The Annals of Statistics
1973, Vol. 1, No. 1, 180-182

ON EQUAL DISTRIBUTIONS

BY MOSHE POLLAK

The Hebrew University of Jerusalem

It is shown that two distributions both of which have a finite expectation are equal if and only if for every $n \geqq 1$ there exists $1 \leqq k \leqq n$ such that the kth order statistics from samples of size n of each distribution have equal expectations.

Similarly, it is shown that a distribution with finite expectation is symmetric about zero if and only if for every $n \geqq 0$ there exists $0 \leqq k \leqq 2n+1$ such that the sum of the expectations of the kth smallest and the kth largest observations in a sample of size $2n+1$ is zero.

The object of this paper is to give a characterization of distributions of random variables whose expectations exist and are finite.

In this paper, $(X_{(1)}^n, \cdots, X_{(n)}^n)$ will denote the order statistic (in increasing order) of n independent observations of the random variable X.

THEOREM 1. *Let X, Y be random variables whose expectations exist and are finite, and let F, G denote their respective distributions. A necessary and sufficient condition for F to equal G is that for every positive integer n there exist a k_n, $1 \leqq k_n \leqq n$, such that*

$$EX_{(k_n)}^n = EY_{(k_n)}^n . \tag{1}$$

PROOF. The necessity of the condition is clear. The sufficiency will be proved by the following two lemmas.

LEMMA 1. *Let X, Y be defined as in Theorem 1. If for every positive integer n there exists a k_n, $1 \leqq k_n \leqq n$, such that* (1) *holds, then for every n and every $1 \leqq i \leqq n$.*

$$EX_{(i)}^n = EY_{(i)}^n . \tag{2}$$

PROOF. By induction on n and i. (2) clearly holds for $n = 1, i = 1$. Suppose that (2) holds for all $1 \leqq i \leqq n$, $n \leqq r$. It suffices to prove that $EX_{(i)}^{r+1} = EY_{(i)}^{r+1}$ for all $1 \leqq i \leqq r+1$. By hypothesis, $EX_{(k_{r+1})}^{r+1} = EY_{(k_{r+1})}^{r+1}$.

Let

$$\begin{aligned}
A_j^n &= \text{the set of all subsets of } \{1, \cdots, n\} \text{ which} \\
&\quad\ \text{contain exactly } j \text{ different elements} \\
(S_{(1)}^\sigma, \cdots, S_{(j)}^\sigma) &= \text{the order statistic (in increasing order) of} \\
&\quad\ \{X_{(i)}^n \mid i \in \sigma,\ \sigma \in A_j^n\} \\
(T_{(1)}^\sigma, \cdots, T_{(j)}^\sigma) &= \text{the order statistic (in increasing order) of} \\
&\quad\ \{Y_{(i)}^n \mid i \in \sigma,\ \sigma \in A_j^n\} \\
V_i^{r+1} &= \textstyle\sum_{\sigma \in A_r{}^{r+1}} S_{(i)}^\sigma \\
W_i^{r+1} &= \textstyle\sum_{\sigma \in A_r{}^{r+1}} T_{(i)}^\sigma
\end{aligned}$$

Received June 3, 1971; revised April 25, 1972.

By the induction hypothesis, $ES^{\sigma}_{(i)} = ET^{\sigma}_{(i)}$ for each $\sigma \in A_r{}^{r+1}$, $1 \leqq i \leqq r$; and so for every $1 \leqq i \leqq r$:

$$EV_i{}^{r+1} = EW_i{}^{r+1}\,. \tag{3}$$

Clearly, for $1 \leqq i \leqq r+1$

$$V_i{}^{r+1} = (r+1-i)X^{r+1}_{(i)} + iX^{r+1}_{(i+1)} \tag{4}$$

$$W_i{}^{r+1} = (r+1-i)Y^{r+1}_{(i)} + iY^{r+1}_{(i+1)}\,. \tag{5}$$

If $1 \leqq k_{r+1} < r+1$, then taking $i = k_{r+1}$ in (3), (4) and (5), one gets that

$$\begin{aligned} iE(X^{r+1}_{(k_{r+1}+1)} - Y^{r+1}_{(k_{r+1}+1)}) &= E(V^{r+1}_{k_{r+1}} - W^{r+1}_{k_{r+1}}) - (r+1-i)E(X^{r+1}_{(k_{r+1})} - Y^{r+1}_{(k_{r+1})}) \\ &= 0 \end{aligned}$$

and so $EX^{r+1}_{(j)} = EY^{r+1}_{(j)}$ for $j = k_{r+1}+1$ and inductively for $j = k_{r+1}+2, \cdots, r+1$.

For $1 < k_{r+1} \leqq r+1$, taking $i = k_{r+1} - 1$ in (3), (4) and (5) in the same way gives $EX^{r+1}_{(j)} = EY^{r+1}_{(j)}$ for $j = k_{r+1} - 1$, and inductively for $j = k_{r+1} - 2, \cdots, 1$.

Therefore (2) holds for $n \leqq r+1$, $1 \leqq i \leqq n$, and by induction for all n, $1 \leqq i \leqq n$, proving Lemma 1.

LEMMA 2. *Let X, Y, F, G be defined as in Theorem* 1. *Let* (2) *hold for every n and every i, $1 \leqq i \leqq n$. Then $F = G$.*

PROOF. Let $0 < \alpha < 1$ be such that there are unique x_α and y_α with $F(x_\alpha -) \leqq \alpha \leqq F(x_\alpha)$ and $G(y_\alpha -) \leqq \alpha \leqq G(y_\alpha)$. To show $F = G$ it is enough to show $x_\alpha = y_\alpha$ for such pairs. Select integers ρ_n such that $\rho_n/n \to \alpha$. Then if F_n is the empirical distribution function based on n observations of X, $F_n(X^n_{(\rho_n)}) = \rho_n/n \to \alpha$ (if there are ties, F_n may jump above ρ_n/n at $X^n_{(\rho_n)}$ but we give the equation $F_n(X^n_{(\rho_n)}) = \rho_n/n$ the obvious interpretation) so that $X^n_{(\rho_n)} \to_{\text{a.s.}} x_\alpha$. Similarly, $Y^n_{(\rho_n)} \to_{\text{a.s.}} y_\alpha$. It is easy to see that $EX^n_{(\rho_n)} \to x_\alpha$ and $EY^n_{(\rho_n)} \to y_\alpha$. But since $EX^n_{(\rho_n)} = EY^n_{(\rho_n)}$ (by hypothesis), this implies that $x_\alpha = y_\alpha$.

This completes the proof of Lemma 2 and thus the proof of Theorem 1.

REMARK. In particular, $F = G$ if $E \max_{i=1,\cdots,K} X_i = E \max_{i=1,\cdots,K} Y_i$ for all K (or for $K \geqq n_0$, as can be shown by similar methods; this would be enough also if $EX = -\infty$ but $E \max_{i=1,\cdots,K} X_i$ is finite for $K \geqq n_0$, n_0 arbitrary).

THEOREM 2. *Let X be a random variable whose expectation exists. Then the distribution F of X is symmetric about zero if and only if for each nonnegative integer n there exists a k_n, $1 \leqq k_n \leqq 2n+1$, such that $E(X^{2n+1}_{(k_n)} + X^{2n+1}_{(2n+2-k_n)}) = 0$.*

PROOF. The necessity of the condition is clear. The sufficiency: Denote $Y = -X$. As in the proof of Theorem 1, one first proves (by induction on n) that

$$EX^n_{(i)} = EY^n_{(i)} = -EX^n_{(n+1-i)} \tag{6}$$

for every i, n; $1 \leqq i \leqq n$. (6) clearly holds for $n = 1$, $i = 1$. Suppose (6) holds for all i, n where $1 \leqq i \leqq n$, $1 \leqq n \leqq 2m+1$. In particular, $EX^{2m+1}_{(m+1)} = -EX^{2m+1}_{(m+1)} = 0$, and so $EV^{2m+2}_{m+1} = 0$ (where $V_i{}^{r+1}$ is as denoted in the proof of

Lemma 1). However $V^{2m+2}_{m+1} = (m+1)X^{2m+2}_{(m+1)} + (m+1)X^{2m+2}_{(m+2)}$. Thus (6) holds also for $n = 2m+2$, $i = m+1$. By reasoning analogous to the proof of Lemma 1, one gets that (6) holds for $n = 2m+2$ and all $i = 1, \cdots, 2m+2$, and likewise, also for $n = 2m+3$ and all $i = 1, \cdots, 2m+3$, and by induction for every i, n, $n > 0$, $1 \leqq i \leqq n$. This shows the conditions of Lemma 2 to hold, and so X and $-X$ have the same distribution, completing the proof of Theorem 2.

REMARK. In particular, F is symmetric about zero if and only if

$$E(\max_{i=1,\cdots,2n+1} X_i + \min_{i=1,\cdots,2n+1} X_i) = 0$$

and likewise if and only if $E \operatorname{median}_{i=1,\cdots,2n+1} X_i = 0$ for all nonnegative integers n, where $X_1, \cdots, X_{2n+1}$ are independent observations of X. Using similar methods, one can prove that if the expectation of X does not exist, but the expectations of the medians of large enough samples do and are equal to zero, then the distribution of X is symmetric (about zero).

Acknowledgments. Yosef Rinott helped to bring Theorem 1 to its present form. The referee's suggestions shortened the proof of Lemma 2 and helped bring Theorem 2 to its present form.

DEPARTMENT OF STATISTICS
THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM, ISRAEL