The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1768–1773

SOME CONVERGENCE THEOREMS FOR RANKS AND WEIGHTED EMPIRICAL CUMULATIVES

BY HIRA LAL KOUL

Michigan State University

0. Summary. In this paper two convergence theorems are proved. One gives a strong law of large numbers for a class of linear rank statistics and the other gives weak convergence of a weighted empirical cumulative process to Gaussian process, concentrated on continuous sample functions. Of course, both of these results are true under some regularity condition on the quantities involved.

1. Strong law of large numbers for linear rank statistics. Let $\{X_{in}\, 1 \leqq i \leqq n\} n \geqq 1$ be sequences of independent random variables with distributions $\{F_{in},\ 1 \leqq i \leqq n\}$ $n \geqq 1$. Let $\{C_{in}\, 1 \leqq i \leqq n\}$ be arbitrary constants such that if

$$\sigma_{nc}^2 = n^{-1} \textstyle\sum_{i=1}^n C_{in}^2.$$

then

$$\textstyle\sum_{n=1}^\infty [\max_{1 \leqq i \leqq n} C_{in}^2 / n\, \sigma_{nc}^2]^2 < \infty. \tag{1.1}$$

$$\lim_{n\to\infty} \max_{1 \leqq i \leqq n} C_{in}^2 / n\, \sigma_{nc}^2 = 0. \tag{1.1'}$$

Let φ be a function on [0, 1] to the real line such that

$$0 < \textstyle\int_0^1 |\varphi^4(u)|\, du < \infty. \tag{1.2}$$

Define for any $-\infty < t < +\infty$

$$H_n(t) = n^{-1} \textstyle\sum_{i=1}^n I(X_{in} \leqq t) \tag{1.3}$$
$$\bar{H}_n(t) = n^{-1} \textstyle\sum_{i=1}^n F_{in}(t).$$

One can prove, using the fourth moment of $|H_n(t) - \bar{H}_n(t)|$ and the Borel–Cantelli theorem, that

$$\sup_{-\infty < t < +\infty} |Z_n(t)| = \sup_{-\infty < t < +\infty} |H_n(t) - \bar{H}_n(t)| \to 0 \quad \text{a.s.} \tag{1.4}$$

as $n \to \infty$.

Let

$$S_n = n^{-1} \textstyle\sum_{i=1}^n C_{in} \varphi[R_{in}/(n+1)]$$
$$T_n = n^{-1} \textstyle\sum_{i=1}^n C_{in} \varphi[\bar{H}_n(X_{in})] \tag{1.5}$$
$$\mu_n = ET_n = n^{-1} \textstyle\sum C_{in} \int_{-\infty}^{\infty} \varphi(\bar{H}_n(x))\, dF_{in}(x),$$

where

$$R_{in} = \textstyle\sum_{j=1}^n I(X_{jn} \leqq X_{in}).$$

Received June 4, 1969; revised April 9, 1970.

THEOREM 1.1. *If* φ *satisfies* (1.2) *and is continuous on* [0, 1], $\{C_{in}\}$ *satisfy* (1.1) *and* $\{F_{in}\ 1 \leqq i \leqq n\} n \geqq 1$ *are all continuous, then*

$$\sigma_{nc}^{-1}(S_n - \mu_n) \to 0 \quad \text{a.s.} \tag{1.6}$$

as $n \to \infty$.

PROOF. We need continuity of $\{F_{in}\}$ for $\{R_{in}\}$ to be properly defined. Secondly note that φ continuous on [0, 1] implies that φ is uniformly continuous also, and hence, we may replace $(n+1)$ in the definition of S_n by n. Next, with this modification, observe that S_n can also be written as

$$S_n = n^{-1} \textstyle\sum_{i=1}^n C_{in} \varphi[H_n(X_{in})].$$

Further using the Cauchy–Schwarz inequality one gets

$$\begin{aligned} |\sigma_{nc}^{-1}(S_n - T_n)| &= |\sigma_{nc}^{-1} n^{-1} \textstyle\sum_{i=1}^n C_{in}\{\varphi[H_n(X_{in})] - \varphi[\bar{H}_n(X_{in})]\}| \\ &\leqq \max_{1 \leqq i \leqq n} |\varphi[H_n(X_{in})] - \varphi[\bar{H}_n(X_{in})]| \\ &\leqq \sup_{-\infty < t < +\infty} |\varphi[H_n(t)] - \varphi[\bar{H}_n(t)]| \end{aligned}$$

which $\to 0$ a.s. in view of (1.7) and (1.4).

Next, (1.2) implies $\int_{-\infty}^{\infty} |\varphi^4(\bar{H}_n(x))|\, d\bar{H}_n(x) < \infty$ for all n, which in turn is equivalent to the fact that $\int_{-\infty}^{\infty} |\varphi^4(\bar{H}_n(x))|\, dF_{in}(x) < \infty \ \ 1 \leqq i \leqq n,\ n \geqq 1$. So if $Z_{in} = \varphi_{in} - \mu_{in}$, with $\varphi_{in} = \varphi(\bar{H}_n(X_{in}))$, $\mu_{in} = E\varphi_{in}$, one has $E|Z_{in}^k| < \infty$, $k = 1, 2, 3, 4$; $1 \leqq i \leqq n$, $n \geqq 1$. Moreover, by repeated use of the Cauchy–Schwarz inequality it can be shown that

$$|E[\sigma_{nc}^{-1}(T_n - \mu_n)]^4| \leqq K(\varphi)[\max_{1 \leqq i \leqq n} C_{in}^2 / n\, \sigma_{nc}^2]^2$$

where $K(\varphi) =$ constant $\int |\varphi|^4$. Consequently by (1.1) and (1.2) it follows that

$$\textstyle\sum_{n=1}^{\infty} E[\sigma_{nc}^{-1}(T_n - \mu_n)]^4 < \infty,$$

and hence $\sigma_{nc}^{-1}(T_n - \mu_n) \to 0$ a.s. This together with the fact $(S_n - T_n)\sigma_{nc}^{-1} \to 0$ a.s. $\Rightarrow$ $(S_n - \mu_n)\sigma_{nc}^{-1} \to 0$ a.s.

2. Weak convergence of weighted empirical processes. This section uses Theorem 12.1 and Theorem 15.5 of [1]. For the sake of completeness we restate Theorem 15.5 here.

Let $\{V_n(t); 0 \leqq t \leqq 1\}$ be a sequence of stochastic processes in $D[0, 1]$ the space of all functions on [0, 1] with discontinuities of type 1. Let for any $\delta > 0$,

$$W(V_n, \delta) = \sup_{|s-t| \leqq \delta} |V_n(s) - V_n(t)|. \tag{2.1}$$

Let $V(t)$, $0 \leqq t \leqq 1$ be another process.

THEOREM 2.1. (Equal to Theorem 15.5 taken together with Theorem 15.1 of [1]). *Suppose that for each* $\eta > 0\ \exists$ *an* a *such that*

$$\Pr[|V_n(0)| > a] < \eta \qquad n \geqq 1. \tag{2.2}$$

Suppose further that for every $\varepsilon > 0$

$$\lim_{\delta\to 0}\lim_{n\to\infty} \Pr[W(V_n, \delta) \geqq \varepsilon] = 0. \tag{2.3}$$

Also suppose

$$\mathscr{L}(V_n(t_j), 1 \leqq j \leqq k) \to \mathscr{L}(V(t_j), 1 \leqq j \leqq k) \tag{2.4}$$

for all continuity points of V. *Then* $V_n \to_D V$ *and* V *is in* $C[0, 1]$ *with probability* 1.

Our objective is to prove that the following sequence of processes L_n are weakly convergent to a continuous limit.

Let $\{X_{in}\}$ be independent random variables in [0, 1] with cdf's $\{F_{in}\}$ all continuous. Let $\{C_{in}\}$ be as in Section 1. Define, for $0 \leqq t \leqq 1$,

$$L_n(t) = (\sigma_{nc}^{-1})\, n^{-\frac{1}{2}} \textstyle\sum_{i=1}^n C_{in}\{I(X_{in} \leqq t) - F_{in}(t)\}. \tag{2.5}$$

LEMMA 2.1.

$$E[|L_n(t)-L_n(t_1)|^2\, |L_n(t_2)-L_n(t)|^2] \leqq 3[G_n(t_2)-G_n(t_1)]^2$$

for all $t_1 \leqq t \leqq t_2$, *and all* n, *where*

$$G_n(t) = \sigma_{nc}^{-2}\, n^{-1} \textstyle\sum_{i=1}^n C_{in}^2 F_{in}(t). \tag{2.6}$$

PROOF. Let

$$\alpha_{in} = I(t_1 < X_{in} \leqq t) - p_{in}(t_1, t)$$

$$\beta_{jn} = I(t < X_{jn} \leqq t_2) - p_{jn}(t, t_2)$$

where

$$p_{in}(u, v) = F_{in}(v) - F_{in}(u) \qquad 0 \leqq u, v \leqq 1.$$

Then

$$L_n(t) - L_n(t_1) = \sigma_{nc}^{-1}\, n^{-\frac{1}{2}} \textstyle\sum_{i=1}^n C_{in}\alpha_{in}$$

$$L_n(t_2) - L_n(t) = \sigma_{nc}^{-1}\, n^{-\frac{1}{2}} \textstyle\sum_{i=1}^n C_{in}\beta_{in}.$$

Using the independence of $\{\alpha_{in}\}$, of $\{\beta_{jn}\}$, and that of α_{in} from β_{jn} $i \neq j$, and the fact that $E\alpha_{in} = E\beta_{in} = 0$ for all i, one can conclude that

$$\begin{aligned} E\,|L_n(t)-L_n(t_1)|^2|L_n(t_2)-L_n(t_1)|^2 &= n^{-2}\sigma_{nc}^{-4}\{\textstyle\sum_{i=1}^n C_{in}^4 E\alpha_{in}^2\beta_{in}^2 + 2\sum\sum_{i\neq j} C_{in}^2 C_{jn}^2 E\alpha_{in}^2 E\beta_{jn}^2 \\ &\quad + \textstyle\sum\sum_{i\neq j} C_{in}^2 C_{jn}^2 E(\alpha_{in}\beta_{in})E(\alpha_{jn}\beta_{jn})\} \end{aligned}$$

which may be easily shown to be

$$\begin{aligned} &\leqq n^{-2}\sigma_{nc}^{-4}\{3\textstyle\sum_{i=1}^n C_{in}^4 p_{in}(t_1, t)p_{in}(t, t_2) + 2\sum\sum_{i\neq j} C_{in}^2 C_{jn}^2 p_{in}(t_1, t)p_{in}(t, t_2) \\ &\qquad + \textstyle\sum\sum_{i\neq j} C_{in}^2 C_{jn}^2 p_{in}(t_1, t)p_{jn}(t, t_2)\} \\ &\leqq 3\sigma_{nc}^{-4}[n^{-1}\textstyle\sum_{i=1}^n C_{in}^2 p_{in}(t_1, t)][n^{-1}\sum_{i=1}^n C_{in}^2 p_{in}(t, t_2)]. \end{aligned}$$

But

$$t_1 \leqq t \leqq t_2 \Rightarrow p_{in}(t_1, t) = F_{in}(t) - F_{in}(t)$$
$$\leqq F_{in}(t_2) - F_{in}(t_1),$$
$$p_{in}(t, t_2) = F_{in}(t_2) - F_{in}(t)$$
$$\leqq F_{in}(t_2) - F_{in}(t_1)$$

and hence

$$E[|L_n(t) - L_n(t_1)|^2 |L_n(t_2) - L_n(t)|^2]$$
$$\leqq 3[\sigma_{nc}^{-2} n^{-1} \textstyle\sum_{i=1}^n C_{in}^2 p_{in}(t_2, t_1)]^2$$
$$= 3[G_n(t_2) - G_n(t_1)]^2.$$

LEMMA 2.2. *For any* $\eta > 0$ *and* $t_1 < t_2$ *fixed, we have*

(2.8) $$\Pr[\sup_{t_1 \leqq s \leqq t_2} |L_n(s) - L_n(t_1)| \geqq \eta]$$
$$\leqq K/\eta^2 [G_n(t_2) - G_n(t_1)]^2 + \Pr[|L_n(t_2) - L_n(t_1)| \geqq \eta/2]$$

for all n. K *is independent of* t_1, t_2 *and* n.

PROOF. In view of Lemma 2.1 above, Theorem 12.1 of [1] is applicable to rv's

$$\xi_j = L_n((j/m)\delta + t_1) - L_n((j-1/m)\delta + t_1) \qquad 1 \leqq j \leqq m$$

with $\gamma = 2$, $\alpha = 1$ and

$$u_j = G_n((j/m)\delta + t_2) - G_n((j-1/m)\delta + t_1) \qquad 1 \leqq j \leqq m.$$

In the above $\delta = t_2 - t_1$.

Finally using inequality (12.4) of [1] and right continuity of L_n for each n, one gets (2.8).

LEMMA 2.3. *Assume* $\{C_{in}\}$ *satisfy* (1.1′) *and* $\{F_{in}\}$ *are continuous and behave in the limit such that for any* $0 = t_0 < t_1 < \cdots < t_r = 1$, $t_j - t_{j-1} \leqq \delta$, $1 \leqq j \leqq r$

(2.9) $$\lim_{\delta \to 0} \lim_{n \to \infty} \max_{1 \leqq i \leqq n} \max_{1 \leqq j \leqq r} |F_{in}(t_j) - F_{in}(t_{j-1})| = 0.$$

Then for every $\varepsilon > 0$

(2.10) $$\lim_{\delta \to 0} \lim_{n \to \infty} \Pr[W(L_n, \delta) \geqq \varepsilon] = 0.$$

PROOF. For a $\delta > 0$ let $0 = t_0 < t_1 < \cdots < t_r = 1$ be a partition such that $t_i - t_{i-1} = \delta$ $1 \leqq i \leqq r$. Then

(2.11) $$\Pr[W_n(L_n, \delta) \geqq \varepsilon] \leqq \textstyle\sum_{i=1}^r \Pr[\sup_{t_{i-1} \leqq s \leqq t_i} |L_n(s) - L_n(t_{i-1})| \geqq \varepsilon/3]$$
$$\leqq K_\varepsilon \textstyle\sum_{i=1}^r [G_n(t_i) - G_n(t_{i-1})]^2 + \textstyle\sum_{i=1}^r \Pr[|L_n(t_i) - L_n(t_{i-1})| \geqq \varepsilon/6].$$

First inequality follows from Corollary 8.3 of [1], and second inequality follows from Lemma 2.2 above.

We shall show that the right-hand side of (2.11) tends to 0 as $n \to \infty$ and $\delta \to 0$. If

$$s_n^2(i) = \sigma_{nc}^{-2} n^{-1} \sum_{j=1}^n C_{jn} p_{jn}(t_{i-1}, t_i)\{1 - p_{jn}(t_{i-1}, t_i)\}$$

which is $\leqq [G_n(t_i) - G_n(t_{i-1})]$, then under (1.1′) it is not hard to see that, for each $1 \leqq i \leqq r$ fixed,

$$(s_n^{-1}(i)\{L_n(t_i) - L_n(t_{i-1})\}) \to N(0, 1) \quad \text{as} \quad n \to \infty$$

and hence for n sufficiently large, using the Markov inequality on $N(0, 1)$ rv with fourth moment, we have

$$\Pr[|L_n(t_i) - L_n(t_{i-1})| \geqq \varepsilon/6] \leqq 3.6^4/\varepsilon^{-4} \cdot s_n^4(i) \leqq K_{1\varepsilon}[G_n(t_i) - G_n(t_{i-1})]^2,$$

$3 = EZ^4$, $\mathscr{L}(Z) = N(0, 1)$. Therefore for n sufficiently large

$$\Pr[W_n(L_n, \delta) \geqq \varepsilon] \leqq (K_\varepsilon + K_{1\varepsilon}) \sum_{i=1}^r [G_n(t_i) - G_n(t_{i-1})]^2.$$

But, since $\sum_{i=1}^r [G_n(t_i) - G_n(t_{i-1})] = 1$, we have

$$\begin{aligned} \sum_{i=1}^r [G_n(t_i) - G_n(t_{i-1})]^2 &\leqq \max_{1 \leqq j \leqq r} [G_n(t_j) - G_n(t_{j-1})] \cdot 1 \\ &\leqq \max_{1 \leqq i \leqq n} \max_{1 \leqq j \leqq r} |F_{in}(t_j) - F_{in}(t_{j-1})| \to 0 \end{aligned}$$

by (2.9) as $n \to \infty$ and $\delta \to 0$.

Hence the lemma.

Note. A sequence $\{t_i\}$ that could be used in the above is $t_i = i\delta$; then $r = [\delta^{-1}]$. Also it is implicitly assumed that for each i, $\lim_{n\to\infty} s_n^2(i)$ exists.

THEOREM 2.2. *Let* $\{X_{in}, 1 \leqq i \leqq n\} n \geqq 1$ *be independent* rv's *with* cdf's $\{F_{in}\}$, *all defined on* [0, 1]. *Let* $\{C_{in}\}$ *be some arbitrary constants satisfying* (1.1′). *Assume* $\{F_{in}\}$ *to be continuous and such that* (2.9) *is satisfied. Also assume that*

$$\begin{aligned} (2.12) \quad K(s, t) &= \lim_{n\to\infty} \operatorname{Cov}(L_n(s), L_n(t)) = \lim_{n\to\infty} K_n(s, t) \\ &= \lim_{n\to\infty} \sigma_{nc}^{-2} n^{-1} \sum_{i=1}^n C_{in}^2 \min[F_{in}(t), F_{in}(s)][1 - \max\{F_{in}(t), F_{in}(s)\}] \end{aligned}$$

exists. Then

$$\mathscr{L}(L_n(t), 0 \leqq t \leqq 1) \to \mathscr{L}(L(t), 0 \leqq t \leqq 1)$$

where $L(t)$ is a Gaussian process tied down to zero at 0 *and* 1, *with*

$$E\,L(t) = 0 \qquad 0 \leqq t \leqq 1$$

$$\operatorname{Cov}(L(s), L(t)) = K(s, t),$$

and with continuous sample functions.

PROOF. Notice that $L_n \in D[0, 1]$ for all n. Under (1.1′) and (2.12) it is easy to verify that for each fixed t,

$$\begin{aligned} \lim_{n\to\infty} \mathscr{L}(L_n(t) K_n^{-\frac{1}{2}}(t, t)) &= \lim_{n\to\infty} \mathscr{L}(L_n(t) K^{-\frac{1}{2}}(t, t)) \\ &= N(0, 1) \end{aligned}$$

or $\mathscr{L}(L_n(t)) \to N(0, K(t, t))$.

Hence the finite dimensional distributions of $L_n(t)$ converge to that of a Gaussian process, say L, at all continuity points of L. Furthermore, since $L_n(0) = 0 = L_n(1)$ for all n, hence $L(0) = 0 = L(1)$ and (2.3) is obviously satisfied by L_n.

Finally, combining Lemmas 2.1, 2.2, 2.3 and the above remarks, we see that the conditions of Theorem 2.1 are satisfied and hence $L_n \to_D L$ with L in $C[0, 1]$. The proof is terminated.

REMARK. (1) Note that $K(t, t)$ could be zero, for some t, but by a degenerate distribution we understand a normal distribution with zero variance.

(2) Let Y_{in} be i.i.d. F, d_{in} be some constants such that $n^{-1}\sum_{i=1}^{n} d_{in}^2 < \infty$ for all n and $\max_{1 \leqq i \leqq n} d_{in}^2/\sum d_{in}^2 \to 0$. Take $X_{in} = Y_{in} - \theta d_{in}$ for some $\theta \in [-a, a]\, a > \infty$.

Then

$$F_{in}(\cdot) = F(\cdot + \theta d_{in}).$$

Then $L_n(x, \theta) = n^{-\frac{1}{2}} \sum_i C_{in}[I(Y_{in} \leqq x + \theta d_{in}) - F(x + \theta d_{in})]\sigma_{nc}^{-1}$ is relatively compact for each θ provided F is continuous.

Actually using this fact, under additional assumption of bounded and continuous density f of F one can prove

$$\mathscr{L}(\sup_x \sup_{-a \leqq \theta \leqq +a} L_n(x, \theta n^{-\frac{1}{2}}))$$

is relatively compact. See [2].

(3) If $C_{in} = 1$ and $F_{in} = F$, $1 \leqq i \leqq n$, $n \geqq 1$ then one gets usual empirical process and clearly results are true for F continuous.

REFERENCES

[1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[2] KOUL, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple regression. *Ann. Math. Statist.* **40** 1950–1979.