The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 2137–2139

A NOTE ON ORDER STATISTICS FOR HETEROGENEOUS DISTRIBUTIONS[1]

BY PRANAB KUMAR SEN

University of North Carolina

0. Summary. For non-identically distributed random variables, some inequalities on the median and the tails of the distribution of a sample order statistic are derived, and a simple condition for the existence of its moments is studied.

1. Statement of the problems. Let $X_1, \cdots, X_n$ be independent random variables with continuous (cumulative) distribution functions (df) $F_1(x), \cdots F_n(x)$, respectively. We denote the ordered values of the X_i by

$$X_{n,1} \leqq \cdots \leqq X_{n,n}. \tag{1.1}$$

Let then $\mathbf{F}_n = (F_1, \cdots, F_n)$ and $\bar{F}_n = n^{-1}\sum_{i=1}^n F_i$. We assume that $\xi_{n,r}$ is a unique solution of

$$\bar{F}_n(\xi_{n,r}) = r/n, \qquad \text{for} \quad 1 \leqq r \leqq n-1,\ \xi_{n,0} = -\infty,\ \xi_{n,n} = \infty; \tag{1.2}$$

that is, $\xi_{n,r}$ is the rth quantile of $\bar{F}_n$, $r = 1, \cdots, n-1$. Also, let

$$P_r(x; \mathbf{F}_n) = P\{X_{n,r} \leqq x; \mathbf{F}_n\} \quad \text{and} \tag{1.3}$$

$$P_r^*(x; \bar{F}_n) = P\{X_{n,r} \leqq x; F_1 = \cdots = F_n = \bar{F}_n\};$$

$$P_r(\zeta_{n,r}; \mathbf{F}_n) = P_r^*(\zeta_{n,r}^*; \bar{F}_n) = \tfrac{1}{2}, \qquad r = 1, \cdots, n. \tag{1.4}$$

Thus, $\zeta_{n,r}$ and $\zeta_{n,r}^*$ are the medians of P_r and P_r^*, $r = 1, \cdots, n$. Finally, let us define (for $\delta \geqq 0$)

$$v_\delta(F_i) = \int_{-\infty}^{\infty} |x|^\delta \, dF_i(x), \qquad i = 1, \cdots, n, \tag{1.5}$$

so that $v_\delta(\bar{F}_n) = n^{-1}\sum_{i=1}^n v_\delta(F_i)$, whenever (1.5) exists. Then, the main theorems of the note are the following.

THEOREM 1.1. *For* $2 \leqq r \leqq n-1$, *for all* $x \leqq \xi_{n,r-1} < \xi_{n,r} \leqq y$,

$$P_r(y; \mathbf{F}_n) - P_r(x; \mathbf{F}_n) \geqq P_r^*(y; \bar{F}_n) - P_r^*(x; \bar{F}_n), \tag{1.6}$$

where the equality sign holds only if $F_1 = \cdots = F_n = F$ *at both* x *and* y, *and for the two extremes*

$$P_1(x; \mathbf{F}_n) \geqq P_1^*(x; \bar{F}_n) \quad \text{and} \quad P_n(x; \mathbf{F}_n) \leqq P_n^*(x; \bar{F}_n), \qquad \text{for all } x, \tag{1.7}$$

with strict inequalities unless $F_1 = \cdots = F_n = F$ *at* x.

THEOREM 1.2. $\xi_{n,r-1} \leqq \zeta_{n,r}, \zeta_{n,r}^* \leqq \xi_{n,r}$ *for all* $2 \leqq r \leqq n-1$, *and hence* $|\zeta_{n,r} - \zeta_{n,r}^*| \leqq [\xi_{n,r} - \xi_{n,r-1}]$, $2 \leqq r \leqq n-1$.

Received April 14, 1969.

[1] Supported by the Army Research Office, Durham, Grant DA-ARO-D-31-124-G746.

THEOREM 1.3. *$v_\delta(\bar{F}_n) < \infty$ (for some $\delta > 0$) entails that for all $r_0 \leqq r \leqq n-r_0+1$, $E|X_{n,r}|^k < \infty$, where $r_0 = [k/\delta]$, $[s]$ being the integral part of s.*

Theorem 1.1 and Theorem 1.2 are based on the results by Hoeffding (1956) and their generalizations by Samuels (1965) and by Anderson and Samuels (1966). Theorem 1.3 extends the results of Blom (1958) and Sen (1959) to non-identical df's.

2. The proofs of the theorems. (i) *Theorem* 1.1. By definition,

$$1-P_r(x;\mathbf{F}_n) = P\{X_{n,r} > x \mid \mathbf{F}_n\} = P\{r-1 \text{ or less of the } X_i \leqq x \mid \mathbf{F}_n\}$$
$$= \sum_{j=0}^{r-1}\sum_{S_j}\prod_{l=1}^{j} F_{i_l}(x)\prod_{l=j+1}^{n}[1-F_{i_l}(x)] = Q_n(x;\mathbf{F}_n), \text{ say}, \tag{2.1}$$

where the summation S_j extends over all possible $(i_1, \cdots, i_n)$, which are permutations of $1, \cdots, n$. Similarly,

$$1-P_r^*(x;\bar{F}_n) = \sum_{j=0}^{r-1}\binom{n}{j}[\bar{F}_n(x)]^j[1-\bar{F}_n(x)]^{n-j} = Q_n^*(x;\bar{F}_n), \text{ say}. \tag{2.2}$$

Now, it follows from the results of Hoeffding (1956) that

$$\begin{aligned} Q_n(x;\mathbf{F}_n) &\geqq Q_n^*(x;\bar{F}_n), \quad \text{if } n\bar{F}_n(x) \leqq r-1,\\ &\leqq Q_n^*(x;\bar{F}_n), \quad \text{if } n\bar{F}_n(x) \geqq r, \end{aligned} \tag{2.3}$$

where the equality signs hold only when $F_1(x) = \cdots = F_n(x)$. Also, for $x \leqq \xi_{n,r-1}$,

$$n\bar{F}_n(x) = \sum_{i=1}^{n} F_i(x) \leqq \sum_{i=1}^{n} F_i(\xi_{n,r-1}) = n\bar{F}_n(\xi_{n,r-1}) = r-1, \tag{2.4}$$

and for $x \geqq \xi_{n,r}$

$$n\bar{F}_n(x) \geqq n\bar{F}_n(\xi_{n,r}) = r. \tag{2.5}$$

Hence, (1.6) readily follows from (2.1) through (2.5). Further,

$$1-P_1(x;\mathbf{F}_n) = P\{\text{all the } X_i \geqq x\} = \prod_{i=1}^{n}[1-F_i(x)]$$
$$\leqq \{n^{-1}\sum_{i=1}^{n}[1-F_i(x)]\}^n = [1-\bar{F}_n(x)]^n = 1-P_1^*(x;\bar{F}_n), \tag{2.6}$$

and similarly,

$$P_n(x;\mathbf{F}_n) = \prod_{i=1}^{n} F_i(x) \leqq [n^{-1}\sum_{i=1}^{n} F_i(x)]^n = [\bar{F}_n(x)]^n = P_n^*(x;\bar{F}_n). \tag{2.7}$$

Hence the theorem.

(ii) *Theorem* 1.2. It follows from Theorem 1.1 that for $2 \leqq r \leqq n-1$,

$$P_r(\xi_{n,r-1};\mathbf{F}_n) \leqq P_r^*(\xi_{n,r-1};\bar{F}_n), \qquad P_r(\xi_{n,r}, \mathbf{F}_n) \geqq P_r^*(\xi_{n,r};\bar{F}_n). \tag{2.8}$$

Now, it is well known that if we have n independent trials with a constant probability success p and if $np = s$ is an integer, then

$$\sum_{j=0}^{s-1}\binom{n}{j}p^j(1-p)^{n-j} < \tfrac{1}{2} < \sum_{j=0}^{s}\binom{n}{j}p^j(1-p)^{n-j}. \tag{2.9}$$

Consequently, upon noting that $n\bar{F}_n(\xi_{n,r-1}) = r-1$ and $n\bar{F}_n(\xi_{n,r}) = r$, we have from (2.2) and (2.9) that

$$P_r^*(\xi_{n,r-1};\bar{F}_n) = 1-\sum_{j=0}^{r-1}\binom{n}{j}[(r-1)/n]^j[1-(r-1)/n]^{n-j} < \tfrac{1}{2}, \tag{2.10}$$

$$P_r^*(\xi_{n,r};\bar{F}_n) = 1-\sum_{j=0}^{r-1}\binom{n}{j}(r/n)^j(1-r/n)^{n-j} > \tfrac{1}{2}. \tag{2.11}$$

From (2.8), (2.10) and (2.11), we have

$$P_r(\xi_{n,r-1};\mathbf{F}_n) \leqq P_r^*(\xi_{n,r-1};\bar{F}_n) < \tfrac{1}{2} < P_r^*(\xi_{n,r};\bar{F}_n) \leqq P_r(\xi_{n,r};\mathbf{F}_n), \tag{2.12}$$

and hence, the theorem follows by simple reasonings.

(iii) *Theorem* 1.3. We shall consider only the case $2 \leqq r \leqq n-1$, as for $r = 1$ or n, we require $\nu_\delta(\bar{F}_n) < \infty$ for $\delta \geqq k$, and hence the proof follows more directly. Now, by Theorem 1.1, the tails of the distribution $P_r(x; \mathbf{F}_n)$ are dominated by the tails of $P_r^*(x; \bar{F}_n)$ and $\xi_{n,r}(1 \leqq r \leqq n-1)$ are all finite. Hence, it suffices to show that under the stated condition

$$\int_{-\infty}^{\xi_{n,r-1}} |x|^k \, dP_r^*(x;\bar{F}_n) \quad \text{and} \quad \int_{\xi_{n,r}}^{\infty} |x|^k \, dP_r^*(x;\bar{F}_n) \quad \text{both exist.} \tag{2.13}$$

For this, we let (without any loss of generality) $\xi_{n,r} = 0$. Then, the second integral of (2.13) reduces to (for $r \leqq n-r_0+1$)

$$\begin{aligned} \int_0^\infty x^k \, dP_r^*(x;\bar{F}_n) &= r\tbinom{n}{r}\int_0^\infty x^k[\bar{F}_n(x)]^{r-1}[1-\bar{F}_n(x)]^{n-r} \, d\bar{F}_n(x). \\ &\leqq r\tbinom{n}{r}\int_0^\infty x^\delta\{x^{k-\delta}[1-\bar{F}_n(x)]^{r_0-1}\}[1-\bar{F}_n(x)]^{n-r_0+1-r} \, d\bar{F}_n(x) \\ &\leqq r\tbinom{n}{r}(1-r/n)^{n-r_0+1-r}\int_0^\infty x^\delta[c_n(x)] \, d\bar{F}_n(x), \end{aligned} \tag{2.14}$$

where $c_n(x) = x^{k-\delta}[1-\bar{F}_n(x)]^{r_0-1} \leqq [x^\delta[1-\bar{F}_n(x)]]^{r_0-1}$ is bounded above by $[\nu_\delta(\bar{F}_n)]^{r_0-1}$ and it tends to zero as $x \to \infty$. Thus, by the hypothesis that $\nu_\delta(\bar{F}_n) < \infty$, the right-hand side of (2.14) exists for all $r \leqq n-r_0+1$. Similarly, the first integral of (2.13) exists when-ever $r \geqq r_0$. Hence the theorem.

3. A few additional remarks. The asymptotic normality of $n^{\frac{1}{2}}(X_{n,r}-\xi_{n,p}^*)$ (where $\bar{F}_n(\xi_{n,p}^*) = p$, and $r = [np]+1$), follows directly from the results of Sen (1968) (when we put $m = 0$ i.e., consider an independent process). Also, by virtue of Theorem 1.1, the results of Theorem 2 of Sen (1959) can be readily extended to show that the kth central moment of $n^{\frac{1}{2}}(X_{n,r}-\xi_{n,p}^*)$ is asymptotically equal to the kth moment of the sequences of normal distribution with means zero and variances

$$\{n^{-1}\textstyle\sum_{i=1}^n F_i(\xi_{n,p}^*)[1-F_i(\xi_{n,p}^*)]/\bar{f}_n^{\,2}(\xi_{n,p}^*)\},$$

where $\bar{f}_n(x) = (d/dx)\bar{F}_n(x)$ is assumed to be continuous at $x = \xi_{n,p}^*$.

REFERENCES

[1] ANDERSON, T. W. and SAMUELS, S. M. (1966). Some inequalities among binomial and Poisson probabilities. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 1–12.

[2] BLOM, G. (1958). *Statistical Estimates and Transformed Beta Variables.* Almqvist and Wiksell, Stockholm.

[3] HOEFFDING, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27** 713–721.

[4] SAMUELS, S. M. (1965). On the number of success in independent trials. *Ann. Math. Statist.* **36** 1272–1278.

[5] SEN, P. K. (1959). On the moments of the sample quantiles. *Calcutta Statist. Assoc. Bull.* **9** 1–20.

[6] SEN, P. K. (1968). Asymptotic normality of sample quantiles for m-dependent processes. *Ann. Math. Statist.* **39** 1724–1730.