A NEW GENERAL INTERPRETATION OF THE STEIN ESTIMATE AND HOW IT ADAPTS: WITH APPLICATIONS

by

Anirban DasGupta*
Purdue University and
University of California
San Diego

and

Bimal K. Sinha
University of Maryland
Baltimore County

Technical Report # 97-22

Department of Statistics
Purdue University
West Lafayette, IN USA

November, 1997

*Research supported by NSA grant MSP 96G187

△

A NEW GENERAL INTERPRETATION OF THE STEIN ESTIMATE AND HOW IT ADAPTS: WITH APPLICATIONS

Anirban DasGupta*
Department of Statistics
Purdue University and
Department of Mathematics
University of California, San Diego

Bimal K. Sinha
Department of Mathematics & Statistics
University of Maryland Baltimore County

Abstract

In this article, we demonstrate that Stein estimators of the form $[1 - a/(b + ||\mathbf{X}||^2)]\mathbf{X}$ arise under a broad variety of location parameter problems from a natural calculation. The assumptions are limited; one does not need independence of the coordinates, nor symmetry, and the loss function is quite general, including the important absolute error loss.

Our method of calculation automatically shows how the Stein estimate could be adapted to a specific parent distribution and a specific loss. A number of examples that illustrate this adaptive calculation result in quite remarkable features. In particular, some very specific and *new* Stein estimates emerge in important cases, specifically for absolute error loss.

The estimates are studied with respect to their risk, both theoretically and via simulation, and the evidence suggests that minimaxity can be expected in generality in 4 or more dimensions. In addition, the specific proposed estimate appears to outperform both $\mathbf{X}$ and the ordinary James-Stein estimate under absolute error loss for the normal case.

These results are different in character from classic previous works of Brown, Casella, Strawderman, and Shinozaki in several ways.

* Research supported by NSA grant MSP 96G187.

Key Words: Absolute error loss, Extreme value distribution, James-Stein estimate, location parameter, risk, Schwartz's inequality, spherically symmetric, squared error loss, weak law of large numbers.

1. Introduction. Let $\mathbf{X} : p \times 1$ be normally distributed with a mean vector θ and covariance matrix $\sigma^2 \mathbf{I}_p$, $\sigma^2 > 0$. It is well known that for estimating θ under the squared error loss

$$L(\delta, \theta) = ||\delta - \theta||^2 = \sum_{i=1}^{p} (\delta_i - \theta_i)^2, \tag{1.1}$$

$\mathbf{X}$ is *inadmissible* for $p \geq 3$. James and Stein (1961) discovered that an estimator which dominates $\mathbf{X}$ is given by

$$\hat{\theta}_{JS} = [1 - \frac{a}{b + ||\mathbf{X}||^2)}]\mathbf{X} \tag{1.2}$$

where $a > 0$, $b \geq 0$ are suitably chosen constants. Since the pioneering work of James and Stein (1961), a huge amount of work has been done on many aspects of the problem of estimation of a normal mean vector with either a known or an unknown covariance matrix, including Bayesian justifications of $\hat{\theta}_{JS}$; see in particular the review artiles of Brandwein and Strawderman (1990) and Stigler (1990), and also Stein (1981), Berger (1986), Gupta and Pena (1991), and Lehmann and Casella (1997).

In this paper we demonstrate that the typical Stein estimator of θ, namely, $\hat{\theta}_{JS}$ given in (1.2), which is primarily derived under the assumption of a normal distribution and squared error loss, arises naturally as an approximation to a locally best estimate of the parameter θ for a variety of other distributions under a quite general loss function, which includes the *absolute error* loss as a special case. The parameter vector θ in these distributions can be a general location parameter. This can be viewed as a *robustness* property of the James-Stein estimator in (1.2). Furthermore, we show how the estimate can be adapted to the loss function and the parent distribution. Note that Stein estimators indeed have been proposed for general location parameters in the literature; Brown (1966) is a significant reference. But the calculations we present below have an entirely different character.

The distributions of X we have considered in this paper are purely nonparametric except for some regularity conditions. In Section 2, we first discuss the case of a univariate normal distribution and a double exponential distribution to motivate our results. These serve as illustrative examples. The case of a multivariate normal distribution, including covariance matrices of the form $\sigma^2 \mathbf{I}_p$ with an independent estimate s^2 of σ^2 being available, is treated in Section 3. In Section 4 we consider an arbitrary location parameter distribution of $\mathbf{X}$. A general power loss is considered in Section 5, and some concluding remarks are made in Section 6.

The principal contributions of this article are the following.

1. We demonstrate, by a seemingly interesting and simple calculation,

that Stein estimators arise naturally under a broad spectrum of parent distributions and loss functions.

2. We show how adaptation to the parent distribution and the loss function occurs automatically in our method of calculation; in particular, we can handle median estimation for skewed distributions and we can also handle absolute error loss.

3. We show a particular *ridge* estimate $[1 - \frac{p}{p+||\mathbf{X}||^2}]\mathbf{X}$ to be especially interesting, and we also give a nice closed form estimate for the spherically symmetric case under the important absolute error loss. The estimate $[1 - \frac{p}{p+||\mathbf{X}||^2}]\mathbf{X}$ emerges as the natural Stein estimate in the normal case with an identity covariance matrix for an arbitrary power loss $\sum_{i=1}^{p} |\delta_i - \theta_i|^\alpha$ for any $\alpha \geq 1$. An exactly similar result also holds in the case of $\sigma^2\mathbf{I}_p$ with unknown σ^2 as the covariance matrix. This is quite interesting.

4. We give many encouraging risk calculations and simulations to support the proposed estimates. In fact, the estimates emerging from our calculations appear to be minimax for $p \geq 4$ in a broad variety of situations, covering skewed distributions and the absolute error loss. Literature on the important absolute error loss in particular is very limited. In all these respects, our results differ from Brown (1966), Casella (1980) and Shinozaki (1984).

5. Due to the canonical nature of our results, they apply to many problems, *e.g.*, the important evolving area of regression under L_1 norm.

2. Two Illustrative Examples.

2.1. Univariate normal: Example 1.

Suppose X is normal with mean θ and variance one, and consider the problem of estimating θ under the absolute error loss

$$L(\delta, \theta) = |\delta - \theta| \tag{2.1}$$

by linear estimates of the form $\delta = cX$. If $|c| > 1$, X dominates cX. Indeed, for $|c| > 1$,

$$\begin{aligned} E_\theta|cX - \theta| &= |c|.E_\theta|X - \frac{\theta}{c}| \\ &> |c|.E_\theta|X - \theta| \\ &> E_\theta|X - \theta|. \end{aligned} \tag{2.2}$$

In view of this observation, we only consider c such that $0 < c \leq 1$. It is interesting to investigate what would be an *optimal* choice of c, if any, under the above loss. Of course, if $\theta = 0$, the optimal c is equal to 1. Writing $c = \frac{1}{1+\lambda}$ with $\lambda \geq 0$, and noting that for $Z \sim N(0,1)$, and any constant a,

$$\begin{aligned} E|Z-a| &= \int_a^\infty (z-a)\phi(z)dz + \int_{-\infty}^a (a-z)\phi(z)dz \\ &= 2\phi(a) - a(1-\Phi(a)) + a\Phi(a) \\ &= 2\{\phi(a) + a\Phi(a)\} - a, \end{aligned} \tag{2.3}$$

where $\phi(.)$ and $\Phi(.)$ denote the standard normal *pdf* and *cdf*, respectively, we have,

$$\begin{aligned} E|\frac{X}{1+\lambda} - \theta| &= \frac{1}{1+\lambda}E|Z - \lambda\theta| \\ &= \frac{1}{1+\lambda}[2(\phi(\lambda\theta) + \lambda\theta\Phi(\lambda\theta)) - \lambda\theta] \\ &= f(\lambda|\theta), say. \end{aligned} \tag{2.4}$$

It is clear from (2.4) that the optimum λ depends on θ. Writing $u = \lambda\theta$, a direct minimization of $f(\lambda|\theta)$ with respect to u for fixed θ shows that, for $\theta \neq 0$, the minimum obtains at u_0 which is a solution of

$$\begin{aligned} \theta\Phi(u_0) &= \frac{\theta}{2} + \phi(u_0) \\ \longleftrightarrow \theta\Phi(u_0) - \phi(u_0) - \frac{\theta}{2} &= 0. \end{aligned} \tag{2.5}$$

A formal first order Taylor expansion of $\theta\Phi(u_0) - \phi(u_0) - \frac{\theta}{2}$ around $u_0 = 0$ gives $u_0 \sim \frac{1}{\theta}$, and hence an approximation to the optimum λ as $\lambda_0 \sim \frac{1}{\theta^2}$. This results in the estimate

$$\begin{aligned} \hat{\theta} &= \frac{1}{1+\frac{1}{\theta^2}}X \\ &= (1 - \frac{1}{1+\theta^2})X. \end{aligned} \tag{2.6}$$

Substituting X for θ in (2.6) produces the estimate

$$\hat{\theta} = (1 - \frac{1}{1+X^2})X, \tag{2.7}$$

a James-Stein type estimate in one dimension, starting with absolute error loss. This is a special case of the estimate

$$\hat{\theta} = [1 - \frac{p}{p + ||\mathbf{X}||^2}]\mathbf{X}, \tag{2.8}$$

in general p dimensions, which, as we shall see later, arises naturally in many interesting situations.

2.2. Double exponential: Example 2.

Suppose next that X follows a double exponential distribution with the *pdf* $f(x|\theta)$ given by

$$f(x|\theta) = \frac{1}{2}e^{-|x-\theta|}, \quad -\infty < x, \theta < \infty. \tag{2.9}$$

Consider again the problem of estimation of θ by $\delta(X) = \frac{X}{1+\lambda}$ under the absolute error loss given by (2.1). Noting that

$$E|Z - a| = \frac{1}{2}(e^a + e^{-a}) \tag{2.10}$$

where $Z = X - \theta$ follows a *standard* double exponential distribution, we immediately have

$$E|\frac{X}{1+\lambda} - \theta| = \frac{1}{2(1+\lambda)}(e^{\lambda\theta} + e^{-\lambda\theta}). \tag{2.11}$$

As in Example 1, minimization of (2.11) with respect to λ for fixed $\theta \neq 0$ results in the derivative equation for $u = \lambda\theta$:

$$(\theta + u_0)(e^{u_0} - e^{-u_0}) - (e^{u_0} + e^{-u_0}) = 0. \tag{2.12}$$

A first order Taylor expansion now gives $u_0 \sim \frac{1}{\theta}$, resulting in the estimate $\hat{\theta} = (1 - \frac{1}{1+\theta^2})X$. Substituting X for θ results in the *same* estimate obtained in Example 1, *i.e.*,

$$\hat{\theta} = [1 - \frac{1}{1 + X^2}]X. \tag{2.13}$$

3. Multivariate normal.

In this section we consider the case when $\mathbf{X} : p \times 1$ follows a multivariate normal distribution with mean vector θ and an identity covariance matrix for $p > 1$. Consider the problem of estimation of θ by $c\mathbf{X}$ under the absolute error loss $L(\delta, \theta)$ given by

$$L(\delta,\theta) = ||\delta - \theta||_1 = \sum_{i=1}^{p} |\delta_i - \theta_i|. \tag{3.1}$$

As in Example 1, we can assume that $0 < c \leq 1$ and write $c = \frac{1}{1+\lambda}$ for some $\lambda \geq 0$. Using (2.4), we can write

$$\begin{aligned} E||\frac{\mathbf{X}}{1+\lambda} - \theta||_1 &= \frac{1}{1+\lambda}\sum_{i=1}^{p}[2\phi(\lambda\theta_i) + 2\lambda\theta_i\Phi(\lambda\theta_i) - \lambda\theta_i] \\ &= \frac{1}{1+\lambda}[2\sum_{i=1}^{p}\phi(\lambda\theta_i) + 2\lambda\sum_{i=1}^{p}\theta_i\Phi(\lambda\theta_i) - \lambda\sum_{i=1}^{p}\theta_i]. \end{aligned} \tag{3.2}$$

Again, if $\theta = \mathbf{0}$, the optimal choice of λ is $\lambda = 0$. Minimization of the above with respect to λ for fixed $\theta \neq \mathbf{0}$ results in the derivative equation

$$f(\lambda|\theta) = \sum_{i=1}^{p}\theta_i\{\Phi(\lambda\theta_i) - \frac{1}{2}\} - \sum_{i=1}^{p}\phi(\lambda\theta_i) = 0. \tag{3.3}$$

Since $f(0|\theta) = -p\phi(0) < 0$ and $f'(\lambda|\theta) = (1+\lambda)\sum_{i=1}^{p}\theta_i^2\phi(\lambda|\theta_i) > 0$ for $\theta \neq \mathbf{0}$, it follows that there is a unique root of $f(\lambda|\theta) = 0$ for any given $\theta \neq \mathbf{0}$. A *first order* approximation of the true root $\lambda_0(\theta)$ can be obtained by expanding $f(\lambda|\theta)$ around $\mathbf{0}$, thus resulting in

$$\begin{aligned} 0 &= f(\lambda_0|\theta) \\ &\sim f(0|\theta) + \lambda_0 f'(0|\theta) \\ &= -p\phi(0) + \lambda_0\phi(0)||\theta||^2 \end{aligned} \tag{3.4}$$

so that $\lambda_0(\theta) \sim \frac{p}{||\theta||^2}$. Replacing $||\theta||^2$ by $||\mathbf{X}||^2$, we end up with the estimator

$$\hat{\theta} = \frac{\mathbf{X}}{1 + \frac{p}{||\mathbf{X}||^2}} = [1 - \frac{p}{p + ||\mathbf{X}||^2}]\mathbf{X} \tag{3.5}$$

which resembles the well known James-Stein estimator of θ.

Remark 3.1. First notice that the estimator in (3.5) does *not* have the usual disadvantage of the traditional James-Stein estimator $\hat{\theta}_{JS} = [1 - \frac{p-2}{||\mathbf{X}||^2}]\mathbf{X}$ of being a ridiculous estimate of θ for $\mathbf{X}$ near $\mathbf{0}$. In other words, the positive part of the estimator in (3.5) is the estimator itself. Figure 1 gives a one dimensional plot of the mean absolute error of both estimates for $p = 4$ and clearly the estimate in (3.5) does better. Thus, we believe our calculations here can serve as a stepping stone to deriving an explicit estimate better than $\mathbf{X}$ under absolute error loss.

FIGURE 1 here

Next, although (3.5) was derived under the assumption of an absolute error loss, it is interesting to explore its performance under squared error loss. A direct evaluation of its unbiased estimate of risk (Stein (1981)) gives

$$E||\hat{\theta}-\theta||^2 = p + E[\frac{p(4-p)-2p||\mathbf{X}||^2}{(||\mathbf{X}||^2+p)^2}]. \tag{3.6}$$

(3.6) immediately shows that $\hat{\theta}$ is *minimax* under squared error loss for $p \geq 4$. It also follows from Theorem 5.1 in Casella (1980) that $\hat{\theta}$ is not minimax for $p = 3$. Table 1 below gives the mean squared error of $\hat{\theta}$ for some selected values of $||\theta||$ for $p = 3$ and 4. Note that even for $p = 3$, $\hat{\theta}$ has a smaller risk than $\mathbf{X}$ for $||\theta|| \leq 50$.

Table 1. Mean Squared Error of $\hat{\theta}$: normal

$\|\|\theta\|\|$	$p = 3$	$p = 4$
0	2.705	3.561
.25	2.701	3.562
.5	2.692	3.561
.75	2.682	3.561
1	2.670	3.562
2	2.651	3.574
3	2.691	3.629
4	2.757	3.704
5	2.816	3.772
10	2.944	3.927
50	2.998	3.997

Remark 3.2. We have compared the risks of the proposed *new* estimator of θ given in (3.5) with the traditional James-Stein estimator of θ given by $\hat{\theta}_{JS} = [1 - \frac{p-2}{||\mathbf{X}||^2}]\mathbf{X}$ under squared error loss, and found that in general neither dominates the other.

Remark 3.3. We have also looked into a *second order* approximation of the true solution $\lambda_0(\theta)$ of $f(\lambda_0|\theta) = 0$ by keeping terms up to the quadratic in λ_0 in (3.4). However, our numerical computations do not show any advantage in doing this extra work. In fact, a first order approximation can give a smaller risk at $\theta = \mathbf{0}$ and also a more aesthetically pleasing form for the estimator itself.

Remark 3.4. In the case of the covariance matrix of $\mathbf{X}$ being $\sigma^2\mathbf{I}_p$ with σ^2 unknown, and an independent estimate s^2 of σ^2 with ν *df* being available, it follows readily from (3.2)-(3.4) that $\lambda_0(\theta) \sim \frac{p\sigma^2}{||\theta||^2}$. Replacing $||\theta||^2$ by $||\mathbf{X}||^2$ and σ^2 by s^2/ν, we immediately arrive at the following estimate of θ:

$$\hat{\theta} = (1 - \frac{ps^2}{ps^2 + \nu||\mathbf{X}||^2})\mathbf{X} \tag{3.7}$$

which is a wellknown version of the James-Stein estimate in this case.

4. General median estimation.

4.1. Derivation of the estimate

In this section we address the general situation when $\mathbf{X} : p \times 1$ follows an arbitrary location family distribution with a joint *pdf* $f(\mathbf{x}|\theta)$ and the marginal of X_i is denoted as $f_i(x_i|\theta_i)$. We assume that θ_i is the median of X_i for all i. Note that we do *not* need to assume that the X_i's are independent; our calculations are general but the independence case is automatically covered.

Consider the problem of estimation of θ by $\frac{\mathbf{X}}{1+\lambda}$, $\lambda \geq 0$, under the absolute error loss $L(\delta, \theta)$ given by (3.1). Write $Z_i = X_i - \theta_i$ so that $Z_i \sim f_i(z)$ with $F_i(z)$ as the *cdf*. Thus, $F_i(0) = 1/2$ for all i. We *further* assume that $f_i(0) > 0$ for all i. Incidentally, this excludes the so-called power family of distributions with densities of the form $f(z) \sim e^{-|z|^b}|z|^c$ with $b, c > 0$. Then, by Fubini, for any real a,

$$\begin{aligned} E|Z_i - a| &= \int_a^{\infty}(z-a)f_i(z)dz + \int_{-\infty}^{a}(a-z)f_i(z)dz \\ &= \int_a^{\infty}(1 - F_i(z))dz + \int_{-\infty}^{a}F_i(z)dz \\ &= h_i(a) + g_i(-a), say, \end{aligned} \tag{4.1}$$

where

$$h_i(a) = \int_a^{\infty}\{1 - F_i(z)\}dz, \;\; g_i(a) = \int_{-\infty}^{-a}F_i(z)dz. \tag{4.2}$$

Thus, we can write

$$\begin{aligned} E||\frac{\mathbf{X}}{1+\lambda} - \theta||_1 &= \frac{1}{1+\lambda}E[\sum_{i=1}^{p}|Z_i - \lambda\theta_i|] \\ &= \frac{1}{1+\lambda}\sum_{i=1}^{p}\{h_i(\lambda\theta_i) + g_i(-\lambda\theta_i)\}. \end{aligned} \tag{4.3}$$

Since $h_i'(a) = F_i(a) - 1$ and $g_i'(a) = -F_i(-a)$, minimization of $E||\frac{\mathbf{X}}{1+\lambda} - \theta||_1$ with respect to λ for fixed θ leads to the derivative equation :

$$(1+\lambda)[\sum_{i=1}^{p}\{2\theta_i F_i(\lambda\theta_i)-\theta_i\}]-\sum_{i=1}^{p} g_i(-\lambda\theta_i)-\sum_{i=1}^{p} h_i(\lambda\theta_i)=0. \tag{4.4}$$

Denoting the LHS of (4.4) by $\psi(\lambda|\theta)$, we easily see that

$$\psi(0|\theta)=-\sum_{i=1}^{p} g_i(0)-\sum_{i=1}^{p} h_i(0)<0 \tag{4.5}$$

$$\psi'(0|\theta)=2\sum_{i=1}^{p}\theta_i^2 f_i(0) \tag{4.6}$$

$$\psi'(\lambda|\theta)=2(1+\lambda)\sum_{i=1}^{p}\theta_i^2 f_i(\lambda\theta_i)>0. \tag{4.7}$$

Thus there is a unique root $\lambda_0(\theta)$ of $\psi(\lambda|\theta)=0$, and its first order approximation is given by

$$\begin{aligned}\lambda_0(\theta) &\sim -\frac{\psi(0|\theta)}{\psi'(0|\theta)}\\ &= \frac{\sum_{i=1}^{p} g_i(0)+\sum_{i=1}^{p} h_i(0)}{2\sum_{i=1}^{p}\theta_i^2 f_i(0)}.\end{aligned} \tag{4.8}$$

Using (4.2), one has, for each i,

$$g_i(0)+h_i(0)=E|Z_i|. \tag{4.9}$$

On using (4.9) and substituting $\mathbf{X}$ for θ in (4.8), one gets the estimate

$$\hat{\theta}=[1-\frac{\sum_{i=1}^{p}\frac{E|Z_i|}{2}}{\sum_{i=1}^{p}\frac{E|Z_i|}{2}+\sum_{i=1}^{p}X_i^2 f_i(0)}].\mathbf{X}. \tag{4.10}$$

In particular, in case of *identical* marginals F_i's (except, of course, for distinct θ_i's), the estimate in (4.10) reduces to

$$\hat{\theta}=[1-\frac{pE|Z|/2f(0)}{||\mathbf{X}||^2+pE|Z|/2f(0)}]\mathbf{X}. \tag{4.11}$$

Remark 4.1. Recall that the estimate $\hat{\theta}=[1-\frac{p}{p+||\mathbf{X}||^2}]\mathbf{X}$ was earlier obtained in (3.5) in the context of the normal distribution. It is interesting to observe that for independent double exponential distributions, $E|Z_i|=1$ and $f_i(0)=1/2$ so that (4.10) simplifies to

$$\hat{\theta} = [1 - \frac{p}{p + ||\mathbf{X}||^2}]\mathbf{X}, \tag{4.12}$$

the same estimate as in (3.5) !

Remark 4.2. Indeed, when $f_i(z)$ is a normal scale mixture, namely,

$$f_i(z) = f(z) = \int_0^\infty \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{z^2}{2\sigma^2}} dG(\sigma), \tag{4.13}$$

it follows easily that $f(0) = \frac{1}{\sqrt{(2\pi)}}E(1/\sigma)$ and $\frac{1}{2}E|Z| = \frac{1}{\sqrt{(2\pi)}}E(\sigma)$. Thus, whenever $G(.)$ is such that $E(1/\sigma) = E(\sigma)$, for all such marginal models, the same form of $\hat{\theta}$ as given in (4.12) obtains. These give evidence of a fairly interesting phenomenon : *this apparently new version (4.12) of the James-Stein estimator arises quite frequently in our calculations*. It seems natural to briefly investigate its risk performance. Table 2 gives the risk of the estimate $\hat{\theta} = [1 - \frac{p}{p+||\mathbf{X}||^2}]\mathbf{X}$ for independent double exponential models at parameters of the form $\theta = c\mathbf{1} = (c, \cdots, c)'$ under absolute error loss. Note that the risk of $\mathbf{X}$ is p. In view of the calculations presented in Table 2, it is not unreasonable to speculate that $\hat{\theta}$ is minimax under absolute error loss for $p \geq 4$.

Table 2. Mean Absolute Error of $\hat{\theta}$: double exponential

c	$p = 3$	$p = 4$
0	2.505	3.391
.25	2.628	3.512
.5	2.782	3.682
1	2.986	3.905
2	3.045	3.979
5	3.033	3.982
10	3.033	3.982
20	3.032	3.983

Example 3. Consider the problem of estimating the vector of medians of independent *Extreme value* distributions, *i.e.*, X_i are independent with

$$f(x_i - \theta_i) = e^{-e^{(x_i-\theta_i+lnln2)}}.e^{x_i-\theta_i+lnln2}, \quad -\infty < x_i, \ \theta_i < \infty. \tag{4.14}$$

Here the median of X_i is θ_i, but obviously f is extremely skewed. To obtain the estimate in (4.11), we will need $f(0)$ and $E|Z|$. Of these, clearly, $f(0) = e^{lnln2}.e^{-e^{lnln2}} = \frac{ln2}{2}$. Also,

$$
\begin{aligned}
E|Z| &= \int_{-\infty}^{\infty} |z| e^{-e^{(z+lnln2)}} e^{z+lnln2} dz \\
&= ln2. \int_{-\infty}^{\infty} |z| e^{-(ln2)e^z} e^z dz \\
&= ln2[\int_0^{\infty} z e^{-(ln2)e^z} e^z dz - \int_{-\infty}^{0} z e^{-(ln2)e^z} e^z dz] \\
&= ln2[\int_1^{\infty} (lnx).e^{-(ln2)x} dx - \int_0^1 (lnx) e^{-(ln2)x} dx] \\
&= ln2[2\int_1^{\infty} (lnx).e^{-(ln2)x} dx - \int_0^{\infty} (lnx) e^{-(ln2)x} dx] \\
&= 2Ei(ln2) + (C + lnln2) \qquad (4.15)
\end{aligned}
$$

where $Ei(x) = \int_x^{\infty} \frac{e^{-t}}{t} dt$ and $C = -\int_0^{\infty} e^{-t}(lnt)dt = lim_{n\to\infty}(1 + 1/2 + ... + 1/n - logn)$ is Euler's constant; see pp. 573 in Gradshteyn and Ryshik (1980). Substitution of (4.15) into (4.11) gives the following James-Stein estimate for the Extreme value distribution under L_1 loss:

$$
\hat{\theta} = [1 - \frac{p\{C + lnln2 + 2Ei(ln2)\}}{p\{C + lnln2 + 2Ei(ln2)\} + (ln2)||\mathbf{X}||^2}]\mathbf{X}. \qquad (4.16)
$$

Table 3 gives a comparison of the mean absolute errors of $\hat{\theta}$ in (4.16) and of $\mathbf{X}$.

Table 3. Mean Absolute Error of $\hat{\theta}$ and $\mathbf{X}$ for $\theta = c\mathbf{1}$: Extreme value

	$p=3$	$p=3$	$p=4$	$p=4$
c	$\hat{\theta}$	$\mathbf{X}$	$\hat{\theta}$	$\mathbf{X}$
0	1.551	2.904	2.091	3.872
.25	1.664	2.904	2.208	3.872
.5	1.875	2.904	2.476	3.872
.75	2.096	2.904	2.767	3.872
1	2.296	2.904	3.034	3.872
3	2.886	2.904	3.832	3.872
5	2.912	2.904	3.868	3.872
10	2.901	2.904	3.871	3.872
20	2.895	2.904	3.871	3.872
50	2.894	2.904	3.871	3.872

By virtue of being a shrinkage estimator, $\hat{\theta}$ in (4.11) will always give a lower risk at $\theta = \mathbf{0}$ than its natural competitor $\mathbf{X}$. Writing $a = \frac{E|Z|}{2f(0)}$, the relative improvement in the risk at $\theta = \mathbf{0}$ is clearly equal to

$$\frac{\sum_{i=1}^{p} E|Z_i| - \sum_{i=1}^{p} E|(1 - \frac{pa}{pa+||\mathbf{Z}||^2})Z_i|}{\sum_{i=1}^{p} E|Z_i|}$$

$$= \frac{pa.E[\frac{\sum_{i=1}^{p}|Z_i|}{pa+||\mathbf{Z}||^2}]}{pE|Z|}$$

$$= \frac{a}{E|Z|}.E[\frac{\sum_{i=1}^{p}|Z_i|}{pa+||\mathbf{Z}||^2}]$$

$$= \frac{a}{E|Z|}.E[\frac{\sum_{i=1}^{p}|Z_i|/p}{a+\frac{||\mathbf{Z}||^2}{p}}]. \tag{4.17}$$

(4.17) does not require independence of X_i's; if they are, then by the *Weak Law of Large Numbers* (see Ferguson, 1996), $\frac{\sum_{i=1}^{p}|Z_i|}{p} \to E|Z|$ and $\frac{\sum_{i=1}^{p}|Z_i^2|}{p} \to E(Z^2)$, in probability, as $p \to \infty$, and by Schwartz's inequality, $\frac{\sum_{i=1}^{p}|Z_i|/p}{a+\frac{||\mathbf{Z}||^2}{p}}$ is a bounded sequence. It then follows from (4.17) that the following proposition on the relative risk improvement at $\theta = \mathbf{0}$ holds:

Proposition. For independent X_i's with identical marginals F_i's (except for distinct θ_i's), under absolute error loss, the relative risk improvement at $\mathbf{0}$ by $\hat{\theta}$ (in (4.11)) over $\mathbf{X}$ converges to $\frac{E|Z|}{E|Z|+2f(0)E(Z^2)}$ as the number p of dimensions tends to ∞.

Table 4 lists this limiting % improvement in mean absolute error for a few important parent distributions.

Table 4. Limiting % Improvement in Mean Absolute Error at $\mathbf{0}$

Distribution	% Improvement
Normal	50%
Double exponential	$33\frac{1}{3}$%
Logistic	45.73%
t(3)	$33\frac{1}{3}$%
t(5)	42.86%
Extreme value	62.90%

4.2. Important Special Case: Adaptation to a Spherically Symmetric Law under Absolute Error Loss.

Stein type estimators for spherically symmetric distributions under squared error (or concave functions of squared error) loss are well studied (Brandwein and Strawderman, 1990, 1991). In this subsection we show how the estimate in (4.10) simplifies for a general spherically symmetric density. Note that the estimate in (4.10) corresponds to absolute error loss, literature on which is extremely limited.

Let $\mathbf{X}|\theta \sim f(\sum_{i=1}^{p}(x_i-\theta_i)^2)$. Let $F(z)$ be a primitive of $f(z)$, *i.e.*, $F'(z) = f(z)$ for all $z \geq 0$ and the primitive is such that $F(\infty) = 1$. The quantities required in (4.10) simplify as follows on familiar calculations. The marginal density of $Z_1 = X_1 - \theta_1$ is

$$\begin{aligned} f_{Z_1}(z) &= \frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \cdot \int_0^\infty x^{\frac{p-3}{2}} f(x+z^2)dx \\ \rightarrow f_{Z_1}(0) &= \frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \cdot \int_0^\infty x^{\frac{p-3}{2}} f(x)dx; \end{aligned} \tag{4.18}$$

$$\begin{aligned} E|Z_1| &= \frac{2\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \cdot \int_0^\infty \int_0^\infty z x^{\frac{p-3}{2}} f(x+z^2)dxdz \\ &= \frac{2\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \cdot \int_0^\infty x^{\frac{p-3}{2}} \int_0^\infty z f(x+z^2)dxdz \\ &= \frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \cdot \int_0^\infty x^{\frac{p-3}{2}} \int_0^\infty [-\frac{d}{dz}\{1-F(x+z^2)\}]dxdz \\ &= \frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \cdot \int_0^\infty x^{\frac{p-3}{2}} \{1-F(x)\}dx \\ &= \frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p+1}{2})} \cdot \int_0^\infty x^{\frac{p-1}{2}} f(x)dx. \end{aligned} \tag{4.19}$$

Substitution of (4.18) and (4.19) into (4.10) results in the following nice closed form estimate adapted to the spherically symmetric density f under absolute error loss:

$$\hat{\theta} = [1 - \frac{p\int_0^\infty x^{\frac{p-1}{2}} f(x)dx}{p\int_0^\infty x^{\frac{p-1}{2}} f(x)dx + ||\mathbf{X}||^2(p-1)\int_0^\infty x^{\frac{p-3}{2}} f(x)dx}]\mathbf{X}. \tag{4.20}$$

Example 4. Adaptation: Normal *vs* t

It is readily verified that (4.20) simplifies to the estimate $\hat{\theta} = [1-\frac{p}{p+||\mathbf{X}||^2}]\mathbf{X}$ when $\mathbf{X}$ is multivariate normal with an identity covariance matrix. This is consistent with what we saw in (3.5).

If $\mathbf{X}$ has the spherically symmetric t density with m degree of freedom, *i.e.*, if

$$f(\mathbf{x}|\theta) = \frac{\Gamma(\frac{m+p}{2})}{2\Gamma(\frac{m}{2}+1)m^{\frac{p}{2}-1}\pi^{\frac{p}{2}}} \cdot \frac{1}{[1+\frac{\sum_{i=1}^{p}(x_i-\theta_i)^2}{m}]^{\frac{m+p}{2}}}, \tag{4.21}$$

then a remarkable calculational simplification occurs, and the estimate (4.20) reduces to

$$\hat{\theta} = [1 - \frac{p}{p+\frac{m-1}{m}||\mathbf{X}||^2}]\mathbf{X}. \tag{4.22}$$

Comparison with the estimate $\hat{\theta} = [1 - \frac{p}{p+||\mathbf{X}||^2}]\mathbf{X}$ shows that according to (4.21), for the heavier tailed multivariate t distribution, a little more shrinkage is recommended than for the normal case, for the important absolute error loss. Figure 2 gives a plot of the mean absolute error of the estimate in (4.22) for $p = 3$ when the degrees of freedom m is also 3. It is clearly encouraging.

FIGURE 2 here

For the general spherically symmetric case, following the spirit of (4.17), a completely general formula for the relative risk improvement under absolute error loss at $\theta = \mathbf{0}$ can be given for the estimate presented in (4.20). Following the derivation of (4.17), this equals

$$\frac{1}{E|Z_1|} \cdot E[\frac{|Z_1|}{1+b||\mathbf{Z}||^2}] \tag{4.23}$$

where $b = \frac{p-1}{p} \cdot \frac{\int_0^\infty x^{\frac{p-3}{2}} f(x)dx}{\int_0^\infty x^{\frac{p-1}{2}} f(x)dx}$. An expression for $E|Z_1|$ was given in (4.19). Transforming to polar coordinates, the other term $E[\frac{|Z_1|}{1+b||\mathbf{Z}||^2}]$, upon some algebra, is seen to be

$$E[\frac{|Z_1|}{1+b||\mathbf{Z}||^2}] = \frac{\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p+1}{2})} \cdot \int_0^\infty \frac{f(x)}{1+bx} dx. \tag{4.24}$$

In the case of the spherically symmetric t distribution with m degrees of freedom, this general formula (4.23), showing the *percentage improvement in mean absolute error at* $\theta = \mathbf{0}$ *for the estimate (4.22)*, reduces to the following:

$$\frac{200p\Gamma(\frac{m+p}{2})_2F_1(\frac{m+p}{2}, 1; \frac{m+p}{2}+1; 1-\frac{p}{m(p-2)})}{(p-2)\Gamma(\frac{p+1}{2})m^{\frac{p+1}{2}}(m+p)\Gamma(\frac{m-1}{2})} \tag{4.25}$$

where ${}_2F_1$ denotes the standard hypergeometric function. Figure 3 gives a plot of (4.25) for $p = 3$ as the degrees of freedom m varies between 3 and 20. It seems to be stable with respect to the degrees of freedom.

FIGURE 3 here

5. α-Power loss.

Finally, we now consider the more general loss function

$$L(\delta,\theta)=\sum_{i=1}^{p}|\delta_i-\theta_i|^\alpha,\ \ \alpha>1 \tag{5.1}$$

and show that indeed the James-Stein estimators arise again. The case $\alpha=1$, as will be clear from (5.8), needs to be considered separately; this was presented in Section 4. As before, we propose to estimate the vector θ in the joint *pdf* $f(\mathbf{x}|\theta)$ of $\mathbf{X}$ by $\frac{\mathbf{X}}{1+\lambda}$ for some $\lambda\geq 0$. Writing $Z_i=X_i-\theta_i$ and noting that

$$\begin{aligned} E|Z_i-a|^\alpha &= \int_a^\infty (z-a)^\alpha f_i(z)dz+\int_{-\infty}^{a}(a-z)^\alpha f_i(z)dz \\ &= h_i(a)+g_i(-a), say, \end{aligned} \tag{5.2}$$

we can express the *risk* of $\frac{\mathbf{X}}{1+\lambda}$ as

$$E[\sum_{i=1}^{p}|\frac{X_i}{1+\lambda}-\theta_i|^\alpha]=\frac{1}{(1+\lambda)^\alpha}\sum_{i=1}^{p}\{h_i(\lambda\theta_i)+g_i(-\lambda\theta_i)\}. \tag{5.3}$$

Using the facts that

$$\begin{aligned} h_i'(a) &= -\alpha\int_a^\infty (z-a)^{\alpha-1}f_i(z)dz \\ g_i'(a) &= -\alpha\int_{-\infty}^{-a}(-a-z)^{\alpha-1}f_i(z)dz, \end{aligned} \tag{5.4}$$

minimization of (5.3) with respect to λ for fixed θ results in the equation

$$\begin{aligned} \psi(\lambda|\theta) &= (1+\lambda)\sum_{i=1}^{p}\theta_i\{\int_{\lambda\theta_i}^\infty (z-\lambda\theta_i)^{\alpha-1}f_i(z)dz \\ &\quad -\int_{-\infty}^{\lambda\theta_i}(\lambda\theta_i-z)^{\alpha-1}f_i(z)dz\}+\sum_{i=1}^{p}\{h_i(\lambda\theta_i)+g_i(-\lambda\theta_i)\} \\ &= 0. \end{aligned} \tag{5.5}$$

Following the same steps outlined in Sections 3 and 4, a first order approximation of the true solution $\lambda_0(\theta)$ of the above equation can be obtained on the basis of $\psi(0|\theta)$ and $\psi'(0|\theta)$ which are given by

$$\psi(0|\theta) = \sum_{i=1}^{p} E|Z_i|^\alpha + \sum_{i=1}^{p} \theta_i\{\int_0^\infty z^{\alpha-1} f_i(z)dz - \int_{-\infty}^{0} (-z)^{\alpha-1} f_i(z)dz\} \quad (5.6)$$

$$\psi'(0|\theta) = (1-\alpha)[\sum_{i=1}^{p} \theta_i^2 E|Z_i|^{\alpha-2} + \sum_{i=1}^{p} \theta_i\{\int_0^\infty z^{\alpha-1} f_i(z)dz - \int_{-\infty}^{0} (-z)^{\alpha-1} f_i(z)dz\}]. \quad (5.7)$$

One then gets, for $\alpha \neq 1$, a first order approximation

$$\lambda_0(\theta) \sim -\frac{\psi(0|\theta)}{\psi'(0|\theta)}. \quad (5.8)$$

Under the assumption that each f_i is symmetric, one has the simplification

$$\int_0^\infty z^{\alpha-1} f_i(z)dz - \int_{-\infty}^{0} (-z)^{\alpha-1} f_i(z)dz\} = 0, \quad for\ all\ i, \quad (5.9)$$

and so (5.8) simplifies to

$$\lambda_0(\theta) \sim \frac{\sum_{i=1}^{p} E|Z_i|^\alpha}{(\alpha-1)\sum_{i=1}^{p} \theta_i^2 E|Z_i|^{\alpha-2}}. \quad (5.10)$$

This ultimately leads to the following James-Stein estimator *adapted to the loss as well as the parent distribution*:

$$\hat{\theta} = [1 - \frac{\sum_{i=1}^{p} E|Z_i|^\alpha}{\sum_{i=1}^{p} E|Z_i|^\alpha + (\alpha-1)\sum_{i=1}^{p} X_i^2 E|Z_i|^{\alpha-2}}]\mathbf{X}. \quad (5.11)$$

Note that (5.11) is obtained under very mild assumptions: one has a general power loss, and the coordinate distributions need not be either independent or identical, but their symmetry is assumed.

For quick comprehension and comparison, we present in Table 5 the adaptive James-Stein estimator for a few important cases.

Table 5. Adaptive Stein Estimates

Parent Distribution	α	$\hat{\theta}$
Independent N(θ_i,1)	1	$\left[1 - \frac{p}{p+\|\mathbf{X}\|^2}\right]\mathbf{X}$
same	2	same
same	4	same
Independent double exponential (θ_i,1)	1	same
same	2	$\left[1 - \frac{2p}{2p+\|\mathbf{X}\|^2}\right]\mathbf{X}$
same	4	$\left[1 - \frac{12p}{12p+\|\mathbf{X}\|^2}\right]\mathbf{X}$

Since, for $Z \sim N(0,1)$, $E|Z|^\alpha = (\alpha-1)E|Z|^{\alpha-2}$ for $\alpha \geq 1$, one gets from (5.11) that in the normal case *the same estimate* $[1 - \frac{p}{p+||\mathbf{X}||^2}]\mathbf{X}$ *emerges for every* $\alpha \geq 1$: quite a robustness property of this estimate!

6. Concluding remarks.

In view of our calculations, it would be interesting to investigate if James-Stein estimators can be given an empirical Bayes interpretation in the kind of generality we worked on, *i.e.*, for a general power loss and a rather general parent distribution. It would also be worthwhile to explore the particular estimate $[1 - \frac{p}{p+||\mathbf{X}||^2}]\mathbf{X}$ more deeply as it arises naturally and frequently according to the calculations we presented. Extension to a smooth loss function $\sum_{i=1}^p L|\delta_i - \theta_i|$ will be similar but we did not consider it here. We believe the adaptation methods we present can be helpful in research on minimax estimation of general location parameters, particularly for absolute error loss.

Acknowledgments. Our sincere thanks are due to Dhiren Ghosh of Washington, D.C. for suggesting the univariate normal problem under absolute error loss, which eventually led to this investigation, and Malay Ghosh for his careful reading of the manuscript and several constructive suggestions. We also had useful discussions with George Casella who made valuable comments on the risk calculations.

References

Berger, J.O. (1986). **Statistical Decision Theory and Bayesian Analysis**. Springer-Verlag, New York.

Brandwein, A., and Strawderman, W.E. (1990). "Stein Estimation : The spherically symmetric case". *Stat. Sc.*, 5, 356-359.

Brandwein, A., and Strawderman, W.E. (1991). "Generalizations of James Stein estimates under spherical symmetry".*Ann. Stat.*, 19, 1639-1650.

Brown, L.D. (1966). "On the admissibility of invariant estimators of one or more location parameters".*Ann. Math. Stat.*, 37, 1087-1136.

Casella. G. (1980). "Minimax ridge regression estimators". *Ann. Stat.*, 8, 1036-1056.

Gradshteyn, I., and Ryzhik, I.V. (1980). *Tables of Integrals, Series, and Products.* Academic Press. New York.

Gupta, A.K., and Pena, E.A. (1991). "A simple motivation for James-Stein estimators". *Statistics & Probability Letters*, 12, 337-340.

Ferguson, T. (1996). **A Course in Large Sample Theory**. Chapman & Hall.

James,W., and Stein,C. (1961). "Estimation with quadratic loss". *Proc. Fourth Berkeley Symp. Prob. Stat.*, 1, 361-380, Univ. Calif. Press, Berkeley.

Lehmann, E.L., and Casella, G. (1997). **Theory of Point Estimation, Third Edition**. Springer-Verlag, New York.

Shinozaki, M. (1984). "Simultaneous estimation of location parameters under quadratic loss". *Ann. Stat.*, 12, 322-335.

Stein, C. (1981). "Estimation of the mean of a multivariate normal distribution". *Ann. Stat.*, 9, 1135-1151.

Stigler, S.(1990). "The 1988 Neyman memorial lecture: a Galtonian perspective on shrinkage estimators". *Stat. Sc.*, 5, 147-155.

FIG. 1

RISK OF ESTIMATE IN (3.5) AND JAMES-STEIN ESTIMATE UNDER ABSOLUTE ERROR LOSS

$p = 4, \underset{\sim}{\theta} = c\underset{\sim}{1}$

DISTRIBUTION = $N(\underset{\sim}{\theta}, I)$

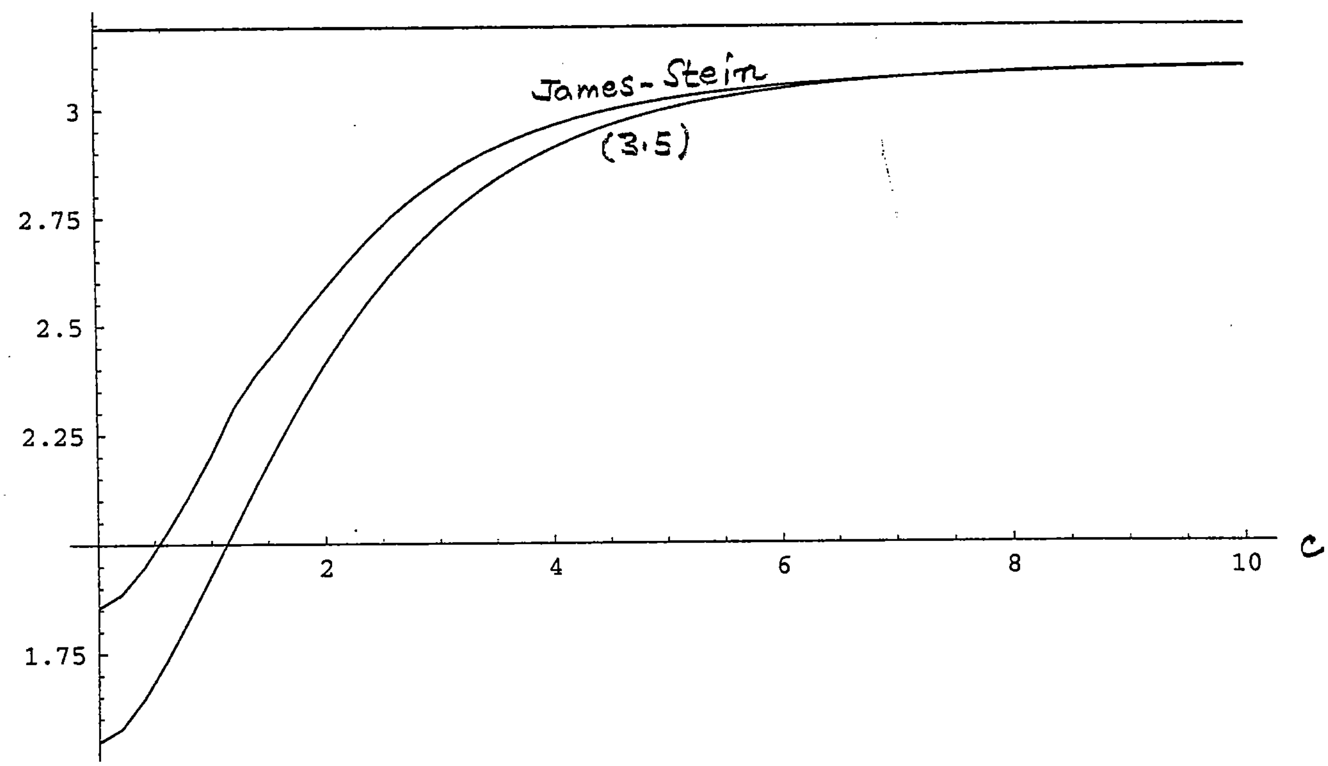

FIG. 2

RISK OF ESTIMATE IN (4.22)
UNDER ABSOLUTE ERROR LOSS
$p = 3$, $\theta = (0, 0, c)$
DISTRIBUTION = t(D.F. = 3)

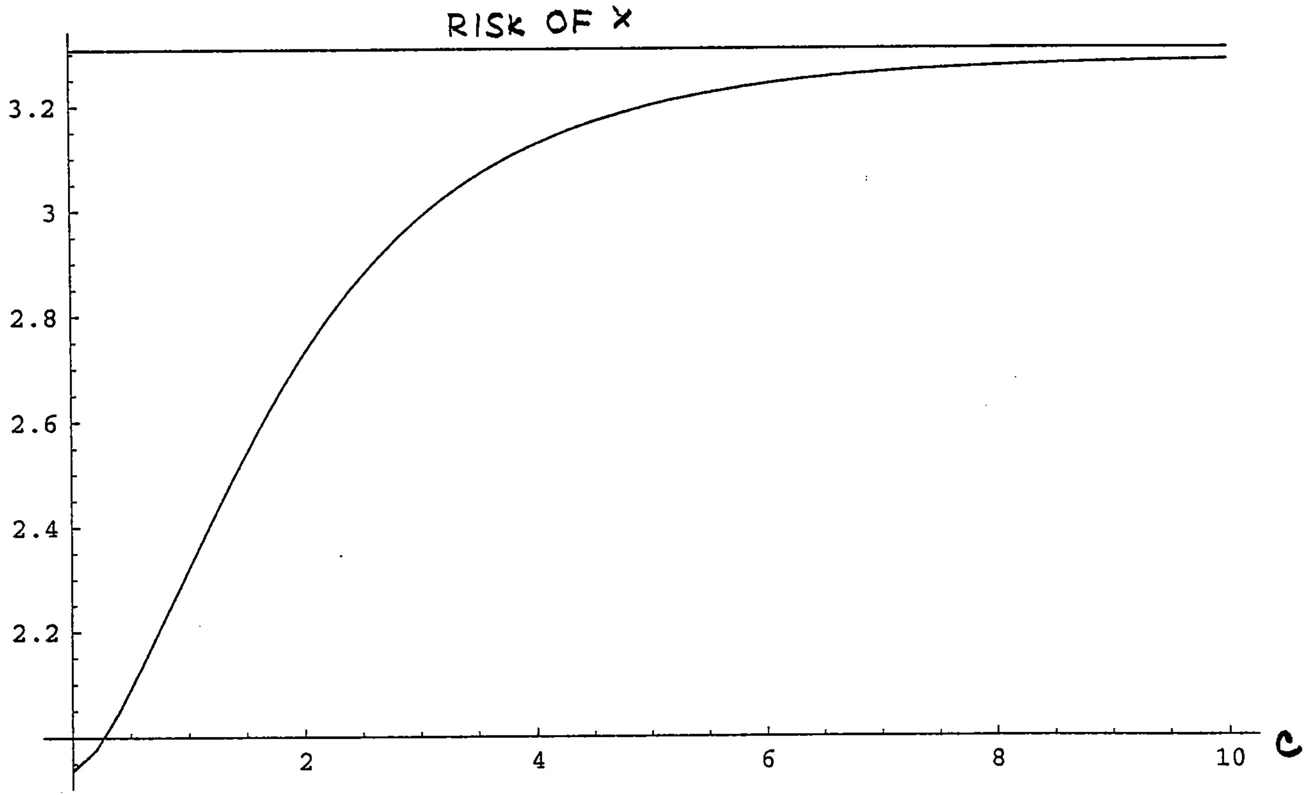

FIG. 3

% improvement in Mean Absolute Error over $\underset{\sim}{X}$

Plot of (4.25)

DISTRIBUTION = t(D.F. = m)