ON LIMITING DISTRIBUTIONS OF INTERMEDIATE ORDER STATISTICS FROM STATIONARY SEQUENCES

by

Vernon Watts, Holger Rootzén and M.R. Leadbetter

Abstract

Let $X_1, X_2 \ldots$ be a sequence of random variables and write $X_k^{(n)}$ for the k^{th} largest among $X_1, X_2, \ldots, X_n$. If $\{k_n\}$ is a sequence of integers such that $k_n \to \infty$, $k_n/n \to 0$, the sequence $\{X_{k_n}^{(n)}\}$ is referred to as the sequence of *intermediate order statistics* corresponding to the *intermediate rank sequence* $\{k_n\}$.

The possible limiting distributions for $X_{k_n}^{(n)}$ have been characterized (under mild restrictions) by various authors when the random variables $X_1, X_2 \ldots$ are independent and identically distributed. In this paper we consider the case when the $\{X_n\}$ form a stationary sequence and obtain a natural dependence restriction under which the "classical" limits still apply.

It is shown in particular that the general dependence restriction applies to *normal* sequences when the covariance sequence $\{r_n\}$ converges to zero as fast as an appropriate power $n^{-\rho}$ as $n \to \infty$.

Key Words and Phrases: order statistics, stationary processes, ranks, intermediate ranks.

Research supported by the Office of Naval Research under Contract N00014-75-C-0809.

1. Introduction.

The problem of finding the asymptotic distribution of the maximum term from a stationary dependent sequence of random variables (r.v.'s) has been extensively investigated in the literature. Of particular interest are the cases in which the concept of "approximate independence" is formulated mathematically in terms of conditions such as "strong mixing" or, for normal sequences, conditions on the rate of decay of the covariances. Loynes (1965) showed that under strong mixing and an additional restriction, the (suitably normalized) maximum of a dependent sequence has the same limiting distribution as the maximum of a corresponding independent and identically distributed (i.i.d.) sequence, provided the latter sequence has a limiting distribution. This limiting distribution is thus necessarily one of the three classical types of extreme value limit laws. For stationary normal sequences Berman (1964) found covariance conditions under which the distribution of the maximum converges to the double-exponential limit law, which arises in the i.i.d. normal case. More recently, Leadbetter (1974) obtained the general result of Loynes under a weaker "distributional mixing" assumption and showed that with Berman's covariance conditions the normal case may be placed into the general framework. Additionally, Leadbetter considered the related high-level exceedance problem for stationary sequences, leading to corresponding limiting results for extreme order statistics.

Our objective in this paper is to obtain analogous results for so-called *intermediate order statistics*. Specifically, for a given sequence of r.v.'s $\{X_n\}$, let $X_k^{(n)}$ denote the k^{th} largest of $X_1,\ldots,X_n$, and let $\{k_n\}$ be integers such that $1 \le k_n \le n$ for each n. Then if

$k_n \to \infty$ but $k_n/n \to 0$, $\{X_{k_n}^{(n)}\}$ is called a sequence of *intermediate order statistics* and $\{k_n\}$ an *intermediate rank sequence*. Wu (1966) found that, subject to the mild restriction that k_n increase monotonically, when the $\{X_n\}$ are i.i.d. the only possible nondegenerate limit laws for the normalized sequence $\{a_n(X_{k_n}^{(n)} - b_n)\}$ are normal and lognormal. In Section 2 we will establish general conditions under which the intermediate order statistic $X_{k_n}^{(n)}$ from a stationary dependent sequence $\{X_n\}$ has the same asymptotic distribution as it would if the $\{X_n\}$ were i.i.d. These conditions parallel those used to obtain the corresponding result in the extreme order statistic problem, a primary difference being that certain more rapid "mixing" rates have to be assumed. Using our procedure it is convenient to deal directly with an appropriate level exceedance problem and to regard that of asymptotic distributions as a specialization. In Section 3 we show that under a certain decay of the covariance function our general conditions are satisfied by a stationary normal sequence $\{X_n\}$; in this instance it is known (see Cheng (1965)) that the asymptotic distribution of $X_{k_n}^{(n)}$ for an independent sequence is itself normal and hence is also normal in the dependent situation considered.

2. The general stationary case.

First suppose that $\{X_n\}$ is an i.i.d. sequence of r.v.'s with marginal distribution function (d.f.) $F(x) = P(X_1 \le x)$ and that $\{k_n\}$ is an intermediate rank sequence. Let $\{u_n\}$ be real numbers, write $S_n = \sum_{i=1}^{n} I_{n,i}$ where $I_{n,i}$ is the indicator of the event $\{X_i > u_n\}$, i.e. $I_{n,i} = 1$ if $X_i > u_n$ and $I_{n,i} = 0$ otherwise, so that S_n is the number of exceedances of the level u_n by $X_1, \ldots, X_n$; and let Φ be the standard normal distribution function. It follows from the Berry-Esseen theorem and the basic equality

$$P(X_{k_n}^{(n)} \le u_n) = P(S_n < k_n)$$

that

(2.1) $$P(X_{k_n}^{(n)} \le u_n) \to \Phi(u) \quad \text{as} \quad n \to \infty$$

if and only if

(2.2) $$1 - F(u_n) = k_n/n - u\sqrt{k_n}/n + o(\sqrt{k_n}/n) \ .$$

Thus, there are constants a_n, b_n $(a_n > 0)$ such that $a_n(X_{k_n}^{(n)} - b_n)$ has a limiting distribution if and only if there exists a function $u(x)$ such that, writing $u_n(x) = x/a_n + b_n$,

(2.3) $$1 - F(u_n(x)) = k_n/n - u(x)\sqrt{k_n}/n + o(\sqrt{k_n}/n)$$

for all continuity points of $\Phi(u(x))$, and furthermore if (2.3) holds then

$$P(a_n(X_{k_n}^{(n)} - b_n) \le x) \to \Phi(u(x)) \quad \text{as} \quad n \to \infty$$

for all continuity points of $\Phi(u(x))$. Wu (1966) proved that if $\{k_n\}$ is nondecreasing then the only possibilities for $u(x)$ are

(i) $u(x) = -\alpha \log |x|$, $\quad x < 0$ $(\alpha > 0)$

$u(x) = \infty$, $\quad x \ge 0$

(ii) $u(x) = -\infty$, $\quad x \le 0$

$u(x) = \alpha \log x$, $\quad x > 0$ $(\alpha > 0)$

(iii) $u(x) = x$

(iv) functions obtained by replacing x by $ax+b$ $(a > 0)$ in (i), (ii), or (iii).

It may be noted that if for example F is continuous then for any real u it is possible to choose levels u_n satisfying (2.2), and hence such that (2.1) holds, but of course these levels may not necessarily constitute a family $u_n(x) = x/a_n + b_n$ which satisfies (2.3) for some function $u(x)$.

Our approach to proving that, say, (2.1) holds for a stationary dependent sequence $\{X_n\}$ is to assume that (2.2) holds and then to use a dependent central limit theorem to prove that

$$P(S_n < k_n) \to \Phi(u) \quad \text{as} \quad n \to \infty$$

and thus that (2.1) holds. Since (2.1) and (2.2) are equivalent for independent sequences, the assumption (2.2) can alternatively be stated as $P(\hat{X}^{(n)}_{k_n} \le u_n) \to \Phi(u)$ where $\hat{X}^{(n)}_{k_n}$ is the k_n^{th} order statistic in the "associated independent sequence" $\hat{X}_1, \hat{X}_2, \ldots$, that is, an i.i.d. sequence which has the same marginal d.f. F as each X_n. For easy reference we start by stating two known results from dependent central limit theory. The first one is Lemma 5.2 of Dvoretzky (1972), while the second one follows for example from Theorem 2.3 of Durrett and Resnick (1978).

Lemma 2.1. *Let* X *be an r.v. on* (Ω, A, P), *write* $\sigma(X)$ *for the* σ*-field generated by* X, *let* B *be a sub-*σ*-field of* A *and define*

$$\alpha = \sup\{|P(AB) - P(A)P(B)| : A \in \sigma(X),\ B \in B\} .$$

If $|X| \le 1$ *then*

$$E|E(X|\mathcal{B}) - E(X)| \le 4\alpha \ .$$

Lemma 2.2. *For* $n \ge 1$ *let* $\{X_{n,i}\}_{i=1}^{N_n}$ *be r.v.'s on the probability space* $(\Omega, \mathcal{B}, P)$ *and let* $\{C_{n,i}\}$ *be sub-σ-fields of* $\mathcal{B}$ *such that* $X_{n,i}$ *is* $C_{n,i}$*-measurable. Suppose further that* $C_{n,i} \subset C_{n,i+1}$ *and that* $E(X_{n,i+1}|C_{n,i}) = 0$ *for* $1 \le i < N_n$. *If* $|X_{n,i}| \le \varepsilon_n$, $1 \le i \le N_n$, *for some constants* $\varepsilon_n \to 0$, *and if*

$$\sum_{i=2}^{N_n} E(X_{n,i}^2|C_{n,i-1}) \overset{P}{\to} \sigma^2 \quad \text{as} \quad n \to \infty \tag{2.4}$$

for some constant $\sigma \ge 0$, *then*

$$P\left(\sum_{i=1}^{N_n} X_{n,i} \le x\right) \to \Phi(x/\sigma) \quad \text{as} \quad n \to \infty$$

for all real x, *where* $\Phi(x/0)$ *is defined to be* 1 *for* $x \ge 0$ *and* 0 *for* $x < 0$.

To be able to give conditions restricting the dependence in the sequence $\{X_n\}$ it is useful to introduce certain "mixing coefficients." Let $\mathcal{B}_{n,k} = \sigma(I_{n,1}, \ldots, I_{n,k})$ be the σ-field generated by $I_{n,1}, \ldots, I_{n,k}$; define

$$\alpha_1(n,k) = \sup\{|P(\{X_{n+i} \le u_n\} \cap B) - P(X_{n+i} \le u_n)P(B)| ;\ i \ge 0,\ B \in \mathcal{B}_{n,n-k}\} ,$$

$$\alpha_2(n,k) = \sup\{|P(\{X_{n+i} \le u_n, X_{n+j} \le u_n\} \cap B) - P(X_{n+i} \le u_n, X_{n+j} \le u_n)P(B)| ;$$
$$i,j \ge 0,\ |i-j| \le k,\ B \in \mathcal{B}_{n,n-k}\} ;$$

and put

$$\bar{\alpha}(n,k) = \max\{\alpha_1(n,k) , \alpha_2(n,k)\} .$$

It is easily checked (by simply listing the events of $\sigma(I_{n,n+i}, I_{n,n+j})$) that

$$4\bar{\alpha}(n,k) \geq \sup\{|P(A\cap B) - P(A)P(B)|; A \in \sigma(I_{n,n+i}, I_{n,n+j}) \text{ for some } i,j \geq 0 , |i-j| \leq k , B \in \mathcal{B}_{n,n-k}\} .$$

Our main dependence condition, to be called $A(u_n)$, depends on the levels u_n and involves sequences $\{\ell_n\},\{\ell_n'\}$ of integers which of course may be chosen to be different for different sequences $\{u_n\}$.

Condition $A(u_n)$ will be said to hold if

$$\frac{n}{k_n} \sum_{i=1}^{[\sqrt{k_n}]} |P(X_1>u_n , X_{1+i}>u_n) - (1-F(u_n))^2| \to 0 \text{ as } n \to \infty ,$$

and if furthermore there exist sequences $\{\ell_n\}$ and $\{\ell_n'\}$ of integers satisfying $\ell_n' \leq \ell_n \leq \sqrt{k_n}$, $\ell_n' = o(\ell_n)$, $\ell_n = o(\sqrt{k_n})$ such that

$$\frac{n}{\sqrt{k_n}} \bar{\alpha}(n,\ell_n) \to 0 \text{ and } \frac{n}{\sqrt{k_n}} \bar{\alpha}(n,\ell_n') \to 0 \text{ as } n \to \infty .$$

The mixing condition in $A(u_n)$ differs from the strong mixing condition which uses the mixing coefficient

$$\alpha(n,k) = \sup\{|P(AB) - P(A)P(B)|; A\in\sigma(X_n,X_{n+1},\dots) , B\in\sigma(X_1,\dots,X_{n-k})\} ,$$

in that substantially fewer events are involved. However, for a strongly mixing sequence, clearly $\alpha(n,k) \geq \bar{\alpha}(n,k')$ if $k \leq k'$, and hence the second part of $A(u_n)$ follows if

$$\frac{n}{\sqrt{k_n}} \alpha(n,\ell_n') \to 0 \text{ as } n \to \infty .$$

However this condition may be harder to check; in particular this seems to be the case when $\{X_n\}$ is normal.

To state the next lemma, which contains the major part of the proof of (2.1) for dependent sequences, we need some further notation. We partition the first n integers into long and short "intervals" $J_1, J_1', J_2, J_2', \ldots, J_{N_n}'$, with $J_1, J_1', \ldots, J_{N_n}$ of alternating lengths $\ell_n, \ell_n', \ldots, \ell_n$ and with J_{N_n}' of length $r \leq \ell_n + \ell_n'$. Clearly

$$N_n \sim n/\ell_n . \tag{2.5}$$

Further, define $C_{n,i} = \sigma(I_{n,j};\, j \in \bigcup_{k=1}^{i} J_k)$ and $C_{n,i}' = \sigma(I_{n,j};\, j \in \bigcup_{k=1}^{i} J_k')$, and put

$$X_{n,i} = \sum_{j \in J_i} \{I_{n,j} - E(I_{n,j} | C_{n,i-1})\}/\sqrt{k_n}$$

and

$$X_{n,i}' = \sum_{j \in J_i'} \{I_{n,j} - E(I_{n,j} | C_{n,i-1}')\}/\sqrt{k_n}$$

for $2 \leq i \leq N_n$.

<u>Lemma 2.3</u>. *Suppose that the stationary sequence* $\{X_n\}$ *satisfies* $A(u_n)$. *Then*

$$(2.6)\qquad \sum_{i=2}^{N_n} \sum_{\substack{j,k \in J_i \\ j \neq k}} |E(I_{n,j} I_{n,k}) - E(I_{n,j})E(I_{n,k})|/k_n \to 0$$

and

$$(2.7)\qquad \sum_{i=2}^{N_n} \sum_{j \in J_i} \{E(I_{n,j}|C_{n,i-1}) - E(I_{n,j})\}/\sqrt{k_n} \overset{L_1}{\to} 0$$

as $n \to \infty$, *and* (2.6) *and* (2.7) *hold also when* J_i *is replaced by* J_i' *and* $C_{n,i}$ *by* $C'_{n,i}$. *If in addition* (2.2) *holds then*

$$(2.8)\qquad \begin{aligned} &\sum_{i=2}^{N_n} E(X_{n,i}^2|C_{n,i-1}) \overset{L_1}{\to} 1 , \\ &\sum_{i=2}^{N_n} E(X'^2_{n,i}|C'_{n,i-1}) \overset{L_1}{\to} 0 \end{aligned}$$

as $n \to \infty$.

<u>Proof</u>. Since $E(I_{n,j}) = 1 - F(u_n)$ and $E(I_{n,j} I_{n,k}) = P(X_j > u_n , X_k > u_n)$ it follows by stationarity that

$$\sum_{i=2}^{N_n} \sum_{\substack{j,k \in J_i \\ j \neq k}} |E(I_{n,j} I_{n,k}) - E(I_{n,j})E(I_{n,k})|/k_n$$

$$\leq N_n \ell_n \sum_{i=1}^{[\sqrt{k_n}]} |P(X_1 > u_n , X_{1+i} > u_n) - (1-F(u_n))^2|/k_n ,$$

which tends to zero as $n \to \infty$ by $A(u_n)$ since $N_n \ell_n / k_n \sim n/k_n$. This proves (2.6).

Next by Lemma 2.1 and stationarity we have for $j \in J_i$ that

$$E|E(I_{n,j}|\mathcal{C}_{n,i-1}) - E(I_{n,j})| \le 4\alpha_1(n,\ell_n') \le 4\bar{\alpha}(n,\ell_n') \ ,$$

and hence by $A(u_n)$ that

$$\begin{aligned}\sum_{i=2}^{N_n} \sum_{j \in J_i} E|E(I_{n,j}|\mathcal{C}_{n,i-1}) - E(I_{n,j})|/\sqrt{k_n} &\le 4N_n \ell_n \bar{\alpha}(n,\ell_n')/\sqrt{k_n} \\ &\le Kn\bar{\alpha}(n,\ell_n')/\sqrt{k_n} \\ &\to 0 \text{ as } n \to \infty \ ,\end{aligned}$$

and (2.7) follows.

To prove the first part of (2.8) we note that

$$(2.9) \qquad E(X_{n,i}^2|\mathcal{C}_{n,i-1}) = \sum_{j,k \in J_i} \{E(I_{n,j} I_{n,k}|\mathcal{C}_{n,i-1}) - E(I_{n,j}|\mathcal{C}_{n,i-1}) \cdot E(I_{n,k}|\mathcal{C}_{n,i-1})/k_n\} \ .$$

Reasoning as above, we have

$$\begin{aligned}(2.10) \qquad \sum_{i=2}^{N_n} \sum_{j,k \in J_i} E|E(I_{n,j} I_{n,k}|\mathcal{C}_{n,i-1}) - E(I_{n,j} I_{n,k})|/k_n &\le 16 N_n \ell_n^2 \bar{\alpha}(n,\ell_n')/k_n \\ &\to 0 \text{ as } n \to \infty \ ,\end{aligned}$$

and furthermore, since $|I_{n,j}| \le 1$,

$$E|E(I_{n,j}|\mathcal{C}_{n,i-1})E(I_{n,k}|\mathcal{C}_{n,i-1}) - E(I_{n,k})E(I_{n,j})|$$

$$= E|\{E(I_{n,j}|\mathcal{C}_{n,i-1}) - E(I_{n,j})\}E(I_{n,k}|\mathcal{C}_{n,i-1})$$

$$+ E(I_{n,j})\{E(I_{n,k}|\mathcal{C}_{n,i-1}) - E(I_{n,k})\}|$$

$$\le E|E(I_{n,j}|\mathcal{C}_{n,i-1}) - E(I_{n,j})| + E|E(I_{n,k}|\mathcal{C}_{n,i-1}) - E(I_{n,k})|$$

$$\le 8\alpha_1(n,\ell_n') \ ,$$

and thus it follows similarly that

$$(2.11) \qquad \sum_{i=2}^{N_n} \sum_{j,k\in J_i} E|E(I_{n,j}|\mathcal{C}_{n,i-1})E(I_{n,k}|\mathcal{C}_{n,i-1}) - E(I_{n,j})E(I_{n,k})|/k_n$$

$$\to 0 \text{ as } n \to \infty \ .$$

Further, by (2.6), (2.5), and (2.2),

$$\sum_{i=2}^{N_n} \sum_{j,k\in J_i} \{E(I_{n,j}\, I_{n,k}) - E(I_{n,j})E(I_{n,k})\}/k_n$$

$$= \sum_{i=2}^{N_n} \sum_{j\in J_i} \{E(I_{n,j}) - E^2(I_{n,j})\}/k_n$$

$$+ \sum_{i=2}^{N_n} \sum_{\substack{j,k\in J_i \\ j\neq k}} \{E(I_{n,j}\, I_{n,k}) - E(I_{n,j})E(I_{n,k})\}/k_n$$

$$= (N_n-1)\ell_n\{(1-F(u_n)) - (1-F(u_n))^2\}/k_n + o(1)$$

$$\to 1 \text{ as } n \to \infty \ ,$$

and together with (2.9) - (2.11) this proves (2.8).

Finally, the proofs of the remaining assertions of the lemma are similar and are left to the reader. □

Our main results now follow easily.

Theorem 2.4. *Let* $\{X_n\}$ *be a stationary sequence of r.v.'s, let* $\{k_n\}$ *be an intermediate rank sequence, and let* S_n *be the number of exceedances of* u_n *by* $X_1, \ldots, X_n$. *If* (2.2) *and* $A(u_n)$ *hold then*

$$P((S_n - E(S_n))/\sqrt{k_n} \le x) \to \Phi(x) \quad \text{as} \quad n \to \infty$$

for all real x, *and therefore*

$$P(X_{k_n}^{(n)} \le u_n) = P(S_n < k_n) \to \Phi(u) \quad \text{as} \quad n \to \infty .$$

Proof. Since $|I_{n,j} - E(I_{n,j}|C_{n,i-1})| \le 2$ we have that $|X_{n,i}| \le 2\ell_n/\sqrt{k_n} \to 0$, and it follows at once from (2.8) and the definition of $\{X_{n,i}\}$ that the conditions of Lemma 2.2 are satisfied (with $\sigma^2 = 1$), and hence that

$$\sum_{i=2}^{N_n} X_{n,i} \overset{d}{\to} \Phi \text{ as } n \to \infty .$$

Similarly it follows that

$$\sum_{i=1}^{N_n} X'_{n,i} \overset{d}{\to} 0 \text{ as } n \to \infty .$$

Together with Lemma 2.3 this implies that

$$(S_n - E(S_n))/\sqrt{k_n} = \sum_{j\in J_1\cup J_1'} (I_{n,j} - EI_{n,j})/\sqrt{k_n} + \sum_{i=2}^{N_n} X_{n,i} + \sum_{i=2}^{N_n} X'_{n,i}$$

$$+ \sum_{i=2}^{N_n} \sum_{j\in J_i} \{E(I_{n,j}|C_{n,i-1}) - E(I_{n,j})\}/\sqrt{k_n}$$

$$+ \sum_{i=2}^{N_n} \sum_{j\in J_i'} \{E(I_{n,j}|C'_{n,i-1}) - E(I_{n,j})\}/\sqrt{k_n}$$

$$\overset{d}{\to} \Phi \text{ as } n \to \infty$$

and thus proves the first part of the theorem.

Next by (2.2)

$$(k_n - E(S_n))/\sqrt{k_n} = (k_n - n(1 - F(u_n)))/\sqrt{k_n}$$

$$\to u \text{ as } n \to \infty ,$$

and, writing

$$P(S_n < k_n) = P((S_n - E(S_n))/\sqrt{k_n} \le (k_n - E(S_n))/\sqrt{k_n}) ,$$

the last part of the theorem follows at once since Φ is continuous. □

Using this result we obtain the following theorem, giving sufficient conditions for $X_{k_n}^{(n)}$ to have an asymptotic distribution, which is the same as if the X_n's were i.i.d.

<u>Theorem 2.5</u>. *Let* $\{X_n\}$ *be stationary and suppose that for some constants* $a_n > 0$, b_n

$$P(a_n(\hat{X}_{k_n}^{(n)} - b_n) \le x) \to \Phi(u(x)) \quad as \quad n \to \infty$$

for all continuity points x *of* u *where* $\{\hat{X}_{k_n}^{(n)}\}$ *is the independent sequence associated with* $\{X_n\}$. *If* $A(u_n)$ *is satisfied for* $u_n = x/a_n + b_n$ *for all continuity points* x *for which* $u(x)$ *is finite, then for such* x

$$P(a_n(X_{k_n}^{(n)} - b_n) \le x) \to \Phi(u(x)) \text{ as } n \to \infty .$$

This then holds for all x *if* u *is continuous (as is the case when for example* k_n *increases monotonically).*

3. The normal case.

In this section the general results obtained above are applied to normal sequences. Let $\{X_n\}$ be a stationary normal sequence which for convenience is assumed to be standardized to have zero means and unit variances. We assume that its covariance function $r_n = EX_1X_{1+n}$ satisfies

$$r_n = O(n^{-\rho}) \tag{3.1}$$

for some constant $\rho > 0$ to be specified later. Write

$$\delta = \sup_{n\ge 1}|r_n| , \quad \delta_n = \sup_{m\ge n}|r_m| .$$

It is easily seen that since $r_n \to 0$ we must have $\delta < 1$, and that (3.1) implies $\delta_n = O(n^{-\rho})$. Further, let $\{k_n\}$ be an intermediate rank sequence and define $\theta = \theta(\{k_n\})$ by

$$\theta = \inf\{\theta' ; k_n = O(n^{\theta'})\} .$$

Clearly $0 \le \theta \le 1$ and $k_n = O(n^{\theta+\varepsilon})$ for all $\varepsilon > 0$.

Now, for x real, suppose that u_n satisfies (2.2) (with u replaced by x), i.e. suppose that

$$1 - \Phi(u_n) = k_n/n - x\sqrt{k_n}/n + o(\sqrt{k_n}/n) \ . \tag{3.2}$$

By making a first order expansion of Φ around the point b_n, it is easily seen that one such u_n is $u_n = x/a_n + b_n$ with

$$b_n = \Phi^{-1}(1 - k_n/n) \ , \ a_n = n\Phi'(b_n)/\sqrt{k_n} \ .$$

Somewhat more generally, $u_n = x/a_n' + b_n'$ for a_n', b_n' satisfying $a_n^{-1} a_n' \to 1$, $a_n^{-1}(b_n' - b_n) \to 0$ also satisfies (3.2). We require the following two useful technical results. First, for $\{u_n\}$ satisfying (3.2) we have

$$k_n/n \sim 1 - \Phi(u_n) \sim (2\pi)^{-\frac{1}{2}} u_n^{-1} e^{-u_n^2/2} \ ,$$

and taking logarithms gives $u_n \sim \sqrt{2 \log n/k_n}$ so that

$$e^{-u_n^2} \sim 4\pi(k_n/n)^2 \log n/k_n \ . \tag{3.3}$$

In the following two lemmas we find conditions on ρ which ensure that $A(u_n)$ is satisfied.

<u>Lemma 3.1</u>. *Suppose that* $\theta < 1$ *and that* $\{r_n\}$ *satisfies* (3.1) *for some* $\rho > \theta$. *Then*

$$\frac{n}{k_n} \sum_{i=1}^{[\sqrt{k_n}]} |P(X_1 > u_n , X_{1+i} > u_n) - (1-\Phi(u_n))^2| \to 0 \text{ as } n \to \infty .$$

Proof. As a special case of a result used by Berman (1964) and others, we have that

$$|P(X_1 > u_n , X_{1+i} > u_n) - (1-\Phi(u_n))^2| \le K|r_i|e^{-u_n^2/(1+|r_i|)}$$

for some constant K (depending only on δ but whose value may change from line to line below). Hence

$$\frac{n}{k_n} \sum_{j=1}^{[\sqrt{k_n}]} |P(X_1 > u_n , X_{1+i} > u_n) - (1-\Phi(u_n))^2|$$

$$\le K \frac{n}{k_n} \sum_{i=1}^{[\sqrt{k_n}]} |r_i| e^{-u_n^2/(1+|r_i|)} ,$$

and we estimate the latter sum by splitting it into two parts: for $1 \le j \le \gamma$ and for $\gamma < j \le [\sqrt{k_n}]$, where $\gamma = [(n/k_n)^{\varepsilon}]$ with $0 < \varepsilon < (1-\delta)/(1+\delta)$. By (3.3)

$$\frac{n}{k_n} \sum_{i=1}^{\gamma} |r_i| e^{-u_n^2/(1+|r_i|)} \le K \frac{n}{k_n} \left(\frac{k_n}{n}\right)^{2/(1+\delta)} (\log \frac{n}{k_n})\gamma$$

$$\to 0 \text{ as } n \to \infty$$

by the choice of γ.

Since $\theta < 1$ and $\delta_n = O(n^{-\rho})$ by the assumption on $\{r_n\}$, we have that $\delta_\gamma u_n^2 \to 0$ as $u \to \infty$, and hence (3.3) gives, for $i > \gamma$,

$$e^{-u_n^2/(1+|r_i|)} \le e^{-u_n^2/(1+\delta_\gamma)} = e^{-u_n^2 + \delta_\gamma u_n^2/(1+\delta_\gamma)}$$

$$\le K\left(\frac{k_n}{n}\right)^2 \log \frac{n}{k_n} \ .$$

Thus, (defining the sum to be zero for $\gamma \ge [\sqrt{k_n}]$

$$\frac{n}{k_n} \sum_{i=\gamma+1}^{[\sqrt{k_n}]} |r_i| e^{-u_n^2/(1+|r_i|)} \le K \frac{k_n}{n} \log \frac{n}{k_n} \sum_{i=\gamma+1}^{[\sqrt{k_n}]} |r_i|$$

$$\le K \frac{k_n}{n} \log \frac{n}{k_n} \sum_{i=\gamma+1}^{[\sqrt{k_n}]} i^{-\rho} \ .$$

For the three separate cases $\rho < 1$, $\rho = 1$, and $\rho > 1$, the last sum is bounded by a constant multiple of $k_n^{(1-\rho)/2}$, $\log \sqrt{k_n}$, and 1 respectively. Therefore in any case the expression on the right-hand side tends to zero since $\rho > \theta$, thus concluding the proof of the lemma. □

To establish the latter part of $A(u_n)$ we shall further extend an important method, due to Slepian, Berman, and Cramér, from the extreme value theory of normal processes. In addition to conditions on ρ, we shall for convenience assume that k_n does not increase too slowly, or more precisely that

(3.4) $$k_n/(\log n)^{2/\rho} \to \infty \text{ as } n \to \infty \ .$$

Lemma 3.2. *Suppose that* $\{r_n\}$ *satisfies* (3.1) *and* (3.4) *with* $\rho > \max(3\theta/2\,,\, 2(2-1/\theta))$ *and that* $\{u_n\}$ *satisfies* (3.2). *Then there exist sequences* $\{\ell_n\}$ *and* $\{\ell_n'\}$ *which satisfy the requirements of* $A(u_n)$.

Proof. We first show that there exists a sequence $\{\ell_n'\}$ with $\ell_n' \le \sqrt{k_n}$ and $\ell_n' = o(\sqrt{k_n})$ such that

$$(3.5) \qquad \frac{n}{\sqrt{k_n}}\, \alpha_1(n,\ell_n') \to 0 \text{ as } n \to \infty .$$

First, by (3.4), a sequence $\{\ell_n'\}$ can be chosen so that $\ell_n'=o(\sqrt{k_n})$, $\ell_n'\le\sqrt{k_n}$ but such that $\ell_n' \ge (\log n)^{1/\rho}$. We shall impose a slight further restriction on ℓ_n' later, but for the moment just assume these properties. Then since $\delta_n \le Kn^{-\rho}$ by (3.1), $u_n^2\,\delta_{\ell_n'} \le K(\log n)(\log n)^{-1} = K$, and hence by (3.3), for $j \ge \ell_n'$,

$$(3.6) \qquad e^{-u_n^2/(1+\delta_j)} = e^{-u_n^2+u_n^2\,\delta_j/(1+\delta_j)} \le K(k_n/n)^2 \log n/k_n .$$

Now let $B \in \sigma(I_{n,1},\ldots, I_{n,n-\ell_n'})$ and $\ell \ge 0$ be fixed. Then B is a disjoint union of sets of the form $\bigcap_{i=1}^{n-\ell_n'} \{I_{n,i} = x_i\}$, where each x_i is zero or one; and hence for any j, $1 \le j \le n - \ell_n'$,

$$B = B_0\{I_{n,j} = 0\} \cup B_1\{I_{n,j} = 1\}$$

where B_0 and B_1 are sets of the same general form as B, except that the j^{th} factor in the intersections are missing. It is evident that

$B = \{(X_1,\ldots,X_{n-\ell'_n}) \in \overline{B}\}$, $B_i = \{(X_1,\ldots,X_{j-1}, X_{j+1},\ldots,X_{n-\ell'_n}) \in \overline{B}_i\}$,

$i = 0,1$,

for some sets $\overline{B} \in R^{n-\ell'_n}$, $\overline{B}_0, \overline{B}_1 \in R^{n-\ell'_n-1}$.

Let R_1 be the covariance matrix of the vector $(X_1,\ldots,X_{n-\ell'_n}, X_{n+\ell})$, let R_0 be the covariance matrix it would have if $(X_1,\ldots,X_{n-\ell'_n})$ and $X_{n+\ell}$ were independent, and define $R_h = hR_1 + (1-h)R_0$. Without loss of generality it may be assumed that R_1 and hence R_h is positive definite, and writing

$$F(h) = \int_{\overline{x}\in\overline{B}}\!\!\cdots\!\int \int_{x_{n+\ell}=-\infty}^{u_n} f_h$$

where $\overline{x} = (X_1,\ldots,X_{n-\ell'_n})$ and f_h is the density function of a zero-mean normal vector with covariance matrix R_h, we have that

$$(3.7) \qquad |P(\{X_{n+\ell} \le u_n\} \cap B) - P(X_{n+\ell} \le u_n)P(B)| = |F(1) - F(0)| \le \int_0^1 |F'| .$$

Proceeding as in Leadbetter, Lindgren, and Rootzén (1978, pp. 46-47), we obtain

$$(3.8) \qquad F'(h) = \sum_{j=1}^{n-\ell'_n} r_{n+\ell-j} \int_{\overline{x}\in\overline{B}}\!\!\cdots\!\int \int_{x_{n+\ell}=-\infty}^{u_n} \frac{\partial^2 f_h}{\partial x_j \partial x_{n+\ell}} .$$

As above, $\{\overline{x} \in \overline{B}\} = \{\overline{x}^* \in \overline{B}_0\}\{x_j \le u_n\} \cup \{\overline{x}^* \in \overline{B}_1\}\{x_j > u_n\}$ where $\overline{x}^* = (x_1,\ldots,x_{j-1}, x_{j+1},\ldots,x_{n-\ell'_n})$, and performing the integrations over x_j and $x_{n+\ell}$ gives

$$\int\cdots\int_{\bar{x}^*\in\bar{B}_0} \int_{x_j=-\infty}^{u_n} \int_{x_{n+\ell}=-\infty}^{u_n} \frac{\partial^2 f_h}{\partial x_j \partial x_{n+\ell}} = \int\cdots\int_{\bar{x}^*\in\bar{B}_0} f_h(x_j = x_{n+\ell} = u_n)$$

$$\leq \int_{-\infty}^{\infty}\!\cdots\!\int f_h(x_j = x_{n+\ell} = u_n)$$

where $f_h(x_j = x_{n+\ell} = u_n)$ is the function of $\bar{x}^*$ which is obtained by putting $x_j = u_n$, $x_{n+\ell} = u_n$ in f_h. The last integral is easily seen to be bounded by $Ke^{-u_n^2/(1+|r_{n+\ell-j}|)}$, with K depending only on δ. Next, making the change of variables $y_i = x_i$, $i \neq j$, $y_j = -x_j$ and writing $\bar{y}^* = (y_1,\ldots,y_{j-1}, y_{j+1},\ldots,y_{n-\ell_n'})$, we have

$$\int\cdots\int_{\bar{x}^*\in\bar{B}_1} \int_{x_j=u_n}^{\infty} \int_{x_{n+\ell}=-\infty}^{u_n} \frac{\partial^2 f_h}{\partial x_j \partial x_{n+\ell}} = -\int\cdots\int_{\bar{y}^*\in\bar{B}_1} \int_{y_j=-\infty}^{-u_n} \int_{y_{n+\ell}=-\infty}^{u_n} \frac{\partial^2 f_h}{\partial y_j \partial y_{n+\ell}}$$

$$= -\int\cdots\int_{\bar{y}^*\in\bar{B}_1} g_h(y_j = -u_n,\ y_{n+\ell} = u_n)$$

where g_h is defined from $(X_1,\ldots,X_{j-1}, -X_j, X_{j+1},\ldots,X_{n-\ell_n'}, X_{n+\ell})$ in the same way as f_h is defined from $(X_1,\ldots,X_{n-\ell_n'}, X_{n+\ell})$. Again, the modulus of the latter integral is seen to be bounded by $Ke^{-u_n^2/(1+|r_{n+\ell-j}|)}$, and it follows that

$$\left|\int\cdots\int_{\bar{x}\in\bar{B}} \int_{x_{n+\ell}=-\infty}^{u_n} \frac{\partial^2 f_h}{\partial x_j \partial x_\ell}\right| \leq Ke^{-u_n^2/(1+|r_{n+\ell-j}|)} .$$

Inserting this into (3.7) and (3.8) gives

$$|P(\{X_{n+\ell} \le u_n\} \cap B) - P(X_{n+\ell} \le u_n)P(B)| \le K \sum_{j=1}^{n-\ell'_n} |r_{n+\ell-j}| e^{-u_n^2/(1+|r_{n+\ell-j}|)}$$

$$\le K \sum_{j=\ell'_n}^{n} \delta_j \, e^{-u_n^2/(1+\delta_j)} .$$

Since the last expression is independent of the particular ℓ and B considered, we have that

$$\alpha_1(n,\ell'_n) \le K \sum_{j=\ell'_n}^{n} \delta_j \, e^{-u_n^2/(1+\delta_j)} .$$

Thus, by again using $r_n = O(n^{-\rho})$ and (3.6), we have

$$(3.9) \qquad \frac{n}{\sqrt{k_n}} \alpha_1(n,\ell'_n) \le K \frac{n}{\sqrt{k_n}} \left(\frac{k_n}{n}\right)^2 \log \frac{n}{k_n} \sum_{j=\ell'_n}^{n} j^{-\rho} .$$

For the three cases $\rho < 1$, $\rho = 1$, and $\rho > 1$, the last sum is bounded by a constant times $n^{1-\rho}$, $\log n$, and $\ell_n'^{\,1-\rho}$ respectively. Thus, since $\rho > \max(3\theta/2\,,\,2(2-1/\theta))$, the right-hand side of (3.9) clearly tends to zero when $\rho \le 1$. For $\rho > 1$ it is readily seen that ℓ'_n may be redefined (by increasing if necessary, keeping $\ell'_n = o(\sqrt{k_n})$, $\ell'_n \le \sqrt{k_n}$) so that (3.9) still tends to zero. Hence (3.5) follows.

The proof that $nk_n^{-\frac{1}{2}} \alpha_2(n,\ell'_n) \to 0$ as $n \to \infty$ for the above choice of ℓ'_n is only notationally more complicated, and together with (3.5) this shows that

$$\frac{n}{\sqrt{k_n}}\,\bar{\alpha}(n,\ell_n') \to 0 \text{ as } n \to \infty .$$

It is now easy to see in the same way that, for any sequence ℓ_n with $\ell_n' \le \ell_n \le \sqrt{k_n}$, we have

$$\frac{n}{\sqrt{k_n}}\,\bar{\alpha}(n,\ell_n) \to 0 \text{ as } n \to \infty ,$$

and this proves the lemma. □

It now follows at once that $A(u_n)$, and hence the results of Theorems 2.4 and 2.5, hold for stationary normal sequences which satisfy the above conditions. To avoid repetition we only state an analog of Theorem 2.5.

Theorem 3.3. *Suppose that* $\{X_n\}$ *is a stationary normal sequence and* $\{k_n\}$ *an intermediate rank sequence such that*

$$r_n = O(n^{-\rho}) , \qquad \textit{some } \rho > \max(3\theta/2 , 2(2-1/\theta)) ,$$

and suppose that in addition $k_n/(\log n)^{2/\rho} \to \infty$. *Then*

$$P(a_n(X_{k_n}^{(n)} - b_n) \le x) \to \Phi(x) \;\; \textit{as} \;\; n \to \infty$$

for all real x, *where* a_n *and* b_n *are defined by* $\Phi(b_n) = 1 - k_n/n$ *and* $a_n = n\Phi'(b_n)/\sqrt{k_n}$.

Finally, it should be remarked that the covariance condition of the theorem does not seem to be optimal. Perhaps even a condition like

$$\frac{k_n}{n} \log \frac{n}{k_n} \sum_{i=1}^{n} |r_i| \to 0 ,$$

or, translated into terms of (3.1), $\rho > \theta$, may be sufficient. In fact, we have been able to show that if X_n can be written as a moving average of independent normal random variables $X_n = \sum_{i=-\infty}^{\infty} c_i Y_{n-i}$, with $c_n = O(n^{-\rho})$ for some $\rho > \max(\theta, \frac{1}{2})$, then the conclusion of Theorem 3.3 holds. In particular, this provides a large class of examples of processes with

$$r_n = O(n^{-\rho}) ,$$

such that $P(a_n(X_{k_n} - b_n) \le x) \to \Phi(x)$ for any $\rho > \max(\theta, \frac{1}{2})$.

References

Berman, S.M. (1964). Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.*, *35*, 502-516.

Cheng, B. (1965). The limiting distributions of order statistics. *Chinese Math.*, *6*, 84-104.

Durrett, R. and Resnick, S.I. (1978). Functional limit theorems for dependent variables. *Ann. Probability*, *6*, 829-846.

Dvoretzky, A. (1972). Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, *2*, 513-535.

Leadbetter, M.R. (1974). On extreme values in stationary sequences. *Z. Wahr. verw. Geb.*, *28*, 289-303.

Leadbetter, M.R., Lindgren, G., and Rootzén, H. (1978). Conditions for the convergence in distribution of maxima of stationary normal processes. *Stochastic Processes Appl.*, *8*, 131-139.

Loynes, R.M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.*, *36*, 993-999.

Watts, J.H.V. (1977). Limit theorems and representations for order statistics from dependent sequences. Ph.D. dissertation, University of North Carolina at Chapel Hill.

Wu, C.Y. (1966). The types of limit distributions for some terms of variational series. *Scientia Sinica, 15*, 749-762.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Limiting Distributions of Intermediate Order Statistics from Stationary Sequences		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
		6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1284
7. AUTHOR(s) Vernon Watts, Holger Rootzén and M.R. Leadbetter		8. CONTRACT OR GRANT NUMBER(s) ONR Contract N00014-75-C-0809
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program (Code 436) Arlington, VA 22217		12. REPORT DATE May 1980
		13. NUMBER OF PAGES 24
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release -- Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

order statistics, stationary processes, ranks, intermediate ranks

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $X_1, X_2 \ldots$ be a sequence of random variables and write $X_k^{(n)}$ for the k^{th} largest among $X_1, X_2, \ldots, X_n$. If $\{k_n\}$ is a sequence of integers such that $k_n \to \infty$, $k_n/n \to 0$, the sequence $\{X_{k_n}^{(n)}\}$ is referred to as the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. sequence of *intermediate order statistics* corresponding to the *intermediate rank sequence* $\{k_n\}$.

The possible limiting distributions for $X_{k_n}^{(n)}$ have been characterized (under mild restrictions) by various authors when the random variables $X_1, X_2 \ldots$ are independent and identically distributed. In this paper we consider the case when the $\{X_n\}$ form a stationary sequence and obtain a natural dependence restriction under which the "classical" limits still apply.

It is shown in particular that the general dependence restriction applies to *normal* sequences when the covariance sequence $\{r_n\}$ converges to zero as fast as an appropriate power $n^{-\rho}$ as $n \to \infty$.