* This research was supported by the National Science Foundation under Grant no. MPS75-07556.

MORE ON INCOMPLETE AND BOUNDEDLY COMPLETE FAMILIES OF DISTRIBUTIONS

by

Wassily Hoeffding

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1071

May, 1976

MORE ON INCOMPLETE AND BOUNDEDLY COMPLETE FAMILIES OF DISTRIBUTIONS*

By Wassily Hoeffding

University of North Carolina at Chapel Hill

Summary. Let $\mathcal{P}$ be a family of probability measures P which satisfy the conditions (*) $\int u_i \, dP = c_i$, $i = 1, \ldots, k$, and let $\mathcal{P}^{(n)}$ be the family of n-fold product measures P^n, $P \in \mathcal{P}$. The author has shown earlier that if $\mathcal{P}$ is sufficiently rich then (i) every symmetric unbiased estimator of zero (s.u.e.z.) for family $\mathcal{P}^{(n)}$ is of a specified form, and (ii) if every nontrivial linear combination of $u_1, \ldots, u_k$ is unbounded, no nontrivial bounded s.u.e.z. exists. In this paper extensions of these results to families of distributions symmetric about zero and to two-sample families are considered, and results by N. I. Fisher on families with conditions (*) replaced by the condition $\int u(x,y) dP^2 = c$ are discussed.

Key words: Family of distributions complete relative to a group of transformations. Boundedly complete family. Invariant unbiased estimator of zero.

* This research was supported by the National Science Foundation under Grant no. MPS75-07556.

1. Introduction. Let $\mathcal{Q}$ be a family of distributions (probability measures) on a measurable space $(\mathcal{Y},\mathcal{B})$ and let Γ be a group of $\mathcal{B}$-measurable transformations of $\mathcal{Y}$. The family $\mathcal{Q}$ is said to be complete relative to Γ if no nontrivial Γ-invariant unbiased estimator of zero for family $\mathcal{Q}$ exists. (A function is called Γ-invariant if it is invariant under all transformations in Γ.) The family $\mathcal{Q}$ is said to be boundedly complete relative to Γ if no bounded nontrivial Γ-invariant unbiased estimator of zero for family $\mathcal{Q}$ exists.

Let $\mathcal{P}$ be a family of distributions on a measurable space $(\mathcal{X},\mathcal{A})$ and let $\mathcal{P}^{(n)} = \{P^n : P\epsilon\mathcal{P}\}$ be the family of the n-fold product measures P^n on the measurable space $(\mathcal{X}^n, \mathcal{A}^{(n)})$ generated by $(\mathcal{X},\mathcal{A})$. The distributions P^n are invariant under the group Π_n of the $n!$ permutations of the coordinates of the points in $\mathcal{X}^n$. It is known that if the family $\mathcal{P}$ is sufficiently rich (for instance, contains all distributions concentrated on finite subsets of $\mathcal{X}$) then $\mathcal{P}^{(n)}$ is complete relative to Π_n (Halmos 1946; Fraser 1954; Lehmann 1959; Bell, Blackwell and Breiman 1960).

Now consider a family, again denoted by $\mathcal{P}$, of distributions P on $(\mathcal{X},\mathcal{A})$ which satisfy the conditions

$$\int u_i \, dP = c_i \, , \qquad i = 1, \ldots ,k, \tag{1.1}$$

where $u_1, \ldots ,u_k$ are given functions and $c_1, \ldots ,c_k$ are given constants. In this case the family $\mathcal{P}^{(n)}$ is, in general, not complete relative to Π_n. Indeed, the statistic

$$(1.2)\quad g(x_1,\dots,x_n) = \sum_{i=1}^{k}\sum_{j=1}^{n} \{u_i(x_j)-c_i\}h_i(x_1,\dots,x_{j-1},\ x_{j+1},\dots,x_n)$$

where $h_1, \dots ,h_k$ are arbitrary $A^{(n-1)}$-measurable, Π_{n-1}-invariant, $P^{(n-1)}$-integrable functions, is a Π_n-invariant estimator of zero. However, $P^{(n)}$ may be boundedly complete, as the following theorems, proved in [5], imply.

If A contains the one-point sets, let P_0 be the family of all distributions P concentrated on finite subsets of X which satisfy conditions (1.1). If μ is a σ-finite measure on (X,A), let $P_0(\mu)$ be the family of all distributions absolutely continuous with respect to μ whose densities $dP/d\mu$ are simple functions (finite linear combinations of indicator functions of sets in A) and which satisfy conditions (1.1).

Theorem 1A. If A contains the one-point sets and P is a convex family of distributions on (X,A) which satisfy conditions (1.1), such that $P_0 \subset P$, then every Π_n-invariant unbiased estimator $g(x_1,\dots,x_n)$ of zero is of the form (1.2).

Theorem 2A. If the conditions of Theorem 1A are satisfied and every nontrivial linear combination of $u_1, \dots ,u_k$ is unbounded then the family $P^{(n)}$ is boundedly complete relative to Π_n .

Theorems 1B and 2B of [5] assert that if P is a convex family of distributions absolutely continuous with respect to a σ-finite measure μ which satisfies conditions (1.1), and $P_0(\mu) \subset P$, then conclusions analogous to those of Theorems 1A and 2A hold, except that $g(x_1,\dots,x_n)$

is of the form (1.2) a.e. $(\mathcal{P}^{(n)})$ and, in the second theorem, every nontrivial linear combination $u(x) = \sum a_i u_i(x)$ is assumed to be $\mathcal{P}$-unbounded in the sense that for every real c there is a $P \in \mathcal{P}$ such that $P(|u(x)| > c) \neq 0$.

In this paper I shall deal with some extensions of these theorems. In Sections 2 and 3, straightforward extensions to some finite groups other than Π_n are briefly discussed. In Section 4 the generalization to the case where conditions (1.1) are replaced by $\iint u(x,y)dP(x)dP(y) = c$ is considered, which has been investigated by N. I. Fisher [2].

2. Distributions Symmetric About 0. Let $\mathcal{P}$ be a family of distributions which satisfy the conditions of one of the Theorems 1A, 1B, 2A, 2B with $\mathcal{X} = R^1$ and $\mathcal{A}$ the Borel sets, and the additional condition that each P in $\mathcal{P}$ is symmetric about zero. In this case the distributions P^n are invariant under the group Γ_n which consists of all permutations of the components of the points $(x_1,\ldots,x_n) \in R^n$ and of all changes of signs of the components. In conditions (1.1) each $u_i(x)$ may be replaced by $\{u_i(x) + u_i(-x)\}/2$. Thus we may assume that $u_i(x) \equiv u_i(-x)$ for all i. Then $g(x_1,\ldots,x_n)$ in (1.2), where each h_i satisfies the additional condition of being Γ_{n-1}-invariant, is a Γ_n-invariant estimator of zero.

The four theorems quoted in the Introduction are true if the distributions in $\mathcal{P}$, $\mathcal{P}_0$ and $\mathcal{P}_0(\mu)$ satisfy the additional condition of being symmetric about zero, if the $u_i(x)$ are symmetric about zero, and if Π_n is replaced by Γ_n throughout. The proof is very simple. If P

is the distribution of the random variable X, symmetric about zero, let P^* denote the distribution of $|X|$. Conditions (1.1) and the condition that $g(x_1,\ldots,x_n)$ is a Γ_n-invariant unbiased estimator of zero can be expressed in terms of the distributions P^*, and the problem is reduced to that of the theorems in the Introduction with X the set of the nonnegative numbers.

3. Two-Sample Families. For $r = 1,2$ let $\mathcal{P}_r$ be a family of distributions P on $(\mathcal{X},\mathcal{A})$ which satisfy the conditions

$$\int u_{r,i}\, dP = c_{r,i}\,, \qquad i = 1, \ldots ,k_r\,. \tag{3.1}$$

Let

$$\mathcal{P}^{(m,n)} = \{P_1^m P_2^n : P_1 \in \mathcal{P}_1,\ P_2 \in \mathcal{P}_2\}.$$

The distributions in $\mathcal{P}^{(m,n)}$ are invariant under the group $\Pi_m\Pi_n$ of those permutations of the coordinates of the points in $\mathcal{X}^{m+n}$ which permute the first m coordinates among themselves and permute the remaining n coordinates among themselves.

To simplify notation (and with no loss of generality) let conditions (3.1) be satisfied with $c_{r,i} = 0$:

$$\int u_{r,i}\, dP = 0, \qquad i = 1, \ldots ,k_r\,. \tag{3.2}$$

Then the statistic

$$g(x_1,\ldots,x_m, y_1,\ldots,y_n) = \tag{3.3}$$
$$\sum_{i=1}^{k_1}\sum_{j=1}^{m} u_{1i}(x_j)h_{1i}(x_1,\ldots,x_{j-1}, x_{j+1},\ldots,x_m, y_1,\ldots,y_n)$$
$$+ \sum_{i=1}^{k_2}\sum_{j=1}^{n} u_{2i}(y_j)h_{2i}(x_1,\ldots,x_m, y_1,\ldots,y_{j-1}, y_{j+1},\ldots,y_n),$$

where each h_{1i} is $\Pi_{m-1}\Pi_n$-invariant and $P^{(m-1,n)}$-integrable and each h_{2i} is $\Pi_m\Pi_{n-1}$-invariant and $P^{(m,n-1)}$-integrable, is a $\Pi_m\Pi_n$-invariant unbiased estimator of zero for family $P^{(m,n)}$. It can be shown by the methods of [5] that the obvious analogs of the four theorems in the introduction are true. For instance, if $P_{10}(P_{20})$ denotes the family of all distributions concentrated on finite subsets of X which satisfy conditions (3.2) with $r = 1$ $(r = 2)$, if A contains the one-point sets, and if $P_1(P_2)$ is a convex family of distributions which satisfy (3.2) with $r = 1$ $(r = 2)$, and such that $P_{r0} \subset P_r$ $(r = 1,2)$ then every $\Pi_m\Pi_n$-invariant estimator of zero is of the form (3.3). If, in addition, every nontrivial linear combination of $u_{r1}, \dots, u_{rk_r}$ is unbounded $(r = 1,2)$ then the family $P^{(m,n)}$ is boundedly complete relative to $\Pi_m\Pi_n$. Details of the proof are omitted.

4. Families Restricted by a Nonlinear Condition. It is natural to replace the conditions $\int u_i \, dP = c_i$ $(i = 1,\dots,k)$, which are linear in P, by one or more conditions of the form

$$\int_{X^s} u(x_1,\dots,x_s)dP^s = c.$$

Here I will consider only a family P of distributions P on (X,A) which satisfy the single condition

$$\int_{X^2} u(x_1,x_2)dP^2 = 0. \tag{4.1}$$

For example, if $X = R^1$ and

(4.2) $$u(x_1,x_2) = (x_1-x_2)^2 - 2\sigma^2 ,$$

condition (4.1) specifies the variance of the distribution P. For

(4.3) $$u(x_1,x_2) = \sum_{i=1}^{k} \{u_i(x_1) - c_i\}\{u_i(x_2) - c_i\}$$

condition (4.1) is equivalent to the conditions (1.1). Some other interesting special cases of (4.1) will be discussed later. We may and shall assume that

(4.4) $$u(x_1,x_2) \equiv u(x_2,x_1).$$

A Π_n-invariant unbiased estimator of zero for family $P^{(n)}$ $(n \geq 2)$ is

(4.5) $$g(x_1,\ldots,x_n) = u(x_1,x_2)h(x_3,\ldots,x_n) + u(x_1,x_3)h(x_2,x_4,\ldots,x_n)+\ldots+u(x_{n-1},x_n)h(x_1,\ldots,x_{n-2}),$$

where h is any Π_{n-2}-invariant, $P^{(n-2)}$-integrable function.

Even if P consists of all distributions P on (X,A) satisfying (4.1), a Π_n-invariant u.e.z., $g(x_1,\ldots,x_n)$, is not necessarily of the form (4.5). Whether it must be of this form depends on the function u. Thus if u is given by (4.3) then g is of the form (1.2), and can not (in general) be expressed in the form (4.5).

Families which satisfy condition (4.1) have been studied by N. I. Fisher in his Ph.D. dissertation [2]. One of Fisher's main results is as follows.

Let P be a convex family of distributions P on (X,A) satisfying (4.1), let A contain the one-point sets, and let $P \supset P_0$, where P_0 is the family of all distributions P concentrated on finite sets and satisfying (4.1). Define

$$L = \sum_{i=1}^{N-1} u_{iN} p_i \, , \qquad Q = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} (u_{iN} u_{jN} - u_{NN} u_{ij}) p_i p_j \, ,$$

where $u_{ij} = u(x_i, x_j)$.

Theorem 3. Suppose that for every $(x_1, \ldots, x_n) \in X^n$ there exist

$$N > n, \quad x_{n+1}, \ldots, x_N \text{ in } X, \quad p_1 > 0, \ldots, p_{N-1} > 0$$

such that

(a) $Q > 0$,

(b) $u_{NN} \neq 0$ and $u_{NN}(Q^{\frac{1}{2}} - L) > 0$,

(c) $Q^{\frac{1}{2}}$, considered as a function of $p_1, \ldots, p_{N-1}$, is irrational.

Then

(i) Every s.u.e.z. for $P^{(n)}$ is of the form (4.5).

(ii) If, in addition, $u(\ )$ is unbounded, then the family $P^{(n)}$ is boundedly complete relative to Π_n.

A similar result, analogous to Theorems 1B and 2B, holds for dominated families satisfying condition (4.1).

To give just one example, the conditions of Theorem 3 are satisfied in the case (4.2).

The condition $u_{NN} \neq 0$ implies $u(x,x) \not\equiv 0$. Fisher shows that the conclusion of Theorem 3 also holds if $u(x,x) \equiv 0$. The conclusion of Theorem 3 also holds in some other cases, for instance if $u(x,y)$ is of the form

$$u(x,y) = v_1(x)v_2(y) + v_1(y)v_2(x).$$

In this case condition (4.1) is equivalent to

$$\int v_1\, dP \cdot \int v_2\, dP = 0.$$

In the case (4.3) it is easy to see that $Q \le 0$, so that condition (a) is not satisfied.

To conclude we consider a special class of functions $u(x,y)$ not (in general) covered by the previously stated results. Let

$$u(x,y) = \sum_{i=1}^{\infty} c_i\, v_i(x) v_i(y), \tag{4.6}$$

where each $v_i(x)$ is bounded and the positive constants c_i are so chosen that $u(x,y)$ is bounded. In this case condition (4.1) is equivalent to the infinite set of conditions

$$\int v_i\, dP = 0, \qquad i = 1,2,\ldots \tag{4.7}$$

Here are a few special cases.

(a) Let $X = R^1$, let P_0 be a given distribution with distribution function $F_0(x)$, and let, I_A denoting the indicator function of set A,

$$v_i(x) = I_{(-\infty, r_i]}(x) - F_0(r_i),$$

where $\{r_i\}$ is a number sequence dense in R^1. Then the family P of all distributions satisfying condition (4.1) consists of the single distribution P_0 .

(b) If $X = R^1$,

$$v_i(x) = I_{(-\infty, r_i]}(x) - I_{(-\infty, r_i]}(-x),$$

then the family of all distributions P on the Borel sets which satisfy (4.1) is the family of all distributions symmetric about 0. In this case the family $P^{(n)}$ is complete relative to the group Γ_n of Section 2.

(c) The family of all bivariate distributions with given marginal distributions can also be characterized in this way with $\{v_i\}$ suitably chosen.

REFERENCES

[1] Bell, C.B., Blackwell, David, and Breiman, Leo (1960). On the completeness of order statistics. *Ann. Math. Statist. 31*, 794-797.

[2] Fisher, Nicholas I. (1976). The theory of unbiased estimation for some nonparametric families of distributions. Inst. of Statist. (Univ. of North Carolina) Mimeo Ser. No. 1051.

[3] Fraser, D.A.S. (1954). Completeness of order statistics. *Canad. J. Math. 6*, 42-45.

[4] Halmos, Paul R. (1946). The theory of unbiased estimation. *Ann. Math. Statist. 17*, 34-43.

[5] Hoeffding, Wassily (1976). Some incomplete and boundedly complete families of distributions. Submitted to *Ann. Statist.*

[6] Lehmann, E.L. (1959). *Testing Statistical Hypotheses*. New York: John Wiley.