

AD-A052 315

FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS
THE WEAK REPRESENTATION OF INTERMEDIATE ORDER STATISTICS.(U)
DEC 77 V WATTS

F/6 12/1

N00014-76-C-0608

UNCLASSIFIED

FSU-STATISTICS-M443

NL

[OF]

AD
A052315



END
DATE
FILMED
5-78
DDC

AD A 052315

AD No. 
UDC FILE COPY



The Florida State University
Department
of
Statistics

Tallahassee, Florida
32306



12

THE WEAK REPRESENTATION OF
INTERMEDIATE ORDER STATISTICS

by

Vernon Watts

FSU Statistics Report M443
ONR Technical Report No. 128

December, 1977
Department of Statistics
The Florida State University
Tallahassee, Florida 32306

Research sponsored by the Office of Naval Research under Contract Nos.
N00014-76-C-0608 and N00014-75-C-0809.

DDC
RECEIVED
APR 7 1978
B

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

The Weak Representation of
Intermediate Order Statistics

by

Vernon Watts

ABSTRACT

It is shown that under certain conditions a sample intermediate order statistic of a sequence of random variables has a representation in terms of the empirical distribution function and a remainder which is stochastically of small order. The methods employed yield readily to the general setting of sequences which are m -dependent and not necessarily identically distributed. Under the assumptions it follows from the representation that the sequence of intermediate order statistics is asymptotically normal.

ACCESSION for	
NTIS	White Section <input checked="" type="checkbox"/>
DDC	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION _____	
BY _____	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	AVAIL. and/or SPECIAL
A	

1. Introduction. In [5] and [6] we showed that under certain conditions a sample intermediate order statistic from an independent and identically distributed sequence of random variables has an almost sure representation in terms of the empirical distribution function and a remainder of suitably small order. The methods we used were parallel to those of Bahadur [1], who established the corresponding representation for sample λ -quantiles.

In this paper we show that similar representations for intermediate order statistics can be obtained under fewer assumptions. However the conclusions are weaker in that the convergence of the remainder term is "in probability" rather than almost sure, and moreover, a less precise order of the remainder term is found. Our procedures follow those used by Ghosh [3] to extend the sample λ -quantile representation to sequences satisfying conditions weaker than those of Bahadur.

Without loss of convenience we achieve a degree of generality by relaxing the requirements of independence and identical distributions. In place of the former we use the condition of m -dependence, and instead of the latter we allow non-identical distributions which are in a sense uniformly close.

2. Preliminary notation. Let $\{X_n\}_{n \geq 1}$ be an m -dependent sequence of random variables on the probability space (Ω, F, P) , with marginal distribution functions $F_n(x) = P(X_n \leq x)$. By m -dependence we mean that $F(X_1, \dots, X_n)$ and $F(X_{n+l}, X_{n+l+1}, \dots)$ are independent σ -fields if $l > m$, where $F(\cdot)$ is generated by the indicated variables. Let $\{k_n\}_{n \geq 1}$ be an intermediate rank sequence of integers, that is, $1 \leq k_n \leq n$ and $k_n/n \rightarrow 0$, and denote by $\{X_{k_n}^{(n)}\}$ the corresponding sequence of intermediate order statistics. For each $n \geq 1$, let $x_0 \geq -\infty$ be

such that $x_0 = \inf\{x: F_n(x) > 0\}$ and $F_n(x_0) = 0$; equivalently, x_0 is the left endpoint of F_n and F_n assigns zero probability to x_0 , for each n . Now define

$$\bar{F}_n(x) = n^{-1} \sum_{i=1}^n F_i(x).$$

Then for $n \geq 1$ and $i = 1, \dots, n$, let $x_{n,i}$ and \tilde{x}_n be such that

$$F_i(x_{n,i}) = k_n/n$$

and

$$\bar{F}_{(n)}(\tilde{x}_n) = k_n/n.$$

Under assumptions to be imposed, $x_{n,i}$ and \tilde{x}_n will be uniquely defined for all sufficiently large n .

In the following two sections we develop the results, examining separately the cases when $x_0 > -\infty$ and $x_0 = -\infty$.

3. The finite endpoint case. Throughout this section we assume that $x_0 > -\infty$. Suppose that in some interval $(x_0, x_0 + \delta)$, each F_n has a continuous derivative F'_n such that $\lim_{x \rightarrow x_0} F'_n(x)$ exists and is positive. Assume further that

$$(3.1) \quad x_{n,i} - \tilde{x}_n = o(k_n^{1/2}/n)$$

uniformly in $i = 1, \dots, n$, as $n \rightarrow \infty$ (which implies that $\tilde{x}_n \rightarrow x_0$), and that

$$(3.2) \quad \lim_n n^{-1} \sum_{i=1}^n F'_i(\tilde{x}_n) = \lim_n \bar{F}'_{(n)}(\tilde{x}_n) > 0.$$

(The limit is assumed to exist.) Finally, assume that

$$(3.3) \quad \overline{\lim}_n k_n/\sigma_n^2 < \infty,$$

where

$$(3.4) \quad \sigma_n^2 = \text{Var}\left(\sum_{i=1}^n I(X_i \leq x_{n,i})\right) \\ = k_n(1 - k_n/n) + 2 \sum_{1 \leq i < j \leq n} (P(X_i \leq x_{n,i}, X_j \leq x_{n,j}) - (k_n/n)^2).$$

Then we have the following result.

Theorem 1. Under the above conditions (3.1) - (3.3),

$$(3.5) \quad X_{k_n}^{(n)} = \tilde{x}_n + \frac{k_n - \sum_{i=1}^n I(X_i \leq x_{n,i})}{n \bar{F}'_{(n)}(\tilde{x}_n)} + R_n,$$

where $n R_n / k_n^{\frac{1}{2}} \xrightarrow{P} 0$.

Proof: We first assume that the $\{X_n\}$ are independent and later indicate the modifications to be made for the more general m -dependent case.

For any real t we have for sufficiently large n ,

$$(3.6) \quad P(n(X_{k_n}^{(n)} - \tilde{x}_n)/k_n^{\frac{1}{2}} \leq t) = P\left(\sum_{i=1}^n I(X_i \leq \tilde{x}_n + tk_n^{\frac{1}{2}}/n) \geq k_n\right) \\ = P\left(\left(1/k_n^{\frac{1}{2}} \bar{F}'_{(n)}(\tilde{x}_n)\right) \sum_{i=1}^n (F_i(\tilde{x}_n + tk_n^{\frac{1}{2}}/n) - I(X_i \leq \tilde{x}_n + tk_n^{\frac{1}{2}}/n)) \geq k_n/n - k_n/n\right) \\ \leq \left(1/k_n^{\frac{1}{2}} \bar{F}'_{(n)}(\tilde{x}_n)\right) \sum_{i=1}^n (F_i(\tilde{x}_n + tk_n^{\frac{1}{2}}/n) - k_n/n) \\ = P(Z_{t,n} \leq t_n),$$

say. Since $\bar{F}_{(n)}(\tilde{x}_n + tk_n^{\frac{1}{2}}/n) = k_n/n + tk_n^{\frac{1}{2}} \bar{F}'_{(n)}(\tilde{x}_n)/n + o(k_n^{\frac{1}{2}}/n)$, it follows that as $n \rightarrow \infty$,

$$(3.7) \quad t_n = (n/k_n^{\frac{1}{2}} \bar{F}'_{(n)}(\tilde{x}_n)) (tk_n^{\frac{1}{2}} \bar{F}'_{(n)}(\tilde{x}_n)/n + o(k_n^{\frac{1}{2}}/n)) \rightarrow t.$$

Suppose $t > 0$. Let

$$W_n = (1/k_n^{1/2} \bar{F}'_{(n)}(\tilde{x}_n)) \sum_{i=1}^n (k_n/n - I(X_i \leq x_{n,i})).$$

By (3.1) we have for large n that $x_{n,i} < \tilde{x}_n + tk_n^{1/2}/n$, so that

$$(3.8) \quad W_n - Z_{t,n} = (1/k_n^{1/2} \bar{F}'_{(n)}(\tilde{x}_n)) \sum_{i=1}^n (I(x_{n,i} < X_i \leq tk_n^{1/2}/n) - p_{n,i}),$$

where $p_{n,i} = F_i(\tilde{x}_n + tk_n^{1/2}/n) - k_n/n$. Since

$$E(W_n - Z_{t,n})^2 \leq (1/k_n \bar{F}'_{(n)}(\tilde{x}_n))^2 \sum_{i=1}^n p_{n,i}$$

and since

$$\begin{aligned} \sum_{i=1}^n p_{n,i} &= n \bar{F}_{(n)}(\tilde{x}_n + tk_n^{1/2}/n) - k_n \\ &= tk_n^{1/2} \bar{F}'_{(n)}(\tilde{x}_n) + o(k_n^{1/2}), \end{aligned}$$

it follows that $E(W_n - Z_{t,n})^2 \rightarrow 0$ as $n \rightarrow \infty$, so that

$$(3.9) \quad W_n - Z_{t,n} \xrightarrow{P} 0.$$

Similarly, (3.9) holds for $t \leq 0$. Also, by the usual central limit theorem for row sums from triangular arrays, it is easily seen that

$$(3.10) \quad W_n \Rightarrow N(0, (\lim_n \bar{F}'_{(n)}(\tilde{x}_n))^{-2}),$$

so in particular the sequence of distributions $\{L(W_n)\}_{n \geq 1}$ is tight. (\Rightarrow denotes convergence in distribution.)

Let $V_n = n(X_{k_n}^{(n)} - \tilde{x}_n)/k_n^{1/2}$. Then by (3.6) - (3.10) we see that V_n and W_n satisfy the conditions of Lemma 1 of Ghosh [3], that is, the collection $\{L(W_n)\}$ is tight, and for all real t and all $\epsilon > 0$,

$$\lim_n P(V_n \leq t, W_n \geq t + \epsilon) = 0$$

and

$$\lim_n P(V_n \geq t + \epsilon, W_n \leq t) = 0.$$

Therefore $V_n - W_n \xrightarrow{P} 0$. Hence for R_n defined by (3.5),

$$\begin{aligned} n R_n / k_n^{\frac{1}{2}} &= n(X_{k_n}^{(n)} - \bar{x}_n) / k_n^{\frac{1}{2}} - (1/k_n^{\frac{1}{2}} \bar{F}'_n(\bar{x}_n)) \sum_{i=1}^n (k_n/n - I(X_i \leq x_{n,i})) \\ &= V_n - W_n \\ &\xrightarrow{P} 0. \end{aligned}$$

We now indicate the modifications to be made in order to establish the theorem for m -dependent sequences. We need to show that the sequence $\{L(W_n)\}$ is tight, and

$$(3.11) \quad E(W_n - Z_{t,n})^2 \rightarrow 0.$$

To obtain (3.11) we use the following result of Rosen [4]: If Y_1, \dots, Y_n are m -dependent ($n \geq m$) and have finite variances, then

$$(3.12) \quad \text{Var}(\sum_{i=1}^n Y_i) \leq (2m + 1) \sum_{i=1}^n \text{Var} Y_i.$$

Applying this to (3.8) yields

$$E(W_n - Z_{t,n})^2 \leq ((2m + 1)/k_n (\bar{F}'_n(\bar{x}_n))^2) \sum_{i=1}^n p_{n,i},$$

which tends to zero as in the independent case.

In order to show that $\{L(W_n)\}$ is tight we employ the following central limit theorem for row sums from triangular arrays with m -dependent rows ($m \geq 0$ fixed), due to Rosen [4]: Let $\{X_{n,i}\}_{n \geq 1, 1 \leq i \leq n}$, with d.f.'s

$F_{n,i}(x) = P(X_{n,i} \leq x)$ be an array of random variables with $X_{n,1}, \dots, X_{n,n}$ m -dependent for each $n \geq m + 1$ and let $S_n = \sum_{i=1}^n X_{n,i}$. If

$$E X_{n,i} = 0 \text{ for all } n, i,$$

$$\text{Var } S_n = 1 \text{ for all } n,$$

$$\lim_n \sum_{i=1}^n \int_{|x|>\epsilon} x^2 dF_{n,i}(x) = 0 \text{ for all } \epsilon > 0,$$

and

$$\overline{\lim}_n \sum_{i=1}^n \text{Var } X_{n,i} < \infty,$$

then

$$S_n \Rightarrow N(0, 1).$$

Now, we first write

$$(3.13) \quad W_n = -(\sigma_n/k_n)^{1/2} F'_{(n)}(\tilde{x}_n) \sum_{i=1}^n (I(X_i \leq x_{n,i}) - k_n/n)/\sigma_n.$$

If (3.3) holds, then it is easily shown that the array $\{X_{n,i}\}$, for n large and $1 \leq i \leq n$, given by

$$X_{n,i} = (I(X_i \leq x_{n,i}) - k_n/n)/\sigma_n,$$

satisfies the assumptions of the above central limit theorem, so that $S_n = \sum_{i=1}^n X_{n,i} \Rightarrow N(0, 1)$. Then using (3.12) again we see that $\sigma_n^2/k_n \leq (2m + 1)(1 + o(1))$, and it is clear from (3.13) that the sequence $\{L(W_n)\}$ is tight (but not necessarily weakly convergent). The proof of the theorem is now complete. \square

For identically distributed sequences $\{X_n\}$ we may replace the $x_{n,i}$, $1 \leq i \leq n$, by $\tilde{x}_n = x_n$, say, and thus disregard (3.1). Additionally in this instance the assumption (3.2) is redundant. If also the $\{X_n\}$ are independent,

then (3.3) is trivially satisfied and hence we have a representation weaker than that given in [5], but using fewer assumptions on the marginal d.f. and having no restriction upon the intermediate rank sequence $\{k_n\}$. Also, for stationary m -dependent sequences ($m \geq 1$), (3.3) becomes

$$\overline{\lim}_n k_n (k_n + 2 \sum_{i=1}^m (n-i) (P(X_1 \leq x_n, X_{i+1} \leq x_n) - (k_n/n)^2))^{-1} < \infty,$$

which holds if $(2/k_n) \sum_{i=1}^m (n-i) (P(X_1 \leq x_n, X_{i+1} \leq x_n) - (k_n/n)^2)$ is bounded away from -1 for all large n .

To conclude this section we point out that the asymptotic normality of $X_{k_n}^{(n)}$ under our assumptions easily follows as a corollary to the representation (3.5). This may be seen by applying the above quoted central limit theorem along with the relation (3.3), and we state the result formally as a theorem.

Theorem 2. With the previous notation and under the same conditions as Theorem 1,

$$(n \bar{F}'_{(n)}(\tilde{x}_n)/\sigma_n) (X_{k_n}^{(n)} - \tilde{x}_n) \Rightarrow N(0, 1).$$

4. The infinite endpoint case. In this section we suppose that the d.f.'s $\{F_n\}$ of the m -dependent sequence $\{X_n\}$ have the common left endpoint $x_0 = -\infty$. The assumptions made upon the marginal d.f.'s are generalizations of those conditions in the independent and identically distributed case under which it is known that $X_{k_n}^{(n)}$ is asymptotically normal (see Cheng [2]). Specifically, assume first that there is a real number c such that $F'_n(x)$ and $F''_n(x)$ exist for $x < c$, for each $n \geq 1$. Then, with the previous notation, assume that there are constants $p \leq 1$ and $M < \infty$ such that

$$(4.1) \quad \bar{F}_{(n)}(\tilde{x}_n) / |\tilde{x}_n|^p \bar{F}'_{(n)}(\tilde{x}_n) < M$$

and

$$(4.2) \quad \bar{F}_{(n)}(\tilde{x}_n) |\bar{F}'_{(n)}(\tilde{x}_n + y_n)| / (\bar{F}'_{(n)}(\tilde{x}_n))^2 < M$$

for all large n , where $\{y_n\}$ is any sequence such that $y_n = o(|x_n|^p)$. Furthermore, suppose that

$$(4.3) \quad x_{n,i} - \tilde{x}_n = o(k_n^{1/2}/n \bar{F}'_{(n)}(\tilde{x}_n)) \text{ uniformly in } i = 1, \dots, n,$$

and finally that (3.3) holds, where σ_n^2 is defined by (3.4).

Theorem 3. Let $\{X_n\}$ be m -dependent with d.f.'s $\{F_n\}$ having the left endpoint $x_0 = -\infty$ and such that the conditions (4.1) - (4.3) and (3.3) are satisfied.

Then

$$X_{k_n}^{(n)} = \tilde{x}_n + \frac{k_n - \sum_{i=1}^n I(X_i \leq x_{n,i})}{n \bar{F}'_{(n)}(\tilde{x}_n)} + R_n,$$

where $n \bar{F}'_{(n)}(\tilde{x}_n) R_n / k_n^{1/2} \xrightarrow{p} 0$.

Proof: We proceed as in the proof of Theorem 1, supposing initially that the $\{X_n\}$ are independent. For any real t ,

$$P((n \bar{F}'_{(n)}(\tilde{x}_n) / k_n^{1/2})(X_{k_n}^{(n)} - \tilde{x}_n) \leq t) = P(Z_{t,n} \leq t_n),$$

where now

$$Z_{t,n} = k_n^{-1/2} \sum_{i=1}^n (F_i(\tilde{x}_n + tk_n^{1/2}/n\bar{F}'_{(n)}(\tilde{x}_n)) - I(X_i \leq \tilde{x}_n + k_n^{1/2}/n\bar{F}'_{(n)}(\tilde{x}_n)))$$

and

$$t_n = k_n^{-\frac{1}{2}} \sum_{i=1}^n (F_i(\tilde{x}_n + tk_n^{\frac{1}{2}}/n\bar{F}'_{(n)}(\tilde{x}_n)) - k_n/n).$$

Let $y_n = tk_n^{\frac{1}{2}}/n\bar{F}'_{(n)}(\tilde{x}_n)$. Since $p \leq 1$ and $k_n \rightarrow \infty$, (4.1) implies that $y_n = o(\tilde{x}_n)$.

Thus for n sufficiently large a Taylor expansion gives

$$\bar{F}_{(n)}(\tilde{x}_n + y_n) = k_n/n + tk_n^{\frac{1}{2}}/n + \frac{1}{2}y_n^2 \bar{F}''_{(n)}(\tilde{x}_n + \theta_n y_n),$$

where $|\theta_n| \leq 1$. It then follows from (4.2) that $t_n \rightarrow t$.

Let $t > 0$. Define

$$W_n = k_n^{-\frac{1}{2}} \sum_{i=1}^n (k_n/n - I(X_i \leq x_{n,i})).$$

By (4.3) we may suppose that $x_{n,i} < \tilde{x}_n + y_n$, so that

$$W_n - Z_{t,n} = k_n^{-\frac{1}{2}} \sum_{i=1}^n (I(x_{n,i} < X_i \leq \tilde{x}_n + y_n) - p_{n,i}),$$

where

$$p_{n,i} = F_i(\tilde{x}_n + y_n) - k_n/n.$$

Therefore

$$\begin{aligned} E(W_n - Z_{t,n})^2 &\leq k_n^{-1} \sum_{i=1}^n p_{n,i} \\ &= (n/k_n)(\bar{F}_{(n)}(\tilde{x}_n + y_n) - k_n/n) \end{aligned}$$

$$(4.4) \quad \rightarrow 0.$$

(4.4) holds similarly for $t \leq 0$, and so $W_n - Z_{t,n} \xrightarrow{P} 0$ for every real t . Also,

$W_n \Rightarrow N(0, 1)$. Then defining

$$V_n = (n\bar{F}'_{(n)}(\tilde{x}_n)/k_n^{\frac{1}{2}})(X_{k_n}^{(n)} - \tilde{x}_n),$$

we have by Lemma 1 of Ghosh [3] that $V_n - W_n \xrightarrow{P} 0$, and the theorem for independent sequences follows. Modifications similar to those made in the finite endpoint case establish the theorem for m -dependent sequences. \square

Just as in Section 3, simplifications of notation and assumptions can be made for identically distributed sequences, but we do not restate these here. However we do point out the following result which is valid in the general setting.

Theorem 4. Under the conditions of Theorem 3,

$$(\sqrt{n} \bar{F}'_{(n)}(\tilde{x}_n)/\sigma_n)(X_{k_n}^{(n)} - \tilde{x}_n) \Rightarrow N(0, 1).$$

REFERENCES

- [1] Bahadur, R. R. (1966). A note on quantiles in large samples. Ann. Math. Statist. 37, 577-580.
- [2] Cheng, B. (1965). The limiting distributions of order statistics. Chinese Math. 6, 84-104.
- [3] Ghosh, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. Ann. Math. Statist. 42, 1957-1961.
- [4] Rosen, B. (1967). On the central limit theorem for sums of dependent random variables. Z. Wahr. verw. Geb. 7, 48-82.
- [5] Watts, V. (1977). An almost sure representation for intermediate order statistics: The finite endpoint case. U.N.C. Mimeo Series No. 1136.
- [6] Watts, V. (1977). Asymptotic behavior of intermediate order statistics: The infinite endpoint case. F.S.U. Statistics Report M436.

14 FSU-STATISTICS-M443,
TR-128-ONR

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M443 ✓ ONR No. 128 ✓	2. GOVT ACCESSION NO. 3. RECIPIENT'S CATALOG NUMBER
4. TITLE 6 The Weak Representation of Intermediate Order Statistics.	5. TYPE OF REPORT & PERIOD COVERED 9 Technical Report.
7. AUTHOR(s) 10 Vernon/Watts	6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M443 8. CONTRACT OR GRANT NUMBER(s) 15 ONR N00014-76-C-0608 ONR N00014-75-C-0809
9. PERFORMING ORGANIZATION NAME & ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 12 15p.
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics and Probability Program Arlington, Virginia 22217	12. REPORT DATE 11 Dec 1977 13. NUMBER OF PAGES 11
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 Intermediate ranks, empirical distribution function, convergence in probability.	
20. ABSTRACT It is shown that under certain conditions a sample intermediate order statistic of a sequence of random variables has a representation in terms of the empirical distribution function and a remainder which is stochastically of small order. The methods employed yield readily to the general setting of sequences which are m-dependent and not necessarily identically distributed. Under the assumptions it follows from the representation that the sequence of intermediate order statistics is asymptotically normal.	

400 277

Gen