

THE EXTREMA OF THE EXPECTED VALUE OF A FUNCTION OF INDEPENDENT
RANDOM VARIABLES¹

by

Wassily Hoeffding
Institute of Statistics, University of North Carolina

Institute of Statistics
Mimeograph Series No. 83
Limited Distribution
October, 1953

1. This research was supported in part by THE UNITED STATES AIR FORCE under Contract AF 18(600)-458 monitored by the Office of Scientific Research.

THE EXTREMA OF THE EXPECTED VALUE OF A FUNCTION OF INDEPENDENT
RANDOM VARIABLES*

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 $EK(X_1, \dots, X_n)$ of a given function K of n random variables X_1, \dots, X_n under the assumption that X_1, \dots, X_n are independent and each X_j has given range and satisfies k conditions of the form $Eg_i^{(j)}(X_j) = c_{ij}$, $i = 1, \dots, k$. 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 \underline{C} be the class of n -dimensional cdfs (cumulative distribution functions) $F(\underline{x}) = F(x_1, \dots, x_n)$ which satisfy the conditions

$$(1.1) \quad F(x_1, \dots, x_n) = F_1(x_1) F_2(x_2) \dots F_n(x_n),$$

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

$$(1.3) \quad F_j(x) = 0 \text{ if } x < A_j, \quad F_j(x) = 1 \text{ if } x > B_j, \quad j = 1, \dots, n,$$

where the functions $g_i^{(j)}(x)$ and the constants c_{ij} , A_j , B_j are given. We

*This research was supported in whole or in part by the United States Air Force under Contract No. AF 18(600)-458 monitored by Headquarters, Air Research and Development Command, P. O. Box 1395, Baltimore 3, Maryland.

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 odfs are continuous on the right.

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

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

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

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

For $n = 1$, $g_1^{(1)}(x) = x^i$, $K(x) = 1$ if $x \leq t$, $K(x) = 0$ otherwise,

the problem was stated and its solution found by Chebyshev. This result was extended to more general function $K(x)$ by Markov, Possé and others.

For references and proofs see Shohat and Tamarkin [6]. An extension to the case $g_1^{(1)}(x) = x^{m_1}$, $A_1 = 0$ was considered by Wald [7]. A recent contribution is due to Royden [5].

For n arbitrary, $k = 1$, $g_1^{(j)} = x$, $A_j = 0$, $B_j = \infty$, $K(\underline{x}) = 1$ or 0

according as $\sum_{i=1}^n x_i \geq t$ or $< t$, Birnbaum, Raymond and Zuckerman [2] showed that when looking for the least upper bound of $\phi(F)$ we need consider only cdfs F_j which are step functions with at most two steps. They gave an explicit solution for $n = 2$. For the case $k = 3$, $g_i^{(j)}(x) = x^i$, $\phi(F)$ the distribution function of the sum $\sum_{i=1}^n x_i$, the inequality of Berry [1] and Esseen [3] gives bounds which are asymptotically best as $n \rightarrow \infty$ 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 $\underline{\underline{C}}^*$ of $\underline{\underline{C}}$ 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 same result holds in the unrestricted problem if (A)

each F in $\underline{\underline{C}}$ can be approximated (in a certain sense) by a step function in $\underline{\underline{C}}^*$, 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. Let $\underline{\underline{C}}$ be the class of cdfs $F(\underline{x})$ defined by (1.1) to (1.3), and suppose that $\phi(F) = \int K(\underline{x}) dF(\underline{x})$ exists for all F in $\underline{\underline{C}}$. Denote by $\underline{\underline{C}}^*$ the class of all $F(\underline{x}) = F_1(x_1) \dots F_n(x_n)$ in $\underline{\underline{C}}$ such that F_1, \dots, F_n are step functions with a finite number of steps. Let $\underline{\underline{C}}_m$ be the subclass of $\underline{\underline{C}}^*$ in which each F_j ($j = 1, \dots, n$) is a step function with at most m steps.

Theorem 2.1. $\sup_{F \in \underline{C}^*} \phi(F) = \sup_{F \in \underline{C}_{k+1}} \phi(F)$.

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

- a) $H_n(x)$ has less than m steps;
- b) $H(\underline{x}) = F_1(x_1) \dots F_{n-1}(x_{n-1}) H_n(x_n)$ is in \underline{C} ;
- c) $\phi(H) \geq \phi(F)$.

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

$$\begin{aligned} F_n(x) &= 0 && \text{if } x < a_1 \\ &= p_1 + \dots + p_r && \text{if } a_r \leq x < a_{r+1}, r = 1, \dots, m-1 \\ &= 1 && \text{if } a_m \leq x, \end{aligned}$$

where

$$A_n \leq a_1 < a_2 < \dots < a_m \leq B_n ,$$

$$p_r > 0, \quad r = 1, \dots, m ,$$

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

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

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

Let

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

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

$$(2.1) \quad p_r + td_r \geq 0, \quad r = 1, \dots, m,$$

$$(2.2) \quad g_{i1}d_1 + g_{i2}d_2 + \dots + g_{im}d_m = 0, \quad i = 0, 1, \dots, k$$

We can write

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

where

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

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

$$g_{01}, \dots, g_{0m}$$

.....

$$g_{k1}, \dots, g_{km}$$

$$K_1, \dots, K_m$$

is less than or equal to $k + 2$. Then the equations (2.2) and

$$\sum_{r=1}^m K_r d_r = \lambda$$

have a solution $(d_1, \dots, d_m) \neq (0, \dots, 0)$. Having thus fixed d_1, \dots, d_m , 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.

Theorem 2.2. Suppose that the following two conditions are satisfied.

A) For every $F(\underline{x}) = F_1(x_1) \dots F_n(x_n)$ in \underline{C} and every $\delta > 0$ there exists a cdf $F^*(\underline{x}) = F_1^*(x_1) \dots F_n^*(x_n)$ in \underline{C}^* such that

$$\sup_x | F_j^*(x) - F_j(x) | < \delta, \quad j = 1, \dots, n.$$

B) For every $F(\underline{x}) = F_1(x_1) \dots F_n(x_n)$ in \underline{C} and every $\varepsilon > 0$ there exists $\delta > 0$ such that for any $G(\underline{x}) = G_1(x_1) \dots G_n(x_n)$ in \underline{C} which satisfies the inequalities

$$\sup_x | F_j(x) - G_j(x) | < \delta, \quad j = 1, \dots, n,$$

we have

$$| \phi(F) - \phi(G) | < \varepsilon$$

Then

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

Proof. Conditions A) and B) imply that

$$\sup_{F \in \underline{C}} \phi(F) = \sup_{F \in \underline{C}^*} \phi(F) .$$

The theorem now follows from Theorem 2.1 .

It is easy to extend Theorems 2.1 and 2.2 to the case where the number k of restrictions $\int g_i^{(j)}(x) dF_j(x) = c_{ij}$, $i = 1, \dots, k$, depends on j .

Conditions A) and B) make use of the distance between two cdfs F and G defined by

$$\max_j \sup_x | F_j(x) - G_j(x) | .$$

Obviously an analogous theorem can be stated with an arbitrary distance $d(F, G)$.

In sections 3 and 4 it will be shown that assumptions A) and B) of Theorem 2.2 are satisfied for certain classes \underline{C} and functions K which are of interest in statistics.

3. Approximation of a cdf by a step function. It will now be shown that condition A) of Theorem 2.2 is satisfied if \underline{C} is the class of distributions of the product type (1.1) with prescribed moments and ranges.

Theorem 3.1. Condition A) of Theorem 2.2 is satisfied if \underline{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 the following lemma.

Lemma 3.1. Let $F(x)$ be a cdf on the real line such that

$$(3.1) \quad \int x^i dF(x) = c_i, \quad i = 1, \dots, s$$

$$(3.2) \quad F(x) = 0 \text{ if } x < A, \quad F(x) = 1 \text{ if } x > B,$$

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 cdf which satisfies conditions (3.1) and (3.2), there exists a cdf 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 [6, Theorems 1.2 and 1.3 and Lemma 3.1]. Cf. also Royden [5, Lemma 2].

Proof of Lemma 3.1. Given $\delta > 0$, we can choose a finite set of points

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

such that

$$p_r = F(a_{r+1} - 0) - F(a_r) < \delta, \quad r = 0, 1, \dots, m.$$

If $p_r \neq 0$, let

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

By Lemma 3.2 there exists a cdf $F_r^*(x)$ which is a step function with

a finite number of steps and such that

$$\int x^i dF_r^*(x) = \int x^i dF_r(x), \quad i = 1, \dots, s,$$

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

Let

$$F^*(x) = F(a_r) + p_r F_r^*(x) \quad \text{if} \quad a_r \leq x < a_{r+1}, \quad r = 0, 1, \dots, m,$$

$$F^*(x) = 0 \text{ if } x < A, \quad F^*(x) = 1 \text{ if } x \geq B.$$

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) of Theorem 2.2. The next theorem shows that assumption B) is satisfied if $\phi(F)$ is the probability that the random vector \underline{X} with cdf F is contained in a set S of a fairly general type.

Theorem 4.1. Condition B) of Theorem 2.2 is satisfied if
 $K(\underline{x}) = 1$ or 0 according as \underline{x} does or does not belong to a Borel-measurable
set S such that every set $S_j = \{ \underline{x}_j : \underline{x} \in S \}$, $j = 1, \dots, n$, is the
union of a finite and bounded number of intervals.

(Here $\{ x : A \}$ denotes the set of all points x which satisfy condition A.)

Proof. Let δ be an arbitrary positive number. Let $F(\underline{x}) = F_1(x_1) \dots F_n(x_n)$ and $G(\underline{x}) = G_1(x_1) \dots G_n(x_n)$ be any two cdfs in \underline{C} such that

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

Let

$$F^{(j)}(\underline{x}) = \left\{ \prod_{h=1}^j F_h(x_h) \right\} \left\{ \prod_{h=j+1}^n G_h(x_h) \right\}, \quad j = 0, 1, \dots, n,$$

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

$$\varphi(F) - \varphi(G) = \sum_{j=1}^n \int \varphi'(F^{(j)}) - \varphi(F^{(j-1)}) \int$$

We may assume that every set S_j is the union of at most N non-overlapping intervals, where N is a fixed number. For a fixed integer j and

fixed values x_h ($h \neq j$) denote these intervals by I_1, I_2, \dots, I_M , where $M \leq N$. We have

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

where

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

Since

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

we get

$$\left| \varphi(F) - \varphi(G) \right| < 2nM\delta \leq 2nN\delta,$$

so that condition B) is satisfied.

Condition B) of Theorem 2.2 is evidently satisfied under more general assumptions than those of Theorem 4.1. Thus if $K(\underline{x}) =$

$d_1 K_1(\underline{x}) + \dots + d_R K_R(\underline{x})$, where d_1, \dots, d_R are arbitrary constants and $\phi_r(F) = \int K_r(\underline{x}) dF(\underline{x})$ is continuous in the sense of assumption B) for $r = 1, \dots, R$, then $\phi(F) = \int K(\underline{x}) dF(\underline{x})$ is also continuous. The same is true if $K(\underline{x})$ can be approximated by a function of the form $\sum d_r K_r(\underline{x})$, uniformly in the range of the distributions F in \underline{C} .

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

Theorem 4.2. Let \underline{C} be the class of cdfs $F(\underline{x}) = F_1(x_1) \dots F_n(x_n)$

such that $F_j(A_j - 0) = 0$, $F_j(B_j + 0) = 1$, $\int x^{m_{ij}} dF_j(x) = c_{ij}$, $i=1, \dots, k$;

$j=1, \dots, n$, where the numbers A_j, B_j, c_{ij} and the integers m_{ij} are given.

Let S be a Borel-measurable set such that every set $\{x_j: x_j \in S\}$ is the

union of a finite and bounded number of intervals. Then

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

5. Concluding remark. The problem considered in this paper can be modified by admitting only those cdfs in \underline{C} for which the marginal distributions F_1, \dots, F_n are identical. It is of interest to note that with this restriction the conditions A), B) of Theorem 2.2 are no longer sufficient in order to reduce the class of competing cdfs 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 [4] showed that under the additional assumption that the cdf is continuous the least upper bound is attained with a continuous cdf when $n \geq 2$. At least for $n = 2$ it can be shown that the Hartley-David bound can not be arbitrarily closely approached with a discrete cdf having a bounded number of steps. On the other hand if the assumption that the random variables are identically distributed is dropped, the conditions of Theorem 2.2 are satisfied, and hence the least upper bound is attained or approached with step functions having at most three steps.

References

- [1] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates", Transactions of the American Mathematical Society, Vol. 49(1941), pp. 122-136.
- [2] 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.
- [3] C. G. Esseen, "Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law", Acta Math., Vol. 77, pp. 1-125 (1944).
- [4] H. O. Hartley and H. A. David, "Universal bounds for mean range and extreme observation",
- [5] H. L. Royden, "Bounds on a distribution function when its first n moments are given", Annals of Mathematical Statistics, Vol. 24 (1953), 361-376.
- [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", Transactions of the American Mathematical Society, Vol. 46 (1939), pp. 280-306.