The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 2144–2149

ASYMPTOTIC NORMALITY OF RANDOM RANK STATISTICS

BY HIRA LAL KOUL

Michigan State University

0. Summary. Asymptotic normality of a class of rank statistics based on random number of observations, called random rank statistics, is proved. Underlying rv's are assumed to be i.i.d.

1. Let Y_i, $i \geqq 1$ be a sequence of i.i.d. rv's with a cdf F, $\{c_i\}$ be a sequence of constants, N_r, $r \geqq 1$ a sequence of positive integer valued rv's and n_r, $r \geqq 1$ a sequence of positive integers. All rv's are defined on the same sample space.

Let

$$R_i = \sum_{j=1}^{N_r} I(|Y_j| \leqq |Y_i|) \qquad 1 \leqq i \leqq N_r. \tag{1.1}$$

Let φ be a score function defined on [0, 1] to real line.

Define

$$S_{N_r} = N_r^{-1} \sum_{i=1}^{N_r} c_i \varphi(R_i/N_r+1)\,\mathrm{sgn}\,(Y_i) \tag{1.2}$$

where $\mathrm{sgn}(x) = I(x \geqq 0) - I(x < 0)$.

Our main result is Theorem 2.1 which gives asymptotic normality of $N_r^{\frac{1}{2}} S_{N_r}$. This result could be used in the following situations.

Suppose we were observing Y_i, $i \geqq 1$, sequentially and stop after observing N_r observations and, would like to test $H_0: \beta = 0$ in the regression model $Y_i = \beta c_i + Z_i$, where Z_i are i.i.d. F symmetric about zero. One could use S_{N_r} as a test statistic to test H_0. Our result below says that under suitable conditions on the stopping variables N_r, F and $\{c_i\}$ the cut-off value for large r may be computed from normal tables. Another situation where asymptotic normality of $N_r^{\frac{1}{2}} S_{N_r}$ is useful in the problem of constructing bounded length confidence interval of prescribed coverage probability for β, using signed rank statistic, for example see [1].

It may be mentioned that our result is more general than the Pyke–Shorack result of [6] in the sense that one-sample and two-sample statistics can be obtained from S_{N_r} above by choosing $\{c_i\}$ appropriately. But our results are valid only under null hypothesis that Y_i are i.i.d. F. Furthermore our proof is simpler.

If we combine the comments mentioned at the end of paper [1] with our Theorem 2.1 here, we can obtain asymptotic normality of $N_r^{\frac{1}{2}} S_{N_r}$ under alternatives $\beta_r = N_r^{-\frac{1}{2}} \beta_0$ for some β_0.

We next state assumption.

Let $F \in \mathscr{F}$, where

$$\mathscr{F} = \{F;\, F \text{ a continuous cdf, } F(x) = 1 - F(-x)\ \forall\, x\}. \tag{1.3}$$

Received August 25, 1969; revised February 25, 1970.

About $\{c_i\}$ we assume that

$$\lim_{n\to\infty} \max_{1\leqq i\leqq n} c_i^2/\sum_{i=1}^n c_i^2 \to 0. \tag{1.4}$$

$\{n_r\}$ and $\{N_r\}$ are such that $n_r \to \infty$ as $r \to \infty$ and

$$N_r/n_r \to 1 \quad \text{in prob. as} \quad r \to \infty. \tag{1.5}$$

In the above and what follows r is thought to be integer, but this is no restriction. One can have r as continuous time parameter also. What is important is that N_r and n_r both be integer valued.

Let

$$G(x) = 2F(x) - 1 \qquad \text{for} \quad x > 0. \tag{1.6}$$

Without loss of generality we may assume that

$$G(x) \equiv x \qquad 0 \leqq x \leqq 1. \tag{1.7}$$

For, if there are any flat spots in G, one can delete these flat spots without changing the order of Y's and hence the distribution of ranks of $|Y|$'s like this one ends with a strictly increasing cdf G which now may be transformed by a strictly increasing transformation to the form given by (1.7).

About φ function, we assume that φ is absolutely continuous and

$$0 < \textstyle\int_0^1 \varphi^2 < \infty; \qquad \|\varphi\| = \int_0^1 |\varphi'(u)|\, du < \infty. \tag{1.8}$$

Let

$$\mu_r(x) = N_r^{-1} \textstyle\sum_{i=1}^{N_r} c_i I(Y_i \leqq x)\operatorname{sgn}(Y_i)$$

$$\bar{\mu}_r(x) = \bar{c}_r[F(x)\operatorname{sgn}(x) - I(x \geqq 0)] \tag{1.9}$$

$$\bar{c}_r = N_r^{-1} \textstyle\sum_{i=1}^{N_r} c_i$$

$$H_r(|x|) = N_r^{-1} \textstyle\sum_{i=1}^{N_r} I(|Y_i| \leqq |x|) \tag{1.10}$$

$$\bar{H}_r(|x|) = G(|x|)$$

$$L_r(x) = N_r^{\frac{1}{2}}[\mu_r(x) - \bar{\mu}_r(x)] \tag{1.11}$$

$$Z_r(|x|) = N_r^{\frac{1}{2}}[H_r(|x| - \bar{H}_r(|x|)]. \tag{1.12}$$

For $0 \leqq y \leqq 1$, we define

$$H_r^{-1}(y) = \inf\{x \geqq 0;\ H_r(x) \geqq y\}. \tag{1.13}$$

Also let

$$\sigma_n^2 = n^{-1} \textstyle\sum_{i=1}^n c_i^2 \cdot \int_0^1 \varphi^2(u)\, du = \sigma_{nc}^2 \sigma_\varphi^2. \tag{1.14}$$

Let

$$W_{N_r}(x) = N_r^{-\frac{1}{2}} \textstyle\sum_{i=1}^{N_r} d_i\{I(Y_i \leqq x) - F(x)\}. \tag{1.15}$$

Note the following relationships:
If $d_i = c_i$ in (1.15) we have

$$(1.16) \qquad \begin{aligned} L_r(x) &= W_{N_r}(x) - 2W_{N_r}(0) && \text{if } x \geqq 0; \\ &= W_{N_r}(x) && \text{if } x < 0. \end{aligned}$$

If $d_i = 1$, we have

$$(1.17) \qquad Z_r(|x|) = W_{N_r}(|x|) - W_{N_r}(-|x|).$$

In what follows all probability statements are computed under the probability measure given by $\{Y_i\}$ and $\{N_r\}$.

Before proceeding further we state a P. Lévy type inequality for rv's in $D[-\infty, +\infty]$ space. Its proof is a straightforward generalization of one appearing on page 45 in [4] under the name of Skorhod inequality after noticing that "sum" is measurable operation in $D[-\infty, +\infty]$ when metrized by Skorhod metric. Also see [2].

If X_i, $1 \leqq i \leqq n$ are independent rv's on $D[-\infty, +\infty]$, i.e. $X_i(t)$, $-\infty \leqq t \leqq +\infty$ is a stochastic process with jumps of first kind for each i, and if $S_n(t) = \sum_{i=1}^{n} X_i(t)$ then

$$(1.18) \qquad \operatorname{Prob}[\max_{1 \leqq j \leqq n} \|S_j\| \geqq 2\varepsilon] \leqq \frac{\operatorname{Prob}[\|S_n\| \geqq \varepsilon]}{1 - \max_{1 \leqq j \leqq n} \operatorname{Prob}[\|S_n - S_j\| \geqq \varepsilon]}$$

where $\|\cdot\|$ is sup norm.

LEMMA 1.1. *The stochastic processes* $\{\sigma_{n_r d}^{-1} W_{N_r}(x), -\infty \leqq x \leqq +\infty\}$, $r \geqq 1$ *are relatively compact as* $r \to \infty$, *with continuous Gaussian process as its limit, provided* $\max_{1 \leqq i \leqq n_r} d_i^2 / \sum_{i=1}^{n_r} d_i^2 \to 0$, *F is continuous and* $N_r/n_r \to 1$ *in probability.*

PROOF. One first shows that the processes $\{\sigma_{n_r d}^{-1} W_{n_r}(x), -\infty \leqq x \leqq +\infty\}$ have a continuous Gaussian process as a limit. The proof of this may be found in Theorem A3 of [5] by putting $t = 0$ in that theorem. However note that in Theorem A3 of [5] we have assumed that F be absolutely continuous with bounded density f and $\sigma_{n_r d}^2$ be bounded in the limit. But if one goes through that proof, one sees that these assumptions are not really needed once we normalize by $\sigma_{n_r d}^{-1}$, but continuity of F is crucial.

Next one compares $\sigma_{n_r d}^{-1} W_{N_r}$ with $\sigma_{n_r d}^{-1} W_{n_r}$. For, for any $\varepsilon > 0$, $\eta > 0$ we have

$$\begin{aligned} &\Pr[\|W_{N_r} - W_{n_r}\| \geqq 2\varepsilon\sigma_{n_r d}] \\ &\quad \leqq \Pr[\max_{n_r \leqq j \leqq m_r} \|W_j - W_{n_r}\| \geqq 2\varepsilon\sigma_{n_r d}] + \Pr[|N_r - n_r| \geqq \eta n_r] + \delta_r(\varepsilon, \eta) \\ &\quad \leqq \frac{\Pr[\|W_{m_r} - W_{n_r}\| \geqq \varepsilon\sigma_{n_r d}]}{1 - \max_{n_r \leqq j \leqq m_r} \Pr[\|W_{n_r} - W_j\| \geqq \varepsilon]} + \Pr[|N_r - n_r| \geqq \eta n_r] + \delta_r(\varepsilon, \eta) \end{aligned}$$

where $m_r = [n_r(1+\eta)+1]$, $[x]$ = greatest integer less than x.

Note that last inequality follows from (1.18). By assumption second term on the right-hand side above can be made small for large r. That first term can be made small is not hard to show (see [1], Lemma (1.3)). $\delta_r(\varepsilon, \eta)$ is similar to the first term on the right-hand side of first inequality where now max is taken over $\nu_r \leqq j \leqq n_r$; $\nu_r = [n_r - n_r\varepsilon]$, which may be shown to be small by argument similar to one used in showing first term is small. Hence the lemma.

LEMMA 1.2. *Under* (1.3), (1.4) *and* (1.5)

$$\sup_{0 \leqq y \leqq 1} \sigma_{n_r c}^{-1} |L_r(H_r^{-1}(y)) - L_r(\bar{H}_r^{-1}(y))| \to 0 \tag{1.19}$$

in probability as $r \to \infty$.

PROOF. Here we use the fact that $\bar{H}_r(x) \equiv G(x) \equiv x$, $0 \leqq x \leqq 1$. Now in view of Lemma 1.1, with $d_i = c_i$, we have in view of (1.16) that for every $\varepsilon > 0$

$$\lim_{\delta \to 0} \lim_{r \to \infty} \operatorname{Prob}[\sup_{|x-y| \leqq \delta} |L_r(x) - L_r(y)| \geqq \varepsilon\sigma_{n_r c}] = 0.$$

Again using (1.17) and Lemma 1.3 with $d_i = 1$, we have for any $\varepsilon > 0$

$$\sup_{-\infty \leqq x \leqq \infty} |H_r(|x|) - \bar{H}_r(|x|)| \to 0$$

in probability as $r \to \infty$.

But since $\bar{H}_r(x) \equiv x$ for $0 \leqq x \leqq 1$ we have, making change of variable,

$$\sup_{0 \leqq y \leqq \infty} |H_r^{-1}(y) - \bar{H}_r^{-1}(y)| = \sup_{0 \leqq y \leqq 1} |H_r^{-1}(y) - y| \to 0$$

in probability.

Hence

$$\begin{aligned} &\lim_{r \to \infty} \operatorname{Prob}[\sup_{0 \leqq y \leqq 1} \sigma_{n_r c}^{-1} |L_r(H_r^{-1}(y)) - L_r(H_r^{-1}(y))| \leqq \varepsilon] \\ &\quad \geqq \lim_{r \to \infty} \operatorname{Prob}[\sup_{|x-y| \leqq \delta} |L_r(x) - L_r(y)| \leqq \varepsilon\,\sigma_{n_r c}, \sup_{0 \leqq z \leqq 1} |H_r^{-1}(z) - z| \leqq \delta] \\ &\quad = 1. \end{aligned}$$

This then concludes the proof.

2. Asymptotic normality of S_{N_r}. In view of (1.2), (1.9) and (1.10) we can write

$$S_{N_r} = \int_{-\infty}^{\infty} \varphi(H_r(|x|))\, d\mu_r(x) \quad \text{a.s.} \tag{2.1}$$

Notice that

$$\int_{-\infty}^{\infty} \varphi(\bar{H}_r(|x|))\, d\bar{\mu}_r(x) = 0 = \int_{-\infty}^{\infty} \varphi(H_r(|x|))\, d\bar{\mu}_r(x) \tag{2.2}$$

holds with probability 1 in view of symmetry of F.

THEOREM 2.1. *Under* (1.3), (1.4), (1.5) *and* (1.8)

$$\mathscr{L}(N_r^{\frac{1}{2}} S_{N_r}/\sigma_{N_r}) \to N(0, 1) \tag{2.3}$$

as $r \to \infty$.

PROOF. The proof uses usual decomposition of S_{N_r} and facts (2.2).

We rewrite

$$\begin{aligned} S_{N_r} &= \int_{-\infty}^{\infty} \varphi(\bar{H}_r(|x|))\, d\{\mu_r(x) - \bar{\mu}_r(x)\} \\ &\quad + \int_{-\infty}^{\infty} [\varphi(H_r(|x|)) - \varphi(\bar{H}_r(|x|))]\, d\{\mu_r(x) - \bar{\mu}_r(x)\} \\ &= B_r + R_r \quad \text{say.} \end{aligned}$$

Recalling definition of L_r from (1.11), we have

$$\begin{aligned}\sigma_{n_r}^{-1}\left|N_r^{\frac{1}{2}}R_{r1}\right| &= \left|\int_0^\infty [\varphi(H_r(|x|))-\varphi(G(|x|))]\,dL_r(x)\right| \\ &= \left|\int_0^1 \{L_r(H_r^{-1}(y))-L_r(y)\}\,\varphi'(y)\,dy\right| \\ &\leqq \sup_{0\leqq y\leqq 1}\left|L_r(H_r^{-1}(y))-L_r(y)\right|\sigma_{n_rc}^{-1}\cdot\|\varphi\|\,\sigma_\varphi^{-1}\end{aligned}$$

which $\to 0$ in probability in view of (1.8) and (1.19).

The term

$$\sigma_{n_r}^{-1}N_r^{\frac{1}{2}}R_{r2} = \sigma_{n_r}^{-1}\int_{-\infty}^0 [\varphi(H_r(|x|))-\varphi(G(|x|))]\,dL_r(x)$$

may be handled similarly.

Therefore $\left|\sigma_{n_rc}^{-1}N_r^{\frac{1}{2}}R_r\right| = \left|\sigma_{n_rc}^{-1}N_r^{\frac{1}{2}}[R_{r1}+R_{r2}]\right| \to 0$ in probability as $r\to\infty$.

Next we show that

$$\begin{aligned}\sigma_{n_r}^{-1}N_r^{\frac{1}{2}}B_r &= \sigma_{n_r}^{-1}\int_{-\infty}^\infty \varphi(\bar{H}_r(|x|)\,dL_r(x) \\ &= \sigma_{n_r}^{-1}N_r^{-\frac{1}{2}}\sum_{i=1}^{N_r} c_i\,\varphi(G|Y_i|)\,\mathrm{sgn}\,(Y_i)\end{aligned}$$

have a limiting normal distribution.

Write $V_n = \sum_{i=1}^n c_i\varphi(G(|Y_i|))\,\mathrm{sgn}\,(Y_i)$. Then $\sigma_{n_r}^{-1}N_r^{\frac{1}{2}}B_r = \sigma_{n_rc}^{-1}N_r^{-\frac{1}{2}}V_{N_r}$. First we show that $\sigma_{n_r}^{-1}n_r^{-\frac{1}{2}}V_{N_r}$ has limiting normal distribution. Then since $N_r/n_r\to 1$ in probability, we can easily conclude that $\mathscr{L}(\sigma_{n_r}^{-1}N_r^{-\frac{1}{2}}V_{N_r})\to N(0,1)$ as $r\to\infty$.

Now for any $\varepsilon>0$, $\eta>0$

$$\begin{aligned}&\mathrm{Prob}\left[\left|n_r^{-\frac{1}{2}}V_{N_r}-n_r^{-\frac{1}{2}}V_{n_r}\right|\geqq\varepsilon\sigma_{n_r}\right] \\ &\leqq \mathrm{Prob}\left[\max_{n_r\leqq k\leqq m_r}|V_k-V_{n_r}|\geqq\varepsilon\sigma_{n_r}n_r^{\frac{1}{2}}\right]+\mathrm{Prob}\left[|N_r-n_r|>n_r\eta\right] \\ &\leqq \frac{\mathrm{Prob}\left[|V_{m_r}-V_{n_r}|\geqq\varepsilon\sigma_{n_r}n_r^{\frac{1}{2}}\right]}{1-\max_{n_r\leqq k\leqq m_r}\mathrm{Pr}\left[|V_{m_r}-V_k|\geqq\varepsilon\sigma_{n_r}n_r^{\frac{1}{2}}\right]}+\mathrm{Prob}\left[|N_r-n_r|>n_r\eta\right]\end{aligned}$$

where m_r is as defined in the proof of Lemma 1.1, and $\Delta_r(\varepsilon,\eta)$ is similar to the first term on the right-hand side of the first inequality where now max is taken $v_r\leqq j\leqq n_r$; $v_r=[n_r-n_r\eta]$.

Last inequality follows from applying (1.18).

Now let $M_r = (V_{m_r}-V_{n_r})\sigma_{n_r}^{-1}n_r^{-\frac{1}{2}}$. Note that M_r is sum of m_r-n_r independent rv's with means equal to zero.

Also observe that

$$\begin{aligned}\mathrm{Var}\,(M_r) &\leqq n_r^{-1}\sigma_{n_r}^{-2}\sum_{i=n_r+1}^{m_r} c_i^2\,E\,\varphi^2(G(|Y_i|)) \\ &= \sigma_{n_rc}^{-2}[(m_r/n_r)\,\sigma_{m_rc}^2-\sigma_{n_rc}^2] \\ &= (m_r/n_r)(\sigma_{m_rc}/\sigma_{n_rc})^2-1\end{aligned}$$

which can be made arbitrarily small for sufficiently large r and arbitrarily small η.

This implies $M_r\to 0$ in probability as $r\to\infty$ and hence first term in (2.3) tends to zero as $r\to\infty$. Second term of (2.3) tends to zero by (1.5). Similarly $\Delta_r(\varepsilon,\eta)$ may be shown to be small. Hence $\sigma_{n_r}^{-1} n_r^{-\frac{1}{2}} |V_{N_r} - V_{n_r}| \to 0$ in probability. But under (1.3), (1.4) and (1.8) it is easy to verify that $\sigma_{n_r}^{-1} n_r^{-\frac{1}{2}} V_{n_r}$ has limiting $N(0, 1)$ distribution [see 3]. Hence $\sigma_{n_r}^{-1} n_r^{-\frac{1}{2}} V_{N_r}$ has limiting normal $N(0, 1)$ distribution. In order to conclude the proof of (2.3) we need to show that

$$(\sigma_{N_r}^2 - \sigma_{n_r}^2)\sigma_{n_r}^2 \to 0 \quad \text{in probability.} \tag{2.5}$$

In order to prove (2.5), it is enough to show that $|\sigma_{N_r c}^2 - \sigma_{n_r c}^2| \sigma_{n_r c}^{-2} \to 0$ in probability. However, since $P([n_r - n_r\eta] \leqq N_r \leqq m_r) \geqq 1 - \varepsilon$ for large r, we have

$$\begin{aligned} &\sigma_{n_r c}^{-2} |\sigma_{N_r c}^2 - \sigma_{n_r c}| \\ &\quad \leqq \sigma_{n_r c}^{-2} |N_r^{-1} \textstyle\sum_{i=n_r+1}^{N_r} c_i^2 + (N_r^{-1} - n_r^{-1}) \sum_{i=1}^{n_r} c_i^2| \\ &\quad \leqq |(v_r^{-1} m_r (\sigma_{m_r c}/\sigma_{n_r c})^2 - n_r/v_r| + \eta \end{aligned}$$

with probability at least $1 - \varepsilon$, where m_r is one that appears above and $v_r = [n_r - n_r\eta]$. Now note that the right-hand side above can be made very small for large r and arbitrarily small η. With (2.5) and limiting normality of $\sigma_{n_r}^{-1} N_r^{\frac{1}{2}} S_{N_r}$ at hand it is easy to conclude (2.3). The proof is terminated.

REFERENCES

[1] Koul, H. L. (1969). Random rank statistics and confidence intervals. RM-234, Michigan State University.

[2] Fernandez, P. (1970). A weak convergence theorem for random sum of independent random variables. *Ann. Math. Statist.* **41** 710–712.

[3] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Test.* Academic Press, New York.

[4] Brieman, L. (1968). *Probability.* Addison-Wesley, Reading.

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

[6] Pyke, R. and Shorack, G. R. (1968). Weak convergence and a Chernoff–Savage theorem for random sample size. *Ann. Math. Statist.* **39** 1675–1685.