ESTIMATES OF LOCATION BASED ON RANK TESTS

By J. L. Hodges, Jr.[1] and E. L. Lehmann[2]

University of California, Berkeley

1. Introduction and summary. A serious objection to many of the classical statistical methods based on linear models or normality assumptions is their vulnerability to gross errors. For certain testing problems this difficulty is successfully overcome by rank tests such as the two Wilcoxon tests or the Kruskal-Wallis H-test. Their power is more robust against gross errors than that of the t- and F-tests, and their efficiency loss is quite small even in the rare case in which the suspicion of the possibility of gross errors is unfounded.

For the corresponding problems of point estimation a beginning has been made to attack the difficulty by modifying the classical estimates either through removal or Winsorization of outlying observations; see for example Tukey (1960) and Anscombe (1960). It is the purpose of the present paper to explore a different approach to these problems of point estimation. In Sections 2–5 point estimates of location or shift parameter are defined in terms of rank test statistics such as the Wilcoxon or normal scores statistic, which are successful in providing robust power for the corresponding testing problems. In Sections 6 and 7, certain regularity and invariance properties of these estimates are proved. The distributions of the estimates are shown in Section 8 to be symmetric with respect to the parameter being estimated—and hence in particular to be unbiased—if the underlying distribution of the observations on which the estimate is based is symmetric. Without this assumption, the estimates are shown in Section 9 to be either exactly or approximately median unbiased for small samples and in Section 10 to be approximately normally distributed about the true parameter value for large samples. The variance of this asymptotic distribution depends of course on the underlying distribution of the observations, so that the estimates are not "distribution-free." In Section 9 there is also established a close relationship between the estimates and the corresponding upper and lower confidence bound for the parameter at confidence level $\frac{1}{2}$, with which the estimate coincide in many cases. Finally, in Section 11, it is proved that the asymptotic relative efficiency of the estimates to the classical linear estimates is the same as the Pitman efficiency of the rank tests on which they are based to the corresponding t-tests.

2. Point estimates based on test statistics. Let $X_1, \cdots, X_m; Y_1, \cdots, Y_n$ be independent random variables with distributions

$$(2.1) \qquad P\{X_i \leq u\} = F(u); \qquad P\{Y_j \leq u\} = F(u - \Delta).$$

Since the variables $X_1, \cdots, X_m$, and the variables $Y_1 - \Delta, \cdots, Y_n - \Delta$ obtained by shifting the Y-sample Δ to the left, are independently, identically distributed, it is natural to estimate Δ by the amount of shift needed to align a closely as possible the two sets $(X_1, \cdots, X_m)$ and $(Y_1 - \Delta, \cdots, Y_n - \Delta)$. A definition of alignment could for example be given with reference to the Wilcoxon statistic, by defining the two sets to be aligned if half of the non-zero differences $(Y_j - \Delta) - X_i$ are positive and half negative. There is either a unique such value of Δ, which would then serve as estimate, or an interval of such values; in the latter case, the midpoint of this interval provides a natural estimate. More generally, if a test of the hypothesis $\Delta = 0$ is based on a statistic whose distribution is symmetric about a point μ, the two sets could be defined to be in alignment when giving to the test statistic the value μ. To formalize these considerations let us either assume that F is known to belong to the class $\mathfrak{F}_0$ of all continuous distributions, or that it is known to belong to the class $\mathfrak{F}_1$ of all continuous distributions that are symmetric about zero. Consider a test statistic

$$h = h(X_1, \cdots, X_m; Y_1, \cdots, Y_n)$$

for the hypothesis $H:\Delta = 0$ against the alternatives $\Delta > 0$. We shall assume throughout that

(A) $h(x_1, \cdots, x_m, y_1 + a, \cdots, y_n + a)$ is a nondecreasing function of a for all x and y,

(B) when $\Delta = 0$, the distribution of $h(X_1, \cdots, X_m; Y_1, \cdots, Y_n)$ is symmetric about a fixed point μ (independent of F), (i) for all $F \varepsilon \mathfrak{F}_0$, or (ii) for all $F \varepsilon \mathfrak{F}_1$.

We shall use the notation $x = (x_1, \cdots, x_m)$ and $y = (y_1, \cdots, y_n)$ with the obvious conventions. Thus $x < x'$ means that the inequality holds for each coordinate; if a is a real number, then $x + a = (x_1 + a, \cdots, x_n + a)$; etc. The notation $P_0\{\cdot\}$ will be used to indicate that the probability in question is being computed for the case $\Delta = 0$. Let

(2.2) $\Delta^* = \sup\{\Delta:h(x, y - \Delta) > \mu\}$ and $\Delta^{**} = \inf\{\Delta:h(x, y - \Delta) < \mu\}$

and let

(2.3) $$\hat{\Delta} = (\Delta^* + \Delta^{**})/2.$$

For suitable functions h we propose $\hat{\Delta}$ as estimate of the shift parameter Δ.

As a second problem suppose that $Z_1, \cdots, Z_N$ are independently distributed with common distribution

(2.4) $$P\{Z_i \leqq u\} = F(u - \theta)$$

where F is continuous and symmetric about zero.

Considerations similar to those in the two-sample problem suggest basing an estimate on a test statistic $h = h(Z_1, \cdots, Z_N)$ for the hypothesis $\theta = 0$ against the alternatives $\theta > 0$. We shall assume throughout that

(C) $h(z_1 + a, \cdots, z_N + a)$ is a nondecreasing function of a for each z,

and

(D) for $\theta = 0$, the distribution of h is symmetric about a fixed point μ (independent of F) for all $F \varepsilon \mathfrak{F}_1$.

If μ is the median of $h(Z)$ when $\theta = 0$, let

$$(2.5) \quad \theta^* = \sup \{\theta : h(z - \theta) > \mu\} \quad \text{and} \quad \theta^{**} = \inf \{\theta : h(z - \theta) < \mu\}$$

and

$$(2.6) \qquad\qquad \hat{\theta} = (\theta^* + \theta^{**})/2.$$

Then we propose $\hat{\theta}$ as the estimate of the location parameter θ.

3. A class of estimates for the two-sample problem. We shall be concerned primarily with estimates based on rank tests. An important class of rank statistics for the two-sample problem is given by

$$(3.1) \qquad\qquad h(x, y) = \sum_{j=1}^{n} E_\Psi[V^{(s_j)}]$$

where $s_1, \cdots, s_n$ denote the ranks of $y_1, \cdots, y_n$ in the combined sample, and where $V^{(1)} < \cdots < V^{(m+n)}$ denote an ordered sample of size $m + n$ from a distribution Ψ.

The function h defined by (3.1) clearly satisfies requirement (A) of the preceding section. Conditions under which h satisfies (B) are given by the following lemma, in which h is not assumed to satisfy (3.1).

LEMMA 1. *The distribution of $h(X, Y)$ is symmetric about μ if any one of the following three conditions holds*:

(i) *h is a function only of the ranks and satisfies*

$$(3.2) \qquad\qquad h(x, y) + h(-x, -y) = 2\mu \qquad\qquad (\text{a.e. } P_0)$$

(ii) *the sample sizes m and n are equal, and h satisfies*

$$(3.3) \qquad\qquad h(x, y) + h(y, x) = 2\mu \qquad\qquad (\text{a.e. } P_0)$$

(iii) *the distribution F is symmetric about zero, and h satisfies (3.2).*

PROOF. (i) If $h(x, y) = g(s_1, \cdots, s_n)$, then $h(-x, -y) = g(m + n - s_n, \cdots, m + n - s_1)$ and $P_0\{h(X, Y) < \mu - a\} = P_0\{h(-X, -Y) > \mu + a\} = P_0\{g(m + n - S_n, \cdots, m + n - S_1) > \mu + a\} = P_0\{g(S_1, \cdots, S_n) > \mu + a\} = P_0\{h(X, Y) > \mu + a\}$. Here P_0 indicates that the probabilities are computed for $\Delta = 0$; the first equality follows from (3.2); the third equality follows from the fact that $(S_1, \cdots, S_n)$ and $(m + n - S_n, \cdots, m + n - S_1)$ have the same joint distribution when $\Delta = 0$.

(ii) When $m = n$ and $\Delta = 0$, the vectors (X, Y) and (Y, X) have the same joint distribution, and hence $P_0\{h(X, Y) < \mu - a\} = P_0\{h(Y, X) < \mu - a\} = P_0\{h(X, Y) > \mu + a\}$ where the second equality follows from (3.3).

(iii) This follows from the first equality in the proof of (i) and the fact that (X, Y) and $(-X, -Y)$ have the same distribution when F is symmetric about zero.

Conditions under which the function h defined by (3.1) satisfies (3.2) or (3.3) are given by the following lemma.

LEMMA 2. *Let h be defined by (3.1). Then*

(i) *if Ψ is symmetric about b, the function h satisfies (3.2) with $\mu = nb$*

(ii) *if $m = n$ and b denotes the expectation of Ψ the function h satisfies (3.3) with $\mu = \frac{1}{2}(m + n)b$.*

PROOF.

(i) $h(x, y) + h(-x, -y) = \sum E_\Psi[V^{(s_j)}] + \sum E_\Psi[V^{(m+n-s_j)}]$

$$= \sum E_\Psi[V^{(s_j)}] + \sum E_\Psi[2b - V^{(s_j)}] = 2nb.$$

(ii) $h(x, y) + h(y, x) = \sum E_\Psi[V^{(s_j)}] + \sum E_\Psi[V^{(r_i)}]$ where $r_1, \cdots, r_m$ de-denotes the ranks of the x's. The right hand side is equal to the sum of the expectations of all the V's and hence to $(m + n)b$.

It follows from Lemmas 1 and 2 that a function h given by (3.1) satisfies Condition B(i) of the preceding section if either Ψ is symmetric or the two sample sizes are equal.

Among the statistics given by (3.1) and satisfying B(i), we shall be particularly interested in the Wilcoxon statistic and the normal scores statistic obtained by taking for Ψ a rectangular or normal distribution respectively. The resulting estimates will be considered in more detail in the next section.

Suppose next that $m = n$. Let us denote by $^-$ the average of the indicated variables and by $^\sim$ or med their median, so that for example

$$(3.4) \qquad \bar{x} = (x_1 + \cdots + x_m)/m$$

and

$$(3.5) \qquad \tilde{x} = \text{med } x = \begin{cases} x^{(k+1)} & \text{if } m = 2k + 1 \\ [x^{(k)} + x^{(k+1)}]/2 & \text{if } m = 2k \end{cases}$$

where $x^{(1)} < \cdots < x^{(m)}$ denote the ordered x's. Then both $h(x, y) = \bar{y} - \bar{x}$ and $h(x, y) = \tilde{y} - \tilde{x}$ satisfy (3.3) with $\mu = 0$ and $\Delta^* = \Delta^{**} = h$. The resulting estimates are therefore $\bar{Y} - \bar{X}$ and $\tilde{Y} - \tilde{X}$ respectively.

More generally suppose that in addition to (3.3), h satisfies the invariance relation

$$(3.6) \qquad h(x, y + a) = h(x, y) + a \quad \text{for all real } a.$$

Assume without loss of generality that $\mu = 0$ since the function $h'(x, y) = h(x, y - \mu)$ satisfies (3.3) with $\mu = 0$. Condition (3.6) then implies that $\Delta^* = \Delta^{**} = h$ since for example

$$\Delta^{**}(x, y) = \inf\{\Delta : h(x, y - \Delta) < 0\} = \inf\{\Delta : h(x, y) < \Delta\} = h(x, y).$$

4. Estimates based on the Wilcoxon and normal scores statistics. Let $h(x, y)$ be the number of pairs (i, j) such that $x_i < x_j$ $(1 \leq i \leq m; 1 \leq j \leq n)$. The test based on this statistic is the Wilcoxon two-sample test in the Mann-Whitney form; it is equivalent to the test based on (3.1) with Ψ the rectangular distribu-

tion on $(0, 1)$. The possible values of the function h, which satisfies assumption
B(i), are the integers, $0, 1, \cdots, mn$. To find an explicit expression for the esti-
mate $\hat{\Delta}$ obtained from h through (2.2) and (2.3) denote by $W^{(1)} < \cdots < W^{(mn)}$
the ordered differences $Y_j - X_i$. It is convenient to distinguish two cases.

Suppose first that mn is odd, $mn = 2k + 1$, say. Then $\mu = k + \frac{1}{2}$, which is
not a possible value of h, and

$$\Delta^{**} = \inf\{\Delta : h(x, y - \Delta) < k + \tfrac{1}{2}\}$$

$$= \inf\{\Delta : \text{Fewer than } k + \tfrac{1}{2} \text{ of the differences } Y_j - X_i \text{ exceed } \Delta\}$$

$$= \inf\{\Delta : W^{(k+1)} \leq \Delta\} = W^{(k+1)}.$$

Similarly,

$$\Delta^* = \sup\{\Delta : \text{More than } k + \tfrac{1}{2} \text{ of the differences } Y_j - X_i \text{ exceed } \Delta\}$$

$$= \sup\{\Delta : W^{(k+1)} \leq \Delta\} = W^{(k+1)},$$

so that $\hat{\Delta} = W^{(k+1)}$.

On the other hand, if mn is even, $mn = 2k$ say, then

$$\Delta^{**} = \inf\{\Delta : W^{(k+1)} \leq \Delta\} = W^{(k+1)}$$

$$\Delta^* = \sup\{\Delta : W^{(k)} > \Delta\} = W^{(k)}$$

and hence $\hat{\Delta} = [W^{(k)} + W^{(k+1)}]/2$.

Thus in both cases,

(4.1) $\hat{\Delta} = \operatorname{med}(Y - X)$

is the median of the set of mn differences $Y_j - X_i$.

Formula (4.1) will frequently not be the most convenient way of computing $\hat{\Delta}$
in practice. To illustrate how this can be done quite expeditiously consider the
case $mn = 2k + 1$. Then k of the differences $(Y_j - \hat{\Delta}) - X_i$ are greater than,
k less than, and one is equal to, zero. To obtain $\hat{\Delta}$, plot the two samples on two
separate strips of paper. Sliding the Y-strip to the left decreases by one the
number of positive differences $Y_j - X_i$ each time a Y moves over an X. By pro-
ceeding in this way, one rapidly finds the position in which k of the differences are
positive, k negative, and one is zero. The difference of the origins of the two
strips in this position is $\hat{\Delta}$. If $mn = 2k$, one finds in a similar way the shift Δ^*
that produces k positive, $k - 1$ negative, and one zero difference, and the shift
Δ^{**} that produces $k - 1$ positive, k negative, and one zero difference, and hence
the estimate $\hat{\Delta} = (\Delta^* + \Delta^{**})/2$.

A somewhat similar procedure can be used to obtain the estimate based on the
normal scores statistic $h = \sum E_\Phi[V^{(s_j)}]$. Suppose the y-strip is slid to the left so
that $Y^{(j)}$ moves across an X. Then s_j is decreased by 1 so that h is decreased by
$E_\Phi[V^{(s_j)}] - E_\Phi[V^{(s_j-1)}]$. Starting out with a position of the two strips which gives
to h an approximately central value, the desired position can quickly be found
using a table of expected normal order statistics.

5. Some estimates for the one-sample problem. Let $Z_1, \cdots, Z_N$ be identically distributed, their distribution given by (2.1) with F symmetric about zero. Let $s_1, \cdots, s_n$ denote the ranks of the positive Z's among the N absolute values $|Z_1|, \cdots, |Z_N|$. Here n is a random variable which for $\Delta = 0$ has the binomial distribution $Bi(N, \frac{1}{2})$. A class of rank tests is based on the test statistic

$$(5.1) \qquad h(z) = \sum_{j=1}^{n} E_{\Psi}[V^{(s_j)}]$$

where $V^{(1)} < \cdots < V^{(N)}$ denote the ordered absolute values of a sample of size N from a distribution Ψ, so that they constitute an ordered sample from the distribution given by $\Psi^*(u) = \Psi(u) - \Psi(-u)$ for $u > 0$.

The function h defined by (5.1) clearly satisfied assumption (C) of Section 2. As in Lemma 1(iii) it is seen that a function h satisfies requirement (D) if

$$(5.2) \qquad h(z) + h(-z) = 2\mu \qquad \qquad \text{(a.e. } P_0) \,.$$

It follows from (5.2) that any function h given by (3.1) satisfies requirement (D). We have in fact

LEMMA 3. *If h is given by (5.1) and $\theta = 0$, the distribution of h is symmetric about $\mu = \frac{1}{2}NE_{\Psi}|Z_1|$ for all $F \, \varepsilon \, \mathfrak{F}_1$* .

PROOF. Let $r_1, \cdots, r_m$ denote the ranks of the negative Z's among $|Z_1|, \cdots, |Z_N|$. Then $h(z) + h(-z) = \sum E_{\Psi}[V^{(s_j)}] + \sum E_{\Psi}[V^{(r_i)}] = NE_{\Psi}|Z_1|$.

An important special case is again that of the Wilcoxon test statistic corresponding to the choice of a rectangular distribution for Ψ. To obtain an explicit expression for the estimate θ in this case, it is convenient to use the equivalent form of the test statistic due to Tukey (1949), namely

$$(5.3) \quad h(z) = \text{Number of pairs } (i,j) \text{ with } 1 \leq i \leq j \leq N \text{ such that } z_i + z_j > 0.$$

The possible values of h are the integers $0, 1, \cdots, N(N+1)/2$. Let $W^{(1)} < \cdots < W^{(K)}$ be the $K = N(N+1)/2$ averages $(z_i + z_j)/2$ with $i \leq j$. Then it is seen quite similarly to the corresponding result in Section 4 that $\hat{\theta}$ is the median of the variables $W^{(1)}, \cdots, W^{(K)}$.

Another class of examples is obtained by taking for h a functon that satisfies (5.2) and is translation invariant in the sense that

$$(5.4) \qquad h(z + a) = h(z) + a \quad \text{for all real } a.$$

As in the corresponding examples for the two-sample problem discussed at the end of Section 3, one can assume without loss of generality that $\mu = 0$ and then finds $\hat{\theta}(z) = h(z)$.

Examples of this are (i) $h(z) = \bar{z}$ and (ii) $h(z) = \tilde{z}$. It is interesting to see that $\tilde{z}$ is also obtained as the estimate corresponding to quite a different function h. For let $h(z)$ be the number of positive z's, so that the test defined by h is the sign test. Then h satisfies (5.2) with $\mu = N/2$ and a computation similar to that of the Wilcoxon case then shows that

$$(5.5) \qquad \qquad \hat{\theta} = \tilde{Z}.$$

6. Regularity properties. In Section 2, estimates $\hat{\Delta}$ and $\hat{\theta}$ of a shift or location parameter were defined in terms of given test statistics, and these estimates were illustrated in Sections 3–5. The remainder of the paper is concerned with general properties of these estimates. In the present section, certain regularity properties are obtained; these are immediate consequences of the following theorem.

THEOREM 1. *Let h be a real-valued function on $(m + n)$-space such that $h(x_1, \cdots, x_m, y_1 + a, \cdots, y_n + a)$ is a non-decreasing function of a for all x and y, and let Δ^*, Δ^{**} be defined by (2.2). Suppose that $(X_1, \cdots, X_m, Y_1, \cdots, Y_n)$ is a random vector with joint distribution H. Then the distribution of Δ^* and Δ^{**} is (absolutely) continuous provided H is (absolutely) continuous.*

PROOF. For any fixed numbers $t_2, \cdots, t_n$, $h(x_1, \cdots, x_m, y_1, y_1 + t_2, \cdots, y_1 + t_n)$ is non-decreasing in y_1. Let $u(x_1, \cdots, x_m, t_2, \cdots, t_n)$ be such that $h(x_1, \cdots, x_m, y_1, y_1 + t_2, \cdots, y_1 + t_n) <$ or $\geqq \mu$ if y_1 is $<$ or $> u(x_1, \cdots, x_m; t_2, \ldots, t_n)$. Then

$$
\begin{aligned}
\Delta^{**}(x_1, &\cdots, x_m, y_1, y_1 + t_2, \cdots, y_1 + t_n) \\
&= y_1 - u(x_1, \cdots, x_m; t_2, \cdots, t_n).
\end{aligned}
$$
(6.1)

Suppose now first that H is continuous. Since

$$
\begin{aligned}
\Delta^{**}(x_1, &\cdots, x_m, y_1, \cdots, y_n) = c \Leftrightarrow \\
&y_1 = u(x_1, \cdots, x_m, y_2 - y_1, \cdots, y_n - y_1) + c,
\end{aligned}
$$
(6.2)

the set $S = \{(x_1, \cdots, x_m, y_1, \cdots, y_n) : \Delta^{**}(x_1, \cdots, x_m, y_1, \cdots, y_n) = c\}$ intersects each line

$$
\begin{aligned}
L(x_1^0, &\cdots, x_m^0, t_2^0, \cdots, t_n^0) : x_1 = x_1^0, \cdots, x_m = x_m^0, \\
&y_2 = y_1 + t_2^0, \cdots, y_n = y_1 + t_n^0
\end{aligned}
$$
(6.3)

in the single point whose y_1-coordinate is $u(x_1^0, \cdots, x_m^0, t_2^0, \cdots, t_n^0) + c$. Consider now the probability of the set S. Each section of S with one of the lines (6.3) consists of a single point and by the assumed continuity of H therefore has probability zero. Since S is measurable, it follows from Fubini's theorem that $P(S) = 0$, as was to be proved.

Suppose now that H is absolutely continuous. Let A be any set on the real line with Lebesgue measure zero and let

$$
S = \{(x_1, \cdots, x_m, y_1, \cdots, y_n) : \Delta^{**}(x_1, \cdots, x_m, y_1, \cdots, y_n) \varepsilon A\}.
$$

Since

$$
\begin{aligned}
\Delta^{**}(x_1, &\cdots, x_m, y_1, \cdots, y_n) \varepsilon A \\
&\Leftrightarrow y_1 - u(x_1, \cdots, x_m, y_2 - y_1, \cdots, y_n - y_1) \varepsilon A,
\end{aligned}
$$

the section of S with each line (6.3) has Lebesgue measure zero, and thus S itself has Lebesgue measure zero. It follows from the absolute continuity of H that $P(S) = 0$, as was to be proved.

In both cases, the proof for Δ^* is completely analogous. It follows therefore that the distribution of the estimate $\hat{\Delta}$ is (absolutely) continuous if F is (absolutely) continuous. The corresponding result for $\hat{\theta}$ follows by letting $m = 0$, $n = N$ in the lemma.

7. Invariance properties. A simple but useful property of the estimates $\hat{\Delta}$ and $\hat{\theta}$ is their translation invariance, namely

$$(7.1) \qquad \hat{\Delta}(x, y + a) = \hat{\Delta}(x, y) + a \quad \text{for all real } a$$

and

$$(7.2) \qquad \hat{\theta}(z + a) = \hat{\theta}(z) + a \quad \text{for all real } a.$$

These are immediate consequences of the definitions (2.2) and (2.5), which show that Δ^*, Δ^{**}, θ^* and θ^{**} are translation invariant.

It follows from (7.1) and (7.2) that

$$(7.3) \qquad P_\Delta(\hat{\Delta} - \Delta \leqq u) = P_0(\hat{\Delta} \leqq u)$$

and

$$(7.4) \qquad P_\theta(\hat{\theta} - \theta \leqq u) = P_0(\hat{\theta} \leqq u)$$

where the notation P_Δ and P_θ indicate, that the probabilities are computed assuming Δ and θ to be the true values of these parameters. Relations (7.3) and (7.4) show that when investigating distributional properties of the estimates, one may assume without loss of generality that $\Delta = 0$ or $\theta = 0$ respectively, since the distribution for the general case is obtained simply by translation.

Typically, the function h in the two-sample problem satisfies the invariance relation

$$(7.5) \qquad h(x + a, y + a) = h(x, y) \quad \text{for all } a.$$

This relation holds in particular for any rank test. It is obvious from the definition that (7.5) implies the corresponding relationship for $\hat{\Delta}$.

8. Symmetry properties. If $\hat{\Delta}$ and $\hat{\theta}$ are to be reasonable estimates of Δ and θ, their distribution should in some sense be centered on the true parameter values. In the present section, we shall give conditions under which the distributions of $\hat{\Delta}$ and $\hat{\theta}$ are actually symmetric with respect to Δ and θ, so that in particular the estimates are unbiased. (These conditions are related to those given by van der Vaart (1950).

THEOREM 2. *The distribution of the estimate $\hat{\Delta}$ defined by (2.2) and (2.3) is symmetric about Δ if either one of the following conditions hold:*

(i) *The distribution F defined in (2.1) is symmetric and h satisfies (3.2) and (7.5);*

(ii) *the two sample sizes m and n are equal, and h satisfies (3.3) and (7.5).*

PROOF.

(i) By the results of Section 7, we may assume without loss of generality that $\Delta = 0$. Further, by (7.5), the distribution F may be assumed to be symmetric

about zero. Since the random vectors (X, Y) and $(-X, -Y)$ then have the same distribution, $\hat{\Delta}(X, Y)$ and $\hat{\Delta}(-X, -Y)$ will also have the same distribution. To prove that $\hat{\Delta}(X, Y)$ and $-\hat{\Delta}(X, Y)$ have the same distribution (which is what the theorem claims), it is therefore only necessary to show that

(8.1) $$\hat{\Delta}(-x, -y) = -\hat{\Delta}(x, y).$$

This equation follows from the relations

(8.2) $\Delta^{**}(-x, -y) = -\Delta^*(x, y)$ and $\Delta^*(-x, -y) = -\Delta^{**}(x, y).$

To see for example the first of these, note that

$$\Delta^{**}(x, y) = \inf\{\Delta : h(-x, y - \Delta) < \mu\} = \inf\{\Delta : h(x, y + \Delta) > \mu\}$$

while

$$-\Delta^*(x, y) = \inf\{-\Delta : h(x, y - \Delta) > \mu\} = \inf\{\Delta : h(x, y + \Delta) > \mu\}.$$

(ii) Assume again that $\Delta = 0$. Since then the vectors (X, Y) and (Y, X) have the same distribution, it is enough to show that

(8.3) $$\hat{\Delta}(y, x) = -\hat{\Delta}(x, y).$$

This equation is an immediate consequence of the relations

$$\Delta^*(y, x) = -\Delta^{**}(x, y)   \text{ and }   \Delta^{**}(y, x) = -\Delta^*(x, y)$$

which follow from (3.3) and (7.5) as the corresponding relations in the proof of (i) followed from (3.2).

COROLLARY. *If h is given by* (3.1), *then the distribution of $\hat{\Delta}$ is symmetric about Δ if either one of the following conditions holds:*

(i) *the distributions F and Ψ are symmetric*

(ii) *the sample sizes m and n are equal.*

PROOF. Since h depends only on the ranks, it satisfies (7.5). The result now follows from Lemma 2 and Theorem 2.

The requirement in part (i) of the Corollary is a restriction on the test, which is satisfied both for the Wilcoxon and the normal scores test. On the other hand, the symmetry condition for F concerns an unknown distribution, and it is therefore much less certain whether it is satisfied in any given situation. The assumption is however frequently not unreasonable if one might be willing to assume normality except for the possibility of symmetric gross errors.

If F is not symmetric, $\hat{\Delta}$ need no longer be either symmetric or unbiased. Consider for example the case that h is the Wilcoxon statistic and that $m = 1$. Then $\hat{\Delta} = \text{med}(Y - X) = \text{med } Y - X_1$. For large n, $\tilde{Y}$ is essentially equal to the median of F, and $\hat{\Delta}$ need clearly not be unbiased. We shall however show in the next section that Δ is typically at least approximately, and in many cases exactly, median unbiased.

Results analogous to Theorem 2 and its Corollary hold also for the one-sample problem.

THEOREM 3. *The distribution of the estimate $\hat{\theta}$ defined by (2.5) and (2.6) is symmetric about θ if*

(i) *F is symmetric about zero and h satisfies (5.2) and hence in particular if*

(ii) *h is given by (5.1).*

PROOF. The proof of (i) is exactly analogous to that of Theorem 2(i); part (ii) follows from (i) and Lemma 3.

As in the two-sample problem, the symmetry assumption for F is frequently not unreasonable. In the particular case of a paired comparison experiment, with $Z_i = Y_i - X_i$, it is of course guaranteed if the assignment of the two subjects within each pair to the two treatments is performed at random.

9. Median unbiasedness. In general, when the distribution of $\hat{\Delta}$ or $\hat{\theta}$ is not symmetric about the true value, the estimate will also not be unbiased, that is, its expectation will not coincide with the parameter being estimated. The estimate will, however, in many cases still be median unbiased in the sense that the median of the distribution will be equal to the true value of the parameter. This follows from the following two lemmas, the first of which will also be used in the succeeding sections.

LEMMA 4. *For any real number a, the estimates $\hat{\Delta}$ and $\hat{\theta}$ satisfy the inequalities*

$$(9.1) \quad P\{h(X, Y - a) < \mu\} \leqq P\{\hat{\Delta} < a\} \leqq P\{h(X, Y - a) \leqq \mu\}$$

and

$$(9.2) \quad P\{h(Z - a) < \mu\} \leqq P\{\hat{\theta} < a\} \leqq P\{h(Z - a) \leqq \mu\}.$$

PROOF. By the definition of Δ^* and Δ^{**}; $\Delta^{**} < a \Rightarrow h(x, y - a) < \mu \Rightarrow \Delta^{**} \leqq a$ and $\Delta^* > a \Rightarrow h(x, y - a) > \mu \Rightarrow \Delta^* \geqq a$. Since Δ^* and Δ^{**} have been shown to have continuous distributions, it follows that

$$(9.3) \quad P\{\Delta^{**} < a\} = P\{h(X, Y - a) < \mu\}$$

and

$$(9.4) \quad P\{\Delta^* < a\} = P\{h(X, Y - a) \leqq \mu\},$$

and these imply (9.1).

The proof of (9.2) is exactly analogous.

LEMMA 5. *Let*

$$(9.5) \quad P_0\{h(X, Y) = \mu\} = \delta, \qquad P_0\{h(Z) = \mu\} = \epsilon.$$

Then

$$(9.6) \quad \tfrac{1}{2} - \delta/2 \leqq P_\Delta\{\hat{\Delta} \leqq \Delta\} \leqq \tfrac{1}{2} + \delta/2$$

and

$$(9.7) \quad \tfrac{1}{2} - \epsilon/2 \leqq P_\Theta\{\hat{\theta} \leqq \theta\} \leqq \tfrac{1}{2} + \epsilon/2.$$

PROOF. The inequalities (9.6) follow directly from (9.3) and (9.4).

As an immediate consequence of Lemma 5, it is seen that $\hat{\Delta}$ and $\hat{\theta}$ are median

unbiased if $P_0\{h(X, Y) = \mu\} = 0$ and $P_0\{h(Z) = \mu\} = 0$ respectively. Examples are the estimates based on the Wilcoxon statistics: in the two-sample problem if m and n are both odd, in the one-sample problem if N is odd, since in both these cases μ is an impossible value of h.

Even if $P_0(h = \mu)$ is not zero, it will typically be small, and it then follows from Lemma 5 that the probabilities of over- and under-estimation with the estimates $\hat{\Delta}$ and $\hat{\theta}$ will be close to $\frac{1}{2}$.

The fact that $\hat{\Delta}$ and $\hat{\theta}$ are either exactly or approximately median unbiased is related to their behaviour as upper or lower confidence bounds for Δ and θ. To discuss this relationship for the two-sample problem (the other case is completely analogous), consider the acceptance regions

$$(9.8) \qquad\qquad A(\Delta_0) = \{(x, y):h(x, y - \Delta_0) < \mu\}$$

for testing $H(\Delta_0):\Delta = \Delta_0$ against $\Delta > \Delta_0$ on the basis of the test statistic h. The associated family of confidence sets is $\{\Delta:h(x, y - \Delta) < \mu\}$, so that Δ^{**} is a lower confidence bound for Δ at confidence level $P_0\{h(X, Y) < \mu\}$. Analogously Δ^* is an upper confidence bound for Δ at confidence level $P_0\{h(X, Y) < \mu\}$.

Suppose now that $\Delta^* = \Delta^{**}$, as is the case for example when h is the Wilcoxon statistic and $m \cdot n$ is odd. It then follows from (9.3) and (9.4) with $a = 0$ that $P_0\{h(X, Y) = \mu\} = 0$ and hence that $\hat{\Delta} = \Delta^* = \Delta^{**}$ is the lower and upper confidence bound for Δ with confidence coefficient $\frac{1}{2}$, based on the test statistic h.

10. Asymptotic normality. So far, we have discussed small-sample properties of the estimates $\hat{\Delta}$ and $\hat{\theta}$; in the remaining sections, we shall be concerned with their large-sample behaviour. For the two-sample problem, let $m(N)$ and $n(N)$ for $N = 1, 2, \cdots$, be a sequence of pairs of sample sizes tending to infinity in such a way that $m(N)/N \to \lambda$, say, and let Δ_N be a sequence of values of the parameter Δ. Also for the one-sample problem consider the sequence of sample sizes $N = 1, 2, \cdots$ and let θ_N be a sequence of values of θ. In both cases, we shall indicate the dependence of h and μ on N by writing h_N and μ_N.

THEOREM 4. *Let* $a, c_1, c_2, \cdots$ *be real constants, and let*

$$(10.1) \qquad\qquad \Delta_N = -a/c_N \quad or \quad \theta_N = -a/c_N.$$

Let G be the continuous distribution function of a random variable with mean zero and unit variance, and suppose

$$(10.2) \qquad \lim_{N\to\infty} P_N\{c_N(h_N - \mu_N) \leq u\} = G((u + aB)/A)$$

where P_N indicates that the probability is computed for the parameter values Δ_N or θ_N and where h_N stands for $h_N(X_1, \cdots, X_{m(N)}; Y_1, \cdots, Y_{n(N)})$ or $h_N(Z_1, \cdots, Z_N)$. Then for any fixed Δ and θ

$$(10.3) \qquad \lim_{N\to\infty} P_\Delta\{c_N(\hat{\Delta}_N - \Delta) \leq a\} = G(aB/A)$$

or

$$(10.4) \qquad \lim_{N\to\infty} P_\theta\{c_N(\hat{\theta}_N - \theta) \leq a\} = G(aB/A).$$

PROOF. It is enough to give the proof for $\hat{\Delta}_N$, and we may in the proof let $\Delta = 0$. It follows from (9.1) that

$$\lim P_0\{c_N\hat{\Delta}_N \leqq a\} = \lim P_0\left\{h\left(X_1, \cdots, X_{m(N)}; Y_1 - \frac{a}{c_N}, \cdots, Y_{n(N)} - \frac{a}{c_N}\right) \leqq \mu_N\right\}$$

$$= \lim P_N\{h(X_1, \cdots, X_{m(N)}; Y_1, \cdots, X_{n(N)}) \leqq \mu_N\} = G\left(\frac{aB}{A}\right),$$

and this completes the proof.

Consider now the family of test statistics h given by (3.1). It then follows from the work of Chernoff and Savage (1958), see for example Theorem 7.1 of Puri (1962), that under suitable regularity conditions on Ψ

$$(10.5) \qquad N^{\frac{1}{2}}[h_N(X, Y) - \mu_N]$$

satisfies the assumptions of Theorem 4, with G the standard normal distribution and with A and B given by

$$(10.6) \qquad A^2 = \lambda(1 - \lambda)\left[\int_0^1 J^2(u)\,du - \left(\int_0^1 J(u)\,du\right)^2\right]$$

and

$$(10.7) \qquad B = \lambda(1 - \lambda)\int \{dJ[F(x)]/dx\}dF(x)$$

where $J = \Psi^{-1}$.

Combining this result with Theorem 4, we have

THEOREM 5. *If h is given by (3.1) with Ψ satisfying the assumptions of Theorem 7.1 of Puri (1963) and if $m(N)/N \to \lambda$ as $N \to \infty$, then $N^{\frac{1}{2}}(\hat{\Delta}_N - \Delta)$ has a limiting normal distribution with mean zero and variance A^2/B^2 where A and B are given by (10.6) and (10.7).*

Consider now in particular the estimate (4.1) which is obtained from (3.1) by taking for Ψ the rectangular distribution on $(0, 1)$. Since then $J(u) = u$ for $0 \leqq u \leqq 1$, we have $A^2 = \lambda(1 - \lambda)/12$ and $B = \lambda(1 - \lambda)\int f^2(x)\,dx$ where f is the density of F. The asymptotic variance of $N^{\frac{1}{2}}(\hat{\Delta} - \Delta)$ in this case is therefore

$$(10.8) \qquad 1\bigg/\left[12\lambda(1 - \lambda)\left(\int f^2(x)dx\right)^2\right].$$

11. Asymptotic efficiency. In basing the estimates $\hat{\Delta}$ and $\hat{\theta}$ on tests with desirable efficiency properties, it was the hope that these properties would be transferred to the estimates. That this hope is fulfilled under suitable regularity conditions follows from a result of Stuart (1954). The following theorem proves it under somewhat different assumptions.

THEOREM 6. *Let $\hat{\Delta}_N$ and $\hat{\Delta}'_N$ (or $\hat{\theta}_N$ and $\hat{\theta}'_N$) be estimates of Δ (or θ) based on sequences of test statistics h_N and h'_N satisfying the assumptions of Theorem 4 for the*

same limiting distribution G. Then the asymptotic relative efficiency of $\hat{\Delta}'_N$ relative to $\hat{\Delta}_N$ (or of $\hat{\theta}'_N$ relative to $\hat{\theta}_N$) in the sense of reciprocal ratio of asymptotic variances, is the same as the corresponding Pitman efficiency of the two sequences of tests based on h'_N and h_N provided the latter exists and $c_N = c'_N = N^{\frac{1}{2}}$.

PROOF. Since G is assumed to have unit variance, it is seen from (10.3) that the asymptotic variance of $N^{\frac{1}{2}}\hat{\Delta}_N$ is A^2/B^2 and that of $N^{\frac{1}{2}}\hat{\Delta}'_N$ is A'^2/B'^2 so that the asymptotic relative efficiency of the second sequence of estimates with respect to the first is $\mathrm{ARE}(\hat{\Delta}'_N ; \hat{\Delta}_N) = (A'^2/B'^2)/(A^2/B^2)$.

Consider on the other hand the power against the alternatives Δ_N given by (10.1) of the sequence of tests of $H : \Delta = 0$ against $\Delta < 0$ with rejection region $h_N < \mu_N$. By (10.2), this power tends to $G(aB/A)$. Similarly, the power of the corresponding sequence of tests based on h' and N' observations against the alternatives $\Delta'_{N'} = -a'/c'_{N'}$ tends to $G(a'B'/A')$. If we want the same limiting power against the same sequence of alternatives, we must have

$$a'/(N')^{\frac{1}{2}} = a/N^{\frac{1}{2}} \quad \text{and} \quad aB/A = a'B'/A'$$

and therefore $N'/N = (a'/a)^2 = (A'^2/B'^2)/(A^2/B^2)$. This completes the proof for $\hat{\Delta}_N$, and that for $\hat{\theta}_N$ is completely analogous.

It follows from this theorem (or more directly from (10.8)) that the asymptotic efficiency of the estimates (4.1) and (5.5) relative to the classical estimates $\bar{Y} - \bar{X}$ and $\bar{Z}$ is $12\sigma^2(\int f^2(x)\,dx)^2$, which in the case of normal F is $3/\pi \sim .955$. It is interesting to compare this value with the corresponding values for very small N. For $N = 1$ and 2, we have $\hat{\theta}_N = \bar{Z}$ so that the efficiency in these cases is 1. For $N = 3$, $\hat{\theta}_N$ is the median of the six quantities $Z_1, Z_2, Z_3, (Z_1 + Z_2)/2, (Z_1 + Z_3)/2, (Z_2 + Z_3)/2$. Let the ordered Z's be denoted by $Z^{(1)} < Z^{(2)} < Z^{(3)}$. Then

$$Z^{(1)} < (Z^{(1)} + Z^{(2)})/2 < Z^{(2)} < (Z^{(2)} + Z^{(3)})/2 < Z^{(3)}$$

and

$$Z^{(1)} < (Z^{(1)} + Z^{(2)})/2 < (Z^{(1)} + Z^{(3)})/2 < (Z^{(2)} + Z^{(3)})/2 < Z^{(3)}.$$

These inequalities show that $\hat{\theta}_3$ is the average of $Z^{(2)}$ and $(Z^{(1)} + Z^{(3)})/2$ so that

$$\hat{\theta}_3 = \tfrac{1}{4}(Z^{(1)} + 2Z^{(2)} + Z^{(3)}).$$

From a table of the covariances of normal order statistics, the efficiency of $\hat{\theta}_3$ is then seen to be .979.

REFERENCES

ANSCOMBE, F. J. (1960). Rejection of outliers. *Technometrics* **2** 123–147.

CHERNOFF, HERMAN and SAVAGE, I. RICHARD (1958). Asymptotic normality and efficiency of certain nonparametric tests. *Ann. Math. Statist.* **29** 972–994.

GOVINDARAJULU, Z. (1960). Central limit theorems and asymptotic efficiency for one-sample nonparametric procedures. Technical Report No. 11. Department of Statistics, Univ. of Minnesota.

LEHMANN, E. L. (1963). A class of selection procedures based on ranks. To appear in *Math. Ann.*

PURI, MADAN LAL (1962). Asymptotic efficiency of a class of c-sample tests. Submitted to *Ann. Math. Statist.*

STUART, ALAN (1954). Asymptotic relative efficiencies of distribution-free tests of randomness against normal alternatives. *J. Amer. Statist. Assoc.* **49** 147–157.

TUKEY, J. W. (1949). The simplest signed-rank tests. Mem. Report 17, Statistical Research Group, Princeton Univ.

TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics*, Essays in honor of Harold Hotelling 448–485.

VAN DER VAART, H. R. (1950). Some remarks on the power function of Wilcoxon's test for the problem of two samples. I and II. *Indag. Math.* **12** 146–172.