THE EXTREMA OF THE EXPECTED VALUE OF A FUNCTION OF INDEPENDENT RANDOM VARIABLES[1]

By Wassily Hoeffding

University of North Carolina

Summary. The problem is considered of determining the least upper (or greatest lower) bound for the expected value $EK(X_1, \cdots, X_n)$ of a given function K of n random variables $X_1, \cdots, X_n$ under the assumption that $X_1, \cdots, X_n$ are independent and each X_j has given range and satisfies k conditions of the form $Eg_i^{(j)}(X_j) = c_{ij}$ for $i = 1, \cdots, k$. It is shown that under general conditions we need consider only discrete random variables X_j which take on at most $k + 1$ values.

1. Introduction. Let $\mathcal{C}$ be the class of n-dimensional dfs (distribution functions) $F(\mathbf{x}) = F(x_1, \cdots, x_n)$ which satisfy the conditions

$$F(x_1, \cdots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n), \tag{1.1}$$

$$\int g_i^{(j)}(x)\, dF_j(x) = c_{ij}, \qquad i = 1, \cdots, k; j = 1, \cdots, n, \tag{1.2}$$

$$F_j(x) = \begin{cases} 0 & \text{if } x < A_j, \\ 1 & \text{if } x > B_j, \end{cases} \qquad j = 1, \cdots, n, \tag{1.3}$$

where the functions $g_i^{(j)}(x)$ and the constants c_{ij}, A_j, and B_j are given. We allow that $A_j = -\infty$ and/or $B_j = \infty$. Here and in what follows, when the domain of integration is not indicated, the integral extends over the entire range of the variables involved. It will be understood that all dfs are continuous on the right.

Let $K(\mathbf{x})$ be a function such that

$$\phi(F) = \int K(\mathbf{x})\, dF(\mathbf{x})$$

exists and is finite for all F in $\mathcal{C}$. The problem is to determine the least upper and the greatest lower bound of $\phi(F)$ when F is in $\mathcal{C}$. It will be sufficient to consider only the least upper bound.

Special cases of statistical interest include $K(\mathbf{x}) = 0$ or 1 according as a function $f(\mathbf{x})$ does or does not exceed a given constant; $K(\mathbf{x}) = \max(x_1, \cdots, x_n)$, etc.; $g_i^{(j)}(x) = x^i$ (given moments up to order k); $g_i^{(j)}(x) = 1$ or 0 according as

Received November 3, 1953, revised November 30, 1954.

[1] This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

$x < b_i$ or $x \geqq b_i$ (given quantiles); etc. For

$$n = 1, \qquad g_i^{(1)}(x) = x^i, \qquad K(x) = \begin{cases} 1 \text{ if } x \leqq t, \\ K(x) = 0 \text{ otherwise} \end{cases}$$

the problem was stated and its solution found by Tchebycheff. This result was extended to more general function $K(x)$ by Markov, Possé, and others. References and proofs are given by Shohat and Tamarkin [9]. An extension to the case $g_i^{(1)}(x) = x^{m_i}$ and $A_1 = 0$ was considered by Wald [10]. Recent contributions are due to Karlin and Shapley [7] and Royden [8].

For n arbitrary, $k = 1$, $g_i^{(j)} = x$, $A_j = 0$, $B_j = \infty$, and $K(x) = 1$ or 0 according as $\sum_1^n x_i \geqq t$ or $< t$, Birnbaum, Raymond, and Zuckerman [3] showed that when looking for the least upper bound of $\phi(F)$ we need consider only dfs F_j which are step-functions with at most two steps. They gave an explicit solution for $n = 2$. For the case $k = 3$, $g_i^{(j)}(x) = x^i$ for $i = 1, 2$, and $g_3^{(j)}(x) = |x - c_{1j}|^3$, with $\phi(F)$ the distribution function of the sum $\sum_1^n x_i$, the inequalities of Berry [2], Esseen [4], and Bergström [1] give bounds which are asymptotically best as $n \to \infty$ but can presumably be improved for finite n.

In the present paper it is shown that if, in the general problem as stated above, we restrict ourselves to the subclass $\mathcal{C}^*$ of $\mathcal{C}$ where the F_j are step-functions with a finite number of steps, then we need consider only step-functions with at most $k + 1$ steps (Theorem 2.1); the number $k + 1$ can in general not be reduced. The same result holds in the unrestricted problem if (A) each F in $\mathcal{C}$ can be approximated (in a certain sense) by a step-function in $\mathcal{C}^*$, and (B) $\phi(F)$ is, in a sense, a continuous function of F (Theorem 2.2). Sufficient conditions for the fulfillment of assumptions (A) and (B) are given in Sections 3 and 4.

2. The main theorems. Denote by $\mathcal{C}^*$ the class of all $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ in $\mathcal{C}$ such that $F_1, \cdots, F_n$ are step-functions with a finite number of steps. Let $\mathcal{C}_m$ be the subclass of $\mathcal{C}^*$ in which each F_j for $j = 1, \cdots, n$ is a step-function with at most m steps.

THEOREM 2.1.

$$\sup_{F \varepsilon \mathcal{C}^*} \phi(F) = \sup_{F \varepsilon \mathcal{C}_{k+1}} \phi(F).$$

PROOF. Let $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ be an arbitrary df in $\mathcal{C}^*$ such that for some j, F_j has more than $k + 1$ steps. It is sufficient to show that there exists a df G in $\mathcal{C}_{k+1}$ such that $\phi(G) \geqq \phi(F)$. This, in turn, will easily follow when we show that if $F_n(x)$ has $m > k + 1$ steps, there exists a df $H_n(x)$ such that

a) $H_n(x)$ has less than m steps;

b) $H(\mathbf{x}) = F_1(x_1) \cdots F_{n-1}(x_{n-1})H_n(x_n)$ is in $\mathcal{C}$;

c) $\phi(H) \geqq \phi(F)$.

By assumption $F_n(x)$ is of the form

$$F_n(x) = \begin{cases} 0 & \text{if} \quad x < a_1 ; \\ p_1 + \cdots + p_r & \text{if } a_r \leqq x < a_{r+1}, \quad r = 1, \cdots, m-1; \\ 1 & \text{if } a_m \leqq x; \end{cases}$$

where

$$A_n \leqq a_1 < a_2 < \cdots < a_m \leqq B_n ;$$

$$p_r > 0, \qquad r = 1, \cdots, m;$$

$$g_{i1}p_1 + g_{i2}p_2 + \cdots + g_{im}p_m = c_{in}, \quad i = 0, 1, \cdots, k;$$

$$c_{0n} = 1; \quad g_{0r} = 1, \qquad r = 1, \cdots, m;$$

$$g_{ir} = g_i^{(n)}(a_r), \quad i = 1, \cdots, k; \quad r = 1, \cdots, m.$$

Let

$$H_n(x) = \begin{cases} 0 & \text{if} \quad x < a_1 ; \\ (p_1 + td_1) + \cdots + (p_r + td_r) & \\ & \text{if } a_r \leqq x < a_{r+1} ; \quad r = 1, \cdots, m-1; \\ 1 & \text{if } a_m \leqq x. \end{cases}$$

In order to satisfy condition b) it is sufficient to choose t, and $d_1, \cdots, d_r$ in such a way that

$$p_r + td_r \geqq 0, \qquad r = 1, \cdots, m; \tag{2.1}$$

$$g_{i1}d_1 + g_{i2}d_2 + \cdots + g_{im}d_m = 0, \quad i = 0, 1, \cdots, k. \tag{2.2}$$

We can write $\phi(H) - \phi(F) = t\sum_{r=1}^{m} K_r d_r$ where

$$K_r = \int K(x_1, \cdots, x_{n-1}, a_r) \quad d\left\{\prod_{j=1}^{n-1} F_j(x_j)\right\}.$$

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

$$\begin{matrix} g_{01}, & \cdots, & g_{0m} \\ \cdots & \cdots & \cdots \\ g_{k1}, & \cdots, & g_{km} \\ K_1, & \cdots, & K_m \end{matrix}$$

is less than or equal to $k + 2$. Then the equations (2.2) and $\sum_1^m K_r d_r = \lambda$ have a solution $(d_1, \cdots, d_m) \neq (0, \cdots, 0)$. Since $\sum_1^m d_r = 0$, at least one component of the solution $d_1, \cdots, d_m$ must be negative. Having thus fixed $d_1, \cdots, d_m$, we choose t as the largest number which satisfies the inequalities (2.1). This number exists and is positive. Conditions a), b) and c) are now satisfied, and the proof is complete.

Let the distance $d(F, G)$ between two dfs $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ and $G(\mathbf{x}) = G_1(x_1) \cdots G_n(x_n)$ be defined by

$$d(F, G) = \max_{1 \leqq j \leqq n} \sup_{-\infty < x < \infty} |F_j(x) - G_j(x)|.$$

We shall make the following assumptions.

ASSUMPTION A. For every F in $\mathcal{C}$ and every $\delta > 0$ there exists an F^* in $\mathcal{C}^*$ such that $d(F, F^*) < \delta$.

ASSUMPTION B. For every F in $\mathcal{C}$ and every $\epsilon > 0$ there exists $\delta > 0$ such that for any G in $\mathcal{C}$ which satisfies $d(F, G) < \delta$ we have $|\phi(F) - \phi(G)| < \epsilon$.

LEMMA 2.1. *If Assumptions* A *and* B *are satisfied,*

$$\sup_{F \varepsilon \mathcal{C}} \phi(F) = \sup_{F \varepsilon \mathcal{C}^*} \phi(F).$$

The proof is obvious. An analogous theorem clearly holds with an arbitrary distance function.

From Theorem 2.1 and Lemma 2.1 we obtain

THEOREM 2.2. *If Assumptions* A *and* B *are satisfied,*

$$\sup_{F \varepsilon \mathcal{C}} \phi(F) = \sup_{F \varepsilon \mathcal{C}_{k+1}} \phi(F).$$

In Sections 3 and 4 it will be shown that Assumptions A and B are satisfied for certain classes $\mathcal{C}$ and functions K which are of interest in statistics.

We conclude this section by showing that Theorem 2.1 cannot be improved without imposing additional restrictions. More precisely, for every k and every n there exist functions K, $g_i^{(j)}$ and constants A_j, B_j, c_{ij} such that the conditions of the theorem are satisfied and $\sup_{\mathcal{C}} \phi(F) > \sup_{\mathcal{C}_m} \phi(F)$ if $m < k + 1$. Furthermore, the functions $g_i^{(j)}$ and the constants A_j, B_j, c_{ij} can be chosen to be independent of j. Let

$$g_i^{(j)}(x) = x^i, \qquad c_{ij} = c_i, \qquad A_j = A, \qquad B_j = B, \qquad j = 1, \cdots, n,$$

where A and B are finite.

First assume $n = 1$. We can choose the constants c_i and $k + 1$ distinct real numbers $a_1, a_2, \cdots, a_{k+1}$ in $[A, B]$ in such a way that the equations

$$a_1^i p_1 + a_2^i p_2 + \cdots + a_{k+1}^i p_{k+1} = c_i, \qquad i = 0, 1, \cdots, k,$$

where $c_0 = 1$, have a unique solution $(p_1, \cdots, p_{k+1})$ with $p_r > 0$ for all r (see, for example, [9]). Then the df F_0, which assigns probability p_r to the point a_r for all r, is in $\mathcal{C}_{k+1}$. Let $K_1(x)$ be a nonnegative continuous function, and let $Q(x)$ be a polynomial of degree k such that $K_1(x) \leqq Q(x)$ for $A \leqq x \leqq B$ and equality holds if and only if $x = a_r$ for $r = 1, 2, \cdots, k + 1$. Let $\phi_1(F) = \int K_1(x)\, dF(x)$. Then for every F in $\mathcal{C}$,

$$\phi_1(F) \leqq \int Q(x)\, dF(x) = \phi_1(F_0),$$

and equality holds only for $F = F_0$.

Now suppose that $\sup_{\mathcal{C}_m} \phi_1(F) = \phi_1(F_0)$ for some $m < k + 1$. Then there exists a sequence $\{F^{(s)}\}$ of cdfs in $\mathcal{C}_m$ such that $\lim \phi_1(F^{(s)}) = \phi_1(F_0)$. Since A and B are finite, there exists, by the Bolzano-Weierstrass theorem, a subsequence $\{F^{(s_j)}\}$ of $\{F^{(s)}\}$ which converges to a function F^* in $\mathcal{C}_m$ at all points of continuity of F^*. Since $K_1(x)$ is continuous, this implies $\lim \phi_1(F^{(s_j)}) = \phi_1(F^*) = \phi_1(F_0)$. This is a contradiction since $F^* \neq F_0$.

For n arbitrary let $K(x_1, \cdots, x_n) = K_1(x_1)K_1(x_2) \cdots K_1(x_n)$, so that $\phi(F) = \phi_1(F_1) \cdots \phi_1(F_n)$. If the conditions for the case $n = 1$ are satisfied, we arrive at the desired conclusion, making use of the condition $K_1(x) \geqq 0$.

The assumptions that K_1 is continuous and A, B are finite are inessential, at least for $k = 2$. This is seen from the fact that the bound of the Bienaymé-Tchebycheff inequality can not in general be arbitrarily closely approached when the distribution is a two-step-function.

3. Approximation of a df by a step-function. It will now be shown that Assumption A is satisfied if $\mathcal{C}$ is the class of distributions of the product type (1.1) with prescribed moments and ranges.

THEOREM 3.1. *Assumption* A *is satisfied if* $\mathcal{C}$ *is the class defined by* (1.1) *to* (1.3) *with* $g_i^{(j)}(x) = x^{m_{ij}}$, *where the* m_{ij} *are arbitrary positive integers.*

The theorem is an immediate consequence of

LEMMA 3.1. *Let* $F(x)$ *be a df on the real line such that*

$$\int x^i \, dF(x) = c_i, \qquad i = 1, \cdots, s; \tag{3.1}$$

$$F(x) = \begin{cases} 0 & \text{if } x < A, \\ 1 & \text{if } x > B, \end{cases} \tag{3.2}$$

where we may have $A = -\infty$ *or* $B = \infty$. *Then for every* $\delta > 0$ *there exists a cdf* $F^*(x)$ *which is a step-function with a finite number of steps, satisfies conditions* (3.1) *and* (3.2), *and for which*

$$\sup_x |F^*(x) - F(x)| < \delta.$$

To prove Lemma 3.1 we shall need

LEMMA 3.2. *If* $F(x)$ *is any df which satisfies conditions* (3.1) *and* (3.2), *there exists a df which is a step-function with a finite number of steps and satisfies the same conditions.*

The statement of Lemma 3.2 is well known. For example, it follows from Shohat and Tamarkin ([9], Theorems 1.2 and 1.3 and Lemma 3.1).

PROOF OF LEMMA 3.1. Given $\delta > 0$, we can choose a finite set of points

$$A = a_0 < a_1 < a_2 < \cdots < a_m < a_{m+1} = B$$

such that $p_r = F(a_{r+1} - 0) - F(a_r) < \delta$ for $r = 0, 1, \cdots, m$. If $p_r \neq 0$, let

$$F_r(x) = \begin{cases} 0 & \text{if } \quad x < a_r \\ p_r^{-1}[F(x) - F(a_r)] & \text{if } \quad a_r \leqq x < a_{r+1} \\ 1 & \text{if } a_{r+1} \leqq x. \end{cases}$$

By Lemma 3.2 there exists a df $F_r^*(x)$ which is a step-function with a finite number of steps and such that

$$\int x^i \, dF_r^*(x) = \int x^i \, dF_r(x), \qquad i = 1, \cdots, s,$$

$$F_r^*(x) = \begin{cases} 0 & \text{if } x < a_r, \\ 1 & \text{if } x > a_{r+1}. \end{cases}$$

Let

$$F^*(x) = \begin{cases} 0 & \text{if} \quad x < A; \\ F(a_r) + p_r F_r^*(x) & \text{if } a_r \leqq x < a_{r+1}, \quad r = 0, 1, \cdots, m; \\ 1 & \text{if} \quad x \geqq B; \end{cases}$$

where $p_r F_r^*(x) = 0$ if $p_r = 0$. It can be verified that $F^*(x)$ has the properties stated in Lemma 3.1.

4. Continuity of $\phi(F)$. In this section we consider sufficient conditions for the continuity of $\phi(F)$ in the sense of Assumption B. The next theorem shows that Assumption B is satisfied if $\phi(F)$ is the probability that the random vector $\mathbf{X}$ with df F is contained in a set S of a fairly general type.

THEOREM 4.1. *Assumption* B *is satisfied if* $K(x) = 1$ *or* 0 *according as* x *does or does not belong to a Borel set* S *such that every set*

$$S_j(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) = \{x_j : \mathbf{x} \in S, x_h \text{ fixed for } h \neq j\}, \quad j = 1, \cdots, n,$$

is the union of a finite and bounded number of intervals.

(Here $\{x \colon \text{C}\}$ denotes the set of all points x which satisfy condition C.)

PROOF. Let δ be an arbitrary positive number. Let $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ and $G(\mathbf{x}) = G_1(x_1) \cdots G_n(x_n)$ be any two dfs in $\mathfrak{C}$ such that

$$\sup_x |F_h(x) - G_h(x)| < \delta, \qquad h = 1, \cdots, n.$$

Let

$$F^{(j)}(\mathbf{x}) = \left\{\prod_{h=1}^{j} F_h(x_h)\right\} \left\{\prod_{h=j+1}^{n} G_h(x_h)\right\}, \qquad j = 0, 1, \cdots, n,$$

so that $F^{(0)}(\mathbf{x}) = G(\mathbf{x})$ and $F^{(n)}(\mathbf{x}) = F(\mathbf{x})$. Then

$$\phi(F) - \phi(G) = \sum_{j=1}^{n} [\phi(F^{(j)}) - \phi(F^{(j-1)})].$$

We may assume that every set S_j is the union of at most N nonoverlapping intervals, where N is a fixed number. For a fixed integer j and fixed values x_h, with $h \neq j$, denote these intervals by $I_1, I_2, \cdots, I_M$, where $M \leqq N$. We have

$$\phi(F^{(j)}) - \phi(F^{(j-1)})$$
$$= \int L(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) \, d\{F_1(x_1) \cdots F_{j-1}(x_{j-1}) G_{j+1}(x_{j+1}) \cdots G_n(x_n)\},$$

where $L(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n) = \sum_{m=1}^{M} \{\int_{I_m} dF_j(x_j) - \int_{I_m} dG_j(x_j)\}$. Since

$$\left| \int_{I_m} dF_j(x_j) - \int_{I_m} dG_j(x_j) \right| < 2\delta,$$

we get $|\phi(F) - \phi(G)| < 2nM\delta \leqq 2nN\delta$, so that Assumption B is satisfied.

Assumption B of Theorem 2.2 is evidently satisfied under more general conditions than those of Theorem 4.1. Thus if $K(\mathbf{x}) = d_1K_1(\mathbf{x}) + \cdots + d_RK_R(\mathbf{x})$, where $d_1, \cdots, d_R$ are arbitrary constants and $\phi_r(F) = \int K_r(\mathbf{x})\, dF(\mathbf{x})$ is continuous in the sense of Assumption B for $r = 1, \cdots, R$, then $\phi(F) = \int K(\mathbf{x})\, dF(\mathbf{x})$ is also continuous. The same is true if $K(\mathbf{x})$ can be approximated by a function of the form $\sum d_rK_r(\mathbf{x})$, uniformly in the range of the distributions in $\mathfrak{C}$.

Using Theorems 3.1 and 4.1 we can state the following corollary of Theorem 2.2.

THEOREM 4.2. *Let $\mathfrak{C}$ be the class of dfs $F(\mathbf{x}) = F_1(x_1) \cdots F_n(x_n)$ such that*

$$F_j(A_j - 0) = 0, \qquad F_j(B_j + 0) = 1, \qquad \int x^{m_{ij}}\, dF_j(x) = c_{ij},$$

$$i = 1, \cdots, k;\ j = 1, \cdots, n,$$

where the numbers A_j, B_j, c_{ij} and the integers m_{ij} are given. Let S be a Borel set such that every set $\{x_j : \mathbf{x} \in S, x_h \text{ fixed for } h \neq j\}$ is the union of a finite and bounded number of intervals. Then

$$\sup_{F \in \mathfrak{C}} \int_S dF = \sup_{F \in \mathfrak{C}_{k+1}} \int_S dF.$$

5. Concluding remark. The problem considered in this paper can be modified by admitting only those dfs in $\mathfrak{C}$ for which the marginal distributions $F_1, \cdots, F_n$ are identical. Some results for this case were obtained in [6]. With this restriction Theorem 2.1 is no longer true, and the assumptions are no longer sufficient in order to reduce the class of competing dfs to step-functions with a bounded number of steps.

For example, consider the problem of the least upper bound for the expected value of the largest of n independent, identically distributed random variables with given mean and variance. Hartley and David [5] showed that under the additional assumption that the df is continuous the least upper bound is attained with a continuous df when $n \geqq 2$. At least for $n = 2$ it can be shown [6] that the Hartley-David bound cannot be arbitrarily closely approached with a discrete df having a bounded number of steps.

On the other hand, if the assumption that the random variables are identically distributed is dropped, Theorem 2.2 implies that the least upper bound is attained or approached with step-functions having at most three steps.

REFERENCES

[1] H. BERGSTRÖM, "On the central limit theorem in the case of not equally distributed random variables", *Skand. Aktuarietids.*, Vol. 32 (1949), pp. 37–62.

[2] A. C. BERRY, "The accuracy of the Gaussian approximation to the sum of independent variates", *Trans. Amer. Math. Soc.*, Vol. 49 (1941), pp. 122–136.

[3] Z. W. BIRNBAUM, J. RAYMOND AND H. S. ZUCKERMAN, "A generalization of Tshebyshev's inequality to two dimensions", *Ann. Math. Stat.*, Vol. 18 (1947), pp. 70–79.

[4] C. G. ESSEN, "Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law", *Acta Math.*, Vol. 77, (1944), pp. 1–125.

[5] H. O. HARTLEY AND H. A. DAVID, "Universal bounds for mean range and extreme observation", *Ann. Math. Stat.*, Vol. 25 (1954), pp. 85–99.

[6] W. HOEFFDING AND S. S. SHRIKHANDE, "Bounds for the distribution function of a sum of independent, identically distributed random variables", *Ann. Math. Stat.*, to be published.

[7] S. KARLIN AND L. S. SHAPLEY, "Geometry of moment spaces", *Mem. Amer. Math. Soc.*, No. 12 (1953).

[8] H. L. ROYDEN, "Bounds on a distribution function when its first n moments are given", *Ann. Math. Stat.*, Vol. 24 (1953), pp. 361–376.

[9] J. A. SHOHAT AND J. D. TAMARKIN, *The Problem of Moments*, Mathematical Surveys, No. II, American Mathematical Society, New York, 1943.

[10] 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.