The Annals of Statistics
1975, Vol. 3, No. 4, 955–958

BOUNDS ON THE VARIANCE OF THE U-STATISTIC FOR SYMMETRIC DISTRIBUTIONS WITH SHIFT ALTERNATIVES

BY BRUCE MCK. JOHNSON

The University of Connecticut

Sharp bounds on the variance of the Wilcoxon–Mann–Whitney U-statistic are obtained for the case of symmetric distributions and shift alternatives.

1. Summary. The two-sample problem with symmetric distributions and shift alternatives ($H_0: F_2 = F_1$, $H_A: F_2 = F_1(\cdot - \delta)$, F_1 symmetric) is a model which is often encountered in practice. In this note we obtain sharp bounds on the variance of the Wilcoxon–Mann–Whitney statistic U, as a function of $p = \int (1 - F_1(x))\, dF_2(x)$. These bounds are more restrictive than those obtained by Birnbaum and Klose (1957) for the more general model of F_1 and F_2 stochastically comparable.

2. Introduction and results. The variance of the statistic U depends on two parameters, γ^2 and ϕ^2 in addition to p and the two sample sizes m and n. However, in the case under consideration $\gamma^2 = \phi^2$. Let $c = \int_{-\infty}^{\infty} (1 - F(x + \delta))\, dF^2(x)$, for $\delta > 0$, say. Then if $\int_{-\infty}^{\infty} (1 - F(x + \delta))\, dF(x) = p$, $c = \gamma^2 + p^2$ in the usual notation, and $\sigma^2(U) = mn[(m + n - 2)(c - p^2) + (1 - p)p]$. The bounds on $\sigma^2(U)$ are obtained from corresponding bounds on c. Let S be the class of continuous symmetric distributions and S_0 the subclass of S with unimodal densities.

THEOREM. *If* $\int_{-\infty}^{\infty} (1 - F(x - \delta))\, dF(x) = p \in (0, \frac{1}{2})$ *and* F *is restricted to* S_0 *or to* S, *then*

$$\frac{1}{3}p^{\frac{3}{2}} \leqq c < p - \frac{1}{6}(1 - (1 - 2p)^{\frac{3}{2}})\,. \tag{1}$$

The lower bound is achieved for F rectangular. The upper bound while not achieved is sharp.

3. Proof of the theorem. To establish the validity of (1) it will suffice to consider only distributions F with support a finite interval and which are continuous and strictly increasing thereon. We introduce classes of functions $S^*(S_0^*)$ on $[0, 1]$ whose members are

$$h(u) = F(F^{-1}(u) + \delta) - u\,,$$

for $F \in S(S_0)$, and $\delta > 0$. Then

$$\begin{aligned} p &= \textstyle\int_{-\infty}^{\infty} (1 - F(x + \delta))\, dF(x) = \frac{1}{2} - \int_0^1 h(u)\, du \\ c &= \textstyle\int_{-\infty}^{\infty} (1 - F(x + \delta))\, dF^2(x) = \frac{1}{3} - 2\int_0^1 h(u)u\, du \end{aligned} \tag{2}$$

Received March 1973; revised December 1974.
AMS 1970 *subject classifications.* Primary 62; Secondary 70.
Key words and phrases. Variance, U-statistic.

and the original problem is equivalent to evaluating the infimum and the supremum of

$$\{\textstyle\int_0^1 h(u)u\,du : h \in S^*(S_0^*),\ \int_0^1 h(u)\,du = q\} \tag{3}$$

for fixed $q \in (0, \frac{1}{2})$.

The following properties of $S^*(S_0^*)$ are required. If $h \in S^*$ and $h \neq 1 - u$

(i) h is nonnegative and continuous, $h(u) + u$ is nondecreasing, and $h(u_2) \geqq h(u_1) - (u_2 - u_1)$ for $0 \leqq u_1 \leqq u_2 \leqq 1$.

(ii) $u = (1 - h(u))/2$ has a unique solution u^*. If $h \in S_0^*$, $h(u^*) = \max h$ and h is monotone to either side of u^*.

(iii) For $0 \leqq u < u^*$, $u^* < 1 - u - h(u) \leqq 1 - h(0)$ and $h(1 - u - h(u)) = h(u)$.

(iv) $h(0) > 0$ and $h(u) = 1 - u$ for $1 - h(0) \leqq u \leqq 1$.

Of the above, (i) and (ii) are obvious, (iii) follows from the symmetry of F, and (iv) follows from F having compact support and (iii).

Let u^* be as in (ii). Then

$$\int_0^1 h(u)\,du = \int_0^{u^*} h(u)\,du + \int_{u^*}^{1-h(0)} h(u)\,du + \int_{1-h(0)}^1 h(u)\,du$$

and with the change of variable, $u = 1 - v - h(v)$, in the second (first) integral it follows from (iii) that

$$\begin{aligned} \int_0^1 h(u)\,du &= 2\int_0^{(1-b)/2} h(u)\,du + b^2/2 \\ &= 2\int_{(1-b)/2}^1 h(u)\,du - b^2/2\,. \end{aligned} \tag{4}$$

where $u^* = (1 - b)/2$ and $h(u^*) = b$. A similar calculation yields

$$\begin{aligned} \int_0^1 h(u)u\,du &= \tfrac{1}{2}\int_0^1 h(u)\,du - \tfrac{1}{2}\int_0^{(1-b)/2} h^2(u)\,du - b^3/12 \\ &= \tfrac{1}{2}\int_0^1 h(u)\,du - \tfrac{1}{2}\int_{(1-b)/2}^1 h^2(u)\,du + b^3/12\,. \end{aligned} \tag{5}$$

We first consider the case of $h \in S_0^*$, or more precisely the subclass of S^* for which $\max h = h(u^*)$. The scheme is to bound the supremum and infimum of

$$\{\textstyle\int_0^{(1-b)/2} h^2(u)\,du;\ h \in S_0^*,\ u^* = (1-b)/2,\ 2\int_0^{(1-b)/2} h(u)\,du + b^2/2 = q\} \tag{6}$$

and

$$\{\textstyle\int_{(1-b)/2}^1 h^2(u)\,du;\ h \in S_0^*,\ u^* = (1-b)/2,\ 2\int_{(1-b)/2}^1 h(u)\,du - b^2/2 = q\} \tag{7}$$

respectively for fixed b and then to make the resulting bounds on c, via (2) and (5), extreme by varying b. As $0 \leqq h(u) \leqq b$ it is obvious from the integral condition of (6) and from (4) that both sets are vacuous for $b \notin [1 - (1 - 2q)^{\frac{1}{2}}, (2q)^{\frac{1}{2}})$. For b in this interval define

$$\begin{aligned} h_1 &= 0\,, \quad 0 \leqq u < u_1 \\ &= b\,, \quad u_1 \leqq u \leqq (1-b)/2\,, \end{aligned}$$

where $u_1 = \frac{1}{2} - q/2b - b/4$, and

$$\begin{aligned} h_2 &= (1+b)/2 - u\,, && (1-b)/2 \leqq u \leqq u_2 \\ &= (1+b)/2 - u_2\,, && u_2 \leqq u \leqq 1 - b/2 + u_2 \\ &= 1 - u\,, && 1 - b/2 + u_2 \leqq u \leqq 1\,, \end{aligned}$$

where $u_2 = (\frac{1}{2} - q)/(1-b)$. It is easily checked that the h_i satisfy the integral condition of the appropriate set and $h_i \leqq b$. Neglecting the question of whether the h_i are restrictions of members of $S_0{}^*$ we assert that

$$\int_0^{(1-b)/2} h_1{}^2(u)\,du = bq/12 - b^3/4$$

is an upper bound for (6) and that

$$\int_{(1-b)/2}^1 h_2{}^2(u)\,du = \tfrac{1}{2}(q - b^2/2)^2/(1-b) + b^3/3$$

is a lower bound for (7). An elementary variational argument, essentially $(s+t)^2 \geqq s^2 + t^2$ for $s, t \geqq 0$, and the condition $h \leqq b$ establishes the first assertion. The same variational argument coupled with $h(u) \geqq b - (u - (1-b)/2)$, $u \geqq (1-b)/2$, which follows from (i), yields the second.

Accordingly, for $1 - (1-2q)^{\frac{1}{2}} \leqq b < (2q)^{\frac{1}{2}}$,

$$(8) \qquad \tfrac{1}{3} - q + (q - b^2/2)/2(1-b) + b^3/6 \leqq c \leqq \tfrac{1}{3} - q + bq/2 - b^3/12\,.$$

The derivative of each side is positive and taking $b = (2q)^{\frac{1}{2}}$ on the right and $b = 1 - (1-2q)^{\frac{1}{2}}$ on the left, with $q = \frac{1}{2} - p$, gives (1).

It remains to show that (1) is sharp. For $b = 1 - (1-2q)^{\frac{1}{2}}$,

$$\begin{aligned} h_2 &= 1 - (1-2q)^{\frac{1}{2}} && (1-2q)/2^{\frac{1}{2}} \leqq u \leqq (1-2q)^{\frac{1}{2}} \\ &= 1 - u && (1-2q)^{\frac{1}{2}} \leqq u \leqq 1 \end{aligned}$$

is the restriction of the $S_0{}^*$ function corresponding to $\delta = 1$ and F uniform on $[-2(1-(1-2q)^{\frac{1}{2}}), 2(1-(1-2q)^{\frac{1}{2}})]$.

For $1 - (1-2q)^{\frac{1}{2}} < b < (2q)^{\frac{1}{2}}$, h_1 is not the restriction of a member of $S_0{}^*$. However, for such b and α sufficiently small

$$\begin{aligned} h_\alpha &= \alpha\,, && 0 \leqq u \leqq u_1 - \alpha \\ &= \alpha + (u - u_1 + \alpha)(b - \alpha)/\alpha\,, && u_1 - \alpha \leqq u \leqq u_1 \\ &= b\,, && u_1 \leqq u \leqq (1-b)/2\,, \end{aligned}$$

where $u_1 = \alpha/2 + (b(1 - b/2) - q)/2(b - \alpha)$, is the restriction of the $S_0{}^*$ function corresponding to $\delta = 1$ and F having density b on $|x| \leqq (1 - 2u_1)/2b$, α on $(1-2u_1)/2b < |x| \leqq (1-2u_1)/2b + u_1/\alpha$, and 0 elsewhere. For fixed b, $h_\alpha \to h_1$ a.e. on $[0, (1-b)/2]$ as $\alpha \to 0$. Thus the upper bound cannot be improved. If there exists $F \in S_0$ and $\delta > 0$ for which the bound is achieved, then there are F_n, of the type we are considering, converging weakly to F for which

$$\begin{aligned} \tfrac{1}{3} - 2\int_0^1 h_n(u)u\,du &= \int (1 - F_n(x+\delta))\,dF_n{}^2(x) \to \tfrac{1}{6}(1 - (2q)^{\frac{3}{2}}) \\ &= \int (1 - F(x+\delta))\,dF^2(x) \end{aligned}$$

and $\int_0^1 h_n(u)\,du \to q$. Let u_n^* be as in (ii) and $h_n(u_n^*) = b_n$. As the derivative of the right side of (8) is positive, $b_n \to (2q)^{\frac{1}{2}}$ and $h_n(u) \to 0$ for $0 \leqq u < (1-(2q)^{\frac{1}{2}})/2$. Now, δ fixed and $h_n(u) \to 0$ easily imply $F_n^{-1}(u) \to -\infty$ for such u, a contradiction. Actually, the assumption that F is continuous may be dropped.

A similar argument applies for $h \in S^*$. From the last assertion of (i) and the same variational argument used above it is clear that

$$\sup\{\textstyle\int_0^{(1-b)/2} h^2(u)\,du;\ h \in S^*,\ u^* = (1-b)/2,\ 2\int_0^{(1-b)/2} h(u)\,du + b^2/2 = q\}$$
$$\leqq \int_0^{(1-b)/2} h_1^2(u)\,du\,,$$

where

$$\begin{aligned} h_1(u) &= 0 & 0 &\leqq u \leqq (1+b)/2 - (b^2/2+q)^{\frac{1}{2}} \\ &= (1+b)/2 - u\,, & (1+b)/2 - (b^2/2+q)^{\frac{1}{2}} &< u \leqq (1-b)/2\,. \end{aligned}$$

For $b < (2q)^{\frac{1}{2}}$ and such that the integral condition is satisfied

$$c \leqq \tfrac{1}{3} - q + \tfrac{1}{3}((b^2/2+q)^{\frac{3}{2}} - b^3/2)\,.$$

A calculation shows the right side to be increasing in b and to have the same value as the right side of (8) at $b = (2q)^{\frac{1}{2}}$. As in the preceding case the bound is not achieved.

If a lower bound which minorizes that obtained in the S_0^* case is possible it must result from $h \in S^*$ with $\max h > h(u^*)$. On the other hand, for fixed b it is evident from the variational argument that $\max h$ should be as small as is possible. This with the growth condition of (i) imply we need only consider h with restrictions to $[0, (1-b)/2]$ of the form

$$\begin{aligned} h_2 &= (1+b)/2 - ((1+b)/2)^2 - q - b^2/2)^{\frac{1}{2}}\,, & 0 &\leqq u \leqq ((1+b)/2)^2 - q - b^2/2)^{\frac{1}{2}} \\ &= (1+b)/2 - u\,, & ((1+b)/2)^2 - q - b^2/2)^{\frac{1}{2}} &\leqq u \leqq (1-b)/2 \end{aligned}$$

for $b \leqq 1 - (1-2q)^{\frac{1}{2}}$. If $b > 1 - (1-2q)^{\frac{1}{2}}$ there are $h \in S^*$ with $\max h = b$ for which the integral of interest minorizes that of any h' with $h'(u^*) = b < \max h'$. For such h_2,

$$\begin{aligned} c = \tfrac{1}{3} - q &+ ((1+b)/2 - ((1+b)/2)^2 - q - b^2/2)^{\frac{1}{2}})^2((1+b)/2)^2 - q - b^2/2)^{\frac{1}{2}} \\ &+ \tfrac{1}{3}(1+b)/2 - ((1+b)/2)^2 - q - b^2/2)^{\frac{1}{2}})^3 - b^3/6\,. \end{aligned}$$

A straightforward calculation shows c is minimized at $b = 1 - (1-2q)^{\frac{1}{2}}$, that is, for $h_2 \equiv 1 - (1-2q)^{\frac{1}{2}}$ on $[0, (1-2q)^{\frac{1}{2}}/2]$. Thus the lower bound for S_0^* is valid for S^*.

REFERENCES

BIRNBAUM, Z. W. and KLOSE, U. M. (1957). Bounds for the variance of the Mann-Whitney statistics. *Ann. Math. Statist.* **28** 933-945.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
STORRS, CONNECTICUT 06268