ON SOME ALTERNATIVE ESTIMATES FOR SHIFT IN THE P-VARIATE ONE SAMPLE PROBLEM[1]

By Peter J. Bickel

University of California, Berkeley

1. Summary. The vector of medians $\mathbf{M}_n$ and the vector of medians of averages of pairs $\mathbf{W}_n$ are investigated as competitors of the vector mean $\dot{\mathbf{X}}_n$ in estimating the location parameters in the p-variate one-sample problem.

These estimates are found to be asymptotically normal and unbiased. Necessary and sufficient conditions for the degeneracy of the asymptotic distribution of $\mathbf{M}_n$ and $\mathbf{W}_n$ are given. For $\mathbf{W}_n$, in the case $p = 2$, these reduce to the condition that one coordinate variable be a monotone function of the other. Sufficient symmetry conditions are given for the asymptotic independence of the coordinates of these estimates.

$\mathbf{W}_n$ and $\mathbf{M}_n$ when compared to $\dot{\mathbf{X}}_n$ in terms of the Wilks generalized variance are robust in the case of asymptotically independent coordinates. But for $p \geqq 3$ they can have arbitrarily small efficiency even in the non-singular p-variate normal case, if the underlying distribution is permitted to approach a suitable degenerate distribution arbitrarily closely. For $p = 2$, in the normal case, $\mathbf{W}_n$ is highly efficient, although $\mathbf{M}_n$ can have arbitrarily small efficiency. However, $\mathbf{W}_n$ is also shown to have arbitrarily small efficiency for a suitable highly correlated family of distributions even in the case $p = 2$. On the other hand, $\mathbf{W}_n$ becomes infinitely more efficient than $\dot{\mathbf{X}}_n$ as a given fixed distribution is mixed with an increasingly heavy gross error distribution. The behavior of these estimates is also considered for other non-normal families.

2. Introduction. If $Z_1, \cdots, Z_n$ is a sample from a univariate population with distribution $F(x - \theta)$, where F and θ are both unknown and F is symmetric about 0 and continuous, then several consistent and asymptotically normal point estimates, $\hat{\theta}_n(Z_1, \cdots, Z_n)$, of θ are known. In a recent paper [3] Hodges and Lehmann proposed a general method of obtaining such estimates from the test statistics used in this problem.

Now suppose $\mathbf{X}_i$, $i = 1, \cdots, n$, where $\mathbf{X}_i = (X_{1i}, \cdots, X_{pi})$, is a sample from a p-variate distribution $F(x_1 - \theta_1, \cdots, x_p - \theta_p)$, where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$ is unknown, F is symmetric about $\mathbf{0}$, and F is continuous. Then a natural estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_n(X_{11}, \cdots, X_{1n}), \cdots, \hat{\theta}_n(X_{p1}, \cdots, X_{pn})). \tag{2.1}$$

We shall be particularly concerned with the following three estimates: the

Received 30 December 1963; revised 13 February 1964.

[1] This research formed a portion of the author's doctoral dissertation submitted to the University of California, Berkeley, and was written with the partial support of the National Science Foundation, Grant GP-10.

"classical estimate,"

$$\hat{\theta}_n(Z_1, \cdots, Z_n) = (Z_1 + \cdots + Z_n)/n, \tag{2.2}$$

$$\hat{\theta}_n(Z_1, \cdots, Z_n) = \text{median}_{i=1,\cdots,n} Z_i, \tag{2.3}$$

and

$$\hat{\theta}_n(Z_1, \cdots, Z_n) = \text{median}_{1 \leqq i \leqq j \leqq n} \tfrac{1}{2}(Z_i + Z_j). \tag{2.4}$$

The corresponding vector estimates are denoted by $\dot{\mathbf{X}}_n$, $\mathbf{M}_n$, and $\mathbf{W}_n$, respectively. The first two estimates are well-known. The last has been shown by Hodges and Lehmann in [3] to share the robustness of the Wilcoxon test, even as the first two estimates are intimately connected with the "t" and sign tests, respectively.

3. Asymptotic Normality of $\mathbf{M}_n$, $\mathbf{W}_n$, $\dot{\mathbf{X}}_n$. The first two parts of the following theorem are well known. Both parts (2) and (3) may be established by a straightforward generalization of the arguments used in [3] to derive the corresponding univariate distributions and by an application of Lehmann's version of the Hoeffding theorem on U statistics (Lehmann [4]).

Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$ be a $1 \times p$ vector, $\mathbf{0}$ be the zero vector, and let $P_{\boldsymbol{\theta}}$ denote the probability measure under which the $\mathbf{X}_i$ are independent and have distribution $F(x_1 - \theta_1, \cdots, x_p - \theta_p)$.

THEOREM 3.1.

(1) *Suppose* $E_{\mathbf{0}}(X_{i1}^2) < \infty$, $i = 1, \cdots, p$, *where* $E_{\boldsymbol{\theta}}$ *indicates that the expected value is taken under* $P_{\boldsymbol{\theta}}$, *and* $\boldsymbol{\theta}$ *is as above. Then*

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}}(n^{\frac{1}{2}}(\dot{\mathbf{X}}_n - \boldsymbol{\theta}) \leqq \mathbf{u}) = \Phi_{[\mathbf{0}, \|\sigma_{ij}\|]}(u_1, \cdots, u_p) \tag{3.1}$$

where $\Phi_{[\mathbf{0},\|\sigma_{ij}\|]}$ *is the cumulative of the p-variate normal distribution with mean* $\mathbf{0}$ *and covariance matrix* $\|\sigma_{ij}\|$ *and* $\sigma_{ij} = E_{\mathbf{0}}(X_{i1}, X_{j1})$ *in our case.*

(2) *Suppose the marginal distribution of* X_{i1}, $F_i(x)$, *is absolutely continuous for* $i = 1, \cdots, p$ *and its derivative at* 0 *denoted by* $f_i(0)$ *exists. Then*

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}}(n^{\frac{1}{2}}(\mathbf{M}_n - \boldsymbol{\theta}) \leqq \mathbf{u}) = \Phi_{[\mathbf{0},\|\tau_{ij}\|]}(u_1, \cdots, u_p)$$

where

$$\begin{aligned} \tau_{ij} &= 1/4f_i^2(0), && i = j, \\ &= \{P_{\mathbf{0}}[X_{i1} > 0, X_{j1} > 0] - \tfrac{1}{4}\}/f_i(0)f_j(0), && i \neq j. \end{aligned} \tag{3.2}$$

(3) *Suppose the marginal distribution of* X_{i1} *is absolutely continuous,* $i = 1, \cdots, p$. *As above its density is denoted by* f_i. *Then*

$$\lim_n P_{\boldsymbol{\theta}}(n^{\frac{1}{2}}(\mathbf{W}_n - \boldsymbol{\theta}) \leqq \mathbf{u}) = \Phi_{[0,\|\gamma_{ij}\|]}(u_1, \cdots, u_p)$$

where

$$\begin{aligned} \gamma_{ij} &= 1/12[\textstyle\int_{-\infty}^{\infty} f_i^2(x)\,dx]^2, && i = j, \\ &= \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_i(x)F_j(y)\,dF_{i,j}(x, y) - \frac{1}{4}}{(\int_{-\infty}^{\infty} f_i^2(x)\,dx)(\int_{-\infty}^{\infty} f_j^2(y)\,dy)} && i \neq j, \end{aligned} \tag{3.3}$$

and $F_{i,j}$ *is the joint distribution of* (X_{i1}, X_{j1}) *under* $P_{\mathbf{0}}$.

We now examine conditions under which the distributions of $\dot{\mathbf{X}}_n$, $\mathbf{M}_n$, $\mathbf{W}_n$ asymptotically (a) have independent coordinates, and (b) are degenerate.

We say that a bivariate random vector (X, Y) is *totally symmetric* if

$$(3.4) \qquad (X, Y), (-X, -Y) \quad \text{and} \quad (X, -Y)$$

have the same distribution.

We then have

THEOREM 3.2. *A sufficient condition for the asymptotic independence of the components of* (1) $\dot{\mathbf{X}}_n$, (2) $\mathbf{M}_n$, (3) $\mathbf{W}_n$ *is the total symmetry of* (X_{i1}, X_{j1}) *for every pair* (i, j).

PROOF. By the hypothesis, every pair of coordinates of $\dot{\mathbf{X}}_n$, $\mathbf{W}_n$, or $\mathbf{M}_n$ must possess the total symmetry property and, by consideration of either characteristic or distribution functions, so must every pair of coordinates of the respective limiting normal distributions. But if (X, Y) is totally symmetric the correlation of X and Y is 0 and our result follows.

Examples of symmetric distributions possessing this property are those with pairwise independent coordinates and those with densities $f(x_1, \cdots, x_p)$ such that $f(x_1, \cdots, x_p) = f(\pm x_1, \cdots, \pm x_p)$ for any choice of $\pm$, for example, a uniform distribution on the p sphere or any mixture of distributions with pairwise independent coordinates.

THEOREM 3.3 *Let* $\mathbf{X}_i$, $1 \leqq i \leqq n$, *be random vectors satisfying the hypotheses of Theorem* 3.1 *on a space* $(\Omega, \mathcal{A})$. *Then the asymptotic distribution of* (1) $\dot{\mathbf{X}}_n$, (2) $\mathbf{W}_n$, (3) $\mathbf{M}_n$ *is k-variate normal,* $k < p$, *if and only if, for some j,* a.s. $[P_0]$, *respectively,*

$$(3.5) \qquad (1) \qquad X_{j1} = \sum_{k \neq j} \alpha_k X_{k1},$$

$$(3.6) \qquad (2) \qquad F_j(X_{j1}) = \sum_{k \neq j} \alpha_k [F_k(X_{k1}) - \tfrac{1}{2}] + \tfrac{1}{2},$$

$$(3.7) \qquad (3) \qquad I(X_{j1}^+) = \sum_{k \neq j} \alpha_k [I(X_{k1}^+) - \tfrac{1}{2}] + \tfrac{1}{2},$$

where the distribution of $\mathbf{X}_i$ *is computed under* P_0, F_i *is the marginal distribution of* X_{i1} *and* $I(X^+) = 1$ *if* X *is positive and* 0 *otherwise.*

PROOF. We observe that the covariance matrices of the limiting distributions of $\dot{\mathbf{X}}_n$, $\mathbf{W}_n$, $\mathbf{M}_n$ are the moment matrices, respectively, of the random vectors $\mathbf{X}_1$, $\{c_1[F_1(X_{11}) - \tfrac{1}{2}], \cdots, c_p[F_p(X_{p1}) - \tfrac{1}{2}]\}$, $c_i \neq 0$, and $\{c_1^*[I(X_{11}^+) - \tfrac{1}{2}], \cdots, c_p^*[I(X_{p1}^+) - \tfrac{1}{2}]\}$, $c_i^* \neq 0$. The result is then a consequence of Frisch's theorem (see Cramér [1], p. 297), which states that the moment matrix of a multivariate distribution is of rank $r \leqq p$, if and only if the distribution is carried by an r dimensional subspace of R^p and r is the minimal dimension of such a subspace.

Consequently we obtain

THEOREM 3.4. *Necessary conditions that the asymptotic distributions of* (1) $\mathbf{M}_n$, (2) $\mathbf{W}_n$ *be degenerate are, respectively, that under* P_0 :

(1) $\mathbf{X}_1$ *has its distribution carried by at most* $2^p - 2$ *of the* 2^p *orthants of* R^p.

(2) $X_{j1} = f(X_{i1}, \text{ all } i \neq j)$ *for some* j. *If* $p = 2$ *Condition* (1) *becomes sufficient for the asymptotic degeneracy of* $\mathbf{M}_n$.

The following Condition (2^*) *becomes both necessary and sufficient for the asymptotic degeneracy of* $\mathbf{W}_n$, *if* $p = 2$.

(2^*) *Condition* (2) *holds and the* f *in Condition* (2) *is strictly monotone on a Borel set* A *such that* $P_0(X_{21} \varepsilon A) = 1$.

PROOF. Suppose (1) does not hold. Then

$$P_0[X_{i1} > 0, i = 1, \cdots, p] > 0$$

which from (3.7) implies that $\sum_k \alpha_k = 1$. But this implies that $P_0[X_{j1} < 0, X_{i1} > 0, \text{ all } i \neq j] = 0$, a contradiction. (2) is a consequence of the following lemma.

LEMMA 3.5. *Let* X *be any random variable with continuous distribution function* F. *Let* $A_x = F^{-1}(\{x\})$, $D(F) = \{x \mid A_x$ *contains more than* 1 *point*$\}$. *Then* $P[X \varepsilon \bigcup_{x \varepsilon D(F)} A_x] = 0$.

PROOF. Since F is continuous and monotone, A_x is a degenerate or nondegenerate closed interval. Then $x \varepsilon D_F$ if and only if A_x is nondegenerate. But $A_x \cap A_y = \phi$ if $x \neq y$. Hence $D(F)$ is countable. But $P[X \varepsilon A_x] = 0$ for every x.

Now let $S = \{F_j(X_{j1}) \varepsilon [0, 1] - D(F_j)\}$. Then $P_0(S) = 1$ by Lemma 3.5. But on $F_j^{-1}\{[0, 1] - D(F_j)\}$ F_j is 1-1. Let $\tilde{F}_j^{-1}$ be the pointwise inverse of F_j on $[0, 1] - D(F_j)$. Then on $S \cap (\Omega - N)$,

$$X_{j1} = \tilde{F}_j^{-1}[\tfrac{1}{2} + \sum_{k \neq j} \alpha_k(F_k(X_{k1}) - \tfrac{1}{2})]$$

where (3.6) holds on $\Omega - N$, and $P_0(N) = 0$. (2) now follows.

The sufficiency Condition (1) is immediate. For, if the distribution of (X_{11}, X_{21}) is carried by two quadrants, $P_0[X_{11} > 0, X_{21} > 0] > 0] = 0$ or $P_0[X_{11} > 0, X_{21} < 0] = 0$ giving, in conjunction with symmetry, $I(X_{11}^+) = I(\pm X_{21}^+)$ respectively. Condition (3) of Theorem 3.3 is then satisfied.

We first show that for $p = 2$ Condition (2^*) is necessary. (3.6) reduces to

$$F_1(X_{11}) = \tfrac{1}{2}(1 - \alpha) + \alpha F_2(X_{21}) \qquad \text{on } \Omega - N. \tag{3.8}$$

Suppose $P_0[X_{11} > 0, X_{21} > 0] > 0$. Then we choose $\{w_n\} \varepsilon \Omega$ such that $F_1[X_{11}(w_n)]$, $F_2[X_{21}(w_n)]$ tend to 1 and (3.6) holds, giving $\alpha = 1$. Now let $A = \{x \mid F_2(x) \varepsilon [0, 1] - D(F_1)\}$. Then $f(x) = \tilde{F}_1^{-1} F_2$ is well defined on A and $X_{11} = f(X_{21})$ on $[\Omega - N] \cap [F_1(X_{11}) \varepsilon [0, 1] - D(F_1)]$. Clearly f is strictly increasing on A and $1 = P_0\{(\Omega - N) \cap [F_1(X_{11}) \varepsilon [0, 1] - D(F_1)]\} = P_0[X_{21} \varepsilon A]$. On the other hand, if $P_0[X_{11} < 0, X_{21} > 0] > 0$, upon applying the above results to $(X_{11}, -X_{21})$ we obtain f strictly decreasing on A. Thus necessity is proved.

Now suppose f is as specified with f strictly increasing on A. Then, for $x \varepsilon f(A)$,

$$P_0[X_{11} \leqq x] = P_0\{[X_{11} \leqq x] \cap [\Omega - N]\} = P_0[X_{21} \leqq f^{-1}(x)]$$

by the monotonicity of f. Hence

$$F_1(X_{11}) = F_2(f^{-1}(X_{11})) \qquad \text{on } [\Omega - N] \cap X_{21}^{-1}(A)$$
$$= F_2(X_{21}),$$

since on $\Omega - N$, $X_{11} = f(X_{21})$.

If f is strictly decreasing, the result follows by considering $(X_{11}, -X_{21})$ and observing that $F_i(x) = 1 - F_i(-x)$, $i = 1, 2$.

We remark that, if $p = 2$, Condition (3.5) implies Condition (2^*) of Theorem 3.4, which implies Condition (1) of Theorem 3.4, and the reverse implications need not hold.

If $p > 2$, then the example of Proposition 4.1 shows that Condition (3.5) need not imply (3.6) and similar examples showing that neither (3.6) nor (3.5) need imply (3.7) can also be constructed.

4. Asymptotic efficiency of $\mathbf{W}_n$, $\mathbf{M}_n$ with respect to $\dot{\mathbf{X}}_n$. We employ as a measure of the asymptotic "scatter" of multivariate estimators which are asymptotically unbiased their asymptotic "generalized variances", a concept introduced by Wilks (see Cramér [3], p. 301). The "generalized variance" of a p-variate random vector $(X_1, \cdots, X_p)'$ with nonsingular covariance matrix, $\|\rho_{ij}\sigma_i\sigma_j\|$, is defined to be

$$\text{Var } \mathbf{X} = \sigma_1^2 \cdots \sigma_p^2 \det \|\rho_{ij}\|, \tag{4.1}$$

where "det" denotes determinant.

Then it is natural to define as the (asymptotic) efficiency of a p-variate (asymptotically) unbiased estimator T of $\boldsymbol{\theta}$ with (asymptotically) nonsingular covariance matrix with respect to a similar estimator $\mathbf{S}$, the limit (if any) of the inverse ratio of sample sizes required to reach equal (asymptotic) generalized variances. Thus if the (asymptotic) covariance matrix of $\mathbf{S}$ is $\|\rho_{ij}\sigma_i\sigma_j\|$ and that of $\mathbf{T}$ is $\|\rho_{ij}^*\tau_i\tau_j\|$, the efficiency of $\mathbf{S}$ with respect to $\mathbf{T}$ is

$$e(\mathbf{S}, \mathbf{T}) = [(\tau_1^2 \cdots \tau_p^2/\sigma_1^2 \cdots \sigma_p^2)(\det \|\rho_{ij}^*\|/\det \|\rho_{ij}\|)]^{p^{-1}}. \tag{4.2}$$

By substituting from (3.2) and (3.3) we therefore have, in the nonsingular case,

$$e(\mathbf{W}_n, \dot{\mathbf{X}}_n) = \left[\prod_{i=1}^{p} 12\sigma_i^2\left(\textstyle\int_{-\infty}^{\infty} f_i^2(x)\,dx\right)^2\right]^{p^{-1}} [\det \|\rho_{ij}\|/\det \|\rho_{ij}^*\|]^{p^{-1}} \tag{4.3}$$

where $\|\rho_{ij}\sigma_i\sigma_j\|$ is the covariance matrix of $\mathbf{X}_i$ and where

$$\rho_{ij}^* = 12\left[\textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_i(x)F_j(y)\,dF_{i,j}(x, y) - \frac{1}{4}\right], \qquad i \neq j,$$
$$= 1, \qquad i = j.$$

Similarly, we have

$$e(\mathbf{M}_n, \dot{\mathbf{X}}_n) = \left[\prod_{i=1}^{p} 4\sigma_i^2 f_i^2(0)\right]^{p^{-1}} [\det \|\rho_{ij}\|/\det \|\rho_{ij}^{**}\|]^{p^{-1}} \tag{4.4}$$

where

$$\rho_{ij}^{**} = 4[P_0[X_{i1} > 0, X_{j1} > 0] - \tfrac{1}{4}].$$

We remark immediately that if $\dot{\mathbf{X}}_n$, $\mathbf{W}_n$, and $\mathbf{M}_n$ have asymptotically independent co-ordinates, for instance under the conditions of Theorem 3.2, we have

$$e(\mathbf{M}_n, \dot{\mathbf{X}}_n) = 4\left(\prod_{i=1}^{p} \sigma_i^2 f_i^2(0),\right)^{p^{-1}} \tag{4.5}$$

$$e(\mathbf{W}_n, \dot{\mathbf{X}}_n) = 12\left(\prod_{i=1}^{p} \sigma_1^2 \left(\textstyle\int_{-\infty}^{\infty} f_i^2(x)\, dx\right)^2\right)^{p^{-1}} \tag{4.6}$$

and hence from a result of Hodges and Lehmann [2]

$$\inf_{F\varepsilon\mathfrak{F}} e(\mathbf{W}_n, \dot{\mathbf{X}}_n) = .864, \tag{4.7}$$

$$\inf_{F\varepsilon\mathfrak{F}^*} e(\mathbf{M}_n, \dot{\mathbf{X}}_n) = .33, \tag{4.8}$$

where $\mathfrak{F}$ is the set of all totally symmetric continuous p = variate distributions and $\mathfrak{F}^*$ is the set of all unimodal totally symmetric continuous p-variate distributions.

Unfortunately the favorable result above is offset by

PROPOSITION 4.1. For $p > 2$,

$$\inf_{\Phi\varepsilon\varphi} e(\mathbf{W}_n, \dot{\mathbf{X}}_n) = 0, \tag{4.9}$$

and for $p \geqq 2$

$$\inf_{\Phi\varepsilon\varphi} e(\mathbf{M}_n, \dot{\mathbf{X}}_n) = 0, \tag{4.10}$$

where φ is the family of all nonsingular p-variate normal distributions.

PROOF. We prove (4.9). The proof of (4.10) is delayed to the next section.

It suffices to prove the result for $p = 3$.

Take $\mathbf{X}_n$ with underlying distribution tri-variate nonsingular normal,

$$\begin{aligned} \rho_{ij} &= 0 & i, j &= 1, 2, \\ &= (1 - \alpha)2^{-\frac{1}{2}} & i &\neq j \text{ otherwise}, \\ &= 1 & i &= j. \end{aligned}$$

As $\alpha \to 0$ clearly $\det \|\rho_{ij}\| \to 0$, $\det \|\rho_{ij}^*\| \not\to 0$ and (4.9) follows.

We conclude that in the independent case the behavior of these estimates is excellent and essentially equivalent to the univariate situation.

Now, Proposition 4.1 indicates that for $p > 2$ in the general normal case $\mathbf{M}_n$ and $\mathbf{W}_n$ are not competitive. However, it is always true that

$$e(\mathbf{M}_n, \dot{\mathbf{X}}_n) \geqq \left[\prod_{i=1}^{p} 4\sigma_i^2 f_i^2(0)\right]^{p^{-1}} [\det \|\rho_{ij}\|]^{p^{-1}} \tag{4.11}$$

and

$$e(\mathbf{W}_n, \dot{\mathbf{X}}_n) \geqq \left[\prod_{i=1}^{p} 12\sigma_i^2 \left(\textstyle\int_{-\infty}^{\infty} f_i^2(x)\, dx\right)^2\right]^{p^{-1}} [\det \|\rho_{ij}\|]^{p^{-1}} \tag{4.12}$$

(see Cramér [1], p. 296).

Hence for distributions that are sufficiently nondegenerate i.e. for which det $||\rho_{ij}||$ is considerably greater than 0 in absolute value, the behavior of $\mathbf{W}_n$ and $\mathbf{M}_n$ is similar to that in the univariate case.

Thus the efficiency of $\mathbf{W}_n$ is large for distributions with heavy tails and, in fact, Theorem 6.1 can clearly be extended to the general p-variate situation.

It also seems clear that, for distributions satisfying (3.6) but not (3.5), $\mathbf{W}_n$ is infinitely more efficient than $\dot{\mathbf{X}}_n$. Simple examples of such distributions are ones concentrating all their mass on non-linear monotone curves contained in one of the co-ordinate planes. Similar remarks apply to $\mathbf{M}_n$.

Yet another special case that is of interest occurs when the distribution of (X_{i1}, X_{j1}) is independent of i and j when $\boldsymbol{\theta} = \mathbf{0}$. In this case, from a well-known formula for determinants, we obtain

$$(4.13) \qquad e(\mathbf{M}_n, \dot{\mathbf{X}}_n) = 4\sigma_1^2 f_1^2(0)\left[\frac{1-\rho_{12}}{1-\rho_{12}^{**}}\right]^{1-p^{-1}}\left[\frac{1+(p-1)\rho_{12}}{1+(p-1\rho_{12}^{**}}\right]^{p^{-1}}$$

and

$$(4.13) \quad e(\mathbf{W}_n, \dot{\mathbf{X}}_n) = 12\sigma_1^2\left(\int_{-\infty}^{\infty} f_1^2(x)\,dx\right)^2\left[\frac{1-\rho_{12}}{1-\rho_{12}^{*}}\right]^{1-p^{-1}}\left[\frac{1+(p-1)\rho_{12}}{1+(p-1)\rho_{12}^{*}}\right]^{p^{-1}}.$$

If, in particular, following Stuart [6] we have $X_{i1} = Z_{i+1} - Z_1$ where the Z_j, $j = 1, \cdots, p+1$ form a set of independent, identically distributed, symmetric random variables, then

$$e(\mathbf{M}_n, \mathbf{X}_n) = 3.2^{p^{-1}}\left[\frac{p+1}{p+2}\right]^{p^{-1}} \sigma_1^2 f_1^2(0)$$

which tends to $3\sigma_1^2 f_1^2(0)$ from above as $p \to \infty$. From Theorem 2 in Lehmann [5] we also see that in this case

$$6\sigma_1^2\left(\int_{-\infty}^{\infty} f_1^2(x)\,dx\right)^2(p+1)^{p^{-1}} < e(\mathbf{W}_n, \dot{\mathbf{X}}_n) \leqq 12\sigma_1^2\left(\int_{-\infty}^{\infty} f_1^2(x)\,dx\right)^2.$$

It follows from Table 7a in Lehmann [5] that, if Z_1 is normal or rectangular, $e(\mathbf{W}_n, \dot{\mathbf{X}}_n)$ is lower bounded by .925 and .88, respectively, for all p.

There would, thus, despite Proposition 4.1 seem to be many situations in which $\mathbf{W}_n$ and $\mathbf{M}_n$ are preferable to $\dot{\mathbf{X}}_n$ even for $p > 2$ and which could bear further investigation. In this paper, however, we from now on restrict ourselves to the simpler case $p = 2$.

5. Efficiency in the case $p = 2$ (Normal case). Suppose that the underlying distribution F is in fact nondegenerate bivariate normal $\Phi(\mathbf{0}, \sigma_1^2, \sigma_2^2, \rho)$. Then the efficiency behavior of $\mathbf{M}_n$ and $\mathbf{W}_n$ with respect to $\dot{\mathbf{X}}_n$ is described by the following theorems.

THEOREM 5.1. *The efficiency of* $\mathbf{M}_n$ *with respect to* $\dot{\mathbf{X}}_n$ *is independent of* σ_1 *and* σ_2 *and is given by*

$$(5.1) \qquad (a) \qquad e(\mathbf{M}_n, \dot{\mathbf{X}}_n)(\rho) = \frac{2}{\pi}\left[\frac{1-\rho^2}{1-(1-(2/\pi)\cos^{-1}\rho)^2}\right]^{\frac{1}{2}}$$

or in the analytically simple form

$$(5.1) \qquad \text{(b)} \qquad e(\mathbf{M}_n, \dot{\mathbf{X}}_n)(u) = \sin u/[u(\pi - u)]^{\frac{1}{2}}$$

where u is determined by the relation $\rho = \cos u$. The function $e(\mathbf{M}_n, \dot{\mathbf{X}}_n)$ is (1) Monotone decreasing for $0 \leqq \rho < 1$ and (2) Symmetric about $\rho = 0$ and hence unimodal.

Finally, $\lim_{|\rho|\to 1} e(\mathbf{M}_n, \dot{\mathbf{X}}_n)(\rho) = 0$.

Proof. In the bivariate case if $\rho_{12} = \rho$, (4.4) reduces to

$$(5.2) \qquad e(\mathbf{M}_n, \mathbf{X}_n) = 4\sigma_1\sigma_2 f_1(0) f_2(0)[(1 - \rho^2)/1 - (4p_{12} - 1)^2]^{\frac{1}{2}}$$

where $p_{12} = P_0[X_{11} > 0, X_{21} > 0]$.

In the normal case we have

$$(5.3) \qquad f_i^2(0) = (2\pi\sigma_i^2)^{-1}$$

and

$$(5.4) \qquad \rho = \cos p\pi \quad \text{where} \quad p = 1 - 2p_{12}.$$

Of these, (5.3) is self-evident and (5.4) is a well-known result due to Sheppard (see [1], p. 290). Simplifying and letting $u = p\pi$ we obtain (5.1)(a) and (b).

To see (1) let $f(u) = \sin^2 u/u(\pi - u)$. Then

$$(5.5) \qquad [\log f(u)]' = 2 \cot u - [u^{-1} - (\pi - u)^{-1}]$$

vanishes for $u = \frac{1}{2}\pi$ and we see

$$(5.6) \qquad [\log f(u)]'' = -2\csc^2 u + u^{-2} + (\pi - u)^{-2} \leqq 0 \qquad \text{for } u \leqq \tfrac{1}{2}\pi,$$

since

$$\csc u = 1/u + u/6 + \cdots + [2(2^{2n} - 1)/2n!]B_n u^{2n-1} + \cdots$$

where the $B_n > 0$.

Statement (1) now follows readily. Statements (2) and (3) are immediate. We note that this third statement is essentially (4.10).

Theorem 5.2. *The efficiency of $\mathbf{W}_n$ with respect to $\dot{\mathbf{X}}_n$ is independent of σ_1 and σ_2 and is given by*

$$(5.8) \qquad \text{(a)} \qquad e(\mathbf{W}_n, \dot{\mathbf{X}}_n)(\rho) = \frac{3}{\pi}\left[\frac{1 - \rho^2}{1 - 9\left(1 - \frac{2}{\pi}\cos^{-1}\frac{\rho}{2}\right)^2}\right]^{\frac{1}{2}}$$

or in the analytically convenient form

$$\text{(b)} \qquad e(\mathbf{W}_n, \dot{\mathbf{X}}_n)(u) = \tfrac{3}{2}\left[\frac{-(1 + 2\cos\frac{2}{3}(\pi + u))}{u(\pi - u)}\right]^{\frac{1}{2}}$$

where u is determined by $\rho = 2\cos(u + \pi)/3$. The function $e(\mathbf{W}_n, \dot{\mathbf{X}}_n)$ is (1) monotone decreasing for $0 \leqq \rho < 1$ and (2) symmetric about $\rho = 0$ and hence uni-

modal. Finally we have

$$\lim_{|\rho|\to 1} e(\mathbf{W}_u, \dot{\mathbf{X}}_u)(\rho) = (3/\pi \sin \pi/3)^{\frac{1}{2}} = .91.$$

PROOF. (4.2) in the bivariate case reduces to

$$(5.9) \qquad e(\mathbf{W}_n, \dot{\mathbf{X}}_n) = 12\sigma_1\sigma_2 \int_{-\infty}^{\infty} f_1^2(x)\,dx \int_{-\infty}^{\infty} f_2^2(x)\,dx[(1-\rho^2)/(1-\gamma_{12}^2)]^{\frac{1}{2}}$$

where γ_{12} is given in (3.3). To prove (5.8) (a) we remark first that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_1(x)F_2(y)\,dF(x,y) = P_0(X - X' > 0, Y - Y'' > 0)$ where X', Y'' are independent of each other and (X, Y) and have the same distribution as X and Y, respectively. Now if (X, Y) has a $\mathbf{\Phi}(\mathbf{0}, \sigma_1^2, \sigma_2^2, \rho)$ distribution, $(X - X', Y - Y'')$ has a $\mathbf{\Phi}(\mathbf{0}, 2\sigma_1^2, 2\sigma_2^2, \rho/2)$ distribution. The result now follows upon employing (5.4) and

$$(5.10) \qquad (1/2\pi\sigma_i^2)\int_{-\infty}^{\infty} \exp\{-x^2/\sigma_i^2\}\,dx = 1/2\pi^{\frac{1}{2}}\sigma_i.$$

We obtain (5.8) (b) by substituting $u = \pi\{3(1 - 2P_0[X - X' > 0, Y - Y'' > 0]) - 1\}$ in (5.9).

To prove (1) consider

$$f(u) = -[1 + 2\cos\tfrac{2}{3}(u + \pi)]/u(\pi - u).$$

Then

$$(5.11) \quad f'(u) = \frac{\frac{4}{3}\sin\frac{2}{3}(u+\pi)u(\pi - u) + (\pi - 2u)[1 + \frac{2}{3}\cos\frac{2}{3}(\pi + u)])}{u^2(\pi - u)^2}$$

vanishes for $u = \pi/2$.

It suffices to show $g(u) \geqq 0$, $0 \leqq u \leqq \pi/2$ where

$$(5.12) \quad g(u) = \tfrac{4}{3}\sin\tfrac{2}{3}(u+\pi)[u(\pi - u)] + (\pi - 2u)[1 + \tfrac{2}{3}\cos\tfrac{2}{3}(\pi + u)].$$

Now

$$(5.13) \qquad g'(u) = -\cos\tfrac{2}{3}(u+\pi)[\tfrac{8}{9}u^2 - \tfrac{8}{9}\pi u + 4] - 2,$$

$g'(0) = 0$ and $g'(\pi/2) < 0$.

Since $g(0) = g(\pi/2) = 0$, it is sufficient to show that $g'(u)$ changes signs exactly once in $[0, \pi/2]$.

Changing variables to $v = \pi - \frac{2}{3}(u + \pi)$ we find

$$(5.14) \qquad g'(u) = h(v) = 2\cos v(v^2 - [\pi^2/9 - 2]) - 2,$$

where $0 \leqq v \leqq \pi/3$.

Now $h(v)$ changes signs if and only if

$$(5.15) \qquad s(v) = \sec v - v^2 + \pi^2/9 - 2$$

does. The theorem is now readily established by triple differentiation of $s(v)$.

Thus in the bivariate case, with an underlying normal law, $e(\mathbf{W}_n, \dot{\mathbf{X}}_n)$ and $e(\mathbf{M}_n, \dot{\mathbf{X}}_n)$ behave quite similarly on the whole. Both depend only on ρ and have a unique maximum when $\rho = 0$. However, as in the univariate case $\mathbf{W}_n$ is appreciably better than $\mathbf{M}_n$ if the underlying distribution is normal and the discrepancy is accentuated as $|\rho| \to 1$. In subsequent robustness considerations we restrict ourselves to $\mathbf{W}_n$.

6. Robustness. In this section we exhibit first some of the advantages of $\mathbf{W}_n$ with respect to $\dot{\mathbf{X}}_n$. In particular, we show that as the underlying distribution of $\mathbf{X}_1$ becomes highly contaminated with "gross errors" $e(\mathbf{W}_n, \dot{\mathbf{X}}_n)$ tends to ∞. More precisely, we have

THEOREM 6.1. *Let* Λ, ψ *be any nonsingular bivariate distributions that are continuous and symmetric and have marginal densities. Let* $0 < \epsilon < 1$, $\boldsymbol{\tau} = (\tau_1, \tau_2)$. *Define*

$$F_{\epsilon,\boldsymbol{\tau}}(x, y) = (1 - \epsilon)\Lambda(x, y) + \epsilon\psi(x/\tau_1, y/\tau_2) \tag{6.1}$$

and

$$e(\epsilon, \boldsymbol{\tau}) = e(\mathbf{W}_n, \dot{\mathbf{X}}_n) \tag{6.2}$$

where $F_{\epsilon,\boldsymbol{\tau}}$ *is the underlying distribution of* $\mathbf{X}_1$ *under* $P_{\mathbf{0}}$.

Then, for every $0 < \epsilon < 1$, $\lim_{\boldsymbol{\tau}\to(\infty,\infty)} e(\epsilon, \boldsymbol{\tau}) = +\infty$.

PROOF. Let $\rho(\epsilon, \boldsymbol{\tau})$ be the correlation coefficient of (X_{11}, X_{21}) and $\sigma_1^2(\epsilon, \boldsymbol{\tau})$, $\sigma_2^2(\epsilon, \boldsymbol{\tau})$ be the variances of X_{11} and X_{21}, respectively, under $F_{\epsilon,\boldsymbol{\tau}}$. Then we have

$$\begin{aligned} \int_{-\infty}^{\infty} f_i^2(\epsilon, \boldsymbol{\tau}, x)\, dx &= (1 - \epsilon)^2 \int_{-\infty}^{\infty} f_i^2(0, \mathbf{1}, x)\, dx \\ &\quad + (\epsilon^2/\tau_i^2) \int_{-\infty}^{\infty} f_i^2(1, \mathbf{1}, x/\tau_i)\, dx \\ &\quad + [2\epsilon(1 - \epsilon)/\tau_i] \int_{-\infty}^{\infty} f_i(0, \mathbf{1}, x) f_i(1, \mathbf{1}, x/\tau_i)\, dx \end{aligned} \tag{6.3}$$

where $\mathbf{1} = (1, 1)$. Applying the Schwarz inequality we find that

$$\int_{-\infty}^{\infty} f_i(0, \mathbf{1}, x) f_i(1, \mathbf{1}, x/\tau_i)\, dx$$

is $o(\tau_i)$ and hence

$$\begin{aligned} \lim_{\boldsymbol{\tau}\to\infty} 12 \int_{-\infty}^{+\infty} f_1^2(\epsilon, \boldsymbol{\tau}, x)\, dx \int_{-\infty}^{\infty} f_2^2(\epsilon, \boldsymbol{\tau}, x)\, dx \\ = 12(1 - \epsilon)^4 \int_{-\infty}^{\infty} f_1^2(0, \mathbf{1}, x)\, dx \int_{-\infty}^{\infty} f_2^2(0, \mathbf{1}, x)\, dx, \end{aligned}$$

which is positive for $\epsilon < 1$.

Upon remarking that $\lim_{\boldsymbol{\tau}\to\infty} \sigma_i^2(\epsilon, \boldsymbol{\tau}) = \infty$ for $\epsilon > 0$ and that $\lim_{\boldsymbol{\tau}\to\infty} \rho(\epsilon, \boldsymbol{\tau}) = \rho(1, \mathbf{1})$, where $|\rho(1, \mathbf{1})| < 1$, we conclude from (5.9) that

$$e(\epsilon, \boldsymbol{\tau}) \geqq (12\sigma_1(\epsilon, \boldsymbol{\tau})\sigma_2(\epsilon, \boldsymbol{\tau}) \int_{-\infty}^{\infty} f_1^2(\epsilon, \boldsymbol{\tau}, x)\, dx \int_{-\infty}^{\infty} f_1^2(\epsilon, \boldsymbol{\tau}, x)\, dx) \to \infty$$

as $\boldsymbol{\tau} \to \infty$.

It may readily be shown that the same result is obtained if we let only one of τ_1, τ_2 tend to ∞ keeping the other fixed. A similar result can clearly be established for $\mathbf{M}_n$.

Unfortunately, we also have the following theorem showing not only that even in the case $p = 2$ there exist distributions for which $e(\mathbf{W}_n, \dot{\mathbf{X}}_n)$ is arbitrarily small but that they may arise from contamination models similar to the one above. These remarks are embodied in the statement and proof of

THEOREM 6.2. *Let $\mathfrak{F}$ be the family of all continuous symmetric bivariate distributions. Then*

$$\inf_{F\epsilon\mathfrak{F}} e(\mathbf{W}_n, \dot{\mathbf{X}}_n) = 0. \tag{6.4}$$

PROOF. Let $\Lambda(x, y)$, $\psi(x, y)$ be two distributions on the line $x = y$ such that both are continuous, symmetric, and have unit variances. Let Λ_1, Λ_2 be the marginal distributions of X_{11}, X_{21}, respectively, under Λ, and similarly let ψ_1, ψ_2 be the marginal distributions of X_{11}, X_{21} under ψ. Denote expressions of the form $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Lambda_1(x)\Lambda_2(y)\, d\Lambda(x, y)$ by $(\Lambda_1, \Lambda_2, \Lambda)$. We observe that there exist Λ, ψ such that

$$3(\Lambda_1, \Lambda_2, \Lambda) - [(\Lambda_1, \psi_2, \Lambda) + (\Lambda_1, \Lambda_2, \psi) + (\psi_1, \Lambda_2, \Lambda)] \neq 0. \tag{6.5}$$

(6.5) is achieved for instance, by taking Λ uniform on the interval $(-(6)^{\frac{1}{2}}, (6)^{\frac{1}{2}})$ on $x = y$, and ψ double exponential on the same line.

Consider the family of bivariate distributions

$$F_{(\epsilon,b)}(x, y) = (1 - \epsilon)\Lambda(x, y) + \epsilon b\psi(x, y) + \epsilon(1 - b)\psi_1(x)\psi_2(y).$$

Let $\sigma_i^2(\epsilon)$, $F_i(\epsilon, x)$, $f_i(\epsilon, x)$ denote the variance, marginal distribution and marginal density of X_{i1} when the joint distribution of (X_{11}, X_{21}) under P_0 is given by $F_{(\epsilon,b)}(x, y)$.

Clearly $\sigma_i^2(\epsilon) = 1$, $F_i(\epsilon, x)$ and $f_i(\epsilon, x)$ are independent of b. Moreover,

$$\int_{-\infty}^{\infty} f_i^2(\epsilon, x)\, dx \rightarrow \int_{-\infty}^{\infty} f_i^2(0, x)\, dx$$

as $\epsilon \rightarrow 0$.

Denote quantities such as

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_1(0, x)F_2(1, y)\, dF_{(0,b)}(x, y) \quad \text{by} \quad (F_1(0), F_2(1), F(0, b)).$$

Now applying L'Hopital's rule and (5.9) we find

$$\lim_{\epsilon\to 0} e(\mathbf{W}_n, \dot{\mathbf{X}}_n)(\epsilon, b) = 12\left(\int_{-\infty}^{\infty} f_1^2(0, x)\, dx \int_{-\infty}^{\infty} f_2^2(0, x)\, dx\right)(1 - b)/H(b) \tag{6.6}$$

where

$$\begin{aligned} H(b) = {} & 3(\Lambda_1, \Lambda_2, \Lambda) \\ & -[(F_1(0), F_2(1), F(0, b)) + (F_1(0), F_2(0), F(1, b)) \\ & + (F_1(1), F_2(1), F(0, b))]. \end{aligned}$$

Now as $b \rightarrow 1$, $H(b)$ tends to the expression in (6.4). Hence we conclude

$$\lim_{b\to 1} \lim_{\epsilon\to 0} e(\mathbf{W}_n, \dot{\mathbf{X}}_n)(b, \epsilon) = 0 \tag{6.7}$$

and our result follows.

Thus it would seem that, despite our encouraging result on gross errors, on the whole $\mathbf{W}_n$ tends to misbehave badly when different highly correlated bivariate random variables are mixed and, hence, should not be used even in the presence of gross errors for $|\rho|$ too close to 1. Of course (4.6) and Theorem 6.1 would suggest that $\mathbf{W}_n$ is a more acceptable estimate than $\dot{\mathbf{X}}_n$ for $|\rho|$ close to 0. This is confirmed by the elementary observation made in Section 4 that $e(\mathbf{W}_n, \dot{\mathbf{X}}_n) \geqq .86 (1 - \rho^2)^{\frac{1}{2}}$. Of course, the remarks made in connection with the case $\rho = \frac{1}{2}$ for general p also apply here.

The pathological behavior of $\mathbf{M}_n$ and $\mathbf{W}_n$ may in part be due to their lack of affine invariance, a property which the problem, of course, possesses.

It is also interesting to note that if quadratic loss, rather than generalized variance, is used as a criterion of efficiency, the univariate results carry over verbatim, as might be expected, since this really reduces to consideration of the case when the components are independent.

In a forthcoming paper we shall examine Hotelling type tests based on these estimates and on the corresponding nonparametric tests, and compare them to Hotelling's T^2 test.

Acknowledgment. I should like to thank Professor Erich Lehmann for suggesting the problem and for his great help and encouragement during the progress of the work.

REFERENCES

[1] Cramér, H. (1956). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[2] Hodges, J. L., Jr. and Lehmann, E. L. (1956). Efficiency of some nonparametric competitors of the *t*-test. *Ann. Math. Statist.* **27** 324–335.

[3] Hodges, J. L., Jr. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

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

[5] Lehmann, E. L. (1964). Asymptotically nonparametric inference in some linear models with one observation per cell. *Ann. Math. Statist.* **35** 726–734.

[6] Stuart, A. (1958). Equally correlated variates and the multinormal integral. *J. Roy. Statist. Soc. Ser. B* **20** 373–378.