The Annals of Mathematical Statistics
1968, Vol. 39, No. 4, 1196–1201

ROBUSTNESS OF THE WILCOXON ESTIMATE OF LOCATION AGAINST A CERTAIN DEPENDENCE[1]

By Arnljot Høyland

The Norwegian Institute of Technology, Trondheim

0. Introduction and summary. Let $Z_1, Z_2, \cdots, Z_N$ be N random variables with common distribution $P(Z_i \leqq u) = F(u - \theta)$ where $F \varepsilon \mathfrak{F}_{cs0}$ which throughout this paper means the class of all continuous distributions symmetric about zero. The distributions considered will furthermore satisfy the regularity conditions of Lemma 3a in [3]. θ is an unknown constant to be estimated.

In the case where $Z_1, \cdots, Z_N$ are independent, the following estimate of θ has been recently investigated by J. L. Hodges Jr. and E. L. Lehmann [4].

$$\theta^* = \text{med}_{r \leqq s} \{(Z_r + Z_s)/2\}, \tag{0.1}$$

.e. the median of the $N + \binom{N}{2}$ averages $(Z_r + Z_s)/2$. The asymptotic efficiency of θ^* relative to the classical estimate

$$\hat{\theta} = \textstyle\sum_{i=1}^N Z_i/N \tag{0.2}$$

in the sense of reciprocal ratio of asymptotic variances has been determined in [4] and shown to be $12\sigma_z^2[\int f^2(z)\, dz]^2$ where f is the density corresponding to F and σ_z^2 denotes the variance of the Z's.

It follows directly from Theorem 2.2 of [6] that θ^*, in case of independent but not necessarily symmetrically distributed observations Z_j, is a consistent estimate of the pseudomedian of F ([6], p. 178), which in general may be different from the median θ.

In this paper we shall consider a situation where only few (c) observations can be collected per day and where the experiments have to be conducted over several (n) days to yield the necessary number of observations. During this period the experimental conditions may easily change, whereby the standard assumption of "independent and identically distributed" observations is violated. The data occur naturally grouped in n blocks, c observations per block, and the possible change of conditions is introduced as a (nuisance) random block effect.

We shall study the behavior of the two estimates θ^* and $\hat{\theta}$ under such conditions to find out how robust they are against this kind of dependence. In particular we shall study their asymptotic behavior as $n \to \infty$ with c fixed, and shall derive a general expression for the asymptotic efficiency of θ^* relative to $\hat{\theta}$. The efficiency is finally computed for normal and gross error models.

1. Notation and model. Let the random variables after grouping in blocks be

Received 24 August 1964; revised 24 August 1967.

[1] The main part of this paper was prepared with partial support of the U.S. Office of Naval Research, Contract Nonr-22-43. This paper in whole or in part may be reproduced for any purpose of the United States Government.

denoted X_{ij} $(i = 1, \cdots, n; j = 1, \cdots, c; nc = N)$. Consider the following model:

$$(1.1) \qquad X_{ij} = U_i + V_{ij} \qquad (i = 1, \cdots, n; \; j = 1, \cdots, c),$$

where $U_1, \cdots, U_n, V_{11}, \cdots, V_{nc}$ are assumed independent with distributions

$$(1.2) \quad P(U_i \leqq u) = G(u - \theta),$$
$$P(V_{ij} \leqq v) = K(v), \qquad (i = 1, \cdots, n; \; j = 1, \cdots, c),$$

and $G \varepsilon \mathfrak{F}_{cs0}$, $K \varepsilon \mathfrak{F}_{cs0}$ have densities g and k and variances τ^2 and σ^2 respectively.

θ^*_{nc} and $\hat{\theta}_{nc}$ shall denote the estimates (0.1) and (0.2) based on nc observations.

In this setup $\hat{\theta}_{nc}$ is easily seen to be symmetrically distributed about θ with variance $(\sigma^2 + c\tau^2)/nc$. Furthermore, as $n \to \infty$, $(nc)^{\frac{1}{2}}(\hat{\theta}_{nc} - \theta)$ has a limiting normal distribution with zero mean and variance $\sigma^2 + c\tau^2$.

2. Properties of θ^*. In the model (1.1), (1.2) θ^* is defined by

$$(2.1) \qquad \theta^* = \text{med}_{(ij) \leqq (kl)} \{(X_{ij} + X_{kl})/2\},$$

i.e. the median of the $\binom{nc}{2} + nc$ averages $(X_{ij} + X_{kl})/2$. Here $(ij) < (kl)$ means either $i < k$ or, if $i = k, j < l$. Similarly $(ij) = (kl)$ means $i = k, j = l$.

θ^* is easily seen to be symmetrically distributed about θ.

THEOREM 1. *Assume the model* (1.1), (1.2) *and let* θ^*_{nc} *be defined by* (2.1). *Then for every fixed* θ

$$(2.2) \qquad \lim_{n \to \infty} P_\theta\{(nc)^{\frac{1}{2}}(\theta^*_{nc} - \theta) \leqq u\} = \phi(Bu/A)$$

where

$$(2.3) \quad A^2 = 4 \cdot [\tfrac{1}{12} + (c - 1)(\lambda_0(G, K) - \tfrac{1}{4})], \qquad B = 2 \cdot [\textstyle\int f^2(x)\, dx]^2,$$

and

$$(2.4) \qquad \lambda_0(G, K) = P_0\{(X_{11} + X_{21}) > 0, (X_{12} + X_{31}) > 0\},$$

P_θ *(as throughout this paper) indicating that the probability is computed for median value* $= \theta$, *and* ϕ *denoting the cumulative distribution function of the standard normal distribution.*

Since Theorem 4 of [4] is checked to be true in this situation, Theorem 1 will immediately follow if one can establish that

$$(2.5) \quad \lim_{n \to \infty} P_{-a/(nc)^{\frac{1}{2}}}\{(nc)^{\frac{1}{2}}(W_{nc} - \tfrac{1}{2} - 1/(nc - 1) \leqq u\} = \phi((u + aB)/A)$$

where A^2 and B are given by (2.3).

According to the model (1.1) (1.2) the vectors $\mathbf{X}_i = (X_{i1}, X_{i2}, \cdots, X_{ic})$; $i = 1, 2, \cdots, n$, are independent. Let $\Psi_0(\mathbf{X}_i, \mathbf{X}_j)$ be defined by

$$(2.6) \quad \Psi_0(\mathbf{X}_i, \mathbf{X}_j) = \textstyle\sum_{r,s} U_{ir,js} \quad \text{with} \quad U_{ir,js} = 1 \quad \text{if} \quad X_{ir} + X_{js} > 0$$
$$= 0 \quad \text{otherwise.}$$

Then

$$W_{nc} = \binom{nc}{2}^{-1} \sum_{i,j} \Psi_0(\mathbf{X}_i, \mathbf{X}_j). \tag{2.7}$$

Consider

$$W_n = \binom{n}{2}^{-1} \sum_{i<j} \Psi_0(\mathbf{X}_i, \mathbf{X}_j). \tag{2.8}$$

W_n is, however, a generalized U-statistic [5]. It can be shown that Lehmann's extension of Hoeffding's theorem on generalized U-statistics [7] is applicable, and finally by Slutsky's theorem ([1], page 254) that $(nc)^{\frac{1}{2}}(W_{nc} - \frac{1}{2} - 1/(nc-1))$ has the same limiting distribution as $c^{-1}n^{\frac{1}{2}}(W_n - \frac{1}{2})$ when $n \to \infty$ (and $\theta_n \to 0$). (2.5) follows, and Theorem 2.1 is proved.

3. The asymptotic efficiency of θ^* relative to $\hat{\theta}$. In the model (1.1), (1.2), $\hat{\theta}$ is defined by

$$\hat{\theta} = \sum_{i,j} X_{ij}/nc. \tag{3.1}$$

By Theorem 1 and well-known properties of $\hat{\theta}$ the following result is immediate.

THEOREM 2. *Assume model* (1.1), (1.2) *and let* θ^* *and* $\hat{\theta}$ *be defined by* (2.1) *and* (3.1). *Then the asymptotic efficiency of* θ^* *relative to* $\hat{\theta}$ *in the sense of reciprocal ratio of asymptotic variances is*

$$\text{ARE }(\theta^*, \hat{\theta}) = 12(\sigma^2 + c\tau^2)[\textstyle\int f^2(x)\,dx]^2\{1 + 12(c-1)[\lambda_0(G,K) - \frac{1}{4}]\}^{-1}, \tag{3.2}$$

or equivalently

$$\text{ARE }(\theta^*, \hat{\theta}) = 12(1 + (c-1)\rho)[\textstyle\int f_s^2(x)\,dx]^2\{1 + 12(c-1)[\lambda_0(G,K) - \frac{1}{4}]\}^{-1}, \tag{3.3}$$

where $\rho = \tau^2/(\tau^2 + \sigma^2)$, $f_s(x)$ *is the standardized density,* i.e. $f(x) = \sigma_x^{-1}f_s(x/\sigma_x)$, *and* $\lambda_0(G, K)$ *is defined by* (2.4).

Note that changes in G and K may affect the ARE $(\theta^*, \hat{\theta})$ through the numerators of (3.2), (3.3) as well as the denominators.

4. Applications.

4.1. *Normal distributions.*

THEOREM 3. *Assume* (1.1), (1.2) *with* G *and* K *representing normal distributions. Let* θ^* *and* $\hat{\theta}$ *be defined by* (2.1) *and* (3.1). *Then*

$$\text{(i) ARE }(\theta^*, \hat{\theta}) = 3\pi^{-1}\cdot\{[1 + 2(c-1)\cdot\rho/2][1 + 6\pi^{-1}(c-1)\text{ Arcsine }(\rho/2)]^{-1}\}, \tag{4.1}$$

where $\rho = \tau^2/(\sigma^2 + \tau^2)$ *is the correlation of* $(X_{11} + X_{21})$ *and* $(X_{12} + X_{31})$.

$$\text{(ii) } 3/\pi \leqq \text{ARE }(\theta^*, \hat{\theta}) < 1. \tag{4.2}$$

PROOF. (i) By a well-known result due to Sheppard (see for example [2]), $\lambda_0(G, K)$ can be evaluated and shown to be equal to $(4\pi)^{-1}[\pi + 2\text{ Arcsine }(\rho/2)]$.

Furthermore $[\int f_s^2(x)\,dx]^2$ in this case is equal to $(4\pi)^{-1}$. These results introduced into (3.3), prove (i).

(ii) First we notice that for $\rho = 0$, ARE $(\theta^*, \hat{\theta}) = 3/\pi$. For $0 < \rho < 1$, $0 < \rho/2 <$ Arcsine $(\rho/2)$, and hence

$$\text{(4.3)} \quad \text{ARE } (\theta^*, \hat{\theta}) < 3\pi^{-1}\cdot\{[1 + 2(c - 1)\cdot\rho/2][1 + 6\pi^{-1}(c - 1)\cdot\rho/2]^{-1}\}.$$

As an increasing function of ρ the right hand side of (4.3) is less than or equal to its value for $\rho = 1$. This proves the right hand side of (4.2). By elementary computations it is easily shown that

$$[1 + 2(c - 1)\cdot\rho/2][1 + 6\pi^{-1}(c - 1) \text{ Arcsine } (\rho/2)]^{-1} \geqq 1 \quad \text{for} \quad 0 \leqq \rho \leqq 1.$$

Hence (4.2) follows. Since $\lim_{c\to\infty}$ ARE $(\theta^*, \hat{\theta}) = (\rho/2)/$Arcsine $(\rho/2)$ may be arbitrarily close to 1 for ρ sufficiently small, it follows that the right hand side inequality of (4.2) cannot be sharpened in general.

As a numerical illustration, the asymptotic efficiency of θ^* relative to $\hat{\theta}$ has been computed for different values of ρ and c.

TABLE 1

ARE $(\theta^*, \hat{\theta})$ *as a function of* ρ *and* c. *Normal distributions*

τ^2/σ^2	ρ	ARE $(\theta^*, \hat{\theta})$						
		$c = 2$	$c = 3$	$c = 4$	$c = 5$	$c = 6$	$c = 10$	$c = 20$
0	0	.955	.955	.955	.955	.955	.955	.955
$\frac{1}{4}$	$\frac{1}{5}$	.962	.967	.971	.974	.976	.982	.989
$\frac{1}{3}$	$\frac{1}{4}$	.963	.969	.973	.976	.978	.984	.990
$\frac{1}{2}$	$\frac{1}{3}$	.965	.971	.975	.978	.980	.985	.990
1	$\frac{1}{2}$	.966	.972	.975	.978	.979	.983	.986
2	$\frac{2}{3}$	.965	.970	.972	.974	.975	.977	.977
3	$\frac{3}{4}$	.964	.967	.969	.970	.971	.973	.974
4	$\frac{4}{5}$	.962	.965	.967	.968	.969	.970	.971
∞	1	.955	.955	.955	.955	.955	.955	.955

4.2. *Gross error models.* A more interesting case is the one where G and K represent gross error models:

$$\text{(4.4)} \qquad G(u) = (1 - \epsilon_1)\Phi(u/\sigma_1) + \epsilon_1\Phi(u/a_1\sigma_1),$$

$$\text{(4.5)} \qquad K(v) = (1 - \epsilon_2)\Phi(v/\sigma_2) + \epsilon_2\Phi(v/a_2\sigma_2).$$

Evaluation of $\int f^2(z)\,dz$ essentially involves computation of integrals of the form $\int \phi(y/a)\phi(y/b)\,dy$, easily shown ([6], (4. 10)) to be equal to $ab/2\pi(a^2 + b^2)$, while evaluation of $\lambda_0(G, K) = \int L^2(w)g(u)\,du$ essentially involves integrals of the form $\int \Phi(u/b)\Phi(u/c)a^{-1}\phi(u/a)\,du$, easily shown ([6], (4.9)) to be equal to $4^{-1}(1 - 2\pi^{-1}$ Arcsine $[a^2/((a^2 + b^2)(a^2 + c^2))^{\frac{1}{2}}])$.

The general expression for the ARE $(\theta^*, \hat{\theta})$ obtained through this approach, is however rather lengthy and will not be given here.

The ARE $(\theta^*, \hat{\theta})$ has however been determined numerically for $a_1 = a_2 = 3$

and values of the parameters ϵ_1, ϵ_2, σ_1^2/σ_2^2, and c selected so as to cover a variety of situations of practical interest. The ARE $(\theta^*, \hat{\theta})$ turns out to be larger than one in almost all cases and considerably larger than one in many. The results indicate that θ^* in general should be preferred to $\hat{\theta}$ in this gross error model.

TABLE 2

ARE $(\theta^*, \hat{\theta})$. *Gross error model* $a_1 = a_2 = 3$

ϵ_1	ϵ_2	$c = 2$					$c = 5$				
		σ_1^2/σ_2^2					σ_1^2/σ_2^2				
		$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4
	.01	1.0050	1.0069	1.0095	1.0109	1.0107	1.0129	1.0186	1.0222	1.0216	1.0184
.01	.02	1.0366	1.0292	1.0227	1.0173	1.0134	1.0341	1.0314	1.0288	1.0247	1.0196
	.05	1.1227	1.0903	1.0593	1.0354	1.0208	1.0932	1.0674	1.0476	1.0333	1.0233
	.01	1.0135	1.0240	1.0367	1.0470	1.0533	1.0279	1.0437	1.0568	1.0633	1.0643
.02	.02	1.0444	1.0454	1.0491	1.0529	1.0555	1.0483	1.0558	1.0629	1.0659	1.0654
	.05	1.1284	1.1042	1.0834	1.0692	1.0617	1.1050	1.0897	1.0801	1.0734	1.0683
	.01	1.0366	1.0695	1.1089	1.1433	1.1667	1.0669	1.1088	1.1472	1.1729	1.1862
.05	.02	1.0652	1.0886	1.1192	1.1475	1.1678	1.0850	1.1190	1.1518	1.1745	1.1865
	.05	1.1433	1.1408	1.1476	1.1591	1.1707	1.1353	1.1472	1.1647	1.1788	1.1874

4.3. *General bounds for* ARE $(\theta^*, \hat{\theta})$. Numerical evaluation of $\lambda_0(G, K)$ may be complicated for specific distributions, G and K. We shall therefore determine bounds for $\lambda_0(G, K)$ and thereby obtain general bounds for ARE $(\theta^*, \hat{\theta})$.

LEMMA 1. *Let T_1, T_2, W_1 and W_2 have continuous distributions.* (i) *If T_1 and T_2 are independently and identically distributed, then for any pair of real numbers* (a, b),

$$P\{(T_1 > a) \cap (T_2 > b)\} \leqq P\{(T_1 > a) \cap (T_1 > b)\}. \tag{4.6}$$

(ii) *If, for given $W_1 = w_1$, $W_2 = w_2$, T_1 and T_2 are independently and identically distributed for all (w_1, w_2), then*

$$P\{(T_1 > W_1) \cap (T_2 > W_2)\} \leqq P\{(T_1 > W_1) \cap (T_1 > W_2)\}. \tag{4.7}$$

Part (i) is obvious, and part (ii) is immediate by conditioning wrt W_1 and W_2 and using (i).

LEMMA 2. *Assume model* (1.1), (1.2). *Then*

$$\tfrac{1}{4} \leqq P_0\{(X_{11} + X_{21} > 0) \cap (X_{12} + X_{31} > 0)\} \leqq \tfrac{1}{3}, \tag{4.8}$$

P_0 indicating that the probability is computed for $\theta = 0$.

PROOF.

$$\lambda_0(G, K) = P_0\{(V_{11} > -U_1 - U_2 - V_{21}) \cap (V_{21} > -U_1 - U_3 - V_{31})\}$$

is by Lemma 1 and the symmetry properties

$$\leqq P_0\{(V_{11} > -U_1 - U_2 - V_{21}) \cap (V_{11} > -U_1 - U_3 - V_{31})\}$$
$$\leqq P_0\{(X_{11} > X_{21}) \cap (X_{11} > X_{31})\} = \tfrac{1}{3}.$$

On the other hand

$$\begin{aligned} \lambda_0(G, K) &= P_0\{(V_{11} + U_2 + V_{21} > -U_1) \cap (V_{12} + U_3 + V_{31} > -U_1)\} \\ &= P_0\{(V_{11} + U_2 + V_{21} < U_1) \cap (V_{12} + U_3 + V_{31} < U_1)\} \qquad (4.9) \\ &= \int L^2(u_1)g(u_1)\,du_1, \end{aligned}$$

where

$$P_0(V_{11} + U_2 + V_{21} \leqq u) = L(u).$$

By Schwarz' inequality

$$\int L^2(u_1)g(u_1)\,du_1 \geqq [\int L(u_1)g(u_1)\,du_1]^2 \geqq [P_0\{(X_{11} + X_{21}) > 0\}]^2 = \tfrac{1}{4}. \text{ q.e.d.}$$

The following result is immediate by combining Theorem 2 and Lemma 2.

THEOREM 4. *Assume model* (1.1), (1.2) *and let* θ^* *and* $\hat{\theta}$ *be defined by* (2.1) *and* (3.1). *Then the asymptotic efficiency of* θ^* *relative to* $\hat{\theta}$ *in the sense of reciprocal ratio of asymptotic variances*, ARE $(\theta^*, \hat{\theta})$ *satisfies the following inequalities*:

$$12c^{-1}(\sigma^2 + c\tau^2)[\int f^2(x)\,dx]^2 \leqq \text{ARE}(\theta^*, \hat{\theta}) \leqq 12(\sigma^2 + c\tau^2)[\int f^2(x)\,dx]^2 \qquad (4.10)$$

5. Acknowledgment. The main portion of this work was done as part of a Ph.D. Thesis at the University of California, Berkeley. I wish to express my sincere thanks to Professor E. L. Lehmann for suggesting the problem and for continued encouragement and guidance.

REFERENCES

[1] CRAMÉR, H. (1945). *Mathematical Methods of Statistics*. Almqvist and Wiksells, Uppsala.

[2] DAVID, F. N. (1953). A note of the evaluation of the multivariate normal integral. *Biometrika* **40** 458–559.

[3] HODGES, J. L., JR. and LEHMANN, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. *Proc. Fourth Berkeley Symposium on Math. Statist. and Prob.* **1** 307–317. Univ. of California Press.

[4] HODGES, J. L., JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

[5] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19** 293–325.

[6] HØYLAND, A. (1964). Robustness of the Hodges-Lehmann estimates for shift. *Ann. Math. Statist.* **36** 174–197.

[7] LEHMANN, E. L. (1963). Robust estimation in analysis of variance. *Ann. Math. Statist.* **34** 957–966.