COMPLETENESS OF ORDER STATISTICS

D. A. S. FRASER

1. Introduction. Under the non-parametric assumption that a set of observations is a sample from an absolutely continuous distribution, the order statistics are known to form a complete sufficient statistic. It is proved in this note that it suffices to have the class of uniform distributions over finite numbers of intervals *or* the class of uniform distributions over sets of a ring which is a basis for the σ-algebra of Borel sets. This result is derived as a particular case of that of several samples from more general distributions.

2. Formulation and statement of results. Let $\mathfrak{X}$, $\mathfrak{A}$ stand for a measurable space: $\mathfrak{X}$ is an arbitrary space and $\mathfrak{A}$ is a σ-algebra of subsets $A \subset \mathfrak{X}$. A class $\mathfrak{B} = \{B\}$ of subsets of $\mathfrak{X}$ will be called a basis of $\mathfrak{A}$ and written $\mathfrak{B} = \mathscr{B}(\mathfrak{A})$ if $\mathfrak{B}$ is a ring and if $\mathfrak{A}$ is the σ-ring generated by $\mathfrak{B}$.

Distributions over the space $\mathfrak{X}$ will be given in terms of a finite measure μ over $\mathfrak{A}$ by means of a density function $f(x)$; that is,

$$P_\eta(A) = \int_A f_\eta(x)\, d\mu(x),$$

and we are thus restricted to distributions which are absolutely continuous with respect to $\mu(x)$.

To describe a sample of n from a distribution P_η, we envisage the product space $\mathfrak{X}^n$ with the σ-algebra generated by $\mathfrak{A}^n$ and the power product measure μ^n. Then, for C a measurable subset of $\mathfrak{X}^n$,

$$P_\eta^n(C) = \int_C \prod_1^n f_\eta(x_i) \prod_1^n d\mu(x_i) \tag{2.1}$$

and the distributions are given by $\{\Pi f_\eta(x_i)\}$.

A statistic based on a sample of n is a measurable function $g(x_1, \ldots, x_n)$. A statistic of interest will be $O(x_1, \ldots, x_n) = \{x_1, \ldots, x_n\}$, that is, the set of x's without regard for the order in which the x's occur in the function $O(x_1, \ldots, x_n)$. Thus $O(x_1, \ldots, x_n)$ is invariant under the group of permutations of the x's and is in fact the maximal invariant function under such transformations. Any statistic $g(x_1, \ldots, x_n)$ which can be written as a function of $O(x_1, \ldots, x_n)$ is a symmetric function of the x's, and conversely.

We define a complete class of measures. *$\{\nu_\eta(C) | \eta \in \Omega\}$ is a complete class if*

$$\int_y g(y)\, d\nu_\eta(y) \underset{\eta}{\equiv} 0 \tag{2.2}$$

Received May 4, 1953.

implies that $g(y) = 0$ *almost everywhere with respect to the measures* ν_η. In this note we prove that the class of distributions of $O(x_1, \ldots, x_n)$ induced by the distributions $\{P_\eta^n | \eta \in \Omega\}$ is complete subject to conditions on Ω.

3. Derivation of results.

THEOREM. *The distributions of* $O(x_1, \ldots, x_n)$ *induced by the distributions* $\{P_\eta^n | \eta \in \Omega\}$ *over* $\mathfrak{X}^n$ *form a complete class if* $\mu(x)$ *is non-atomic and if* $\{f_\eta^n | \eta \in \Omega\}$ *consists of uniform distributions over the sets* B *of a basis* $\mathscr{B}(\mathfrak{A})$.

Proof. The density function $f_\eta(x)$ of a uniform distribution will have the simple form

$$f_\eta(x) = c(\eta)\, \phi_\eta(x), \tag{3.1}$$

where $\phi_\eta(x)$ is the characteristic function of a set $B_\eta \in \mathscr{B}(\mathfrak{A})$ and $c(\eta)$ is a normalizing constant such that

$$1 = c(\eta) \int_{B_\eta} d\mu(x) = c(\eta)\, \mu(B_\eta).$$

To show that the distributions of $O(x_1, \ldots, x_n)$ form a complete class, we show that any measurable function of $O(x_1, \ldots, x_n)$ satisfying (2.2) is zero almost everywhere μ^n. However, such a function is necessarily a symmetric function $h(x_1, \ldots, x_n)$ of the x's and hence (2.2) gives

$$\int_{\mathfrak{X}^n} h(x_1, \ldots, x_n) \prod_1^n f_\eta(x_i) \prod_1^n d\mu(x_i) = 0. \tag{3.2}$$

Since $c(\eta) \neq 0$, we have

$$\int_{B^n} h(x_1, \ldots, x_n) \prod_1^n d\mu(x_i) = 0. \tag{3.3}$$

Let $B_1, \ldots, B_n$ be any n disjoint sets belonging to $\mathscr{B}(\mathfrak{A})$. Since $\mathscr{B}(\mathfrak{A})$ is a ring, any sum of B's will belong to $\mathscr{B}(\mathfrak{A})$ and therefore

$$\int_{(B_{i_1} \cup \ldots \cup B_{i_r})^n} h(x_1, \ldots, x_n) \prod_1^n d\mu(x_i) = 0. \tag{3.4}$$

If we define

$$I(i_1, \ldots, i_n) = \int_{B_{i_1} \times \ldots \times B_{i_n}} h(x_1, \ldots, x_n) \prod_1^n d\mu(x_i),$$

then the symmetry of h implies that I is symmetric. From (3.4) with $r = 1$ we obtain

$$I(1, \ldots, 1) = 0, \;\; I(2, \ldots, 2) = 0;$$

and with $r = 2$,

$$\sum_{j_1=1}^{2} \ldots \sum_{j_n=1}^{2} I(j_1, \ldots, j_n) = 0.$$

Then by subtraction,

$$\sum_{\alpha(1,2)} I(j_1, \ldots, j_n) = 0$$

where $\alpha(1,2)$ is all n-tuples $(j_1, \ldots, j_n)$ containing both and only 1's and 2's.

Proceeding inductively, we obtain finally

$$\sum_{\alpha(1,\ldots,n)} I(j_1, \ldots, j_n) = 0$$

where $\alpha(1, \ldots, n)$ is the set of permutations of $(1, \ldots, n)$. From the symmetry of I it follows that

$$I(1, \ldots, n) = 0,$$

that is,

$$\int_{B_1 \times \ldots \times B_n} h(x_1, \ldots, x_n) \prod_1^n d\mu(x_i) = 0. \tag{3.4}$$

Since μ is non-atomic it follows from Halmos **(1)** that $\mathfrak{X}$ can be divided into a finite number of disjoint sets $\{S_i^\epsilon\}$ each of which has μ-measure less than or equal to ϵ. Consequently the diagonal space of $\mathfrak{X}^n$,

$$D(\mathfrak{X}^n) = \{(x_1, \ldots, x_n) | \exists i, j \ni x_i = x_j, i \neq j\}$$

can for any positive δ be enclosed in a set $E_\delta(\mathfrak{X}^n)$ having μ^n-measure less than or equal to δ. For example,

$$E_\delta(\mathfrak{X}^n) = \bigcup_{i,P} \{S_i^\epsilon \times S_i^\epsilon \times \mathfrak{X}^{n-2}\},$$

where ϵ is chosen sufficiently small $(= \epsilon(\delta))$ and where the union is taken over i and over the image sets under permutations P of the coordinates.

Since the sets $\mathscr{B}(\mathfrak{A})$ form a basis for $\mathfrak{A}$, the sets $\{B_1 \times \ldots \times B_n\}$ with disjoint B's belonging to $\mathscr{B}(\mathfrak{A})$ form a basis for the σ-algebra

$$\mathfrak{A}_\delta^n = \{C \cap (\mathfrak{X}^n - E_\delta(\mathfrak{X}^n)) | C \in \mathfrak{A}^n\}.$$

By extending the signed measure (3.4), we obtain

$$\int_{C^*} h(x_1, \ldots, x_n) \prod_1^n d\mu(x_i) = 0,$$

where $C^* \in \mathfrak{A}_\delta^n$. The Radon-Nikodym theorem then establishes that $h = 0$ almost everywhere μ^n in $\mathfrak{X}^n - E_\delta(\mathfrak{X}^n)$. Since δ can be arbitrarily small, the theorem follows.

Corollary. *The distributions of* $O(x_1, \ldots, x_n), \ldots, O(y_1, \ldots, y_m)$ *induced by the distributions* $\{P_\eta^n \times \ldots \times P_\xi^m | (\eta, \ldots, \xi) \in \Omega \times \ldots \times \Xi\}$ *over* $\mathfrak{X}^n \times \ldots \times \mathscr{Y}^n$ *form a complete class if* $P_\eta, \ldots, P_\xi$ *satisfy the conditions in the* Theorem.

The proof is a straight extension of that for the Theorem.

Example. Consider the space $\mathfrak{X} = \{(z, y)\} =]-\infty, 0[\times]0, \infty[$ and take as a basis for the σ-algebra of Borel sets the class of sets consisting of finite

https://doi.org/10.4153/CJM-1954-006-0 Published online by Cambridge University Press

unions of rectangles. If we take as probability measures the uniform distributions over sets of the basis above, then for a sample of n the theorem gives the completeness of the probability measures of the statistic

$$\{(z_1, y_1), \ldots, (z_n, y_n)\}.$$

Reference

1. P. Halmos, *The range of a vector measure*, Bull. Amer. Math. Soc., *54* (1948), 416–421.

University of Toronto

https://doi.org/10.4153/CJM-1954-006-0 Published online by Cambridge University Press