AD-A052 314 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1

ASYMPTOTIC BEHAVIOR OF INTERMEDIATE ORDER STATISTICS: THE INFIN--ETC(U)

OCT 77 V WATTS N00014-76-C-0608

UNCLASSIFIED FSU-STATISTICS-M436 NL

| OF |

AD A052314

END

DATE FILMED

5 -78

DDC

The Florida State University

Department of Statistics

Tallahassee, Florida
32306

12

ASYMPTOTIC BEHAVIOR OF INTERMEDIATE ORDER STATISTICS: THE INFINITE ENDPOINT CASE

by

Vernon Watts

FSU Statistics Report M436
ONR Technical Report No. 125

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

Research sponsored by the Office of Naval Research under Contract No. N00014-76-C-0608.

ASYMPTOTIC BEHAVIOR OF INTERMEDIATE ORDER STATISTICS:
THE INFINITE ENDPOINT CASE

by

Vernon Watts

ABSTRACT

Suppose $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variables with marginal distribution function $F(x) = P(X_1 \le x)$ satisfying $F(x) > 0$ for all real x. Let $X_{k_n}^{(n)}$ denote the k_nth smallest order statistic of the sample $X_1, \ldots, X_n$, where $k_n/n \to 0$ as $n \to \infty$. An almost sure representation of $X_{k_n}^{(n)}$ in terms of the empirical distribution function is established. The conditions imposed upon F include [illegible] under which it is known that $X_{k_n}^{(n)}$ is asymptotically normal. From the representation the law of the iterated logarithm for $X_{k_n}^{(n)}$ is obtained. Examples illustrating the general result are presented.

1. Introduction. Suppose $\{X_n\}_{n\geq 1}$ is a sequence of independent and identically distributed (i.i.d.) random variables on (Ω, F, P) with marginal distribution function (d.f.) $F(x) = P(X_1 \leq x)$. Denote the k^{th} smallest order statistic of the sample $X_1, \ldots, X_n$ by $X_k^{(n)}$. Let $\{k_n\}_{n\geq 1}$ be a sequence of integers satisfying $1 \leq k_n \leq n$ for each n and $k_n/n \to 0$ as $n \to \infty$; $\{k_n\}$ is then called an intermediate rank sequence and $\{X_{k_n}^{(n)}\}_{n\geq 1}$ a sequence of intermediate order statistics.

In this paper we study the limiting properties of sequences $\{X_{k_n}^{(n)}\}_{n\geq 1}$. Our consideration is restricted to the case in which $F(x) > 0$ for all real x, that is, F has an infinite left endpoint. We first state a known result giving conditions under which $X_{k_n}^{(n)}$ is asymptotically normal. Then our primary interest is the development of an almost sure representation for $X_{k_n}^{(n)}$ in terms of the empirical d.f. of $X_1, \ldots, X_n$, which is analogous to our previous result obtained when $F(x_o) = 0$ for some $x_o > -\infty$ (Watts (1977)). The procedure for deriving this essentially follows that used by Bahadur (1966) to establish the corresponding representation for sample λ-quantiles. Furthermore, we obtain the law of the iterated logarithm for $X_{k_n}^{(n)}$, and finally we illustrate the general representation with two examples.

2. Asymptotic normality and the almost sure representation. The basis for our investigation is the following theorem of Cheng (1965) giving sufficient conditions for the intermediate order statistic $X_{k_n}^{(n)}$ to be asymptotically normal. Here and throughout the paper we let x_n be defined by $F(x_n) = k_n/n$. The conditions upon the d.f. F insure that x_n is determined uniquely for all large n.

Theorem 1. Suppose for some real number c the two derivatives F' and F" exist with $F'(x) > 0$, for $x < c$. Also assume that for some $p \leq 1$ and some $M < \infty$,

$$(2.1) \qquad \frac{F(x_n)}{|x_n|^p\, F'(x_n)} < M \text{ and } \frac{F(x_n)\,|F''(x_n+y_n)|}{(F'(x_n))^2} < M$$

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

$$(2.2) \qquad P\left\{\frac{n\,F'(x_n)}{k_n^{\frac{1}{2}}}(X_{k_n}^{(n)} - x_n) \le x\right\} \to \Phi(x)$$

as $n \to \infty$, for all real x, where Φ is the standard normal d.f.

<u>Proof</u>. For any x, in order that (2.2) hold it is necessary and sufficient that

$$(2.3) \qquad \frac{n}{k_n^{\frac{1}{2}}}(F(u_n) - k_n/n) \to x$$

as $n \to \infty$, where $u_n = u_n(x) = x_n + x\,k_n^{\frac{1}{2}}/n\,F'(x_n)$. (See Smirnov (1967).) Let $y_n = x\,k_n^{\frac{1}{2}}/n\,F'(x_n) = x\,F(x_n)/k_n^{\frac{1}{2}}\,F'(x_n)$. Since $p \le 1$ and $k_n \to \infty$ we have $y_n = o(x_n)$ by (2.1), so that $u_n \to -\infty$. Thus we may write

$$F(u_n) = F(x_n) + x\,k_n^{\frac{1}{2}}/n + \tfrac{1}{2}\,y_n^2\,F''(x_n + \theta_n y_n),$$

where $|\theta_n| \le 1$. Then since $y_n = o(|x_n|^p)$ and $F(x_n) = k_n/n$, the second condition in (2.1) leads to (2.3). $\square$

From now on we assume that F satisfies the conditions stated in Theorem 1. Moreover we impose the mild restriction that $k_n/\log^3 n \to \infty$. Let

$$a_n = C_o\,\frac{k_n^{\frac{1}{2}}\,\log^{\frac{1}{2}} n}{n\,F'(x_n)}$$

for some $C_o > 0$ and let $\{b_n\}$ be a sequence of positive integers satisfying

$$b_n \sim (k_n/\log n)^{\frac{1}{4}} \to \infty.$$

From (2.1) it follows that $a_n = o(x_n)$, so that $x_n + a_n \to -\infty$. For large n let $I_n = [x_n - a_n, x_n + a_n]$ and define

$$L_n = \sup\{F'(x): \ x \in I_n\}$$

and

$$M_n = \inf\{F'(x): \ x \in I_n\}.$$

Assume that

$$\text{(2.4)} \qquad L_n/F'(x_n) \text{ and } F'(x_n)/M_n \text{ are bounded as } n \to \infty.$$

Let $F_n(x, \omega) = n^{-1} \sum_{i=1}^n I(X_i \le x)$ denote the empirical d.f. of $X_1, \ldots, X_n$. In the remainder of this section we establish the following result.

<u>Theorem 2.</u> Let $\{X_n\}$ be an i.i.d. sequence with marginal d.f. F satisfying the conditions of Theorem 1, and in addition, (2.4), where the intermediate rank sequence $\{k_n\}$ is such that $k_n/\log^3 n \to \infty$. Then

$$\text{(2.5)} \qquad X_{k_n}^{(n)}(\omega) = x_n + \frac{k_n - n F_n(x_n, \omega)}{n F'(x_n)} + R_n(\omega),$$

where $R_n(\omega) = O\left(\frac{k_n^{\frac{1}{2}} \log^{3/4} n}{n F'(x_n)}\right)$ with probability one.

For large n let

$$G_n(x, \omega) = F_n(x, \omega) - F(x) - F_n(x_n, \omega) + k_n/n$$

and

$$H_n(\omega) = \sup\{|G_n(x, \omega)|: \ x \in I_n\}.$$

Most of the steps in the proof of Theorem 2 are contained in the following two lemmas.

Lemma 2.1. With probability one,

$$H_n(\omega) = O(F'(x_n)\, a_n\, b_n^{-1}) = O(n^{-1} k_n^{\frac{1}{2}} \log^{3/4} n)$$

as $n \to \infty$.

Proof: Let N_1 be such that x_n is uniquely determined for $n \geq N_1$. For $n \geq N_1$ and any integer r let $\eta_{r,n} = x_n + r\, a_n\, b_n^{-1}$. By the monotonicity of F and F_n we have for $x \in [\eta_{r,n}, \eta_{r+1,n}]$ that

$$F_n(\eta_{r,n}, \omega) - F(\eta_{r+1,n}) \leq F_n(x, \omega) - F(x) \leq F_n(\eta_{r+1,n}, \omega) - F(\eta_{r,n}).$$

Therefore

$$H_n(\omega) \leq \max\{|G_n(\eta_{r,n}, \omega)|: \ -b_n \leq r \leq b_n\}$$

$$+ \max\{F(\eta_{r+1,n}) - F(\eta_{r,n}): \ -b_n \leq r \leq b_n - 1\}$$

$$= K_n(\omega) + \alpha_n, \tag{2.6}$$

say. We have $\eta_{r+1,n} - \eta_{r,n} = a_n b_n^{-1}$ and $\eta_{b_n,n} = x_n + a_n \to -\infty$. Then by (2.4),

$$\alpha_n = O(F'(x_n)\, a_n\, b_n^{-1}). \tag{2.7}$$

Now for any $n \geq N_1$, and any r,

$$G_n(\eta_{r,n}) = \begin{cases} n^{-1} \sum_{i=1}^n I(x_n < X_i \leq \eta_{r,n}) - p_{r,n}, & r \geq 0 \\ -n^{-1} \sum_{i=1}^n I(\eta_{r,n} < X_i \leq x_n) + p_{r,n}, & r \leq -1, \end{cases}$$

where $p_{r,n} = |F(\eta_{r,n}) - k_n/n|$. Let $t_n = C_1\, F'(x_n)\, a_n\, b_n^{-1}$ for some $C_1 > 0$. According to Bernstein's inequality (see Bahadur (1966)) we have

$$P(|G_n(\eta_{r,n})| \geq 2\exp(-h_{r,n}),$$

where

$$h_{r,n} = \frac{n^2 t_n^2}{2\{n\, p_{r,n}(1-p_{r,n}) + \frac{nt_n}{3}\max\{p_{r,n},\, 1-p_{r,n}\}\}} .$$

We may choose $N_2 \geq N_1$ such that $n \geq N_2$ and $|r| \leq b_n$ imply $t_n \leq L_n a_n$ and $p_{r,n} \leq L_n a_n$. Thus for $n \geq N_2$ and $|r| \leq b_n$,

$$P(|G_n(\eta_{r,n})| \geq t_n) \leq 2\exp(-\delta_n),$$

where

$$\delta_n = \frac{nt_n^2}{3\, L_n a_n} = \frac{C_o\, C_1^2\, F'(x_n)}{3\, L_n} \log n(1 + o(1)).$$

If C_o and C_1 are chosen sufficiently large, then

$$P(|G_n(\eta_{r,n})| \geq t_n) \leq 2\, n^{-(1+\varepsilon)}$$

for some $\varepsilon > 0$ and for all $n \geq N_3$, say, so that

$$\sum_{n\geq N_1} P(K_n \geq t_n) < \infty.$$

Therefore $P(K_n \geq t_n$ infinitely often$) = 0$ by the Borel-Cantelli lemma, and since $t_n = C_1 F'(x_n)\, a_n\, b_n^{-1}$, the statement of the lemma follows from (2.6) and (2.7). □

<u>Lemma 2.2.</u> With probability one, $X_{k_n}^{(n)} \varepsilon I_n$ for all large n.

<u>Proof:</u> We first have

$$P(X_{k_n}^{(n)} < x_n - a_n) \leq P(\textstyle\sum_{i=1}^n I(X_i \leq x_n + a_n) - np_n \geq k_n - np_n),$$

where $p_n = F(x_n - a_n)$. The right side does not exceed 2 exp(-h), where

$$h = \frac{n^2(F(x_n) - F(x_n - a_n))^2}{2k_n} \geq \frac{n^2 M_n{}^2 a_n{}^2}{2k_n} \geq (1 + \varepsilon) \log n$$

for some $\varepsilon > 0$, if C_o is sufficiently large. Therefore

$$\sum_{n \geq N_1} P(X_n < x_n - a_n) < \infty. \tag{2.8}$$

Similarly,

$$P(X_{k_n}^{(n)} > x_n + a_n) \leq P(\textstyle\sum_{i=1}^{n} I(X_i > x_n + a_n) - np_n > n(1 - p_n) - k_n)$$

$$\leq 2 \exp(-h),$$

where now $p_n = 1 - F(x_n + a_n)$ and

$$h = \frac{n^2 M_n{}^2 a_n{}^2}{2n(1-p_n) + n L_n a_n}.$$

Since $F(x_n + a_n) \leq F(x_n) + L_n a_n$ and $\log n = o(k_n)$, and by (2.4), we have that

$$h \geq \frac{n^2 M_n{}^2 a_n{}^2}{5k_n} \geq (1 + \varepsilon) \log n$$

for some $\varepsilon > 0$ and for all large n. Thus

$$\sum_{n \geq N_1} P(X_n > x_n + a_n) < \infty. \tag{2.9}$$

Combining (2.9) with (2.8) and applying the Borel-Cantelli lemma completes the proof. □

<u>Proof of Theorem 2</u>: From Lemmas 2.1 and 2.2 it follows that

$$F_n(X_{k_n}^{(n)}(\omega), \omega) - F(X_{k_n}^{(n)}(\omega)) - F_n(x_n, \omega) + k_n/n = O(F'(x_n)\, a_n\, b_n{}^{-1})$$

with probability one. Also, we have with probability one that for all large n, $F_n(X_{k_n}^{(n)}(\omega), \omega) = k_n/n$ and

$$F(X_{k_n}^{(n)}(\omega)) = k_n/n + (X_{k_n}^{(n)}(\omega) - x_n)\, F'(x_n)$$

$$+ \tfrac{1}{2}(X_{k_n}^{(n)}(\omega) - x_n)^2\, F''(x_n + a_n\, \theta_n(\omega)),$$

where $|\theta_n(\omega)| \leq 1$. But $a_n b_n F''(x_n + a_n \theta_n(\omega))/F'(x_n) \to 0$, by (2.1). Therefore

$$(X_{k_n}^{(n)}(\omega) - x_n)\, F'(x_n) - k_n/n + F_n(x_n, \omega) = O(F'(x_n)\, a_n\, b_n^{-1}),$$

and (2.5) follows. □

It may be noted that the above derivation of (2.5) requires that $\{k_n\}$ only satisfy $k_n/\log n \to \infty$. However if $k_n/\log^3 n \to \infty$, then (2.5) verifies the asymptotic normality of $X_{k_n}^{(n)}$, which follows from the usual central limit theory for row sums of triangular arrays of random variables, applied to $\{X_{n,i}\}$ defined by $X_{n,i} = (k_n/n - I(X_i \leq x_n))/k_n^{\frac{1}{2}}$, $i = 1, \ldots, n$. Then it is clear that without the stronger restriction $k_n/\log^3 n \to \infty$ the rate of convergence of the remainder term R_n indicated in Theorem 2 may not be sufficiently rapid to insure that (2.5) is meaningful.

3. <u>The law of the iterated logarithm</u>. Kiefer (1972, Theorems 5 and 6) has established the law of the iterated logarithm for the intermediate order statistic $U_{k_n}^{(n)}$ from an i.i.d. sequence $\{U_n\}_{n\geq 1}$ uniformly distributed on the unit interval. Specifically, in his Theorem 5 he has shown that if $k_n \to \infty$ and $k_n/n \to 0$ monotonically, and if $k_n/\log\log n \to \infty$, then

$$P(\limsup_{n\to\infty} \pm \frac{T_n - k_n/n}{(2\, k_n \log\log n)^{\frac{1}{2}}} = 1) = 1,$$

where $T_n = \sum_{i=1}^{n} I(U_i \le k_n/n)$. Suppose the marginal d.f. F of the sequence $\{X_n\}$ satisfies our conditions in the previous section, and moreover, is everywhere continuous, so that each $F(X_n)$ has the uniform distribution. Then using the almost sure relation

$$F_n(x_n, \omega) = n^{-1} \sum_{i=1}^{n} I(F(X_i) \le k_n/n),$$

valid for large n, along with the representation (2.5) we obtain the following result.

Theorem 3. Let $\{X_n\}$ be an i.i.d. sequence with continuous marginal d.f. F satisfying the conditions of Theorem 2. If $\{k_n\}$ satisfies $k_n \uparrow \infty$ and $k_n/n \downarrow 0$ as $n \to \infty$, along with the restriction $k_n(\log\log n)^2/\log^3 n \to \infty$, then

$$P\left(\limsup_{n\to\infty} \pm \frac{n\,F'(x_n)(X_{k_n}^{(n)} - x_n)}{(2\,k_n \log\log n)^{\frac{1}{2}}} = 1\right) = 1.$$

4. Examples. We now illustrate the representation (2.5) with two examples. For each of these the conditions of Theorem 2 can be verified in a straightforward manner.

Example 1.
$$F(x) = \begin{cases} e^x, & x < 0 \\ 1, & x \ge 0. \end{cases}$$

Let $\{k_n\}$ satisfy $k_n/\log^3 n \to \infty$. For (2.1) we take $p = 0$. Then (2.5) holds, where

$$R_n(\omega) = 0\left(\left(\frac{\log n}{k_n}\right)^{3/4}\right)$$

with probability one.

<u>Example 2.</u> $F(x) = \Phi(x).$

We have the basic relation $F(x) \sim -x^{-1} F'(x)$ as $x \to -\infty$, from which it follows that

$$x_n \sim -(2 \log n/k_n)^{\frac{1}{2}},$$

for any $\{k_n\}$. Moreover, (2.1) is satisfied with $p = -1$. Then (2.5) holds, where

$$R_n(\omega) = 0\left(\left(\frac{\log n}{k_n}\right)^{3/4} \left(\log \frac{n}{k_n}\right)^{-\frac{1}{2}}\right)$$

with probability one.

REFERENCES

Bahadur, R. R. (1966). A note on quantiles in large samples. Ann. Math. Statist. 37, 577-580.

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

Kiefer, J. (1972). Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. Proc. Sixth Berkeley Symp. Math. Statist. Probabil. 1, 227-244.

Smirnov, N. V. (1967). Some remarks on limit laws for order statistics. Theory Probabil. Appl. 12, 336-337.

Watts, V. (1977). An almost sure representation for intermediate order statistics: The finite endpoint case. Inst. of Stat. Mimeo Series No. 1136, Univ. of N.C.

(14) FSU-STATISTICS-M436, TR-125-ONR

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M436 ONR No. 125	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (6) Asymptotic Behavior of Intermediate Order Statistics: The Infinite Endpoint Case.		5. TYPE OF REPORT & PERIOD COVERED (9) Technical Rep[redacted]t. 6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M436
7. AUTHOR (s) (10) Vernon Watts		8. CONTRACT OR GRANT NUMBER(s) (15) ONR N00014-76-C-0608
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) 14,
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics and Probability Program Arlington, Virginia 22217		12. REPORT DATE (11) Oct[redacted]77 13. NUMBER OF PAGES 10
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

X(k sub n) superscript n

20. ABSTRACT

(k sub n)th

Suppose $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variables with marginal distribution function $F(x) = P(X_1 \le x)$ satisfying $F(x) > 0$ for all real x. Let $X_{k_n}^{(n)}$ denote the k_nth smallest order statistic of the sample $X_1, \ldots, X_n$, where $k_n/n \to 0$ as $n \to \infty$. An almost sure representation of $X_{k_n}^{(n)}$ in terms of the empirical distribution function is established. The conditions imposed upon F include those under which it is known that $X_{k_n}^{(n)}$ is asymptotically normal. From the representation the law of the iterated logarithm for $X_{k_n}^{(n)}$ is obtained. Examples illustrating the general result are presented.

sub n approaches infinity

400 277