Ann. Inst. Statist. Math.
Vol. 49, No. 1, 117–121 (1997)

A NOTE ON MAXIMUM VARIANCE OF ORDER STATISTICS FROM SYMMETRIC POPULATIONS

NICKOS PAPADATOS

Department of Mathematics, Section of Statistics and Operations Research, University of Athens, Panepistemiopolis, 157 84 Athens, Greece

(Received March 22, 1995; revised October 2, 1995)

Abstract. The maximum variance of order statistics from a symmetrical parent population is obtained in terms of the population variance. The proof is based on a suitable representation for the variance of order statistics in terms of the parent distribution function.

Key words and phrases: Order statistics, maximum variance, symmetric populations.

1. Introduction

Let $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ be the order statistics corresponding to the sample X_j, $j = 1, 2, \ldots, n$, from an arbitrary df F with finite variance σ^2. Yang (1982) proved that the variance of the sample median cannot exceed σ^2 and Papadatos (1995) extended this result to any order statistic showing that

$$\operatorname{Var}[X_{i:n}] \le \sigma_n^2(i) \cdot \sigma^2, \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

where $\sigma_n^2(i)$ are given constants depending only on i and n (of course, Yang's result is equivalent to $\sigma_{2n-1}^2(n) = 1$; see Papadatos (1995) for more details).

Although the bounds (1.1) are attainable, they may be too large compared with the exact values of a specific df. Thus, in order to improve the results, we have to impose further conditions on the parent df F; a natural simple condition is the symmetry of F about some point c. In the remainder of this note we assume that F is symmetric about 0 (that is $F(x) = 1 - F(-x_-)$ for all x) and we derive the modified bounds for the variance of order statistics in this case (note that symmetry about c is equivalent to symmetry about 0, since the transformation $X \to X + c$ does not change the variance of order statistics). The proof of the main result (Theorem 2.1) is based on a suitable representation for the variance of order statistics (Lemma 2.1), combined with the monotone behavior of the special bivariate function D defined by (2.2) below (see Lemma 2.2 and Fig. 1).

2. Main result

Fix n and i with $(n+1)/2 < i < n$. Then we have the following

LEMMA 2.1.

(i) $\sigma^2 = 4 \iint_{0 \le x \le y} (1 - F(y)) dy dx.$

(ii) $\mathrm{Var}[X_{i:n}]$

$$= 2 \iint_{0 \le x \le y} \{[G(F(x)) + G(R(x))][1 - G(F(y)) - G(R(y))] + 2G(R(y))\} dy dx,$$

where $R(x) = 1 - F(x)$ and $G(x) = I_x(i, n+1-i)$ ($I_x(a,b)$ denotes, as usually, the incomplete beta function with parameters $a > 0$ and $b > 0$).

PROOF. (i) Follows immediately from the relation (c.f. Papadatos (1995))

$$\sigma^2 = 2 \iint_{x \le y} F(x)(1 - F(y)) dy dx \tag{2.1}$$

and the fact that $F(x) = 1 - F(-x)$ a.e. in $(-\infty, +\infty)$; note that (2.1) is an immediate consequence of Hoeffding's identity for the covariance of a random pair (c.f. Lehmann (1966), Lemma 2).

(ii) Follows, by using similar arguments, from

$$\mathrm{Var}[X_{i:n}] = 2 \iint_{x \le y} G(F(x))(1 - G(F(y))) dy dx,$$

since the df of $X_{i:n}$ is simply $G(F(x))$.

For x, y satisfying $\frac{1}{2} \le x \le y < 1$, consider the function (see Fig. 1)

$$D(x,y) = \frac{[G(x) + G(1-x)][1 - G(y) - G(1-y)] + 2G(1-y)}{2(1-y)} \tag{2.2}$$

and set $D(x) = D(x,x)$, $\frac{1}{2} \le x < 1$.

The following lemma is required for the main result.

LEMMA 2.2. (i) *If $x < y$ ($x \ge 1/2$) then*

$$D(x,y) < D(y).$$

(ii) $D(1/2) = 4G(1/2)(1 - G(1/2))$, $D(1_-) = 0$.

PROOF. (i) Since

$$\frac{\partial}{\partial x} D(x,y) = \frac{[g(x) - g(1-x)][1 - G(y) - G(1-y)]}{2(1-y)} > 0,$$

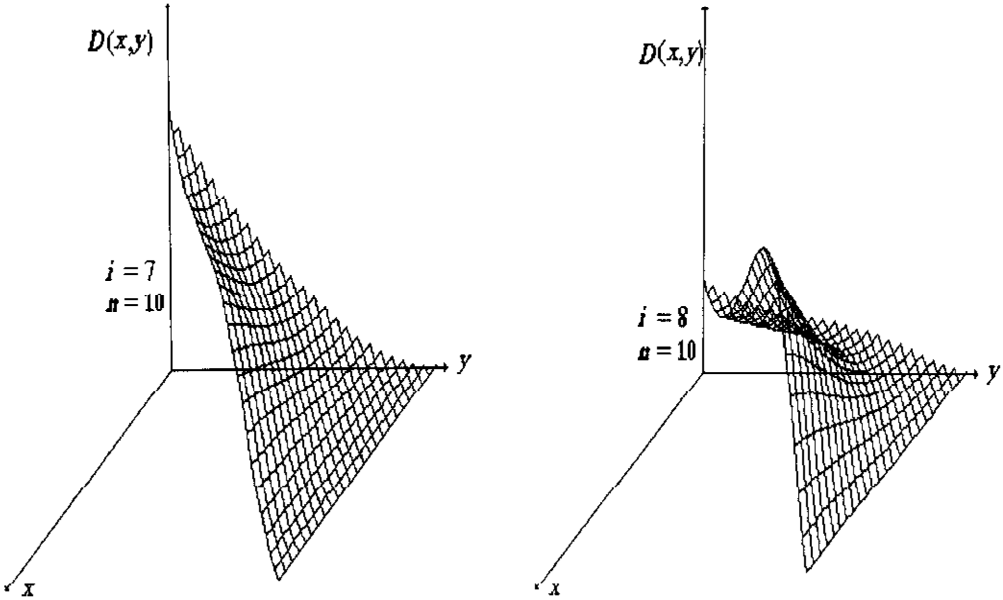

Fig. 1. Three dimensional graphs of the function $D(x,y)$ for $n = 10$ and $i = 7$ and 8.

for $x \in (1/2, y]$ (because it is easily verified that $g(x) > g(1-x)$ and $1 - G(y) > G(1-y)$ for all $x > 1/2$, $y > 1/2$), we conclude the desired result.

(ii) Follows immediately from the definition of $D(x)$.

Remark 1. From the continuity of $D(x)$ in $[1/2, 1]$ follows that D attains its maximum value for some point $y_0 \neq 1$ (this point y_0 needs not be unique). This observation, together with Lemma 2.2, leads to the inequality

$$[G(F(x)) + G(R(x))][1 - G(F(y)) - G(R(y))] + 2G(R(y)) \leq 2D(y_0) \cdot R(y), \tag{2.3}$$

for all $0 < x \leq y < +\infty$ (since $F(x) \geq 1/2$ for $x > 0$). Equality in (2.3) is equivalent to either $F(y) = 1$ or $F(x) = F(y) = y_0$, where y_0 is any (if more than one) of the maximizing points of D.

Remark 2. A point y_0 in $[1/2, 1)$ may be found as either a root of the equation $D'(x) = 0$ or $y_0 = 1/2$. Since $D'(x) = 0$ reduces to a polynomial equation of degree $2n-2$, there are at most $2n-1$ possible maximizing values of D (including $1/2$). These values may be calculated numerically in order to find the maximizing point(s) of D in $[1/2, 1)$.

Consider now any such point y_0 $(= y_0(i,n))$. Certainly the maximum value of D (i.e. $D(y_0)$) does not depend on the particular choice of y_0 and therefore we can prove the main result of

THEOREM 2.1.

$$\mathrm{Var}[X_{i:n}] \leq D(y_0) \cdot \sigma^2, \tag{2.4}$$

with equality if and only if

$$\text{(2.5)}\quad \mathrm{P}\left[X_j = \pm\frac{\sigma}{\sqrt{2(1-y_0)}}\right] = 1-y_0, \quad \mathrm{P}[X_j = 0] = 2y_0-1, \qquad j = 1,2,\ldots,n.$$

(If there exist k (>1) different values of y_0 (see Remarks 1 and 2), then (2.5) leads to k different df's, one for each y_0.)

PROOF. Integrating (2.3) over $S = \{0 < x \le y < +\infty\}$ and using Lemma 2.1 we have (2.4). In order to hold the equality in (2.4) it is necessary and sufficient that the equality in (2.3) holds a.e. in S; that is, either $F(y) = 1$ or $F(x) = F(y) = y_0$ for almost all $(x,y) \in S$. Taking into account the symmetry of F and the fact that $\mathrm{Var}[X_j] = \sigma^2$, we conclude (2.5) (since F is right continuous and non-decreasing) and the proof is complete. (It is not hard to show that for the equality in (2.4) it is necessary that the set $\{F(x), x \in [0,+\infty)\}$ takes *only* two values: the values 1 and y_0, where y_0 is any (fixed) maximizing point of D.)

Remark 3. The same bound (2.4) holds true for the variance of $X_{n+1-i:n}$, since for any symmetrical parent df we have $\mathrm{Var}[X_{i:n}] = \mathrm{Var}[X_{n+1-i:n}]$.

Remark 4. The case $y_0 = 1/2$ leads to a two-valued symmetric df. If $y_0 > 1/2$, then a three-valued parent df attains the maximum variance of order statistics.

3. Further remarks

The bounds (2.4) are, as expected, lower than the bounds (1.1) for any $i \neq (n+1)/2$ (of course, the case $i = (n+1)/2$ does not effect any improvement on Yang's bound σ^2, since the maximizing df in this case is already symmetric). However, the improvement is better as long as the distance between i and $(n+1)/2$ is large. In the following short table (Table 1) we give for illustration the sharp bounds for the variances of order statistics, both in the symmetric and in the completely arbitrary case (see also Fig. 2).

Table 1. Least upper bounds of $\mathrm{Var}[X_{i:10}]/\sigma^2$ for $i = 6,7,8,9^{(*)}$.

i	6	7	8	9
Upper bound (1.1) (any population)	1.0111	1.1096	1.3778	2.1608
Upper bound (2.4) (any symmetric population)	0.9394	0.5693	0.5463	0.9725

$^{(*)}$The values have been truncated to four decimal points. For $i = 6,7,8,9$, y_0 equals to 1/2, 1/2, 0.8008 and 0.9042, respectively.

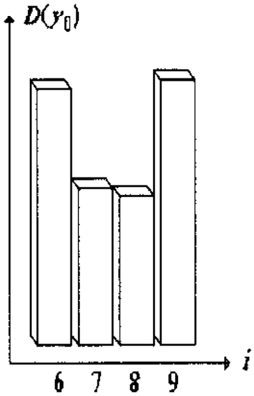

Fig. 2. Two dimensional chart for the values $D(y_0)$ $(n = 10)$.

For the extremes ($i = 1$ and $i = n$) the problem has been already treated by Moriguti (1951), who showed that the best upper bound for symmetrical parent population is $\mathrm{Var}[X_{n:n}] < (n/2) \cdot \sigma^2$ (note that the respective upper bound for any population is $n \cdot \sigma^2$).

The corresponding lower bounds are trivial for $1 < i < n$, taking the form $\mathrm{Var}[X_{i:n}] > 0$. However, these trivial bounds are the best possible, since for the symmetrical df F taking the values $-\sigma/\sqrt{2\epsilon}$, 0, $\sigma/\sqrt{2\epsilon}$ with probability ϵ, $1 - 2\epsilon$, ϵ, respectively, we have $\mathrm{Var}[X_{i:n}] \to 0$ as $\epsilon \to 0$, while the population variance remains σ^2. The non-trivial case $i = n$ (equivalently $i = 1$) treated by Moriguti, who showed (see also David (1981)) that

$$\mathrm{Var}[X_{n:n}] \geq \lambda_n \cdot \sigma^2,$$

where λ_n is some specific value (depending only on n) such that $\lambda_n = (\pi/2^n) \cdot [1 + O(1/n)]$ as $n \to \infty$. However, the previous formulas hold as long as F is symmetric, since the best lower bound for the variance of the extremes from an arbitrary (not necessarily symmetric) population is again the trivial 0.

References

David, H. A. (1981). *Order Statistics*, 2nd ed., Wiley, New York.

Lehmann, E. L. (1966). Some concepts of dependence, *Ann. Math. Statist.*, **37**, 1137–1153.

Moriguti, S. (1951). Extremal properties of extreme value distributions, *Ann. Math. Statist.*, **22**, 523–536.

Papadatos, N. (1995). Maximum variance of order statistics, *Ann. Inst. Statist. Math.*, **47**, 185–193.

Yang, H. (1982). On the variances of median and some other order statistics, *Bull. Inst. Math. Acad. Sinica*, **10**, 197–204.