The Annals of Statistics
1975, Vol. 3, No. 3, 754–758

CHARACTERIZATION OF DISTRIBUTIONS BY SETS OF MOMENTS OF ORDER STATISTICS

BY BARRY C. ARNOLD AND GLEN MEEDEN

Iowa State University

It is known that the set $M = \{\mu_{r,n} : r = 1, 2, \cdots, n; n = 1, 2, 3, \cdots\}$ of expectations of order statistics of samples from a distribution F completely determines F (see, for example, Hoeffding (1953)). In addition Chan (1967), Konheim (1971) and Pollak (1973) have shown that certain subsets of M determine F. In this note conditions are derived which are sufficient for a proper subset of M to determine F.

1. Introduction. Let F be a specified probability distribution on the real line. For each positive integer m and each positive integer $r \leqq m$, let $X_{r:m}$ denote the rth order statistic of a sample of size m from the distribution F. If F has a finite kth moment then $\mu_{r:m}^k = E(X_{r:m}^k)$ exists for every $r \leqq m$. A convenient reference for this fact and for other results quoted below is Chapter 3 of David (1970). Define

$$M_n{}^k = \{\mu_{r:m}^k : r = 1, 2, \cdots, m; m = 1, 2, \cdots, n\}$$

and

$$M^k = \bigcup_{n=1}^\infty M_n{}^k .$$

It may be verified, using the results of Hoeffding (1953), that for any odd k, M^k determines F. To see that the characterization fails for even k, it is sufficient to consider two distributions $F_1(x) = \delta_1(x)$ and $F_2(x) = \frac{1}{2}(\delta_{-1}(x) + \delta_1(x))$, where $\delta_a(x)$ is a distribution with a jump of 1 at the point a. It is readily verified that $\mu_{r:m}^k = 1$ for all r, m if k is even, for both F_1 and F_2.

Chan (1967) showed that, when $k = 1$, the sequence $\{\mu_{n:n}^1\}_{n=1}^\infty$ determines F. This was accomplished by showing that $\{\mu_{n:n}^1\}$ determines the Fourier transform of F^{-1} defined by $F^{-1}(u) = \inf\{x : F(x) \geqq u\}$). Subsequently an alternative proof that $\{\mu_{n:n}^1\}_{n=1}^\infty$ determines F was provided by Konheim (1971). Pollak (1973) obtained a more general result. He showed that, when $k = 1$, any sequence of the type $\{\mu_{r_n:n}^1 : n = 1, 2, 3, \cdots\}$ (where $r_n \leqq n$ for every n) determines M^1 and thus determines F. Pollak, incidentally, supplies an alternative proof that M^1 determines F to that based on Hoeffding's paper. In this note we will generalize Pollak's result. That is, we will determine conditions that if satisfied by a proper subset of M^k imply that the subset determines M^k and thus, if k is odd, determines F.

2. Triangular arrays. Consider Δ the infinite triangular array of points

Received April 1973; revised May 1974.
AMS 1970 *subject classifications.* Primary 62E10; Secondary 62 630.
Key words and phrases. Moments of order statistics, characterization.

shown in (1).

$$
\begin{array}{ccccccccc}
 & & & & X & & & & \\
 & & & X & & X & & & \\
 & & X & & X & & X & & \\
 & X & & X & & X & & X & \\
\cdot & & \cdot & & & & \cdot & & \cdot \\
\vdots & \cdot & & & & & & \cdot & \cdot \\
\cdot & \cdot & & & & & & \cdot & \cdot
\end{array} \tag{1}
$$

For each n, we denote by Δ_n the finite array consisting of the first n rows of Δ. A typical point in Δ will be denoted by (i, m) representing the ith point in the mth row. Let Γ be a configuration of points in Δ. We describe a method of adding points to Γ based on what we call the "triangle rule." Consider a triangular set of three contiguous points in Δ of the form

$$
\begin{array}{ccc}
 & X & \\
X & & X
\end{array} \tag{2}
$$

where the base must contain two of the points. The triangle rule specifies that if any two points of a triangular array of the form (2) belong to a configuration then the third point may be added to the configuration. Thus, given any configuration Γ there exists a smallest configuration which contains Γ and which is closed under the triangle rule. Denote this set by $T(\Gamma)$. Hence $T(\Gamma)$ is the intersection of all the configurations which contain Γ and are closed under the triangle rule. If $\emptyset$ denotes the empty set, then $T(\emptyset) = \emptyset$. If $\Gamma \neq \emptyset$, then $T(\Gamma)$ is a union of equilateral triangles (provided that we consider a single point to be an equilateral triangle). It is convenient to define the size of an equilateral triangular array to be the number of points in any side.

A configuration Γ is said to be connected if $T(\Gamma)$ consists of a single triangle.

For each $a \in \Gamma$, a subset of Δ, define $\Gamma_a = \Gamma - \{a\}$. Γ will be said to be independent if, for each $a \in \Gamma$, $a \notin T(\Gamma_a)$. Note that any subset of an independent configuration is independent.

With these definitions we may prove the following lemma.

LEMMA 1. *Let Γ be an independent connected set containing n points; then $T(\Gamma)$ is an equilateral triangle of size n.*

PROOF. The lemma follows easily by induction utilizing the following two observations.

First, if Γ is an independent configuration of size $n > 2$ then there exist nonempty disjoint subsets of Γ, say Γ_1 and Γ_2, such that $T(\Gamma_1) \cap T(\Gamma_2) = \emptyset$ and $T(T(\Gamma_1) \cup T(\Gamma_2)) = T(\Gamma)$. To see that this is so, consider Γ as a collection of n disjoint triangles of size 1. Since Γ is connected, two of these triangles must be contiguous and can be used to generate a new triangle of size 2 by the triangle rule. There remain $n - 1$ disjoint triangles (since the original array was independent). The process of reduction can be continued until we are left with two disjoint triangles, say T_1 and T_2 whose union generates $T(\Gamma)$. For each i, let Γ_i be the subset of Γ which generated T_i.

Secondly, let T_m and T_n be triangles of size m and n respectively. If T_m and T_n are disjoint and if $T_m \cup T_n$ is connected then the size of $T(T_m \cup T_n)$ is $m + n$.

The statement of the lemma is trivially true for $n = 1$ and, if it is assumed true for all $m < n$, the two observations imply it must be true for n.

It is easy to give examples of independent connected sets in Δ. A simple example which generates Δ_n is an array consisting of exactly one point from each of the first n rows of Δ. A second example, generating Δ_n, is the nth row of Δ.

In the next lemma we give a necessary and sufficient condition on the set Γ so that $T(\Gamma) = \Delta$.

LEMMA 2. *For a configuration Γ of Δ, $T(\Gamma) = \Delta$ if and only if $\Gamma \cap \Delta_n$ has an independent connected subset of size n for infinitely many n's.*

PROOF. If $\Gamma \cap \Delta_n$ has an independent connected subset of size n then, by Lemma 1, $T(\Gamma \cap \Delta_n) = \Delta_n$. Thus $\Delta_n \subset T(\Gamma)$ for infinitely many n's and consequently $\Delta = T(\Gamma)$.

To prove the converse, we first observe that for any configuration Γ $T(\Gamma) = \bigcup_{n=1}^{\infty} T(\Gamma \cap \Delta_n)$. Clearly $T(\Gamma) \supset \bigcup_{n=1}^{\infty} T(\Gamma \cap \Delta_n)$. On the other hand, $T(\Gamma) \subset \bigcup_n T(\Gamma \cap \Delta_n)$, since $\bigcup_n T(\Gamma \cap \Delta_n)$ is a set containing Γ which is closed under the triangle rule.

Suppose $T(\Gamma) = \Delta$. Then $\Delta = \bigcup_n T(\Gamma \cap \Delta_n)$, and to complete the proof it suffices to show that for any positive integer k there exists an $n' \geqq k$ such that $T(\Gamma \cap \Delta_{n'}) = \Delta_{n'}$. Clearly there exists an $n(k)$ such that $n(k) \geqq k$ and $T(\Gamma \cap \Delta_{n(k)}) \supset \Delta_k$. Since $T(\Gamma \cap \Delta_{n(k)})$ is closed under the triangle rule all or none of the elements in the $(k + 1)$st row of Δ belong to this set. If the $(k + 1)$st row is present the next row is either completely present or completely absent, and so on. Let n' be the row which immediately precedes the first empty row of $T(\Gamma \cap \Delta_{n(k)})$. If no such empty row exists let $n' = n(k)$. Now $T(\Gamma \cap \Delta_{n'}) = \Delta_{n'}$ and the proof is complete.

3. Characterizations by moments. We now return to the original problem of finding subsets of M^k (i.e. sets of kth moments of order statistics) which determine the underlying distribution F. Observe that the elements of M^k may be indexed by Δ, and the elements of M_n^k indexed by Δ_n.

The kth moments of order statistics of samples from an arbitrary distribution are known to satisfy

$$(m - r)\mu_{r:m}^k + \mu_{r+1:m}^k = m\mu_{r:m-1}^k . \tag{3}$$

For our purposes, the important fact expressed in (3) is that within certain triples of expected values of order statistics, knowledge of any two determines the third. There are many other known recursive relations among the $\mu^k_{r:m}$'s but, in essence, they follow by repeated application of (3).

If Λ is a subset of M^k indexed by Γ, i.e. $\Lambda = \{\mu^k_{r:m} : (r, m) \in \Gamma\}$, then from equation (3) we may deduce that knowledge of the elements of Λ implies knowledge of $\mu^k_{r':m'}$ for any (r', m') which can be obtained from Γ by repeated application of the triangle rule. Thus $\Lambda = \{\mu^k_{r:m} : (r, m) \in \Gamma\}$ determines $\{\mu^k_{r:m} : (r, m) \in T(\Gamma)\}$. Consequently a subset Λ of M^k indexed by Γ will determine M^k and hence, for odd k, determine the distribution function F, provided that $T(\Gamma) = \Delta$. Combining these observations with Lemma 2, we may state the following theorem.

THEOREM. *Let Λ be a subset of M^k indexed by Γ. Λ determines M^k (and thus, for odd k, determines F) if for infinitely many n's, $\Gamma \cap \Delta_n$ has an independent connected subset of size n.*

Thus, a configuration which includes one entry in each row of M^k will determine M^k (as mentioned earlier, Pollak (1973) obtained this result for $k = 1$). Another configuration adequate to determine M^k is one which includes infinitely many rows of M^k. This configuration includes more points than does a "one entry per row" configuration, but it does point out the fact that many rows can be unrepresented in a determining configuration.

Let Λ be a subset on M^k indexed by Γ. It is possible that Λ determines a unique F even though $T(\Gamma) \neq \Delta$. For example, let $\Lambda_s{}^k = \{\mu^k_{r,m} : s < r \leqq m - s$ and $m = 2s + 1, 2s + 2, \cdots\}$ where s is a positive integer. A minor modification of Hoeffding's argument will verify that such a configuration, for k odd, determines F. When F is the Cauchy distribution, $\Lambda_1{}^1$ exists and determines F while M^1 is not even well defined.

It may be observed that equation (3) remains valid under the assumption that $\mu^k_{r:m}$ is the kth moment of the rth largest of the first m coordinates of an exchangeable sequence $X_1, X_2, \cdots$. The array M^k will still be determined by a subset Λ if the conditions of the theorem are met. Unfortunately the array M^k (even for k odd) does not necessarily characterize the common distribution function of the coordinates of an exchangeable sequence. To see this, consider M^1 for a trivial exchangeable sequence $X_1 = X_2 = X_3 = \cdots$ where X_1 is normal $(0, \sigma^2)$. For any choice of σ^2 all the entries in M^1 will be zero.

In closing we consider a related problem. Suppose we are given a sequence of real numbers $\{a_{r(i),n(i)}\}$, where for each i $r(i)$ and $n(i)$ are positive integers such that $r(i) \leqq n(i)$. We are interested in determining whether or not there exists a distribution F such that

$$E(X^k_{r(i):n(i)}) = a_{r(i):n(i)} \qquad \text{for} \quad i = 1, 2, \cdots .$$

The sequence $\{a_{r(i),n(i)}\}$ defines a subset, say Γ, of Δ in (1), in a natural way. If $T(\Gamma) = \Delta$, then $\{a_{r(i):n(i)}\ i = 1, 2, \cdots\}$ generates a set $S = \{a_{r,n} : 1 \leqq r \leqq n, n = 1, 2, \cdots\}$. The problem then becomes that of determining when a distribution function exists such that

$$E(X^k_{r,n}) = a_{r,n}, \quad \forall\, r \leqq n .$$

Kadane (1971) gives a condition on S which is necessary and sufficient that there exist a distribution F, satisfying $F(0_-) = 0$, with kth moments of its order statistics given by S. Mallows (1973) solves the problem when the support of F is the real numbers. Both Kadane and Mallows treat the case $k = 1$ but their results readily extend to the case of arbitrary k.

REFERENCES

[1] CHAN, L. K. (1967). On a characterization of distribution by expected values of extreme order statistics. *Amer. Math. Monthly* **74** 950-951.

[2] DAVID, HERBERT A. (1970). *Order Statistics*. Wiley, New York.

[3] HOEFFDING, WASSILY (1953). On the distribution of the expected values of the order statistics. *Ann. Math. Statist.* **24** 93-100.

[4] KADANE, JOSEPH B. (1971). A moment problem for order statistics. *Ann. Math. Statist.* **42** 745-751.

[5] KONHEIM, A. G. (1971). A note on order statistics. *Amer. Math. Monthly* **78** 524.

[6] MALLOWS, C. L. (1973). Bounds on distributions functions in terms of expectations of order-statistics. *Ann. Probability* **1** 297-303.

[7] POLLAK, MOSHE (1973). On equal distribution. *Ann. Statist.* **1** 180-182.

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50010