The Annals of Probability
1977, Vol. 5, No. 6, 926–939

A LOCAL LIMIT THEOREM FOR THE WILCOXON RANK SUM

BY DANIEL THORBURN

National Central Bureau of Statistics, Sweden

The normalized probability function of the Wilcoxon rank sum statistic is shown to converge to the normal density under very mild conditions on the two sample distributions. This is done by studying the conditional distribution of the rank sum given the first sample and by a rather heavy use of characteristic functions.

1. Introduction. Let $X = (X_1, \cdots, X_n)$ and $Y = (Y_1, \cdots, Y_m)$ be independent samples from the distibutions F_X and F_Y, respectively. Rank the variables in increasing order from 1 to $n + m$ and consider the rank sum W of the Y_i.

In 1945 Wilcoxon suggested this sum as a test statistic for the hypothesis $F_X = F_Y$ against $F_X \neq F_Y$. This test had been presented independently by many writers prior to Wilcoxon. Wald and Wolfowitz had the year before given a theorem that immediately proved the asymptotic normality of the rank sum under the null-hypothesis. The local limit version was proved under the null-hypothesis by Bennedichs (1973), who studied sampling from a finite population, and independently by Viźková (1973). We will prove the same result under quite general alternatives. The idea of our proof is inspired by the local limit theorem for sums of i.i.d. random variables and its proof given in the book by Ibragimov and Linnik (1971).

2. Notations and some lemmas. Since a monotone transformation does not change the rank sum, we may, without loss of generality, assume that the distribution of X_i is uniform on the unit interval, $F_X(x) = x$, and that the value of Y_i will lie in the same interval but with the distribution $F_Y = F$. This statement will be proved more rigorously in the proof of Theorem 3.3.

Let the expected value and the variance of Y_i be denoted by M and V_1. This means that

$$M = 1 - E(F(X_i)) = P(X_i \leqq Y_i) .$$

Further let V_2 and $D_{n,m}$ be defined by

$$V_2 = \text{Var}\,(F(X_i)) = \textstyle\int_0^1 F^2(x)\,dx - (1 - M)^2$$
$$D_{n,m} = (nm(nV_1 + mV_2))^{\frac{1}{2}} .$$

The number of pairs (X_i, Y_j) such that $X_i \leqq Y_j$ is a simple linear function of the rank sum

$$R_{n,m} = \#\{(i, j);\, X_i \leqq Y_j\} = W - m(m + 1)/2 . \tag{2.1}$$

Received May 10, 1976; revised March 7, 1977.
AMS 1970 *subject classifications*. 60F99, 62G99.
Key words and phrases. Local limit theorem, Mann–Whitney, rank sum, Wilcoxon.

In the sequel we will discuss under what conditions

$$\sup_k |D_{n,m}P(R_{n,m} = k) - \phi((k - nmM)/D_{n,m})| \to 0 ,$$

where ϕ is the standard normal density. The expected value of $R_{n,m}$ is nmM and its exact variance is (e.g., Noether 1967) $D^2_{n,m} + nm(M - M^2 - V_1 - V_2)$, which is of the same order as $D^2_{n,m}$.

Denoting the number of X_i that are less than or equal to Y_j by r_j, we have

$$\sum_1^m r_j = R_{n,m} . \tag{2.2}$$

For given X the rank sum $R_{n,m}$ is thus a sum of i.i.d. random variables with the distribution

$$P(r_j \leqq k \mid X) = F(X_{(k+1)}) , \tag{2.3}$$

where $X_{(k)}$ is the kth order statistic and where $F(X_{(n+1)}) = 1$. (The right-hand side of (2.3) ought to be $F(X_{(k+1)} - 0)$, but this equals $F(X_{(k+1)})$ a.s., since X is uniformly distributed.) In the sequel we will also use the conventions $F(X_{(0)}) = 0$, $X_{(0)} = 0$ and $X_{(n+1)} = 1$. The idea of this paper is to prove the ordinary local limit theorem for the sum of the variables r_j and then to remove the condition by taking the expected value.

To do so we need the following additional notations: for given X, let $F_n(\cdot)$, $\Psi_n(\cdot)$, a_n, b_n and c_n denote the distribution function, characteristic function, mean, variance and third central absolute moment of r_j/n, respectively. We emphasize that F_n and Ψ_n are random functions and that a_n, b_n and c_n are random variables. In the following lemmata we will study some of their properties.

LEMMA 2.1.

(a) $F_n(x) \to F(x)$ a.s. *as* $n \to \infty$ *at all continuity points* x *of* F.

(b) $n^{\frac{1}{2}}(a_n - M)/(V_2)^{\frac{1}{2}}$ *tends in law to the standard normal distribution as* $n \to \infty$.

(c) $b_n \to V_1$ a.s. *as* $n \to \infty$.

(d) $\Psi_n(t) \to \Psi(t)$ *uniformly on every compact* a.s. *as* $n \to \infty$, *where* Ψ *is the characteristic function of* Y_i.

(e) *The limits in* (a), (c) *and* (d) *hold in the mean too.*

PROOF. (a) It is well known from order statistics that $X_{([nx+1])} \to x$ a.s. as $n \to \infty$. Since $|F_n(x) - F(x)| = |F(X_{([nx+1])}) - F(x)|$, part (a) follows.

(b) Since a_n is the mean of a variable with distribution F_n, a step-function with jumps $F(X_{(i+1)}) - F(X_{(i)})$ at the points i/n, we can write

$$\begin{aligned} a_n &= \sum_{i=0}^n \left(\frac{i}{n}(F(X_{(i+1)}) - F(X_{(i)}))\right) \\ &= \sum_{i=1}^{n+1} \frac{i-1}{n} F(X_{(i)}) - \sum_{i=0}^n \frac{i}{n} F(X_{(i)}) \\ &= 1 - \frac{1}{n} \sum_{i=1}^n F(X_{(i)}) = 1 - \frac{1}{n} \sum_{i=1}^n F(X_i) . \end{aligned}$$

It is now expressed as a sum of bounded i.i.d. random variables with mean M and variance V_2. The result follows from the central limit theorem.

(c) and (d) are simple consequences of part (a). It is also clear that all finite moments of F_n will converge to the corresponding moments of F. They are, in fact, asymptotically normal with a variance of order $1/n$ (see David (1970) and the references given there).

(e) Since F_n, b_n and Ψ_n all are bounded, this follows from almost sure convergence. □

In the next section b_n will sometimes appear in the denominator. In Lemma 2.3 it is shown that b_n stays away from zero with a large probability. The idea behind all the technical details of the proof is that b_n is small when all the X_i are near zero or one. In the proof of Lemma 2.3 we need the following result, which is a special case of Theorem 1 in Hoeffding (1963).

LEMMA 2.2. *Let Z have a binomial distribution with parameters n and p. If $p > q$, then*

$$P(Z \leqq nq) \leqq \exp(-2n(p-q)^2) \leqq a^n ,$$

for some $a < 1$.

LEMMA 2.3. *If F has a positive density in some interval, there exist numbers $\alpha < 1$, $\beta > 0$ and n_0 such that*

$$P(b_n < \beta) \leqq \alpha^n \qquad \textit{for all} \quad n > n_0 .$$

PROOF. Find two numbers, γ and δ, such that

$$0 < F(\gamma) < F(\delta) < 1 \qquad \text{and} \qquad \delta - \gamma < \tfrac{1}{2} .$$

Let n_1 and n_2 denote the number of X-observations in $(0, \gamma]$ and $(\gamma, \delta]$, respectively. Break up b_n into three sums

$$\begin{aligned} b_n = {} & \textstyle\sum_0^{n_1} (F(X_{(i+1)}) - F(X_{(i)})) \left(\frac{i}{n} - a_n\right)^2 \\ & + \textstyle\sum_{n_1+1}^{n_1+n_2-1} (F(X_{(i+1)}) - F(X_{(i)})) \left(\frac{i}{n} - a_n\right)^2 \\ & + \textstyle\sum_{n_1+n_2}^{n} (F(X_{(i+1)}) - F(X_{(i)})) \left(\frac{i}{n} - a_n\right)^2 , \end{aligned}$$

where we have assumed that $n_2 > 0$. If $a_n \leqq (2n_1 + n_2)/2n$, the third of these sums is greater than $(1 - F(X_{(n_1+n_2)}))(n_2/2n)^2 \geqq (1 - F(\delta))(n_2/2n)^2$; otherwise the first sum is greater than $F(\gamma)(n_2/2n)^2$. These two implications give that

$$b_n \geqq \left(\frac{n_2}{n}\right)^2 \tfrac{1}{4} \min\{F(\gamma), 1 - F(\delta)\} . \tag{2.4}$$

This formula holds trivially also when $n_2 = 0$. We thus have that

$$P(b_n < \beta) \leqq P(n_2 < 2n(\beta/\min\{F(\gamma), 1 - F(\delta)\})^{\frac{1}{2}}) .$$

The result now follows from Lemma 2.2. □

3. Main theorem. We first prove a local limit theorem in the situation described in the previous section, i.e., when the distribution of X_i is uniform.

THEOREM 3.1. *If Y in some interval has a density which is bounded away from zero, then*

$$\sup_k |D_{n,m}P(R_{n,m}=k)-\phi((k-nmM)/D_{n,m})|\to 0$$

as n and m tend to infinity so that $\liminf n/m>0$ *and* $\limsup n^\kappa/m=0$ *for some* $\kappa>0$.

PROOF. The proof consists of four parts. The first one contains the whole proof, except for the estimation of some integrals. These estimations are made in the remaining parts.

(i) *Main part.* For a given X, the statistic $R_{n,m}$ is distributed as a sum of m i.i.d. random variables, r_j. The inversion theorem for characteristic functions and a change of variables give

$$\begin{aligned} n(mb_n)^{\frac12}P(R_{n,m}=k\mid X) &= \frac{n(mb_n)^{\frac12}}{2\pi}\int_{-\pi}^{\pi} e^{-itk}(\Psi_n(tn))^m\,dt \\ &= \frac{1}{2\pi}\int_{-\pi n(mb_n)^{\frac12}}^{\pi n(mb_n)^{\frac12}} \exp\left(-it\frac{k-nma_n}{n(mb_n)^{\frac12}}\right)\exp\left(-it\frac{a_n m^{\frac12}}{(b_n)^{\frac12}}\right) \\ &\quad\times\left(\Psi_n\left(\frac{t}{(mb_n)^{\frac12}}\right)\right)^m dt\,. \end{aligned}$$

For the normal distribution there exists a similar inversion formula:

$$\phi\left(\frac{k-nma_n}{n(mb_n)^{\frac12}}\right)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(-it\frac{k-nma_n}{n(mb_n)^{\frac12}}\right)\exp(-t^2/2)\,dt\,.$$

The absolute difference between these two expressions can be estimated in the following way:

$$\begin{aligned} \sup_k&\left|n(mb_n)^{\frac12}P(R_{n,m}=k\mid X)-\phi\left(\frac{k-nma_n}{n(mb_n)^{\frac12}}\right)\right| \\ &\leqq \frac{1}{2\pi}\sup_k\left\{\int_{-\pi n(mb_n)^{\frac12}}^{\pi n(mb_n)^{\frac12}}\left|\exp\left(-it\left(\frac{k-nma_n}{n(mb_n)^{\frac12}}\right)\right)\right|\cdot\left|\exp\left(-it\left(\frac{a_n m^{\frac12}}{(b_n)^{\frac12}}\right)\right)\right.\right. \\ &\quad\times\left.\left(\Psi_n\left(\frac{t}{(mb_n)^{\frac12}}\right)\right)^m-\exp(-t^2/2)\right|dt \\ &\quad+\left.\int_{|t|>\pi n(mb_n)^{\frac12}}\left|\exp\left(-it\left(\frac{k-nma_n}{n(mb_n)^{\frac12}}\right)\right)\right|\exp(-t^2/2)\,dt\right\} \\ &\leqq \frac{1}{2\pi}\int_{-m^{0.5-\delta}}^{m^{0.5-\delta}}\left|\exp\left(-it\left(\frac{a_n m^{\frac12}}{(b_n)^{\frac12}}\right)\right)\left(\Psi_n\left(\frac{t}{(mb_n)^{\frac12}}\right)\right)^m-\exp(-t^2/2)\right|dt \\ &\quad+\frac{1}{2\pi}\int_{m^{0.5-\delta}<|t|<\pi n(mb_n)^{\frac12}}\left|\Psi_n\left(\frac{t}{(mb_n)^{\frac12}}\right)\right|^m dt+\frac{1}{2\pi}\int_{|t|\geqq m^{0.5-\delta}}\exp(-t^2/2)\,dt \\ &= I_1+I_2+I_3\,,\quad\text{say.} \end{aligned}$$

In this expression it is assumed that $m^{0.5-\delta} < \pi n(mb_n)^{\frac{1}{2}}$. This will hold for m and n sufficiently large. In order to make the formula hold for all m and n we can define $I_2 = 0$, if $m^{0.5-\delta} \geqq \pi n(mb_n)^{\frac{1}{2}}$. In the sequel δ will be assumed to lie strictly between 0 and $\frac{1}{6}$.

In (iii) and (iv) below we will show that the expected values of I_1 and I_2 tend to zero as n and m tend to infinity. It is trivial that this holds for I_3. If we accept these facts, we have proved that

$$E\left\{\sup_k \left| n(mb_n)^{\frac{1}{2}} P(R_{n,m} = k \mid X) - \phi\left(\frac{k - nma_n}{n(mb_n)^{\frac{1}{2}}}\right)\right|\right\} \to 0 \qquad \text{as} \quad n, m \to \infty .$$

Changing the order between the expected and the absolute value, we get

$$\sup_k \left| E\{n(mb_n)^{\frac{1}{2}} P(R_{n,m} = k \mid X)\} - E\left\{\phi\left(\frac{k - nma_n}{n(mb_n)^{\frac{1}{2}}}\right)\right\}\right| \to 0 \qquad \text{as} \quad n, m \to \infty .$$

The next step of the proof is to replace b_n by V_1. By Lemma 2.3 b_n is larger than some number $\beta > 0$ except with a probability α^n, where $\alpha < 1$. This tends to zero faster than $nm^{\frac{1}{2}}$. Thus we can assume that b_n is larger than some number β except with a negligible probability. Since b_n is also bounded by 1 and tends to V_1 a.s. by Lemma 2.1 c, we can replace b_n by V_1 in the first term. For the second term we observe that

$$\sup_k \left| E\left\{\phi\left(\frac{k - nma_n}{n(mb_n)^{\frac{1}{2}}}\right) - \phi\left(\frac{k - nma_n}{n(mV_1)^{\frac{1}{2}}}\right)\right\}\right|$$
$$\leqq \sup_k \left| E\left\{\phi'\left(\frac{k - nma_n}{n(m\theta)^{\frac{1}{2}}}\right)\left(\frac{k - nma_n}{n(m\theta)^{\frac{1}{2}}}\right)\frac{1}{2\theta}(b_n - V_1)\right\}\right| ,$$

where θ is a number between b_n and V_1. It is well known that $\phi'(x)x$ is bounded. Since we have assumed that $\theta > \beta$, we get from Lemma 2.1 c that the whole expression tends to zero. We thus have that

$$\sup_k \left| E\{n(mV_1)^{\frac{1}{2}} P(R_{n,m} = k \mid X)\} - E\left\{\phi\left(\frac{k - nma_n}{n(mV_1)^{\frac{1}{2}}}\right)\right\}\right| \to 0 \qquad \text{as} \quad n, m \to \infty .$$

The expected value of the first term is $n(mV_1)^{\frac{1}{2}} P(R_{n,m} = k)$ and we get

$$\sup_k \left| n(mV_1)^{\frac{1}{2}} P(R_{n,m} = k) - E\left(\phi\left(\frac{k - nma_n}{n(mV_1)^{\frac{1}{2}}}\right)\right)\right| \to 0 \qquad \text{as} \quad n, m \to \infty .$$

We know by Lemma 2.1 b that a_n has an asymptotically normal distribution. If a_n had been exactly normal with mean E and variance V_2/n, the following would have been true:

$$E\left(\phi\left(\frac{k - nma_n}{n(mV_1)^{\frac{1}{2}}}\right)\right) = \frac{n(mV_1)^{\frac{1}{2}}}{D_{n,m}}\phi\left(\frac{k - nmM}{D_{n,m}}\right). \tag{3.1}$$

In (ii) below we will show that the difference between the two sides tends to zero as n and m tend to infinity. Accepting this, we get

$$\sup_k \left| n(mV_1)^{\frac{1}{2}} P(R_{n,m} = k) - \frac{n(mV_1)^{\frac{1}{2}}}{D_{n,m}} \phi\left(\frac{k - nmM}{D_{n,m}}\right)\right| \to 0 \quad \text{as} \quad n, m \to \infty .$$

Since F is not degenerate, $V_1 \neq 0$. We can now easily get the result of the theorem.

(ii) *Estimation of* $E\{\phi((k - nma_n)(n^{-1}(mV_1)^{-\frac{1}{2}}))\}$. Let a_n' be a normal variable with mean M and variance V_2/n, such that a_n' increases whenever a_n increases. If a_n has a continuous distribution function, G, say, such a variable may be constructed by $a_n' = M + (V_2/n)^{\frac{1}{2}}\Phi^{-1}(G(a_n))$. If G is not continuous a similar construction can be made. (Φ denotes the standard normal distribution function.)

By the Berry-Esseen theorem we have that both

$$|\Phi((a_n - M)(V_2/n)^{-\frac{1}{2}}) - \Phi((a_n' - M)(V_2/n)^{-\frac{1}{2}})| \quad \text{and}$$
$$|\Phi((a_n - M)(V_2/n)^{-\frac{1}{2}}) - G(a_n)|$$

are less than or equal to $cn^{-\frac{1}{2}}$, where c is a constant that may depend on F.

Suppose that both a_n and a_n' belong to the interval

$$M \pm \Phi^{-1}(1 - n^{-\lambda})(V_2/n)^{\frac{1}{2}} \quad \text{where} \quad 0 < \lambda < \tfrac{1}{2} .$$

By the mean value theorem of integral calculus, we then have

$$|a_n - a_n'| \leqq c(V_2)^{\frac{1}{2}}/(n\phi(\Phi^{-1}(1 - n^{-\lambda}))) .$$

The probability of a_n or a_n' outside the interval is at most

$$P\{|(a_n' - M)(V_2/n)^{-\frac{1}{2}}| > \Phi^{-1}(1 - n^{-\lambda})\} + P\{|(a_n - M)(V_2/n)^{-\frac{1}{2}}| > \Phi^{-1}(1 - n^{-\lambda})\}$$
$$\leqq 4n^{-\lambda} + 2cn^{-\frac{1}{2}} .$$

Since (3.1) holds for a_n', we have

$$\left|E\left(\phi\left(\frac{k - nma_n}{n(mV_1)^{\frac{1}{2}}}\right)\right) - \phi\left(\frac{k - nmM}{D_{n,m}}\right)\frac{n(mV_1)^{\frac{1}{2}}}{D_{n,m}}\right|$$
$$= \left|E\left\{\phi\left(\frac{k - nma_n}{n(mV_1)^{\frac{1}{2}}}\right) - \phi\left(\frac{k - nma_n'}{n(mV_1)^{\frac{1}{2}}}\right)\right\}\right|$$
$$\leqq (m/V_1)^{\frac{1}{2}}c(V_2)^{\frac{1}{2}}/(n\phi(\Phi^{-1}(1 - n^{-\lambda}))) + 4n^{-\lambda} + 2cn^{-\frac{1}{2}}$$
$$\leqq c(V_2/V_1)^{\frac{1}{2}}m^{\frac{1}{2}}n^{-1+\lambda} + 4n^{-\lambda} + 2cn^{-\frac{1}{2}} .$$

This tends to zero as n and m tend to infinity in the prescribed way.

(iii) *Estimation of* $E(I_1)$. In order to estimate

$$I_1 = \frac{1}{2\pi}\int_{-m^{0.5-\delta}}^{m^{0.5-\delta}} |\exp(-it(a_n m^{\frac{1}{2}}(b_n)^{-\frac{1}{2}}))(\Psi_n(t(mb_n)^{-\frac{1}{2}}))^m - \exp(-t^2/2)|\, dt ,$$

we use the following lemma given in Chung (1968, page 210). It is given in our notation; e.g., c_n is the third absolute central moment of r_j/n.

LEMMA 3.2. *If*

$$|t| \leqq m^{\frac{1}{2}}(b_n)^{\frac{3}{2}}/(4c_n)\,, \tag{3.2}$$

then

$$|\exp(-it((a_n m^{\frac{1}{2}})(b_n)^{-\frac{1}{2}}))(\Psi_n(t(mb_n)^{-\frac{1}{2}}))^m - \exp(-t^2/2)| \leqq \frac{16|t|^3 \exp(-t^2/3)c_n}{m^{\frac{1}{2}}(b_n)^{\frac{3}{2}}}\,.$$

This lemma yields an estimate for such outcomes that (3.2) holds in the range of integration. The probability of this is

$$\begin{aligned} P(|t| \leqq m^{\frac{1}{2}}(b_n)^{\frac{3}{2}}/(4c_n) \ \text{ for all } \ |t| < m^{0.5-\delta}) &= P(m^{0.5-\delta} \leqq m^{\frac{1}{2}}(b_n)^{\frac{3}{2}}/(4c_n)) \\ &\geqq P(b_n \geqq (4m^{-\delta})^{\frac{2}{3}}) \\ &\geqq 1 - \alpha^n\,, \end{aligned}$$

where α is given by Lemma 2.3, if m and n are large enough. We have used that c_n is less than 1.

If (3.2) holds we estimate I_1 as follows:

$$\begin{aligned} 2\pi I_1 &\leqq \int_{-m^{0.5-\delta}}^{m^{0.5-\delta}} \frac{16|t|^3 \exp(-t^2/3)c_n}{m^{\frac{1}{2}}(b_n)^{\frac{3}{2}}}\,dt \\ &\leqq \int_{-\infty}^{\infty} 16|t|^3 \exp(-t^2/3) m^{\delta-0.5}/4\,dt = 36m^{\delta-0.5}\,. \end{aligned}$$

We have here used (3.2) in order to replace $c_n/(m^{\frac{1}{2}}(b_n)^{\frac{3}{2}})$ by $m^{\delta-0.5}/4$. The expected value of I_1 is thus less than

$$\frac{1}{2\pi}(36m^{\delta-0.5} + 4\alpha^n m^{0.5-\delta})\,,$$

which tends to zero as n and m tend to infinity in the prescribed way.

(iv) *Estimation of* $E(I_2)$. The integral

$$I_2 = \frac{1}{2\pi}\int_{m^{0.5-\delta}<|t|<\pi n(mb_n)^{\frac{1}{2}}} |\Psi_n(t(mb_n)^{-\frac{1}{2}})|^m\,dt$$

is harder to estimate. The main idea is to show that the supremum of $|\Psi_n(t)|$ is less than $1 - m^{-\frac{1}{2}}$, except with a negligible probability $P_{n,m}$, say.

$$\begin{aligned} P_{n,m} &= P(\sup_{m^{-\delta}<t<\pi n} |\Psi_n(t)| \geqq 1 - m^{-\frac{1}{2}}) \\ &= P(\sup_{m^{-\delta}<t<\pi n} \sup_{z\in[0,2\pi)} \operatorname{Re}(\exp(iz)\Psi_n(t)) \geqq 1 - m^{-\frac{1}{2}})\,. \end{aligned} \tag{3.3}$$

Let Z_m be defined by

$$Z_m = \{km^{-\frac{1}{4}};\ \ k \ \text{ integer}\} \cap [0, 2\pi)\,.$$

If the supremum in (3.3) over all $z \in [0, 2\pi)$ is larger than $1 - m^{-\frac{1}{2}}$, then the supremum over Z_m is larger than $\cos(m^{-\frac{1}{4}}/2)(1 - m^{-\frac{1}{2}}) \geqq 1 - 2m^{-\frac{1}{2}}$. This gives that

$$P_{n,m} \leqq P(\sup_{m^{-\delta}<t<\pi n} \sup_{z\in Z_m} \operatorname{Re}(\exp(iz)\Psi_n(t)) \geqq 1 - 2m^{-\frac{1}{2}})\,.$$

Using the inequality $P(\bigcup_1^n A_k) \leqq \sum_1^n P(A_k)$ on the events

$$A_k = \{\sup_t \operatorname{Re}(\exp(ikm^{-\frac{1}{4}})\Psi_n(t)) \geqq 1 - 2m^{-\frac{1}{2}}\},$$

we get

$$\begin{aligned} P_{n,m} &\leqq 2\pi m^{\frac{1}{4}} \sup_z P(\sup_{m^{-\delta}<t<\pi n} \operatorname{Re}(\exp(iz)\Psi_n(t)) \geqq 1 - 2m^{-\frac{1}{2}}) \\ &= 2\pi m^{\frac{1}{4}} \sup_z P(\sup_{m^{-\delta}<t<\pi n} \operatorname{Re}\{\textstyle\sum_{j=0}^n \exp(i(z + tj/n)) \\ &\qquad \times (F(X_{(j+1)}) - F(X_{(j)}))\} \geqq 1 - 2m^{-\frac{1}{2}}) . \end{aligned}$$

Arguing in the same way for t as we did for z, we take the $\sup_t$ outside the probability sign:

$$\begin{aligned} (3.4) \qquad P_{n,m} &\leqq 2\pi^2 m^{\frac{1}{2}} n \sup_z \sup_{m^{-\delta}<t<\pi n} P(\textstyle\sum_{j=0}^n \operatorname{Re}(\exp(i(z + tj/n)) \\ &\qquad \times (F(X_{(j+1)}) - F(X_{(j)}))) \geqq 1 - 4m^{-\frac{1}{2}}) . \end{aligned}$$

We denote the event $\{j; \operatorname{Re}(\exp(i(z + tj/n))) > 1 - m^{-0.5+\delta}\}$ by $A_{z,t}$. If

$$\textstyle\sum_{j\in A_{z,t}} F(X_{(j+1)}) - F(X_{(j)}) < 1 - 4m^{-\delta},$$

then the sum in (3.4) is less than

$$(1 - m^{-0.5+\delta})4m^{-\delta} + 1(1 - 4m^{-\delta}) = 1 - 4m^{-\frac{1}{2}} .$$

Using this we have

$$\begin{aligned} (3.5) \qquad P_{n,m} &\leqq 2\pi^2 m^{\frac{1}{2}} n \sup_z \sup_{m^{-\delta}<t<\pi n} P(\textstyle\sum_{j\in A_{z,t}} F(X_{(j+1)}) \\ &\qquad - F(X_{(j)}) \geqq 1 - 4m^{-\delta}) . \end{aligned}$$

According to the assumptions there exists an interval, (a, b) say, such that Y has a density, which is bounded away from zero, in this interval. Define

$$\begin{aligned} G(x) &= 0 && \text{if } x \leqq a \\ &= 1 && \text{if } x \geqq b \\ &= \frac{F(x) - F(a)}{F(b) - F(a)} && \text{if } a < x < b . \end{aligned}$$

The density of G has a positive infimum, in the interval (a, b). If all the terms, where $X_{(j)}$ does not belong to (a, b), are included in the sum in (3.5) the probability will be larger:

$$P_{n,m} \leqq 2\pi^2 m^{\frac{1}{2}} n \sup_z \sup_{m^{-\delta}<t<\pi n} P\left(\textstyle\sum_{j\in A_{z,t}} G(X_{(j+1)}) - G(X_{(j)}) \geqq 1 - \frac{4m^{-\delta}}{F(b) - F(a)}\right) .$$

We shall now estimate the number of terms in the sum. The set $\{i; \operatorname{Re}(\exp(iz)) \geqq 1 - m^{-0.5+\delta}\}$ corresponds to an arc of the unit circle, which is shorter than $4m^{-0.25+\delta/2}$. A shift of j to $j + 1$ in $\operatorname{Re}(\exp(i(z + tj/n)))$ corresponds to the arc $t/n > m^{-\delta}/n$. If there are $n(b - a)/2$ consecutive j-values, this corresponds to an arc that is longer than $(b - a)m^{-\delta}/2$. If m is large enough, it is easy to see that

$$4m^{-0.25+\delta} \leqq (b-a)m^{-\delta}/2 .$$

If there are $n_2 > n(b-a)/2$ observations in (a, b) and if m is large enough, we have thus proved that there are at most $2n_2/3$ $X_{(j)}$-values in (a, b), such that $j \in A_{z,t}$. Which observations that are included does not depend on their values (except that they must lie in the interval). Suppose that they are $X_{(j_1)}, X_{(j_2)}, \cdots$. We find that

$$P_{n,m} \leqq 2\pi^2 m^{\frac{1}{2}} n \Big(P(n_2 \leqq n(b-a)/2) + \sup_z \sup_{m^{-\delta}<t<\pi n} P\Big(\sum_k (G(X_{(j_k+1)}) - G(X_{(j_k)})) \geqq 1 - \frac{4m^{-\delta}}{F(b)-F(a)} \Big| n_2 > n(b-a)/2 \Big) \Big) .$$

The first of these probabilities is by Lemma 2.2 less than β^n for some $\beta < 1$. The other one can be estimated as follows:

$$\text{(3.6)} \quad P\Big(\sum_k (G(X_{(j_k+1)}) - G(X_{(j_k)})) \geqq 1 - \frac{4m^{-\delta}}{F(b)-F(a)} \Big| n_2 > n(b-a)/2 \Big) \leqq P\Big(d \sum_{j \notin \{j_k\}} (X_{(j+1)} - X_{(j)}) \leqq \frac{4m^{-\delta}}{F(b)-F(a)} \Big| n_2 > n(b-a)/2 \Big) ,$$

where d is the positive infimum of the density of G in (a, b). We found above that there are at least $n_2/3$ terms in the sum. The differences $X_{(j+1)} - X_{(j)}$ have an exchangeable distribution for $j = 0, 1, \cdots$. We can thus replace those in the sum by those with the smallest j-values without changing the probability. Combining these facts, we find that (3.6) is at most

$$P\Big(d \sum_0^{[n_2/3]} (X_{(j+1)} - X_{(j)}) \leqq \frac{4m^{-\delta}}{F(b)-F(a)} \Big| n_2 > n(b-a)/2 \Big) = P\Big(dX_{([n_2/3])} \leqq \frac{4m^{-\delta}}{F(b)-F(a)} \Big| n_2 > n(b-a)/2 \Big) ,$$

where n_2 has changed its meaning after the reordering and now stands for the number of observations less than $b - a$. By Lemma 2.2 this expression is less than γ^n for some positive $\gamma < 1$ (an intermediary result is $(\gamma_1)^{n_2}$, but it is assumed that $n_2 > n(b-a)/2$). This does not depend on z or t, so the bound is uniform. We thus have

$$\text{(3.7)} \quad P_{n,m} \leqq 2\pi^2 m^{\frac{1}{2}} n(\beta^n + \gamma^n) .$$

We can now estimate the expected value of I_2:

$$E(I_2) \leqq (1/2\pi) 2\pi n (mb_n)^{\frac{1}{2}} ((1 - m^{-\frac{1}{2}})^m + 2\pi^2 m^{\frac{1}{2}} n(\beta^n + \gamma^n)) \leqq nm^{\frac{1}{2}}(1 - m^{-\frac{1}{2}})^m + 2\pi^2 mn^2(\beta^n + \gamma^n) \to 0$$

as n and m tend to infinity in the prescribed way.

The proof of Theorem 3.1 is now complete. □

In Theorem 3.1 we assumed that $F_X(x) = x$. Making a monotone transformation we get the following slightly more general theorem.

THEOREM 3.3. *Let X and Y be two independent samples of* i.i.d. *random variables with sizes n and m, respectively and let their distribution functions, F_X and F_Y, be absolutely continuous with positive and bounded densities on some interval. Then*

$$\sup_k |D_{n,m} P(R_{n,m} = k) - \phi((k - nmM)/D_{n,m})| \to 0$$

as n and m tend to infinity not faster than powers of each other, i.e., *so that there exists a $\kappa > 0$ with* $\limsup n^\kappa/m = \limsup m^\kappa/n = 0$. *Here $M = P(X \leqq Y)$ and* $D^2_{n,m} = nm(nV(F_X(Y)) + mV(F_Y(X)))$.

PROOF. First suppose that $\liminf n/m > 0$.

Construct two variables ξ and η, which are uniformly distributed on the unit interval. Define

$$F_X^{-1}(a) = \inf \{x; F_X(x) \geqq a\}$$
$$F_Y^{-1}(a) = \inf \{y; F_Y(y) \geqq a\} .$$

It is easy to see that $F_X^{-1}(\xi)$ and $F_Y^{-1}(\eta)$ have the distribution functions F_X and F_Y, respectively. It is also easy to see that the following two statements are equivalent:

$$F_X^{-1}(\xi) \leqq F_Y^{-1}(\eta) \qquad \text{and} \qquad \xi \leqq F_X(F_Y^{-1}(\eta)) . \tag{3.8}$$

If we can prove the theorem for the distribution functions x and $F_X(F_Y^{-1}(x))$, we can also prove it for the distribution functions $F_X(x)$ and $F_Y(x)$, since by (3.8) the rank sum is the same. The result for x and $F_X(F_Y^{-1}(x))$ follows, however, directly from Theorem 3.1.

When $\liminf m/n > 0$, the result follows from the previous paragraph, if we multiply the two samples by minus one and let them change places.

Any sequence of (m, n) can be divided into two subsequences, where $n \leqq m$ in one and $n > m$ in the other. Since the theorem holds for both subsequences according to the previous paragraphs, it must also hold for the original sequence. □

In the situation of Theorem 3.3 there may appear ties. The assumptions say only that there exists one interval, where there a.s. cannot appear ties, but where the distributions are continuous. That the local limit theorem holds in this situation is largely due to the definition (2.1) of the rank sum. In Section 5 we shall discuss this further. In the next section we shall assume that both the distributions are continuous and this means, of course, that there cannot be any ties.

4. Modifications of the theorem. The assumptions made in Theorems 3.1 and 3.3, concerning the way in which n and m tend to infinity, are sometimes unnecessary. We give here a theorem in which this assumption is omitted, but instead the class of distributions is restricted. Before stating the theorem we give and prove a lemma.

LEMMA 4.1. *Let F have a density f such that*

$$f(z) < C(z^k + (1-z)^k) \quad \text{for some} \quad k > -\tfrac{1}{2} \quad \text{and some constant} \quad C .$$

Then there exists a constant Q independent of n such that

$$E(n \textstyle\sum_0^n (F(X_{(j+1)}) - F(X_{(j)}))^2) \leqq Q .$$

PROOF. First we study the sum without the two end terms.

$$\begin{aligned} E(n \textstyle\sum_{j=1}^{n-1} & (F(X_{(j+1)}) - F(X_{(j)}))^2) \\ &= n \textstyle\sum_{j=1}^{n-1} \iint_{1>x_{j+1}>x_j>0} (F(x_{j+1}) - F(x_j))^2 \\ &\quad \times \frac{n!\, x_j^{j-1}(1-x_{j+1})^{n-j-1}}{(j-1)!\,(n-j-1)!} \, dx_j \, dx_{j+1} \\ &= n \textstyle\sum_{j=1}^{n-1} \iint_{0<x<1; 0<y<1-x} (F(x+y) - F(x))^2 \\ &\quad \times \frac{n!}{(j-1)!\,(n-j-1)!} x^{j-1}(1-x-y)^{n-j-1} \, dx \, dy \\ &= n^2(n-1) \textstyle\iint_{0<x<1; 0<y<1-x} (F(x+y) - F(x))^2 (1-y)^{n-2} \, dy \, dx . \end{aligned} \tag{4.1}$$

First we have changed the order of summation and integration (the density can be found in David (1970), page 9). The second equality is a change of variables and the third follows from the binomial theorem.

If k is positive, it is easy to see that (4.1) is bounded, otherwise we can estimate $F(x+y) - F(x)$ using the mean value theorem. The expression (4.1) is less than

$$\begin{aligned} n^2(n-1) \textstyle\int_0^1 \int_0^1 & C^2 y^2 (x^k + (1-x)^k)^2 (1-y)^{n-2} \, dy \, dx \\ &\leqq 4C^2 n/(n+1) \textstyle\int_0^1 (x^{2k} + (1-x)^{2k}) \, dx \\ &= 8nC^2/((n+1)(1+2k)) . \end{aligned}$$

It remains to show that the end terms of the sum are bounded. This, however, is quite simple and is left to the reader. □

THEOREM 4.2. *Let X_i be uniformly distributed on the interval* $(0, 1)$ *and let Y_i have the distribution function F, such that* $F(0) = 1 - F(1) = 0$. *If F has a density f, such that $f(z) < C(z^k + (1-z)^k)$ for some $k > -\frac{1}{2}$ and some constant C, then*

$$\sup_k |D_{n,m} P(R_{n,m} = k) - \phi((k - nmM)/D_{n,m})| \to 0$$

as n and m tend to infinity, so that $\liminf n/m > 0$.

PROOF. We will here only pick out those parts of the proof of Theorem 3.1 where changes have to be made.

The assumption $\limsup n^\kappa/m = 0$ for some $\kappa > 0$ was used only in estimating $E(I_2)$ in part (iv). Here we use Parseval's formula when estimating that integral:

$$\begin{aligned} E(2\pi I_2) &= E(\textstyle\int_{m^{0.5-\delta}<|t|<\pi n(mb_n)^{\frac{1}{2}}} |\Psi_n(t(mb_n)^{-\frac{1}{2}})|^m \, dt) \\ &\leqq E(\sup_{m^{-\delta}<t<\pi n} |\Psi_n(t)|^{m-2} \textstyle\int_{-\pi n}^{\pi n} |\Psi_n(t)|^2 (mb_n)^{\frac{1}{2}} \, dt) \\ &= E(\sup_{m^{-\delta}<t<\pi n} |\Psi_n(t)|^{m-2} 2\pi n \textstyle\sum_{j=0}^n (F(X_{(j+1)}) - F(X_{(j)}))^2 (mb_n)^{\frac{1}{2}}) . \end{aligned} \tag{4.2}$$

As in the proof of Theorem 3.1 (formula 3.7) we have that

$$P_{n,m} \leqq 2\pi^2 m^{\frac{1}{2}} n(\beta^n + \gamma^n) ,$$

where $\beta < 1$ and $\gamma < 1$. If n and m are sufficiently large, we see that $P_{n,m} < \frac{1}{2}$. With this result in mind, we can easily from Lemma 4.1 deduce that

$$E(n \sum_0^n (F(X_{(j+1)}) - F(X_{(j)}))^2 \,|\, \sup_{\pi-\delta<t<\pi n} |\Psi_n(t)| \leqq 1 - m^{-\frac{1}{2}}) \leqq 2Q .$$

If we use the last two formulas in (4.2) and note that $b_n \leqq 1$, we get that

$$E(2\pi I_2) \leqq (1 - m^{-\frac{1}{2}})^{m-2} 4\pi m^{\frac{1}{2}} Q + 2\pi^2 m^{\frac{1}{2}} n(\beta^n + \gamma^n) 2\pi n m^{\frac{1}{2}} .$$

This expression tends to zero, when n and m tend to infinity so that $\liminf n/m > 0$.

The rest of the proof of Theorem 3.1 is not affected by the change in the conditions. Theorem 4.2 is thus proved. □

Theorem 4.2 is given in this unsymmetric form, since the class of allowable distributions will be even more restricted, when the conditions on n and m are totally removed. Its symmetric version, which is given below, is proved exactly as Theorem 3.3.

THEOREM 4.3. *Let F_X and F_Y be absolutely continuous with respect to each other and to Lebesgue measure. Let $F_X \circ F_Y^{-1}$ and $F_Y \circ F_X^{-1}$ be differentiable with derivatives f and g, respectively, such that for some $k > -\frac{1}{2}$ and some constant C*

$$\max(f(z), g(z)) \leqq C(z^k + (1 - z)^k) .$$

Then as $n, m \to \infty$,

$$\sup_k |D_{n,m} P(R_{n,m} = k) - \phi((k - nmM)/D_{n,m})| \to 0 .$$

5. Comments on the theorem. In this section we give some examples, where the local limit theorem is false.

EXAMPLE 5.1. In the previous theorems the probability of ties was zero. The usual definition of the rank sum, when dealing with ties, is

$$R'_{n,m} = \#\{(i, j); X_i < Y_j\} + \tfrac{1}{2}\#\{(i, j); X_i = Y_j\} . \tag{5.1}$$

If the distributions are such that ties can appear at one and only one value with positive probability, then a local limit theorem cannot hold.

Let us for instance define

$$\begin{aligned} F_X(t) = F_Y(t) &= 1 - \exp(-t)/2 \qquad &&\text{if } t \geqq 0 \\ &= 0 &&\text{if } t < 0 . \end{aligned}$$

It is quite easy to see that

$$\begin{aligned} &P(R'_{n,m} \text{ is an integer}) \\ &\quad = P(\#\{i; X_i = 0\} \text{ is even or } \#\{j; Y_j = 0\} \text{ is even}) \\ &\quad \to 1 - 0.5 \cdot 0.5 \qquad \text{as } n \text{ and } m \text{ tend to infinity.} \end{aligned}$$

Thus a local limit theorem cannot hold in this particular example. If it did the probability of an integer rank sum would tend to one-half.

Whenever there are a finite but positive number of values at which ties can appear, this contradiction exists. In fact it can be shown that

$$P(R'_{n,m} \text{ is an integer}) \to \tfrac{1}{2} + 1/2^{k+1}, \tag{5.2}$$

where k is the number of possible tie values, if the distributions of X and Y have positive densities at some interval. To show this, first prove that the probability of an integer rank sum at one fixed tie value is $\frac{3}{4}$. Then show that the events of integer rank sums at different tie values are asymptotically independent. The number of tie values with integer rank sums will thus follow the binomial law. A summation of every second term in the binomial distribution gives (5.2), which is easy to show by induction.

In our specific example it can be proved that

$$\sup_k |D_{n,m}P(R'_{n,m} = k) - \tfrac{3}{4}\phi((k - nmM)/D_{n,m})| \to 0$$

and

$$\sup_k |D_{n,m}P(R'_{n,m} = k + \tfrac{1}{2}) - \tfrac{1}{4}\phi((k + \tfrac{1}{2} - nmM)/D_{n,m})| \to 0$$

as n and m tend to infinity. The conditions and notations are the same as in Theorem 3.3. The proof follows the same lines as Section 3, the only change being that the cases with odd and even number of X_i at the possible tie value have to be separated.

EXAMPLE 5.2. The local limit theorem does not hold when the two distributions have disjoint supports, i.e., when the two measures are orthogonal.

Let X have a uniform distribution on $(0.25, 0.5) \cup (0.75, 1)$ and Y on $(0, 0.25) \cup (0.5, 0.75)$. After simple calculations we find that

$$\begin{aligned} &P(R_{n,m} \text{ is an even integer}) \\ &\quad = 1 - P(\#\{i;\, X_i \in (0.25, 0.5)\} \text{ is odd}) \cdot P(\#\{j;\, Y_j \in (0.5, 0.75)\} \text{ is odd}) \\ &\quad \to 1 - 0.5 \cdot 0.5 \qquad \text{as} \quad n, m \to \infty\,. \end{aligned}$$

If a local limit theorem held this limit ought to be 0.5.

This idea can be extended to more than two intervals. If the supports can be divided into $2k$ alternating disjoint intervals then

$$P(R_{n,m} \text{ is even}) \to \tfrac{1}{2} + 1/2^{k+1}, \tag{5.3}$$

where we have assumed that the largest interval is an X-interval. There is nothing peculiar about odd and even in this case. We could as well have studied the probability of $R_{n,m}$ being a multiple of a fixed prime number.

EXAMPLE 5.3. Let us now consider the very common situation, where the observations are measured with a certain precision (e.g., two figures after the decimal point). Let us also suppose that we know that all observations lie in a certain interval (e.g., we have a bounded scale on an instrument). In this situation the local limit theorem can neither hold when the rank sum is defined as in (2.1), nor when it is defined as in (5.1). This can be shown as (5.3) or almost as (5.2). If, however, $R_{n,m}$ is defined with a randomizing device in order to break the ties, the theorem will of course continue to hold.

Example 5.3 holds true for any discrete random variable, which can take only a finite number of values. If the range is infinite, none of the theorems in this paper is applicable, but I guess that a local limit theorem holds also in that situation, regardless of the definition of the rank sum.

6. Possible generations. The results of this paper cannot easily be generalized to cover other rank statistics. For our proof it is essential that $R_{n,m}$ can be written as a sum of random variables, which are independent given one of the samples. A general rank statistic cannot be written in that way. It is, however, possible to treat statistics of the following type:

$$Q_{n,m} = \sum_1^m g_n(r_j) ,$$

where g_n is a function that may depend on n but not on X_i.

It is often of interest to know something about the rate of convergence. Our proof behaves badly in that respect. It only says that the difference in Theorem 3.1 is at most of order $n^{-\frac{1}{4}}$ if n and m are of the same size. This can be pushed down by using better estimates in parts (ii) and (iii). It will, however, never reach $n^{-\frac{1}{2}}$ since we have not used the exact variance in the limiting distribution.

REFERENCES

BENNEDICHS, M. (1973). Local central limit theorem for sampling without replacement. *Trit. Mat.* 14, Dept. of Mathematics, KTH, Stockholm.

CHUNG, K. L. (1968). *A Course in Probability.* Harcourt, Brace and World, New York.

DAVID, H. A. (1970). *Order Statistics.* Wiley, New York.

HÁJEK, J. and ŜIDAK, Z. (1967). *Theory of Rank Tests.* Academic Press, Prague.

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

IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. of California Publ. Statist.* 3, No. 2.

NOETHER, G. E. (1967). *Elements of Nonparametric Statistics.* Wiley, New York.

VIŽKOVÁ, Z. (1973). Unpublished thesis, Charles University, Prague.

NATIONAL CENTRAL BUREAU OF STATISTICS
FACK
S-102 50 STOCKHOLM 27
SWEDEN