The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 2140–2143

ON SOME CONVERGENCE PROPERTIES OF ONE-SAMPLE RANK ORDER STATISTICS[1]

BY PRANAB KUMAR SEN

University of North Carolina

1. Summary and introduction. For a broad class of one-sample rank order statistics, almost sure (a.s.) convergence and exponential bounds for the probability of large deviations, when the basic random variables are not necessarily identically distributed, are established here. In this context, extending a result of Brillinger (1962) to the case of non-iidrv (independent and identically distributed random variables), a result on the a.s. convergence of sample means for a double sequence of random variables is derived. These results are of importance for the study of the properties of sequential tests and estimates based on rank order statistics.

2. Statement of the results. Consider a double sequence of random variables $\{X_{n1}, X_{n2}, \cdots\}$, $n \geqq 1$, where the random variables of the same row are independently distributed with continuous distribution functions (df) $\{F_{n1}, F_{n2}, \cdots\}$, $n \geqq 1$. For a sample $(X_{n1}, \cdots, X_{nn})$, consider the usual one-sample rank order statistic (cf. [5], [6])

$$T_n = n^{-1} \sum_{i=1}^{n} c(X_{ni}) E_n(R_{ni}); \qquad R_{ni} = \sum_{j=1}^{n} c(|X_{ni}| - |X_{nj}|), \qquad i = 1, \cdots, n, \tag{2.1}$$

where $c(u)$ is 1 or 0 according as u is $\geqq$ or < 0, and the rank scores $E_n(1), \cdots, E_n(n)$ are generated by a continuous score-function $J(u)$: $0 < u < 1$, in the following manner:

$$E_n(i) = J(i/(n+1)), \qquad i = 1, \cdots, n. \tag{2.2}$$

We may also work with $E_n(i) = J_n(i/(n+1))$, where the function $J_n(u)$: $0 < u < 1$, satisfies

$$\lim_{n \to \infty} \int_0^1 |J_n(u) - J(u)| \, du = 0. \tag{2.3}$$

Assume that $J \in L^r$, i.e., $\int_0^1 |J(u)|^r \, du < \infty$, for some positive r, to be specified later on. Let then

$$\bar{F}_n = n^{-1} \sum_{i=1}^{n} F_i, \qquad \bar{H}_n(x) = \bar{F}_n(x) - \bar{F}_n(-x), \qquad x \geqq 0; \tag{2.4}$$

$$\tau_n = \int_0^\infty J(\bar{H}_n(x)) \, d\bar{F}_n(x). \tag{2.5}$$

Note that if the X_{ni} $(= X_i)$ are iidrv with a df $F(x)$, independent of n, then $\bar{F}_n = F$, $\bar{H}_n(x) = H(x) = F(x) - F(-x)$, $x \geqq 0$, and $\tau_n = \tau = \int_0^\infty J(H(x)) \, dF(x)$. Finally, define $M(t)$ by

$$M(t) = \int_0^1 \{\exp [tJ(u)]\} \, du, \qquad \text{whenever it exists.} \tag{2.6}$$

Received August 4, 1969; revised April 27, 1970.
[1] Work supported by the National Institutes of Health, Grant GM-12868.

Then, we have the following theorems.

THEOREM 1. *If* $J \in L^r$ *for some* $r > 2$, *then* $\lim_{n\to\infty}(T_n - \tau_n) = 0$ a.s.; *if the* $X_{ni}\,(= X_i)$ *are* iidrv, *then* $\lim_{n\to\infty} T_n = \tau$ a.s. *whenever* $J \in L^1$.

THEOREM 2. *If for some* $T(>0)$, $M(t) < \infty$, $\forall |t| \leqq T$, *then for every* $\varepsilon > 0$, $\exists$ *a* $\rho(\varepsilon): 0 < \rho(\varepsilon) < 1$, *and an* $n_0(\varepsilon)$, *such that for all* $n \geqq n_0(\varepsilon)$,

$$P\{|T_n - \tau_n| > \varepsilon\} \leqq 2\{\rho(\varepsilon)\}^n, \quad \textit{uniformly in} \quad \{F_{n1}, \cdots, F_{nn}\}. \tag{2.7}$$

The proofs of these theorems are sketched in Section 3. For non-iidrv's, the task of establishing the a.s. convergence of sample means of a double sequence of rv's, needed in proving Theorem 1, is accomplished by the following theorem whose proof is given in the appendix. Here we let $\bar{X}_n = n^{-1}\sum_{i=1}^n X_{ni}$, $v_{ni}^{(r)} = \int_{-\infty}^{\infty} |x|^r\, dF_{ni}(x)$, $i = 1, \cdots, n$, and $\bar{v}_n^{(r)} = n^{-1}\sum_{i=1}^n v_{ni}^{(r)} = \int_{-\infty}^{\infty} |x|^r\, d\bar{F}_n(x)$.

THEOREM 3. *For a double sequence* $\{X_{n1}, X_{n2}, \cdots\}$, $n \geqq 1$, *of (row-wise independent)* rv'*s, if* $\sup_n \bar{v}_n^{(r)} \leqq K < \infty$, *for some* $r > 1$, *then for every* $\varepsilon > 0$, $\exists$ *a positive* C_ε *and an* $n_0(\varepsilon)$, *such that*

$$P\{|\bar{X}_n - E\bar{X}_n| > \varepsilon\} \leqq C_\varepsilon n^{-s}, \quad \textit{for all} \quad n \geqq n_0(\varepsilon), \tag{2.8}$$

where $s = r-1$, $1 < r \leqq 2$, *and* $s = \min(r-1, k)$ *for* $2(k-1) < r \leqq 2k$, $k \geqq 2$.

Note that for iidrv's, when $r > 2$, Brillinger (1962) has obtained in (2.8), $s = r/2$, whereas our $s \geqq r/2$ and the result holds even for non-iidrv's.

3. Derivation of the main results.

Define the empirical df's by

$$F_n(x) = n^{-1}\sum_{i=1}^n c(x - X_{ni}), \tag{3.1}$$

$$-\infty < x < \infty, \quad H_n(x) = F_n(x) - F_n(-x-), \quad x \geqq 0,$$

and express T_n in (2.1) equivalently as

$$T_n = \int_0^\infty J(H_n(x))\, dF_n(x). \tag{3.2}$$

If we let $J_n(u) = J(i/(n+1))$ for $(i-1)/n < u \leqq i/n$, $i = 1, \cdots, n$, as $J \in L^1$ (in any case), $\lim_{n\to\infty}\int_0^1 |J_n(u)|\, du = \int_0^1 |J(u)|\, du < \infty$. Hence, for every $\varepsilon > 0$, there exists a $\delta > 0$ and an $n_0(\varepsilon)$, such that for all $n \geqq n_0(\varepsilon)$,

$$\int_0^\delta + \int_{1-\delta}^1 |J_n(u)|\, du < \varepsilon/4 \left(\Rightarrow \int_0^\delta + \int_{1-\delta}^1 |J(u)|\, du < \varepsilon/4\right). \tag{3.3}$$

Also, $J(u)$ is continuous on (0, 1). Hence, for every $\varepsilon > 0$, there exist δ_1, δ_2, where $0 < \delta_2 < \delta_1/2$, $\delta_1 + \delta_2 = \delta$, such that

$$\sup_{|v| < \delta_2} \sup_{\delta_1 < u < 1-\delta_1} |J(u+v) - J(u)| < \varepsilon/4. \tag{3.4}$$

Defining then $a_n = \sup\{x : \bar{H}_n(x) \leqq \delta_1\}$, $b_n = \inf\{x : \bar{H}_n(x) \geqq 1 - \delta_1\}$, noting that $d\bar{F}_n \leqq d\bar{H}_n$, $dF_n \leqq dH_n$, and making use of (2.5), (3.2)–(3.4), we obtain that

$$|T_n - \tau_n| < \sum_{r=1}^4 I_n^{(r)}, \tag{3.5}$$

where $I_n^{(1)} = \int_0^{a_n} + \int_{b_n}^{\infty} |J(\bar{H}_n)| \, d\bar{F}_n \leqq \int_0^{\delta_1} + \int_{1-\delta_1}^{1} |J(u)| \, du < \varepsilon/4$, and

$$I_n^{(2)} = \int_0^{a_n} + \int_{b_n}^{\infty} |J(nH_n/(n+1))| \, dF_n \leqq \int_0^{a_n} + \int_{b_n}^{\infty} |J(nH_n/(n+1))| \, dH_n, \tag{3.6}$$

$$I_n^{(3)} = \int_{a_n}^{b_n} |J(nH_n/(n+1)) - J(\bar{H}_n)| \, dF_n, \tag{3.7}$$

$$I_n^{(4)} = \left|\int_{a_n}^{b_n} J(\bar{H}_n) \, d[F_n - \bar{F}_n]\right| = \left|n^{-1} \textstyle\sum_{i=1}^{n} (Z_{ni} - EZ_{ni})\right|, \tag{3.8}$$

where $Z_{ni} = J(\bar{H}_n(X_{ni}))$ if $a_n \leqq X_{ni} \leqq b_n$ and $Z_{ni} = 0$, otherwise, for $i = 1, \cdots, n$. Since, $n^{-1}\sum_{i=1}^{n} E|Z_{ni}|^r \leqq n^{-1}\sum_{i=1}^{n} \int_0^{\infty} |J(\bar{H}_n)|^r \, dF_{ni} \leqq \int_0^1 |J(u)|^r \, du < \infty$, for $J \in L^r$, by Theorem 3, when $r > 2$, $I_n^{(4)} \to 0$ a.s., as $n \to \infty$. When the X_{ni} $(= X_i)$ are iidrv, $a_n = a = H^{-1}(\delta_1)$ and $b_n = b = H^{-1}(1-\delta_1)$ do not depend on n, and thus, the Z_{ni} $(= Z_i)$ are iidrv with a df not depending on n. Hence, by the Kintchine law of large numbers, $I_n^{(4)} \to 0$ a.s. as $n \to \infty$.Hence, to prove Theorem 1, it suffices to show that for every $\varepsilon > 0$, $I_n^{(2)} + I_n^{(3)} \leqq \varepsilon/2$ a.s. as $n \to \infty$. It can be easily shown that for every $0 < \delta_2' < \delta_2$, as $n \to \infty$, $\sup_x |F_n(x) - \bar{F}_n(x)| < \delta_2'/2$ a.s.; this along with (3.1) and the fact that $(n+1)^{-1} < \delta_2 - \delta_2'$ for n sufficiently large leads to

$$\sup_{x \geqq 0} |nH_n(x)/(n+1) - \bar{H}_n(x)| < \delta_2 \quad \text{a.s.} \quad \text{as} \quad n \to \infty. \tag{3.9}$$

Thus, $nH_n(a_n)/(n+1) < \delta$ a.s. and $nH_n(b_n)/(n+1) > 1 - \delta$ a.s. (as $n \to \infty$), and hence, by (3.3), (3.6) and (3.9), $I_n^{(2)} < \varepsilon/4$ a.s. as $n \to \infty$. Similarly, by (3.4), (3.7) and (3.9), $I_n^{(3)} < \varepsilon/4$ a.s. as $n \to \infty$. Hence the proof is completed.

To prove Theorem 2, we note that by Theorem 1 of Hoeffding (1953) and proceeding as in Sethuraman (1964), for every $\delta_2 > 0$, $\exists$ a $\rho(\delta_2): 0 < \rho(\delta_2) < 1$, and an $n_0(\delta_2)$, such that for all $n \geqq n_0(\delta_2)$,

$$P\{\sup_x |nH_n(x)/(n+1) - \bar{H}_n(x)| > \delta_2\} \leqq [\rho(\delta_2)]^n. \tag{3.10}$$

Then, (3.3), (3.4) and (3.9) imply that $I_n^{(2)} + I_n^{(3)} \geqq \varepsilon/2$, with probability $\leqq [\rho(\delta_2)]^n$ for all $n \geqq n_0(\delta_2)$. Also, for all $|t| \leqq T(>0)$,

$$\begin{aligned} M_n(t) &= E[\exp(t\textstyle\sum_{i=1}^{n} Z_{ni})] = \prod_{i=1}^{n} E[\exp(tZ_{ni})] \\ &\leqq [n^{-1}\textstyle\sum_{i=1}^{n} E\{\exp(tZ_{ni})\}]^n \\ &\leqq [n^{-1}\textstyle\sum_{i=1}^{n} \int_0^{\infty} \{\exp[tJ(\bar{H}_n(x))]\} \, dF_{ni}(x)]^n \\ &\leqq [\int_0^1 \{\exp(tJ(u))\} \, du]^n \leqq [M(t)]^n, \end{aligned} \tag{3.11}$$

uniformly in $\{F_{n1}, \cdots, F_{nn}\}$ (as $d\bar{F}_n \leqq d\bar{H}_n$). Hence, by the same technique as in Bahadur and Rao (1960), it follows that for all $n \geqq n_0(\varepsilon)$, $I_n^{(4)} > \varepsilon/4$, with probability $\leqq [\rho_1(\varepsilon)]^n$, where $0 < \rho_1(\varepsilon) < 1$. The proof of Theorem 2 is then completed by letting $\rho(\varepsilon) = \max[\rho(\delta_2), \rho_1(\varepsilon)]$.

REMARK. For $E_n(i)$ satisfying (2.3), $|T_n - n^{-1}\sum_{i=1}^{n} c(X_{ni})J(R_{ni}/(n+1))| \to 0$ as $n \to \infty$, and hence, Theorem 1 and Theorem 2 remain valid.

4. Appendix: proof of Theorem 3. We present the proof only for $2 < r \leqq 4$; for $r > 4$, the proof proceeds on parallel lines, while for $1 < r \leqq 2$, the result follows from Lemma 1 of Chatterji (1969). Let $Y_{ni} = X_{ni}$ or 0 according as $|X_{ni}| < n$ or not, $i = 1, \cdots, n$, and let $\bar{Y}_n = n^{-1}\sum_{i=1}^n Y_{ni}$. Then,

$$P[\bar{X}_n \neq \bar{Y}_n] \leqq P\{|X_{ni}| > n \text{ for at least one } i = 1, \cdots, n\} = 1 - \prod_{i=1}^n [F_{ni}(n) - F_{ni}(-n)]. \tag{4.1}$$

By the Chebichev inequality, $\sup_{x \geqq 0} \{x^r[1-F_{ni}(x)] \text{ or } x^r F_{ni}(-x)\} \leqq v_{ni}^{(r)}$, and hence,

$$F_{ni}(-n) \leqq v_{ni}^{(r)}/n^r, \qquad 1-F_{ni}(n) \leqq v_{ni}^{(r)}/n^r, \qquad \text{for all } i = 1, \cdots, n. \tag{4.2}$$

Also, $\sup_n \bar{v}_n^{(r)} < K < \infty \Rightarrow v_{ni}^{(r)} \leqq Kn$, $\forall i$, and hence, for $r > 1$ and $n \geqq (2K)^{1/(r-1)}$, by (4.1) and (4.2),

$$P[\bar{X}_n \neq \bar{Y}_n] \leqq 1 - \prod_{i=1}^n [1-2n^{-r}v_{ni}^{(r)}] \leqq 2n^{-(r-1)}\bar{v}_n^{(r)} \leqq 2K/n^{r-1}. \tag{4.3}$$

Finally, note that

$$|E(\bar{X}_n - \bar{Y}_n)| \leqq n^{-1}\sum_{i=1}^n \int_{|x|>n} |x|\, dF_{ni}(x) = \int_{|x|>n} |x|\, d\bar{F}_n(x) \leqq n^{-(r-1)}\int_{|x|>n} |x|^r\, d\bar{F}_n(x) \leqq Kn^{-(r-1)}; \tag{4.4}$$

$$\begin{aligned} E(\bar{Y}_n - E\bar{Y}_n)^4 &= n^{-4}[\textstyle\sum_{i=1}^n E(Y_{ni}-EY_{ni})^4 \\ &\quad + \textstyle\sum_{i\neq j=1}^n E(Y_{ni}-EY_{ni})^2 E(Y_{nj}-EY_{nj})^2] \\ &\leqq 16n^{-(r-1)}(n^{-1}\textstyle\sum_{i=1}^n E\,|Y_{ni}|^r) + n^{-2}(n^{-1}\textstyle\sum_{i=1}^n EY_{ni}^2)^2 \\ &\leqq 16\bar{v}_n^{(r)}/n^{r-1} + (\bar{v}_n^{(r)})^{4/r}/n^2 \leqq K^* n^{-s}, \qquad K^* < \infty, \end{aligned} \tag{4.5}$$

where s is defined after (2.8). Hence, by the Chebichev inequality,

$$P[|\bar{Y}_n - E\bar{Y}_n| > \varepsilon] \leqq K^*/\varepsilon^4 n^s = K_\varepsilon n^{-s}. \tag{4.6}$$

The proof of (2.8) follows directly from (4.3), (4.4) and (4.6), while the use of the Borel–Cantelli lemma on (2.8) yields the desired a.s. convergence result.

Acknowledgments. The author is grateful to the referee and to Dr. M. Ghosh for their useful comments on the paper.

REFERENCES

[1] BAHADUR, R. R., and RAO, R. RANGA (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

[2] BRILLINGER, D. R. (1962). A note on the rate of convergence of a mean. *Biometrika* **49** 574–576

[3] CHATTERJI, S. D. (1969). An L^p-convergence theorem. *Ann. Math. Statist.* **40** 1068–1070.

[4] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[5] PURI, M. L. and SEN, P. K. (1969). On the asymptotic normality of one-sample rank order test statistics. *Teoria Veroi. i. Primen* **14** 167–172.

[6] SEN, P. K. (1970). On the distribution of the one-sample rank order statistics. *Proc. First Internat. Symp. Nonparametric Tech.*, ed. M. L. Puri. Cambridge Univ., 53–72.

[7] SETHURAMAN, J. (1964). On the probability of large deviations of families of sample means. *Ann. Math. Statist.* **35** 1304–1316.