)} x^l = \prod_{j \neq i} \frac{(x - x_j)}{(x_i - x_j)}.$$

(If Q has no multiple zeros we have $Q^{(i)}(x) = (Q(x))/(Q'(x_i)(x - x_i))$.)

For a proof the reader may either refer to ([4], p. 43) or construct one analogous to that given here in Sections 2 and 3.

2. The Stieltjes case. In this section we consider the sharpening of the foregoing inequalities which is possible in the case of a positive random variable Let $M_0, M_1, \cdots, M_n$, be the first n moments of a positive random variable X with the cumulative distribution function $F(X)$, that is, $M_k = \int_{0-}^{\infty} t^k \, dF(t)$. Then it is known [4] that the determinants $\Delta_{2k} = | M_{i+j} |_{i,j=0}^{k}$ and $\Delta_{2k+1} = | M_{i+j+1} |_{i,j=0}^{k}$ satisfy

$$\Delta_l > 0, \qquad l = 0, \cdots, n$$

unless the moments $\{M_k\}$ belong to an arithmetic distribution function having mass at $n/2$ or fewer points, counting the point zero with multiplicity one-half. In this latter case, that is, when $\Delta_n = 0$, the distribution function having these moments is unique, and the upper and lower bounds are trivial. Consequently, we shall assume that the Δ_k are all positive.

The set of distribution functions which have moments of order n may be considered as a set of positive linear functionals on the space $\mathfrak{N}$ consisting of those continuous functions φ on $[0, \infty]$ for which $\lim_{t\to\infty} \varphi(t)/t^n$ exists.

We say that a sequence of distribution functions F_i converges to the distribution functions F if $\lim \int_{0-}^{\infty} \varphi\, dF_i = \int_{0-}^{\infty} \varphi\, dF$ for all $\varphi \varepsilon \mathfrak{N}$. It follows at once from the definition of the Stieltjes integral that every distribution function is the limit of arithmetic distribution functions.

Let $\mathfrak{F} = \mathfrak{F}(n, M)$ be the set of arithmetic distribution functions which have mass on at most $n + 1$ points and for which $\int_{0-}^{\infty} t^n\, dF(t) \leqq M$. We shall find it convenient to include in $\mathfrak{F}$ the positive linear functionals on $\mathfrak{N}$ defined by

$$L[\varphi] = \lambda \lim_{t\to\infty} \frac{\varphi(t)}{t^n}, \qquad \lambda \geqq 0.$$

We say that L arises by placing an "infinitesimal" mass λ at infinity. Thus if F is an arithmetic distribution function having masses m_i at x_i and an infinitesimal mass λ at infinity we write

$$\int_{0-}^{\infty} \varphi\, dF = \sum m_i \varphi(x_i) + \lambda \lim \frac{\varphi(t)}{t^n}.$$

LEMMA 1. *The set $\mathfrak{F}$ is sequentially compact.*

PROOF. Given a sequence $F^{(j)} \varepsilon \mathfrak{F}$, we may choose a subsequence such that (a) each spectral point $x_i^{(j)}$ of $F^{(j)}$ converges to some point $x_i \varepsilon [0, \infty]$. (b) If $x_i \neq \infty$, the masses $m_i^{(j)}$ which $F^{(j)}$ has at $x_i^{(j)}$ converge to some number m_i. (c) The integrals $I_j = \int_0^{\infty} t^n\, dF^{(j)}$ converge to some number I.

Any function $\varphi \varepsilon \mathfrak{N}$ may be written as $\varphi = \varphi_0 + \alpha t^n$ where $\lim_{t\to\infty} \varphi_0(t)/t^n = 0$ and $\lim_{t\to\infty} \varphi(t)/t^n = \alpha$. Given $\epsilon > 0$, we may choose $\mathfrak{I}$ so large that

$$\frac{\varphi_0(t)}{t^n} < \epsilon/M \qquad \text{for all} \qquad t \geqq \mathfrak{I}.$$

Consequently,

$$\int_{\mathfrak{I}}^{\infty} \varphi_0(t)\, dF^{(j)} \leqq \epsilon/M \int_{\mathfrak{I}}^{\infty} t^n\, dF^{(j)} \leqq \epsilon.$$

Since φ_0 is uniformly continuous on $[0, \mathfrak{I}]$, we may choose j so large that

$$\Big| \sum_{x_i < \mathfrak{I}} m_i \varphi_0(x_i) - m_i^{(j)} \varphi_0(x_i^{(j)}) \Big| < \epsilon.$$

Then

$$\left| \int_{0-}^{\infty} \varphi_0(t)\, dF^{(j)}(t) - \sum_{x_i < \mathfrak{J}} m_i \varphi_0(x_i) \right| < 2\epsilon$$

whence, taking j so large that $| I - I_j | < \epsilon$,

$$\left| \int_{0-}^{\infty} \varphi_0(t)\, dF^{(j)}(t) - \sum_{x_i < \mathfrak{J}} m_i \varphi_0(x_i) - \alpha I \right| < 3\epsilon.$$

Since ϵ was arbitrary and $\mathfrak{J} \to \infty$ as $\epsilon \to 0$,

$$\lim \int_{0-}^{\infty} \varphi_0(t)\, dF^{(j)} = \sum m_i \varphi_0(x_i) + \alpha I.$$

Also

$$I = \lim \int_{0-}^{\infty} t^n\, dF^{(j)}(t) \geqq \lim \int_{0-}^{\mathfrak{J}} t^n\, dF^{(j)}(t) = \sum_{x_i \leqq \mathfrak{J}} m_i (x_i)^n.$$

Since $\mathfrak{J}$ is arbitrary $I \geqq \sum m_i(x_i)^n$, whence $\lambda = I - \sum m_i(x_i)^n \geqq 0$. Since $\varphi(x_i) = \varphi_0(x_i) + \alpha x_i^n$, we have

$$\lim \int_{0-}^{\infty} \varphi(t)\, dF^{(j)} = \sum m_i \varphi(x_i) + \alpha\lambda = \int_{0-}^{\infty} \varphi(t)\, dF(t).$$

where F is that distribution function in $\mathfrak{F}$ which has mass m_i at x_i and an infinitesimal mass λ at infinity. Thus $F^{(j)} \to F$ proving the lemma.

LEMMA 2. *If $\{M_1, \cdots, M_n\}$ are the moments of some distribution function, they are the moments of a distribution in $\mathfrak{F} = \mathfrak{F}(n, M)$, with $M_n < M$.*

PROOF. Every distribution function with n moments is the limit of arithmetic distributions by the definition of the Stieltjes integral. Since $\mathfrak{F}$ is sequentially compact and the moments are continuous, it suffices to show that if $M_1, \cdots, M_n$ are the moments of an arithmetic distribution, they are the moments of a distribution $\mathfrak{F}$. Let A be the region in n-dimensional Euclidean space whose points are moments of arithmetic distributions. Then A is the convex hull of the curve

$$C: M_i = t^i \qquad i = 1, \cdots, n; \quad 0 \leqq t \leqq \infty.$$

Thus every point of A must be in the simplex determined by some $n + 1$ points of C, that is,

$$M_i = \sum_{j=0}^{n} m_j (t_j)^i; \qquad m_j \geqq 0; \qquad \sum m_j = 1.$$

But this is just the statement that the M_i are the moments of the arithmetic distribution F which has mass m_j at t_j, and this belongs to $\mathfrak{F}$ as soon as $M_n < M$.

LEMMA 3. *The mass $m = m(d)$, which a distribution F has at a given fixed point d is upper semi-continuous as a function of F.*

PROOF. Let $F^{(j)}$ with mass $m^{(j)}$ at d converge to F with mass m at d. Let φ be a positive continuous function belonging to $\mathfrak{N}$, which is one at d and vanishes at the other spectral points of F. Then

$$m = \int_{0-}^{\infty} \varphi\, dF = \lim \int_{0-}^{\infty} \varphi\, dF^{(j)} \geqq m^{(j)},$$

as was to be proved.

DEFINITION. Let $M_0, M_1, \cdots, M_n$, be the Stieltjes moments of some distribution function. An arithmetic distribution function belonging to this set is said to be characteristic for d if its spectrum contains at most $n/2$ points different from d where zero and infinity are counted with multiplicity one-half.

THEOREM 1. *If $M_0, \cdots, M_n$ are the Stieltjes moments of some distribution function and d is a positive number, then there is an arithmetic distribution function belonging to $M_0, \cdots, M_n$ which is characteristic for d.*

PROOF. By Lemmas 1, 2, and 3, there is an arithmetic distribution function F_d belonging to $M_0, \cdots, M_n$ which has the largest possible mass m_d at the point d. Suppose the spectrum of F_d contained more than $n/2$ points. Then if n is odd we must have one of the cases A or B below.

CASE A. Positive masses m_i at the k points x_i where $k = (n + 1)/2$. Then the Jacobian of the moments $M_0, \cdots, M_n$ with respect to changes in m_i and x_i is

$$\begin{vmatrix} 1 & \cdot\cdot & 1 & 0 & \cdot & 0 \\ x_i & \cdot & x_k & m_1 & \cdot & m_k \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_1^n & \cdot & x_k^n & nm_1x_1^{n-1} & \cdot & nm_kx_k^{n-1} \end{vmatrix} = m_1 \cdots m_k \prod_{i \neq j} (x_i - x_j)^4 \neq 0,$$

and we can increase the mass at d by a small amount and change the m_i and x_i so that the moments $M_0, \cdots, M_n$ remain the same. Thus case A is impossible.

CASE B. Positive masses m_i at the k points x_i, mass m_0 at zero and infinitesimal mass λ at infinity, where $k = (n - 1)/2$. Here the Jacobian is

$$\begin{vmatrix} 1 & 0 & 1 & \cdot & 1 & 0 & \cdot & 0 \\ 0 & 0 & x_1 & \cdot & x_k & m_1 & \cdot & m_k \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & x_1^n & \cdot & x_k^n & nm_1x_1^{n-1} & \cdot & nm_kx_k^{n-1} \end{vmatrix}$$

$$= m_1 \cdots m_k x_1^2 \cdots x_k^2 \prod (x_i - x_j)^4 \neq 0.$$

and Case B is impossible.

Using similar arguments in the case of even n, we see that the spectrum of F_d can contain at most $n/2$ points other than d, thus proving the theorem.

THEOREM 2. *Let $M_0, \cdots, M_n$ be the Stieltjes moments of a distribution function $F(x)$. Let $F_d(x)$ belong to $M_0, \cdots, M_n$ and be characteristic for d. Then $F_d(d-) \leqq F(d) \leqq F_d(d)$.*

PROOF. We consider four cases:

CASE A. Even n and the spectrum of F_d consists of d and $k \leqq n/2$ other points $x_1, \cdots, x_k$ all different from zero and infinity. We construct a polynomial $P(x)$ of degree $2k + 1$ which satisfies the following conditions:

$$\begin{aligned} P'(x_i) &= 0 && i = 1, \cdots, k \\ P(x_i) &= 1 && x_i < d \\ P(d) &= 1 && \\ P(x_i) &= 0 && x_i > d. \end{aligned}$$

Such a polynomial exists, since these conditions are $2k + 1$ linear equations for the $2k + 1$ coefficients of the polynomial and the determinant of this system is

$$\prod_i (d - x_i)^2 \prod_{i \neq j} (x_i - x_j)^4 \neq 0.$$

It is easily verified that $P(x) \geqq 1$ for $x < d$ and $P(x) \geqq 0$ for all x. Since F_d has the same first $2k + 1$ moments as F we have

$$F_d(d) = \int_{0-}^{\infty} P(x)\, dF(x) = \int_{0-}^{\infty} P(x)\, dF(x) \geqq \int_{0-}^{d} P(x)\, dF(x) \geqq \int_{0-}^{d} dF(x) = F(d).$$

Similarly $F_d(d-) \leqq F(d)$.

CASE B. Even n and the spectrum of F_d consists of d and $k = (n/2) - 1$ points other than zero and infinity. We construct a polynomial of degree $2k + 1$ which satisfies the conditions

$$\begin{aligned} P'(x_i) &= 0 && i = 1, \cdots, k \\ P(0) &= 1 && \\ P(x_i) &= 1 && x_i < d \\ P(d) &= 1 && \\ P(x_i) &= 0 && x_i > d. \end{aligned}$$

Such a polynomial exists as before, and also $P(x) \geqq 0$ for $x \geqq d$, and $P(x) \geqq 1$ for $0 \leqq x \leqq d$. Hence, since P is of degree less than n, we have

$$F_d(d) = \int_{0-}^{\infty} P(x)\, dF_d(x) = \int_{0-}^{\infty} P(x)\, dF(x) \geqq \int_{0-}^{d} P(x)\, dF(x) \geqq \int_{0-}^{d} dF(x) = F(d).$$

Similarly $F_d(d-) \leqq F(d)$.

CASE C. Odd n and the spectrum of F_d consists (possibly) of 0, d and $k \leqq (n - 1)/2$ other points all different from infinity. Construct $P(x)$ of degree $2k + 1$ such that

$$\begin{aligned} P'(x_i) &= 0 & i &= 1, \cdots, k. \\ P(0) &= 1 & & \\ P(x_i) &= 1 & x_i &< d \\ P(d) &= 1 & & \\ P(x_i) &= 0 & x_i &> d. \end{aligned}$$

Such a polynomial exists as before and $P(x) \geqq 0$ for $x \geqq d$ and $P(x) \geqq 1$ for $0 \leqq x \leqq d$. Hence

$$F_d(d) = \int_{0-}^{\infty} P(x)\, dF_d(x) = \int_{0-}^{\infty} P(x)\, dF(x) \geqq F(d),$$

and similarly $F_d(d-) \leqq F(d)$.

CASE D. Odd n and the spectrum of F_d consists (possibly) of d, infinity and $k \leqq (n - 1)/2$ other points all different from zero. Construct $P(x)$ of degree $2k$ such that

$$\begin{aligned} P'(x_i) &= 0 & i &= 1, \cdots, k. \\ P(x_i) &= 1 & x_i &> d. \\ P(d) &= 1 & & \\ P(x_i) &= 0 & x_i &< d. \end{aligned}$$

Here $P(x) \geqq 1$ for $x \leqq d$ and $P(x) \geqq 0$. Hence $F_d(d) \geqq F(d)$ and $F_d(d-) \leqq F(d)$ as before.

This plethora of cases establishes the theorem.

3. Determination of F_d.

LEMMA 4. *Let n be greater than or equal to $2k$, and let the roots of*

$$D(\lambda) \equiv \begin{vmatrix} \lambda & M_0 & \cdot & M_{k-1} \\ \lambda^2 & M & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \lambda^k & M_k & \cdot & M_{2k-1} \end{vmatrix} = 0,$$

be $\lambda_1 \leqq \lambda_2 \leqq , \cdots , \leqq \lambda_k$. *Let the roots of*

$$\Delta(\theta) \equiv \begin{vmatrix} \theta & M_1 & \cdot & M_k \\ \theta^2 & M_2 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \theta^{k+1} & M_{k+1} & \cdot & M_{2k} \end{vmatrix} = 0$$

be $\theta_0 \leqq \theta_1 \leqq \cdots \leqq \theta_k$. *Then* $0 = \theta_0 < \lambda_1 < \theta_1 < \cdots \lambda_k < \theta_k$, *and the polynomial*

$$Q_d(x) = \begin{vmatrix} 1 & 1 & M_0 & \cdot & M_{k-1} \\ x & d & M_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x^{k+1} & d^{k+1} & M_{k+1} & \cdot & M_{2k} \end{vmatrix}$$

has all positive roots if either $\theta_{i-1} < d < \lambda_i$ *for some* i *or* $\theta_k < d$ *and has one negative root if*

$$\lambda_i < d < \theta_i$$

for some i.

PROOF. We first note that $D(\lambda)$ is (apart from a constant factor) the orthogonal polynomial degree k with respect to dF. Hence ([6], p. 43) the λ_i are all positive and distinct. By Propositions 1 and 3 there is an arithmetic distribution G_d defined on $(-\infty, \infty)$ whose spectrum consists of the zeros of $Q_d(x)$, and which has $M_0, \cdots, M_{2k}$ as its (algebraic) moments. By Proposition 2 we see that the zeros of $Q_d(x)$ and $Q_d'(x)$ separate each other. Consequently, as d increases, the zeros of $Q_d(x)$ must increase until the largest becomes $+\infty$. Proposition 2 also shows that Q_d can have at most one negative zero. Hence for $d < 0$, d is the only negative root of Q_d. As d increases all roots of Q_d increase and so for d slightly larger than zero all roots are positive and remain so until the largest becomes infinite, but this happens only when the coefficient of x^{k+1} in Q_d vanishes, that is, when $D(d) = 0$. Thus all roots are positive if $0 < d < \lambda_1$. For d slightly larger than λ_1 we have a large negative root which increases to zero as d increases, and Q_d has a negative root for $\lambda_1 < d < \theta_{i_1}$ where θ_{i_1} is the smallest root of $\Delta(\theta)$ which is larger than d. Continuing this process we see that the roots of $\Delta(\theta)$ separate those of $D(\lambda)$ and that the lemma is true.

THEOREM 3. *Let* $n = 2k$ *be even, and let* λ_i *and* θ_i *be defined as in Lemma* 4. *Then if* $\theta_{i-1} < d < \lambda_i$ or $\theta_n < d$, *the spectrum of* F_d *consists of the roots* x_i *of*

$$Q_d(x) = \begin{vmatrix} 1 & 1 & M_0 & \cdot & M_{x-1} \\ x & d & M_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x^{k+1} & d^{k+1} & M_{k+1} & \cdot & M_{2k} \end{vmatrix} = 0.$$

The mass m_i concentrated at x_i is given by

$$m_i = \sum_{l=0}^{k} c_l^{(i)} M_l \tag{2}$$

where

$$Q^{(i)}(x) \equiv \sum_{l=0}^{k} c_l^{(i)} x^l \equiv \frac{Q_d(x)}{Q'(x_i)(x - x_i)}.$$

If $\theta_i < d < \lambda_i$ the spectrum of F_d consists of ∞ and the roots x_i of

$$R_d(x) = R_{d,k}(x) = \begin{vmatrix} x & d & M_1 & \cdot & M_{k-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x^{k+1} & d^{k+1} & M_{k+1} & \cdot & M_{2k-1} \end{vmatrix} = 0.$$

The mass m_i at x_i is given by

$$m_i = \sum_{l=0}^{k} b_l^{(i)} M_l \tag{3}$$

where

$$R^{(i)}(x) \equiv \sum_{l=0}^{k} b_l^{(i)} x^l \equiv \frac{R_d(x)}{R_d'(x_1)(x - x_i)}.$$

Proof. If

$$\theta_{i-1} < d < \lambda_i \qquad \text{or} \qquad \theta_k < d$$

then the characteristic distribution F_d of Proposition 1 is by Proposition 3 a characteristic distribution for the Stieltjes problem, and the masses are as stated. Suppose on the other hand that F_d as guaranteed by Theorem 2 had mass at d and the k other points $x_i, \cdots, x_k$ all different from zero or infinity. Then $Q_d(x)$ is orthogonal to all polynomials of degree $k - 1$ or less. Thus

$$0 = \int_{0-}^{\infty} Q_d(x)(x - x_1) \cdots (x - x_{k-1})\, dF_d(x) = m_k Q_d(x_k)$$

since $Q_d(d) = 0$. Hence x_k is a root of $Q_d(x)$ and so must the other x_i be, and we must have all roots of $Q_d(x)$ nonnegative in this case. Thus in the case of a negative root of Q_d we must have the spectrum of F_d consisting of at most 0, d, infinity and the $k - 1$ points $x_1, \cdots, x_{k-1}$. Now $R_d(x)$ is orthogonal to all polynomials of degree $k - 2$ or less and vanishes at zero and d. Consequently

$$0 = \int_{0-}^{\infty} R_d(x)(x - x_1) \cdots (x - x_{k-2})\, dF_d(x) = m_{k-1} R_d(x_{k-1})$$

and x_{k-1} is a root of R_d. Similarly with the remaining x_i's.

Thus we need only verify the expression for the masses m_i, which follows immediately from the fact that $R^{(i)}(x)$ is of degree less than n and vanishes at all spectral points of F_d except x_i where it has the value one.

A similar argument gives

THEOREM 3′. *Let n be odd and $k = (n - 1)/2$. Then if*

$$\theta_{i-1} < d < \lambda_i \qquad or \qquad \theta_k < d$$

with λ_i and θ_i as in Lemma 4, the spectrum of $F_d(x)$ consists of infinity and the roots x_i of $Q_{d,k}(x) = 0$, and the masses m_i at x_i are given by (2). If $\lambda_i < d < \theta_i$ the spectrum of F_d consists of the roots x_i of $R_{d,k+1}(x) = 0$ and the masses m_i at x_i are given by (3) with $R_{d,k}(x)$ replaced by $R_{d,k+1}(x)$.

4. Some special cases. If X is a random variable with algebraic moments M_1 and M_2 given, we can use Propositions 1, 2, and 3 to calculate the inequalities

$$0 \leqq \Pr.\,(x \leqq d) \leqq \frac{M_2 - M_1^2}{M_2 - M_1^2 + (M_1 - d)^2} \qquad d \geqq M_1$$

and

$$1 - \frac{M_2 - M_1}{M_2 - M_1^2 + (M_1 - d)^2} \leqq \Pr.\,(X \leqq d) \leqq 1 \qquad d \geqq M_1.$$

This is the well known inequality of Tchebycheff.

If we know in addition that X is a positive random variable (say the absolute value of another variable) then Theorems 1, 2, and 3 enable one to calculate

$$0 \leqq \Pr.\,(X \leqq d) \leqq \frac{M_2 - M_1^2}{M_2 - M_1^2 + (M_1 - d)^2} \qquad 0 \leqq d \leqq M_1$$

$$1 - \frac{M_1}{d} \leqq \Pr.\,(X \leqq d) \leqq 1 \qquad M_1 \leqq d \leqq \frac{M_2}{M_1}$$

$$1 - \frac{M_2 - M_1^2}{M_2 - M_1^2 + (M_1 - d)^2} \leqq \Pr.\,(X \leqq d) \leqq 1, \qquad \frac{M_2}{M_1} \leqq d$$

which are the Cantelli inequalities [7]. We see that they are an improvement over the Tchebycheff inequalities in the region $M_1 \leqq d \leqq M_2/M_1$.

5. Unimodal distributions. In this section we determine the (sharp) upper and lower bounds for $\Pr.\,(X \leqq d)$ of a unimodal distribution when its first two moments about the mode are given. It should be emphasized that this is a distinct problem from that in which the moments of orders α and 2α are given, since one of the transformations

$$X^* = X^{1/\alpha}; \qquad X = (X^*)^{1/\alpha}$$

does not take unimodal distributions into unimodal distributions if $d \neq 1$, although the methods used below should work equally well in the case of any two moments.

Mathematically, we consider distribution functions of the form

$$F(X) = F(0) + \int_0^X \varphi(t)\, dt \tag{4}$$

where $\varphi(t)$ is a nonincreasing function of t, and we are given

$$\int_{0-}^{\infty} dF(X) = 1, \qquad \int_{0-}^{\infty} X\, dF(X) = M_1, \qquad \int_{0-}^{\infty} X^2\, dF(x) = M_2. \tag{5}$$

Every distribution function F of the form (4) is the limit of distributions of the form

$$F = \sum m_i F_i, \qquad m_i \geqq 0, \qquad \sum m_i = 1, \tag{6}$$

where the F_i are rectangular distributions, that is,

$$F_i(X) = \begin{cases} \dfrac{X}{t_i} & X \leqq t_i \\ 1 & X \geqq t_i \end{cases} \tag{7}$$

or

$$F_0 = \begin{cases} 0 & X = 0 \\ 1 & X > 0. \end{cases}$$

For brevity we call these the rectangular distribution at t_i or at zero, and say that F has rectangular distributions of strength m_i at t_i.

We consider the space $\mathfrak{F}$ of all convex combinations of five rectangular distributions whose second moment does not exceed some fixed constant M and compactify it (as in Section 2) by the addition of an infinitesimal distribution at infinity. Then we have the following lemma by the method of Lemma 2.

LEMMA 5. *Let $F(X)$ be a unimodal distribution on $[0, \infty]$ and d a given positive number. Then there is a distribution F^* in $\mathfrak{F}$ for which*

$$F(d) = F^*(d),$$

$$\int_{0-}^{\infty} X\, dF(X) = \int_{0-}^{\infty} X\, dF^*(X), \qquad \int_{0-}^{\infty} X^2\, dF(X) = \int_{0-}^{\infty} X^2\, dF^*(X).$$

Thus we need only consider convex combinations of a finite number of rectangular distributions. Since the upper limit for Pr. $(X \leqq d)$ is a continuous function of d, we consider only cases in which there is a rectangular distribution at a finite point larger than d and take 1 for the sharp upper limit of Pr. $(X \leqq d)$ after this upper limit once becomes one.

If X is a positive random variable with a unimodal distribution function and moments M_1 and M_2, we normalize by taking $x = X/2M_1$. Then x has median $\frac{1}{2}$ and second moment $M_2/4M_1^2 = \Lambda/3$. It is known [1] that a necessary and sufficient condition for M_1 and M_2 to be the first two moments of a unimodal distribution is that $\Lambda \geqq 1$. If now $F = \sum m_i F_i + \lambda F_\infty$, then

$$1 = \sum m_i \qquad 1 = \sum m_i t_i \qquad \Lambda = \sum m_i^2 + \lambda.$$

Suppose that F had rectangular distributions on at least two finite points t_1 and t_2 which are greater than d. Then

$$P = \text{Pr.}\ (x \leqq d) = \left(\frac{m_1}{t_1} + \frac{m_2}{t_2}\right) d + c$$

where c is independent of m_1, m_2, t_1 and t_2. In order for P to be a maximum or a minimum with the total mass, first and second moments fixed, the Jacobian of these quantities with respect to m_1, m_2, t_1 and t_2 must vanish. But this Jacobian is

$$\begin{vmatrix} \frac{d}{t_1} & \frac{d}{t_2} & \frac{-dm_1}{t_1^2} & \frac{dm_1}{t_2^2} \\ 1 & 1 & 0 & 0 \\ t_1 & t_2 & m_1 & m_2 \\ t_1^2 & t_2^2 & 2m_1 t_1 & 2m_2 t_2 \end{vmatrix} = \frac{m_1 m_2 d}{t_1^2 t_2^2} (t_1 - t_2)^2 \neq 0.$$

Hence an extremal distribution function must have a rectangular distribution on exactly one finite point $t > d$. Suppose now that an extremal distribution had rectangular distributions at two points t_1, t_2 which are less than or equal to d, and one of which (say t_1) is different from zero and d. Then the Jacobian with respect to t_1, m_1, m_2, and m_3 would have to vanish. But this Jacobian is

$$\begin{vmatrix} 0 & 1 & 1 & \frac{d}{t} \\ 0 & 1 & 1 & 1 \\ m_1 & t_1 & t_2 & t \\ 2m_1 t & t_1^2 & t_2^2 & t^2 \end{vmatrix} = -m\left(1 - \frac{d}{t}\right)(t_1 - t_2)^2 \neq 0.$$

In a similar manner it is readily verified that an extremal distribution which has a rectangular distribution at a point t_1, $0 < t_1 < d$, can have no rectangular distribution at infinity, and that no extremal distribution can have rectangular distributions at zero, d and infinity. Thus an extremal distribution must belong to one of the following cases.

CASE 1. Rectangular distributions at exactly two points t_1 and t with $0 < t_1 < d < t$. Then

$$P = \Pr_{\cdot t}(x \leqq d)\, m_1 + \frac{m_2 d}{t} \tag{8}$$

and

$$1 = m_1 + m_2 \qquad 1 = m_1 t_1 + m_2 t \qquad \Lambda = m_1 t_1^2 + m_2 t^2. \tag{9}$$

In order for P to be an extremum

$$0 = \begin{vmatrix} 1 & \frac{d}{t} & 0 & \frac{-m_2 d}{t^2} \\ 1 & 1 & 0 & 0 \\ t_1 & t & m_1 & m_2 \\ t_1^2 & t^2 & 2m_1 t_1 & 2m_2 t \end{vmatrix} = \frac{m_1 m_2 d}{t^2}(2t^2 - 3td + t_1 d),$$

and hence

$$d = \frac{3t^2}{3t - t_1}. \tag{10}$$

In order for there to be numbers m_1 and m_2 for which equations (9) are satisfied, we must have

$$0 = \begin{vmatrix} 1 & 1 & 1 \\ t_1 & t & 1 \\ t_1^2 & t^2 & \Lambda \end{vmatrix} = \Lambda - t - t_1(1 - t),$$

whence $t_1 = (t - \Lambda)/(t - 1)$. In order for t_1 to be positive and less than t, t must be greater than Λ, solving (9)

$$m_1 = \frac{(t-1)^2}{\Lambda - 1 + (t-1)^2} \qquad m_2 = \frac{\Lambda - 1}{\Lambda 1 + (t-1)^2}$$

and thus

$$P = \frac{\Lambda - 1}{3t^2 - 4t + \Lambda} \tag{11}$$

and

$$d = 2\,\frac{t^3 - t^2}{3t^2 - 4t + \Lambda}.$$

Since $t \geqq \Lambda$, Case 1 can only arise when $d \geqq 2\Lambda/3$.

Case 2. Rectangular distributions at zero, d, t and possibly infinity. In order to have an extremal distribution we would need

$$\begin{vmatrix} 1 & 1 & \frac{d}{t} & \frac{-m_3 d}{t^2} \\ 1 & 1 & 1 & 0 \\ 0 & d & t & m_3 \\ 0 & d^2 & t^2 & 2m_3 t \end{vmatrix} = -\frac{2m_3}{t}(d-t)^2 \neq 0.$$

Thus Case 2 never occurs.

Case 3. Rectangular distributions at d, t and infinity. Then for an extremal distribution

$$0 = \begin{vmatrix} 1 & \frac{d}{t} & \frac{-m_2 d}{t^2} & 0 \\ 1 & 1 & 1 & 0 \\ d & t & m_2 & 0 \\ d^2 & t^2 & 2m_2 t & 1 \end{vmatrix} = \frac{m_2(t-d)^2}{t^2} \neq 0.$$

Thus Case 3 can never arise.

Case 4. Rectangular distributions at d and t only. Then

$$(12) \qquad m_1 + m_2 = 1 \qquad m_1 d + m_2 t = 1 \qquad m_1 d^2 + m_2 t^2 = \Lambda$$

and

$$0 = \begin{vmatrix} 1 & 1 & 1 \\ d & t & 1 \\ d^2 & t^2 & \Lambda \end{vmatrix} = \Lambda - d - t(1-d),$$

or

$$(13) \qquad t = \frac{\Lambda - d}{1 - d}.$$

In order to have $0 < d < t$, we must have $d < 1$. Solving (12) using (13), gives

$$m_1 = \frac{\Lambda - 1}{\Lambda - 1 + (1-d)^2} \qquad m_2 = \frac{\Lambda - 1}{\Lambda - 1 + (1-d)^2}$$

and
$$P = \text{Pr.}\,(x \leqq d) = m_1 + m_2 \frac{d}{t} = 1 - \frac{(1-d)^2}{\Lambda - d}.$$

Case 5. Rectangular distributions at zero, t and infinity with $0 < d < t$. Then

$$m_0 + m_1 = 1 \qquad m_1 t = 1 \qquad m_1 t^2 + \lambda = \Lambda$$

whence $m_1 = 1/t$, and consequently $1 \leqq t \leqq \Lambda$. Now

$$P = \Pr.\ (x \leqq d) = m_0 + m_1 \frac{d}{t} = 1 - \frac{1}{t} + \frac{d}{t^2},$$

and P can be an extremum only if $t = 1$, $t = \Lambda$, or $dP/dt = 0$. In the first alternative we have

$$P = d, \qquad d \leqq 1.$$

The second alternative gives

$$P = 1 - \frac{1}{\Lambda} + \frac{d}{\Lambda^2}, \qquad d \leqq \Lambda.$$

The last alternative arises when $1/t^2 = 2d/t^3$ or $d = t/2$. Then

$$P = 1 - 1/4\,d, \qquad \tfrac{1}{2} \leqq d \leqq \Lambda/2.$$

We summarize these results in the following theorem.

THEOREM 4. *If $F(x)$ is a unimodal distribution whose first and second moments are $1/2$ and $\Lambda/3$, then*

$$F(d) \leqq 1 - \frac{(1-d)^2}{\Lambda - d} \qquad \textit{for } 0 \leqq d \leqq 1$$

$$F(d) \leqq 1 \qquad d \geqq 1$$

and

$$F(d) \geqq d \qquad 0 \leqq d \leqq \frac{1}{2}$$

$$F(d) \geqq 1 - \frac{1}{4d} \qquad \frac{1}{2} \leqq d \leqq \frac{\Lambda}{2}$$

$$F(d) \geqq 1 - \frac{1}{\Lambda} + \frac{d}{\Lambda^2} \qquad \frac{\Lambda}{2} \leqq d \leqq \frac{2\Lambda}{3}$$

$$F(d) \geqq 1 - \frac{\Lambda - 1}{3t^2 - 4t + \Lambda} \qquad \frac{2\Lambda}{3} \leqq d$$

where

$$d = 2\frac{t^3 - t^2}{3t^2 - 4t + \Lambda}.$$

These inequalities are the best possible.

Removing the normalization from F, we get the following theorem.

THEOREM 4′. *If $F(x)$ is unimodal distribution with moments M_1 and M_2, then*

$$F(d) \leqq 1 - \frac{(2M_1 - d)^2}{3M_2 - dM_1} \qquad 0 \leqq d \leqq 2M_1$$

$$F(d) \leqq 1 \qquad 2M_1 \leqq d$$

and

$$F(d) \geqq \frac{d}{2M_1} \qquad 0 \leqq d \leqq M_1$$

$$F(d) \geq 1 - \frac{M_1}{2d} \qquad M_1 \leqq d \leqq \frac{3M_2}{4M_1}$$

$$F(d) \geqq 1 - \frac{4M_1^2}{3M_2} + \frac{8M_1^3 d}{9M_2^2} \qquad \frac{3M_2}{4M_1} \leqq d \leqq \frac{M_2}{M_1}$$

$$F(d) \geqq 1 - \frac{\Lambda - 1}{3t^2 - 4t + \Lambda} \qquad \frac{M_2}{M_1} \leqq d$$

where

$$\Lambda = \frac{3M_2}{4M_1^2} \qquad \text{and} \qquad d = 4M_1 \frac{t^3 - t^2}{3t^2 - 4t + \Lambda}.$$

REFERENCES

[1] N. L. Johnson and C. A. Rogers, "The moment problem for unimodal distributions," *Ann. Math. Stat.* Vol. 22 (1951), pp. 433–439.

[2] A. Markoff, "Demonstration de certaines inégalites de Tchebycheff," *Math. Ann.*, Vol. 24 (1884), pp. 172–180.

[3] A. Markoff, "Sur une question de maximum et de minimum proposée par M. Tchebycheff," *Acta Math.*, Vol. 9 (1886), pp. 57–70.

[4] J. Shohat and J. Tamarkin, *The Problem of Moments*, American Mathematical Society, 1943.

[5] T. J. Stieltjes, "Quelques recherches sur les quadratures dites mécaniques," *Ann. Sci. École Norm. Sup.*, Vol. 3 (1884), pp. 409–426.

[6] G. Szegö, *Orthogonal Polynomials*, American Mathematical Society, 1939.

[7] A. Wald, "Generalization of the inequality of Markoff," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 244–255.

[8] A. Wald, "Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments," *Trans. Amer. Math. Soc.*, Vol. 46 (1939), pp. 280–306.

[9] M. G. Krein, "The ideas of P. L. Chebyshev and A. A. Markov in the theory of limiting values of integrals and their further development," *Uspehi Matem. Nauk*, (N.S.) 6, No. 4, Vol. 44, (1951), pp. 3–120.