The Annals of Statistics
1975, Vol. 3, No. 5, 1070–1100

THE BEHAVIOR OF ROBUST ESTIMATORS ON DEPENDENT DATA

BY JOSEPH L. GASTWIRTH[1] AND HERMAN RUBIN[2]

George Washington University and Purdue University

This paper investigates the effect of serial dependence in the data on the efficiency of some robust estimators. When the observations are from a stationary process satisfying certain mixing conditions, linear combinations of order statistics and the Hodges–Lehmann estimator are shown to be asymptotically normally distributed. Gaussian processes are studied in detail and it is shown that when all the serial correlations (ρ_n) are $\geqq 0$, the efficiency of the robust estimators relative to the mean is greater than in the case of independent observations.

1. Introduction and summary. Many robust estimators of the location parameter of a symmetric unimodal distribution have been proposed in the past several years. All of them have the desirable property of being relatively insensitive to outliers or "wild observations." This paper investigates the effect of serial dependence in the data on the efficiency of the following estimators: the mean, median, trimmed mean, the average of two symmetric percentiles and the Hodges–Lehmann estimator. We study the estimators when the observations are assumed to come from a strongly mixing strictly stationary process (S.S.P.). Gaussian processes are studied in great detail but we also study the behavior of some of the estimators on a first order autoregressive process with a double-exponential marginal distribution (F.O.A.D.P.). One general result (Theorem 4.1) states that for any Gaussian process for which all the serial correlations $\{\rho_n\}$ are nonnegative, the efficiency of any linear combination of the order statistics relative to the mean is greater than the corresponding efficiency in the case of independent observations. The same result holds for the efficiency of the Hodges–Lehmann estimator. On the other hand, on first order autoregressive Gaussian processes (F.O.A.G.P.'s) as ρ approaches -1, the efficiency of any *finite* linear combination of sample percentiles relative to the mean approaches 0. The corresponding efficiencies of the Hodges–Lehmann estimator and the trimmed mean have a nonzero limit.

On the F.O.A.D.P., the median is the most efficient estimator studied, although

---

Received September 1970; revised October 1974.

[1] This research was supported by the National Science Foundation Research Grants No. GP71108 and GP20527 awarded to the Department of Statistics, George Washington University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] This research was supported in part by the Office of Naval Research Contract N00014-67-A-0226-0008, and project number NRO42-216 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

AMS 1970 *subject classifications.* Primary 62G35; Secondary 62E20, 60J99.

Key words and phrases. Order statistics, robust estimation, strong mixing, stationary processes, relative efficiency.

it is not the best possible one. For all values of ρ, the median is twice as efficient as the mean while the Hodges-Lehmann estimator (HL) is always more efficient than the mean but less efficient than the median. In contrast with the Gaussian case, the relative efficiency of the HL estimator gets worse as $\rho \to +1$ and better as $\rho \to -1$.

Using the probabilistic results of [5] the asymptotic theory of all estimators studied is derived. Linear combinations of order statistics are discussed in Section 2 and the Hodges-Lehmann estimator in Section 3.

Sections 4 and 5 are devoted to a study of the relative efficiency of our estimators on Gaussian and double-exponential processes.

The last section (6) of the paper is concerned with various models of contamination which lead to dependent processes. Recently, Hoyland [10] studied the behavior of the HL estimator in a contamination model which is a special case of our first model. The second model leads to a stationary process which has a contaminated normal (in the sense of Tukey [16]) marginal distribution.

2. The asymptotic distribution of a general linear estimator. In [5] we proved that the empiric cdf formed from a strong mixing Δ_s process converges to a Gaussian process. This implies that any sufficiently smooth linear combination of the order statistics is asymptotically normally distributed. In this section we derive an expression for the asymptotic variance of any linear combination of order statistics. We do not, however, investigate the exact conditions required for its validity. After deriving some general formulas we specialize to Gaussian processes in order to illustrate their use on Gaussian processes.

If $X_{(1)} < \cdots < X_{(n)}$ are the order statistics from a sample of size n from a S.S.P., a linear estimator W is a statistic of the form ([4])

$$W = n^{-1} \sum_{i=1}^{n} w_i X_{(i)} , \tag{2.1}$$

where

$$w_i = n\nu\left[\frac{i-1}{n}, \frac{i}{n}\right], \qquad i = 1, \cdots, n \tag{2.2}$$

and ν is a measure of variation 1 and finite total variation on $[0, 1]$. If $\lambda \in (0, 1)$, then $X_{(i)}$ for $i = [n\lambda]$ is the sample λth fractile. If W_1 and W_2 denote the αth and βth sample fractiles and $x = F^{-1}(\alpha)$, $y = F^{-1}(\beta)$ the population fractiles, the asymptotic joint distribution of W_1 and W_2 is given by

THEOREM 2.1. *If f is continuous at x and y, then the sample fractiles from a strongly mixing Δ_s process are asymptotically jointly normally distributed with means x and y and covariance given by*

$$n^{-1} \sum_{q=-\infty}^{\infty} \frac{P[X_0 < x, X_q < y] - \alpha\beta}{f(x)f(y)} . \tag{2.3}$$

PROOF. For each observation X_i define the indicator rv's

$$\begin{aligned} Y_i(\alpha) &= 1 , \quad X_i < x & \qquad Y_i(\beta) &= 1 , \quad X_i < y \\ &= 0 , \quad \text{otherwise,} & &= 0 , \quad \text{otherwise.} \end{aligned} \tag{2.4}$$

Using the Bahadur [1] approach to the representation of a quantile (see also [6], [12], [15])

$$(2.5)\qquad W_1 - x \sim -[nf(x)]^{-1}S_1\,, \qquad W_2 - y \sim -[nf(y)]^{-1}S_2\,,$$

where

$$(2.6)\qquad S_1 = \sum_{i=1}^{n} [Y_i(\alpha) - \alpha] \quad \text{and} \quad S_2 = \sum_{i=1}^{n} [Y_i(\beta) - \beta]\,.$$

The rv's S_1 ond S_2 are jointly asymptotically normally distributed and the calculation of their covariance proceeds as follows:

$$(2.7)\qquad \begin{aligned} \operatorname{Cov}(S_1, S_2) &= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \operatorname{Cov}[Y_i(\alpha), Y_j(\beta)] \\ &= n^{-1} \sum_{q=-(n-1)}^{(n-1)} (n - |q|) \operatorname{Cov}[Y_0(\alpha), Y_q(\beta)] \\ &\to \sum_{q=-\infty}^{+\infty} \operatorname{Cov}[Y_0(\alpha), Y_q(\beta)] \qquad \text{as } n \to \infty\,. \end{aligned}$$

As $\operatorname{Cov}[Y_0(\alpha), Y_q(\beta)] = P[X_0 < x, X_q < y] - \alpha\beta$, substituting (2.7) into (2.5) yields (2.3).

In particular we have

COROLLARY 2.1. *If* $Z_{(i)}$, $i = 1, \cdots, k$ *are the* λ_ith *sample percentiles and if* $\sum_1^k w_i = 1$, *then* $W = \sum_{i=1}^k w_i Z_{(i)}$ *is asymptotically normally distributed with mean* $\sum_{i=1}^k w_i F^{-1}(\lambda_i)$, *and variance*

$$(2.8)\qquad \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \sum_{q=-\infty}^{\infty} \frac{P[X_0 < F^{-1}(\lambda_i), X_q < F^{-1}(\lambda_j)] - \lambda_i \lambda_j}{f[F^{-1}(\lambda_i)] f[F^{-1}(\lambda_j)]}$$

provided that $\{X_i\}$ *is a strongly mixing* Δ_s S.S.P. *with a continuous density at* $F^{-1}(\lambda_i)$, $i = 1, \cdots, k$.

REMARK. Actually Theorem 2.1 and its Corollary are valid for S.S.P.'s satisfying the conditions of Theorem 2.1 of [5].

At this point we shall operate heuristically. Assuming that ν is sufficiently regular, the general linear estimator W based on the measure ν will have asymptotic variance

$$(2.9)\qquad V(n^{\frac{1}{2}}W) \sim \sum_{q=-\infty}^{\infty} \int_0^1 \int_0^1 \frac{P[X_0 < F^{-1}(\alpha), X_q < F^{-1}(\beta)] - \alpha\beta}{f[F^{-1}(\alpha)] f[F^{-1}(\beta)]}\, d\nu(\alpha)\, d\nu(\beta)\,.$$

REMARK. If ν is a positive measure and all rv's X_0 and X_q are positive quadrant dependent ([13]) then the variance of W is greater than in the case of independent observations.

The interchange of summation and integration is valid in the case of a finite number of sample percentiles but requires justification in general. For estimators such as the trimmed mean, if the density is sufficiently smooth at the trimming points, formula (2.9) holds. At this point it may be instructive to note that the term in (2.9) for $q = 0$ is the variance in the case of independent observations so that (2.9) can be regarded as the variance in the independent case with correction terms for each "qth order dependence."

For purposes of calculation it is often convenient to express the measure ν on

(0, 1) in terms of an equivalent measure μ on $(-\infty, \infty)$ defined by

$$d\nu[F(x)] = f(x)\, d\mu(x) .$$

In terms of μ, (2.9) becomes

$$(2.10) \qquad V(n^{\frac{1}{2}}W) \sim \sum_{q=-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{P[X_0 < x, X_q < y] - P[X_0 < x][P[X_q < y]\}\, d\mu(x)\, d\mu(y) .$$

In order to use formula (2.10) one needs a reasonable expression for $P[X_0 < x, X_q < y] - P[X_0 < x]P[X_q < y]$. Fortunately, for normal rv's we have

LEMMA 2.1. *If X and Y are two correlated standard normal* rv's *with correlation coefficient η, then*

$$(2.11) \qquad P[X < a, Y < b] - P[X < a]P[Y < b] = \frac{1}{2\pi} \exp[-(a^2 + b^2)/2] \sum_{k=1}^{\infty} H_{k-1}(a)H_{k-1}(b) \frac{\eta^k}{k!} ,$$

where $H_k(a)$ is the kth Hermite polynomial.

PROOF. The bivariate normal density function can be expressed in terms of the Hermite polynomials as follows ([8]):

$$(2.12) \qquad \begin{aligned} &(2\pi)^{-1}(1 - \eta^2)^{-\frac{1}{2}} \exp[-(x^2 + y^2 - 2\eta xy)/2(1 - \eta^2)] \\ &= (2\pi)^{-1} \exp[-(x^2 + y^2)/2] \sum_{k=0}^{\infty} \frac{H_k(x)H_k(y)\eta^k}{k!} \\ &= \psi(x, y, \eta) . \end{aligned}$$

The probability desired is

$$(2.13) \qquad \begin{aligned} &\int_{-\infty}^{a} \int_{-\infty}^{b} [\psi(x, y, \eta) - \varphi(x)\varphi(y)]\, dx\, dy \\ &= \int_{-\infty}^{a} \int_{-\infty}^{b} (2\pi)^{-1} \exp[-(x^2 + y^2)/2] \sum_{k=1}^{\infty} \frac{H_k(x)H_k(y)\eta^k}{k!}\, dx\, dy , \end{aligned}$$

where $\varphi(x)$ denotes the standard normal density. Integrating (2.13), after interchanging the summation and integration operations, yields the right side of (2.11).

REMARK. Formula (2.11) remains valid when $\eta = \pm 1$.

Before discussing some examples we introduce an assumption on the measure μ which allows us to freely interchange summation and integration operations. Letting

$$(2.14) \qquad \begin{aligned} \mu(x) &= \int_0^x d\mu(t) , \quad x > 0 , \\ &= \int_{-x}^{0} d\mu(t) , \quad x < 0 , \end{aligned}$$

and its "total variation function"

$$(2.15) \qquad \begin{aligned} \mu^*(x) &= \int_0^x |d\mu|(t) , \quad x > 0 , \\ &= \int_{-x}^{0} |d\mu|(t) , \quad x < 0 , \end{aligned}$$

we shall assume that $\mu^*(x)$ is in L^2 w.r.t. the normal density function $\varphi(x) = (2\pi)^{-\frac{1}{2}}e^{-x^2/2}$. Substituting (2.11) into (2.12) yields

PROPOSITION 2.1. *The asymptotic variance of a general linear estimator on Gaussian processes such that $\sum_k |\rho_k| < \infty$ is given by*

$$V(n^{\frac{1}{2}}W) \sim \sum_{q=-\infty}^{\infty} \sum_{k=1}^{\infty} \frac{c_k}{k!} (\rho_{|q|})^k = \sum_{k=1}^{\infty} \frac{c_k}{k!} \sum_{q=-\infty}^{\infty} (\rho_{|q|})^k , \tag{2.16}$$

where

$$c_k = [\textstyle\int_{-\infty}^{\infty} H_{k-1}(x)\varphi(x)\, d\mu(x)]^2 .$$

REMARK 1. The assumption that $\mu(x)$ is in L^2 guarantees that $\sum_{k=1}^{\infty} c_k/k!$ converges.

REMARK 2. If $d\mu(x)$ is a symmetric measure, i.e., the original measure ν gives equal weight to the ith and $(n + 1 - i)$st order statistics, that its odd Fourier–Hermite coefficients vanish, so that only terms involving c_{2k+1} appear in (2.16).

REMARK 3. As the exact conditions for the validity of expression (2.16) for the asymptotic variance are not known in complete generality, we note that whenever the regularity conditions of Theorem 2.1 and $\sum |\rho_k| < \infty$, the mean, trimmed mean and any finite linear combination of sample percentiles are asymptotically normally distributed with the stated asymptotic variance.

We now discuss several special estimators on normal processes.

EXAMPLE 1. Consider the mean, $\bar{X}$. Here $d\mu(x) = 1$. Since $H_0(x) = 1$, $c_1 = 1$ while $c_k = 0$ if $k \neq 1$. Thus, $V(n^{\frac{1}{2}}\bar{X}) = \sum \rho_{|k|}$ as is well known.

EXAMPLE 2. The median M is represented by a measure $d\mu(x)$ which places an atom of mass $(2\pi)^{\frac{1}{2}}$ at 0 so that $c_{2k+1} = H_{2k}^2(0) = (2k - 1)!!$ while $c_{2k} = 0$. For each q in the right side of (2.16), we have

$$\sum_{k=0}^{\infty} \frac{(2k-1)!!}{(2k+1)(2k)!!} \rho_{|q|}^{2k+1} = \arcsin \rho_{|q|} \tag{2.17}$$

and summing over q yields

$$V(M) \sim \frac{1}{n} \sum_{k=-\infty}^{\infty} \arcsin \rho_{|k|} . \tag{2.18}$$

To illustrate the use of (2.18) consider a simple moving average process. Let $\{Z_i\}$ be i.i.d. standard normal rv's and let $X_1 = (m + 1)^{-\frac{1}{2}}(Z_0 + m + Z_m)$, etc. Then $\rho_{-k} = \rho_k = 1 - |k|/(m + 1)$ if $|k| < m + 1$ and 0 otherwise. Thus (2.18) becomes

$$V(M) \sim \frac{1}{n}\left(2 \sum_{j=1}^{m} \arcsin\left(1 - \frac{j}{m+1}\right) + \frac{\pi}{2}\right) \tag{2.19}$$

which can be regarded as the variance in the case of independent observations plus a correction factor. As the variance of the sample mean, $\bar{X}$, is $(m + 1)/n$, the reciprocal of the efficiency of the median to the mean is

$$\frac{V(M)}{V(\bar{X})} = \frac{2\sum_1^m \arcsin(j/(m+1)) + \pi/2}{m+1}. \tag{2.20}$$

When m goes to infinity at a smaller rate than n, formula (2.20) is a Riemann approximation to

$$2\int_0^1 \arcsin x\,dx = \pi - 2 \sim 1.14. \tag{2.21}$$

It is interesting to notice that the efficiency of M to $\bar{X}$ as m approaches ∞ is about 87.7% which is much higher than the efficiency (63.6%) when the observations are independent.

We now give an example which shows that Gaussian processes exist for which the asymptotic efficiency of the median relative to the mean can be arbitrarily close to one. We choose the $\rho_k > 0$ in a manner that the piecewise linear function connecting them will be convex. Then, by Pólya's theorem, the $\{\rho_k\}$ will be the correlation sequence of a stationary process. Of course, $\rho_0 = 1$. For $k \geqq 1$ define

$$\rho_k = \rho\frac{(K+1)^{1+\varepsilon}}{(K+k)^{1+\varepsilon}} = \rho\left(\frac{K+1}{K+k}\right)^{1+\varepsilon}, \quad 0 < \rho \leqq \frac{1}{2 - \left(\frac{K+1}{K+2}\right)^{1+\varepsilon}} \tag{2.22}$$

where ε will be chosen to be arbitrarily small and K large. Using Riemann approximations we obtain

$$\begin{aligned}\sum_{-\infty}^{\infty} \arcsin \rho_k &= \frac{\pi}{2} + 2\sum_1^\infty \arcsin \rho\left(\frac{K+1}{K+k}\right)^{1+\varepsilon} \\ &\sim \frac{\pi}{2} + 2K\int_1^\infty \arcsin \rho x^{-(1+\varepsilon)}\,dx\end{aligned} \tag{2.23}$$

and

$$\sum_{-\infty}^{\infty} \rho_k \sim 1 + 2K\int_1^\infty \frac{\rho}{x^{1+\varepsilon}}\,dx. \tag{2.24}$$

By choosing K large, the ratio $\lim_{n\to\infty} V(M)/V(\bar{X})$ can be made arbitrarily close to

$$\frac{\int_1^\infty \arcsin \rho x^{-(1+\varepsilon)}\,dx}{\rho\int_1^\infty x^{-(1+\varepsilon)}\,dx}. \tag{2.25}$$

As $\arcsin y \leqq y + (\frac{1}{2}\pi - 1)y^3$, the ratio (2.25) is

$$\leqq \frac{\rho\varepsilon^{-1} + (\frac{1}{2}\pi - 1)\rho^3(2+3\varepsilon)^{-1}}{\rho\varepsilon^{-1}} = 1 + \left(\frac{\pi}{2} - 1\right)\rho\varepsilon(2+3\varepsilon)^{-1}, \tag{2.26}$$

which can be made arbitrarily close to 1, by choosing ε sufficiently small.

EXAMPLE 3. Consider a *finite* combination of sample percentiles, i.e., let ν give weight w_i to $x_{(i)}$, where $i = [n\lambda_i]$, $0 < \lambda_1 < \cdots < \lambda_r < 1$ and let $a_i = \Phi^{-1}(\lambda_i)$, then $d\mu(x) = 0$ if $x \neq a_i$ and $d\mu(a_i) = w_i\varphi(a_i)^{-1}$. In this case $c_k = [\sum_1^r w_i H_{k-1}(a_i)]^2$. Later we shall study the average of two symmetric percentiles, $W(a)$, where

$\lambda_1 = \alpha$, $\lambda_2 = 1 - \alpha$, $a_2 = -a_1$ and $w_1 = w_2 = \frac{1}{2}$. Since $H_{2k}(-a) = H_{2k}(a)$, while $H_{2k+1}(-a) = -H_{2k+1}(a)$, $c_{2k} = 0$ and $c_{2k+1} = H_{2k}^2(a)$, so that

$$V(n^{\frac{1}{2}}W(\alpha)) \sim \sum_{k=0}^{\infty} \frac{H_{2k}^2(a)}{(2k+1)!} \left(\sum_{q=-\infty}^{\infty} \rho_{|q|}^{2k+1}\right) . \tag{2.27}$$

EXAMPLE 4. One of the most widely studied robust estimators is the trimmed mean, defined by

$$T = (1 - 2\alpha)^{-1} \int_{A_n}^{B_n} x \, dF_n(x) , \tag{2.28}$$

where A_n is the sample αth quantile, B_n is the sample $(1 - \alpha)$th quantile and $F_n(x)$ is the empiric cdf. In terms of measures, $d\nu(u) = (1 - 2\alpha)^{-1} du$ for $\alpha < u < 1 - \alpha$ and 0 elsewhere, while $d\mu(x) = (1 - 2\alpha)^{-1}$. Thus, $c_k = (1 - 2\alpha)^{-2}[\int_{-a}^{a} H_{k-1}(x)\varphi(x)\, dx]^2$ which is 0 for even k and equals $4(1 - 2\alpha)^{-2}H_{k-2}^2(a)\varphi^2(a)$ for odd $k > 1$, while $c_1 = 1$. Thus

$$V(n^{\frac{1}{2}}T(a)) \sim T_0 + 2 \sum_{q=1}^{\infty} (\rho_q) + \frac{4e^{-a^2}}{(1-2a)^2\pi} \sum_{j=1}^{\infty} \frac{H_{2j-1}^2(a)}{(2j+1)!} \sum_{q=1}^{\infty} (\rho_q)^k , \tag{2.29}$$

where $T_0 = (1 - 2a)^{-1}[\int_{-a}^{a} x^2 \, dF(x) + 2\alpha a^2]$ is the variance in the independent case.

3. The Hodges–Lehmann estimator. One robust estimator which has received much attention recently is the Hodges–Lehmann [9] estimator which is derived from the Wilcoxon test. If $X_1, \cdots, X_n$ are n observations from $F(x)$, the Hodges–Lehmann estimator, HL, is the median of all the pairwise averages of the X's, i.e.,

$$\mathrm{HL} = \mathrm{med} \left\{\frac{X_i + X_j}{2}\right\}_{i,j=1}^{n} . \tag{3.1}$$

In this section we shall give general conditions for the asymptotic normality of HL and show that they are satisfied by strongly mixing Gaussian processes such that $\sum |\rho_k| < \infty$. Moreover, we shall evaluate explicitly the asymptotic variance of HL for these Gaussian processes and the autoregressive double-exponential process so that the effect of serial correlation in the data can be explored numerically.

Instead of working with the Hodges–Lehmann estimator it is more convenient to discuss an asymptotically equivalent estimator which is defined in terms of the empiric cdf $F_n(t)$ by

$$M^* = 2\,\mathrm{HL} = \mathrm{median}\,\{F_n(t) * F_n(t)\} , \tag{3.2}$$

where * denotes convolution. The estimator HL* is the median of all pairwise averages, where the average of a pair of distinct observations is counted twice while the individual observations are counted once. This is just a consequence of the fact that $F_n * F_n$ places mass $2n^{-2}$ at the $\binom{n}{2}$ points of the form $x_i + x_j$ if $i \neq j$, and places mass n^{-2} at the n points of the form $2x_i$. If M^* denotes the median of $F_n * F_n$, i.e., $F_n * F_n(M^*) = \frac{1}{2}$, then $\mathrm{HL}^* = M^*/2$. The idea underlying the proof of asymptotic normality of HL* is similar to the proof of asymptotic normality of the sign test. The number of pairs of observations X_i, X_j such that $X_i + X_j < 0$ is $n^2 F_n * F_n(0)$ and should be asymptotically normally distributed. Using the density $g(0)$ of $F * F$ at 0 in place of $f(0)$, one can convert the asymptotic normality of $F_n * F_n(0)$ into the asymptotic normality of M^*.

The results in [5] imply that HL* is asymptotically normal; however, HL* is asymptotically normal under weaker conditions. In terms of the conditions on the empiric process

$$G_n(t) = n^{\frac{1}{2}}[F_n(t) - F(t)] \tag{3.3}$$

in a neighborhood of 0, we have

THEOREM 3.1. *The estimator*, HL*, *is asymptotically normally distributed whenever the following conditions are satisfied*:

$$(G_n * F)(0) \quad \textit{is asymptotically normally distributed,} \tag{3.4a}$$

(3.4b) *there exist two sequences of reals* w_n *and* λ_n *such that*

$$w_n \to 0\,, \quad \lambda_n \to \infty \quad \textit{but} \quad \lambda_n n^{-\frac{1}{2}} \to 0\,,$$

$$\sup_{|x|<\lambda_n n^{-\frac{1}{2}}} P\{|G_n * F(x) - (G_n * F)(0)| > w_n\} \to 0\,,$$

and

$$\sup_{|x|<\lambda_n n^{-\frac{1}{2}}} P\{|n^{-\frac{1}{2}} G_n * G_n(x)| > w_n\} \to 0 \qquad \textit{as} \quad n \to \infty\,, \tag{3.4c}$$

$$(F * F)'(0) \quad \textit{exists.} \tag{3.4d}$$

Moreover,

$$n^{\frac{1}{2}} H^* + G_n * F(0)/(F * F)'(0) \to_p 0 \tag{3.5}$$

and

$$n^{\frac{1}{2}} H^* \sim N(0, \sigma^2) \tag{3.6}$$

where σ^2 *is the variance of* $(G_n * F)(0)/(F * F)'(0)$.

Before giving the proof we should like to discuss the assumptions. Clearly,

$$F_n * F_n = F * F + 2n^{-\frac{1}{2}} G_n * F + n^{-1} G_n * G_n\,. \tag{3.7}$$

Assumptions (b) and (c) state that there is a neighborhood about zero of order larger than $n^{-\frac{1}{2}}$ such that in the neighborhood $G_n * F$ is essentially $(G_n * F)(0)$ while $n^{-1}(G_n * G_n)$ is essentially 0. Thus, in a neighborhood of order greater than $n^{-\frac{1}{2}}$, $(F_n * F_n) - (F * F)$ differs from the random variable $2n^{-\frac{1}{2}} G_n * F(0)$ by terms of order $o_p(n^{-\frac{1}{2}})$. Asymptotic normality of the sign-type statistic then follows from the first assumption and the fourth assumption (d) guarantees the asymptotic normality of H^*.

PROOF. Letting α denote $(F * F)'(0)$, by the definition of a derivative and assumption (d), there exists a sequence $\rho_n \to 0$ such that

$$|(F * F)(x) - (F * F)(0) - \alpha x| \leqq \rho_n |x| \tag{3.8}$$

for $|x| < \lambda_n n^{-\frac{1}{2}}$. Next select a sequence $Q_n \to \infty$ slowly so that

(3.9) $$Q_n \sup_{|x|<\lambda_n n^{-\frac{1}{2}}} P\{|(G_n * F)(x) - (G_n * F)(0)| > w_n\} \to 0 ,$$

(3.10) $$Q_n \sup_{|x|<\lambda_n n^{-\frac{1}{2}}} P\{|n^{-\frac{1}{2}}(G_n * G_n(x))| > w_n\} \to 0 ,$$

and $Q_n \rho_n \to 0$. The existence of a sequence Q_n satisfying conditions (3.9) and (3.10) follows from assumptions (b) and (c) respectively. Finally, select a sequence $h_n \to 0$ such that $Q_n h_n \to \infty$, $Q_n h_n < \lambda_n$ and $Q_n h_n \rho_n \to 0$. (The interval $(-Q_n h_n n^{-\frac{1}{2}}, Q_n h_n n^{-\frac{1}{2}})$ is the desired neighborhood of zero which is larger than $n^{-\frac{1}{2}}$.) At all points of the form

(3.11) $$x = \frac{kh_n}{n^{\frac{1}{2}}}, \quad Q_n \leqq k \leqq Q_n ,$$

relations (3.9) and (3.10) imply that

(3.12) $$|F_n * F_n(x) - F * F(x) - 2n^{-\frac{1}{2}} G_n * F(0)| \leqq 3w_n n^{-\frac{1}{2}}$$

except with small probability. Substituting (3.8) in (3.12), shows that with large probability (w.l.p.)

(3.13) $$|F_n * F_n(x) - F * F(0) - \alpha x - 2n^{-\frac{1}{2}}(G_n * F)(0)| < \rho_n|x| + 3w_n n^{-\frac{1}{2}} .$$

In particular, it follows from (3.13) that w.l.p. if $F_n * F_n(x) \leqq \frac{1}{2}$, then

(3.14) $$\alpha x + 2n^{-\frac{1}{2}}(G_n * F)(0) \leqq \rho_n|x| + 3w_n n^{-\frac{1}{2}}$$

while if $F_n * F_n(x) \geqq \frac{1}{2}$, then

(3.15) $$\alpha x + 2n^{-\frac{1}{2}}(G_n * F)(0) \geqq -\rho_n|x| - 3w_n n^{-\frac{1}{2}} .$$

Thus, w.l.p.

(3.16) $$n^{\frac{1}{2}}[(F_n * F_n)(Q_n h_n n^{-\frac{1}{2}}) - \tfrac{1}{2}] \geqq \alpha Q_n h_n + 2(G_n * F)(0) - \rho_n Q_n h_n - 3w_n > 0$$

since $(G_n * F)(0)$ has bounded variance, $\rho_n Q_n h_n \to 0$, $w_n \to 0$ but $\alpha Q_n h_n \to \infty$. Similarly w.l.p.

(3.17) $$n^{\frac{1}{2}}[(F_n * F_n)(-Q_n h_n n^{-\frac{1}{2}}) - \tfrac{1}{2}] < 0 ,$$

so that w.l.p. M^* lies in the interval $[-Q_n h_n/n^{\frac{1}{2}}, Q_n h_n/n^{\frac{1}{2}}]$ and there exists a k such that

(3.18) $$n^{-\frac{1}{2}}kh_n \leqq M^* \leqq n^{-\frac{1}{2}}(k+1)h_n , \quad -Q_n \leqq k \leqq Q_n .$$

Let

(3.19) $$Y = n^{\frac{1}{2}}(\alpha M^* + 2n^{-\frac{1}{2}}(G_n * F)(0)) .$$

Since $(F_n * F_n)((k+1)h_n n^{-\frac{1}{2}}) \geqq \frac{1}{2}$, it follows from (3.11) that w.l.p.

(3.20) $$Y \geqq \rho_n Q_n h_n - 3w_n - h_n .$$

As $(F_n * F_n)(kh_n n^{-\frac{1}{2}}) \leqq \frac{1}{2}$, (3.14) implies that w.l.p.

(3.21) $$Y \leqq \rho_n Q_n h_n + 3w_n + h_n .$$

As h_n, w_n and $\rho_n Q_n h_n$ all approach 0 as $n \to \infty$, (3.20) and (3.21) imply that $(Y) \to 0$ in probability. Hence

$$n^{\frac{1}{2}}M^* + \frac{2(G_n * F)(0)}{\alpha} \to_P 0 \tag{3.22}$$

or

$$n^{\frac{1}{2}}\mathrm{HL}^* + \frac{2(G_n * F)(0)}{\alpha} \to_P 0 \tag{3.23}$$

and the asymptotic distribution of $n^{\frac{1}{2}}\mathrm{HL}^*$ is that of $(G_n * F)(0)/\alpha$.

In the appendix we verify that the assumptions of Theorem 3.1 are satisfied by strongly mixing Gaussian processes such that $\sum |\rho_k| < \infty$. A similar (but simpler) argument will show that the first order autoregressive double-exponential process also obeys Theorem 3.1 so that HL is asymptotically normally distributed for data from the processes we shall discuss. We now proceed to calculate the variance of the asymptotic distribution of HL for observations from these processes.

In order to calculate the variance of $F * G_n(0)$, we express it as

$$F * G_n(0) = n^{-\frac{1}{2}} \sum_{i=1}^n \{F(X_i) - E[F(-X_i)]\}\,. \tag{3.24}$$

In particular, when f is symmetric about 0,

$$F * G(0) = n^{-\frac{1}{2}} \sum_{i=1}^n \{F(X_i) - \tfrac{1}{2}\} \tag{3.25}$$

and

$$\mathrm{Var}\,[F * G_n(0)] = \frac{1}{n} \sum_i \sum_j E[F(X_i) - \tfrac{1}{2}][F(X_j) - \tfrac{1}{2}]\,. \tag{3.26}$$

A non-trivial use of the representation (3.26) occurs in the derivation of the asymptotic variance of HL on double-exponential first order autoregressive data. Specifically, we have

THEOREM 3.2. *When $\{X_i\}$ is a* F.O.A.D.P. *the asymptotic variance of the Hodges–Lehmann estimator is given by*

$$nV[\mathrm{HL}] \sim \tfrac{4}{3} + \frac{3\rho}{1-\rho} - 3\left(\sum_{j=1}^{\infty} \frac{\rho^{3j}}{(2+\rho^j)^2}\right) \quad \text{if} \quad \rho > 0 \tag{3.27}$$

and

$$nV[\mathrm{HL}] \sim \tfrac{3}{2} \sum_{j=-\infty}^{\infty} \left(\rho^{|j|} - \frac{\rho^{3j}}{(2+|\rho|^{|j|})^2}\right) \quad \text{if} \quad \rho < 0\,. \tag{3.28}$$

PROOF (Outline). From (3.23), $nV[\mathrm{HL}] = 16V[F * G_n(0)]$ on double-exponential data. Since the cdf of the double-exponential distribution is

$$F(u) = \tfrac{1}{2} + \tfrac{1}{2}(1 - e^{-|u|})\,\mathrm{sgn}\, u\,, \tag{3.29}$$

(3.26) becomes

$$nV[F * G_n(0)] = (\tfrac{1}{4}) \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} E[(1 - e^{-|Y_j|})(1 - e^{-|Y_k|})\,\mathrm{sgn}\,(Y_j Y_k)]\,. \tag{3.30}$$

Using the Markov nature of the process and letting $n \to \infty$, one obtains

$$V[F * G_n(0)] \sim (\tfrac{1}{4}) \textstyle\sum_{j=-\infty}^{\infty} E[(1 - e^{-|Y_0|})(1 - e^{-|Y_j|}) \operatorname{sgn}(Y_0 Y_j)] \,. \tag{3.31}$$

The integrals in (3.31) are evaluated using the representation given in [6], [7].

For Gaussian processes it is more convenient to make a direct computation in terms of the rv's which indicate whether $(X_i + X_j)$ is less than 0 (the median) and convert this to obtain the asymptotic variance of M^*. Of course $\text{HL}^* = M^*/2$.

Let

$$\begin{aligned} I_{ij} &= +1\,, \quad \text{if} \quad (X_i + X_j) > 0\,, \\ &= -1\,, \quad \text{if} \quad (X_i + X_j) < 0\,, \end{aligned} \tag{3.32}$$

and $S = \sum_{\text{all pairs}} I_{ij}$.

The random variables X_i, X_k, X_j, X_l are jointly Gaussian so the covariance matrix of the two random variables $X_i + X_j$ and $X_k + X_l$ is

$$\begin{pmatrix} 2(1 + \rho_{|j-i|}) & \gamma \\ \gamma & 2(1 + \rho_{|k-l|}) \end{pmatrix}, \tag{3.33}$$

where $\gamma = \rho_{|k-i|} + \rho_{|j-l|} + \rho_{|i-l|} + \rho_{|j-k|}$ and the correlation between $X_i + X_j$ and $X_k + X_l$ is

$$\rho^* = \frac{\gamma}{2(1 + \rho_{|j-i|})^{\frac{1}{2}}(1 + \rho_{|k-l|})^{\frac{1}{2}}} \,. \tag{3.34}$$

Using

$$\begin{aligned} \operatorname{Cov}(I_{ij}, I_{kl}) &= 4(P[X_i + X_j > 0,\, X_k + X_l > 0] - \tfrac{1}{4}) \\ &= \frac{2}{\pi} \arcsin \rho^* \,, \end{aligned}$$

and counting the contribution of the various terms to Var (S) one can show that

$$n^{-3} \operatorname{Var}(S) \sim \frac{8}{\pi} \textstyle\sum_h \arcsin \dfrac{\rho_{|h|}}{2} \,. \tag{3.35}$$

Now the statistic corresponding to the sign test statistic is the number S^* of pairwise means (or pairwise sums) that are < 0. Essentially, $S^* = \frac{1}{2}n^2 - \frac{1}{2}S$ so that

$$n^{-3} \operatorname{Var}(S^*) \sim \frac{2}{\pi} \textstyle\sum_h \arcsin \dfrac{\rho_{|h|}}{2} \,. \tag{3.36}$$

The same derivation as given in Section 2, with n replaced by n^2 and $f(0)$ replaced by the density of $X_i + X_j$ at 0, which is $\frac{1}{2}\pi^{\frac{1}{2}}$ in the normal case by Lemma 3.1 yields

$$\operatorname{Var}(M^*) = n^{-1} 8 \textstyle\sum_h \arcsin \dfrac{\rho_{|h|}}{2} \,. \tag{3.37}$$

As $\text{HL}^* = M^*/2$, the asymptotic variance of $n^{\frac{1}{2}}\text{HL}$ is given by

$$\operatorname{Var}(n^{\frac{1}{2}}\text{HL}) \sim 2 \textstyle\sum_{h=-\infty}^{\infty} \arcsin \dfrac{\rho_{|h|}}{2} \,. \tag{3.38}$$

4. The efficiency of the estimators relative to the mean in Gaussian processes. In this section we study the efficiency of our estimators relative to $\bar{X}$ Gaussian processes. In particular, their behavior on the F.O.A.G.P. is analyzed in detail. We first show that all linear estimators are efficiency-robust against positive dependence (all $\rho_k \geqq 0$). We then specialize to data from a F.O.A.G.P. and evaluate the relative efficiency of our estimators for various values of ρ. A short table, Table 4.1 is presented which summarizes the behavior of the median M, the Hodges–Lehmann estimator HL, the mid-mean (25% trimmed mean), the 5% trimmed mean and the average of the 25th and 75th percentiles for various values of ρ. A more extensive survey of the behavior of the α-trimmed mean, $T(\alpha)$, and the average of two symmetric percentiles, $W(\alpha)$, as α (the fractile used for trimming or averaging) varies is presented in Table 4.2. For the estimator $W(\alpha)$ it turns out that the optimum choice of α in the case of *independent* observations remains nearly optimum for small values of ρ. The behavior of the relative efficiencies of our estimators as $\rho \to -1$ is also quite interesting. As $\rho \to -1$, the A.R.E. of the median or any *finite linear combination* of sample percentiles approaches 0 while that of HL or $T(\alpha)$ approaches a finite limit (see the Appendices). This is in sharp contrast with the case of independent observations where the efficiency of M to $\bar{X}$ is always $\geqq \frac{1}{3}$ provided that the density sampled is symmetric and unimodal.

In order to discuss the efficiency of linear estimators we require

LEMMA 4.1. *If S is any estimator such that*

$$V(n^{\frac{1}{2}}S) \sim \sum_{q=-\infty}^{\infty} g(\rho_q)\,, \tag{4.1}$$

where $g(\rho)$ is a function satisfying $g(\rho) \leqq \rho g(1)$, then

$$\lim_{n\to\infty} \frac{V(n^{\frac{1}{2}}S)}{V(n^{\frac{1}{2}}\bar{X})} = \frac{\sum_{q=-\infty}^{\infty} g(\rho_q)}{\sum_{-\infty}^{\infty} \rho_q} \leqq g(1)\,. \tag{4.2}$$

In particular, Lemma 4.1 is applicable whenever $g(\rho)/\rho$ is an increasing function of ρ. Typically $g(1)$ is the asymptotic variance of $n^{\frac{1}{2}}S$ in the case of independent observations. We next apply Lemma 4.1 to derive

THEOREM 4.1. *The efficiency of any unbiased linear combination of the order statistics (obeying (2.9)) relative to $\bar{X}$ on strong mixing Gaussian process such that $p_k \geqq 0$ for all k and $\sum |p_k| < \infty$ is always greater than or equal to its value when the observations are independent.*

PROOF. The variance of $n^{\frac{1}{2}}W$, for any linear combination W, is given by (2.17). Setting

$$g(\rho) = \sum_{k=1}^{\infty} \frac{c_k \rho^k}{k!}\,, \tag{4.3}$$

and recalling that $c_k \geqq 0$, the result follows from Lemma 4.1.

REMARK. This efficiency-robustness result depends heavily on the assumption that the process is Gaussian. The fact that on F.O.A.D.P.'s, the efficiency of M to $\bar{X}$ always is 2 suggests that this result will not be generally true.

Using formula (3.38) the analog for the Hodges–Lehmann estimator can be obtained. We state

THEOREM 4.2. *The efficiency of the Hodges–Lehmann estimator*, HL, *to the sample mean* $\bar{X}$, *on strongly mixing Gaussian* S.S.P.*'s such that* $\rho_k \geqq 0$ *for all* k *and* $\sum \rho_k < \infty$, *is always greater than or equal to its value*, $3/\pi$, *in the case of* i.i.d. *Gaussian observations.*

For the remainder of the section we shall assume that the $\{X_i\}$ are a F.O.A.G.P. Two important linear estimators are the average of the αth and $(1-\alpha)$th quantile, $W(\alpha)$, and the α-trimmed mean, $T(\alpha)$, the average of all observations between the αth and $(1-\alpha)$th quantiles. For convenience we specialize the results in Section 2 in

PROPOSITION 4.1. *On* F.O.A.G.P.*'s,*

$$V(n^{\frac{1}{2}}W(\alpha)) \sim \pi\alpha e^{a^2} + 2\sum_{j=0}^{\infty} \frac{\rho^{2j+1}}{1-\rho^{2j+1}} \cdot \frac{H_{2j}^2(a)}{(2j+1)!} \tag{4.4}$$

$$V(n^{\frac{1}{2}}T(\alpha)) \sim T_0 + \frac{2\rho}{1-\rho} + \frac{4e^{-a^2}}{(1-2\alpha)^2\pi}\sum_{j=1}^{\infty} \frac{H_{2j-1}^2(a)}{(2j+1)!} \cdot \frac{\rho^{2j+1}}{1-\rho^{2j+1}} . \tag{4.5}$$

An interesting and readily derived general result is

PROPOSITION 4.2. *On* F.O.A.G.P.*'s, the asymptotic variance of any unbiased linear estimator approaches* ∞ *as* $\rho \to +1$ *and the asymptotic variance of any symmetric estimator approaches* 0 *as* $\rho \to -1$.

Usually robustness results study the efficiency of an estimator relative to the optimum one. As $\bar{X}$ is asymptotically efficient for Gaussian processes one needs the *reciprocal* of the *efficiency* of any unbiased linear estimator (W) which is given by

$$\lim_{n\to\infty} \frac{V(n^{\frac{1}{2}}W)}{V(n^{\frac{1}{2}}\bar{X})} = \sum_{k=1}^{\infty} \frac{c_k}{k!} \cdot \frac{1+\rho^k}{1-\rho^k} \cdot \frac{1-\rho}{1+\rho} \tag{4.6}$$

as $V(n^{\frac{1}{2}}\bar{X}) = (1-\rho)/(1+\rho)$. An interesting monotonicity property of the relative efficiency is based on the following elementary

LEMMA 4.2.

For all $l > 0$ $\dfrac{1+\rho^l}{1-\rho^l} \cdot \dfrac{1-\rho}{1+\rho}$ *decreases as* ρ *goes from* 0 *to* 1 .

For odd $l > 0$ $\dfrac{1+\rho^l}{1-\rho^l} \cdot \dfrac{1-\rho}{1+\rho}$ *decreases as* ρ *goes from* -1 *to* $+1$.

Applying Lemma 4.2 to expression (4.6) yields

COROLLARY 4.1. *The efficiency of any unbiased linear estimator W relative to $\bar{X}$ on data from a* F.O.A.G.P. *is an increasing function of ρ for $0 < \rho < 1$. If W is a symmetric linear estimator the relative efficiency is a monotonically increasing function of ρ for $-1 < \rho < 1$.*

PROOF. The first part of the Corollary is trivial as $c_k \geqq 0$ and each term in (4.6) decreases as ρ increases. The second part of the Corollary follows as $c_{2k} = 0$ for symmetric estimators.

Corollary 4.1 implies that for any symmetric linear combination of the order statistics the situations when ρ approaches $+1$ and -1 yield bounds for the relative efficiency. Using L'Hospital's rule we can derive

THEOREM 4.3. *The reciprocal of the efficiency of any symmetric linear estimator relative to $\bar{X}$ as $\rho \to +1$ or -1 is given by*

$$\lim_{\rho\to 1} \frac{V(n^{\frac{1}{2}}W)}{V(n^{\frac{1}{2}}\bar{X})} = 1 + \sum_{j=1}^{\infty} \frac{c_{2j+1}}{(2j+1)(2j+1)!}\,, \tag{4.7}$$

and

$$\lim_{\rho\to -1} \frac{V(n^{\frac{1}{2}}W)}{V(n^{\frac{1}{2}}\bar{X})} = \sum_{j=0}^{\infty} \frac{c_{2j+1}}{(2j)!}\,. \tag{4.8}$$

Expression (4.8) may be infinite.

Expressions (4.7) and (4.8) can be evaluated for the median without recourse to the explicit values of c_{2j+1}. As this analysis also applies to the HL estimator we formally present

THEOREM 4.4. *If $\rho > 0$, then $V(\bar{X}) < V(H) < V(M)$,*

$$\lim_{\rho\to +1} \frac{V(\bar{X})}{V(M)} = \frac{2}{\pi \log 2} \sim .9184 \tag{4.9}$$

and

$$\lim_{\rho\to +1} \frac{V(\bar{X})}{V(H)} \sim .9853\,. \tag{4.10}$$

PROOF. As $nV(M) \sim \sum \arcsin \rho^k$, and $nV(H) \sim 2 \sum \arcsin (\rho^k/2)$ the first assertion follows from the elementary inequality: $x \leqq 2 \arcsin (x/2) \leqq \arcsin x$. The limiting efficiencies are evaluated by using the fact that the arcsin function has a Taylor series, i.e.,

$$\arcsin x = \sum_{j=1}^{\infty} a_j x^{2j+1}\,. \tag{4.11}$$

In terms of this expansion,

$$nV(M) = \sum_{k=-\infty}^{\infty} \arcsin \rho^{|k|} = \sum_j a_j \sum_k (\rho^{2j+1})^{|k|} = \sum a_j \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}}\,. \tag{4.12}$$

Thus, the reciprocal of the efficiency is asymptotically

$$\frac{V(\bar{M})}{V(\bar{X})} = \sum_j a_j \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} \cdot \frac{1-\rho}{1+\rho}\,. \tag{4.13}$$

As $\rho \to +1$, the jth term approaches $a_j/(2j+1)$, so that

$$\lim_{\rho\to 1} \frac{V(\bar{X})}{V(M)} = \sum_j \frac{a_j}{(2j+1)} . \tag{4.14}$$

From the expansion (4.11), it follows that

$$\sum_j \frac{a_j x^{2j}}{(2j+1)} = \frac{1}{x} \int_0^x \frac{\arcsin y}{y} \, dy = \frac{1}{x} \int_0^{\sin^{-1} x} z \cot z \, dz . \tag{4.15}$$

Evaluation of (4.15) at $x = 1$ yields

$$\lim_{\rho\to 1} \frac{V(M)}{V(\bar{X})} = \sum \frac{a_j}{(2j+1)} = \int_0^{\pi/2} z \cot z \, dz = \frac{\pi}{2} \log 2 . \tag{4.16}$$

Similarly, the asymptotic variance of the Hodges–Lehmann estimator is expressible as

$$V(H) = 2 \sum_{k=-\infty}^{\infty} \arcsin \frac{\rho^{|k|}}{2} = \sum_j a_j 2^{-2j} \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} . \tag{4.17}$$

Proceeding as before the reciprocal of the limiting relative efficiency is obtained by evaluating expression (4.15) at $x = \frac{1}{2}$. Thus,

$$\lim_{\rho\to 1} \frac{V(H)}{V(\bar{X})} = 2 \int_0^{\pi/6} z \cot z \, dz = 2\left[\frac{\pi}{6} - \sum_{j=1}^{\infty} B_j \frac{(\pi/3)^{2j+1}}{(2j+1)!}\right], \tag{4.18}$$

where the B_j are the Bernoulli numbers.

From formula (4.17) we see that the reciprocal of the efficiency of the Hodges–Lehmann estimator is given by a formula of the same type as (4.6). As $\rho \to -1$, the asymptotic efficiencies of M and HL are given by

THEOREM 4.5. *If* $\rho < 0$, *then* $V(\bar{X}) < V(\text{HL}) < V(M)$,

$$\lim_{\rho\to -1} \frac{V(\bar{X})}{V(M)} = 0 , \tag{4.19}$$

and

$$\lim_{\rho\to -1} \frac{V(\bar{X})}{V(\text{HL})} = \frac{3^{\frac{1}{2}}}{2} . \tag{4.20}$$

The behavior of the median as $\rho \to -1$ is quite interesting because n times its variance decreases to zero as $\rho \to -1$ and yet its efficiency relative to $\bar{X}$ approaches 0. In the Appendices we show that this characteristic of any *finite* linear combination of sample percentilés. As robustness studies are usually concerned with the sensitivity of procedures to small departures from the basic assumptions we present Table 4.1 of the efficiencies of several robust estimators for various values of ρ. All these estimators, which are robust against outliers, are robust against positive serial correlation. For all ρ, the 5% trimmed mean is the most robust as one would expect as it is the estimator which is "nearest" $\bar{X}$. For small ρ, i.e., $-.3 \leq \rho \leq +.3$, the relative efficiency of the HL estimator is within 3.5% of value in the case of i.i.d. observations. The efficiencies of the other estimators appear to be more sensitive.

TABLE 4.1
The asymptotic efficiency of some estimators relative to $\bar{X}$ on first order autoregressive Gaussian processes

ρ	M	H	5% Trimmed Mean	Mid-Mean	Average of 25th and 75th Percentiles
−1.0	.0000	.8660	.9000	.5000	.0000
−.9	.1517	.8673	.9129	.5418	.2423
−.8	.2214	.8714	.9200	.5688	.3608
−.7	.2801	.8783	.9267	.5995	.4559
−.6	.3344	.8875	.9337	.6332	.5323
−.5	.3868	.8985	.9408	.6686	.5952
−.4	.4384	.9106	.9482	.7047	.6489
−.3	.4895	.9229	.9556	.7407	.6960
−.2	.5399	.9348	.9626	.7753	.7377
−.1	.5892	.9456	.9690	.8076	.7749
0	.6363	.9549	.9744	.8367	.8079
.1	.6815	.9628	.9790	.8622	.8370
.2	.7234	.9691	.9826	.8838	.8627
.3	.7619	.9740	.9854	.9018	.8853
.4	.7967	.9778	.9875	.9164	.9051
.5	.8278	.9806	.9891	.9230	.9225
.6	.8549	.9826	.9903	.9367	.9377
.7	.8780	.9839	.9910	.9430	.9510
.8	.8968	.9847	.9915	.9471	.9624
.9	.9109	.9851	.9918	.9493	.9715
1.0	.9184	.9853	.9919	.9501	.9766

The results in Table 4.1 do not provide a comprehensive survey of the behavior of the estimators $T(\alpha)$ and $W(\alpha)$ since one is interested in the optimal choice of α and how this value changes with ρ. In the case of independent normal observations, Mosteller [14] showed that the optimal choice of α for $W(\alpha)$ is about .27. Of course, the optimal choice of α for $T(\alpha)$ is 0 since $T(0)$ is the asymptotically efficient estimator $\bar{X}$.

It is interesting to observe that the optimal choice of α for $W(\alpha)$ does not vary very much from its value in the independent case. When $\rho = .9$, the optimal value for α is about .20 and when $\rho = -.9$, the optimal choice for α is about .35. Moreover, for $.2 \leqq \alpha \leqq .4$, the efficiency of $W(\alpha)$ is always higher than the efficiency of the median. Since $W(.27)$ or any approximation to it such as $W(.25)$, is only slightly harder to compute than the median and is quite a bit more efficient than the median for independent and first order autoregressive Gaussian data, its use in practice as a quick estimator can be recommended. For small ρ, $-.2 \leqq \rho \leqq .2$, an interpolation showed that $W(.27)$ remains nearly optimum. Finally, a glance at Table 4.1 shows that $W(.25)$ behaves very similarly to $T(.25)$ on most first order autoregressive processes so that our results also support the claims in recent literature [2], [4], [11], [16] concerning the robustness properties of the mid-mean, $T(.25)$.

The results reported above are based on Table 4.2.

TABLE 4.2
The asymptotic efficiency of $W(\alpha)$ and $T(\alpha)$ relative to $\bar{X}$

ρ	$T(.10)$	$W(.10)$	$T(.2)$	$W(.2)$	$T(.3)$	$W(.3)$	$T(.4)$	$W(.4)$
−1	.8000	.0000	.6000	.0000	.4000	.0000	.2000	.0000
−.9	.8221	.1341	.6363	.2131	.4476	.2683	.2739	.2817
−.8	.8345	.1967	.6574	.3146	.4833	.3936	.3323	.3622
−.7	.8464	.2511	.6806	.4037	.5226	.4807	.3872	.4212
−.6	.8592	.3044	.7068	.4823	.5634	.5470	.4384	.4737
−.5	.8732	.3581	.7352	.5507	.6049	.6021	.4883	.5234
−.4	.8882	.4120	.7649	.6100	.6466	.6509	.5374	.5717
−.3	.9034	.4652	.7947	.6619	.6877	.6952	.5856	.6186
−.2	.9180	.5173	.8234	.7075	.7274	.7358	.6326	.6638
−.1	.9314	.5677	.8500	.7479	.7647	.7727	.6775	.7067
0	.9430	.6160	.8736	.7838	.7987	.8059	.7196	.7463
.1	.9528	.6621	.8940	.8158	.8289	.8353	.7581	.7823
.2	.9608	.7060	.9111	.8446	.8550	.8610	.7927	.8142
.3	.9670	.7473	.9250	.8703	.8772	.8831	.8232	.8421
.4	.9720	.7858	.9361	.8936	.8956	.9021	.8496	.8660
.5	.9754	.8210	.9446	.9147	.9103	.9183	.8720	.8860
.6	.9780	.8524	.9509	.9332	.9219	.9320	.8904	.9025
.7	.9797	.8794	.9554	.9497	.9304	.9434	.9050	.9156
.8	.9808	.9016	.9582	.9637	.9362	.9530	.9158	.9253
.9	.9814	.9180	.9598	.9745	.9395	.9607	.9227	.9320
1.0	.9816	.9269	.9602	.9804	.9406	.9652	.9252	.9356

5. The behavior of some estimators on the first order autoregressive double-exponential process. So far we have only studied Gaussian processes. In order to investigate the effect that the marginal distribution has on our results, in this section we summarize the behavior of some robust estimators on the first order autoregressive process with double-exponential marginal (F.O.A.D.P.). As this process is a rather special one we shall omit proofs. Not surprisingly the results indicate that the median, M, remains the best estimator of the four estimators studied. More interesting is the fact that the HL estimator is more efficient than the mean, $\bar{X}$, for all values of ρ so that it retains its desirable robustness property. In contrast with the Gaussian situation, however, the efficiency of the HL estimator to M (the best one considered) decreases as $\rho \to +1$ and increases as $\rho \to -1$. In the case of independent observations from a double-exponential distribution the efficiency of the HL estimator to the median is 75%. In the case of observations from a F.O.A.D.P., as $\rho \to +1$ this efficiency drops to 69.8% while it rises to 90% as $\rho \to -1$.

The basic tool needed to derive the asymptotic variances of the estimators is given in [6] and [7]. Briefly, the result says that the F.O.A.D.P. satisfies a stochastic difference equation of the form

$$X_i = \rho X_{i-1} + (q \cdot 0 + (1 - q)\varepsilon_1)\,,$$

where 0 stands for a rv which is degenerate at the origin and ε_1 is an independent double-exponential rv. Thus, a F.O.A.D.P. exists and the error term is a mixture of a degenerate rv and a double-exponential rv. Using this representation and formula (2.3) one obtains

THEOREM 5.1. *The average $W(\alpha)$ of the upper and lower* 100αth *percentiles on a* F.O.A.D.P. *is asymptotically normally distributed with expectation the mean of the process and asymptotic variance given by*

$$\lim_{n\to\infty} nV[W(\alpha)] = \sum_{k=-\infty}^{\infty} (\rho^{|k|} \cosh \nu\rho^{|k|} + \sinh \nu\rho^{|k|})$$
$$= e^{\nu} + 2\sum_{j=0}^{\infty} \frac{\nu^{2j}}{(2j)!} \frac{\rho^{2j+1}}{1-\rho^{2j+1}} + 2\sum_{j=0}^{\infty} \frac{\rho^{2j+1}}{1-\rho^{2j+1}} \frac{\nu^{2j+1}}{(2j+1)!},$$

where ν is the upper αth *point of the double-exponential distribution.*

Our first result shows that, as is the case of independent observations, the median is the most efficient estimator of the form $W(2)$. Specifically, one can derive

PROPOSITION 5.1. *The median has the minimum asymptotic variance of any average of two symmetric percentiles on any* F.O.A.D.P.

When we compare the efficiency of any estimator $W(\alpha)$ to the median as ρ varies we obtain the following analog of Corollary 4.1:

PROPOSITION 5.2. *The efficiency of any average of two symmetric order statistics, $W(\alpha)$, to M on* F.O.A.D.P.*'s is an increasing function of ρ.*

The asymptotic variance of the Hodges–Lehmann estimator was given in (3.27). It is always more efficient than $\bar{X}$ but less efficient than the median. Using the methods of the previous section one obtains

THEOREM 5.2. *On data from a* F.O.A.D.P. *the limiting ratios of the asymptotic variances of the estimators considered, as $\rho \to \pm 1$ are given by*

$$\lim_{\rho\to\pm1} V(\bar{X})/V(M) = 2,$$
$$\lim_{\rho\to+1} V(H)/V(M) = 2[3\log(\tfrac{3}{2}) - \tfrac{1}{2}] \approx 1.4328,$$
$$\lim_{\rho\to+1} V[W(\alpha)]/V(M) = \frac{\sinh\nu}{\nu} + \sum_{j=0}^{\infty} \frac{\nu^{2j+1}}{(2j+1)(2j+1)!},$$

where ν is the upper αth *fractile of the double-exponential* cdf,

$$\lim_{\rho\to-1} V(H)/V(M) = 10/9$$

and

$$\lim_{\rho\to-1} V[W(\alpha)]/V(M) = \cosh\nu + \nu e^{\nu}.$$

In Table 5.1 we present the asymptotic efficiency of the HL estimator and several averages of symmetric percentiles, $W(.45)$, $W(.4)$, $W(.25)$, and $W(.1)$ relative to M for various choices of ρ. One interesting observation is that all the estimators seem rather more sensitive to small values of ρ than in the Gaussian case. For instance, for Gaussian data the HL estimator has efficiency .995 at

TABLE 5.1
The A.R.E. *of the estimators H,* $W(.45)$, $W(.4)$, $W(.25)$ *and* $W(.1)$ *relative to M*

ρ	$V(M)/V(H)$	$V(M)/V(W(.45))$	$V(M)/V(W(.4))$	$V(M)/V(W(.25))$	$V(M)/V(W(.1))$
−1	.9000	.8908	.7669	.3793	.0939
−.9	.8991	.8908	.7672	.3804	.0948
−.8	.8961	.8912	.7684	.3842	.0975
−.5	.8653	.8938	.7777	.4142	.1192
−.3	.8210	.8965	.7871	.4476	.1460
−.1	.7709	.8990	.7962	.4837	.1810
0	.7500	.9000	.8000	.5000	.2000
+.1	.7336	.9008	.8031	.5142	.2139
.3	.7134	.9020	.8075	.5355	.2527
.5	.7038	.9027	.8099	.5483	.2775
.8	.6986	.9031	.8114	.5564	.2958
.9	.6981	.9031	.8116	.5572	.2978
1.0	.6979	.9031	.8116	.5574	.2984

$\rho = 0$, .866 at $\rho = -1$ and .923 at $\rho = -.3$ so that about 36% of the total change in efficiency is attained when $\rho = -.3$. In the double-exponential case 47% of the total change in efficiency already occurs at $\rho = -.3$. This behavior is characteristic of all the estimators.

Probably the most basic conclusion that can be drawn from Table 5.1 is that the efficiency of the HL estimator decreases as ρ increases which is the *exact opposite* of its behavior in the Gaussian case. This suggests that it is not possible to find one estimator which will be robust against positive serial correlations for all autoregressive procosses. As the relative efficiency of the HL estimator to the median appears to be a monotonically decreasing function of ρ achieving its minimum value .6979 at $\rho = 1$, it appears to be more suitable than the ordinary mean for general use. Moreover, the efficiency of the HL estimator on Gaussian processes is much superior to the median or the average of two symmetric percentiles (especially when ρ is negative).

6. Two models allowing for contaminated observations. In this section we study two processes which are models of a basic sequence of i.i.d. rv's $\{\varepsilon_i\}$ which are subject to possible contamination. The first process assumes that the observations come in groups. Each group has a common contaminant and the size of a group is determined by a discrete renewal process. This model can be considered as a generalization of Hoyland's [10] results when each group has the same size (c). The second model assumes that the process alternates between groups of observations with a common contaminant and groups of uncontaminated observations. The stationary marginal distribution of the second process is a contaminated normal distribution, in the sense of Tukey [16] when the $\{\varepsilon_i\}$ and the contaminant are normally distributed (with different variances)·

Our general results imply that for the first process the efficiency of any linear combination of the order statistics or the HL estimator, relative to $\bar{X}$ is greater than its relative efficiency in the case of independent observations whenever the ε_i and the contaminant are normally distributed.

Process 1. Assume that a process is composed of phases whose lengths (L_1, $L_2, \cdots$) form a discrete renewal process, i.e., the probability that a phase lasts for j observations is $p_j = P(L = j)$. When a phase begins a contaminant U is added to a basic sequence ε_i of i.i.d. rv's throughout that phase. Letting $N(i)$ denote the number of renewals (phases) that have occurred by time i, the process X_i is representable as

$$X_i = U_{N(i)} + \varepsilon_i \,. \tag{6.1}$$

Whenever $E(L)$ is finite the process will be asymptotically stationary. Indeed, by choosing the stationary distribution for the renewal process as the time until the first renewal occurs, the process can be made strictly stationary from time 0. We denote the generating function of L by $\varphi(z) = \sum_1^\infty p_j z^j$ and discuss the asymptotic behavior of $\bar{X}$, M and HL when $\varepsilon \sim \eta(0, e^2)$ and $U \sim \eta(0, \sigma^2)$. It will be convenient to assume that $e^2 + \sigma^2 = 1$ and we let $r = \sigma^2/\sigma^2 + e^2$. The asymptotic behavior of the three estimators is given in

PROPOSITION 6.1. *When the observations are from Process* 1, *as* $n \to \infty$

$$V(\bar{X}) \sim \frac{1}{n}(1 + r\varphi''(1)/\omega'(1)) \tag{6.2}$$

$$V(M) \sim \frac{\pi}{2n}\left(1 + \frac{2}{\pi}\frac{\varphi''(1)}{\varphi'(1)} \arcsin r\right) \tag{6.3}$$

and

$$V(\mathrm{HL}) \sim \frac{\pi}{3n}\left(1 + \frac{6}{\pi}\frac{\varphi''(1)}{\varphi'(1)} \arcsin \frac{r}{2}\right). \tag{6.4}$$

PROOF. As usual

$$n^2 V(\bar{X}) = \textstyle\sum_{i=1}^n V(X_i) + \sum\sum_{i \neq j} \mathrm{Cov}\,(X_i, X_j)\,. \tag{6.5}$$

If X_i and X_j are not in the same phase, they are independent. When X_i and X_j are in the same phase, $\mathrm{Cov}\,(X_i, X_j) = r$. As $n \to \infty$, the number of different phases is approximately $n/E(L) = n/\varphi'(1)$. The average number of ordered pairs of distinct observations that occur in a phase is $E(L(L-1)) = \varphi''(1)$. Hence, as $n \to \infty$

$$\textstyle\sum\sum_{i \neq j} \mathrm{Cov}\,(X_i, X_j) \sim \dfrac{n\varphi''(1)r}{\varphi'(1)}\,, \tag{6.6}$$

and

$$V(\bar{X}) \sim \frac{1}{n}(1 + r\varphi''(1)/\varphi'(1))\,. \tag{6.7}$$

To derive (6.3), consider the sign test statistic S_n. Its variance is

$$V(S_n) = \textstyle\sum_{i=j=1}^n (\frac{1}{4}) + \sum_{i \neq j} [P(X_i > 0, X_j > 0) - \frac{1}{4}]\,. \tag{6.8}$$

The terms in the second sum on the right side of (6.8) are 0 unless i and j are in the same phase, in which case $P(X_i > 0, X_j > 0) - \frac{1}{4} = 1/2\pi \arcsin r$. Counting the number of correlated X_i's as in the preceding paragraph yields

$$V(S_n) \sim \frac{n}{4}\left(1 + \frac{\varphi''(1)}{\varphi'(1)} \frac{2}{\pi} \arcsin r\right) \tag{6.9}$$

and (6.3) follows from (6.9).

The derivation of (6.4) follows the one in Section 3. There is one simplification. In calculating the variance of $\sum I_{ij}$, one need only calculate

$$\textstyle\sum_{i,j} [P(X_i + Y > 0, X_j + Z > 0) - \frac{1}{4}] \tag{6.10}$$

where Y, Z are independent of X_i, X_j and of each other and have the stationary distribution of the process. The number of other term is of lower order. When $i = j$, $P(X + Y > 0, X + Z > 0) = \frac{1}{3}$ as all the rv's are symmetric. If i and j are not in the same phase $P(X_i + Y > 0, X_j + Z > 0) = \frac{1}{4}$. When X_i and X_j are in the same phase they are correlated and $X_i + Y$, $X_j + Z$ will be jointly normally distributed (conditionally) with correlation $r/2$. Thus, $P(X_i + Y > 0, X_j + Z > 0) - \frac{1}{4} = 1/2\pi \arcsin r/2$ when i, j are in the same phase, and the variance of the appropriate sign test statistic is

$$n^3\left(\frac{1}{12} + \frac{1}{2\pi} \frac{\varphi''(1)}{\varphi'(1)} \arcsin \frac{r}{2}\right). \tag{6.11}$$

Converting this to the variance of the HL estimator as before yields (6.4).

From formulas (6.2), (6.3) and (6.4) it follows that for small r, if $\varphi''(1)/\varphi'(1)$ approaches ∞, the three estimators exhibit the same behavior. As $n\varphi''(1)/\varphi'(1)$ is the expected number of correlations it is apparent that the variances of the estimators really depend on the total amount of correlation between the observations rather than just on r. In other words, a few really large groups (i.e. $\varphi''(1)$ large) has a greater effect than many small groups. Also, for any fixed value of r, the efficiency of both M and HL relative to $\bar{X}$ is an increasing function of $\varphi''(1)/\varphi'(1)$.

Usually we are concerned with values of $r > \frac{1}{2}$, i.e., the contaminant ordinarily has a larger variance than the underlying i.i.d. rv's. As $r \to 1$, all the terms in the parentheses in formulas (6.2), (6.3) and (6.4) approach $[1 + \varphi''(1)/\varphi'(1)]$ so that their relative efficiencies reduce to their efficiencies in the case of independent observations. This is expected as each group is essentially one observation from the contaminant.

In order to illustrate the results, we present the A.R.E.'s of M and HL to $\bar{X}$ for various values of r and $\tau = \varphi'(1)/[\varphi'(1) + \varphi''(1)]$ in Table 6.1. Of course, for any r the A.R.E. is monotone decreasing in τ as τ is a decreasing function of $\varphi''(1)/\varphi'(1)$.

In all cases, except $\tau = 0$, the efficiencies approach the case of i.i.d. observations where $r = 0$ or 1. For moderate values of τ, i.e., $\varphi''(1)/\varphi'(1)$ is *not* large, the A.R.E.'s of both estimators are not greatly increased compared to the independent case. However, for small values of τ the effect of large values of $\varphi''(1)$ takes over. In the limiting case, $\tau = 0$, when r is small the A.R.E.'s approach one.

Process 1 can be regarded as a generalization of Hoyland's [10] model, in which L is a constant c. He studied the Hodges–Lehmann estimator but gives a formula for arbitrary distributions which essentially replaces $1/2\pi \arcsin r/2$ by $P(X_i + Y > 0, X_j + Z > 0) - \frac{1}{4}$, where X_i and X_j are in the same contamination period and used the general conversion factor of the density of $X_i + X_j$ at 0, i.e., $\int_{-\infty}^{\infty} f^2(t)\, dt$ in the conversion of the sign type statistic to the Hodges–Lehmann estimate. The asymptotic variance of the median given in (6.3) can also be generalized to an arbitrary marginal distribution and is

$$\frac{1}{4f^2(0)n}\left(1 + \frac{\varphi''(1)}{\varphi'(1)}\lambda\right), \tag{6.3*}$$

where $\lambda = P[X_i > 0, X_j > 0] - \frac{1}{4}$ when X_i and X_j are in the same contamination period. Of course, in Hoyland's case, $\varphi''(1)[\varphi'(1)]^{-1} = c - 1$.

It may be instructive to compare the effect of having random group sizes in place of constant group sizes on the variances of our estimators. If we take a geometric distribution for L with average group size c, then $p = 1/c$ and $\varphi''(1)/\varphi'(1) = 2(c - 1)$. Thus the factor due to dependence is doubled if the size of a group is geometrically distributed with mean c rather than always equal to c.

TABLE 6.1

The A.R.E.'*s of M and* HL *for Process* 1

$V(\bar{X})/V(M)$

r	τ						
	0	.1	.2	.3	.5	.7	1.
0	1.0	.637	.637	.637	.637	.637	.637
.1	.998	.767	.710	.683	.658	.646	.637
.3	.985	.858	.789	.745	.693	.663	.637
.5	.955	.875	.819	.775	.716	.676	.637
.7	.903	.854	.813	.779	.725	.683	.637
.9	.804	.781	.760	.741	.706	.676	.637
1·0	.637	.637	.637	.637	.637	.637	.637

$V(\bar{X})/V(\text{HL})$

r	τ						
	0	.1	.2	.3	.5	.7	1.
0	1.0	.955	.955	.955	.955	.955	.955
.1	.999+	.976	.967	.963	.959	.957	.955
.3	.996	.985	.977	.972	.964	.960	.955
.5	.989	.983	.978	.973	.966	.961	.955
.7	.979	.975	.972	.970	.965	.960	.955
.9	.964	.963	.962	.961	.960	.957	.955
1.0	.955	.955	.955	.955	.955	.955	.955

Process 2. The second model assumes the process alternates between two types of periods (or phases). During the first phase X_i equals e_i where the ε_i are i.i.d. standard normal rv's. During the second phase $X_i = \varepsilon_i + U$, where U is a $N(0, \eta)$ random variate independent of the ε_i and is a common component of the observations during the phase. Thus, we have a period of pure observations followed by a period of contaminated ones and then another period of pure observations, etc. We shall call the phases in which uncontaminated observations occur pure periods and the other phases contaminated periods.

While we were motivated by a model which assumed that the lengths of the two types of periods were determined by two independent renewal processes (the so-called alternating renewal process) our analysis is more general. Denoting the starting times of the pure periods by s_j and the starting times of the contaminated periods by t_j our analysis will be valid provided that (s_j, t_j) is a stationary difference process. Let p equal the (stationary) probability that an index i is in a contamination period and let γ denote the expected number of indices $j \neq i$ which lie in the same contamination period as the index i. The number γ is the average number of other observations with which X_i has nonzero correlation. In the case when the lengths of the two periods are determined by renewal processes with generating functions $\Psi(z)$ for the length of pure periods and $\varphi(z)$ for the length of contamination periods we have

$$p = \frac{\varphi'(1)}{\varphi'(1) + \Psi'(1)} \quad \text{and} \quad \gamma = \frac{\varphi''(1)}{\varphi'(1) + \Psi'(1)} = p\,\frac{\varphi''(1)}{\varphi'(1)}\,. \tag{6.12}$$

The marginal distribution of X_i is the contaminated normal,

$$qN(0, 1) + pN(0, 1 + \eta) \tag{6.13}$$

since with probability q, X_i lies in a pure period and has a unit normal distribution while with probability p, X_i lies in a contamination period and has a distribution with mean 0 and variance $1 + \eta$. The asymptotic behavior of $\bar{X}$, M and HL for this process is derived by the same methods used in proving Proposition 6.1. The result is

PROPOSITION 6.2. *When observations come from Process* 2 *the estimators* $\bar{X}$, M *and* HL *are asymptotically normally distributed with asymptotic variances given by*

$$V(\bar{X}) \sim \frac{1}{n}(1 + (p + \gamma)\eta) \tag{6.14}$$

$$V(M) \sim \frac{1}{n}\frac{\pi}{2}\frac{1}{(q + p/(1 + \eta)^{\frac{1}{2}})^2}\left[1 + \frac{2\gamma}{\pi}\arcsin\frac{\eta}{1 + \eta}\right] \tag{6.15}$$

and

$$V(\mathrm{HL}) \sim \frac{1}{n}\frac{\pi}{3}\left\{\left(1 + \frac{6\gamma}{\pi}\left[q^2\arcsin\frac{\eta}{1+\eta} + 2pq\arcsin\frac{\eta}{[(2+\eta)(2+2\eta)]^{\frac{1}{2}}} + p^2\arcsin\frac{\eta}{2+2\eta}\right]\right)\Big/\left(q^2 + \frac{2pq}{(1+\eta/2)^{\frac{1}{2}}} + \frac{p^2}{(1+\eta)^{\frac{1}{2}}}\right)^2\right\}. \tag{6.16}$$

REMARKS. As the variance, η, of the contaminating distribution approaches infinity the variance of $\bar{X}$ approaches infinity while

$$nV(M) \to \frac{\pi}{2}\frac{(1+\gamma)}{q^2}, \tag{6.17}$$

and

$$nV(H) \to \frac{\pi}{3q^4}(1+\gamma(3q^2+3pq+p^2)). \tag{6.18}$$

Thus, the efficiency of HL to the median becomes

$$\tfrac{3}{2}q^2\frac{(1+\gamma)}{1+\gamma(3q^2+3pq+p^2)}. \tag{6.19}$$

In particular, if $p > 1 - 2^{\frac{1}{2}}/3$, the efficiency of HL to the median is < 1, regardless of the value of γ just as in the case of independent contaminated normal observations [3]. However, the Hodges–Lehmann estimator is more sensitive to contamination in this model since (6.19) increases as a function of γ.

REMARK. It should be noted that the analysis given for the two processes in this section really depended only on the following assumptions on the process generating the periods. For Process 1 as long as the lengths l_i and their squares l_i^2 obey a law of large numbers the results, suitably interpreted, will hold. For Process 2 the lengths l_i and their squares, l_i^2, generating the contamination periods must obey a law of large numbers while only the lengths, l_i, of the process generating the pure periods need obey a law of large numbers.

APPENDICES

Appendix to Section 3. *Verification of the conditions of Theorem* 3.1 *for strongly mixing Gaussian processes such that* $\sum |\rho_k| < \infty$. We shall show, in detail, that conditions (3.4a) and (3.4b) are satisfied. The argument showing the condition (3.4c) is satisfied is a tedious calculation which is similar to those in [5] and we omit it. Condition (3.4d) is obviously satisfied.

Verification of condition (a). Letting $\{X_k\}$ denote the rv's of the process we have derived the representation

$$F * G_n(0) = n^{-\frac{1}{2}} \textstyle\sum_{k=1}^{n} \{F(-X_k) - E[F(-X_k)]\}. \tag{3.1*}$$

As the $\{X_k\}$ are strongly mixing, the rv's $F(-X_k)$ are also strongly mixing as they are functions of X_k. The asymptotic normality of the right side of (3.1*) follows from the Blum–Rosenblatt central limit theorem.

Verification of condition (b). Let

$$\beta(x) = \sup_t |F(x+t) - F(t)|. \tag{3.2*}$$

When $F(t)$ is the normal cdf, $\beta(x) \leqq |x|\pi^{\frac{1}{2}}$.

Now

$$\begin{aligned}(3.3^*)\qquad &\operatorname{Var}[F * G_n(x) - F * G_n(0)] \\ &= n^{-1} \textstyle\sum_i \sum_j \operatorname{Cov}[F(x - X_i) - F(-X_i)][F(x - X_j) - F(X_j)] \\ &= n^{-1} \textstyle\sum_i \sum_j \iint [F(x - x_i) - F(-x_i)][F(x - x_j) - F(-x_j)] \\ &\quad \times [dP_{ij}(x_i, x_j) - dP_i(x_i)\, dP_j(x_j)]\,,\end{aligned}$$

where P_{ij} denotes the joint cdf of X_i and X_j while P_i and P_j denote the respective marginal cdf's. Applying (3.2*) to the integrand shows that in a small neighborhood of 0,

$$\begin{aligned}(3.4^*)\qquad &\operatorname{Var}(F * G_n(x) - F * G_n(0)) \\ &\leqq n^{-1}\beta^2(x) \textstyle\sum_i \sum_j \iint |dP_{ij}(x_i, x_j) - dP_i(x_i)\, dP_j(x_j)|\,.\end{aligned}$$

Since $\iint dP_{ij}(x_i, x_j) - dP_j(x_i)\, dP_j(x_j)|$ is a function of $|\rho_{i-j}|$ which is bounded by a constant K times $|\rho_{i-j}|$ as long as the $|\rho_k|$ are bounded away from one. Thus, the right side of (3.4*) is

$$(3.5^*)\qquad \leqq n^{-1}\beta^2(x) \textstyle\sum_i \sum_j K|\rho_{i-j}| \leqq \beta^2(x) \sum_{r=-(n-1)}^{(n-1)} \dfrac{n - |r|}{n} K|\rho_r| \leqq K'\beta^2(x)$$

where K' is another constant. Hence the variance of $[G_n * F(x) - G_n * F(0)]$ can be made uniformly small in a neighborhood of the origin. Applying Chebyshev's inequality yields

$$(3.6^*)\qquad P\{|G_n * F(x) - G_n * F(0)|\} > \frac{K'\beta^2(x)}{w_n^{\,2}}\,.$$

As $\beta(x) \to 0$ as $x \to 0$, in fact at the same rate, any sequence $w_n \to 0$ at a slowei rate than $\lambda_n n^{-\frac{1}{2}} \to 0$ will satisfy the conditions.

REMARK. In order to verified condition (3.4b) for data from an arbitrary continuous distribution one must find a sequence w_n approaching 0 at a slower rate than $\beta(\lambda_n n^{-\frac{1}{2}})$. Then the same argument, replacing ρ_k by $||\Delta(0, k)||_1$, applies in general. The customary tedious fourth order moment argument shows that if $\sum \Delta_k < \infty$, then $P(|(I/n^{\frac{1}{2}})G_n * G_n(x)| > \varepsilon) \to 0$ uniformly in x. In the Gaussian case, we can use the condition $\sum |\rho_k| < \infty$ instead.

Appendix to Section 4. This appendix is concerned with the computation of the limiting A.R.E. of any symmetric linear combination of the order statistics, W, relative to $\bar{X}$ on F.O.A.G.P.'s as $\rho \to -1$. First we need

LEMMA 4.1*. *The reciprocal of the efficiency of any symmetric estimator relative to $\bar{X}$ as $\rho \to -1$, is given by*

$$(4.1^*)\qquad \lim_{t\to 1} \iint \frac{1}{2\pi(1 - t^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2(1 - t^2)}(x^2 - 2txy + y^2)\right] d\mu(x)\, d\mu(y)\,.$$

PROOF. Since $c_{2j} = 0$ for symmetric estimators, (4.8) is just

$$(4.2^*)\qquad \textstyle\sum_{k=1}^{\infty} \dfrac{c_k}{(k - 1)!} = \lim_{t\to 1} \dfrac{d}{dt}\left(\sum_{k=0}^{\infty} \dfrac{c_k t^k}{k!}\right)$$

and

$$\sum \frac{c_k t^k}{k!} = \iint \{P_t[X < x, Y < y] - \Phi(x)\Phi(y)\} \, d\mu(x) \, d\mu(y) \,, \tag{4.3*}$$

where P_t denotes the joint cdf of two standard normal rv's with correlation t. Differentiating (4.3*) with respect to t yields

$$\sum_{k=1}^{\infty} \frac{c_k}{(k-1)!} = \lim_{t\to 1} \iint \frac{1}{2\pi(1-t^2)^{\frac{1}{2}}} \times \exp\left[-\frac{(x^2 - 2txy + y^2)}{2(1-t^2)}\right] d\mu(x) \, d\mu(y) \,. \tag{4.4*}$$

In order to illustrate the use of Lemma 4.1* we prove

COROLLARY 4.1*. *The efficiency of the α-trimmed mean, relative to $\bar{X}$ on F.O.A.G.P.'s, equals* $(1 - 2\alpha)$ *as* $\rho \to -1$.

PROOF. For the α-trimmed mean, let $B = \Phi^{-1}(1 - \alpha)$ so that (4.4*) becomes

$$\lim_{n\to\infty} \frac{V(n^{\frac{1}{2}}T(\alpha))}{V(n^{\frac{1}{2}}\bar{X})} = (1 - 2\alpha)^{-2} \lim_{t\to 1} \int_{-B}^{B} \int_{-B}^{B} \frac{1}{2\pi(1-t^2)^{\frac{1}{2}}} \times \exp\left[-\frac{(x^2 - 2txy + y^2)}{2(1-t^2)}\right] dx \, dy \,. \tag{4.5*}$$

The integral is just the probability that two standard normal rv's with correlation t are both in $(-B, B)$. As $t \to 1$, this approaches the probability that a single standard normal rv is in $(-B, B)$ which is $(1 - 2\alpha)$.

In order to derive the limiting efficiency of a general linear estimator as $\rho \to -1$ we need

LEMMA 4.2*. *For any* $\varepsilon > 0$,

$$\lim_{t\to 1} \iint_{|x-y|>\varepsilon} f_t(x, y) \, d\mu(x) \, d\mu(y) = 0 \,. \tag{4.6*}$$

PROOF. As $f_t(x, y) \geqq 0$, (4.6*) certainly holds if the same limit with μ replaced by μ^* is valid. Since

$$f_t(x, y) = (2\pi(1 - t^2)^{\frac{1}{2}})^{-1} \times \exp\left[-\frac{(x^2 + y^2)}{2(1+t)}\right] \exp\left[-\frac{t}{2(1-t^2)}(x-y)^2\right], \tag{4.7*}$$

$$\begin{aligned} &\iint_{|x-y|>\varepsilon} f_t(x, y) \, d\mu^*(x) \, d\mu(y) \\ &\quad \leqq (1 - t^2)^{-\frac{1}{2}} \exp\left[-\frac{t\varepsilon^2}{2(1-t^2)}\right] \\ &\quad\quad \times \left\{\int (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{x^2}{2(1+t)}\right] d\mu^*(x)\right\}^2 . \end{aligned} \tag{4.8*}$$

Integration by parts and applying the Schwarz inequality yields

$$
\begin{aligned}
&\left\{\int (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{x^2}{2(1+t)}\right] d\mu^*(x)\right\}^2 \\
(4.9^*)\qquad &= \left\{\int \frac{\mu^*(x)}{(2\pi)^{\frac{1}{2}}} \frac{x}{1+t} \exp\left[-\frac{x^2}{2(1+t)}\right] dx\right\}^2 \\
&= \left\{\int \mu^*(x) e^{-x^2/4}(x) \frac{\exp[-(x^2/4)((1-t)/(1+t))]}{1+t} dx\right\} \\
&\leqq \left\{\int \mu^{*2}(x) \frac{e^{-x^2/2}}{(2\pi)^{\frac{1}{2}}} dx\right\} \left\{\frac{1}{(2\pi)^{\frac{1}{2}}(1+t)^2} x^2 \exp\left[-\frac{x^2}{2}\frac{(1-t)}{(1+t)}\right] dx\right\}.
\end{aligned}
$$

As the last term of (4.4*) is $(1-t)^{-\frac{3}{2}}(1+t)^{-\frac{1}{2}}$, the right side of (5.8*) is

$$
(4.10^*)\qquad \leqq K(1+t)^{-1}(1-t)^{-2} \exp\left[-\frac{\varepsilon^2 t}{2(1-t^2)}\right]
$$

which $\to 0$ as $t \to 1$.

In contrast to the limiting behavior of the trimmed mean any estimator based on a measure with an atom at any single order statistic has limiting efficiency 0 (relative to $\bar{X}$). This is shown as follows. As any symmetric unimodal density is a mixture of uniform densities

$$
(4.11^*)\qquad \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{x^2}{2\sigma^2}\right] = \int_{-\infty}^{\infty} \frac{1}{2y} \chi_y(x)\, d\eta_\sigma(y),
$$

where η is a probability measure and $\chi_y(x)$ is the indicator of the set $\{x: |x| \leqq y\}$. Substituting (4.11*) into (4.7*) and (4.1*) means that we must prove that

$$
\begin{aligned}
(4.12^*)\qquad &\lim_{t\to\infty} \frac{1}{t^{\frac{1}{2}}} \iiint \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{(x^2+y^2)}{2(1+t)}\right] \frac{\chi_w(x-y)}{2w}\, dy_t(w)\, d\mu(x)\, d\mu(y) \\
&= \infty .
\end{aligned}
$$

By the lemma we can restrict ourselves to the region $|x-y| < \varepsilon$. Letting $z = \min(\varepsilon, w)$, (4.12*) becomes

$$
\begin{aligned}
(4.13^*)\qquad &\lim_{t\to 1} \frac{1}{t^{\frac{1}{2}}} \int \left\{\iint \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{(x^2+y^2)}{2(1+t)}\right]\right. \\
&\qquad\qquad \left. \times \frac{\chi_z(x-y)}{2x}\, d\mu(x)\, d\mu(y)\right\} \frac{z}{w}\, d\eta_t(w) .
\end{aligned}
$$

As $t \to 1$, $\eta_t(w)$ places more of its mass in a small neighborhood of the origin so that the limit in (4.13*)

$$
(4.14^*)\qquad \lim_{t\to 1;\, z\to 0} \iint \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{(x^2+y^2)}{2(1+t)}\right] \frac{\chi_z(x-y)}{2z}\, d\mu(x)\, d\mu(y) .
$$

When $\mu = \alpha\mu_0 + \beta\mu_1$, where μ_0 is a unit mass at ζ and ζ is *not* an atom of μ_1, evaluation of the double integral in (4.14*) yields

$$(4.15^*)\qquad \frac{\alpha^2 e^{-\zeta^2/(1+t)}}{(2\pi)^{\frac{1}{2}}(2z)} + \alpha\,\frac{\beta e^{-\zeta^2/2(1+t)}}{z} \int_{\zeta-z}^{\zeta+z} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp[-y^2/2(1+t)]\,d\mu_1(y)$$
$$+ \beta^2 \iint \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{(x^2+y^2)}{2(1+t)}\right] \frac{\chi_z(x-y)}{2z}\,d\mu_1(y)\,d\mu_1(y)\,.$$

The third term is $\geqq 0$ as it is essentially the limiting efficiency of an estimator based on μ_1. As $t \to 1$ and $z \to 0$ the first term is $O(1/z)$ while the second is $O(1/z)o(1)$ as μ_1 is "smooth" near ζ (ζ not an atom). Thus the reciprocal of the relative efficiency approaches ∞. In particular the Winsorized mean and a finite linear combination of sample percentiles have this property.

A more general condition for the existence of a positive limiting relative efficiency is formulated in

THEOREM 4.1*. *Let $A(\mu)$ be the Riemann-Hellinger integral*

$$A(\mu) = \int \frac{e^{-x^2/2}}{(2\pi)^{\frac{1}{2}}} \frac{(d\mu(x))^2}{dx}\,.$$

Then

(i) *if μ is a positive measure ($\mu = \mu^*$) and $A(\mu) = \infty$, the limiting relative efficiency is* 0.

(ii) *if $A(\mu)$ exists and is finite and if $B(\mu^*)$, the upper Riemann-Hellinger integral corresponding to A, is finite, then the reciprocal of the limiting relative efficiency is $A(\mu)$.*

PROOF. Let

$$(4.16^*)\qquad H_\mu(z, t, w, v) = \sum_{n=1}^{\infty} \int_{z+nw}^{z+(n+1)w} \int_{z+v+nw}^{z+v+(n+1)w} \frac{1}{(2\pi)^{\frac{1}{2}}}$$
$$\times \exp\left[-\frac{1}{2(1+t)}(x^2+y^2)\right] d\mu(x)\,d\mu(y)\,.$$

Clearly

$$(4.17^*)\qquad \lim_{t\to 1;\, w\to 0} \frac{1}{w} H_\mu(z, t, w, 0) = A(\mu)$$

if $A(\mu)$ exists and

$$(4.18^*)\qquad \lim_{t\to 1;\, w\to 0} \frac{1}{w} H_\mu(z, t, w, 0) \leqq B(\mu)\,.$$

Now let μ be positive. Then

$$(4.19^*)\qquad \frac{1}{2w} H_\mu(z, t, w, 0) \leqq \iint \frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left[-\frac{x^2+y^2}{2(1+t)}\right] \frac{\chi_w(x-y)}{2w}\,d\mu(x)\,d\mu(y)\,,$$

(in Figure 1 the right integral is over the whole strip while the left side is over the shaded area) which, from (4.17*), proves (i).

To prove (ii), let $0 < \delta < 1$ be a fixed number. Define

$$(4.20^*)\qquad \bar{H}_\mu(t, w) = \int_0^w H_\mu(z, t, w, 0)\,dz\,.$$

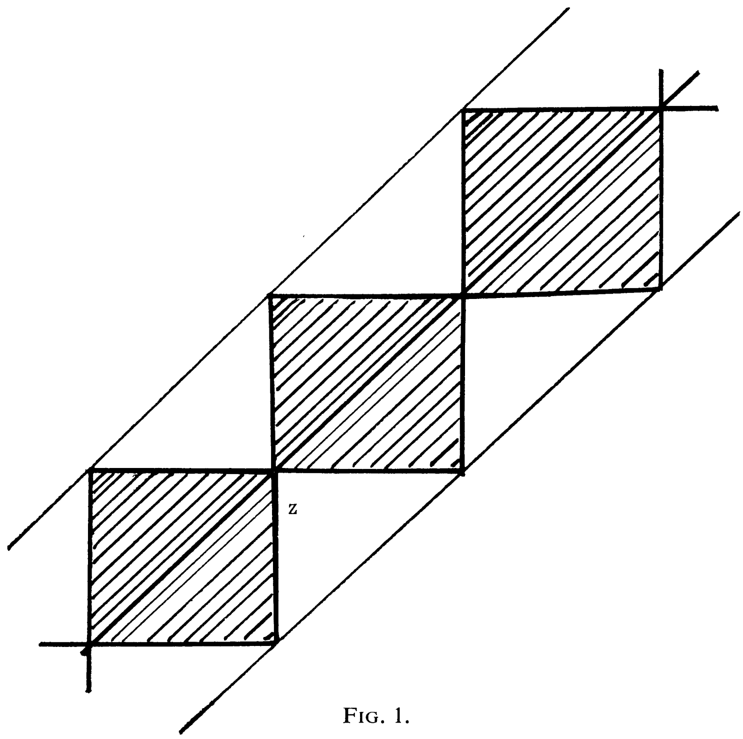

FIG. 1.

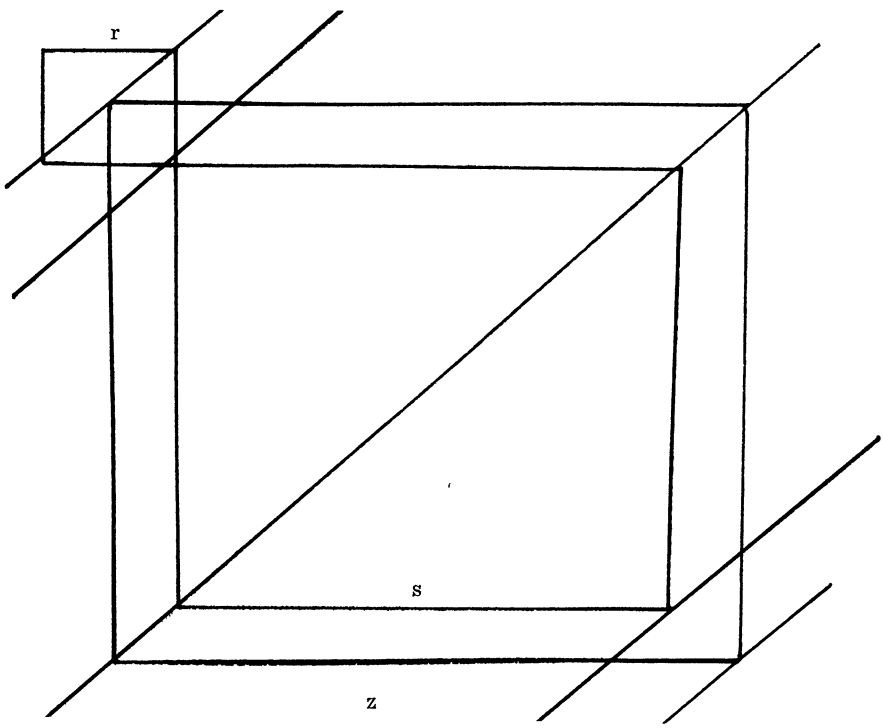

FIG. 2.

Let $s < z$, $r = z - s$. Then (see Figure 2)

$$(4.21^*)\qquad \int_0^w H_\mu(z, t, w, 0)\,dz - \int_0^s H_\mu(z, t, s, 0)\,dz = r \iint \chi_w(x-y)\exp\left[-\frac{x^2+y^2}{2(1+t)}\right] d\mu(x)\,d\mu(y) - R(w, s, t)\,,$$

where

$$(4.22^*)\qquad \begin{aligned} |R(w, s, t)| &= 2\left|\int_0^r \int_{z+(n-1)r}^{z+nr} \int_{z+s+nr}^{z+s+(n+1)r} \chi_w(x-y)\exp\left[-\frac{x^2+y^2}{2(1+t)}\right] d\mu(x)\,d\mu(y)\right| \\ &\leqq 2\int_0^r \int_{z+(n-1)r}^{z+nr} \int_{z+s+nr}^{z+s+(n+1)r} \exp\left[-\frac{x^2+y^2}{2(1+t)}\right] d\mu^*(x)\,d\mu^*(y) \\ &\leqq 2\bar{H}_{\mu^*}(t, r)\,. \end{aligned}$$

Now let $s = (1-\delta)w$, $r = \delta w$. From (4.21*) and (4.22*),

$$(4.23^*)\qquad \left|\frac{1}{2w}\iint \chi_w(x-y)\exp\left[\frac{x^2+y^2}{2(1+t)}\right] d\mu(x)\,d\mu(y) - \frac{1}{2rw}(\bar{H}_\mu(t, w) - \bar{H}_\mu(t, s))\right| \leqq \frac{1}{rw}\bar{H}_{\mu^*}(t, r)\,.$$

Now

$$(4.24^*)\qquad \begin{aligned} \bar{H}_\mu(t, r) &= w^2 A(\mu) + o(w^2)\,, \\ \bar{H}_\mu(t, s) &= s^2 A(\mu) + o(w^2)\,, \\ \bar{H}_{\mu^*}(t, r) &= r^2 B(\mu^*) + o(w^2)\,, \end{aligned}$$

so

$$(4.25^*)\qquad \left|\frac{1}{2w}\iint \chi_w(x-y)\exp\left[-\frac{x^2+y^2}{2(1+t)}\right] d\mu(x)\,d\mu(y) - \left(1-\frac{\delta}{2}\right)A(\mu)\right| \leqq \delta B^*(\mu) + o(1)\,.$$

The result follows easily from (4.25*).

REFERENCES

[1] Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

[2] Crow, E. L. and Siddiqi, M. M. (1967). Robust estimation of location. *J. Amer. Statist. Assoc.* **62** 353–389.

[3] Gastwirth, J. L. and Cohen, M. L. (1970). The small sample behavior of some robust linear estimators of location. *J. Amer. Statist. Assoc.* **65** 946–973.

[4] Gastwirth, J. L. and Rubin, H. (1969). On robust linear estimators. *Ann. Math. Statist.* **40** 24–39.

[5] Gastwirth, J. L. and Rubin, H. (1975). Asymptotic distribution theory of the empiric c.d.f. for mixing stochastic processes. *Ann. Statist.* **3** 809–824.

[6] Gastwirth, J. L., Rubin, H. and Wolff, S. S. (1967). The effect of autoregressive dependence on a non-parametric test. *IEEE Trans. Information Theory* **IT-13** 311–313.

[7] Gastwirth, J. L. and Wolff, S. S. (1965). A characterization of the Laplace distribution. Department of Statistics Report 28, The Johns Hopkins Univ.

[8] Hochstadt, H. (1961). *Special Functions of Mathematical Physics.* Holt, Rinehart and Winston, New York.

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

[10] HOYLAND, A. (1968). Robustness of the Wilcoxon estimate of location against a certain dependence. *Ann. Math. Statist.* **39** 1196-1201.

[11] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.

[12] KIEFER, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38** 1323-1342.

[13] LEHMANN, E. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137-1153.

[14] MOSTELLER, F. (1946). On some useful inefficient statistics. *Ann. Math. Statist.* **17** 377-408.

[15] SEN, P. K. (1968). Asymptotic normality of sample quantiles for *m*-dependent processes. *Ann. Math. Statist.* **39** 1724-1730.

[16] TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling.* (Olkin, I. *et al.*, eds.). Stanford Univ. Press, 448-485.

DEPARTMENT OF STATISTICS
GEORGE WASHINGTON UNIVERSITY
2201 G ST. N.W.
WASHINGTON D.C. 20006

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
W. LAFAYETTE, INDIANA 47907