AD-A052 318 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
ON THE SECOND MOMENT OF THE REMAINDER TERM APPEARING IN THE INT--ETC(U)
FEB 78 V WATTS N00014-76-C-0608
UNCLASSIFIED FSU-STATISTICS-M457 NL

| OF |

AD
A052318

END
DATE
FILMED
5-78
DDC

he Florida State University

(13)

# ON THE SECOND MOMENT OF THE REMAINDER TERM APPEARING IN THE INTERMEDIATE ORDER STATISTIC REPRESENTATION

by

Vernon Watts

FSU Statistics Report M457
ONR Technical Report No. 131

February, 1978
Department of Statistics
The Florida State University
Tallahassee, Florida 32306

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

On the Second Moment of the Remainder Term Appearing in the Intermediate Order Statistic Representation

by

Vernon Watts

## ABSTRACT

Under certain conditions a sample intermediate order statistic from a sequence of independent and identically distributed random variables has an almost sure representation involving the empirical distribution and a remainder term of small order. In this paper an asymptotic approximation of the second moment of the remainder term is obtained. It is assumed that the marginal distribution function of the independent and identically distributed sequence has a finite left endpoint.

1. Introduction. Suppose that $\{X_n\}_{n\geq 1}$ are independent and identically distributed random variables on $(\Omega, F, P)$ with marginal distribution function (d.f.) $F(x) = P(X_1 \leq x)$. Let $\{k_n\}_{n\geq 1}$ be integers such that $1 \leq k_n \leq n$ for each $n$ and $k_n \to \infty$ but $k_n/n \to 0$ as $n \to \infty$, and denote by $X_{k_n}^{(n)}$ the $k_n$ th smallest of $X_1, \ldots, X_n$. Then $\{X_{k_n}^{(n)}\}_{n\geq 1}$ is called a sequence of intermediate order statistics.

Define the left endpoint $x_0$ of $F$ by $x_0 = \inf\{x: F(x) > 0\}$, which we assume to be finite. Let $F_n(x, \omega) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x)$ be the empirical d.f. of the sample $X_1, \ldots, X_n$. Then under certain conditions on $F$, $X_{k_n}^{(n)}$ has the representation

$$(1.1) \qquad X_{k_n}^{(n)}(\omega) = x_n' + \frac{k_n/n - F_n(x_n', \omega)}{F'(x_n')} + R_n'(\omega),$$

where $x_n'$ satisfies $F(x_n') = k_n/n$ and where $R_n'(\omega) = O(n^{-1}k_n^{1/4}\log^{3/4} n)$ as $n \to \infty$ with probability one. (See Watts (1977).) In this paper we develop an asymptotic approximation to the second moment of the remainder term $R_n'$, retaining the same conditions used to establish (1.1) and imposing some additional restrictions. Our procedure follows that employed by Duttweiler (1973) to approximate the second moment of the remainder term appearing in the Bahadur (1966) representation of sample $\lambda$-quantiles.

Since the expected value of the $k$ th smallest order statistic of a sample of $n$ variables uniformly distributed on the unit interval (0, 1) is $k/(n + 1)$ rather than $k/n$, it turns out that a substantial simplification in the procedure is effected by dealing with a representation for $X_{k_n}^{(n)}$ which is slightly different from (1.1). We consider instead

$$(1.2) \qquad X_{k_n}^{(n)}(\omega) = x_n + \frac{k_n/(n + 1) - F_n(x_n, \omega)}{F'(x_n)} + R_n(\omega),$$

where $x_n$ is defined by $F(x_n) = k_n/(n + 1)$. Then the methods used to prove (1.1) also show that $R_n'(\omega) = O(n^{-1}k_n^{1/4}\log^{3/4} n)$ with probability one.

2. <u>Derivation of the approximation</u>. The following preliminary result, which we state as a lemma, provides an exact expression for the second moment of $R_n$ when F is the uniform (0, 1) distribution. The details of proof are given by Duttweiler (1973).

<u>Lemma</u>. Let $U_k^{(n)}$ be the $k^{th}$ smallest order statistic for the set $U_1, \ldots, U_n$ of n independent random variables uniformly distributed on (0, 1). Define

$$\hat{U}_k^{(n)} = k/(n + 1) + \left(\textstyle\sum_{i=1}^{n} I(U_i \geq k/(n + 1)) - n(1 - k/(n + 1))\right)/n \tag{2.1}$$

and let $R = U_k^{(n)} - \hat{U}_k^{(n)}$. Then

$$E(R^2) = \frac{2k}{n(n + 1)}\{I_{k/(n+1)}(k, n + 1 - k) - I_{k/(n+1)}(k + 1, n + 1 - k)\} - \frac{2k(1 - k/(n + 1}{n(n + 1)(n + 2}$$

where $I_x(a, b)$ is the incomplete beta function, defined by

$$I_x(a, b) = \frac{(a + b - 1)!}{(a - 1)!(b - 1)!}\int_0^x y^{a-1}(1 - y)^{b-1}\, dy.$$

Now consider a general d.f. F for which $x_0 > -\infty$. Suppose that $F(x_0) = 0$ and in an interval $(x_0, x_0 + \delta)$ F is twice differentiable with F" bounded, and such that $\lim_{x \downarrow x_0} F'(x)$ exists and is positive. These assumptions insure that $x_n$ is uniquely determined for large n. If $k_n/\log n \to \infty$ as $n \to \infty$, then (1.2) holds with the indicated order of $R_n$. For the following result we make the additional restriction that the $\{X_n\}$ have a finite second moment, that is, that $\int x^2\, dF(x) < \infty$, as well as a slightly stronger requirement on $\{k_n\}$.

<u>Theorem</u>. Suppose the $\{X_n\}$ have marginal d.f. F satisfying the above conditions and that $k_n/n^\theta \to \infty$ for some $\theta > 0$. Then for $R_n$ defined by (1.2),

$$E(R_n^2) \sim (2/\pi)^{1/2} k_n^{1/2}/(nF'(x_n))^2, \text{ as } n \to \infty.$$

<u>Proof</u>. Let $U_1, \ldots, U_n$ be independent uniform (0, 1) variables with $k_n^{th}$ smallest order statistic $U_{k_n}^{(n)}$. Define the quantile function $Q(\cdot)$ by $Q(u) = \sup\{x\colon\ F(x) \le u\}$. Then each $Q(U_i)$ has d.f. F, and

$$R_n \overset{\sim}{\sim} Q(U_{k_n}^{(n)}) - x_n - \frac{\frac{nk_n}{n+1} - \sum_{i=1}^n I(Q(U_i) \le x_n)}{nF'(x_n)},$$

where $\overset{\sim}{\sim}$ indicates having identical distributions.

We may suppose that $F'(x)$ is positive in the interval $A = (x_0, x_0 + \delta)$. Let $B = \{u\colon\ u = F(x),\ x \in A\}$. Then Q restricted to B is the inverse of F restricted to A. Let $p_n = k_n/(n + 1)$. There is an integer $N_1$ such that $p_n \in B$ if $n \ge N_1$, and we have that

$$Q(p_n) = x_n, \tag{2.2}$$

$$Q'(p_n) \text{ exists and equals } (F'(x_n))^{-1}, \tag{2.3}$$

and

$$Q(u) \le x_n \text{ if and only if } u \le p_n. \tag{2.4}$$

Also, Q" exists and is bounded in B, and a straightforward derivation shows that

$$\int_0^1 Q^2(u)\, du = EX_1^2 < \infty.$$

For $n \ge 1$ and $u \in [0, 1]$ define

$$H_n(u) = Q(u) - Q(p_n) - Q'(p_n)(u - p_n).$$

Then by (2.2)-(2.4) for $n \geq N_1$ we have that

$$R_n \stackrel{\sim}{\sim} Q(p_n) + Q'(p_n)(U_{k_n}^{(n)} - p_n) + H_n(U_{k_n}^{(n)}) - x_n - \frac{np_n - \sum_{i=1}^n I(U_i \leq p_n)}{n\,F'(x_n)}$$

(2.5)

$$\stackrel{\sim}{\sim} H_n(U_{k_n}^{(n)}) + (F'(x_n))^{-1}(U_{k_n}^{(n)} - \hat{U}_{k_n}^{(n)}),$$

where $\hat{U}_{k_n}^{(n)}$ is given by (2.1). Now by Lemma 1,

$$E(U_{k_n}^{(n)} - \hat{U}_{k_n}^{(n)})^2 = \frac{2k_n}{n(n+1)}\{I_{p_n}(k_n, n+1-k_n) - I_{p_n}(k_n+1, n+1-k_n)\} + O(k_n/n^3).$$

Applying the relation

$$I_x(a, b) - I_x(a+1, b) = \frac{(a+b-1)!}{a!(b-1)!}\, x^a(1-x)^b$$

for positive integers a and b (see Abramowitz and Stegun (1964, Equation 26.5.16)) and Stirling's formula

$$n! = e^{-n}\, n^{n+1/2}(2\pi)^{1/2}(1 + O(n^{-1})),$$

we obtain

(2.6) $$E(U_{k_n}^{(n)} - \hat{U}_{k_n}^{(n)})^2 = (2/\pi)^{1/2}(k_n^{1/2}/n^2)(1 + o(1)).$$

Therefore to complete the proof of the theorem it is sufficient from (2.5) and (2.6) and by the Schwarz inequality to show that

(2.7) $$E(H_n^2(U_{k_n}^{(n)})) = o(k_n^{1/2}/n^2).$$

Choose $\alpha$, $0 < \alpha < 1/8$, and let $\varepsilon_n = k_n^{\alpha}/n^{1/2}$ and $I_n = (\max\{0, p_n - \varepsilon_n\}, p_n + \varepsilon_n)$. We may assume that $I_n \subseteq B$. Denoting the probability density of $U_{k_n}^{(n)}$ by $g_n$, we have

$$E\,H_n^2(U_{k_n}^{(n)}) = \Big(\int_{u\in I_n} + \int_{u\notin I_n}\Big) H_n^2(u)\, g_n(u)\, du.$$

Let $H_{n,\max} = \sup_{u\in I_n} |H_n(u)|$ and $g_{n,\max} = \sup_{u\notin I_n} g_n(u)$. Then

$$E\,H_n^2(U_{k_n}^{(n)}) \le H_{n,\max}^2 + g_{n,\max} \int_0^1 H_n^2(u)\, du.$$

Also, it follows from the inequality $(a + b + c)^2/3 \le a^2 + b^2 + c^2$ that

$$(1/3)\int_0^1 H_n^2(u)\, du \le E\,X_1^2 + Q^2(p_n) + (Q'(p_n))^2.$$

Since $E\,X_1^2 < \infty$, by (2.2) and (2.3), and since $F'(x_n)$ tends to a non-zero limit, there is a constant $C_1 < \infty$ such that for $n \ge N_1$,

$$\int_0^1 H_n^2(u)\, du \le C_1.$$

Hence for $n \ge N_1$,

$$E\,H_n^2(U_{k_n}^{(n)}) \le H_{n,\max}^2 + C_1\, g_{n,\max}. \tag{2.8}$$

Then letting $C_2 = \sup_{u\in B} \{|Q''(u)|\} < \infty$, we have by Taylor's expansion that $|H_n(u)| \le C_2(u - p_n)^2/2$ for $u \in I_n$, and therefore

$$H_{n,\max}^2 \le C_2^2\, \varepsilon_n^4/4 \tag{2.9}$$

if $n \ge N_1$.

Next, we observe that the density

$$g_n(u) = \frac{n!}{(k_n - 1)!(n - k_n)!}\, u^{k_n - 1} (1 - u)^{n-k_n},\ 0 < u < 1,$$

has mode $m_n = (k_n - 1)/(n - 1)$ and decreases monotonically on both sides. Since

$|m_n - p_n| = o(\varepsilon_n)$, $m_n \in I_n$ for $n \geq N_2 \geq N_1$. Let $\kappa \geq 1$. We have

$$E(U_{k_n}^{(n)} - p_n)^{2\kappa} \geq \int_{m_n}^{p_n+\varepsilon_n} (u - p_n)^{2\kappa}\, g_n(u)\, du$$

$$\geq g_n(p_n + \varepsilon_n)(2\kappa + 1)^{-1}(\varepsilon_n^{2\kappa+1} - (m_n - p_n)^{2\kappa+1}),$$

and since $n\varepsilon_n \to \infty$ whereas $m_n - p_n = O(n^{-1})$, there exists $C_3 < \infty$, depending on $\kappa$, such that

$$g_n(p_n + \varepsilon_n) \leq C_3\, \varepsilon_n^{-(2\kappa+1)}\, E(U_{k_n}^{(n)} - p_n)^{2\kappa}$$

for $n \geq N_3 \geq N_2$. In a similar manner there exists $C_4 < \infty$ such that

$$g_n(p_n - \varepsilon_n) \leq C_4\, \varepsilon_n^{-(2\kappa+1)}\, E(U_{k_n}^{(n)} - p_n)^{2\kappa}$$

for $n \geq N_3$. Then letting $C_5 = \max\{C_3, C_4\}$ gives

$$g_{n,\max} \leq C_5\, \varepsilon_n^{-(2\kappa+1)}\, E(U_{k_n}^{(n)} - p_n)^{2\kappa}.$$

Since (see Blom (1958, p. 42)) there is a constant $C_6 < \infty$ independent of $n$, $k_n$, and $\kappa$ such that

$$E(U_{k_n}^{(n)} - p_n)^{2\kappa} \leq C_6\, n^{-\kappa},$$

it follows that

$$g_{n,\max} \leq C_6\, C_5\, n^{-\kappa}\, \varepsilon_n^{-(2\kappa+1)} \tag{2.10}$$

for $n \geq N_3$. Then (2.8), (2.9), and (2.10) lead to

$$E\, H_n^2(U_{k_n}^{(n)}) \leq C_2^2\, \varepsilon_n^4/4 + C_1 C_6 C_5\, n^{-\kappa}\, \varepsilon_n^{-(2\kappa+1)}$$

for $n \geq N_3$. Finally we may suppose that $k_n \geq n^\theta$ for some $\theta > 0$, so that

$$n^{-\kappa}\, \varepsilon_n^{-(2\kappa+1)} \leq n^{1/2-\theta\alpha(2\kappa+1)}$$

and by choosing $\kappa$ sufficiently large we obtain (2.7). □

## REFERENCES

[1] Abramowitz, M., and Stegun, I. A. (1964). *Handbook of Mathematical Functions*. New York, Dover Publications.

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

[3] Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables*. New York, John Wiley and Sons.

[4] Duttweiler, D. L. (1973). The mean-square error of Bahadur's order-statistic approximation. *Ann. Statist.* 1 446-453.

[5] Watts, V. (1977). An almost sure representation for intermediate order statistics: The finite endpoint case. Inst. of Statist. Mimeo Series No. 1136, University of North Carolina.

(14) FSU-STATISTICS-M457, TR-131-ONR

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

| | | |
|---|---|---|
| 1. REPORT NUMBER<br>FSU No. M457<br>ONR No. 131 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (6)<br>On the Second Moment of the Remainder Term Appearing in the Intermediate Order Statistic Representation. | | 5. TYPE OF REPORT & PERIOD COVERED<br>(9) Technical Rep[redacted]rt,<br>6. PERFORMING ORG. REPORT NUMBER<br>FSU Statistics Report M457 |
| 7. (10) AUTHOR(s)<br>Vernon Watts | | 8. CONTRACT OR GRANT NUMBER(s)<br>(15) ONR N00014-76-C-0608 |
| 9. PERFORMING ORGANIZATION NAME & ADDRESS<br>The Florida State University<br>Department of Statistics<br>Tallahassee, Florida 32306 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>(12) 11p. |
| 11. CONTROLLING OFFICE NAME & ADDRESS<br>Office of Naval Research<br>Statistics and Probability Program<br>Arlington, Virginia 22217 | | 12. REPORT DATE<br>(11) Feb[redacted]78<br>13. NUMBER OF PAGES<br>7 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS (of this report)<br>Unclassified<br>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

Almost sure representation, empirical distribution function.

20. ABSTRACT

Under certain conditions a sample intermediate order statistic from a sequence of independent and identically distributed random variables has an almost sure representation involving the empirical distribution and a remainder term of small order. In this paper an asymptotic approximation of the second moment of the remainder term is obtained. It is assumed that the marginal distribution function of the independent and identically distributed sequence has a finite left endpoint.

400 277 Gu