BOUNDS ON ASYMPTOTIC RELATIVE EFFICIENCIES OF ROBUST ESTIMATES OF LOCATIONS FOR CONTAMINATIONS BY SCALE MIXTURES

by

Anirban DasGupta
Purdue University

Technical Report #90-64

Department of Statistics
Purdue University

December 1990
Revised November 1991
Revised June 1993

BOUNDS ON ASYMPTOTIC RELATIVE EFFICIENCIES OF ROBUST ESTIMATES OF LOCATIONS FOR CONTAMINATIONS BY SCALE MIXTURES*

by

Anirban DasGupta
Purdue University

Abstract

We consider robust estimation of a location parameter θ when the CDF of the error is of the form $F(x) = (1-\varepsilon_0)H(x) + \varepsilon_0 H(\frac{x}{\sqrt{s}})$, where H is symmetric and absolutely continuous. In practice, H is often taken to be the standard normal CDF Φ, as in the pioneering article of Tukey (1960). We assume s is random but we have the information that $E(s) = s_0$. Infimums and supremums of the asymptotic relative efficiencies of the Hodges–Lehmann estimate W, the median M and the general trimmed mean $\overline{X}_\alpha$ with respect to the sample mean are calculated under the assumption $s \geq s_1$. Here ε_0, s_0, s_1 are assumed known, but are otherwise arbitrary.

Parseval's relation is used to prove that if either $H(\sqrt{x})$ or $\psi(\sqrt{\lambda})$ is log convex (for $x, \lambda > 0$), where ψ is the characteristic function of H, then $\inf e(W, \overline{X})$ and $\inf e(M, \overline{X})$ are attained at $F_0(x) = (1-\varepsilon_0)H(x) + \varepsilon_0 H(\frac{x}{\sqrt{s_0}})$ over the class of distributions described above. This establishes, in a sense, the Tukey distribution as an extreme point of the family of distributions

$$\mathcal{F} = \left\{F: F(x) = \int \left\{(1-\varepsilon_0)H(x) + \varepsilon_0 H\left(\frac{x}{\sqrt{s}}\right)\right\} dG(s)\right\},$$

where G satisfies the constraints given above. This result is not necessarily valid for a general trimmed mean but it is proved that the mixing distribution on s corresponding to the infimum as well as the supremum is supported on at most 3 points. Extensive numerical studies indicate that the mixing distribution corresponding to the supremum is a two point distribution and that corresponding to the infimum is usually degenerate (at $s = s_0$) but is occasionally two point. It is also demonstrated that the infimum as well as the supremum of the efficiency remain bounded over $\mathcal{F}$ for each of the three estimates

* Research supported by NSF grant DMS 89–230–71.

and in fact they are very close to each other. Thus the efficiency remains stable over the entire class $\mathcal{F}$. The family $\mathcal{F}$ thus behaves like a singleton set. These results also enable us to answer a general version of a question raised by John Tukey (Tukey (1960)): given s_0, how large an $\varepsilon_0 > 0$ is required for the infimum of the efficiency over $\mathcal{F}$ to exceed 1?

Key Words: Asymptotic relative efficiency, contamination, Hodges–Lehmann estimate, median, mean, trimmed mean, convex, log convex, Schur convex, infimum, supremum.

1. Introduction

1.1. The basic setup. A common assumption in much of statistical theory is that the observed samples $x_1, x_2, \ldots$ are iid from a normal distribution. In practice, real data are rarely exactly normal and yet many standard statistical procedures are "optimal" provided the population is in fact normal. The realization that real data do not exactly satisfy the model assumptions that guarantee the optimal behavior of commonly used procedures led to the study of robustness and construction of robust procedures. There is now a vast literature on this topic; see Huber (1973) and Hampel et al (1986) for a general exposition and many other references.

In this article, we confine ourselves to the topic of robust estimation of a single location parameter. Formally, then, $X_1, X_2, \ldots$ are iid with $X_i = \theta + Z_i$ where Z_i are iid, distributed according to a symmetric distribution F on the real line. We assume F is symmetric about zero. There is a truly rich literature on this topic. It has long been well known that the sample mean is not a safe estimate of the location parameter θ if heavy tails are possible. Variety of alternative estimators are known and have been studied. We would not go into a discussion of the merits and demerits of such estimators; the literature includes the fundamental articles of Chernoff and Savage (1958) and Hodges and Lehmann (1963) and the latter works of many authors including Andrews et al (1972), Bickel (1965), Bickel and Lehmann (1975) etc. Lehmann (1983) is a particularly good source for a broad exposition.

A particularly attractive way of modelling deviation of F from the standard normal is the well known contaminated structure in which we assume that F is of the form $F(x) = (1-\varepsilon_0)\Phi(x)+\varepsilon_0 H(x)$ where Φ is the standard normal CDF and H is another CDF belonging to a suitable family of distributions on $\mathbb{R}$. If F is symmetric, so is H and vice versa. In one of the most fundamental articles on contaminated distributions, Tukey (1960) takes H to be the CDF of another normal distribution, although with a variance larger than 1. This can be justified as a model for a small fraction of outlying observations; Tukey (1960) discusses this in detail. Formally, then, Tukey (1960) takes $F(x) = (1-\varepsilon_0)\Phi(x)+\varepsilon_0 \Phi\left(\frac{x}{\sqrt{s}}\right)$ where s is usually greater than 1 or even substantially greater than 1 (although s near zero can be thought of as a continuous approximation to a point mass at zero and thus seems realistic also). Tukey (1960) allows ε_0 and s_0 to be arbitrary, but assumes they are fixed. We consider generalized Tukey models, with F belonging to the family of CDF's

$$\mathcal{F} = \left\{ F{:}F(x) = \int \left\{ (1-\varepsilon_0)\Phi(x) + \varepsilon_0 \, \Phi\left(\frac{x}{\sqrt{s}}\right) \right\} dG(s), \right.$$
$$\left. \text{where } G[s_1, \infty) = 1, \ \int s dG = s_0 \right\}, \tag{1.1}$$

with ε_0, s_0, s_1 fixed. As will be seen, s_0 and s_1 both play a prominent role in determining the merits and demerits of different estimators. The assumption $G[s_1, \infty) = 1$ reflects the belief that s is large.

An even greater generalization would be to use a more general CDF H instead of Φ in the definition of F in (1.1); for example, one can potentially contemplate situations in which most of the data come from a standard Double Exponential distribution but a small fraction comes from a stretched Double Exponential. Indeed, barring our results on the trimmed mean in Section 4, all other results are for the generalization of (1.1) when Φ is replaced by a more general H.

1.2. Discussion on $\mathcal{F}$. Several things need to be pointed out here. First, if s_1 is 0, then the contamination is an arbitrary scale mixture of normals with a given mean. The moment restriction $\int s dG = s_0$ puts a bound on the heaviness of the tail of F; also notice, denoting the density of F by f, that

$$f(0) = \frac{1}{\sqrt{2\pi}} \left(1 - \varepsilon_0 + \varepsilon_0 \int \frac{1}{\sqrt{s}} \, dG(s) \right),$$

so that if $s_1 > 0$, then the density at the center also remains in control. We have thus deliberately kept both the tail and the center of F in control; our setup thus relates to the structure in Loh (1984). It is also reassuring that $\mathcal{F}$ is a convex family of CDF's and each F in $\mathcal{F}$ is symmetric and unimodal. Following the arguments in Efron and Olshen (1978), one can also check that for ε_0 and s_0 not too small, $\mathcal{F}$ is quite broad in the sense that the probabilities of fixed intervals on the real line vary in a reasonably wide range as F varies over $\mathcal{F}$. $\mathcal{F}$ can thus be regarded as a moderately broad convex subclass of the class of symmetric unimodal distributions.

The ARE of the Hodges-Lehmann estimate W was earlier considered in Sen (1968). The family (1.1) can be regarded as a semiparametric extension of the setup there as well.

1.3. Objective. The main technical goal of the article is to find the infimum and the supremum of the asymptotic relative efficiency (ARE) of the sample mean with respect to three commonly used estimates of location: $W =$ the Hodges–Lehmann estimate, $M =$ the sample median and $\overline{X}_\alpha =$ the trimmed mean of order α. The median is treated separately because there are a number of technical simplifications that can be done for the median but seem difficult for a general trimmed mean. As in Loh (1984), we demonstrate that the infimum and the supremum of the ARE of each estimator with respect to the sample mean remain bounded; furthermore, we give evidence that often the infimum and the supremum are quite close. Other robust alternatives to the sample mean, such as winsorized means, also are worth investigating. We refrain from doing so due to space considerations.

We also prove a result which establishes, in a specific sense, the CDF $F_{\varepsilon_0,s_0}(x) = (1-\varepsilon_0)\Phi(x) + \varepsilon_0 \Phi\left(\frac{x}{\sqrt{s_0}}\right)$ as an extreme point of the family $\mathcal{F}$. Specifically, we show that for each of the Hodges–Lehmann estimate and the median, the infimum of the ARE with respect to the mean over all CDF's in $\mathcal{F}$ is attained at F_{ε_0,s_0}. This result did not seem to be easily provable for the trimmed mean but we do show that there is a broad generalization of this phenomenon in the following sense: if $\mathcal{F}$ is redefined with $\Phi(\cdot)$ replaced by $H(\cdot)$ in (1.1) where H is symmetric and either $H(\sqrt{x})$ or $\psi(\sqrt{\lambda})$ (where $\psi(\lambda)$ is the characteristic function of H) is log convex, then the same phenomenon holds with $\Phi(\cdot)$ replaced by H; i.e., the infima mentioned above are each attained at $(1-\varepsilon_0)H(x) + \varepsilon_0 H\left(\frac{x}{\sqrt{s_0}}\right)$. In particular, if H itself is any scale mixture of normals, then $H(\sqrt{x})$ as well as $\psi(\sqrt{\lambda})$ are log convex, although in general the log convexity of one does not seem to imply that of the other. We also explicitly answer a question originally raised in Tukey (1960), namely, given $s_0 > 0$ what is the smallest $\varepsilon_0 > 0$ required for the ARE of each estimate with respect to the mean to exceed 1? In fact, we answer a more general question: we describe the joint set of (ε_0, s_0) such that the ARE exceeds 1. Here the computations are not new but the explicit description of the set is new and should be useful.

The main results achieved in this article are thus the following:

i. establishing Tukey type models as extreme points of more general contamination families;

ii. demonstrating that the family (1.1) behaves like a singleton set in the sense of stable ARE's across the whole family; this is very intriguing;

iii. formally answering Tukey's question about the percentage of contamination needed for the (infimum) ARE over the mean to exceed 1;

iv. extension of earlier results of Sen (1968).

1.4. Outline. Section 2 treats the Hodges–Lehmann estimate W, section 3 treats the median and the general trimmed mean is discussed in section 4. The question of the relative efficiency of these three estimators between themselves is also of importance. We do not give any technical results on this but in section 5, we present some numerical comparison of these three estimates. Section 6 contains some concluding remarks.

Throughout the entire article, $e(W,\overline{X})$, $e(M,\overline{X})$ and $e(\overline{X}_\alpha,\overline{X})$ denote the ARE of W, M and $\overline{X}_\alpha$ respectively with respect to $\overline{X}$. The reader is reminded that these also correspond to the Pitman efficiencies of the corresponding tests with respect to the mean test.

2. Bounds on $e(W,\overline{X})$.

Denote the class of measures G such that $\int_{[s_1,\infty]} dG(s) = 1$ and $\int s dG = s_0$ by $\mathcal{G}$. For future reference, let

$$q(s) = \int h(x) h\left(\frac{x}{\sqrt{s}}\right)\, dx$$
$$Q(s,t) = \int h\left(\frac{x}{\sqrt{s}}\right)\, h\left(\frac{x}{\sqrt{t}}\right)\, dx, \tag{2.1}$$

where h is the density of H, a fixed symmetric and absolutely continuous distribution. We will make the following assumption about H:

Assumption (A): Assume H is symmetric and absolutely continuous with density h. Let

$\psi(\lambda)$ be the characteristic function of H. Let $h(x) = h_0(x^2)$ and $\psi(\lambda) = \psi_0(\lambda^2)$. Assume either $h_0(\cdot)$ or $\psi_0(\cdot)$ is log convex.

Lemma 2.1. Under assumption (A), $\frac{q(s)}{\sqrt{s}}$ and $\frac{Q(s,s)}{s}$ are both convex functions of s.

Proof: We will only prove that $r(s) = \frac{q(s)}{\sqrt{s}}$ is convex; the proof that $\frac{Q(s,s)}{s}$ is convex is similar. It is well known that because of measurability, it is enough to prove that $r\left(\frac{s_1+s_2}{2}\right) \le \frac{r(s_1)+r(s_2)}{2}$.

First note that
$$r(s) = \frac{q(s)}{\sqrt{s}} = \frac{1}{\sqrt{s}}\int h(x)h\left(\frac{x}{\sqrt{s}}\right)dx = \int h(x\sqrt{s})h(x)dx = \int h_0(sx^2)h_0(x^2)dx$$

$$\begin{aligned}
\therefore\quad & r\left(\frac{s_1+s_2}{2}\right) \\
&= \int h_0\left(\left(\frac{s_1+s_2}{2}\right)x^2\right)h_0(x^2)dx \\
&\le \int \sqrt{h_0(s_1x^2)h_0(s_2x^2)}\,h_0(x^2)dx \\
&= \int \sqrt{h_0(s_1x^2)h_0(x^2)}\sqrt{h_0(s_2x^2)h_0(x^2)}dx \\
&\le \int \frac{h_0(s_1x^2)h_0(x^2)+h_0(s_2x^2)h_0(x^2)}{2}dx \\
&= \frac{r(s_1)+r(s_2)}{2}.
\end{aligned}$$

In the above, the first inequality follows if h_0 is log convex and the second inequality is a consequence of the inequality $\sqrt{ab} \le \frac{a+b}{2}$ for $a,\ b \ge 0$.

We will now prove that $r(s)$ is convex if ψ_0 is log convex. But this follows from the fact that the map $h \to \psi$ is an isometry and hence by Parseval's identity (see Rudin (1973)), $\int h(x)h(x\sqrt{s})dx = \int \psi(\lambda)\psi(\lambda\sqrt{s})d\lambda$; the rest of the proof follows by a straightforward repetition of the argument given above.

Cor. The function $u(s) = \frac{2q(s)}{\sqrt{s}} + \frac{\alpha Q(s,s)}{s}$ is convex if $\alpha \ge 0$.

Lemma 2.2. Under assumption (A), $\inf_G \int u(x)dG(x) = u(s_0)$.

Proof: By a standard theorem (see Karlin and Studden (1966)),

$$\inf_G \int u(x)dG(x) = \sup\left\{as_0 + b\colon \text{ for } x \ge x_1,\ u(x) \ge ax+b\right\}. \tag{2.2}$$

The result now follows from (2.2) on noting that $u(x)$ is convex and the line tangent to $u(x)$ at the point s_0 is below $u(x)$ for all x.

Lemma 2.3. Under assumption (A), the function

$$f(s,t) = \frac{q(s)}{\sqrt{s}} + \frac{q(t)}{\sqrt{t}} + \frac{\alpha Q(s,t)}{\sqrt{st}} \text{ is Schur convex.}$$

Proof: Since $\frac{q(s)}{\sqrt{s}}$ is convex, it follows that $\frac{q(s)}{\sqrt{s}} + \frac{q(t)}{\sqrt{t}}$ is Schur convex (see Marshall and Olkin (1979)). It will then be sufficient to prove that $v(s,t) = \frac{Q(s,t)}{\sqrt{st}}$ is also Schur convex.

$$\begin{aligned}
\text{However, } v(s,t) &= \frac{1}{\sqrt{st}} \int h\left(\frac{x}{\sqrt{s}}\right) h\left(\frac{x}{\sqrt{t}}\right) dx \\
&= \int h(x\sqrt{s}) h(x\sqrt{t}) dx \\
&= \int h_0(sx^2) h_0(tx^2) dx \\
&\geq \int h_0^2 \left(\frac{(s+t)x^2}{2}\right) dx \\
&= \int h^2 \left(\sqrt{\frac{s+t}{2}} x\right) dx \\
&= v\left(\frac{s+t}{2}, \frac{s+t}{2}\right),
\end{aligned}$$

where the only inequality follows if h_0 is log convex. The proof when ψ_0 is log convex follows by again using Parseval's identity in the lines of Lemma 2.1. This completes the proof.

Theorem 2.4. Let $\mathcal{F}$ be defined as in (1.1) with $\Phi(\cdot)$ replaced by $H(\cdot)$. Under assumption (A), $\inf_{F \epsilon \mathcal{F}} e(W, \overline{X})$ is attained at $F_{\varepsilon_0, s_0}(x) = (1-\varepsilon_0)H(x) + \varepsilon_0 H\left(\frac{x}{\sqrt{s_0}}\right)$.

Proof: Recall that $e(W, \overline{X})$ equals $12 \cdot \sigma_f^2 \cdot (\int_{-\infty}^{\infty} f^2(x)dx)^2$ where f denotes the density of F and σ_f^2 denotes the variance of X under F. If h denotes the density of H, then by direct calculations,

$$\begin{aligned}
e(W, \overline{X}) &= 12 \cdot c \cdot (1 - \varepsilon_0 + \varepsilon_0 s_0) \\
&\times \left[(1-\varepsilon_0)^2 \int h^2(x)dx + 2\varepsilon_0(1-\varepsilon_0) \int\int \frac{1}{\sqrt{s}} h(x) h\left(\frac{x}{\sqrt{s}}\right) dG(s)dx \right. \qquad (2.3) \\
&\left. + \varepsilon_0^2 \int\int\int \frac{1}{\sqrt{st}} h\left(\frac{x}{\sqrt{s}}\right) h\left(\frac{x}{\sqrt{t}}\right) dG(s)dG(t)dx \right]^2,
\end{aligned}$$

where $c = \int_{-\infty}^{\infty} x^2 h(x)dx$.

If we let $\alpha = \frac{\varepsilon_0}{1-\varepsilon_0}$, then the problem of computing $\inf e(W, \overline{X})$ is equivalent to computing

$$\begin{aligned}
&\inf \left[2 \int \frac{q(s)}{\sqrt{s}} dG(s) + \alpha \int \int \frac{Q(s,t)}{\sqrt{st}} dG(s)dG(t)\right] \\
&= \inf \left[\int \int \left\{\frac{q(s)}{\sqrt{s}} + \frac{q(t)}{\sqrt{t}} + \frac{\alpha Q(s,t)}{\sqrt{st}}\right\} dG(s)dG(t)\right] \\
&\overset{\text{def}}{=} \inf \int \int f(s,t)dG(s)dG(t)
\end{aligned}$$

over $\mathcal{G}$.

We claim the infimum is attained at the measure degenerate at s_0. Towards this end, note that

$$\begin{aligned}
&f(s_0, s_0) \\
&\geq \inf_{\mathcal{G}} \int \int f(s,t)dG(s)dG(t) \\
&\geq \inf_{\mathcal{G}} \int \int f\left(\frac{s+t}{2}, \frac{s+t}{2}\right) dG(s)dG(t) \\
&\geq \inf_{\mathcal{G}} \int \int f(x,x)dG(x) \\
&= f(s_0, s_0);
\end{aligned}$$

here the first inequality is trivial, the second inequality is a consequence of Lemma 2.3, the third inequality follows from the fact that the marginal distribution of $\frac{s+t}{2}$ also belongs to $\mathcal{G}$ and the last inequality is guaranteed by Lemma 2.2. This proves the theorem.

Remark: Assumption (A) holds if H is any scale mixture of normal distributions with mean zero. In particular, the following result holds:

Cor. Let $\mathcal{F}$ be defined as in (1.1). Then $\inf_{F\varepsilon\mathcal{F}} e(W, \overline{X})$ is attained at $F_{\varepsilon_0, s_0}(x) = (1-\varepsilon_0)\Phi(x) + \varepsilon_0 \Phi\left(\frac{x}{\sqrt{s_0}}\right)$ and hence equals

$$e_w(\varepsilon_0, s_0) \overset{\text{def.}}{=} \frac{3}{\pi}(1 + (s_0 - 1)\varepsilon_0)\left[(1-\varepsilon_0)^2 + \frac{2\sqrt{2}\varepsilon_0(1-\varepsilon_0)}{\sqrt{1+s_0}} + \frac{\varepsilon_0^2}{\sqrt{s_0}}\right]^2. \tag{2.4}$$

Cor. Let $\mathcal{F}$ be defined as in (1.1) with Φ replaced by H. If H is any scale mixture of normals, then,

$$\inf_{F\epsilon\mathcal{F}} e(W,\overline{X}) = 12c(1+(s_0-1)\varepsilon_0)\cdot\Bigg[(1-\varepsilon_0)^2\int h^2(x)dx + \frac{2\varepsilon_0(1-\varepsilon_0)}{\sqrt{s_0}}\int h(x)h\left(\frac{x}{\sqrt{s_0}}\right)dx$$
$$+\frac{\varepsilon_0^2}{s_0}\int h^2\left(\frac{x}{\sqrt{s_0}}\right)dx\Bigg]^2,$$

where $c = \int x^2 h(x)dx$.

Discussion: One question of interest here is what is the overall infimum of $e(W,\overline{X})$ over all families $\mathcal{F}$ when H is also allowed to vary over arbitrary scale mixtures of normals; thus, if one lets $h(x) = \int \frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}d\nu(t)$, then by using the above corollary, one wishes to minimize $\int\int\left(\frac{1}{\sqrt{s+t}}+\frac{b}{\sqrt{s_0 s+t}}+\frac{b}{\sqrt{s_0 t+s}}\right)d\nu(s)d\nu(t)$ where b is a suitable positive real. One may assume that $\int t d\nu(t) = 1$ for the purpose of this minimization because of scale invariance. The argument used in Theorem 2.4 can now be used essentially word for word to prove that the minimum occurs at the measure ν_0 degenerate at $t=1$. The implication of this is that the overall infimum is attained at $H=\Phi$ and thus equals the expression in (2.4). This is, however, an expected result.

We next derive $\sup_{F\epsilon\mathcal{F}} e(W,\overline{X})$ where $\mathcal{F}$ is as in the statement of Theorem 2.4.

Theorem 2.5. Under assumption (A) and the additional assumption that $h(x)$ is uniformly bounded by some finite number,

$$\sup_{F\epsilon\mathcal{F}} e(W,\overline{X}) = 12c(1-\varepsilon_0+s_0\varepsilon_0)\times\Bigg[(1-\varepsilon_0)^2\int_{-\infty}^{\infty}h^2(x)dx+\varepsilon_0(1-\varepsilon_0)\cdot$$
$$\left\{\frac{2q(s_1)}{\sqrt{s_1}}+\frac{\alpha Q(s_1,s_1)}{s_1}\right\}\Bigg]^2. \tag{2.5}$$

Again, the proof of this requires several lemmas. We state them below with a short sketch of the proof in each case.

Lemma 2.6. Let $J(s)$ be defined as

$$J(s) = \frac{q(2s-s_1)}{\sqrt{2s-s_1}}+\frac{\alpha Q(s_1,2s-s_1)}{\sqrt{s_1(2s-s_1)}},\ s\geq s_1.$$

Then under assumption (A), $J(s)$ is convex.

Proof: We prove that $\frac{q(2s-s_1)}{\sqrt{2s-s_1}}$ and $\frac{Q(s_1,2s-s_1)}{\sqrt{s_1(2s-s_1)}}$ are separately convex. That $\frac{q(2s-s_1)}{\sqrt{2s-s_1}}$ is convex follows from Lemma 2.1. The convexity of $\frac{Q(s_1,2s-s_1)}{\sqrt{s_1(2s-s_1)}}$ follows on noting that

$$\begin{aligned}
N(s) &\overset{\text{def.}}{=} \frac{Q(s_1, 2s - s_1)}{\sqrt{s_1(2s - s_1)}} \\
&= \int h(x\sqrt{s_1})h(x\sqrt{2s - s_1})dx \\
&= \int h_0(s_1x^2)h_0((2s - s_1)x^2)dx,
\end{aligned}$$

and hence

$$\begin{aligned}
&\frac{N(u) + N(v)}{2} \\
&= \int h_0(s_1x^2)\frac{h_0((2u - s_1)x^2) + h_0((2v - s_1)x^2)}{2}dx \\
&\geq \int h_0(s_1x^2)\sqrt{h_0((2u - s_1)x^2)h_0((2v - s_1)x^2)}dx \\
&\geq \int h_0(s_1x^2)h_0((u + v - s_1)x^2)dx \\
&= N\left(\frac{u + v}{2}\right).
\end{aligned}$$

Lemma 2.7. If $h(x)$ is uniformly bounded, then for any sequence $b_n \to \infty$, $\frac{q(b_n)}{\sqrt{b_n}}$, $\frac{Q(s_1,b_n)}{\sqrt{s_1 b_n}}$ and $\frac{Q(b_n,b_n)}{b_n}$ converge to zero as $n \to \infty$.

Proof: Follows trivially from the definitions of $q(s)$ and $Q(s,t)$.

Lemma 2.8. Consider the distribution G_n belonging to $\mathcal{G}$ with $G_n\{s_1\} = p_n$, $G_n\{b_n\} = 1 - p_n$, where $b_n = \frac{s_0 - p_n s_1}{1-p_n}$. Then, there exists a sequence $p_n \to 1$ such that $\int\int f(s,t)dG_n(s)d\,G_n(t)$ converges to $\frac{2q(s_1)}{\sqrt{s_1}} + \frac{\alpha Q(s_1,s_1)}{s_1}$ as $n \to \infty$.

Proof: We assume $s_0 > s_1$ for if $s_0 = s_1$, then the result follows on taking $p_n \equiv 1$. Actually, for $s_0 > s_1$ we will prove that Lemma 2.8 holds for any sequence $p_n \to 1$.

By definition of $f(s,t)$,

$$\begin{aligned}
&\int\int f(s,t)dG_n(s)dG_n(t) \\
&= p_n^2\left(\frac{2q(s_1)}{\sqrt{s_1}} + \frac{\alpha Q(s_1,s_1)}{s_1}\right)
\end{aligned}$$

$$+ 2p_n(1-p_n)\left(\frac{q(s_1)}{\sqrt{s_1}} + \frac{q(b_n)}{\sqrt{b_n}} + \frac{\alpha Q(s_1, b_n)}{\sqrt{s_1 b_n}}\right)$$
$$+ (1-p_n)^2\left(\frac{2q(b_n)}{\sqrt{b_n}} + \frac{\alpha Q(b_n, b_n)}{b_n}\right). \tag{2.6}$$

The result now follows from Lemma 2.7.

Proof of Theorem 2.5: Clearly, from (2.3), it follows that it is enough to show that

$$\sup_{\mathcal{G}} \int\int f(s,t)dG(s)dG(t) = \frac{2q(s_1)}{\sqrt{s_1}} + \frac{\alpha Q(s_1, s_1)}{s_1}.$$

This follows on noting that

$$\frac{2q(s_1)}{\sqrt{s_1}} + \frac{\alpha Q(s_1, s_1)}{s_1}$$
$$= \lim_{n\to\infty} \int\int f(s,t)dG_n(s)dG_n(t) \qquad \text{(Lemma 2.8)}$$
$$\leq \sup_{\mathcal{G}} \int\int f(s,t)dG(s)dG(t) \qquad (\because G_n \varepsilon \mathcal{G})$$
$$\leq \sup_{\mathcal{G}} \int\int \left\{\frac{q(s_1)}{\sqrt{s_1}} + \frac{q(s+t-s_1)}{\sqrt{s+t-s_1}} + \frac{\alpha Q(s_1, s+t-s_1)}{\sqrt{s_1(s+t-s_1)}}\right\} dG(s)dG(t)$$

(Lemma 2.3 and the fact that (s,t) is weakly majorized by $(s_1, s+t-s_1)$)

$$= \frac{q(s_1)}{\sqrt{s_1}} + \sup_{\mathcal{G}} \int\int J\left(\frac{s+t}{2}\right) dG(s)dG(t) \qquad \text{(definition of } J(\cdot))$$
$$\leq \frac{q(s_1)}{\sqrt{s_1}} + \sup_{\mathcal{G}} \int J(s)dG(s) \qquad (\because \text{ the distribution of } \tfrac{s+t}{2} \text{ also belongs to } \mathcal{G})$$
$$= \frac{q(s_1)}{\sqrt{s_1}} + \inf\{as_0 + b\colon J(s) \leq as + b \text{ for all } s \geq s_1\}$$

(by a standard moment theorem; see Karlin and Studden (1966))

$$= \frac{q(s_1)}{\sqrt{s_1}} + J(s_1) \qquad \text{(Lemma 2.6)}$$
$$= \frac{2q(s_1)}{\sqrt{s_1}} + \frac{\alpha Q(s_1, s_1)}{s_1} \qquad \text{(definition of } J(\cdot)).$$

This proves the theorem.

Cor. If $H(x) = \Phi(x)$, then

$$\sup_{F\varepsilon\mathcal{F}} e(W, \overline{X}) = 12(1 + (s_0 - 1)\varepsilon_0)\left[\frac{(1-\varepsilon_0)^2}{2\sqrt{\pi}} + \frac{\varepsilon_0(1-\varepsilon_0)}{\sqrt{2\pi}}\left\{\frac{2}{\sqrt{1+s_1}} + \frac{\frac{\varepsilon_0}{1-\varepsilon_0}}{\sqrt{2s_1}}\right\}\right]^2 \tag{2.7}$$

The following table gives the values of $\inf e(W,\overline{X})$ and $\sup e(W,\overline{X})$ for various values of ε_0, s_1 and s_0 when H is the standard normal CDF. The values in parentheses are the suprema.

Table 1: $\inf e(W,\overline{X})(\sup e(W,\overline{X}))$ for contaminations of $N(0,1)$

		ε_0		
s_1	s_0	.01	.05	.1
0	.01	.962(∞)	1.022(∞)	1.143(∞)
	1	.955(∞)	.955(∞)	.955(∞)
	4	.969(∞)	1.020(∞)	1.071(∞)
1	4	.969(.984)	1.020(1.098)	1.071(1.241)
	9	1.009(1.031)	1.196(1.337)	1.373(1.719)
	16	1.067(1.098)	1.463(1.671)	1.825(2.387)
4	9	1.009(1.016)	1.196(1.242)	1.373(1.483)
	16	1.067(1.082)	1.463(1.552)	1.825(2.060)
	25	1.150(1.167)	1.814(1.952)	2.412(2.801)

Notice that in general, unless s_1 is small, the infimum and the supremum of the efficiency are rather close.

Since an explicit formula for $\inf e(W,\overline{X})$ is available, we can also answer the following interesting question: given s_0, for which values of ε_0, is $\inf e(W,\overline{X}) \geq 1$? The following table gives the two values of $\varepsilon_0 = \varepsilon_0(s_0)$ for which the infimum equals 1 for some specific values of s_0. The infimum is 1 or more for ε_0 between these two values.

Table 2: Roots of $\inf e(W,\overline{X}) = 1$ for given s_0

s_0	.01	.1	.25	.5	1.5	2	4	9	16	25
Lower root	.2098	.4367	.6	.7861	.1151	.0507	.0159	.0059	.0031	.0020
Upper root	.9995	.9946	.9829	.9292	.5940	.6978	.7936	.8507	.8791	.8973

Thus, for example, if the variance of the contaminating distribution is thought to be about 4, only about 1.6% contamination is needed for the Hodges–Lehmann estimate to be more efficient than the sample mean.

3. Bounds on $e(M,\overline{X})$. The derivation of the infimum and the supremum of $e(M,\overline{X})$ is technically much easier. Also, it turns out that the infimum and the supremum for different H are multiples of each other, making computations much simpler too.

Theorem 3.1. Let $\mathcal{F}$ be defined as in (1.1) with Φ replaced by H. If H is symmetric and absolutely continuous with a bounded density, then $\inf_{F \epsilon \mathcal{F}} e(M, \overline{X})$ is attained at $F_{\varepsilon_0, s_0}(x) = (1-\varepsilon_0)H(x) + \varepsilon_0 H\left(\frac{x}{\sqrt{s_0}}\right)$.

Proof: Recall that $e(M, \overline{X}) = (2f(0)\sigma_f)^2$ where f denotes the density of F and σ_f^2 denotes the variance of X under f. This simplifies to

$$
\begin{aligned}
&e(M, \overline{X}) \\
&= 4c(1-\varepsilon_0 + s_0\varepsilon_0) \cdot h^2(0)(1-\varepsilon_0+\varepsilon_0 \int \frac{1}{\sqrt{s}} dG(s))^2, \qquad (3.1) \\
&\text{where } c = \int_{-\infty}^{\infty} x^2 h(x) dx.
\end{aligned}
$$

The problem therefore reduces to minimizing $\int \frac{1}{\sqrt{s}} dG(s)$ for G in $\mathcal{G}$. This infimum is attained at the distribution degenerate at s_0 since

$$
\left(E\left(\frac{1}{\sqrt{s}}\right)\right)^2 = \frac{\left(E\left(\frac{1}{\sqrt{s}}\right) \cdot E(\sqrt{s})\right)^2}{(E(\sqrt{s}))^2} \geq \frac{1}{(E(\sqrt{s}))^2} \geq \frac{1}{E(s)} = \frac{1}{s_0},
$$

thus giving the Theorem.

Cor. Let $H = \Phi$ in (1.1); then

$$
\inf_{F \epsilon \mathcal{F}} e(M, \overline{X}) = \frac{2}{\pi}(1 + (s_0 - 1)\varepsilon_0)\left(1 - \varepsilon_0 + \frac{\varepsilon_0}{\sqrt{s_0}}\right)^2. \qquad (3.2)
$$

Cor. For an arbitrary H satisfying the assumptions of Theorem 3.1,

$$
\inf_{F \epsilon \mathcal{F}} e(M, \overline{X}) = 4ch^2(0)(1 + (s_0 - 1)\varepsilon_0)\left(1 - \varepsilon_0 + \frac{\varepsilon_0}{\sqrt{s_0}}\right)^2. \qquad (3.3)
$$

Remark:

1. The corollary above implies that for different H, $\inf_{F \epsilon \mathcal{F}} e(M, \overline{X})$ are multiples of each other.

2. Suppose that each family $\mathcal{F}$ corresponding to a fixed H is considered as a model. Then (3.3) gives an expression for the minimum efficiency of the median with respect to the mean. If we now consider all the models generated by different symmetric unimodal H,

then (3.3) gives the fact that the minimum efficiency of the median with respect to the mean simultaneously over all such models equals $\frac{1}{3}\cdot(1+(s_0-1)\varepsilon_0)\times\left(1-\varepsilon_0+\frac{\varepsilon_0}{\sqrt{s_0}}\right)^2$ since the minimum over the symmetric unimodal densities is $\frac{1}{3}$.

Theorem 3.2. Under the assumptions of Theorem 3.1,

$$\sup_{F\epsilon\mathcal{F}} e(M,\overline{X}) = 4ch^2(0)(1+(s_0-1)\varepsilon_0)\left(1-\varepsilon_0+\frac{\varepsilon_0}{\sqrt{s_1}}\right)^2. \tag{3.4}$$

Proof: Again, from (3.1), we need to find $\sup\int\frac{1}{\sqrt{s}}dG(s)$ over $\mathcal{G}$. That this equals $\frac{1}{\sqrt{s_1}}$ is proved by taking G_n as in Lemma 2.8 and using the inequalities

$$\begin{aligned}
&\frac{1}{\sqrt{s_1}}\\
&\quad=\lim_{n\to\infty}\int\frac{1}{\sqrt{s}}dG_n(s)\\
&\quad\le\sup_{\mathcal{G}}\int\frac{1}{\sqrt{s}}dG(s)\\
&\quad=\inf\left\{as_0+b\colon as+b\ge\frac{1}{\sqrt{s}}\text{ for all } s\ge s_1\right\}\\
&\quad=\frac{1}{\sqrt{s_1}}\left(\because\frac{1}{\sqrt{s}}\text{ is convex in } s\right),
\end{aligned}$$

thus giving the Theorem.

Cor. Let $H=\Phi$ in (1.1); then

$$\sup_{F\epsilon\mathcal{F}} e(M,\overline{X}) = \frac{2}{\pi}(1+(s_0-1)\varepsilon_0)\left(1-\varepsilon_0+\frac{\varepsilon_0}{\sqrt{s_1}}\right)^2. \tag{3.5}$$

Remark:

1. Again, $\sup e(M,\overline{X})$ are multiples of each other for different H.
2. Notice that unlike the infimum, the supremum is not exactly attained at any distribution.

The following table gives the values of $\inf e(M,\overline{X})$ and $\sup e(M,\overline{X})$ for various combinations of ε_0, s_1 and s_0 when H equals Φ. As in Table 1, the values in the parentheses are the suprema.

Table 3: inf $e(M,\overline{X})$(sup $e(M,\overline{X})$) for contaminations of $N(0,1)$

		ε_0		
s_1	s_0	.01	.05	.1
0	.01	.749(∞)	1.272(∞)	2.071(∞)
	1	.634(∞)	.634(∞)	.634(∞)
	4	.649(∞)	.696(∞)	.747(∞)
1	4	.649(.656)	.696(.732)	.747(.828)
	9	.678(.688)	.833(.891)	.998(1.146)
	16	.721(.732)	1.032(1.114)	1.362(1.592)
4	9	.678(.681)	.833(.847)	.998(1.034)
	16	.721(.725)	1.032(1.059)	1.362(1.436)
	25	.777(.782)	1.291(1.331)	1.832(1.953)

Again, inspection of the values in Table 3 show that except when s_1 is small, the infimum and the supremum of $e(M,\overline{X})$ are amazingly close. Also, comparison with Table 1 shows that generally speaking, except when the contamination is of the form of a near point mass at 0 with a moderate amount of contamination, the Hodges–Lehmann estimate is probably to be preferred to the median, although an explicit statement to this effect requires computation of $e(W,M)$.

Again, because of Theorem 3.1, we can find out the smallest $\varepsilon_0 > 0$ required for inf $e(M,\overline{X})$ to be 1 or more for a given s_0. The following table gives the values of ε_0 at which the infimum equals 1 for given s_0. For ε_0 between these values, the infimum is 1 or more.

Table 4: Roots of inf $e(M,\overline{X}) = 1$ for given s_0

s_0	.01	.1	.20	.5	1.5	2	4	9	16	25
Lower root	.03028	.16599	.49992	*	*	*	*	.09800	.04560	.02667
Uppper root	.99406	.91412	.62549	*	*	*	*	.81500	.89285	.92469

(an asterisk (*) represents no root of the equation; this is an indicator of the fact that if s_0 is "close" to 1, then the median stays less efficient than the mean).

4. Bounds on $e(\overline{X}_\alpha, \overline{X})$. The results of sections 2 and 3 imply that the minimum efficiencies of the Hodges–Lehmann estimate and the median are attained at a degenerate mixing

distribution G. However, a different feature of the results in this section is that in general it was only possible to prove that the mixing distribution at which the minimum occurs is supported at at most three points and there are in fact some cases in which the support *has* two points. Usually, however, we found that the support again has only one point in which case the distribution is automatically degenerate at s_0. The technique we apply does not seem easily adaptable to the case of general H; so we restrict ourselves to the case $H = \Phi$ in this section. Notice that the results in Bickel and Lehmann (1975) and also Loh (1984) would imply that for every F in $\mathcal{F}$, $e(\overline{X}_\alpha, \overline{X})$ exceeds the value of this efficiency under Φ because of tail orderings between F and Φ. However, Φ itself is not necessarily a member of $\mathcal{F}$.

Recall that for any F,

$$e(\overline{X}_\alpha, \overline{X}) = \frac{\sigma_f^2}{\sigma_\alpha^2},$$

where σ_f^2 is the variance of x under f and

$$\sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\left[\int_0^{F^{-1}(1-\alpha)} t^2 f(t)dt + \alpha\left\{F^{-1}(1-\alpha)\right\}^2\right]. \tag{4.1}$$

The problem of minimizing $e(\overline{X}_\alpha, \overline{X})$ then is equivalent to maximizing σ_α^2 over F in $\mathcal{F}$, or equivalently over G in $\mathcal{G}$. This can be done by considering the subclass of distributions in $\mathcal{F}$ for which $F^{-1}(1-\alpha)$ equals a fixed constant γ and maximizing $\int_0^\gamma t^2 f(t)dt$ in this subclass. The overall maximum can then be obtained by varying γ in its appropriate range.

The restriction $F^{-1}(1-\alpha) = \gamma$ is equivalent to the restriction

$$\int \Phi\left(\frac{\gamma}{\sqrt{s}}\right) dG(s) = \frac{(1-\alpha) - (1-\varepsilon)\Phi(\gamma)}{\varepsilon}. \tag{4.2}$$

Also, maximizing $\int_0^\gamma t^2 f(t)dt$ is easily shown to be equivalent to minimizing $\int_\gamma^\infty t^2 f(t)dt$ and straightforward calculations using integration by parts give that

$$\int_\gamma^\infty t^2 f(t)dt$$
$$= (1-\varepsilon_0)\left[\frac{1}{\sqrt{2\pi}}\gamma e^{-\frac{\gamma^2}{2}} + 1 - \Phi(\gamma)\right] + \varepsilon_0 \int \left\{\frac{1}{\sqrt{2\pi}}\gamma\sqrt{s}\, e^{-\frac{\gamma^2}{2s}} + s\left(1 - \Phi\left(\frac{\gamma}{\sqrt{s}}\right)\right)\right\} dG(s), \tag{4.3}$$

so that we require to minimize $\int f_2(s)dG(s)$ subject to the restrictions (4.2) and $\int sdG(s) = s_0$ (the restriction that $\int_{[s_1,\infty)} dG(s) = 1$ is automatically imposed by only considering $s \geq s_1$), where

$$f_2(s) = \frac{1}{\sqrt{2\pi}} \gamma\sqrt{s}\, e^{-\frac{\gamma^2}{2s}} - s\Phi\left(\frac{\gamma}{\sqrt{s}}\right). \tag{4.4}$$

We also let $f_1(s) = \Phi\left(\frac{\gamma}{\sqrt{s}}\right)$.

The following preparatory result is needed.

Lemma 4.1. For any real b, the second derivative of the function

$$f_3(s) = \frac{\gamma\sqrt{s}\, e^{-\frac{\gamma^2}{2s}}}{\sqrt{2\pi}} - (s+b)\Phi\left(\frac{\gamma}{\sqrt{s}}\right) \tag{4.5}$$

has at most one sign change.

Proof: First observe that

$$f_3(s) = -[2s^2 f_1'(s) + (s+b)f_1(s)].$$

So it suffices to prove that the second derivative of

$$f_4(s) = -2s^2 f_1'(s) + (s+b)f_1(s) \tag{4.6}$$

has at most one sign change.

Differentiation yields

$$f_4''(s) = 6f_1'(s) + 8sf_1''(s) + (s+b)f_1''(s) + 2s^2 f_1^{(3)}(x). \tag{4.7}$$

Also, $f_1'(s) = -\frac{\gamma}{2s^{3/2}}\phi\left(\frac{\gamma}{\sqrt{s}}\right)$,

$$f_1''(s) = -\frac{\gamma^3}{4s^{7/2}}\phi\left(\frac{\gamma}{\sqrt{s}}\right) + \frac{3\gamma}{4s^{5/2}}\phi\left(\frac{\gamma}{\sqrt{s}}\right),$$

$$\text{and } f_1^{(3)}(s) = -\frac{\gamma^5}{8s^{11/2}}\phi\left(\frac{\gamma}{\sqrt{s}}\right) + \frac{10\gamma^3}{8s^{9/2}}\phi\left(\frac{\gamma}{\sqrt{s}}\right) - \frac{15\gamma}{8s^{7/2}}\phi\left(\frac{\gamma}{\sqrt{s}}\right), \tag{4.8}$$

where $\phi(\cdot)$ denotes the density of the standard normal distribution.

Substitution of (4.8) into (4.7) and algebra now yields

$$f_4''(s) = \frac{1}{s^{7/2}} \cdot \frac{\gamma}{4} s^2[(\gamma^2 + 3b)s - (b\gamma^2 + \gamma^4)]. \tag{4.9}$$

It follows that $f_4''(s)$ has at most one sign change and the desired conclusion follows.

Theorem 4.2. Let $\mathcal{F}$ be defined as in (1.1). Then $\inf e(\overline{X}_\alpha, \overline{X})$ over $\mathcal{F}$ equals $\inf e(\overline{X}_\alpha, \overline{X})$ over the subclass of distributions corresponding to those G in $\mathcal{G}$ which are supported at at most 3 points.

Proof: From the discussion preceding Lemma 4.1, it follows that we need to find $\inf \int f_2(s)dG(s)$ subject to $\int f_1(s)dG(s) = c_1$ (say) and $\int sdG(s) = c_0 = s_0$, where $s \geq s_1$. However, this infimum equals

$$\sup\{bc_1 + ac_0 + c: f_2(s) \geq bf_1(s) + as + c \text{ for all } s \geq s_1\}$$

and furthermore it is sufficient to consider only such G which are supported on contact sets of the form

$$\{s: f_2(s) = bf_1(s) + as + c\}. \tag{4.10}$$

Since $f_2(s) = bf_1(s) + as + c$

$$\Longleftrightarrow f_2(s) - bf_1(s) = as + c$$

$$\Longleftrightarrow f_3(s) = as + c,$$

it follows from Lemma (4.1) that the contact set (4.10) can have at most 3 points, which proves the theorem.

Since Theorem 4.2 is less explicit than Theorem 2.4 and Theorem 3.1, the calculation of $\inf e(\overline{X}_\alpha, \overline{X})$ has to be done numerically, by minimizing $e(\overline{X}_\alpha, \overline{X}) = \frac{\sigma_f^2}{\sigma_\alpha^2}$ over 3 point distributions in $\mathcal{G}$. This is done by taking s_2, s_3, s_4 and a probability vector p_2, p_3, p_4 such that $\sum_{i=2}^{4} p_i s_i = s_0$ and solving for γ satisfying $F^{-1}(1-\alpha) = \gamma$, which is equivalent to

$$\sum_{i=2}^{4} p_i \Phi\left(\frac{\gamma}{\sqrt{s_i}}\right) = \frac{1 - \alpha - (1-\varepsilon_0)\Phi(\gamma)}{\varepsilon_0}. \tag{4.11}$$

One then minimizes

$$e(\overline{X}_\alpha, \overline{X}) = \frac{1+(s_0-1)\varepsilon_0}{\int_0^\gamma t^2 f(t)dt + \alpha\gamma^2} \cdot \frac{(1-2\alpha)^2}{2}, \tag{4.12}$$

where $\int_0^\gamma t^2 f(t)dt$

$$= \int_0^\infty t^2 f(t)dt - \int_\gamma^\infty t^2 f(t)dt$$
$$= \frac{1+(s_0-1)\varepsilon_0}{2} - \int_\gamma^\infty t^2 f(t)dt, \tag{4.13}$$

where $\int_\gamma^\infty t^2 f(t)$ is given by (4.3) with G described as above.

Even though Theorem 4.2 only enables us to assert that the mixing distribution G is supported at at most three points, in the examples we considered the mixing distribution turned out to be either two point or one point.

Following exactly the same arguments of Theorem 4.2, one can also prove the following Theorem. We omit the proof.

Theorem 4.3. Let $\mathcal{F}$ be defined as in (1.1). Then $\sup e(\overline{X}_\alpha, \overline{X})$ over $\mathcal{F}$ equals $\sup e(\overline{X}_\alpha, \overline{X})$ over the subclass of distributions corresponding to those G in $\mathcal{G}$ which are supported at at most 3 points.

The numerical project for evaluating the supremum is exactly the same as before, except one now maximizes (4.12) using (4.11), (4.13) and (4.3). Again, the mixing distribution corresponding to the supremum turned out to be at most two point in all the examples that we considered. The following tables give the values of $\inf e(\overline{X}_\alpha, \overline{X})$ and $\sup e(\overline{X}_\alpha, \overline{X})$ for various values of α, ε_0, s_1 and s_0. Again, the values in parentheses are the values of the suprema.

Table 5: $\inf e(\overline{X}_\alpha, \overline{X})(\sup e(\overline{X}_\alpha, \overline{X}))$ for contaminations of $N(0,1)$: $\varepsilon_0 = .01$

s_1	s_0	α = .01	.05	.1	.125	.25	.375
1.1	1.5	.996(.998)	.975(.977)	.944(.946)	.927(.929)	.838(.840)	.741(.742)
	2	.997(1.003)	.977(.982)	.946(.951)	.929(.934)	.840(.844)	.743(.746)
5	8	1.033(1.040)	1.020(1.024)	.991(.993)	.974(.977)	.882(.884)	.781(.782)
	10	1.049(1.059)	1.037(1.043)	1.008(1.012)	.991(.995)	.898(.901)	.795(.797)
9	15	1.091(1.097)	1.082(1.086)	1.052(1.054)	1.035(1.037)	.938(.940)	.830(.832)
	20	1.134(1.145)	1.127(1.133)	1.097(1.100)	1.079(1.082)	.978(.981)	.866(.868)
16	20	1.134(1.137)	1.127(1.128)	1.697(1.697)	1.079(1.08)	.978(.979)	.866(.867)
	25	1.179(1.184)	1.173(1.175)	1.142(1.143)	1.123(1.125)	1.019(1.02)	.902(.903)
25	30	1.223(1.226)	1.219(1.220)	1.187(1.187)	1.168(1.168)	1.059(1.060)	.938(.938)
	40	1.315(1.320)	1.312(1.314)	1.278(1.279)	1.257(1.259)	1.141(1.142)	1.010(1.011)

Table 6: $\inf e(\overline{X}_\alpha, \overline{X})(\sup e(\overline{X}_\alpha, \overline{X}))$ for contaminations of $N(0,1)$: $\varepsilon_0 = .05$

		α					
s_1	s_0	.01	.05	.1	.125	.25	.375
	1.5	.997(1.008)	.977(.989)	.947(.958)	.931(.941)	.841(.851)	.744(.752)
1.1	2	1.001(1.030)	.985(1.011)	.956(.980)	.940(.963)	.852(.871)	.754(.771)
	8	1.128(1.177)	1.175(1.198)	1.156(1.176)	1.145(1.160)	1.05(1.06)	.933(.941)
5	10	1.175(1.262)	1.249(1.286)	1.236(1.263)	1.222(1.246)	1.122(1.138)	.999(1.011)
	15	1.285(1.381)	1.44(1.466)	1.433(1.451)	1.418(1.434)	1.306(1.317)	1.163(1.172)
9	20	1.372(1.582)	1.633(1.681)	1.632(1.664)	1.617(1.645)	1.492(1.51)	1.329(1.343)
	20	1.372(1.429)	1.633(1.643)	1.632(1.638)	1.617(1.622)	1.492(1.495)	1.329(1.332)
16	25	1.441(1.605)	1.829(1.853)	1.832(1.848)	1.816(1.83)	1.678(1.687)	1.496(1.503)
	30	1.500(1.569)	2.025(2.033)	2.032(2.038)	2.016(2.020)	1.864(1.867)	1.663(1.665)
25	40	1.597(1.873)	2.418(2.446)	2.434(2.452)	2.416(2.431)	2.238(2.248)	1.997(2.004)

Table 7: $\inf e(\overline{X}_\alpha, \overline{X})(\sup e(\overline{X}_\alpha, \overline{X}))$ for contaminations of $N(0,1)$: $\varepsilon_0 = .15$

		α					
s_1	s_0	.01	.05	.1	.125	.25	.375
	1.5	.999(1.03)	.982(1.016)	.953(.986)	.937(.969)	.849(.878)	.752(.777)
1.1	2	1.007(1.094)	1.000(1.082)	.976(1.052)	.961(1.034)	.875(.937)	.777(.829)
	8	1.103(1.302)	1.376(1.483)	1.436(1.509)	1.439(1.503)	1.365(1.408)	1.23(1.263)
5	10	1.118(1.482)	1.508(1.697)	1.603(1.728)	1.612(1.722)	1.541(1.613)	1.392(1.477)
	15	1.144(1.457)	1.825(1.998)	2.023(2.121)	2.048(2.132)	1.983(2.037)	1.798(1.839)
9	20	1.160(1.791)	2.120(2.477)	2.442(2.632)	2.483(2.646)	2.427(2.528)	2.206(2.283)
	20	1.160(1.275)	2.120(2.201)	2.442(2.480)	2.483(2.516)	2.427(2.447)	2.206(2.221)
16	25	1.172(1.493)	2.391(2.621)	2.857(2.960)	2.918(3.003)	2.87(2.921)	2.613(2.652)
	30	1.180(1.276)	2.640(2.736)	3.271(3.307)	3.350(3.379)	3.312(3.329)	3.020(3.033)
25	40	1.191(1.584)	3.069(3.484)	4.092(4.227)	4.211(4.321)	4.194(4.260)	3.833(3.881)

The most striking feature of these tables is the proximity of the infimum and the supremum. If the expected degree of contamination is small, then a small amount of trimming seems to be optimal; on the other hand, a moderate amount of trimming (10 to 15%) is optimal if the expected degree of contamination is larger. Also, for any trimming proportion, the trimmed mean gets more efficient as the average variance of the contamination distribution increases.

Table 8 gives the values of ε for which $\inf e(\overline{X}_\alpha, \overline{X})$ exceeds 1 for various values of α and s_0 (although we cannot say it conclusively, in all the examples we looked at the numerical value of $\inf e(\overline{X}_\alpha, \overline{X})$ did not depend on s_1 and only depended on ε_0 and s_0; one may thus conjecture that as in the cases of the Hodges–Lehmann estimate and the median, for the trimmed mean also $\inf e(\overline{X}_\alpha, \overline{X})$ is only a function of ε_0 and s_0; Table 8 is correct subject to the numerical accuracy of our computations).

Table 8: Roots of $\inf e(\overline{X}_\alpha, \overline{X}) = 1$ for given s_0

α		s_0: .01	.1	2	3	4	9	16	25
	Lower root	.10012	.11042	.03438	.00822	.00391	.00083	.00036	.00020
.01	Uppper root	.99996	.99929	.78161	.84876	.86668	.88815	.89411	.89666
	Lower root	.11957	.13239	.14881	.03644	.01862	.00464	.00214	.00125
.05	Uppper root	.99973	.99601	.63866	.80404	.83617	.86583	.87327	.87643
	Lower root	.11634	.13022	*	.09897	.05044	.01316	.00623	.00368
.125	Uppper root	.99917	.98862	*	.74837	.81542	.86758	.87613	.87948
	Lower root	.08812	.11338	*	.29647	.12565	.03169	.01508	.00895
.25	Uppper root	.99798	.97257	*	.56306	.75854	.88096	.90179	.90753
	Lower root	.04870	.11868	*	*	.27904	.05832	.02739	.01620
.375	Uppper root	.99635	.94971	*	*	.60568	.86999	.91539	.93320

Note: an asterisk (*) signifies that for this s_0, $e(\overline{X}_\alpha, \overline{X})$ remains below 1 for all ε_0; this is not unexpected since for s_0 close to 1, F is close to Φ and therefore $e(\overline{X}_\alpha, \overline{X})$ is expected to be less than 1.

5. General discussion and comparisons. The results in sections 2 and 3 imply that the contaminated normal distribution $F_{\varepsilon_0,s_0}(x) = (1-\varepsilon_0)\Phi(x) + \varepsilon_0\Phi\left(\frac{x}{\sqrt{s_0}}\right)$ has an extreme point property in the family of more general mixture distributions (1.1). Also, this result holds for a CDF H significantly more general than Φ. For the trimmed mean, however, this extreme point property does not necessarily hold since we did find a few cases in which the mixing distribution G for which $\inf e(\overline{X}_\alpha, \overline{X})$ is attained is two point. The supremum was always attained at a two point distribution.

A comparison of Tables 1, 3, 5, 6 and 7 indicates that in general the trimmed mean with a moderate amount of trimming seems to provide the best overall protection. For example, if s_0 is 9, $\inf e(W, \overline{X})$ varies between 1.009 and 1.373, $\inf e(M, \overline{X})$ varies between .678 and .998, while $\inf e(\overline{X}_\alpha, \overline{X})$ varies approximately between .97 and 1.6 if α is about .1 and s_0 is between 8 and 10.

A second important feature of the results in sections 2, 3, and 4 is that over the entire family of distributions $\mathcal{F}$, the efficiency of each of W, M and $\overline{X}_\alpha$ remains amazingly stable for practically all combinations of ε_0, s_1 and s_0.

Deviation from normality in real data often occur in the form of a slight skewness. Asymmetric contaminations provide natural avenues for addressing this problem. See Jaeckel (1971), Collins (1976), Andrews et al (1972) for some related results.

The restriction that $\int s dG(s)$ equals $s_0 < \infty$ keeps the variance of F in control. This may be viewed as somewhat restrictive. But our results are intended to apply to precisely these situations where the center as well the tail are controlled. There may be other ways to say that s_0 is a guess for the value of s, like a median restriction instead of a mean restriction. Conceptually, the moment techniques still apply. But the derivation of the infimum and the supremum is harder now.

Acknowledgement. I am deeply grateful to S. Basu, S. Mukhopadhyay and M.M. Zen for the numerical work in the paper. Conversations with J.K. Ghosh and Bill Studden were immensely useful at various stages of the work. I am also indebted to the Co-editor who gave many more pertinent references and helped put the results in perspective.

References

1. Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972). Robust estimates of location: survey and advances, Princeton University Press, Princeton, New Jersey.
2. Bickel, P.J. (1965). On some robust estimates of location, *Ann. Math. Stat.* **36**, 847–858.
3. Bickel, P.J. and Lehmann, E.L. (1975). Descriptive statistics for nonparametric models II, *Ann. Stat.* **3**, 1045–1069.
4. Chernoff, H. and Savage, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics, *Ann. Math. Stat.* **29**, 972–994.
5. Collins, J.R. (1976). Robust estimation of a location parameter in the presence of asymmetry, *Ann. Stat.* **4**, 68–85.
6. Efron, B. and Olshen, R.A. (1978). How broad is the class of normal scale mixtures, *Ann. Stat.* **6**, 1159–1164.
7. Hampel, F.R., Rousseeuw, P.J., Ronchetti, E.M. and Stahel, W.A. (1986). Robust statistics: the approach based on influence functions.
8. Hodges, J.L. (Jr.) and Lehmann, E.L. (1963). Estimates of locations based on rank tests, *Ann. Math. Stat.* **34**, 598–611.
9. Huber, P.J. (1973). Robust Statistics, John Wiley, New York.
10. Jaeckel, L.A. (1971). Robust estimates of location: symmetric and asymmetric contaminations, *Ann. Math. Stat.* **42**, 1020–1034.
11. Karlin, S. and Studden, W.J. (1966). Tchebycheff systems: with applications in analysis and statistics, Wiley Interscience, New York.
12. Lehmann, E.L. (1983). Theory of Point Estimation, John Wiley, New York.
13. Loh, W.L. (1984). Bounds on AREs for restricted classes of distributions defined via tail orderings, *Ann. Stat.* **12**, 685–701.
14. Marshall, A.W. and Olkin, I. (1979). Inequalities: Theory of majorization and its applications, Academic Press, New York.
15. Rieder, H. (1980). Estimates derived from robust tests, *Ann. Statist.*, **8**, 106–115.
16. Rieder, H. (1981). Robustness of rank tests against gross errors, *Ann. Statist.*, **9**, 245–265.
17. Rudin, W. (1973). Functional Analysis, McGraw Hill, New York.
18. Sen P.K. (1968). On a further robustness property of the test and estimator based on Wilcoxon's signed rank statistic, *Ann. Math. Stat.*, **39**, 282–285.
19. Sen, P.K. (1968). Robustness of some nonparametric procedures in Linear models,

Ann. Math. Stat., **39**, 1913–1922.

20. Tukey, J.W. (1960). A survey of sampling from contaminated distributions, In Contributions to Probability and Statistics, Ed. I. Olkin, Stanford University Press, Stanford, California.

21. Wilcoxon (1949). *Some rapid approximate statistical procedures*, American Cyanamid Co., Stamford, Connecticut.