BOUNDS FOR THE DISTRIBUTION FUNCTION OF A SUM OF INDEPENDENT, IDENTICALLY DISTRIBUTED RANDOM VARIABLES[1]

by

Wassily Hoeffding and S. S. Shrikhande
University of North Carolina and University of Nagpur

Institute of Statistics
Mimeograph Series No. 84
Limited Distribution
October, 1953

---

[1]This research was supported by the United States Air Force under Contract No. AF 18(600)-458 monitored by Headquarters, Air Research and Development Command.

Bounds for the distribution function of a sum of independent, identically distributed random variables.*

By Wassily Hoeffding and S. S. Shrikhande,
University of North Carolina and University of Nagpur

Summary. The problem is considered of obtaining bounds for the cumulative distribution function of the sum of n independent, identically distributed random variables with k prescribed moments and given range. For $n = 2$ it is shown that the best bounds are attained or arbitrarily closely approached with discrete random variables which take on at most $2k + 2$ values. Explicit bounds are obtained for the case of nonnegative random variables with given mean when $n = 2$; for arbitrary values of n bounds are given which are asymptotically best in the "tail" of the distribution. Some of the results contribute to the more general problem of obtaining bounds for the expected value of a given function of independent, identically distributed random variables when the expected values of certain functions of the individual variables are given.

1. Introduction. The problem considered in this paper is part of the following general problem. Let $\underline{\underline{D}}$ be the class of all cdfs (cumulative distribution functions) $F(x)$ on the real line which satisfy the conditions

---

*This research was supported by the United States Air Force under Contract No. AF 18(600)-458 monitored by Headquarters, Air Research and Development Command.

$$(1.1) \qquad \int g_i(x)\, dF(x) = c_i, \qquad i = 1, \ldots, k,$$

$$(1.2) \qquad F(x) = 0 \text{ if } x < A, \qquad F(x) = 1 \text{ if } x > B ,$$

where the functions $g_1(x), \ldots, g_k(x)$ and the constants $c_1, \ldots, c_k$, A, B are given. We allow that $A = -\infty$ and/or $B = \infty$. Here and in what follows, when the domain of integration is not indicated, the integral extends over the entire range of the variables involved.

Let $K(x_1, \ldots, x_n)$ be a function such that

$$\psi(F) = \int \cdots \int K(x_1, \ldots, x_n)\, dF(x_1) \ldots dF(x_n)$$

exists for all F in $\underline{\underline{D}}$ in the sense that the multiple integral is equal to the repeated integral taken in an arbitrary order. The problem is to determine upper and lower bounds for $\psi(F)$ when F is in $\underline{\underline{D}}$.

For $n = 1$, $g_i(x) = x^i$, $K(x) = 1$ or 0 according as $x \leq t$ or $> t$, as well as for other functions $K(x)$, an extensive literature on the subject exists, associated with the names of Bienaymé, Chebyshev, Markov, Stieltjes, and others; for references see Shohat and Tamarkin [6]. The case $n = 1$, $g_i(x) = x^{m_i}$, $A = 0$, $B = \infty$ was treated by Wald [7].

For n arbitrary, Robbins [5] showed that the Bienaymé - Chebyshev bound for Prob. ($|X_1 + \ldots + X_n| \geq t$), where the X_i are independent and identically distributed with zero mean and given variance, can be improved when $n > 1$. Plackett [4] and Hartley and David [2] obtained the best possible bounds for the expected sample range and the expected value of the largest observation in the case where the mean and the variance are given, assuming that the common cdf is continuous. A problem analogous to the general problem stated above, but without the assumption that the n variables are identically distributed, was considered by one of the authors [3] who showed that under general conditions the best bounds are attained or arbitrarily closely approached with step functions in $\underline{\underline{D}}$ which have at most $k + 1$ steps.

In the present paper the attention is concentrated on the case where $k = 1$ or 0 according as a given function $f(x_1, \ldots, x_n)$ is or is not contained in a given set. The method used permits to obtain the closest bounds only for $n = 2$. If $f = x_1 + \ldots + x_n$, n is even, and $g_i(x) = x^i$, the bounds for $n = 2$ can be applied in an obvious way, but will not in general be the best ones. More general functions K are considered only insofar as they can be handled by the same method.

Theorem 2.1 states conditions under which we need consider only step functions in $\underline{\underline{D}}$. Theorems 2.2 and 2.3 show that for functions $K(x,y)$ of a certain type we may restrict our attention to step functions with a bounded number of steps. In Theorem 3.1 an explicit expression for the least upper bound of Prob. ($X + Y \geq t$) is obtained when X and Y are independent, identically distributed and nonnegative, with given mean. In section 4 bounds for the analogous case with n summands are considered.

2. The least upper bound of $\iint K(x,y)\, dF(x)\, dF(y)$. Let $\underline{\underline{D}}$ be the class of cdfs $F(x)$ which satisfy the conditions (1.1) and (1.2). Let $K(x,y)$ be a function such that

$$\psi(F) = \iint K(x,y)\, dF(x)\, dF(y) \tag{2.1}$$

exists for all F in $\underline{\underline{D}}$, in the sense that the double integral equals the repeated integral. The problem is to determine the least upper and the greatest lower bound of $\psi(F)$ for all F in $\underline{\underline{D}}$. As long as the function $K(x,y)$ is not specified it is sufficient to consider the least upper bound.

Let $\underline{\underline{D}}^*$ be the class of all F in $\underline{\underline{D}}$ which are step functions with a finite number of steps. The following obvious theorem gives sufficient conditions for confining our attention to functions in $\underline{\underline{D}}^*$.

Theorem 2.1. Suppose that

A) for every F in $\underline{\underline{D}}$ and every $\delta > 0$ there exists an F^* in $\underline{\underline{D}}^*$ such that

$$\sup_x \left| F^*(x) - F(x) \right| < \delta \quad ;$$

B) for every F in $\underline{\underline{D}}$ and every $\varepsilon > 0$ there exists a $\delta > 0$ such that for any G in $\underline{\underline{D}}$ for which

$$\sup_x \left| F(x) - G(x) \right| < \delta$$

we have

$$\left| \psi(F) - \psi(G) \right| < \varepsilon.$$

Then

$$\sup_{F \varepsilon \underline{\underline{D}}} \psi(F) = \sup_{F \varepsilon \underline{\underline{D}}^*} \psi(F).$$

The proof is omitted.

Assumption A) is satisfied if $g_i(x) = x^{m_i}$, where $m_1, \ldots, m_k$ are arbitrary integers, that is, if $\underline{D}$ is the class of cdfs with k prescribed moments and given range. This follows from Lemma 3.1 in [3].

A particular case in which assumption B) is satisfied is where $K(x,y) = 1$ or 0 according as (x,y) is or is not contained in a Borel set S with the property that the sets of all points x such that $(x,y) \varepsilon S$ and the sets of all y such that $(x,y) \varepsilon S$ are unions of a finite and bounded number of intervals. This is a special case of Theorem 4.1 in [3].

We shall now derive sufficient conditons under which, given a step function F in $\underline{\underline{D}}$ with m steps, we can construct a step function G in $\underline{\underline{D}}$ with less than m steps such that $\psi(G) \geq \psi(F)$.

A step function F in $\underline{\underline{D}}$ with exactly m steps is of the form

(2.2) $$F(x) = P_j \quad \text{if } a_j \leq x < a_{j+1}, \quad j=0,1,\ldots,m,$$

where

(2.3) $$-\infty = a_0 < a_1 < a_2 < \ldots < a_m < a_{m+1} = \infty ,$$

$$A \leq a_1 , \qquad a_m \leq B ,$$

(2.4) $$0 = P_0 < P_1 < \ldots < P_{m-1} < P_m = 1 ,$$

(2.5) $$\sum_{j=1}^{m} g_i(a_j)(P_j - P_{j-1}) = c_i, \qquad i = 1, \ldots, k .$$

If we define

$$h_{ij} = g_i(a_j) - g_i(a_{j+1}), \quad i = 1, \ldots, k; \ j = 1, \ldots, m-1,$$

the last equations can be written

$$\sum_{j=1}^{m-1} h_{ij} P_j = c_i - g_i(a_m), \qquad i = 1, \ldots, k .$$

Let

$$(2.7) \qquad G(x) = P_j + tD_j \quad \text{if } a_j \leq x < a_{j+1}, \qquad j=0,1,\ldots,m.$$

In order that G(x) be a cdf in $\underline{\underline{D}}$ it is sufficient that the numbers t, D_j satisfy the conditions

$$(2.8) \qquad D_0 = D_m = 0,$$

$$(2.9) \qquad 0 \leq P_1 + tD_1 \leq P_2 + tD_2 \leq \ldots \leq P_{m-1}+tD_{m-1} \leq 1,$$

$$(2.10) \qquad \sum_{j=1}^{m-1} h_{ij} D_j = 0, \qquad i=1,\ldots,k.$$

If F and G are defined by (2.2) and (2.7), we have

$$(2.11) \qquad \psi(F) = \sum_{i=1}^{m} \sum_{j=1}^{m} K_{ij} (P_i - P_{i-1})(P_j - P_{j-1}),$$

where

$$(2.12) \qquad K_{ij} = K(a_i, a_j),$$

and

$$(2.13)\quad \psi(G) - \psi(F) = t \sum_{j=1}^{m-1} L_j D_j + t^2 \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} L_{ij} D_i D_j ,$$

where

$$(2.14)\quad L_j = \sum_{i=1}^{m} (K_{ij} + K_{ji} - K_{i,j+1} - K_{j+1,i}) (P_i - P_{i-1}),$$

$$(2.15)\quad L_{ij} = K_{ij} - K_{i+1,j} - K_{i,j+1} + K_{i+1,j+1} \;.$$

Lemma 2.1. *Let* F *be a step function in* $\underline{D}$ *with exactly* m *steps, defined by* (2.2) *to* (2.5), *where* $m > k+1$. *Suppose that the integers* $u_1, \ldots, u_{k+1}$ *can be so chosen that*

$$1 \leq u_1 < u_2 < \ldots < u_{k+1} \leq m-1$$

and the equations

$$(2.16)\quad \sum_{r=1}^{k+1} h_{iu_r} x_r = 0, \qquad i = 1, \ldots, k,$$

imply

$$(2.17) \qquad \sum_{r=1}^{k+1} \sum_{s=1}^{k+1} L_{u_r u_s} x_r x_s \geq 0.$$

Then there exists a step function G in D with less than m steps, for which $\psi(G) \geq \psi(F)$.

Proof. Let $G(x)$ be defined by (2.7), and let

$$D_j = 0 \qquad \text{for } j \neq u_1, \ldots, u_{k+1}.$$

Let $\lambda = 1$ or 0 according as the rank of the matrix

$$\begin{matrix} h_{1u_1} & \cdots & h_{1u_{k+1}} \\ \cdots\cdots & & \\ h_{ku_1} & \cdots & h_{ku_{k+1}} \\ L_{u_1} & \cdots & L_{u_{k+1}} \end{matrix}$$

is equal to or less than k+1. Then the equations (2.16) and

$$\sum_{r=1}^{k+1} L_{u_r} x_r = \lambda$$

have a solution $(D_{u_1}, \ldots, D_{u_{k+1}}) \neq (0, \ldots, 0)$. Having thus fixed the D_j, let t be the largest number which satisfies the inequalities (2.9). This number exists and is positive. With this choice of the numbers t and D_j, G is a step function in $\underline{\underline{D}}$ with less than m steps. Furthermore, by (2.13),

$$\Psi(G) - \Psi(F) = t\lambda + t^2 \sum_{r=1}^{k+1} \sum_{s=1}^{k+1} L_{u_r u_s} D_{u_r} D_{u_s} \geq 0.$$

The proof is complete.

The next theorem shows that if K(x,y) is of a certain form, and if we restrict ourselves to the class $\underline{\underline{D}}^*$ of step functions in $\underline{\underline{D}}$ with a finite number of steps, we need consider only step functions with a bounded number of steps.

Let $\underline{\underline{D}}_m$ be the class of all F in $\underline{\underline{D}}$ which are step functions with at most m steps.

<u>Theorem 2.2.</u> <u>Suppose that</u> K(x,y) <u>is of the form</u>

$$K(x,y) = \sum_{i=0}^{k} \sum_{j=0}^{k} a_{tij}\, g_i(x)\, g_j(y)$$

$$\underline{\text{if}}\ b_{t-1} \leq f(x,y) < b_t, \qquad t=1,\ldots,s,$$

<u>where</u> $g_0(x) = 1$, <u>the</u> a_{tij} <u>are arbitrary constants</u>,

$$-\infty = b_0 < b_1 < \ldots < b_{s-1} < b_s = \infty,$$

<u>and</u> $f(x,y)$ <u>is a strictly increasing function in each of its arguments when the other argument is fixed. Then</u>

$$\sup_{F \varepsilon \underline{\underline{D}}^*} \psi(F) = \sup_{F \varepsilon \underline{\underline{D}}_{sk+s}} \psi(F).$$

<u>The theorem remains true if in the inequalities</u> $b_{t-1} \leq f(x,y) < b_t$ <u>some signs</u> $\leq$ <u>are replaced by</u> $<$ <u>or vice versa, provided that the</u> s <u>sets defined by the inequalities cover the entire plane.</u>

Proof. Let $F(x)$, as defined by (2.2) to (2.5), be an arbitrary step function in $\underline{\underline{D}}$ with exactly m steps, where $m > sk + s$. It is sufficient to construct a step function G in $\underline{\underline{D}}$ with less than m steps such that $\psi(G) \geq \psi(F)$. Let m_t $(t=1,\ldots,s)$ denote the number of indices u $(1 \leq u \leq m)$ for which

$$b_{t-1} \leq f(a_u, a_u) < b_t.$$

Then

$$s \max (m_t) \geq m_1 + \dots + m_s = m > s(k+1).$$

Hence there exists a t for which $m_t \geq k + 2$ and an integer n such that

$$b_{t-1} \leq f(a_n, a_n) < f(a_{n+k+1}, a_{n+k+1}) < b_t .$$

The assumption about f(x,y) implies that

$$K_{vw} = \sum_{i=0}^{k} \sum_{j=0}^{k} a_{tij}\, g_i(a_v)\, g_j(a_w) \qquad \text{for } n \leq v,\ w \leq n+k+1.$$

By (2.15) and (2.6) this implies

$$L_{vw} = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{tij}\, h_{iv}\, h_{jw} \qquad \text{for } n \leq v,\ w \leq n+k.$$

Hence if we let $u_r = n+r-1$, $r=1,2,\dots,k+1$, the conditions of Lemma 2.1 are satisfied. The proof is complete.

If $g_i(x) = x^i$, that is, if $\underline{\underline{D}}$ is the class of distributions with given moments up to order k and given range, the assumption of Theorem 2.2 means that K(x,y) is piecewise polynomial, of bounded degrees, in sections of the plane separated by curves of negative slope. It is of interest to note that if K(x,y) is piecewise polynomial in sections separated by curves of positive slope, a similar reduction of the problem to the case of step functions with a bounded number of steps is in general impossible. For example, let

$$K(x,y) = \max (x,y),$$

and let D be the class of cdfs F with

$$\int x \, dF(x) = 0, \qquad \int x^2 \, dF(x) = 1.$$

Under the restriction to continuous functions F(x) this is a special case of a problem considered by Hartley and David [2]. For an arbitrary cdf F(x) we can write

$$\psi(F) = 2\int x \, \bar{F}(x) \, dF(x),$$

where

$$\bar{F}(x) = (F(x-0)+F(x+0))/2.$$

Using Schwarz's inequality, we have for any constant c and any F in $\underline{\underline{D}}$

$$\psi(F) + c = 2 \int (x+c)\, \bar{F}(x)\, dF(x)$$

$$\leq 2 \left\{ \int (x+c)^2\, dF(x) \right\}^{1/2} \left\{ \int \bar{F}(x)^2\, dF(x) \right\}^{1/2}$$

$$= 2(1+c^2)^{1/2} \left\{ \int \bar{F}(x)^2\, dF(x) \right\}^{1/2}.$$

If F(x) is continuous, $\int \bar{F}(x)^2\, dF(x) = 1/3$, and the bound

$$\psi\,(F) \leq \min_c \left\{ 2.3^{-1/2}\,(1 + c^2)^{1/2} - c \right\}$$

is attained with a continuous cdf in $\underline{\underline{D}}$, as shown by Hartley and David.

Now let F(x) be a step function with at most m steps which takes on the values $0 = P_0 \leq P_1 \leq \dots \leq P_{m-1} \leq P_m = 1$. Then

$$4 \int \bar{F}(x)^2\, dF(x) = \sum_{j=1}^{m} (P_{j-1} + P_j)^2\, (P_j - P_{j-1}) \quad ,$$

and this can be written

$$12 \int \bar{F}(x)^2\, dF(x) = 4 - \sum_{j=1}^{m} p_j^3 \quad , \qquad p_j = P_j - P_{j-1} \quad .$$

The conditions $\Sigma p_j = 1$, $p_j \geq 0$ imply $\Sigma p_j^3 \geq m^{-2}$. Hence

$$\int \bar{F}(x)^2 \, dF(x) \leq \frac{1}{3} - \frac{1}{12m^2} ,$$

and the Hartley-David bound cannot be arbitrarily closely approached with a step function in $\underline{\underline{D}}$ having a bounded number of steps.

Combining Theorems 2.1 and 2.2 we can state that if the conditions of both theorems are satisfied, then

$$\sup_{F \varepsilon \underline{\underline{D}}} \psi(F) = \sup_{F \varepsilon \underline{\underline{D}}_{sk+s}} \psi(F) .$$

In particular the conditions of Theorem 2.2 are fulfilled if $\psi(F) = P_F \left\{ f(X,Y) \geq c \right\}$ or $P_F \left\{ |f(X,Y)| \geq c \right\}$, etc., where $P_F \left\{ \ldots \right\}$ is the probability of the event in braces when X and Y are independent with the common cdf F, and f(x,y) has the property stated in the theorem. Referring to the remarks following the statement of Theorem 2.1, we obtain

<u>Theorem 2.3</u>. <u>Let $\underline{\underline{D}}$ be the class of cdfs</u> F(x) <u>which satisfy the conditions</u>

$$\int x^{m_i} \, dF(x) = c_i \quad , \qquad i = 1, \ldots, k \quad ,$$

$$F(x) = 0 \text{ if } x < A, \quad F(x) = 1 \text{ if } x > B ,$$

with given integers $m_1, \ldots, m_k$ and given numbers $c_1, \ldots, c_k$, A, B, where we may have $A = -\infty$ and/or $B = \infty$. Let $f(x,y)$ be a strictly increasing function in each of its arguments when the other argument is fixed. Then

$$\sup_{F \varepsilon \underline{\underline{D}}} P_F \left\{ f(X,Y) \geq c \right\} = \sup_{F \varepsilon \underline{\underline{D}}_{2k+2}} P_F \left\{ f(X,Y) \geq c \right\} .$$

3. The least upper bound of $P(X + Y \geq t)$ when X and Y are independent, identically distributed and nonnegative with given mean. As an application of the results of section 2 we shall prove the following theorem.

Theorem 3.1. Let X and Y be two independent random variables with common cdf $F(x)$. Let $\underline{\underline{D}}$ be the class of cdfs F with $\int x \, dF(x) = \mu$ and $F(x) = 0$ for $x < 0$, where $\mu > 0$. Then

$$\sup_{F \varepsilon \underline{\underline{D}}} P_F(X + Y \geq c\mu) = 1 \qquad \text{if } c \leq 2 ,$$

(3.1)
$$= \frac{4}{c^2} \qquad \text{if } 2 \leq c \leq \frac{5}{2} ,$$

$$= \frac{2}{c} - \frac{1}{c^2} \qquad \text{if } \frac{5}{2} \leq c .$$

The three bounds are attained with the respective distributions

$$P(X = \mu) = 1 \quad ;$$

$$P(X = 0) = 1 - \frac{2}{c} \quad , \quad P(X = \frac{c}{2}\mu) = \frac{2}{c} \quad ;$$

$$P(X = 0) = 1 - \frac{1}{c} \quad , \quad P(X = c\mu) = \frac{1}{c} \quad .$$

Theorem 3.1 should be compared with the solution by Birnbaum, Raymond and Zuckerman [1] of the analogous problem without the restriction that X and Y be identically distributed. Let $M(t, \lambda, \mu)$ be the least upper bound of $P(X + Y \geq t)$ when X and Y are independent, nonnegative, and have the respective means $\lambda \leq \mu$. According to [1, Theorem 1.2] we have

$$M(t, \lambda, \mu) = 1 \qquad \text{if} \quad t \leq \lambda + \mu$$

$$(3.2) \qquad = \frac{\mu}{t-\lambda} \qquad \text{if} \quad \lambda + \mu \leq t \leq \tfrac{1}{2}(\lambda+2\mu+\sqrt{\lambda^2+4\mu^2}\,)$$

$$= \frac{\lambda+\mu}{t} - \frac{\lambda\mu}{t^2} \qquad \text{if} \quad \tfrac{1}{2}(\lambda+2\mu+\sqrt{\lambda^2+4\mu^2}\,) \leq t \quad .$$

Obviously

$$\sup_{F \in \underline{\underline{D}}} P_F(X + Y \geq c\mu) \leq M(c\mu, \mu, \mu) = M(c, 1, 1) ,$$

and we have

$$M(c, 1, 1) = 1 \qquad \text{if} \quad c \leq 2$$

$$= \frac{1}{c-1} \qquad \text{if} \quad 2 \leq c \leq \frac{1}{2}(3 + \sqrt{5})$$

$$= \frac{2}{c} - \frac{1}{c^2} \qquad \text{if} \quad \frac{1}{2}(3 + \sqrt{5}) \leq c \quad .$$

Hence the bound (3.1) is smaller than the Birnbaum-Raymond-Zuckerman bound if and only if $2 < c < (3 + \sqrt{5}) / 2$.

Proof of Theorem 3.1. We may and shall assume that $\mu = 1$. By Theorem 2.3 we need consider only cdfs F in $\underline{\underline{D}}$ which are step functions with $m \leq 4$ steps. Then F is of the form

$$F(x) = P_j \text{ if } a_j \leq x < a_{j+1} , \qquad j = 0, 1, \ldots, m,$$

where

$$0 = a_0 \leq a_1 < a_2 < \ldots < a_m < a_{m+1} = \infty \quad ,$$

$$0 = P_0 < P_1 < P_2 < \dots < P_m = 1 \quad ,$$

$$\Sigma_1^m a_j(P_j - P_{j-1}) = 1 \quad .$$

We have

$$K(x,y) = 1 \qquad \text{if} \quad x + y \geq c \quad ,$$

$$= 0 \qquad \text{if} \quad x + y < c \quad .$$

Hence the numbers $K_{ij} = K(a_i, a_j)$ satisfy the conditions

$$K_{ij} = 0 \text{ or } 1 \ , \qquad K_{ij} = K_{ji} \quad ,$$

$$K_{ij} \leq K_{i'j} \quad \text{if} \quad i < i' \quad .$$

The sequence $(K_{11}, K_{22}, \dots, K_{mm})$ consists of a sequence of zeros followed by a sequence of ones. The reasoning used in the proof of Theorem 2.2 shows that any distribution for which there are more than two consecutive zeros or more than two consecutive ones in this sequence can be replaced by a distribution with less than m steps which does not decrease the value of $\psi(F)$.

Hence for $m = 4$ we need consider only matrices $\| K_{ij} \|$ of the four types

$$\begin{matrix} 0 & 0 & 0 & . \\ 0 & 0 & 0 & . \\ 0 & 0 & 1 & 1 \\ . & . & 1 & 1 \end{matrix} \qquad \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{matrix} \qquad \begin{matrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{matrix} \qquad \begin{matrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{matrix}$$

where the numbers represented by dots need not be specified.

The corresponding matrices $\| L_{ij} \|$ are

I	II	III	IV
$\begin{matrix} 0 & 0 & . \\ 0 & 1 & . \\ . & . & 0 \end{matrix}$	$\begin{matrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ -1 & 0 & 0 \end{matrix}$	$\begin{matrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{matrix}$

We shall apply Lemma 2.1 to show that in every case there exists a cdf in $\underline{\underline{D}}$ with at most three steps which does not decrease the value of $\psi(F)$. It is sufficient to find integers u, v $(1 \leq u < v \leq 3)$ such that the equation

$$(3.3) \qquad (a_u - a_{u+1})\, x + (a_v - a_{v+1}) y = 0$$

implies

$$(3.4) \qquad L_{uu}x^2 + 2L_{uv}xy + L_{vv}y^2 \geq 0 \quad .$$

In case I inequality (3.4) is satisfied with $u = 1$, $v = 2$, and in cases II and IV with $u = 1$, $v = 3$. In case III let $u = 1$, $v = 3$. Then the left hand side of (3.4) is $-2xy$, and this is nonnegative by (3.3) since $a_j - a_{j+1} < 0$.

Hence we may confine our attention to step functions in $\underline{\underline{D}}$ with $m \leq 3$ steps.

If $m = 3$, we have to consider the matrices $\| K_{ij} \|$ of the forms

0 0 0	0 0 0	0 0 1	0 0 0	0 0 1	0 1 1
0 0 0	0 0 1	0 0 1	0 1 1	0 1 1	1 1 1
0 0 1	0 1 1	1 1 1	0 1 1	1 1 1	1 1 1

The corresponding matrices $\| L_{ij} \|$ are

A	B	C	D	E	F
0 0	0 1	0 0	1 0	1 -1	-1 0
0 1	1 -1	0 -1	0 0	-1 0	0 0

In applying Lemma 2.1 we have to take $u = 1$, $v = 2$ and to show that (3.3) implies (3.4). This is true for the matrices A, D and E. In the cases B, C and F Lemma 2.1 is not applicable.

In case C, $\psi(F) = 1 - P_2^2$. If G(x) is defined by (2.7) with m = 3, we have

$$\psi(G) - \psi(F) = - tD_2 [2(P_2+tD_2) - tD_2] ,$$

where t and D_2 satisfy the conditions

(3.5) $$(a_1 - a_2)D_1 + (a_2 - a_3)D_2 = 0 ,$$

(3.6) $$0 \leq P_1 + tD_1 \leq P_2 + tD_2 \leq 1 .$$

Let $D_2 = -1$. Then D_1 is given by (3.5). Let t be the largest number which satisfies (3.6). Then t > 0, G is in $\underline{\underline{D}}_2$, and $\psi(G) - \psi(F) \geq 0$.

In case F, $\psi(F) = 1 - P_1^2$, and a similar reasoning shows that this case also can be reduced to a step function with at most two steps.

The only remaining case with m = 3 is case B. Here we can write

$$\psi(F) = 2p_2p_3 + p_3^2 ,$$

where (admitting the possibility that F has less than three steps)

(3.7) $$p_1 + p_2 + p_3 = 1, \qquad a_1p_1 + a_2p_2 + a_3p_3 = 1 ,$$

$$(3.8) \qquad p_1 \geq 0 , \qquad p_2 \geq 0 , \qquad p_3 \geq 0 ,$$

$$(3.9) \qquad 0 \leq a_1 \leq a_2 \leq a_3 \quad ,$$

$$(3.10) \qquad a_1 + a_3 < c, \qquad 2\, a_2 < c, \qquad c \leq a_2 + a_3.$$

Expressing $\psi(F)$ in terms of a_1, a_2, a_3, p_1, we get

$$\psi(F) = (1 - p_1)^2 - p_2^2 ,$$

where

$$p_2 = \frac{a_3 - 1 - (a_3 - a_1)p_1}{a_3 - a_2} .$$

If a_2, a_3 p_1 are held fixed, $\psi(F)$ is a decreasing function of a_1. Hence we maximize $\psi(F)$ by choosing the least possible value for a_1. This is the greatest of the bounds given by the inequalities $p_j \geq 0$ and $a_1 \geq 0$. If this bound is given by one of the equations $p_j = 0$, we get a distribution in $\underline{\underline{D}}_2$. Hence we may assume that the least value is $a_1 = 0$. With $a_1 = 0$,

$$p_2 = \frac{a_3 - 1 - a_3 p_1}{a_3 - a_2} \quad ,$$

and $\psi(F)$ is a decreasing function of a_2 when a_3 and p_1 are fixed. The only lower bound for a_2 which does not necessarily correspond to a distribution in $\underline{\underline{D}}_2$ is $a_2 = c - a_3$. In this case

$$p_2 = \frac{(1-p_1)a_3-1}{2a_3-c} = \frac{1-p_1}{2} + \frac{c(1-p_1)-2}{2(2a_3-c)} ,$$

which is a monotonic function of a_3 (possibly a constant) when p_1 is held fixed. Hence the maximum is attained at one of the endpoints of the range of a_3. This range is given by the inequalities (3.8) to (3.10) with $a_1 = 0$, $a_2 = c - a_3$. Its endpoints correspond either to distributions in $\underline{\underline{D}}_2$ or (if given by $a_1 + a_3 = c$ or $2a_2 = c$) to cases where the value of $\psi(F)$ exceeds $2p_2p_3 + p_3^2$ and which already have been disposed of.

Thus we need consider only cdfs in $\underline{\underline{D}}_2$.

If $c \leq 2$, we have $\psi(F) = 1$ for the cdf in $\underline{\underline{D}}_1$ which has a single step at $x = 1$. Thus

$$\sup_{F \in \underline{\underline{D}}} \psi(F) = 1 \qquad \text{if} \quad c \leq 2 . \tag{3.11}$$

Henceforth we assume that $c > 2$.

A distribution F in $\underline{\underline{D}}_2$ assigns to the points a_1, a_2 the respective probabilities

$$p_1 = \frac{a_2-1}{a_2-a_1} \quad , \qquad p_2 = \frac{1-a_1}{a_2-a_1} \quad ,$$

where

$$0 \leq a_1 \leq 1 \leq a_2 \quad .$$

If $c \leq 2a_1$, we have $c \leq 2$, a case already disposed of. If $c > 2a_2$, then $\Psi(F) = 0$, a case which may be disregarded. We are left with the two cases

(i) $2a_1 < c \leq a_1 + a_2$, (ii) $a_1 + a_2 < c \leq 2a_2$.

In case (i),

$$\Psi(F) = 1-p_1^2 \quad ,$$

a decreasing function of a_1. The lower bound for a_1 is $\max(0, c-a_2)$.

If $a_1 = 0 \geq c-a_2$, then

$$p_1 = 1 - \frac{1}{a_2} \quad ,$$

so that $\Psi(F)$ is a decreasing function of a_2. The lower bound for a_2 is max (1, c) = c, and we obtain

$$\psi(F) = 1 - (1 - \frac{1}{c})^2 = \frac{2}{c} - \frac{1}{c^2} .$$

If $a_1 = c - a_2 \geq 0$, then

$$p_1 = \frac{a_2 - 1}{2a_2 - c} = \frac{1}{2} + \frac{c-2}{2(2a_2 - c)} ,$$

so that $\psi(F)$ is an increasing function of a_2. Since $a_2 \leq c$, we obtain the same maximum of $\psi(F)$ as in the previous case.

In case (ii),

$$\psi(F) = p_2^2 ,$$

a decreasing function of a_2. Hence we let $a_2 = c/2$. Then $\psi(F)$ is a decreasing function of a_1, and hence is maximized for $a_1 = 0$. We get

$$\psi(F) = \frac{4}{c^2} .$$

Hence

$$(3.12) \qquad \sup_{F \in \underline{\underline{D}}} \psi(F) = \max \left\{ \frac{2}{c} - \frac{1}{c^2} , \frac{4}{c^2} \right\} \quad \text{if} \quad c > 2 .$$

Theorem 3.1 now follows from (3.11), (3.12) and the stated conditions under which the bounds are attained.

4. Bounds for $P(X_1 + \dots + X_n \geq c)$. Let $\bar{X}_n = n^{-1}(X_1 + \dots + X_n)$, and let $\omega_n(t)$ denote the least upper bound of $P(\bar{X}_n \geq t\mu)$ when $X_1, \dots, X_n$ are independent, identically distributed and nonnegative with mean μ. It is easily seen that for every n

$$\omega_n(t) = 1 \qquad \text{if } t \leq 1 ,$$

$$\omega_{sn}(t) \leq \omega_s(t), \qquad s = 1, 2, \dots$$

By Markov's inequality

$$\omega_1(t) = \frac{1}{t} \qquad \text{if } 1 \leq t.$$

By Theorem 3.1

$$\omega_2(t) = \frac{1}{t^2} \qquad \text{if } 1 \leq t \leq \frac{5}{4} ,$$

$$= \frac{1}{t} - \frac{1}{4t^2} \text{ if } \frac{5}{4} \leq t .$$

Let $\omega_n^*(t)$ be the least upper bound of $P(\overline{X}_n \geq t\mu)$ when $X_1, \ldots, X_n$ are independent and nonnegative with common mean μ. Clearly

$$\omega_n(t) \leq \omega_n^*(t) .$$

It can be verified (cf. also [1, Corollary 2.2]) that

$$\omega_n^*(t) \leq M\,(nt,\ m,\ n-m), \qquad 2m \leq n,$$

where $M(t, \lambda, \mu)$ is defined by (3.2). In particular, with $n = 2m$,

$$\omega_n^*(t) \leq \frac{1}{t} - \frac{1}{4t^2} \qquad \text{if } \frac{3+\sqrt{5}}{4} \leq t, \text{ n even,}$$

and with $n = 2m + 1$,

$$\omega_n^*(t) \leq \frac{1}{t} - \frac{n^2-1}{n^2}\ \frac{1}{4t^2} \quad \text{if} \quad \frac{3n+1+\sqrt{5n^2+6n+5}}{4n} \leq t ,$$

which holds for n odd and, a fortiori, for n even.

On the other hand, for any random variables X_i which satisfy our assumptions, $P(\overline{X}_n \geq t\mu)$ is a lower bound for $\omega_n(t)$. In particular, if $nt \geq 1$ and $X_i = 0$ or nt with respective probabilities $1 - (nt)^{-1}$ and $(nt)^{-1}$, we get

$$\omega_n(t) \geq 1 - \left(1 - \frac{1}{nt}\right)^n > \frac{1}{t} - \frac{n-1}{2n}\,\frac{1}{t^2} \quad .$$

Hence we have for all positive integers n

$$(4.1) \quad \omega_n(t) = \frac{1}{t} - \frac{1+\theta}{4}\,\frac{n-1}{n}\,\frac{1}{t^2} \quad \text{if} \quad \frac{3n+1+\sqrt{5n^2+6n+5}}{4n} \leq t \ ,$$

where

$$\frac{1}{n} \leq \theta < 1 \quad ,$$

and

$$(4.2) \qquad \omega_{2n}(t) = \frac{1}{t} - \frac{1+\theta'}{4}\,\frac{1}{t^2} \qquad \text{if} \quad \frac{5}{4} \leq t \quad ,$$

where

$$0 \leq \theta' < 1 - \frac{2}{n} \quad .$$

Equation (4.1) is also true for $\omega_n^*(t)$, and (4.2) holds for

$\omega_n^*(t)$ if $(3 + \sqrt{5})/4 \leq t$.

Thus for large values of t the known bounds for $\omega_n(t)$ and $\omega_n^*(t)$ cannot be substantially improved.

[1] Z. W. Birnbaum, J. Raymond, and H. S. Zuckerman, "A generalization of Tshebyshev's inequality to two dimensions", Annals of Mathematical Statistics, Vol. 18 (1947), pp. 70-79.

[2] H. O. Hartley and H. A. David "Universal bounds for mean range and extreme observation", Annals of Math. Stat., ...

[3] W. Hoeffding, "The extrema of the expected value of a function of independent random variables", ...

[4] R. L. Plackett, "Limits of the ratio of mean range to standard deviation", Biometrika, Vol. 34 (1947), pp. 120-122.

[5] H. Robbins, "Some remarks on the inequality of Tchebycheff", Courant Anniversary Volume, 1948, pp. 345-350.

[6] J. A. Shohat and J. D. Tamarkin, The Problem of Moments, American Mathematical Society, 1943.

[7] 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.