The behavior of robust estimators on dependent data

by

Joseph L. Gastwirth[1]

The Johns Hopkins University

and

Herman Rubin[2]

Purdue University

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series No. 197

August 1969

Distribution of this document is unlimited.

---

[1] This research was supported by the National Science Foundation Research Grant No. GP-7118 awarded to the Department of Statistics, The Johns Hopkins 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 NR042-216 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

The behavior of robust estimators on dependent data

by

Joseph L. Gastwirth[1]
The Johns Hopkins University

and

Herman Rubin[2]
Purdue University

1. Introduction and Summary

Many robust estimators of the location parameter of a symmetric unimodal distribution have been proposed in the past several years, e.g. [12], [14], [23]. 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 6.1) states that for any Gaussian process for which all the serial correlations $\{\rho_n\}$

[1]This research was supported by the National Science Foundation Research Grant No. GP-7118 awarded to the Department of Statistics, The Johns Hopkins 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 NR042-216 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

are non-negative, 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 <u>finite</u> 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 non-zero limit.

On the F.O.A.D.P., the median is the most efficient estimator studied, although 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 $\rho \to -1$.

In order to study the asymptotic properties of estimators one needs to obtain their asymptotic distribution. Cogburn [4] showed that the sample percentiles are asymptotically normally distributed for data from a general uniform mixing S.S.P. but his regularity conditions seem hard to verify. As all the estimators we consider are functions of the empiric c.d.f. we introduce a notion of mixing which is stronger than Rosenblatt's strong mixing [3], [19] and verify that the empiric c.d.f. does converge to a Gaussian process. Our notion of mixing is weaker than the concept of φ mixing introduced by Doob [6]. In particular, it is satisfied by all the first order autoregressive processes we study in detail. For Gaussian processes, much weaker conditions suffice, and this is discussed in Section 3.

Section 4 is devoted to obtaining a general expression for the asymptotic variance of a general linear combination of the order statistics. Its use is illustrated on Gaussian processes.

The Hodges-Lehmann estimator is discussed in detail in Section 5. A special result is given to insure its asymptotic normality under weaker conditions than are required for the empiric c.d.f. to converge to a Gaussian process. Sections 6 and 7 are devoted to a study of the relative efficiency of our estimators on Gaussian and double-exponential processes. Both first order autoregressive processes are studied exhaustively.

The last section (8) of the paper is concerned with various models of contamination which lead to dependent processes. Recently, Hoyland [13] studied the behavior of the HL estimator in a contamination model which is a special case of the first model we consider. The second model leads to a stationary process which has a contaminated normal (in the sense of Tukey [24]) marginal distribution.

2. The Median

Let $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ denote an ordered sample of n observations $\{X_i\}$ from a strictly stationary stochastic process (S.S.P.) with continuous marginal c.d.f., $F(x)$. The median, M, is defined by

$$M = \begin{cases} x_{(m+1)} & , \text{ if } n = 2m + 1 \\ \frac{1}{2}(x_{(m)} + x_{(m+1)}) & , \text{ if } n = 2m \ . \end{cases} \tag{2.1}$$

In this section we adapt one of the standard methods, [1] and [15], of deriving the asymptotic normality of the median in the case of independent observations to the case of dependent observations. We illustrate some of its properties on Gaussian sequences and also on a first order autoregressive double-exponential process (F.O.A.D.P.). This section serves as an introduction to the methods used in the paper and the type of results we obtain.

If $\{X_i\}$ is an S.S.P. with median ν (which will be assumed to equal 0 in this section), the ordinary sign test statistic S_n can be expressed as

$$S_n = \sum_{i=1}^{n} Y_i \quad , \tag{2.2}$$

where

$$Y_i = \begin{cases} 1 \;, & x_i < \nu \\ 0 \;, & x_i > \nu \;. \end{cases} \tag{2.3}$$

Under suitable mixing conditions, S_n is asymptotically normally distributed and it is the sum of dependent binomial random variables. Indeed if $\{X_i\}$ is an S.S.P. obeying Rosenblatt's [19] or Cogburn's [4] uniform strong mixing conditions or Rozanov's [20] complete regularity condition, S_n will be asymptotically normally distributed. In the appendix to this section we verify that S_n obeys the other conditions of the Blum-Rosenblatt [3] central limit theorem for Gaussian S.S.P.'s where the correlation coefficients satisfy $\Sigma\,|\rho_k| < \infty$.

The asymptotic behavior of the median, M, is deduced from that of S_n as follows. Suppose $M < \nu$, then $|M - \nu|$ is the length Δ of an interval containing the observations between M and ν. The number of observations between M and ν is $S_n - n/2$ which is approximately normally distributed, i.e.,

$$S_n - n/2 \sim \varepsilon \sqrt{\mathrm{Var}(S_n)} \quad , \tag{2.4}$$

where ε is a standard normal r.v. Note that the sign of ε specifies whether $M < \nu$ or is $> \nu$. Since the number of observations in a small interval of length of Δ about ν is approximately

$$nf(\nu)\Delta \ , \tag{2.5}$$

equating (2.4) and (2.5) yields

$$M - \nu = \Delta \sim \frac{1}{n} \frac{\sqrt{V(S_n)}}{f(\nu)} \epsilon \quad . \tag{2.6}$$

This ''heuristic'' argument can be made rigorous as long as the c.d.f. $F(x)$ has a non-zero derivative at ν.

Throughout this paper the asymptotic variance of an estimator will always mean the variance of its asymptotic distribution and the asymptotic efficiency of two estimators will be the reciprocal of the ratio of their asymptotic variances. The asymptotic variance of the median on Gaussian processes is given in

Proposition 2.1. *If* $\{X_i\}$ *is a stationary Gaussian process such that* $\Sigma\,|\rho_k| < \infty$, *then*

$$V(M) \sim \frac{1}{n} \sum_{k=-\infty}^{\infty} \arcsin \rho_{|k|} \quad . \tag{2.7}$$

Proof: Clearly

$$V(S_n) = \sum_{i=1}^{n} V(Y_i) + \sum_{i,j} \mathrm{Cov}(Y_i, Y_j) \quad . \tag{2.8}$$

Recall that if X and Z are jointly normal with zero mean and correlation ρ, then $P(X > 0,\ Z > 0) = 1/4 + (1/2\pi)\arcsin \rho$. As the correlation between two random variables k time units apart is $\rho_{|k|}$, $\mathrm{Cov}(Y_i, Y_{i\pm k}) = (2\pi)^{-1}\arcsin \rho_{|k|}$ and $V(S_n) = (2\pi)^{-1} \sum_{k=-n}^{n} (n - |k|) \arcsin \rho_{|k|}$. As $\Sigma |\rho_k| < \infty$,

$$\lim_{n \to \infty} V(S_n/n) \sim \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \arcsin \rho_{|k|} \quad , \tag{2.9}$$

and (2.7) follows from (2.9) and (2.6) .

We illustrate the proposition by considering a simple moving average process. Let $\{Z_i\}$ be i.i.d. standard normal r.v.'s and let $X_i = (m+1)^{-1/2} [Z_0 + \ldots + Z_m]$, $X_2 = (m+1)^{-1/2} [Z_1 + \ldots + Z_{m+1}]$, etc. As

$$\rho_{-k} = \rho_k = \begin{cases} 1 - |k|/(m+1), & \text{if } |k| < m+1 \;, \\ 0 \qquad , & \text{otherwise} \;, \end{cases} \tag{2.10}$$

(2.7) becomes

$$V(M) \sim \frac{1}{n} \left(2 \sum_{j=1}^{m} \arcsin \left(1 - \frac{j}{m+1}\right) + \frac{\pi}{2}\right) \tag{2.11}$$

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

$$\frac{V(M)}{V(\bar{X})} = \frac{2 \sum_{1}^{m} \arcsin \left(\frac{j}{m+1}\right) + \pi/2}{m+1} \quad . \tag{2.12}$$

When m goes to infinity at a smaller rate than n, formula (2.12) is a Riemann approximation to

$$2 \int_0^1 \arcsin x \ dx = \pi - 2 \sim 1.14 \ . \tag{2.13}$$

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 Polya's theorem, the $\{\rho_k\}$ will be the correlation sequence of a stationary process. Of course, $\rho_0 = 1$. For $k \geq 1$ define

$$\rho_k = \rho \frac{(K+1)^{1+\epsilon}}{(K+k)^{1+\epsilon}} = \rho \left(\frac{K+1}{K+k} \right)^{1+\epsilon}, \ 0 < \rho \leq \frac{1}{2 - \left(\frac{K+1}{K+2}\right)^{1+\epsilon}} \tag{2.14}$$

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

$$\sum_{-\infty}^{\infty} \arcsin \rho_k = \frac{\pi}{2} + 2 \sum_{1}^{\infty} \arcsin \rho \left(\frac{K+1}{K+k}\right)^{1+\epsilon} \sim \frac{\pi}{2} + 2K \int_1^{\infty} \arcsin \rho x^{-(1+\epsilon)} \ dx \tag{2.15}$$

and

$$\sum_{-\infty}^{\infty} \rho_k \sim 1 + 2K \int_1^{\infty} \frac{\rho}{x^{1+\epsilon}} \ dx \ . \tag{2.16}$$

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

$$(2.17) \qquad \frac{\int_1^\infty \arcsin \rho x^{-(1+\epsilon)}\, dx}{\rho \int_1^\infty x^{-(1+\epsilon)}\, dx} .$$

As $\arcsin y \leq y + (\frac{\pi}{2} - 1)y^3$, the ratio (2.17) is

$$(2.18) \qquad \leq \frac{\rho\epsilon^{-1} + (\frac{\pi}{2} - 1)\, \rho^3 (2 + 3\epsilon)^{-1}}{\rho\epsilon^{-1}} = 1 + (\frac{\pi}{2} - 1)\, \rho\epsilon\, (2 + 3\epsilon)^{-1} ,$$

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

Since most stationary processes discussed in the time series literature are Gaussian it is interesting to study a non-Gaussian process. We shall use the first order autoregressive double-exponential process to illustrate the effect the marginal distribution can have on isospectral processes.

The double-exponential process is treated by exploiting the following characterization, due to Gastwirth and Wolff [7], of its characteristic function. Specifically they proved:

Lemma 2.1. If $\varphi(u) = \varphi(u)\,[q + (1-q)\,\varphi(u)]$, $-\infty < u < \infty$, with $0 < q < 1$ and ρ a real number, $0 < |\rho| < 1$, then $\varphi(u)$ is of the form

$$(2.19) \qquad \varphi(u) = (1 + \alpha^2 u^2)^{-1} ,$$

where $\alpha^2 > 0$ is a scale parameter. Furthermore, $q = \rho^2$ and $\alpha^2 = \sigma^2/2$.

This characterization is more meaningful when it is interpreted in terms of an underlying stochastic process with independent errors. The process satisfies the stochastic difference equation

$$X_i = \rho X_{i-1} + [q\cdot 0 + (1-q)\epsilon_1] \quad , \tag{2.20}$$

where 0 stands for the random variable which is degenerate at the origin. In the stationary case, Lemma 2.1 states that X_i and ϵ_i obey the same law if and only if both have a double-exponential distribution with the same parameter. Thus, a first order autoregressive process with double-exponential marginals exists and the error term is a mixture of a degenerate random variable and a double-exponential random variable. If $|\rho| \neq 0$, then with probability one the error r.v. will be zero at three consecutive times. Once this occurs, ρ and the mean can be estimated perfectly from the curve joining the three observed values of the process. Thus, the estimation problem is a non-regular case.

In order to calculate the asymptotic variance of the sign test statistic note that it follows from Lemma 2.1 that $X_{i+k} = \rho^k X_i + \eta_k$, where $\eta_k = \rho^{2k}\cdot 0 + (1 - \rho^{2k})\epsilon$ and ϵ is an independent double-exponential r.v. so that $P[X_i > 0, X_{i+k} > 0] = 1/4 + \rho^k/4$, and (see [8])

$$\lim_{n \to \infty} V(S_n/n) = \frac{1}{4}\frac{1+\rho}{1-\rho} \quad . \tag{2.21}$$

Using (2.4), (2.6) and the asymptotic normality of the sign test [8] we obtain the following

Proposition 2.2. *When $\{X_i\}$ is a first order autoregressive process with double-exponential marginals the median is asymptotically normally distributed with mean 0 and variance*

$$\frac{1}{n}\frac{1+\rho}{1-\rho} \quad . \tag{2.22}$$

Since the asymptotic variance of the mean on first order autoregressive data is $\frac{\sigma^2}{n} \frac{1+\rho}{1-\rho}$, Proposition 2.2 implies that the median is twice as efficient asymptotically as the mean for all values of ρ on double-exponential first order autoregressive data. In Section 6 we shall see that the Gaussian situation is quite different. Indeed as $\rho \to -1$, the mean is infinitely more efficient than the median.

Another interesting consequence of inequality (2.18) is formalized in

Theorem 2.1. *The efficiency of the median to the mean on completely regular stationary Gaussian processes such that $\rho_k \geq 0$ for all k and $\Sigma \rho_k < \infty$ is always greater than or equal to its value, $2/\pi$, in the case of i.i.d. r.v.'s.*

Proof: As $0 \leq \rho_k \leq 1$, substituting the bound (2.18) in formula (2.17) shows that

$$nV(M) \sim \frac{1}{n} \sum_k \arcsin \rho_k \leq \Sigma \rho_k + (\frac{\pi}{2} - 1) \Sigma \rho_k .$$

As $nV(\bar{X}) \sim \Sigma \rho_k$, the reciprocal of the efficiency is $\lim_{n \to \infty} \frac{V(M)}{V(\bar{X})} \leq \frac{\pi}{2}$.

Appendix: A Summary of Mixing Conditions and Their Application to Statistics of the Form $\sum f(X_i)$

A key step in the proofs of the asymptotic normality of the median and the convergence of an empiric c.d.f. of a general S.S.P. to a Gaussian process is that the sign test statistic is asymptotically normally distributed whenever the basic S.S.P. is strongly mixing [3] (or, equivalently completely regular). In order to prove asymptotic normality one can use the concept of maximal correlation between sets of r.v.'s [23], however these conditions are hard to verify in non-Gaussian cases. We therefore develop analogs of Rosenblatt's mixing number and introduce another measure of dependence, between Rosenblatt's and Doob's which is readily computable.

If $\{X_i\}$, i.e. I and $\{X_j\}$ $i \in I$ are two indexed families of r.v.'s the mixing number measuring the dependence between them is

$$(2.1^*) \qquad \alpha(I;J) = \sup_{A,B} \left| P(AB) - P(A)P(B) \right| ,$$

where the range of A is the Borel field generated by the $\{X_i\}$ $i \in I$ and the range of B is the Borel field generated by the $\{X_j\}$ $j \in J$. Rosenblatt's mixing number for stationary processes will be denoted by α_n and is defined as

$$(2.2^*) \qquad \alpha_n = \alpha(I;J)$$

where $I = \{i;\ i \leq 0\}$ and $J = \{j;\ j \geq n\}$. For finite sets of r.v.'s $I = \{i_1,\ldots,i_q\}$ and $J = \{j_1,\ldots,j_r\}$ the mixing number (2.1) will be denoted by

$$(2.3^*) \qquad \alpha(i_1,\ldots,i_q;\ j_1,\ldots,j_r).$$

Our measure of dependence is defined in terms of a function measuring the conditional deviation from independence. Let $Z = \{X_i, i \varepsilon I\}$ have distribution Q and $W = \{X_j, j \varepsilon J\}$ have distribution R, and (Z,W) have distribution P, then

$$(2.4^*) \qquad \Delta(I; J; y) = \int | P(dz|y) - Q(dz) | .$$

All of our conditions involve the L_S norms of the function $\Delta(I; J; y)$. Since

$$(2.5^*) \qquad || \Delta(I; J) ||_s = [\sup \sum_k | \sum_\ell P(F_\ell | E_k) - Q(F_\ell) |^s R(E_k)]^{1/s}$$

where F is a partition in the σ-field generated by the $\{X_i, i \varepsilon I\}$ and $\{E_k\}$ is a partition in the σ-field generated by the $\{X_j, j \varepsilon J\}$,it is not necessary to have a bonafide conditional probability. The analog of (2.3) is

$$(2.6^*) \qquad \Delta(i_1,\ldots,i_q; j_1,\ldots,j_r; y) = \Delta(I; J; y) ,$$

where $I = \{i_1,\ldots,i_y\}$ and $J = \{j_1,\ldots,j_r\}$.

We also define

$$(2.7^*) \qquad \Delta(n; y) = \Delta(I; J; y)$$

where $I = \{i: i \leq 0\}$ and $J = \{j: j \geq n\}$, and

$$(2.8^*) \qquad \Delta_n = \int \Delta(n; y) dR(y) = || \Delta(n) ||_1 .$$

For stationary processes, if $i_1 \leq \ldots \leq i_q < j_1 \leq \ldots \leq j_r$, it is obvious that

$$(2.9^*) \qquad || \Delta(i_1 \ldots, i_q; j_1, \ldots, j_m) ||_s \leq || \Delta(n) ||_s,$$

where $n = j_1 - i_q$.

Remarks: The quantity $|| \Delta(I; J)||_1$ equals the total variation of P - QxR. Moreover,

$$(2.10^*) \qquad \alpha(I; J) \leq 1/4 \; || \Delta(I,J) ||$$

On the other hand conditions using the s-norms of the function (2.3) are weaker than those depending on the concept of ϕ mixing introduced by Doob [6] and developed by Billingsley [2] and Serfling [22]. Indeed, for stationary processes Billingsley's ϕ_n is $1/2 \; || \Delta(n) ||_\infty$.

At this stage it seems appropriate to illustrate the computability of reasonable bounds for the s-norms of $\Delta(k)$ for first order autoregressive processes. For Markov processes $\Delta(k) = \Delta(0; k)$ and $\Delta_k = || \Delta(k) ||$. Using the representation

$$(2.11^*) \qquad X_k = \rho^k X_0 + U_k$$

and denoting the stationary density of X_k by f and mass of the non-absolutely continuous component of the distribution of U_k by M_k and the density of the absolutely component by f_k we have

$$(2.12^*) \qquad \Delta(0; k; y) = M_k + \int | \; f(x) - f_k(x - \rho^k y) \; | \; dx$$

If the density f is not supported on a bounded interval of the real line, then

$$\sup_y \Delta(0;\ k;\ y) = 2. \tag{2.13*}$$

This shows that autoregressive processes are not ϕ mixing. We now discuss three examples.

Example 1: Our first example is the double-exponential process. Here $m_k = \rho^{2k}$ and $f_k(x) = (\frac{1}{2})\ (1-\rho^{2k})\ e^{-|x-\rho^k y|}$ so that

$$\begin{aligned}\Delta(0,K,y) &= \rho^{2k} + \int (1/2)\left|\ e^{-|x|} + (1-\rho^{2k})\ e^{-|x-\rho^k y|}\ \right| dx \\ &\leq \rho^{2k} + 1/2 \int \rho^{2k}\ e^{-|x|} dx + \int (1/2)(1-\rho^{2k})\ \left|\ e^{-|x|} - e^{-|x-\rho^k y|}\ \right| dx \\ &\leq 2\rho^{2k} + \frac{3}{2}\rho^k |y|\ (1-\rho^{2k}).\end{aligned} \tag{2.14*}$$

Moreover, the Minkowski inequality implies that

$$\|\ \Delta(k)\ \|_s \leq 2\rho^{2k} + |\rho^k|\ (\tfrac{3}{2})\ (s!)^{1/s} \tag{2.15*}$$

so that $\Delta_k \leq C|\rho^k|$.

Example 2: For Gaussian Markov processes we must bound

$$\Delta(0;\ k;\ y) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \left|\ \frac{e^{-(x-\rho y)^2/2(1-\rho^2)}}{(1-\rho^2)^{1/2}} - e^{-x^2/2}\ \right| dx\ , \tag{2.16*}$$

which is bounded by

$$(2.17^*)\quad (2\pi)^{-1/2}\int_{-\infty}^{\infty}\left|\frac{e^{-(x-\rho y)^2/2(1-\rho^2)}}{(1-\rho^2)^{1/2}}-\frac{e^{-x^2/2(1-\rho^2)}}{(1-\rho^2)^{1/2}}\right|dx+$$

$$(2\pi)^{-1/2}\int_{-\infty}^{-\infty}\left|\frac{e^{-x^2/2(1-\rho^2)}}{(1-\rho^2)^{1/2}}-e^{-x^2/2}\right|dx$$

The left side of (2.17*) is treated by noting that both functions are probability densities which are equal when $x = \rho y/2$. When $\rho > 0$ and $y > 0$ this integrand is positive when $x > \rho y/2$ (when $y < 0$, the reverse is true but the same bound occurs by symmetry) so that

$$(2.18^*)\quad (2\pi)^{-1/2}\int\left|\frac{e^{-(x-\rho y)^2/2(1-\rho^2)}}{(1-\rho^2)^{1/2}}-\frac{e^{-x^2/2(1-\rho^2)}}{(1-\rho^2)^{1/2}}\right|dx$$

$$=\frac{2}{[(2\pi)(1-\rho^2)]^{1/2}}\int_{x>\frac{\rho y}{2}}[e^{-(x-\rho y)^2/2(1-\rho^2)}-e^{-x^2/2(1-\rho^2)}]\,dx$$

$$=\frac{2}{[(2\pi)(1-\rho^2)]^{1/2}}\int_{-\rho y/2}^{\rho y/2}e^{-t^2/2(1-\rho^2)}\,dt=$$

$$\frac{4}{(1-\rho^2)^{1/2}}\frac{1}{(2\pi)^{1/2}}\int_0^{\rho y/2}e^{-t^2/2(1-\rho^2)}\,dt\le\frac{K|\rho y|}{(1-\rho^2)^{1/2}}$$

The right side of (2.17) is bounded by observing that the integrand is bounded by $C\,x^2\rho^2 e^{-x^2/2}$ when $|\,e^{-x^2\rho^2/2(1-\rho^2)}-(1-\rho^2)^{1/2}\,| < 0$. When $|\,e^{-x^2\rho^2/2(1-\rho^2)}-(1-\rho^2)^{1/2}\,| > 0$, an application of Taylor's theorem yields the bound

$$(2.19^*)\qquad |\,(1-\rho^2)^{-1/2}-1\,|\le B\rho^2/(1-\rho^2)^{1/2}$$

which implies that the integrand is bounded by $B\rho^2 e^{-x^2/2}$. In any case the integrand in the right side of (2.17*)is bounded by

$$(2.20^*) \qquad (Ax^2 + B)\rho^2 e^{-x^2/2}/(1-\rho^2)^{1/2}$$

so that the right side of (2.17*)is bounded by $K\rho^2$ so expression (2.17) is bounded by

$$(2.21^*) \qquad \frac{K\rho^2}{(1-\rho^2)^{1/2}} + \frac{K^1 \, |\rho y|}{(1-\rho^2)^{1/2}} = C_1\rho^2 + C_2|\rho y|$$

where $(1-\rho^2)^{-1/2}$ is absorbed in the constants C_1, C_2 if $|\rho|$ is bounded away from 1. Thus,

$$(2.22^*) \qquad || \; \Delta(k) \; || = C_1\rho^{2k} + \rho^k C_2 \int |y| \frac{e^{-y^2/2}dy}{(2\pi)^{1/2}} = C_1\rho^{2k} + C_2\rho^k.$$

More generally, applying the Minkowski inequality yields

$$(2.23^*) \qquad || \; \Delta(k) \; ||_s \leq C_1\rho^{2k} + C_2|\rho|^k \; [\mu_s]^{1/s},$$

where μ_s is the s^{th} absolute moment of a unit normal r.v.

Notice that (2.23*) is very similar to the result (2.15*)obtained for the double-exponential process and one might be tempted to conjecture that for Markov processes $\Delta_k = \Delta(0,k) \leq C|\rho^k|$. This is not true as can be seen from considering our final example.

Example 3: For the Cauchy process U_n is a Cauchy r.v. with scale parameter $(1-\rho^k)$ so that

$$(2.24^*)\quad \Delta(0;\,k;\,y) = \pi^{-1}\int_{\infty}^{\infty} \left| \frac{1}{1+x^2} - \frac{1}{(1-\gamma)\ (1 + (\frac{x-\gamma y}{1-\gamma})^2} \right| dx\ ,$$

where $\gamma = \rho^k$. Expression (2.24) is less than equal to

$$(2.25^*)\quad \pi^{-1} \int | (1+x^2)^{-1} - (1-\gamma)^{-1}(1 + x^2/(1-\gamma)^2)^{-1} | \, dx \, +$$

$$\pi^{-1} \int | (1-\gamma)^{-1}(1+x^2/(1-\gamma)^2)^{-1} - (1-\gamma)^{-1}(1+(x-\gamma y)^2/(1-\gamma)^2)^{-1}| \, dx$$

We discuss the case where $\gamma>0$. The left side of (2.25) is a difference of two probability densities which are equal when $x = \pm (1-\gamma)^{1/2}$ and the factor $(1+x^2)^{-1}$ is larger when $x > (1-\gamma)^{1/2}$. Hence the left side of (2.25) is

$$(2.26^*)\quad \leq \frac{2}{\pi} \int_{|x|>(1-\gamma)^{1/2}}^{\infty} |(1+x^2)^{-1} - (1-\gamma)^{-1}(1+x^2/(1-\gamma)^2)^{-1}| dx$$

$$\leq \frac{4}{\pi} \int_{x>(1-\gamma)^{1/2}} [(1+x^2)^{-1} - (1-\gamma)^{-1}(1+x^2/(1-\gamma)^2)^{-1}] dx$$

$$= \frac{4}{\pi}[\arctan\,(1-\gamma)^{-1/2} - \arctan\,(1-\gamma)^{1/2}] \leq 2\gamma.$$

The second integral in (2.25*)is handled by substituting $w = x(1-\gamma)^{-1}$, $z = \gamma y(1-\gamma)^{-1}$ yielding

$$(2.27^*)\qquad \pi^{-1} \int \left| \frac{1}{(1+(w-z)^2)} - \frac{1}{1+w^2} \right| dw$$

which is the difference of two Cauchy densities with locations z and 0 respectively and the densities are equal when $w = z/2$. The same type of argument used above shows that expression (2.27*)is

$$(2.28^*) \quad \leq 4\pi^{-1} \arctan \left|\frac{\gamma y}{2(1-\gamma)}\right| \leq \frac{4}{\pi(1-\gamma)} \arctan \ (\gamma|y|) \leq C \arctan . \gamma|y|,$$

if $\gamma < 1/2$. Thus,

$$(2.29^*) \quad \Delta(0;\ k;\ y) \leq 2\gamma + C \arctan \ \gamma|y|$$

and

$$(2.30^*) \quad ||\Delta(k)||_s \leq 2\gamma + C \left[\int | \arctan \ (\gamma y)|^s \ \frac{1}{\pi(1+y^2)} dy \right]^{1/s} .$$

By decomposing the range of integration and using elementary inequalities one can show that

$$(2.31^*) \quad || \Delta(k)||_s \leq 2\gamma + K_s \gamma^{1/2}$$

when $\gamma = \rho^k$ and K_s is constant depending on s. More interesting is the case where $s = 1$.

Here

$$(2.32^*) \int_{-\infty}^{\infty} | \arctan \gamma y|(1+y^2)^{-1}dy \leq 2 \left[\int_0^{\gamma-1} \frac{2\gamma y}{(1+y^2)} dy + \frac{\pi}{2} \int_{\gamma-1}^{\infty} (1+y^2)^{-1} dy \right]$$

where we use the bound $\arctan x \leq x$ for $y \leq (\gamma)^{-1}$. The right side of (2.32*)is

$$(2.33^*) \leq 2[(1/2)\gamma \log(1+(\gamma^2)^{-1}) + \frac{\pi\gamma}{2}] = \gamma + \gamma \log(1 + (\gamma^2)^{-1})$$

so that

$$(2.34^*) \quad || \Delta(k)||_1 \leq C_1\rho^k + C_2\rho^k \log (1 + \rho^{-2k}) .$$

For large x, log (1+x) is essentially log x so that log $(1+\rho^{-2k})$ is nearly $2k \log (1/\rho)$ which implies that

$$\|\Delta(k)\|_1 \leq A\rho^k + Bk\rho^k. \tag{2.35*}$$

Remark: If one refines the argument to get lower bounds on $\Delta(k)$, one can show that the k term in (2.35) cannot be eliminated.

In order to prove the asymptotic normality of the sign test statistic we need the following Lemma due to Ibragimov [14a]:

Lemma 2.1*. *If U and V are bounded by C_1 and C_2 respectively and U is measurable w.r.t. the Borel field generated by X_i for $i \in I$ and V is measurable w.r.t. the Borel field generated by X_j, $j \in J$, then*

$$\mathrm{Cov}(U,V) \leq 4C_1C_2\alpha\ (I;\ J). \tag{2.36*}$$

We now formally state

Theorem 2.1*: *Whenever $\{X_i\}$ is a strongly mixing S.S.P. such that*

$$\sum_{k=1}^{\infty} \alpha(0,k) < \infty \tag{2.37*}$$

$$\sum_{j+k=1}^{n} \sum^{n} \min\ (\alpha(0,j;\ k),\ \alpha(0;\ j,k),\ \alpha(0,k;\ j)) = O(n) \tag{2.38*}$$

and

$$\sum_{i+j+k \leq n} \sum \sum \min(\alpha(0;\ i,i+j,i+j+k),\ \alpha(0,i;\ i+j;\ i+j+k),\ \alpha(0,i,i+j;\ i+j+k) = O(n) \tag{2.39*}$$

then any statistic of the form $S_n = \sum_{i=1}^{n} f(X_i)$, where f is a bounded function is asymptotically normally distributed, i.e.

$$(2.40^*) \qquad n^{-1/2}\,[S_n - E(S_n)] \xrightarrow{\mathcal{L}} N(0,\sigma^2)\ ,$$

where $\sigma^2 = \lim_{n\to\infty} n^{-1}\, V(S_n)$.

Remark: These conditions show that asymptotic normality is determined by the strength of dependence of various subsets of four r.v's which suggests that strong mixing should not be necessary.

Proof: The result will follow once the fourth moment condition of the Blum-Rosenblatt Theorem is verified, i.e., letting $A_i = f(X_i) - E[f(X_i)]$, $E|\Sigma A_i|^4$ must be $O(n^2)$. Expanding $E|\sum A_i|^4 = \Sigma\Sigma\Sigma\Sigma E(A_iA_jA_kA_\ell)$ one obtains

$$(2.41^*) \quad \sum_i EA_i^4 + 4\sum_i\sum_j E(A_i^3A_j) + 6\sum_{i<j}\sum E(A_i^2\,A_j^2) + 12\sum_{i<j}\sum_k\sum E(A_iA_jA_k^2)$$

$$+ 24\sum_{i<j<k<\ell}\sum\sum\sum E(A_iA_jA_kA_\ell).$$

As the A_i are bounded with mean 0, $\Sigma\, E(A_i^4) \le Kn$, $\Sigma\,\Sigma\, E(A_i^3\, A_j) \le Kn^2$ and $\Sigma\,\Sigma\, E(A_i^2\, A_j^2) \le Kn^2$, when K is an appropriate constant. Now

$$(2.42^*) \quad E(A_iA_jA_k^2) = \begin{cases} E(A_iA_j)E(A_k^2) + Cov(A_iA_j;\, A_k)\ , \\ Cov(A_i;\, A_jA_k^2)\ , \\ Cov(A_iA_k^2;\, A_j)\ . \end{cases}$$

Bounding the covariances by Lemma 2.1* yields

$$(2.43^*)\quad |E(A_iA_jA_k^2)| \le \begin{cases} |E(A_iA_j)|E(A_k^2)| + K\,\alpha(i,j;\,k) \\ K\,\alpha(i;\,j,k) \\ K\,\alpha(i,k;\,j). \end{cases}$$

Since $E(A_iA_j) = E(A_i)E(A_j)+\mathrm{Cov}(A_i,A_j)$, $|E(A_iA_j)| \le K\,\alpha(0,\ j-i)$,

$$(2.44^*)\qquad \sum_{i<j}\sum_k |E(A_iA_j)\ |E(A_k^2) \le n^2 K \sum_{k=1}^{\infty} \alpha(0,k).$$

As $E(A_iA_jA_k^2) \le |E(A_iA_j)|E(A_k)^2 + K\min(\alpha(i,j;\,k),\ \alpha(i;\,j,k),\ \alpha(i,k;\,j))$

$$(2.45^*)\quad \sum_{i<j<k}\sum\sum E(A_iA_jA_k^2) \le n^2 K \sum_{k=1}^{\infty} \alpha(0,k)$$

$$+\ nK \sum_j\sum_k \min(\alpha(0,j;\,k),\ \alpha(0;\,j,k),\ \alpha(0,k;\,j))$$

where j-i and k-j are now denoted by j and k. The fourth order terms are handled by noting that if $i < j < k < \ell$

$$(2.46^*)\quad |E(A_iA_jA_kA_\ell)| \le \begin{cases} |E(A_iA_j)|\ |E(A_kA_\ell)| + K\,\alpha(i,j;\,k,\ell)\ , \\ K\,\alpha(i;\,j,k,\ell)\ , \\ K\,\alpha(i,j,k;\,\ell). \end{cases}$$

As

$$(2.47^*)\quad \sum_{i<j<k<\ell}\sum\sum\sum |E(A_iA_j)|\ |E(A_kA_\ell)| \le n^2 K^2 \left(\sum_{r=1}^{\infty} \alpha(0,r)\right)^2 ,$$

$$(2.48^*)\quad \sum_{i<j<k<\ell}\sum\sum\sum |E(A_iA_jA_kA_\ell)| \le n^2 K' \,[\Sigma\, \alpha(0,r)]^2 +$$

$$K \sum_{i<j<k<\ell}\sum\sum\sum \min(\alpha(i,j;\,k,\ell),\ \alpha(i;\,j,k,\ell),\ \alpha(i,j,k;\,\ell)\ .$$

The second term on the right side of (2.48) is

$$\le Kn \sum\sum\sum_{j+k+\ell\le n} \min\,\{\alpha(0,j;\,j+k,\ j+k+\ell),\ \alpha(0;\,j,j+k,j+k+\ell),\ \alpha(0,j,j+k;\,j+k+\ell)$$

when j, k and ℓ now denote $j-i$, $k-j$ and $\ell-k$. Thus, assumptions (2.37*), (2.38*) and (2.39*) imply that (2.45*) and (2.48*) are $O(n^2)$ so that the conditions of the Blum-Rosenblatt theorem hold.

An important Corollary of Theorem 2.1* is

Corollary 2.1*. Let $\{X_i\}$ obey the conditions of Theorem 2.1*, then the finite dimensional marginal distributions of the empiric process $\sqrt{n}\,[F_n(t)-F(t)]$ converges to a multivariate normal distribution.

Proof: For any t, $F_n(t) = n^{-1}\sum_1^n I_i(t)$, where $I_i(t)$ is 1 if $X_i < t$ and 0 otherwise.

For any set $t_1,\ldots,t_k$ of k values of t any linear combination $\sum_{j=1}^k a_jF_n(t_j)$ is a function of the form $n^{-1}\sum_{i=1}^n f(X_i)$ and Theorem 2.1* applies.

Remarks: 1) Corollary 2.1* includes the asymptotic normality of the sign test statistic.

2) The condition of Theorem 2.1* hold whenever $\Sigma\, k^2\, \alpha_k = O(n)$.

3) Since $\alpha_k \le \Delta_k$, Theorem 2.1* is valid whenever $\Sigma\, k^2\, \Delta_k = O(n)$. This condition is easily verified for the three Markov processes discussed in this appendix as $\Sigma k^2\Delta_k$ converges in all three examples.

We conclude this section by showing that for Gaussian processes the sign test statistic is asymptotically normally distributed if $\Sigma|\rho_k| < \infty$. Since bounded functions f have finite variance and the normal distribution is determined by its moments we can approximate f in L^2(w.r.t. the normal distribution) by a polynomial P_ε so that $||f - P_\varepsilon|| = ||f_\varepsilon|| < \varepsilon$. The statistic $n^{-1/2} \sum_{i=1}^{n} P_\varepsilon(X_i)$ is asymptotically normally distributed by Sun's Theorem [23]. The variance of

$$n^{-1/2} \sum_{i=1}^{n} f_\varepsilon(X_i) < \varepsilon \sum_{-n}^{n} |\rho_k|$$

wherewe have bounded $\mathrm{Cov}(f_\varepsilon(X_i),\ f_\varepsilon(X_j))$ by $\varepsilon|\rho_{i-j}|$ by Sarmanov's Lemma [20]. By the Mann-Wald Theorem [16] the result follows.

3. The Convergence of the Empiric c.d.f. to a Gaussian Process

In this section we show that the empiric c.d.f. of a strong mixing S.S.P. converges to a Gaussian process provided that the Δ(or α) functions defined in the Appendix to Section 2 obey some regularity conditions. These conditions are weaker than Doob's concept of φ mixing so that our result is stronger than Billingsley's Theorem 22.1. In particular, the Gaussian, Cauchy and double-exponential first order autoregressive processes are not φ mixing but satisfy our conditions. For Gaussian processes, a special result is derived showing that $\Sigma |\rho_k| < \infty$ suffices to guarantee the convergence of the empiric c.d.f. to a Gaussian process.

The first step in the proof is an application of a lemma of Rubin [21] which states verifiable conditions which imply Prokhorov's necessary and sufficient condition for processes to converge to a limiting process with a.s. continuous sample paths. The next step is to apply Theorem 2.1* to prove that the finite dimensional marginal distributions converge to the appropriate multivariate Gaussian distribution.

Before stating our main result we prove a generalization of Doob's Lemma 7.1 [6] which applies to our Δ functions. Specifically we have

Lemma 3.1. If f and g are functions such that $E(|f(X)|^q) < \infty$, $E(|g(Y)|^r) < \infty$ and if $\frac{1}{r} + \frac{1}{q} + \frac{1}{s} = 1$ and $1 \leq q, r, s \leq \infty$, then

$$(3.1) \qquad |\mathrm{Cov}\ (f(X), g(Y)| \leq 2^{1/q} ||f||_q\ ||g||_r\ ||\Delta^{\frac{1}{r}+\frac{1}{s}}||_s \ ,$$

where Δ is defined in the appendix to section 2.

Proof: Writing

$$(3.2) \qquad \mathrm{Cov}\ [f(X)g(Y)] = \int g(y) \left\{ \int f(x)[dP(x|y) - dQ(x)] \right\} dR(y)$$

and applying the Hölder inequality to $g(y)$ and $h(y) = \int f(x)[dP(x|y) - dQ(x)]$ yields

$$(3.3) \qquad |\mathrm{Cov}\ f(X)g(Y)| \leq \left[\int |g(y)|^r\ dR(y) \right]^{\frac{1}{r}} \cdot \left[\int |h(y)|^t\ dR(y) \right]^{\frac{1}{t}}$$

Applying the Hölder inequality to $h(y)$ and 1 yields

$$(3.4) \qquad |h(y)|^t \leq \left(\int |f(x)|^t\ |dP(x|y) - dQ(x)| \right) \Delta^{\frac{t}{r}}(y) \ .$$

Setting $\frac{1}{t} = \frac{1}{q} + \frac{1}{s}$ applying the Hölder inequality again yields

$$(3.5) \qquad \int |f(x)|^t\ |dP(x|y) - dQ(x)| \leq \left(\int |f(x)|^q\ |dP(x|y) - dQ(x)| \right)^{\frac{t}{q}} \cdot [\Delta(y)]^{\frac{t}{s}}$$

so that the second factor on the right side of (3.3) is bounded by the t^{th} root of

$$(3.6) \qquad \int \left(\int |f(x)|^q\ |dP(x|y) - dQ(x)| \right)^{\frac{t}{q}} \cdot [\Delta(y)]^{t(\frac{1}{s} + \frac{1}{r})} dR(y) \ ,$$

Applying the Hölder inequality once more shows that (3.6) is

$$(3.7)\quad \le \Big[\int \Big(\int |f(x)|^q \, |dP(x|y) - dQ(x)| \Big) dR(y) \Big]^{\frac{t}{q}} \cdot \Big[\int \Delta^{1+\frac{s}{r}}(y) \, dR(y) \Big]^{\frac{t}{s}}$$

and (3.1) follows by taking t^{th} roots and noting that

$$(3.8)\quad \int\int |f(x)|^q \, |dP(x|y) - dQ(x)| \, dR(y) \le \int |f(x)|^q \int [|dP(x|y) + dQ(x)] \, dR(y)$$

$$\le 2 \int |f(x)|^q \, dQ(x) \quad .$$

A further useful generalization is

Lemma 3.2. If f_i has support F_i, g_i has support G_i, where the sets F_i are pairwise disjoint and the sets G_i are pairwise disjoint, then

$$(3.9)\qquad \Sigma \, \mathrm{Cov}(f_i, g_i) \le 2^{1/q} \, ||f||_q \, ||g||_r \, || \Delta^{\frac{1}{r}+\frac{1}{s}} ||_s$$

where $f = \Sigma f_i$, $g = \Sigma g_i$ and g,r,s are as in Lemma 3.1.

Proof: Observe that in deriving Lemma 3.1 only that part of $\Delta(y)$ for y in the support of g is used. Letting X_{G_i} denote the indicator function of the set G_i and $\Delta_i = \Delta X_{G_i}$, Lemma 3.1 implies that

$$(3.10)\qquad \mathrm{Cov}\,(f_i, g_i) \le 2^{1/q} \, ||f_i||_q \, ||g_i||_s \, ||\Delta_i^{\frac{1}{r}+\frac{1}{s}}||_s .$$

The conclusion follows by applying Hölders inequality to the series and using the fact that the supporting sets of each function are disjoint.

The main technical result of the section is given by

Theorem 3.1. *Whenever* $\{X_i\}$ *is a* S.S.P *such that*

$$\sum_{k=1}^{\infty} ||\Delta(0,q)||_1 < \infty , \tag{3.11}$$

$$\sum_{j+k\leq n} \min(||\Delta(0,j;\ j+k)||_1,\ ||\Delta(0;\ j,j+k)||_1) = O\ (n(\log n)^{-5}) , \tag{3.12}$$

$$\sum_{i+j+k\leq n} \min(||\Delta(0;\ i,i+j,i+j+k)||_1,\ ||\Delta(0,i;\ i+j,i+j+k)||_1,\ ||\Delta(0,i,i+j;\ i+j+k)||_1) = O(n(\log n)^{-5}) \tag{3.13}$$

and either

$$\sum_{k=1}^{\infty} ||\Delta(0,k)||_s < \infty \text{ for some } s > 1 \tag{3.14a}$$

or

$$\alpha(0,k) = O(k^{-1}\ (\log k)^{-6-\varepsilon}) , \tag{3.14b}$$

then the empiric process $n^{1/2}\ [F_n(t)-F(t)]$ *obeys the conditions of Prokhorov's continuity theorem, i.e., as* $n \to \infty$ *it converges to a process with* a.s. *continuous paths if the finite dimensional marginals converge.*

Before proceeding to the proof of the theorem we recall some useful results due to Rubin [21].

Lemma 3.3 (Rubin). *Let* X_n *be a separable process defined on* [0,1] *such that*

$$X_n(t+u) - X_n(t) > - \psi_n(u) \text{ for } u > 0 , \tag{3.15}$$

where ψ_n is increasing on (0,1), and for some $\lambda > 0$

$$(3.16) \qquad \sum_{j=1}^{2^i} E \mid X_n(j/2^i) - X_n(\tfrac{j-1}{2^i}) \mid^{\lambda} \le \gamma_{in} < \infty .$$

For any $\varepsilon > 0$, let $R_n(\varepsilon)$ be the smallest integer such that

$$(3.17) \qquad \psi_n(2^{-R_n(\varepsilon)}) < \varepsilon.$$

The Prokhorov continuity condition is satisfied if for every $\varepsilon > 0$ and $\eta > 0$

$$(3.18) \qquad \overline{\lim_n} \sum_{i=\ell}^{R_n(\varepsilon)} \gamma_{in}^{1/\lambda+1} < u.$$

A frequently useful corollary is

Lemma 3.4. If X_n is a separable process satisfying (3.15) and

$$(3.19) \qquad E \mid X_n(t+u) - X_n(t) \mid^{\lambda} < \phi_n(u).$$

then Prokhorov's condition is satisfied if for every $\varepsilon > 0$ and $\eta > 0$ there is an ℓ such that

$$(3.20) \qquad \overline{\lim_n} \sum_{i=\ell}^{R_n(\varepsilon)} (2^i \phi_n(2^{-i}))^{1/\lambda+1} < \eta ,$$

where R is as in Lemma 3.3.

Remark: If γ_{in} in Lemma 3.3 or ϕ_n in Lemma 3.4 can be written as a sum of a fixed finite number of functions satisfying (3.18) or (3.20) respectively, the conclusion follows.

Proof of Theorem 3.1: Letting $Y_i = F(X_i)$ one can transform the empiric process to the unit interval and we shall assume that this has been done. We shall verify that the conditions of Lemma 3.3 are satisfied where $\lambda = 4$, $\psi_n(u) = n^{1/2} u$ and $R_n(\varepsilon)$ is the smallest integer greater than $\log (1/\varepsilon) + 1/2 \log n$, where logarithms are taken to the base 2. Let $B_i(t,u) = 1$ if $Y_i \varepsilon [t,t+u)$ and 0 otherwise and let $A_i(t,u) = B_i(t,u) - u$. We shall omit the arguments u and t where no confusion will arise. The computation of the bounds on the fourth moments required to verify condition (3.18) is similar to the derivation of Theorem 2.1*. First we note that if $V(t,u) = \prod_{k=1}^{K} A_{j_k}(t,u)$, $W(t,u) = \prod_{g=1}^{G} A_{i_g}(t,u)$, then $|V| \leq 1$, $E|V|^q \leq 2u$ for all $q \geq 1$, $E|W|^r < 2u$ for $r \geq 1$, and Lemma 3.1 implies that

$$(3.21) \quad |\mathrm{Cov}(V,W)| \leq \min \left(2^{1/q} (2u)^{\frac{1}{q} + \frac{1}{r}} || \Delta^{1-1/q}||_s, Cu\right) ,$$

where C is a constant.

Note that any product of A_i's is a constant plus a linear combination of indicator functions (of sets whose probabilities are $\leq u$), and if the intervals $[t_i, t_i + u)$ are disjoint

$$(3.22) \qquad \Sigma \, |\mathrm{Cov}(V(t_i, t_i + u), W(t_i, t_i + u))| \leq M\, 2^{1/q} || \Delta^{1-1/q}||_s .$$

Clearly (3.22) is minimized when $q = \infty$ and $s = 1$. (For the duration of this proof $||\cdot||$ without a subscript will denote $||\cdot||_1$.)

We now bound the terms in the expansion of $n^{-2} \sum_{j=1}^{2^i} E\{\sum_{k=1}^{n} A_k (\frac{j-1}{2^i}, \frac{j}{2^i})\}^4$.

Arguing as in section 2 but using more powerful bounds we obtain

$$(3.23) \quad n^{-2} \sum_{j=1}^{2^i} \sum_{k=1}^{n} E(A_k^4) \leq 2/n ,$$

$$(3.24)\quad n^{-2}\sum_{j=1}^{2^i}\sum_{\substack{k=1\\k\neq\ell}}^{n}\sum_{\ell=1}^{n} E(A_k^3\, A_\ell) = n^{-2}\sum_{j=1}^{2^i}\sum_{\substack{k=1\\k\neq\ell}}^{n}\sum_{\ell=1}^{n} \mathrm{Cov}\,(A_k^3,\, A_\ell)$$

$$\le n^{-2}\sum_{k=1}^{n}\sum_{\ell=1}^{n} M\,||\Delta(k,\ell)|| \le 2\,Mn^{-1}\sum_{k=1}^{n}||\Delta(0,k)||,$$

and

$$(3.25)\quad n^{-2}\sum_{j=1}^{2^i}\sum_{\substack{k=1\\k\neq\ell}}^{n}\sum_{\ell=1}^{n} E(A_k^2\, A_\ell^2) = n^{-2}\sum_{j=1}^{2^i}\sum_{k\neq\ell}\sum E(A_k^2)\, E(A_\ell^2) + n^{-2}\sum_j\sum_{k\neq\ell}\sum \mathrm{Cov}(A_k^2\, ,A_\ell^2)$$

$$\le 4.2^{-i} + 2Mn^{-1}\sum_{k=1}^{n}||\Delta(0,k)||.$$

Next

$$(3.26)\quad n^{-2}\Big|\sum_{j=1}^{2^i}\sum_{k\neq\ell\neq m}\sum\sum E(A_k^2\, A_\ell A_m)\Big| \le n^{-2}\sum\sum\sum\sum E(A_k^2)|\; E(A_\ell A_m)|$$

$$+\, n^{-2}\sum\sum\sum\sum \min\,(|\mathrm{Cov}(A_k^2, A_\ell A_m)|, |\mathrm{Cov}(A_k^2 A_\ell, A_m)|,\; |\mathrm{Cov}(A_k^2 A_m, A_\ell)|)$$

The first term is bounded by $4.2^{-i}\sum_\ell ||\Delta(0,\ell)||$ and the second is bounded by

$$(3.27)\qquad Mn^{-2}\sum\sum\sum \min(||\Delta(k;\ \ell,m)||,\ \Delta(k,\ell;\ m)||,\ ||\Delta(k,m;\ \ell)||)$$

$$\le 6\, Mn^{-1}\sum_{k<\ell}\sum \min(||\Delta(0;\ k,\ell)||,\ ||\Delta(0,k;\ell)||).$$

The most complicated term is

$$(3.28) \quad 24n^{-2}\left|\sum_j \sum_{k<\ell<m<h}\sum\sum\sum E(A_kA_\ell A_mA_h)\right| \leq 24n^{-2}\sum |E(A_kA_\ell)\ E(A_mA_h)|$$

$$+\ 24n^{-2}\sum\ \sum\sum\sum\sum \min\left(|Cov(A_k,A_\ell,A_mA_h)|,\ |Cov(A_k,A_\ell A_mA_h)|,\ |Cov(A_kA_\ell A_m,A_h)|\right).$$

Treating the second term in a manner similar to that used for the second term of (3.26) shows that it is

$$(3.29) \quad \leq 24Mn^{-1}\sum_{i+j+k\leq n} \min(||\Delta(0;i,i+j,i+j+k)||,||\Delta(0,i;i+j,i+j+k)||,\ ||\Delta(0,i,i+j;\ i+j+k)||)$$

For the first term in (3.28) it is necessary to use an individual bound for one factor and Lemma 3.2 for the other. Thus, the first term is

$$(3.30) \quad \leq 48n^{-2}(2^{-i})^{1-1/s}\sum\sum\sum\sum\sum ||\Delta(k;\ell)||_s\ |E(A_m,A_h)|$$

$$\leq 48Mn^{-2}(2^{-i})^{1-1/s}\sum\sum\sum\sum ||\Delta(k;\ \ell)||_s\ ||\Delta(m;\ h)||_1$$

$$\leq K(2^{-i})^{1-1/s}\ \sum_k ||\Delta(0,k)||_s \sum_m ||\Delta(0,m)||_1.$$

If we use Ibragimov's lemma for the individual factor, we obtain the bound

$$(3.31) \qquad M\sum_{\ell=1}^{n} \min(2^{-i},\ \alpha(0,\ell))\sum_{k\leq n} ||\Delta(0,k)||.$$

Putting terms together we see that

$$(3.32) \qquad \sum_{j=1}^{2^i} E\,|X(j/2^i) - X(j-1)/2^i|^4 \le \sum_{r=1}^{8} \zeta_{rin} ,$$

where the ζ's are the various bounds derived above. To check that (3.18) is satisfied note that there are $O(\log n)$ terms and each summand ζ_{rin} in the bound for γ_{in} is either $o((\log n)^{-5})$ uniformly in i or the series $\sum_{i=1}^{\infty} (\zeta_{rin})^{1/5}$ converges uniformly in n.

Remark: **If the original $\{X_i\}$ are strongly mixing, the conditions of Theorem 3.1** are stronger than the conditions of Theorem 2.1*, so that the finite dimensional marginals of the empiric process converge to a multivariate normal distribution and the process converges to a Gaussian process.

A useful Corollary is

Corollary 3.1: *Whenever $\{X_i\}$ is a strongly mixing s.s.p. such that $\Delta_k = o(k^{-2} (\log k)^{-5})$, $n^{1/2} [F_n(t)-F(t)]$ converges to a Gaussian process with a.s. continuous paths.*

Proof: Since $\alpha(0,k)$ and $||\Delta(0,k)|| \le \Delta_k$, (3.11) and (3.14b) are satisfied. The left side of (3.12) is less than $2 \sum_{k=1}^{\infty} k\Delta_k$ and the left side of (3.13) is less than $3\, \Sigma\, k^2\Delta_k$ so the result will follow once $\Sigma k^2\Delta_k$ is shown to be $o(n(\log n)^{-5})$. As Δ_k is $o(k^{-2}(\log k)^{-5})$ and as the logarithm function is slowly varying we are done.

Remark: By the monotonicity of Δ_i no better result of this type can be obtained.

Remark: Theorem 3.1 applies to the three Markov processes discussed in the appendix to section 2.

Remark: In the future we shall call any process $\{X_i\}$ obeying the conditions of Theorem 3.1 a strongly mixing Δ_s process.

An alternate Theorem using only the mixing numbers is

Theorem 3.2: *Whenever X_i is a strongly mixing s.s.p. such that $a_k = O(k^{-5/2})$, then the empiric process $n^{1/2}[F_n(t) - F(t)]$ converges to a Gaussian process.*

Proof: The finite dimensional marginals converge to Gaussian marginals by Theorem 2.1*. The verification of Prokhorov's conditions proceeds as before except that Lemma 3.4 is used and the covariances are bounded by K min (u,α_k). In the expansion of $n^{-2}\,E(\Sigma A_i)^4$ the worst terms, as in Theorem 3.1 are those in the expansion of

$$\sum_{k<\ell<m<h}\sum\sum\sum E(A_k\, A_\ell\, A_m\, A_h)\ ,$$

where the bound is

$$(3.33)\qquad M\,(\sum_k \min(u,\ \alpha_k))^2 + Mn^{-1}\,\Sigma\, k^2 \min\,(u,\ \alpha_k).$$

It can be shown that in the second term the u does not improve the bound appreciably, therefore to satisfy (3.20) we require that

$$(3.34)\qquad \sum_{i=\ell}^{R_n(\varepsilon)} (2^i\, Mn^{-1} \sum_k k^2\, \alpha_k)^{1/5} = (2^{R_n(\varepsilon)}\, n^{-1}\, \Sigma\, k^2\, \alpha_k)^{1/5}\, \frac{(1-2^{\frac{1}{5}(\ell-1-R_n(\varepsilon))}}{1-2^{-1/5}}$$

is $<n$. The second factor approaches a limit and $2^{R_n(\varepsilon)} < 2n^{1/2}/\varepsilon$. This is equivalent to the first factor being o(1) or $\Sigma\, k^2\, \alpha_k = o(n^{1/2})$. As α_k is a monotonically decreasing sequence this reduces to $\alpha_k = o(k^{-5/2})$.

Now we examine the first term in (3.33) which depends on

$$(3.35)\qquad \sum_{k=1}^{n} \min\,(2^{-i},\ \alpha_k) \leq \sum_{k=1}^{2^{.4i}} 2^{-i} + \sum_{k>2^{.4i}} \alpha_k = 2^{-.6i} + o(2^{-.6i})$$

which is $O\,(2^{-.6i})$, where we have used the assumption that $\alpha_k = o(k^{-5/2})$. Hence

$$(3.36)\qquad \sum_{i=\ell}^{R_n(\varepsilon)} [2^i(\sum_{k=1}^{n} \min(2^{-i},\ \alpha_k)^2]^{1/5} \leq \sum_{i=\ell}^{\infty} 2^{-.04i}$$

which can be made less than η if ℓ is chosen sufficiently large.

As before a much better result can be obtained when the original r.v.'s $\{X_i\}$ are a Gaussian process. We require

Lemma 3.5. For every $M > 0$, $\delta > 0$, $m > 0$ there exists a number $c > 0$ and polynomials S, R, R independent of M, such that whenever $X_1, X_2, Y_1, \ldots, Y_m$ are jointly normal with means 0, variances 1, $E(X_1X_2) = \rho$, $E(X_iX_j) = \tau_{ij}$ and the covariance matrix of the (X,Y) vector

<u>has determinant exceeding</u> δ , $0 \le a < b \le a+M$, $B_i = I_{(a,b)}(X_i)$, $\alpha = E(B_i)$, <u>then for all</u> y , $a \le y_i \le b$ <u>for</u> $i = 1, \ldots, m$,

$$(3.37)\qquad |E((B_1-\alpha)(B_2-\alpha)|Y=y) - E((B_1-\alpha)(B_2-\alpha))|$$

$$\le \alpha^c S(b)\Big(|\rho| \sum_{i=1}^{2} \sum_{j=1}^{m} |\tau_{ij}| + \prod_{i=1}^{2} \sum_{j=1}^{m} |\tau_{ij}|\Big)$$

$$\text{if} \quad \min_i \sum_j |\tau_{ij}| < \frac{\delta}{2} \quad \text{or} \quad m = 1 ,$$

$$(3.38)\qquad |E((B_i-\alpha)(B_j-\alpha)|Y=y) - E((B_i-\alpha)(B_j-\alpha))|$$

$$\le M^2 R(b)\Big(|\rho| \sum_{i=1}^{2} \sum_{j=1}^{m} |\tau_{ij}| + \prod_{i=1}^{2} \sum_{j=1}^{m} |\tau_{ij}|\Big) \quad \text{otherwise} .$$

<u>Proof</u>: The conditional distribution of X_1 and X_2 given $Y_1, \ldots, Y_m$ is normal with means $\tau_i' Q^{-1} Y$ and covariance matrix

$$(3.39)\qquad \Sigma(\rho,\tau_1,\tau_2) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} \tau_1' \\ \tau_2' \end{pmatrix} Q^{-1}(\tau_1 \ \tau_2) ,$$

where Q is the covariance matrix of the Y's . By the uniform nonsingularity, all elements of Q^{-1} are bounded by $1/\delta$ and all elements of $\Sigma^{-1}(\rho,\tau_1,\tau_2)$ as well as the determinant of that matrix are bounded by $1/\delta$.

Let

$$(3.40)\qquad P(\rho,\tau_1,\tau_2,y) =$$

$$\int_a^b \int_a^b \frac{1}{2\pi\, |\Sigma(\rho,\tau_1,\tau_2)|^{1/2}} \exp - \tfrac{1}{2}(x-\tau' Q^{-1} y)\Sigma^{-1}(\rho,\tau_1,\tau_2)(x-\tau' Q^{-1} y)\,dx_1 dx_2 .$$

Then expanding,

$$(3.41)\quad E((B_1-\alpha)(B_2-\alpha)|Y=y) - E((B_1-\alpha)(B_2-\alpha))$$

$$= P(\rho,\tau_1,\tau_2,y) - P(0,0,\tau_2,y) - P(0,\tau_1,0,y) - P(\rho,0,0,y) + 2P(0,0,0,y) \quad ,$$

since $\alpha = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$.

Now the right-hand expression in (3.41) is

$$(3.42)\quad \int_0^1 [P_\rho(t\rho,t\tau_1,t\tau_2,y) - P_\rho(t\rho,0,0,y)]\,\rho$$
$$+ [P_{\tau_1}(t\rho,t\tau_1,t\tau_2,y) - P_{\tau_1}(0,t\tau_1,0,y)]\,\tau_1$$
$$+ [P_{\tau_2}(t\rho,t\tau_1,t\tau_2,y) - P_{\tau_2}(0,0,t\tau_2,y)]\,\tau_2\,dt$$

$$= \int_0^1\int_0^t [P_{\rho\tau_1}(t\rho,s\tau_1,s\tau_2,y) + P_{\rho\tau_1}(s\rho,t\tau_1,s\tau_2,y)]\,\rho\tau_1$$
$$+ [P_{\rho\tau_2}(t\rho,s\tau_1,s\tau_2,y) + P_{\rho\tau_2}(s\rho,s\tau_1,t\tau_2,y)]\,\rho\tau_2$$
$$+ [P_{\tau_1\tau_2}(s\rho,t\tau_1,s\tau_2,y) + P_{\tau_1\tau_2}(s\rho,s\tau_1,t\tau_2,y)]\,\tau_1\tau_2\,ds\,dt \quad .$$

Since τ_1 and τ_2 are vectors, by $P_{\tau_1}\tau_1$ we mean $\sum_i P_{\tau_{1i}}\tau_{1i}$, etc.

Hence the expression is bounded by

$$(3.43)\quad |\rho|\left(\Sigma\,|\tau_{1i}|\,\sup\left|\frac{\partial^2 P(r,t_1,t_2)}{\partial r\,\partial t_{1i}}\right| + \Sigma\,|\tau_{2i}|\,\sup\left|\frac{\partial^2 P(r,t_1,t_2)}{\partial r\,\partial t_{2i}}\right|\right)$$
$$+ \Sigma\,\Sigma\,|\tau_{1i}|\,|\tau_{2j}|\,\sup\left|\frac{\partial^2 P(r,t_1,t_2)}{\partial t_{1i}\,\partial t_{2j}}\right| \quad ,$$

where the sup is over the smallest convex set generated by the covariance matrices. An elementary algebraic argument shows that all matrices involved in the argument are uniformly non-singular, so that we can differentiate under the integral sign and obtain the result that all second derivatives of the integrand of (3.40) are bounded by a polynomial $R(b)$, which proves (3.38).

If $\min_i \sum_j |\tau_{ij}| < 1/2\delta$ or $m = 1$, the sum of the regression coefficients will not be near 1 for both variables, and hence for all x, y with $a \le x_i \le b$, $a \le y_j \le b$

$$(3.44) \qquad (x - \tau' Q^{-1} y)' \Sigma(\rho, \tau_1, \tau_2)^{-1}(x - \tau' Q^{-1} y) \ge rb^2 - s .$$

We may assume $r \le 2$.

Now

$$(3.45) \qquad \int_a^b \int_a^b e^{-\frac{1}{2} rb^2} dx_1 dx_2 = \left(\int_a^b e^{-\frac{1}{4} rb^2} dx \right)^2$$

$$\le \left(\int_a^b e^{-\frac{1}{4} rx^2} dx \right)^2 \le \left(\int_a^b e^{-\frac{1}{2} x^2} dx \right)^{\frac{r}{2}} (b - a)^{2 - \frac{r}{2}}$$

from which (3.37) follows with $c = r/2$, $S = 2\pi M^{2-r/2} e^s R$.

Now let us verify the Prokhorov continuity condition for stationary Gaussian process with $\Sigma |\rho_i| < \infty$. Instead of using the usual probability integral transform, let us transform to density $6t(1-t)$, i.e., $P(Y_i \le t) = 3t^2 - 2t^3$. We now have $\sqrt{n}(\{F_{nY}(t+u) - F_Y(t+u)\} - \{F_{nY}(t) - F_Y(t)\}) > -6\sqrt{n}\, u$, which is not essentially different from the case of the probability integral transform.

The proof of Rubin's Lemma involves showing that

$$(3.46)\qquad \sum_{i=k}^{R_n} \sum_{j=0}^{2^i} P\left(\left|X_n\left(\frac{j+1}{2^i}\right) - X_n\left(\frac{j}{2^i}\right)\right| > \zeta_i\right) \to 0 \quad ,$$

where $\Sigma \zeta_i < \epsilon$. We shall bound all except the extreme terms for each i by a 4^{th} moment Markov inequality, and the extreme ones by the Tchebychev inequality. We can even use a fixed number instead of ζ_i for the extreme terms. Now

$$(3.47)\qquad E\,[(X_n(2^{-i}) - X_n(0))^2] = \frac{1}{n} \Sigma\Sigma\, E(A_i A_j) \quad ,$$

where $A_i = I_{(-\infty,\xi)}(X_i) - \alpha$, $\alpha = \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 3\cdot 2^{-2i} - 2^{1-3i}$.

Now $E(A_i^2) = \alpha(1-\alpha)$ and by Sarmonov's lemma [20] , $|E(A_i A_j)| \le \alpha|\rho_{i-j}|$. Thus

$$(3.48)\qquad E\,[(X_n(2^{-i}) - X_n(0))^2] \le 3(1 + 2\sum_{i=1}^{\infty} |\rho_i|)2^{-2i} \quad ,$$

and hence $P(|X_n(2^{-i}) - X_n(0)| > \epsilon) \le K\,\epsilon^{-2}2^{-2i}$, which is more than adequate for our purposes.

For the remaining terms, we proceed as before to verify Rubin's lemma. We will illustrate the use of Lemma 3.5 only on the fourth-order terms; the others are even easier.

Let $B_i = I_{t,t+u}(Y_i)$, $\alpha = E(B_i)$, $A_i = B_i - \alpha$, and $t \ge \frac{1}{2}$. The terms for $t < \frac{1}{2}$ give an equal contribution. Let $i < j < k < \ell$ and consider

$$(3.49)\qquad \begin{aligned} E(A_iA_jA_kA_\ell) - E(A_iA_j)E(A_kA_\ell) = {} & E([E(A_iA_j|X_kX_\ell) - E(A_iA_j)]B_kB_\ell) \\ & - \alpha\, E([E(A_iA_j|X_k) - E(A_iA_j)]B_k) \\ & - \alpha\, E([E(A_iA_j|X_\ell) - E(A_iA_j)]B_\ell) \quad . \end{aligned}$$

If $|\rho_{ik}| + |\rho_{i\ell}| < \frac{\delta}{2}$ or $|\rho_{jk}| + |\rho_{j\ell}| < \frac{\delta}{2}$, we can apply (3.37) to all terms on the right side of (3.49), obtaining

$$(3.50) \qquad |E(A_iA_jA_kA_\ell) - E(A_iA_j)E(A_kA_\ell)| \le 3\,\alpha^{1+c}Q(b)(|\rho_{ij}|(|\rho_{ik}|+|\rho_{i\ell}|+|\rho_{jk}|+|\rho_{j\ell}|) + (|\rho_{ik}|+|\rho_{i\ell}|)(|\rho_{jk}|+|\rho_{j\ell}|)) \; .$$

If $|\rho_{ik}| + |\rho_{i\ell}| \ge \frac{\delta}{2}$ and $|\rho_{jk}| + |\rho_{j\ell}| \ge \frac{\delta}{2}$, we just use $6M^2\alpha\, Q(b)(|\rho_{ik}|+|\rho_{i\ell}|)(|\rho_{jk}|+|\rho_{j\ell}|)$ as a bound. Furthermore, $\alpha < 6(1-t)u$. Also, since the normal tail drops off rapidly, $(1-t)^{1+c}Q(b)$ and $(1-t)R(b)$ are uniformly bounded for all terms. Hence, except for a multiple of n terms, if $i < j < k < \ell$

$$(3.51) \qquad \begin{aligned} |E(A_iA_jA_kA_\ell)| &\le |E(A_iA_j)|\;|E(A_kA_\ell)| \\ &\quad + Ku^{1+c}(|\rho_{ij}|(|\rho_{ik}|+|\rho_{i\ell}|+|\rho_{jk}|+|\rho_{j\ell}|) \\ &\quad + (|\rho_{ik}|+|\rho_{i\ell}|)(|\rho_{jk}|+|\rho_{j\ell}|)) \\ &\le u^2|\rho_{ij}|\;|\rho_{k\ell}| + Ku^{1+c}(|\rho_{ij}|(|\rho_{ik}|+|\rho_{jk}|+|\rho_{i\ell}|+|\rho_{j\ell}|) \\ &\quad + (|\rho_{ik}|+|\rho_{i\ell}|)(|\rho_{jk}|+|\rho_{j\ell}|)) \; . \end{aligned}$$

The sum of these terms is bounded by

$$n^2M(u^2+u^{1+c}) \qquad \text{since} \qquad \Sigma|\rho_{i-j}| < \infty \; .$$

The remaining terms are bounded by $K\alpha$, or $6KM^2(1-t)R(b)u$, so that their sum is bounded by Cnu. We obtain similar results for the other terms. Consequently, we can apply Rubin's lemma with $\varphi_n(u) = M(u^2+u^{1+c}) + D\,u/n$.

4. The Asymptotic Distribution of a General Linear Estimator

In Section 3 we proved that the empiric c.d.f. 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.

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 [9]

$$W = n^{-1} \sum_{i=1}^{n} w_i X_{(i)} \quad , \tag{4.1}$$

where

$$w_i = n\nu \left[\frac{i-1}{n}, \frac{i}{n}\right] \quad , \quad i = 1, \ldots, n \tag{4.2}$$

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 λ^{th} fractile. If W_1 and W_2 denote the α^{th} and β^{th} 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

<u>Theorem 4.1.</u> <u>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</u>

$$n^{-1} \sum_{q=-\infty}^{\infty} \frac{P\,[X_0 < x,\, X_q < y] - \alpha\beta}{f(x)\, f(y)} \quad . \tag{4.3}$$

<u>Proof</u>: For each observation X_i define the indicator r.v.'s

$$(4.4)\qquad Y_i(\alpha) = \begin{cases} 1\,, & X_i < x \\ 0\,, & \text{otherwise}\;; \end{cases} \qquad Y_i(\beta) = \begin{cases} 1\,, & X_i < y \\ 0\,, & \text{otherwise}\;. \end{cases}$$

As in the derivation of the asymptotic distribution of the median

$$(4.5)\qquad W_1 - x \sim -[nf(x)]^{-1}S_1\;, \qquad W_2 - y \sim -[nf(y)]^{-1}S_2\;,$$

where

$$(4.6)\qquad S_1 = \sum_{i=1}^{n} [Y_i(\alpha) - \alpha] \quad \text{and} \quad S_2 = \sum_{i=1}^{n} [Y_i(\beta) - \beta]\;.$$

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

$$\begin{aligned}(4.7)\quad \mathrm{Cov}\,(S_1,S_2) &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}\,[Y_i(\alpha),Y_j(\beta)] \\ &= n^{-1}\sum_{q=-(n-1)}^{(n-1)} (n-|q|)\;\mathrm{Cov}\,[Y_o(\alpha),Y_q(\beta)] \\ &\to \sum_{q=-\infty}^{+\infty} \mathrm{Cov}\,[Y_o(\alpha),Y_q(\beta)] \quad \text{as} \quad n\to\infty\;.\end{aligned}$$

As $\mathrm{Cov}\,[Y_o(\alpha),Y_q(\beta)] = P\,[X_o < x,\; X_q < y] - \alpha\beta$, substituting (4.7) into (4.5) yields (4.3) .

In particular we have

<u>Corollary 4.1</u>. <u>If</u> $Z_{(i)}$, $i = 1, \ldots, k$ <u>are the</u> λ_i^{th} <u>sample percentiles and if</u> $\sum_1^k w_i = 1$, <u>then</u> $W = \sum_{i=1}^{k} w_i Z_{(i)}$ <u>is asymptotically normally distributed with mean</u> $\sum_{i=1}^{k} w_i F^{-1}(\lambda_i)$, <u>and variance</u>

$$(4.8) \qquad \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{k} w_i w_j \sum_{q=-\infty}^{\infty} \frac{P[X_o < F^{-1}(\lambda_i),\ X_q < F^{-1}(\lambda_j)] - \lambda_i \lambda_j}{f[F^{-1}(\lambda_i)]\ f[F^{-1}(\lambda_j)]}$$

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

Remark: Actually Theorem 4.1 and its Corollary are valid for S.S.P.'s satisfying the conditions of Theorem 2.1*.

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

$$(4.9) \qquad V(\sqrt{n}\, W) \sim \sum_{q=-\infty}^{\infty} \int_0^1 \int_0^1 \frac{P[X_o < F^{-1}(\alpha),\ X_q < F^{-1}(\beta)] - \alpha\beta}{f[F^{-1}(\alpha)]\ f[F^{-1}(\beta)]}\, d\nu(\alpha)\, d\nu(\beta) \quad .$$

The interchange of summation and integration is valid in the case of a finite number of sample percentiles but requires justification in general. At this point it may be instructive to note that the term in (4.9) for $q = 0$ is the variance in the case of independent observations so that (4.9) can be regarded as the variance in the independent case with correction terms for each "q^{th} 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

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

In terms of μ, (4.9) becomes

$$(4.10) \qquad V(\sqrt{n}\, W) \sim \sum_{q=-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{P[X_o < x,\ X_q < y] - P[X_o < x]P[X_q < y]\}\, d\mu(x)\, d\mu(y) \quad .$$

In order to use formula (4.10) one needs a reasonable expression for $P[X_o < x, X_q < y] - P[X_o < x]P[X_q < y]$. Fortunately, for normal r.v.'s we have

Lemma 4.1. *If* X *and* Y *are two correlated standard normal r.v.'s with correlation coefficient* η , *then*

$$(4.11) \qquad P[X < a, Y < b] - P[X < a]P[Y < b] = \frac{1}{2\pi} e^{-(a^2+b^2)/2} \sum_{k=1}^{\infty} H_{k-1}(a)H_{k-1}(b) \frac{\eta^k}{k!} ,$$

where $H_k(a)$ *is the* k^{th} *Hermite polynomial.*

Proof: The bivariate normal density function can be expressed in terms of the Hermite polynomials as follows [11]:

$$(4.12) \quad (2\pi)^{-1}(1-\eta^2)^{-1/2} e^{-(x^2+y^2-2\eta xy)/2(1-\eta^2)} = (2\pi)^{-1} e^{-(x^2+y^2)/2} \sum_{k=0}^{\infty} \frac{H_k(x)H_k(y)\eta^k}{k!}$$

$$= \psi(x,y,\eta) \quad .$$

The probability desired is

$$(4.13) \quad \int_{-\infty}^{a} \int_{-\infty}^{b} [\psi(x,y,\eta)-\varphi(x)\varphi(y)]dx\, dy = \int_{-\infty}^{a} \int_{-\infty}^{b} (2\pi)^{-1} e^{-(x^2+y^2)/2} \sum_{k=1}^{\infty} \frac{H_k(x)H_k(y)\eta^k}{k!} dx\, dy,$$

where $\varphi(x)$ denotes the standard normal density. Integrating (4.13), after interchanging the summation and integration operations, yields the right side of (4.11).

Remark: Formula (4.11) 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

$$(4.14) \qquad \mu(x) = \begin{cases} \int_0^x d\mu(t) \quad , & x > 0 \; , \\ \int_{-x}^0 d\mu(t) \quad , & x < 0 \; , \end{cases}$$

and its ''total variation function''

$$(4.15) \qquad \mu^*(x) = \begin{cases} \int_0^x |d\mu|(t) \quad , & x > 0 \; , \\ \int_{-x}^0 |d\mu|(t) \quad , & x < 0 \; , \end{cases}$$

we shall assume that $\mu^*(x)$ is in L^2 w.r.t. the normal density function $\varphi(x) = (2\pi)^{-1/2} e^{-x^2/2}$. Substituting (4.11) into (4.12) yields

<u>Proposition 4.1.</u> <u>The asymptotic variance of a general linear estimator on Gaussian processes such that</u> $\sum_k |\rho_k| < \infty$ <u>is given by</u>

$$(4.16) \qquad V(\sqrt{n}\, W) \sim \sum_{q=-\infty}^{\infty} \sum_{k=1}^{\infty} \frac{c_k}{k!} (\rho_{|q|})^k = \sum_{k=1}^{\infty} \frac{c_k}{k!} \sum_{q=-\infty}^{\infty} (\rho_{|q|})^k \quad ,$$

<u>where</u> $$c_k = \left[\int_{-\infty}^{\infty} H_{k-1}(x)\, \varphi(x)\, d\mu(x) \right]^2 .$$

<u>Remark 1</u>: The assumption that $\mu(x)$ is in L^2 guarantees that $\sum_{k=1}^{\infty} \frac{c_k}{k!}$ converges.

<u>Remark 2</u>: If $d\mu(x)$ is a symmetric measure, i.e., the original measure ν gives equal weight to the i^{th} and $(n+1-i)^{st}$ order statistics, then its odd Fourier-Hermite coefficients vanish, so that only terms involving c_{2k+1} appear in (4.16).

<u>Remark 3</u>: At times it may be convenient to separate the $q = 0$ term in (4.16) as it is the asymptotic variance of $\sqrt{n}\, W$ when the observations are independent, i.e.,

$$(4.17) \qquad V(\sqrt{n}\, W) \sim \sum_{k=1}^{\infty} \frac{c_k}{k!} + 2 \sum_{k=1}^{\infty} \frac{c_k}{k!} \sum_{q=1}^{\infty} (\rho_q^k) \quad .$$

This implies that positive serial correlation ($\rho_q > 0$ for all q) increases the variance of a general linear estimator.

Remark 4: As the exact conditions for the validity of expression (4.16) for the asymptotic variance are not known in complete generality we note that whenever the regularity conditions of Theorem 4.1 and $\Sigma\, |\rho_k| < \infty$, 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 examples.

Example 1. As a check, consider the mean, $\bar{X}$. Here $d\mu(x) = 1$. Since $H_o(x) = 1$, $c_1 = 1$ while $c_k = 0$ if $k \neq 1$. Thus, $V(\sqrt{n}\, \bar{X}) = \Sigma\, \rho_{|k|}$ as is well known.

Example 2. The median M is represented by a measure $d\mu(x)$ which places an atom of mass $(2\pi)^{1/2}$ 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 (4.16), we have

$$(4.18) \qquad \sum_{k=0}^{\infty} \frac{(2k-1)!!}{(2k+1)(2k)!!}\, \rho_{|q|}^{2k+1} = \arcsin\, \rho_{|q|}$$

and summing over q yields (2.7).

Example 3. Consider a <u>finite</u> combination of sample percentiles, i.e., let ν give weight w_i to $x_{(i)}$, where $i = [n\lambda_i]$, $0 < \lambda_1 < \ldots < \lambda_r < 1$ 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, $W(\alpha)$, where $\lambda_1 = \alpha$, $\lambda_2 = 1-\alpha$, $a_2 = -a_1$ and $w_1 = w_2 = 1/2$. 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

$$(4.19) \qquad V(\sqrt{n}\, W(\alpha) \sim \sum_{k=0}^{\infty} \frac{H_{2k}^2(a)}{(2k+1)!} \left(\sum_{q=-\infty}^{\infty} \rho_{|q|}^{2k+1} \right) .$$

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

$$(4.20) \qquad T = (1 - 2\alpha)^{-1} \int_{A_n}^{B_n} x dF_n(x) ,$$

where A_n is the sample α^{th} quantile, B_n is the sample $(1-\alpha)^{th}$ quantile and $F_n(x)$ is the empiric c.d.f. 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

$$(4.21) \qquad V(\sqrt{n}\, T(\alpha)) \sim T_o + 2 \sum_{q=1}^{\infty} (\rho_q) + \frac{4e^{-a^2}}{(1-2\alpha)^2 \pi} \sum_{j=1}^{\infty} \frac{H_{2j-1}^2(a)}{(2j+1)!} \cdot \sum_{q=1}^{\infty} (\rho_q)^k ,$$

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

It should be mentioned that one can derive the asymptotic normality of the trimmed mean by showing that (4.20) is representable by

$$(4.22) \qquad T(\alpha) = -(1 - 2\alpha)^{-1} \int_{-a}^{a} G_n(x)dx + o_p(n^{-1/2}) ,$$

where $G_n(x) = F_n(x) - F(x)$, provided that the underlying density is continuous at $\pm a$. This gives an explicit representation of $T(\alpha)$ in terms of the empiric c.d.f. which is asymptotically a Gaussian process.

Since we shall study, in detail, the behavior of the linear estimators on F.O.A.G.P.'s for reference we specialize the general results in

Proposition 4.2. On F.O.A.G.P.'s ,

$$(4.23) \qquad V(\sqrt{n}\, W(\alpha)) \sim \pi\alpha e^{a^2} + 2 \sum_{j=0}^{\infty} \frac{\rho^{2j+1}}{1-\rho^{2j+1}} \cdot \frac{H_{2j}^2(a)}{(2j+1)!}$$

$$(4.24) \qquad V(\sqrt{n}\, T(\alpha)) \sim T_o + \frac{2\rho}{1-\rho} + \frac{4e^{-a^2}}{(1-2\alpha)^2 \pi} \sum_{j=1}^{\infty} \frac{H_{2j-1}^2(a)}{(2j+1)!} \cdot \frac{\rho^{2j+1}}{1-\rho^{2j+1}} \quad .$$

Finally, we have

Proposition 4.3. On F.O.A.G.P.'s, the asymptotic variance of any unbiased estimator approaches ∞ as $\rho \to +1$ and the asymptotic variance of any symmetric estimator approaches 0 as $\rho \to -1$.

Proof: As $c_k \geq 0 \;\; \forall\, k$ the first result is an immediate consequence of formula (3.16) . For symmetric estimators,

$$(4.25) \qquad V(\sqrt{n}\, W) \sim \sum_{k=0}^{\infty} \frac{c_{2k+1}}{(2k+1)!} \cdot \frac{1+\rho^{2k+1}}{1-\rho^{2k+1}} \quad .$$

As $\rho \to -1$ each term $(1+\rho^{2k+1})/(1-\rho^{2k+1})$ approaches 0 and the result follows as $\Sigma\, c_k/k! < \infty$.

Remark: From (4.25) it is clear that the variance of any symmetric linear estimator is an increasing function of ρ . The variance of a non-symmetric estimator remains an increasing function of ρ for $0 \leq \rho < 1$.

5. The Hodges-Lehmann Estimator

One robust estimator which has received much attention recently is the Hodges-Lehmann [12] estimator which is derived from the Wilcoxon test. If $X_1, \ldots, X_n$ 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.,

(5.1) $$HL = \text{med}\ \{ \frac{X_i+X_j}{2} \}_{i,j=1}^{n} \quad .$$

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 $\Sigma |\rho_k| < \infty$. 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 c.d.f. $F_n(t)$ by

(5.2) $$M^* = 2\ HL = \text{median}\ \{F_n(t)*F_n(t)\} \quad ,$$

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., $F_n*F_n(M^*) = 1/2$, then $HL^* = M^*/2$. 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 $n^2F_n*F_n(0)$ 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^* just as in Section 2.

Theorem 3.2 can be applied to show that HL^* is asymptotically normal, however, the HL^* is asymptotically normal under weaker conditions. In terms of the conditions on the empiric process

(5.3) $$G_n(t) = n^{1/2}[F_n(t) - F(t)]$$

in a neighborhood of 0 , we have

<u>Theorem 5.1</u>. <u>The estimator,</u> HL^* , <u>is asymptotically normally distributed whenever the following conditions are satisfied:</u>

(5.4a) $(G_n * F)(0)$ <u>is asymptotically normally distributed</u>

(5.4b) <u>there exist two sequences of reals</u> w_n <u>and</u> λ_n <u>such that</u>

$$w_n \to 0 \,, \quad \lambda_n \to \infty \ \underline{\text{but}}\ \lambda_n n^{-1/2} \to 0 \,,$$

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

<u>and</u>

(5.4c) $$\sup_{|x| < \lambda_n n^{-1/2}} P\{|n^{-1/2} G_n * G_n(x)| > w_n\} \to 0 \text{ as } n \to \infty$$

(5.4d) $(F*F)'(0)$ <u>exists.</u>

<u>Moreover,</u>

(5.5) $$n^{1/2} H^* + G_n * F(0) \,/\, (F*F)'(0) \ \xrightarrow{P}\ 0$$

and

(5.6) $$n^{1/2} H^* \sim N(0, \sigma^2)$$

<u>where</u> σ^2 <u>is the variance of</u> $(G_n*F)(0) \,/\, (F*F)'(0)$.

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

(5.7) $$F_n*F_n = F*F + 2n^{-1/2}\, G_n*F + n^{-1}\, G_n*G_n \ .$$

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

<u>Proof</u>: 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

(5.8) $$|(F*F)(x) - (F*F)(0) - \alpha x| \le \rho_n |x|$$

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

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

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

and $Q_n\rho_n \to 0$. The existence of a sequence Q_n satisfying conditions (5.9) and (5.10) 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 $(-Q_n h_n n^{-1/2}\ ,\ Q_n h_n n^{-1/2})$ is the desired neighborhood of zero which is larger than $n^{-1/2}$.) At all points of the form

$$x = \frac{k\,h_n}{n^{1/2}}\ , \quad -Q_n \le k \le Q_n\ , \tag{5.11}$$

relations (5.9) and (5.10) imply that

$$|F_n*F_n(x) - F*F(x) - 2n^{-1/2}\,G_n*F(0)| \le 3w_n n^{-1/2} \tag{5.12}$$

except with small probability. Substituting (5.8) in (5.12), shows that with large probability (w.l.p.)

$$|F_n*F_n(x) - F*F(0) - \alpha x - 2n^{-1/2}(G_n*F)(0)| < \rho_n|x| + 3w_n n^{-1/2}\ . \tag{5.13}$$

In particular, it follows from (5.13) that w.l.p. if $F_n*F_n(x) \le 1/2$, then

$$\alpha x + 2n^{-1/2}(G_n*F)(0) \le \rho_n|x| + 3w_n n^{-1/2} \tag{5.14}$$

while if $F_n*F_n(x) \ge 1/2$, then

$$\alpha x + 2n^{-1/2}(G_n*F)(0) \ge -\rho_n|x| - 3w_n n^{-1/2}\ . \tag{5.15}$$

Thus, w.l.p.

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

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.

$$(5.17) \qquad n^{1/2}[(F_n*F_n)(-Q_n h_n n^{-1/2}) - 1/2] < 0 ,$$

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

$$(5.18) \qquad n^{-1/2} kh_n \leq M^* \leq n^{-1/2}(k+1)h_n \quad , \quad -Q_n \leq k \leq Q_n \quad .$$

Let

$$(5.19) \qquad Y = n^{1/2}(\alpha M^* + 2n^{-1/2}(G_n*F)(0)) \quad .$$

Since $(F_n*F_n)((k+1)h_n n^{-1/2}) \geq 1/2$, it follows from (5.11) that w.l.p.

$$(5.20) \qquad Y \geq -\rho_n Q_n h_n - 3w_n - h_n \quad .$$

As $(F_n*F_n)(kh_n n^{-1/2}) \leq 1/2$, (5.14) implies that w.l.p.

$$(5.21) \qquad Y \leq \rho_n Q_n h_n + 3w_n + h_n \quad .$$

As h_n, w_n and $\rho_n Q_n h_n$ all approach 0 as $n \to \infty$, (5.20) and (5.21) imply that $(Y) \to 0$ in probability. Hence

$$(5.22) \qquad n^{1/2} M^* + \frac{2(G_n*F)(0)}{\alpha} \xrightarrow{P} 0$$

or

$$(5.23) \qquad n^{1/2} HL^* + \frac{(G_n*F)(0)}{\alpha} \xrightarrow{P} 0$$

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

In the appendix to this section we verify that the assumptions of Theorem 5.1 are satisfied by strongly mixing Gaussian processes such that $\Sigma |\rho_k| < \infty$. A similar (but simpler) argument will show that the first order autoregressive double-exponential process also obeys Theorem 5.1 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. The following elementary and well known lemma will be of use:

Lemma 5.1. *If* $F'(x) = f(x)$ *is symmetric about* 0, *then*

$$(F{*}F)'(0) = \int_{-\infty}^{\infty} f^2(t)\,dt \quad . \tag{5.24}$$

In order to calculate the variance of $F{*}G_n(0)$, it is often convenient to express $F{*}F_n(x)$ as follows:

$$F{*}F_n(x) = \int F(x-t)\,dF_n(t) = n^{-1}\sum_{i=1}^{n} F(x-X_i) \quad , \tag{5.25}$$

so that

$$F{*}G_n(0) = n^{-1/2}\sum_{i=1}^{n} \{F(-X_i) - E\,[F(-X_i)]\} \quad . \tag{5.26}$$

In particular, when f is symmetric about 0,

$$F{*}G_n(0) = n^{-1/2}\sum_{i=1}^{n} \{F(X_i) - 1/2\} \tag{5.27}$$

and

$$\mathrm{Var}\,[F{*}G_n(0)] = \frac{1}{n}\sum_i\sum_j E\,[F(X_i) - \tfrac{1}{2}][F(X_j) - \tfrac{1}{2}] \quad . \tag{5.28}$$

In the case of i.i.d. symmetric r.v.'s, we have the following

<u>Lemma 5.2</u>. <u>When the r.v.'s</u> $X_1, X_2, \ldots$ <u>are i.i.d. symmetric (about</u> 0), <u>the variance of</u> $F*G_n(0) = \frac{1}{2}$.

<u>Proof</u>: Clearly,

$$\text{(5.29)} \qquad n \operatorname{Var}[F*G_n(0)] = E\,[F(X) - \tfrac{1}{2}]^2 = \int F^2\,dF - \int F\,dF + \tfrac{1}{4} = \tfrac{1}{12}\ .$$

The known fact that the asymptotic variance of the Hodges-Lehmann estimator is given by $[12\,(\int f^2(t)\,dt^2]^{-1}$ in the case of i.i.d. symmetric r.v.'s is an immediate consequence of formulas (5.23), (5.24) and (5.29).

A non-trivial use of the representation (5.28) occurs in the derivation of the asymptotic variance of HL on double-exponential first order autoregressive data. Specifically, we have

<u>Theorem 5.2</u>. <u>When</u> $\{X_i\}$ <u>is a F.O.A.D.P. the asymptotic variance of the Hodges-Lehmann estimator is given by</u>

$$\text{(5.30)} \qquad n\,V[HL] \sim \frac{4}{3} + \frac{3\rho}{1-\rho} - 3\Big(\sum_{j=1}^{\infty} \frac{\rho^{3j}}{(2+\rho^j)^2}\Big) \qquad \text{if } \rho > 0$$

<u>and</u>

$$\text{(5.31)} \qquad n\,V[HL] \sim \frac{3}{2} \sum_{j=-\infty}^{\infty} \Big(\rho^{|j|} - \frac{\rho^{3|j|}}{(2+|\rho|^{|j|})^2}\Big) \qquad \text{if } \rho < 0\ .$$

<u>Proof</u>: From (5.23), $n\,V[HL] = 16\,V[F*G_n(0)]$ on double-exponential data. Since the c.d.f. of the double-exponential distribution is

$$\text{(5.32)} \qquad F(u) = \frac{1}{2} + \frac{1}{2}\,(1 - e^{-|u|})\ \operatorname{sgn} u$$

(5.28) becomes

$$(5.33)\qquad n\,V[F^*G_n(0)] = (\tfrac{1}{4})\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} E\,[(1-e^{-|Y_j|})(1-e^{-|Y_k|})\,\mathrm{sgn}\,(Y_jY_k)] \;.$$

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

$$(5.34)\qquad V[F^*G_n(0)] \sim (\tfrac{1}{4})\sum_{j=-\infty}^{\infty} E\,[(1-e^{-|Y_0|})(1-e^{-|Y_j|})\,\mathrm{sgn}\,(Y_0Y_j)] \;.$$

To compute the expectation in (5.34) note that one can assume that $Y_0 > 0$. Suppose $\rho > 0$, then $\lambda = \rho^j$ and using the characterization Lemma (2.1), one obtains, for $j \neq 0$,

$$(5.35)\qquad E\,[(1-e^{-|Y_0|})(1-e^{-|Y_j|})\,\mathrm{sgn}\,(Y_0Y_j)] = \int_0^{\infty}(1-e^{-x})e^{-x}\{\lambda^2(1-e^{-\lambda x})$$

$$+ (1-\lambda^2)[\int_0^{\infty}\frac{e^{-y}}{2}(1-e^{-\lambda x-y})dy + \int_{-\lambda x}^{0}\frac{e^{y}}{2}(1-e^{-\lambda x-y})\,dy$$

$$- \int_{-\infty}^{-\lambda x}\frac{e^{y}}{2}(1-e^{y+\lambda x})\,dy]\}\,dx \quad ,$$

which equals

$$\int_0^{\infty} e^{-x}(1-e^{-x})\{\lambda^2(1-e^{-\lambda x}) + (\tfrac{1}{2})(1-\lambda^2)(2(1-e^{-\lambda x}) - \lambda x e^{-\lambda x})\}\,dx$$

$$= \frac{3}{8}\lambda - \frac{3}{8}\frac{\lambda^3}{(2+\lambda)^2} \quad .$$

When $j = 0$, the expectation is $\frac{1}{3}$ so that for $\rho > 0$

$$(5.36)\qquad 4\,V[F^*G_n(0)] = \frac{1}{3} + \frac{3}{4}\sum_{j=1}^{\infty}(\rho^j - \frac{\rho^{3j}}{(2+\rho^j)^2})$$

and

$$(5.37) \qquad n\,V[HL] \sim \frac{4}{3} + \frac{3\rho}{1-\rho} - 3\left(\sum_{j=1}^{\infty} \frac{\rho^{3j}}{(2+\rho^{j})^{2}}\right) \quad .$$

The derivation of (5.31) when $\rho < 0$ is similar and is omitted.

For Gaussian processes it is more convenient to make a direct computation in terms of the r.v.'s which indicate whether $(X_i + X_j)$ is less than 0 (the median) and convert this to obtain the asymptotic variance of M^* just as the sign test yielded the asymptotic variance of the median. Of course $HL^* = M^*/2$. Let

$$(5.38) \qquad I_{ij} = \begin{cases} +1\ , & \text{if } (X_i + X_j) > 0\ , \\ -1\ , & \text{if } (X_i + X_j) < 0\ , \end{cases}$$

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

The random variables X_i, X_k, X_j, X_ℓ are jointly Gaussian with covariance matrix

$$(5.39) \qquad \begin{pmatrix} 1 & \rho_{|k-i|} & \rho_{|j-i|} & \rho_{|\ell-i|} \\ \rho_{|k-i|} & 1 & \rho_{|k-j|} & \rho_{|k-\ell|} \\ \rho_{|j-i|} & \rho_{|k-j|} & 1 & \rho_{|j-\ell|} \\ \rho_{|\ell-i|} & \rho_{|k-\ell|} & \rho_{|j-\ell|} & 1 \end{pmatrix}$$

Thus, the covariance matrix of the two random variables $X_i + X_j$ and $X_k + X_\ell$ is

$$(5.40) \qquad \begin{pmatrix} 2(1 + \rho_{|j-i|}) & \gamma \\ \gamma & 2(1 + \rho_{|k-\ell|}) \end{pmatrix} ,$$

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

$$(5.41)\qquad \rho^* = \frac{\gamma}{2(1+\rho_{|j-i|})^{1/2}\,(1+\rho_{|k-\ell|})^{1/2}} \ .$$

Since

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

$$(5.42)\qquad \mathrm{Var}(S) = \sum_{\text{all pairs}} \mathrm{Cov}(I_{ij}, I_{k\ell}) = \frac{2}{\pi} \sum_{i,j,k,\ell} \arcsin \rho^* .$$

Note that

$$|\mathrm{Cov}\,(I_{ij}, I_{k\ell})| \le |\rho_{|k-i|}| + |\rho_{|j-\ell|}| + |\rho_{|k-\ell|}| + |\rho_{|j-k|}|$$

so that

$$(5.43)\qquad n^{-3} \Sigma\,\Sigma\,\Sigma\,\Sigma\ \mathrm{Cov}(I_{ij}, I_{k\ell}) = \lim_{H\to\infty} \sum_{|h|\le H} n^{-3} \Sigma\,\Sigma\,\Sigma\ \mathrm{Cov}(I_{ij}, I_{k\ell}) \ ,$$

uniformly in n, where the last summation is over those indices where the first of $i-k$, $j-\ell$, $i-\ell$, $j-k$ which is smallest in absolute value is h. Consequently

$$(5.44)\qquad \lim_{n\to\infty} n^{-3} \Sigma\,\Sigma\,\Sigma\,\Sigma\ \mathrm{Cov}(I_{ij}, I_{k\ell}) = \sum_{h=-\infty}^{\infty} \lim_{n\to\infty} n^{-3} \Sigma\,\Sigma\,\Sigma\ \mathrm{Cov}(I_{ij}, I_{k\ell}) \ ,$$

where the last summation is as before. Let us now investigate one of the latter limits. Assume for example, that $h = i-k$. We break the sum up for fixed i and k into those terms for which one of the remaining differences is at most T in magnitude (where $T > h$) and into those which all the k remaining differences exceed T. The first sum has at most $(6T+1)n$ terms and so gives a negligible contribution. Each of the remaining terms is close to $(2/\pi)$ arcsin $(\rho_{|h|}/2)$. Since there are $(n-|h|)$ pairs i,k with $h = i-k$ and there are four choices for the pair with the smallest spacing, the h^{th} term in the right side of (5.44) approaches $8\pi^{-1}$ arcsin $(\rho_{|h|}/2)$. Thus

$$n^{-3}\,\mathrm{Var}(S) \sim \frac{8}{\pi} \sum_h \arcsin \frac{\rho_{|h|}}{2}\,. \tag{5.45}$$

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

$$n^{-3}\,\mathrm{Var}(S^*) \sim \frac{2}{\pi} \Sigma_h \arcsin \frac{\rho_{|h|}}{2}\,. \tag{5.46}$$

The sam 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 $1/2\sqrt{\pi}$ in the normal case by Lemma 5.1 yields

$$(5.47) \qquad \mathrm{Var}(M^*) = n^{-1}\, 8 \sum_h \arcsin \frac{\rho|h|}{2} .$$

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

$$(5.48) \qquad \mathrm{Var}\,(\sqrt{n}\,HL) \sim 2 \sum_{h=-\infty}^{\infty} \arcsin \frac{\rho|h|}{2} .$$

In Section 2, we noted that for Gaussian processes with non-negative serial correlation the efficiency of the median to the mean was always greater than or equal to its efficiency $(\frac{2}{\pi})$ in the case of independent observations. An analogous result is true for the Hodges-Lehmann estimator and is obtained from the following elementary

Lemma 5.4. *For* $0 \leq y \leq 1/2$, $\arcsin y \leq (\pi/3)y$.

The result is an immediate consequence of the fact that

Proof: $(\arcsin y)/y$ is an increasing function on $[0, 1/2]$ so that it is always $\leq 2 \arcsin (1/2) = \pi/3$.

We now formally present

Theorem 5.3. *The efficiency of the Hodges-Lehmann estimator,* HL, *to the sample mean* $\bar{X}$, *on strongly mixing Gaussian* S.S.P.'s *such that* $\rho_k \geq 0$ *for all* k *and* $\Sigma\, \rho_k < \infty$, *is always greater than or equal to its value,* $3/\pi$, *in the case of* i.i.d. *Gaussian observations.*

Proof: Lemma 5.4 implies that

$$(5.49) \qquad 2 \arcsin (\rho_k/2) \leq (\pi/3)\, y .$$

Using formula (5.48) and the fact that $V(\sqrt{n}\,\bar{X}) \sim \Sigma\, \rho_k$ it follows that the reciprocal of the efficiency is

$$(5.50) \qquad \lim_{n \to \infty} \frac{V(HL)}{V(\bar{X})} \leq \frac{\Pi}{3} .$$

Appendix: Verification of the Conditions of the Theorem for Strongly Mixing Gaussian Processes Such That $\Sigma |\rho_k| < \infty$.

In this appendix we shall show, in detail, that conditions (5.4a) and (5.4b) are satisfied. The argument showing that condition (5.4c) is satisfied is a tedious calculation which is similar to the proof in the appendix of Section 2 and we omit it. Condition (5.4d) is obviously satisfied.

Verification of Condition (a):

Letting $\{X_k\}$ denote the r.v.'s of the process we have derived the representation

$$(5.1^*) \qquad F*G_n(0) = n^{-1/2} \sum_{k=1}^{n} \{F(-X_k) - E[F(-X_k)]\} .$$

As the $\{X_k\}$ are strongly mixing, the r.v.'s $F(-X_k)$ are also strongly mixing as they are functions of X_k. The asymptotic normality of the right side of (5.1^*) follows from the Blum-Rosenblatt central limit theorem.

Verification of Condition (b):

Let

$$(5.2^*) \qquad \beta(x) = \sup_t |F(x+t) - F(t)| .$$

When $F(t)$ is the normal c.d.f., $\beta(x) \leq |x| \sqrt{\pi}$.

Now

$$(5.3^*)\quad \mathrm{Var}[F*G_n(x) - F*G_n(0)] = n^{-1} \sum_i \sum_j \mathrm{Cov}[F(x-X_i) - F(-X_i)][F(x-X_j) - F(X_j)]$$

$$= n^{-1} \sum_i \sum_j \int\int [F(x-x_i)-F(-x_i)][F(x-x_j)-F(-x_j)][dP_{ij}(x_i,x_j)-dP_i(x_i)dP_j(x_j)] ,$$

where P_{ij} denotes the joint c.d.f. of X_i and X_j while P_i and P_j denote the respective marginal c.d.f.'s. Applying (5.2^*) to the integrand shows that in a small neighborhood of 0 ,

$$(5.4^*)\ \mathrm{Var}(F*G_n(x)-F*G_n(0)) \le n^{-1}\beta^2(x) \sum_i \sum_j \iint |dP_{ij}(x_i,x_j)-dP_i(x_i)dP_j(x_j)| \quad .$$

Since $\iint |dP_{ij}(x_i,x_j) - dP_i(x_i)dP_j(x_j)|$ 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 (5.4^*) is

$$(5.5^*)\qquad \le n^{-1}\ \beta^2(x) \sum_i \sum_j K|\rho_{i-j}| \le \beta^2(x) \sum_{r=-(n-1)}^{(n-1)} \frac{n-|r|}{n}\ K|\rho_r| \le K'\beta^2(x)$$

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

$$(5.6^*)\qquad P\ \{|G_n*F(x) - G_n*F(0)|\ > \frac{K'\beta^2(x)}{w_n^2} \quad .$$

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

Remark: In order to verify condition (5.4b) for data from an arbitrary continuous distribution one must find a sequence w_n approaching 0 at a slower rate than $\beta(\lambda_n n^{-1/2})$. 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 $\Sigma\ \Delta_k < \infty$, then

$P(|\frac{1}{\sqrt{n}}\ G_n{*}G_n(x)| > \epsilon) \to 0$ uniformly in x . In the Gaussian case, we can instead use the condition $\Sigma\ |\rho_k| < \infty$.

6. The Efficiency of the Estimators Relative to the Mean in Gaussian Processes

In this section we study the efficiency of our estimators relative to $\bar{X}$ on 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 robust against positive dependence (all $\rho_k \geq 0$), just as the median and the Hodges-Lehmann estimators are. 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 (6.1) 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 25^{th} and 75^{th} 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 6.2. For the estimator $W(\alpha)$ it turns out that the optimum choice of α in the case of <u>independent</u> 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 efficiency of the median or any *finite linear combination* of sample percentiles approaches 0 while that of HL or $T(\alpha)$ approaches a finite limit. This is in sharp contrast with the case of independent observations where the efficiency of M to $\bar{X}$ is always $\geq 1/3$ provided that the density sampled is symmetric and unimodal.

In order to discuss the efficiency of linear estimators we isolate the following result which was the basis of the proofs of Theorems 2.1 and 5.3. Specifically, we have

Lemma 6.1. If S *is any estimator such that*

$$V(\sqrt{n}\,S) \sim \sum_{q=-\infty}^{\infty} g(\rho_q) \,, \tag{6.1}$$

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

$$\lim_{n \to \infty} \frac{V(\sqrt{n}\,S)}{V(\sqrt{n}\,\bar{X})} = \frac{\sum_{q=-\infty}^{\infty} g(\rho_q)}{\sum_{-\infty}^{\infty} \rho_q} \leq g(1) \,. \tag{6.2}$$

In particular, Lemma 6.1 is applicable whenever $g(\rho)/\rho$ is an increasing function of ρ. Typically $g(1)$ is the asymptotic variance of $\sqrt{n}\,S$ in the case of independent observations. We next apply Lemma 6.1 to derive

Theorem 6.1. The efficiency of any unbiased linear combination of the order statistics relative to $\bar{X}$ *on strong mixing Gaussian process such that* $\rho_k \geq 0$ *for all* k *and* $\Sigma\,|\rho_k| < \infty$ *is always greater than or equal to its value when the observations are independent.*

Proof: The variance of $n^{1/2}W$, for any linear combination W, is given by (4.16). Setting

$$(6.3) \qquad g(\rho) = \sum_{k=1}^{\infty} \frac{c_k(\rho)^k}{k!} ,$$

and recalling that $c_k \geq 0$, the result follows from Lemma 6.1.

Remark: This efficiency-robustness result depends heavily on the assumption that the process is Gaussian. Our analysis relies heavily on the orthogonal expansion of the bivariate normal density function (Lemma 3.1). 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.

For the remainder of the section we shall assume that the $\{X_i\}$ are a F.O.A.G.P. Then the reciprocal of the efficiency of any unbiased linear estimator W is given by

$$(6.4) \qquad \lim_{n \to \infty} \frac{V(\sqrt{n}\, W)}{V(\sqrt{n}\, \bar{X})} = \sum_{k=1}^{\infty} \frac{c_k}{k!} \cdot \frac{1+\rho^k}{1-\rho^k} \cdot \frac{1-\rho}{1+\rho}$$

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

Lemma 6.2. For all $\ell > 0$ $\frac{1+\rho^\ell}{1-\rho^\ell} \cdot \frac{1-\rho}{1+\rho}$ decreases as ρ goes from 0 to 1 .

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

Applying Lemma 6.2 to expression (6.4) yields

Corollary 6.1. 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 $c_k \geq 0$ and each term in (6.4) decreases as ρ increases. The second part of the Corollary follows as $c_{2k} = 0$ for symmetric estimators.

Corollary 6.1 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 6.2. *The reciprocal of the efficiency of any symmetric linear estimator relative to* $\bar{X}$ *as* $\rho \to +1$ *or* -1 *is given by*

$$\text{(6.5)} \qquad \lim_{\rho \to 1} \frac{V(\sqrt{n}\, W)}{V(\sqrt{n}\, \bar{X})} = 1 + \sum_{j=1}^{\infty} \frac{c_{2j+1}}{(2j+1)(2j+1)!} ,$$

and

$$\text{(6.6)} \qquad \lim_{\rho \to -1} \frac{V(\sqrt{n}\, W)}{V(\sqrt{n}\, \bar{X})} = \sum_{j=0}^{\infty} \frac{c_{2j+1}}{(2j)!} .$$

Expression (6.6) may be infinite.

Expressions (6.5) and (6.6) 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 6.3. *If* $\rho > 0$, *then* $V(\bar{X}) < V(H) < V(M)$,

$$\text{(6.7)} \qquad \lim_{\rho \to +1} \frac{V(\bar{X})}{V(M)} = \frac{2}{\pi \log 2} \sim .9184$$

and

$$\text{(6.8)} \qquad \lim_{\rho \to +1} \frac{V(\bar{X})}{V(H)} \sim .9853 .$$

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

$$\arcsin x = \sum_{j=1}^{\infty} a_j x^{2j+1} . \tag{6.9}$$

In terms of this expansion,

$$nV(M) = \sum_{k=-\infty}^{\infty} \arcsin \rho^{|k|} = \sum_j a_j \sum_k (\rho^{2j+1})^{|k|} = \Sigma a_j \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} . \tag{6.10}$$

Thus, the reciprocal of the efficiency is asymptotically

$$\frac{V(\bar{M})}{V(\bar{X})} = \sum_j a_j \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} \cdot \frac{1-\rho}{1+\rho} . \tag{6.11}$$

As $\rho \to +1$, the j^{th} term approaches $a_j/(2j+1)$, so that

$$\lim_{\rho \to 1} \frac{V(\bar{X})}{V(M)} = \sum_j \frac{a_j}{(2j+1)} . \tag{6.12}$$

From the expansion (6.9), it follows that

$$\sum_j \frac{a_j x^{2j}}{(2j+1)} = \frac{1}{x} \int_0^x \frac{\arcsin y}{y} dy = \frac{1}{x} \int_0^{\sin^{-1} x} z \cot z \; dz . \tag{6.13}$$

Evaluation of (6.13) at $x = 1$ yields

$$\lim_{\rho \to 1} \frac{V(M)}{V(\bar{X})} = \sum \frac{a_j}{(2j+1)} = \int_0^{\pi/2} z \cot z \; dz = \frac{\pi}{2} \log 2 . \tag{6.14}$$

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

$$(6.15) \qquad V(H) = 2\sum_{k=-\infty}^{\infty} \arcsin \frac{\rho^{|k|}}{2} = 2\sum_{-\infty}^{\infty}\sum_{j} a_j \frac{\rho^{(2j+1)|k|}}{2^{2j+1}}$$

$$= \sum_j a_j \, 2^{-2j} \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} \quad .$$

Thus,

$$(6.16) \qquad \lim_{\rho \to 1} \frac{V(H)}{V(\bar{X})} = \lim_{\rho \to 1} \sum_j a_j \, 2^{-2j} \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} \cdot \frac{1+\rho}{1+\rho} = \sum_{j=1}^{\infty} a_j \frac{2^{-2j}}{(2j+1)} \quad .$$

The right side of equation (6.16) is obtained by evaluating expression (6.13) at $x = 1/2$. Thus,

$$(6.17) \qquad \lim_{\rho \to 1} \frac{V(H)}{V(\bar{X})} = 2\int_0^{\pi/6} z \cot z \; dz = 2\left[\frac{\pi}{6} - \sum_{j=1}^{\infty} B_j \frac{(\pi/3)^{2j+1}}{(2j+1)!} \right] ,$$

where the B_j are the Bernoulli numbers.

The asymptotic relative efficiencies of M and HL as ρ approaches -1 is given in

<u>Theorem 6.4</u>. <u>If</u> $\rho < 0$, <u>then</u> $V(\bar{X}) < V(HL) < V(M)$,

$$(6.18) \qquad \lim_{\rho \to -1} \frac{V(\bar{X})}{V(M)} = 0 \quad ,$$

<u>and</u>

$$(6.19) \qquad \lim_{\rho \to -1} \frac{V(\bar{X})}{V(HL)} = \frac{\sqrt{3}}{2} \quad .$$

Proof: By L'Hospital's rule,

$$\text{(6.20)} \qquad \lim_{\rho \to -1} \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} \cdot \frac{1-\rho}{1+\rho} = (2j+1) \quad .$$

Substituting (6.20) into the formula (6.11) yields

$$\text{(6.21)} \qquad \lim_{\rho \to -1} \frac{V(M)}{V(\bar{X})} = \Sigma\, a_j\, (2j+1) \quad .$$

Similarly,

$$\text{(6.22)} \qquad \lim_{\rho \to -1} \frac{V(H)}{V(\bar{X})} = \Sigma\, a_j\, (2j+1)\, 2^{-2j} \quad .$$

Since

$$\text{(6.23)} \qquad \Sigma\, a_j x^{2j+1} = \arcsin x = \int_0^x \frac{dy}{\sqrt{1-y^2}} = \int \Sigma\, b_j x^{2j} \quad ,$$

it follows that $a_j = b_j/(2j+1)$. Thus,

$$\text{(6.24)} \qquad \lim_{\rho \to -1} \frac{V(M)}{V(\bar{X})} = \Sigma\, b_j = \lim_{j \to 1} \frac{1}{\sqrt{1-j^2}} = \infty \quad ,$$

and

$$\text{(6.25)} \qquad \lim_{\rho \to -1} \frac{V(H)}{V(\bar{X})} = \Sigma\, b_j 2^{-2j} = \frac{1}{\sqrt{1-(1/2)^2}} = \frac{2}{\sqrt{3}} \quad .$$

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. Later we shall see that this is characteristic of any finite linear combination of sample percentiles. As robustness studies are usually concerned with the sensitivity of procedures to small departures from the basic assumptions we present Table 6.1 of the efficiencies of several robust extimators 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 6.1: 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 25^{th} and 75^{th} Percentiles
- 1	.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 6.1 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 [17] 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 $.2 \leq \alpha \leq .4$, 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, $-.2 \leq \rho \leq .2$, an interpolation showed that $W(.27)$ remains nearly optimum. Finally, a glance at Table 6.1 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 [5], [9], [14], [24] concerning the robustness properties of $T(.25)$.

The results reported above are based on Table 6.2.

Table 6.2: The Asymptotic Efficiency of $W(\alpha)$ and $T(\alpha)$ Relative to $\overline{X}$

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

We conclude this chapter with the analysis of the behavior of the efficiency of any symmetric linear combination of the order statistics, W, relative to $\bar{X}$ as $\rho \to -1$. First we need

Lemma 6.3. The reciprocal of the efficiency of any symmetric estimator relative to $\bar{X}$ as $\rho \to -1$, is given by

$$(6.26) \qquad \lim_{t \to 1} \int\int \frac{1}{2\pi\sqrt{1-t^2}} e^{-\frac{1}{2(1-t^2)}(x^2-2txy+y^2)} d\mu(x)\, d\mu(y) \quad .$$

Proof: Since $c_{2j} = 0$ for symmetric estimators, (6.16) is just

$$(6.27) \qquad \sum_{k=1}^{\infty} \frac{c_k}{(k-1)!} = \lim_{t \to 1} \frac{d}{dt}\left(\sum_{k=0}^{\infty} \frac{c_k t^k}{k!} \right)$$

and

$$(6.28) \qquad \sum \frac{c_k t^k}{k!} = \int\int \{P_t[X < x, Y < y] - \Phi(x)\Phi(y)\}\, d\mu(x)\, d\mu(y) \quad ,$$

where P_t denotes the joint c.d.f. of two standard normal r.v.'s with correlation t. Differentiating (6.28) with respect to t yields

$$(6.29) \qquad \sum_{k=1}^{\infty} \frac{c_k}{(k-1)!} = \lim_{t \to 1} \iint \frac{1}{2\pi(1-t^2)^{1/2}} e^{-(x^2-2txy+y^2)/2(1-t^2)}\, d\mu(x)\, d\mu(y) \, .$$

In order to illustrate the use of Lemma 6.3 we prove

Corollary 6.2. The efficiency of the α-trimmed mean, relative to $\bar{X}$ on P.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 (6.29) becomes

$$(6.30)\quad \lim_{n\to\infty} \frac{V(\sqrt{n}\,T(\alpha))}{V(\sqrt{n}\,\bar{X})} = (1-2\alpha)^{-2} \lim_{t\to 1} \int_{-B}^{B}\int_{-B}^{B} \frac{1}{2\pi(1-t^2)^{1/2}}\, e^{-\frac{(x^2-2txy+y^2)}{2(1-t^2)}}\, dx\, dy \;.$$

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

<u>Lemma 6.4</u>. <u>For any</u> $\epsilon > 0$,

$$(6.31)\qquad \lim_{t\to 1} \iint_{|x-y|>\epsilon} f_t(x,y)\, d\mu(x)\, d\mu(y) = 0 \;.$$

<u>Proof</u>: As $f_t(x,y) \geq 0$, (6.31) certainly holds if the same limit with μ replaced by μ^* is valid. Since

$$(6.32)\qquad f_t(x,y) = (2\pi\sqrt{1-t^2}\,)^{-1}\, e^{-(x^2+y^2)/2(1+t)}\; e^{-\frac{t}{2(1-t^2)}(x-y)^2} \;,$$

$$(6.33)\qquad \iint_{|x-y|>\epsilon} f_t(x,y)\, d\mu^*(x)\, d\mu^*(y) \;\leq\; \frac{e^{-\frac{t\epsilon^2}{2(1-t^2)}}}{\sqrt{1-t^2}} \left[\int (2\pi)^{-1/2}\, e^{-\frac{x^2}{2(1+t)}}\, d\mu^*(x)\right]^2 .$$

Integration by parts and applying the Schwarz inequality yields

$$(6.34)\quad \left[\int (2\pi)^{-1/2} e^{-\frac{x^2}{2(1+t)}} d\mu^*(x)\right]^2 = \left[\int \frac{\mu^*(x)}{\sqrt{2\pi}} \frac{x}{1+t} e^{-\frac{x^2}{2(1+t)}} dx\right]^2$$

$$= \left[\int \mu^*(x) e^{-\frac{x^2}{4}} x \frac{e^{-\frac{x^2}{4}\left(\frac{1-t}{1+t}\right)}}{1+t} dx\right] \le \left[\int \mu^{*2}(x) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx\right]\left[\int \frac{1}{\sqrt{2\pi}(1+t)^2} x^2 e^{-\frac{x^2}{2}\frac{(1-t)}{(1+t)}} dx\right].$$

As the right side of (6.34) is $K\cdot(1-t)^{-3/2}(1+t)^{-1/2}$, the right side of (6.33) is

$$(6.35)\qquad \le K(1+t)^{-1}(1-t)^{-2} e^{-\frac{\epsilon^2 t}{2(1-t^2)}}$$

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

$$(6.36)\qquad \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} = \int_{-\infty}^{\infty} \frac{1}{2y} \chi_y(x)\, d\eta_\sigma(y) \quad ,$$

where η is a probability measure and $\chi_y(x)$ is the indicator of the set $\{x: |x| \le y\}$. Substituting (6.36) into (6.32) and (6.26) means that we must prove that

$$(6.37)\qquad \lim_{t\to\infty} \frac{1}{\sqrt{t}} \int\int\int \frac{1}{\sqrt{2\pi}} e^{-\frac{(x^2+y^2)}{2(1+t)}} \frac{\chi_w(x-y)}{2w} dy_t(w)\, d\mu(x)\, d\mu(y) = \infty \ .$$

By the lemma we can restrict ourselves to the region $|x-y| < \epsilon$. Letting $z = \min(\epsilon, w)$, (6.37) becomes

$$(6.38) \qquad \lim_{t \to 1} \frac{1}{\sqrt{t}} \int \left[\int \int \frac{1}{\sqrt{2\pi}} e^{-\frac{(x^2+y^2)}{2(1+t)}} \frac{\chi_z(x-y)}{2z} \, d\mu(x) \, d\mu(y) \right] \frac{z}{w} \, d\eta_t(w) \quad .$$

As $t \to 1$, $\eta_t(w)$ places more of its mass in a small neighborhood of the origin so that the limit in (6.38) is

$$(6.39) \qquad \lim_{\substack{t \to 1 \\ z \to 0}} \int \int \frac{1}{\sqrt{2\pi}} e^{-\frac{(x^2+y^2)}{2(1+t)}} \frac{\chi_z(x-y)}{2z} \, d\mu(x) \, d\mu(y) \quad .$$

When $\mu = \alpha\mu_o + \beta\mu_1$, where μ_o is a unit mass at ζ and ζ is <u>not</u> an atom of μ_1 , evaluation of the double integral in (6.39) yields

$$(6.40) \qquad \frac{\alpha^2 e^{-\zeta^2/(1+t)}}{(\sqrt{2\pi})(2z)} + \alpha \frac{\beta \, e^{-\zeta^2/2(1+t)}}{z} \int_{\zeta-z}^{\zeta+z} \frac{1}{\sqrt{2\pi}} \, e^{-y^2/2(1+t)} \, d\mu_1(y)$$

$$+ \beta^2 \int \int \frac{1}{\sqrt{2\pi}} \, e^{-\frac{(x^2+y^2)}{2(1+t)}} \frac{\chi_z(x-y)}{2z} \, d\mu_1(y) \, d\mu_1(y) \quad .$$

The third term is ≥ 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

<u>Theorem 6.4. Let</u> $A(\mu)$ <u>be the Riemann-Hellinger integral</u>

$$\int \frac{e^{-x^2/2}}{\sqrt{2\pi}} \frac{(d\mu(x))^2}{dx} \quad . \qquad \underline{\text{Then}}$$

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

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

<u>Proof</u>: Let

$$(6.41) \quad H_\mu(z,t,w,v) = \sum_{n=1}^{\infty} \int_{z+nw}^{z+(n+1)w} \int_{z+v+nw}^{z+v+(n+1)w} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2(1+t)}(x^2+y^2)} \, d\mu(x)\, d\mu(y) \quad .$$

Clearly

$$(6.42) \qquad \lim_{\substack{t \to 1 \\ w \to 0}} \frac{1}{w} H_\mu(z,t,w,0) = A(\mu)$$

if $A(\mu)$ exists and

$$(6.43) \qquad \lim_{\substack{t \to 1 \\ w \to 0}} \frac{1}{w} H_\mu(z,t,w,0) \le B(\mu) \quad .$$

Now let μ be positive. Then

$$(6.44) \quad \frac{1}{2w} H_\mu(z,t,w,0) \le \int\int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2(1+t)}} \frac{\chi_w(x-y)}{2w} \, d\mu(x)\, d\mu(y) \quad ,$$

which, from (6.42), proves (i) . (See Figure 1.)

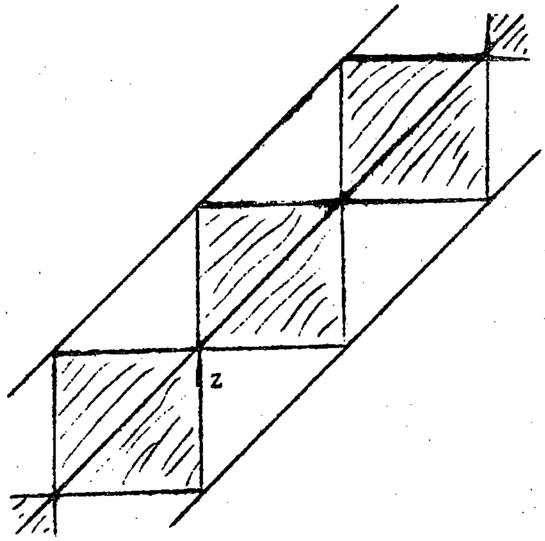

Figure 1

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

$$(6.45) \qquad \overline{H}_\mu(t,w) = \int_0^w H_\mu(z,t,w,0)\,dz \;.$$

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

$$(6.46) \qquad \int_0^w H_\mu(z,t,w,0)\,dz - \int_0^s H_\mu(z,t,s,0)\,dz$$

$$= r \iint \chi_w(x-y)\, e^{-\frac{x^2+y^2}{2(1+t)}}\, d\mu(x)\, d\mu(y) - R(w,s,t) \quad ,$$

where

$$(6.47) \qquad |R(w,s,t)| = 2 \left| \int_0^r \int_{z+(n-1)r}^{z+nr} \int_{z+s+nr}^{z+s+(n+1)r} \chi_w(x-y)\, e^{-\frac{x^2+y^2}{2(1+t)}}\, d\mu(x)\, d\mu(y) \right|$$

$$\leq 2 \int_0^r \int_{z+(n-1)r}^{z+nr} \int_{z+s+nr}^{z+s+(n+1)r} e^{-\frac{x^2+y^2}{2(1+t)}}\, d\mu^*(x)\, d\mu^*(y)$$

$$\leq 2\, \overline{H}_{\mu^*}(t,r) \quad .$$

Now let $s = (1-\delta)w$, $r = \delta w$. From (6.46) and (6.47) ,

$$(6.48) \qquad \left| \frac{1}{2w} \iint \chi_w(x-y)\, e^{-\frac{x^2+y^2}{2(1+t)}}\, d\mu(x)\, d\mu(y) \;-\; \frac{1}{2rw}(\overline{H}_\mu(t,w) - \overline{H}_\mu(t,s)) \right|$$

$$\leq \frac{1}{rw}\, \overline{H}_{\mu^*}(t,r) \quad .$$

Now

$$(6.49) \qquad \begin{aligned} \overline{H}_\mu(t,r) &= w^2 A(\mu) + o(w^2) \quad , \\ \overline{H}_\mu(t,s) &= s^2 A(\mu) + o(w^2) \quad , \\ \overline{H}_{\mu^*}(t,r) &= r^2 B(\mu^*) + o(w^2) \quad , \end{aligned}$$

so

$$(6.50) \qquad \left| \frac{1}{2w} \iint \chi_w(x-y)\, e^{-\frac{x^2+y^2}{2(1+t)}}\, d\mu(x)\, d\mu(y) \;-\; (1-\tfrac{\delta}{2})\, A(\mu) \right|$$

$$\leq \delta\, B^*(\mu) + o(1) \quad .$$

The result follows easily from (6.50).

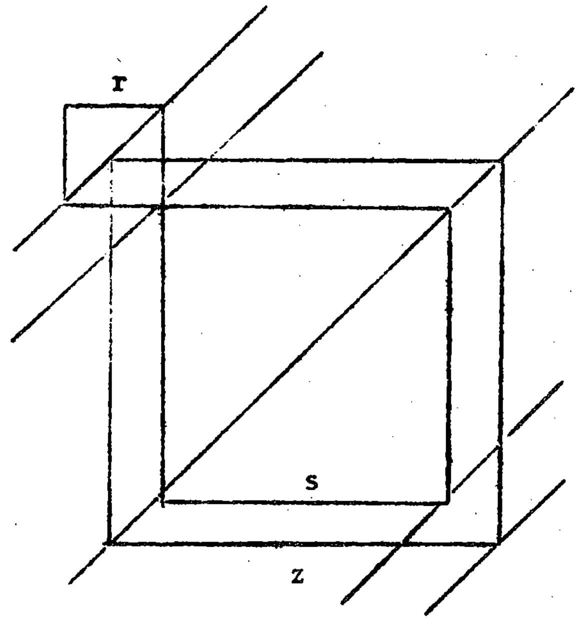

Figure 2

7. <u>The Behavior of Some Estimators on the First Order Autoregressive Double-Exponential Process</u>

In this section we shall summarize the behavior of the mean, $\bar{X}$, median, M, the average of two symmetric percentiles, $W(\alpha)$, and the Hodges-Lehmann estimator, HL, when the observations come from a F.O.A.D.P. 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 $\rho \to -1$.

We begin the section by deriving the asymptotic variance of the estimator $W(\alpha)$:

<u>Theorem 7.1</u>. <u>The average $W(\alpha)$ of the upper and lower $100\ \alpha^{th}$ percentiles on a F.O.A.D.P. is asymptotically normally distributed with expectation the mean of the process and asymptotic variance given by</u>

$$(7.1)\qquad \lim_{n\to\infty} nV\,[W(\alpha)] = \sum_{k=-\infty}^{\infty} (\rho^{|k|}\cosh \nu\rho^{|k|} + \sinh \nu\rho^{|k|})$$

$$= e^{\nu} + 2\sum_{j=0}^{\infty} \frac{\nu^{2j}}{(2j)!}\,\frac{\rho^{2j+1}}{1-\rho^{2j+1}} + 2\sum_{j=0}^{\infty} \frac{\rho^{2j+1}}{1-\rho^{2j+1}}\,\frac{\nu^{2j+1}}{(2j+1)!} \quad ,$$

<u>where ν is the upper α^{th} point of the double-exponential distribution</u>.

<u>Proof</u>: The only part of the statement that does not follow directly from the results in the appendix to Section 2 is the calculation of the asymptotic variance. We evaluate it using formula (4.3) where $x = -\nu$, $y = \nu$, $\beta = 1-\alpha$, $f(\nu) = f(-\nu) = \frac{1}{2}\,e^{-\nu} = \int_{\nu}^{\infty} \frac{1}{2}\,e^{-x}dx = \alpha$. Formula (4.3) depends on $P[X_o < -\nu,\ X_k < \nu]$ which one evaluates by noting that

$$(7.3)\qquad P\,[X_o < -\nu,\ X_k < \nu] = \frac{1}{2}\,e^{-\nu} - P\,[X_o < -\nu,\ X_k > \nu]\ .$$

From Lemma 2.1, it follows that

$$(7.4)\qquad P\,[X_o < -\nu,\ X_k > \nu] = \int_{-\infty}^{-\nu} \frac{1}{2}\,e^{x}\left\{(1-\rho^{2k})\int_{\nu-\rho^k x}^{\infty} \frac{1}{2}\,e^{-y}\,dy\right\}dx$$

$$= \frac{(1-\rho^k)e^{-2\nu}\,e^{-\nu\rho^k}}{4} \quad ,$$

so that

$$(7.5) \qquad \mathrm{Cov}\,[Y_o(-\nu),\, Y_k(\nu)] = \frac{e^{-2\nu}}{4}(1 - e^{-\nu\rho^k} + \rho^k e^{-\nu\rho^k}) \ .$$

Then (4.3) becomes

$$(7.6) \qquad n^{-1} \sum_{k=-\infty}^{\infty} (1 + \rho^{|k|} e^{-\nu\rho^{|k|}} - e^{-\nu\rho^{|k|}}) \ .$$

As both sample percentiles have the same asymptotic variance, which is

$$(7.7) \qquad n^{-1} \sum_{k=-\infty}^{\infty} (e^{\nu\rho^{|k|}} + \rho^{|k|} e^{\nu\rho^{|k|}} - 1) \ ,$$

the asymptotic variance of $W(\alpha)$ is

$$(7.8) \qquad \frac{1}{2} n^{-1} \{ \sum_{k=-\infty} \rho^{|k|} (e^{\nu\rho^{|k|}} + e^{-\nu\rho^{|k|}}) + \sum_{k=-\infty}^{\infty} (e^{\nu\rho^{|k|}} - e^{-\nu\rho^{|k|}}) \}$$

which is equivalent to (7.1).

As a first application of Theorem 7.1 we prove

<u>Corollary 7.1</u>. <u>The median has the minimum asymptotic variance of any average of two symmetric percentiles on any</u> F.O.A.D.P.

<u>Proof</u>: The variance of the average of two symmetric percentiles can be regarded as a function of ν for $0 \leq \nu < \infty$. For any $\rho > 0$, the function (7.2) is increasing in ν so its minimum is attained when $\nu = 0$. For negative values of ρ, differentiating (7.1) yields

$$(7.9) \qquad \Sigma\, \rho^{|2k|} \sinh \nu\rho^{|2k|} + \Sigma\, \rho^{|k|} \cosh \nu\rho^{|k|} \ .$$

The second term in expression (7.9) equals

$$(7.10) \quad \Sigma\, \rho^{|k|} \cosh \nu\rho^{|k|} = \Sigma\, \Sigma\, \rho^{|k|} \frac{(\nu\rho^{|k|})^{2j}}{(2j)!} = \sum_j \frac{\nu^{2j}}{(2j)!} \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}}$$

which increases as a function of ν for any value of ρ. The first term in expression (7.9) also increases as a function of ν for all values of ρ so that (7.9), the derivative of the asymptotic variance is a positive increasing function of ν. Hence, the asymptotic variance attains its minimum when $\nu = 0$ which corresponds to $\alpha = \frac{1}{2}$ or the median.

When we compare the efficiency of any estimator $W(\alpha)$ to the median as ρ varies we obtain the following analog of Corollary 6.1:

Corollary 7.2. *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 .

Proof: Expanding the term e^{ν} in formula (7.2) and collecting terms shows that the reciprocal of the efficiency of $W(\alpha)$ to M is given by

$$(7.11) \quad \sum_{j=0}^{\infty} \frac{\nu^{2j}}{(2j)!} \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} + \sum_{j=0}^{\infty} \frac{\nu^{2j+1}}{(2j+1)!} \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} \frac{1-\rho}{1+\rho} .$$

By the second part of Lemma 6.2 each function

$$(7.12) \qquad \frac{1+\rho^{2j+1}}{1-\rho^{2j+1}} \cdot \frac{1-\rho}{1+\rho}$$

decreases as ρ increases from -1 to 1 . Hence, the reciprocal of the efficiency decreases and the result follows.

The Hodges-Lehmann estimator, HL , is always more efficient than $\bar{X}$ but is always less efficient than the median. This can be proved using Theorem 5.2

but we omit the details. Using the same methods which yielded Theorems 6.3 and 6.4 one obtains

Theorem 7.2. 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

$$\lim_{\rho \to \pm 1} V(\bar{X})/V(M) = 2 ,$$

$$\lim_{\rho \to +1} V(H)/V(M) = 2\left[3 \log \left(\tfrac{3}{2}\right) - \tfrac{1}{2}\right] \approx 1.4328 ,$$

$$\lim_{\rho \to +1} V[W(\alpha)]/V(M) = \frac{\sinh \nu}{\nu} + \sum_{j=0}^{\infty} \frac{\nu^{2j+1}}{(2j+1)(2j+1)!} ,$$

where ν is the upper α^{th} fractile of the double-exponential c.d.f. ,

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

and

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

In Table 7.1 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 $\rho = 0$, .866 at $\rho = -1$ and .923 at $\rho = -.3$ so that about 36 % of the total change in efficiency is achieved when $\rho = -.3$. In the double-exponential case 47 % of the total change in efficiency is achieved at $\rho = -.3$. This behavior is characteristic of all the estimators.

Probably the most basic conclusion that can be drawn from Table 7.1 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 processes. 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).

Table 7.1: The A.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

8. Two Models Allowing for Contaminated Observations

In this section we study two processes which are models of a basic sequence of i.i.d. r.v.'s $\{\epsilon_i\}$ 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 [13] 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 [24] when the $\{\epsilon_i\}$ 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 $(L_1, L_2, \ldots)$ 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. r.v.'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

$$X_i = U_{N(i)} + \epsilon_i \quad . \tag{8.1}$$

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 $\overline{X}$, M and H-L when $\epsilon \sim \eta(0, e^2)$ and 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

<u>Proposition 8.1: When the observations are from Process 1, as</u> $n \to \infty$

$$V(\overline{X}) \sim \frac{1}{n}(1 + r\,\varphi''(1)/\varphi'(1)) \tag{8.2}$$

$$V(M) \sim \frac{\pi}{2n}\left(1 + \frac{2}{\pi}\frac{\varphi''(1)}{\varphi'(1)} \arcsin r\right) \tag{8.3}$$

<u>and</u>

$$V(H\text{-}L) \sim \frac{\pi}{3n}\left(1 + \frac{6}{\pi}\frac{\varphi''(1)}{\varphi'(1)} \arcsin \frac{r}{2}\right) . \tag{8.4}$$

<u>Proof</u>: As usual

$$n^2 V(\overline{X}) = \sum_{i=1}^{n} V(X_i) + \sum_{i \neq j}\sum \mathrm{Cov}(X_i, X_j) . \tag{8.5}$$

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 $n \to \infty$

$$\sum_{i \neq j}\sum \mathrm{Cov}(X_i, X_j) \sim \frac{n\,\varphi''(1) r}{\varphi'(1)} , \tag{8.6}$$

and

$$(8.7) \qquad V(\bar{X}) \sim \frac{1}{n}(1 + r\,\varphi''(1)/\varphi'(1)) \quad .$$

To derive (8.3), consider the sign test statistic S_n. Its variance is

$$(8.8) \qquad V(S_n) = \sum_{i=j=1}^{n} (1/4) \quad + \sum_{i \neq j} [P(X_i > 0,\ X_j > 0) - 1/4] \quad .$$

The terms in the second sum on the right side of (8.8) are 0 unless i and j are in the same phase in which case $P(X_i > 0,\ X_j > 0) - 1/4 = \frac{1}{2\pi}$ arcsin r. Counting the number of correlated X_i's as in the preceeding paragraph yields

$$(8.9) \qquad V(S_n) \sim \frac{n}{4}\left(1 + \frac{\varphi''(1)}{\varphi'(1)}\frac{2}{\pi} \text{ arcsin } r\right)$$

and (8.3) follows from (8.9).

The derivation of (8.4) follows the one in Section 5. There is one simplification. In calculating the variance of $\Sigma\, I_{ij}$, one need only calculate

$$(8.10) \qquad \sum_{i,j} [P(X_i + Y > 0,\ X_j + Z > 0) - 1/4]$$

where Y, Z are independent of X_i, Y_j and of each other and have the stationary distribution of the process. The number of other terms is of lower order. When $i = j$, $P(X + Y > 0,\ X + Z > 0) = 1/3$ as all the r.v.'s are symmetric. If i and j are not in the same phase $P(X_i + Y > 0,\ X_j + Z > 0) = 1/4$. 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 covariance matrix

$$(8.11) \qquad 2\begin{pmatrix} 1 & \frac{r}{2} \\ \frac{r}{2} & 1 \end{pmatrix}$$

Thus, $P(X_i + Y > 0,\ X_j + Z > 0) - 1/4 = \frac{1}{2\pi} \arcsin r/2$, when i,j are in the same phase and the variance of the appropriate sign test statistic is

$$(8.12) \qquad n^3\left(\frac{1}{12} + \frac{1}{2\pi}\frac{\varphi''(1)}{\varphi'(1)} \arcsin r/2\right) .$$

Converting this to the variance of the H-L estimator as before yields

$$(8.13) \qquad \frac{\pi}{3n}\left(1 + \frac{6}{\pi}\frac{\varphi''(1)}{\varphi'(1)} \arcsin r/2\right) .$$

From formulas (8.2), (8.3) and (8.4) 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 $\varphi''(1)/\varphi'(1)$. This is essentially an application of Lemma 6.1 but can be seen directly here as $(\arcsin r)/r$ and $\arcsin(r/2)/(r/2)$ are increasing functions of r .

Usually we are concerned with values of $r > 1/2$, i.e., the contaminant ordinarily has a larger variance than the underlying i.i.d. r.v.'s . As $r \to 1$, all the terms in the parentheses in formulas (8.2), (8.3) and (8.4) 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 8.1. Of course, for any r the efficiency 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 <u>not</u> 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. At first glance this may seem surprising. However $\tau \to 0$ is equivalent to $\varphi''(1)/\varphi'(1) \to \infty$ and for small r , $2 \arcsin(r/2) \sim r \sim \arcsin r$.

Model 1 can be regarded as a generalization of Hoyland's [13] model, in which L is a constant c . He studied the Hodges-Lehmann estimator but gives a formula for arbitrary distributions which essentially replaces $\frac{1}{2\pi} \arcsin r/2$ by $P(X_i + Y > 0, X_j + Z > 0) - 1/4$, where X_i and X_j are in the same contamination period in formula (8.12) 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 (8.3) can also be generalized to an arbitrary marginal distribution and is

$$(8.3^*) \qquad \frac{1}{4f^2(0)n} \left(1 + \frac{\varphi''(1)}{\varphi'(1)} \lambda\right) ,$$

where $\lambda = P[X_i > 0, X_j > 0] - 1/4$ 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, so that $P(L=j) = q^{j-1}p$, $j=1, 2, \ldots$ and $\varphi(z) = (pz)(1-qz)^{-1}$, $\varphi'(1) = p^{-1}$ and $\varphi''(1) = 2qp^{-2}$. If we assume that the average group size is c, then $p = 1/c$ so that $\varphi''(1)/\varphi'(1)=2(c-1)$. 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. Finally, we note that in Hoyland's model $\tau = 1/c$, so that relatively small values of τ are probably of interest and for any value of r the relative efficiency of both M and HL increases with the size c of each group. His Table 1 of the A.R.E. of HL to $\bar{X}$ is consistent with this.

Table 8.1: 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

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

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 ϵ_i where the ϵ_i are i.i.d. standard normal r.v.'s. During the second phase $X_i = \epsilon_i + U$, 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 process (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 non-zero 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

$$(8.14) \qquad p = \frac{\varphi'(1)}{\varphi'(1) + \Psi'(1)} \quad \text{and} \quad \gamma = \frac{\varphi''(1)}{\varphi'(1) + \Psi'(1)} = p\,\frac{\varphi''(1)}{\varphi'(1)} .$$

The marginal distribution of X_i is the contaminated normal,

$$q\,N(0,\,1) \;+\; p\,N(0,\,1+\eta) \tag{8.15}$$

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 normal distribution with mean 0 and variance $1+\eta$. The asymptotic behavior of $\bar{X}$, M and H-L are given in the following:

<u>Proposition 8.2</u>: <u>When observations come from Process 2 the estimators</u> $\bar{X}$, M <u>and</u> H-L <u>are asymptotically normally distributed with asymptotic variances given by</u>

$$V(\bar{X}) \sim \frac{1}{n}\,(1 + (p+\gamma)\eta) \tag{8.16}$$

$$V(M) \sim \frac{1}{n}\,\frac{\pi}{2}\,\frac{1}{(q+p/(1+\eta)^{1/2})^2}\left[1 + \frac{2\gamma}{\pi}\;\arcsin\frac{\eta}{1+\eta}\right] \tag{8.17}$$

and

$$V(\text{H-L}) \sim \frac{1}{n}\,\frac{\pi}{3}\,\frac{\left(1+\frac{6\gamma}{\pi}\left[q^2\arcsin\frac{\eta}{1+\eta} + 2pq\arcsin\frac{\eta}{\sqrt{(2+\eta)(2+2\eta)}} + p^2\arcsin\frac{\eta}{2+2\eta}\right]\right)}{\left(q^2 + \frac{2pq}{\sqrt{1+\eta/2}} + \frac{p^2}{\sqrt{1+\eta}}\right)^2}\;. \tag{8.18}$$

<u>Proof</u>: The method of proof is the same as that of Proposition 8.1. In the derivation of $V(\bar{X})$ the individual variance, $V(X_i)$ is $1 + p\eta$ and $\mathrm{Cov}\,(X_i, X_j) = 0$ unless X_i and X_j are in the same contamination period. The proof is the same as that of (8.2) with r replaced by η and $\varphi''(1)/\varphi'(1)$ replaced by γ. The asymptotic variances of the median and H-L estimators are also found in the same manner. The only part which is more difficult is

the calculation of (8.10) the variance of the sign test statistic for the H-L estimator. When X_i and X_j are in the same contamination period they are (conditionally) jointly normal with covariance matrix

$$(8.19) \qquad \begin{pmatrix} 1+\eta & \eta \\ \eta & 1+\eta \end{pmatrix}$$

As Y and Z are independent r.v.'s with a contaminated normal distribution (8.15), the r.v.'s $X_i + Y$ and $X_j + Z$ are mixtures of bivariate normals with mean 0 and the following covariance structures (and mixture weights)

$$(8.20) \qquad \Sigma = \begin{pmatrix} (2+\eta) & \eta \\ \eta & 2+\eta \end{pmatrix}, \quad \text{with probability } q^2$$

$$\Sigma = \begin{pmatrix} 2+\eta & \eta \\ \eta & 2+2\eta \end{pmatrix}, \quad \text{with probability } 2pq$$

$$\Sigma = \begin{pmatrix} 2+2\eta & \eta \\ \eta & 2+2\eta \end{pmatrix}, \quad \text{with probability } p^2$$

Thus,

$$(8.21) \qquad P(X_i + Y > 0,\ X_j + Z > 0) - 1/4 =$$

$$\frac{1}{2\pi}\left(q^2 \arcsin \frac{\eta}{1+\eta} + 2pq \arcsin \frac{\eta}{\sqrt{(2+\eta)(2+2\eta)}} + p^2 \arcsin \frac{\eta}{2+2\eta}\right)$$

so that the variance of the appropriate sign test statistic is

$$(8.22) \quad n^3\left[\frac{1}{12} + \frac{\gamma}{2\pi}\left(q^2 \arcsin \frac{\eta}{1+\eta} + 2pq \arcsin \frac{\eta}{\sqrt{(2+\eta)(2+2\eta)}} + p^2 \arcsin \frac{\eta}{2+2\eta}\right)\right] .$$

Since

$$(8.23) \qquad g(0) = \frac{1}{\sqrt{2\pi}}\left(\frac{q^2}{\sqrt{2}} + \frac{2pq}{\sqrt{2+\eta}} + \frac{p^2}{\sqrt{2+2\eta}}\right)$$

the variance of H-L is given by (8.18).

<u>Remarks</u>: As the variance , η , of the contaminating distribution approaches infinity the variance of $\bar{X}$ approaches infinity while

$$(8.24) \qquad nV(M) \to \frac{\pi}{2}\,\frac{(1+\gamma)}{q^2} ,$$

and

$$(8.25) \qquad nV(H) \to \frac{\pi}{3q^4}\left(1 + \gamma(3q^2 + 3pq + p^2)\right) .$$

Thus, the efficiency of H-L to the median becomes

$$(8.26) \qquad \frac{3}{2}\, q^2 \,\frac{(1+\gamma)}{1 + \gamma(3q^2 + 3pq + p^2)} .$$

In particular, if $p > 1 - \sqrt{2/3}$, the efficiency of H-L to the median is < 1 , regardless of the value of γ just as in the case of independent contaminated normal observations [10]. However, the Hodges-Lehmann estimator is more sensitive to contamination in this model since (8.26) increases as a function of γ .

<u>Remark</u>: 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.

REFERENCES

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

[2] Billingsley, P. (1968). Convergence of Probability Measure. John Wiley, New York.

[3] Blum, J. R. and Rosenblatt, M. (1956). A class of stationary processes and a central limit theorem. Proc. Nat. Acad. Sci. (43), 412-423.

[4] Cogburn, R. (1960). Asymptotic properties of stationary sequences. Univ. Calif. Pub. Stat. (3), 99-146.

[5] Crow, E. L. and Siddiqi, M. M. (1967). Robust estimation of location. J. Amer. Stat. Assoc. (62), 353-389.

[6] Doob, J. L. (1953). Stochastic Processes, John Wiley, New York.

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

[8] Gastwirth, J. L. , Rubin, H. and Wolff, S. S. (1967). The effect of autoregressive dependence on a non-parametric test. I.E.E.E. Transactions on Information Theory IT-13, 311-313.

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

[10] Gastwirth, J. L. and Cohen, M. L. (1968). The small sample behavior of some robust linear estimators of location. Department of Statistics Report 91. The Johns Hopkins Univ. Baltimore, Md.

[11] Hochstadt, H. (1961). Special Functions of Mathematical Statistics. Holt, Rinehart and Winston, New York.

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

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

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

[14a] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. Theor. Prob. Appl. 7 349-382.

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

[16] Mann, H. B. and Wald, A. (1943). On stochastic limit and order relations. Ann. Math. Statist. (14), 217-226.

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

[18] Prohorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory. Theo. Probability Appl. (1), 157-214.

[19] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. (42), 43-47.

[20] Rozanov, Yu. A. (1967). Stationary Random Processes. Holden Day. San Francisco, Calif.

[21] Rubin, H. (1968). A useful device for continuity theorems. Abstract 17. Ann. Math. Statist. (39), 1090.

[22] Serfling, R. J. (1968). Contributions to central limit theory for dependent variables. Ann. Math. Statist. (39), 1158-1175.

[23] Sun, T. C. (1963). A central limit theorem for non-linear functions of a normal stationary process. Jour. of Math. and Mech. 12 945-977.

[24] 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., ed. Stanford Univ. Press, Calif. 448-485.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) Purdue University	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3 REPORT TITLE
"The Behavior of Robust Estimators on Dependent Data"

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report August 1969

5 AUTHOR(S) (Last name, first name, initial)
Gastwirth, Joseph L. and Rubin, Herman

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1969	101	25

8a. CONTRACT OR GRANT NO. N00014-67-A-0226-0008 b. PROJECT NO. NR042-216 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Mimeograph Series # 197 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 AVAILABILITY/LIMITATION NOTICES
Distribution of this document is unlimited.

11 SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research Washington, D.C.

13 ABSTRACT

This report studies the effect of serial dependence on the efficiency of various robust estimators of the location parameter. In order to show that the asymptotic distribution of these estimators is a normal distribution we introduce a slightly stronger mixing condition than Rosenblatt's strong mixing and show that the empiric c.d.f. formed from such a process approaches a Gaussian process. In particular, first order autoregressive processes with Gaussian, Cauchy and double-exponential marginal distributions are shown to obey our conditions.

The behavior of robust estimators on Gaussian processes is studied in greater detail. One general result states that for any Gaussian process with serial correlation $\rho_k > 0$ and $\Sigma\rho_k < \infty$, the efficiency of any linear combination of the order statistics relative to the sample mean is greater than its efficiency in the case of independent observations. The same result holds for the Hodges-Lehmann estimator. These results are applied to two models of contamination and show that the estimators which have been developed to be robust against outliers are robust against dependence.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification

Unclassified
Security Classification

14 KEY WORDS	LINK A ROLE	WT	LINK B ROLE	WT	LINK C ROLE	WT
robust estimation						
strong mixing						
contaminated data						
asymptotic distribution theory						
weak convergence						
autoregressive processes						
empiric c.d.f.						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________ ."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________ ."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________ ."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

D 3551

Unclassified
Security Classification