Inadmissibility of the Usual Estimators of Scale parameters in Problems with Unknown Location and Scale Parameters
Author(s): L. Brown
Source: *The Annals of Mathematical Statistics*, Vol. 39, No. 1 (Feb., 1968), pp. 29-48
Published by: Institute of Mathematical Statistics
Stable URL: http://www.jstor.org/stable/2238908
Accessed: 25/03/2010 14:49

http://www.jstor.org

INADMISSIBILITY OF THE USUAL ESTIMATORS OF SCALE PARAMETERS IN PROBLEMS WITH UNKNOWN LOCATION AND SCALE PARAMETERS[1]

By L. Brown

Cornell University

1. Introduction. During the past two decades the admissibility of common estimators of location parameters (e.g. the population mean) has been studied by several authors, see Brown (1966) for a short bibliography. By utilizing a logarithmic transformation these results can also be applied to the problem of estimating unknown scale parameters when these are the only unknown parameters (in particular, when the location parameters are known). See Farrell (1964) for details.

In a sense the next more complicated problem is that of estimating any power of the scale parameter (e.g. the population variance) when the location parameter is unknown. This problem was studied in Stein (1964) where inadmissibility of the usual estimator was demonstrated in one important case (estimating the variance of independent identically distributed normal observations, when using squared error loss function).

The results contained in our paper cover general situations with regard to both the distributions and the loss functions involved, but with one major restriction (the compact support assumption in Section 5). Our methods of proof are different from those used in Stein's paper (op. cit.), but the general nature of the results are the same—namely, there is a scale invariant estimator which is better than the usual location and scale invariant estimator (the formula is given in (4.2)). In addition it is true that our estimators may also be motivated by the following empirical Bayes argument. If the estimate, say $\bar{x}$, of the location parameter is small compared with the estimate, say s^2, of the variance then this indicates that the location parameter, μ, is near 0. However, when μ is known to be near 0 *and $\bar{x}$ is also near* 0 the best estimator of the variance is a smaller multiple of s^2 than it is when μ is unknown. Thus, by this reasoning one should use the usual estimator for σ^2 when $|\bar{x}|/s$ is large and a somewhat smaller multiple when $|\bar{x}|/s$ is small. (The italicized portion of the above argument is not needed to justify the estimator (6) of Stein [op. cit.] but is needed to justify estimators such as our δ_1 (see (4.2)).) Actually the proofs in Sections 4 and 5 use a different though related idea; namely that for any μ the value of S given $s > |\bar{x}|/K$ (for fixed K, $K > 0$) is stochastically larger than the unconditional values of S. In

---

order to make this into a proof we will need to be able to make such a statement uniformly in μ, and the loss function will have to satisfy certain reasonable conditions.

Although the estimators proposed in this paper are better than the best invariant estimator, they are themselves inadmissible. We have made no attempt here to propose estimators which in practice afford significant improvement over the usual estimators. Our specific aim in this paper has been to prove certain mathematical theorems in order to provide increased insight into the decision theory of statistical problems which contain unknown location and scale parameters. Nevertheless in a spirit of idle curiosity we computed the risks of our estimators in a few special cases and found—as we had suspected—that they have only slightly smaller risk than the usual procedure. The greatest improvement we found was approximately 9%. We do not know if this figure can be greatly or only slightly increased by the use of admissible procedures. These computations are reported very briefly in Section 6.

The squared error loss function has been traditionally used in a wide variety of theoretical statistical investigations. This is partly because of its mathematical simplicity. But in the case of the location parameter problem other justifications exist as well. Chief among these is probably the fact that when squared error is used the best invariant estimator is unbiased in that problem. However it is well known that in the problem of estimating an unknown scale parameter the best invariant estimator for squared error loss is biased, see Pitman (1938), p. 406. (For example, if $X_1, \cdots, X_n$ are independent identically distributed normal random variables the best invariant estimator of σ^2 for squared error loss is $(n + 1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$, while the unbiased estimator is $(n - 1)^{-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$.) (The biased nature of the best invariant estimator may be imperfectly justified intuitively by the fact that the squared error loss function is out of balance for the scale problem, in the sense that the maximum loss for estimating σ^2 too small is finite, while the maximum loss for estimating σ^2 too large is infinite.)

Because of these possibly disturbing features of squared error loss in the scale parameter problem, it seems reasonable to discuss the simple properties of some other reasonable loss functions and of the corresponding best invariant estimators. Section 3 of the paper is devoted to such a discussion.

The main result of Section 3 (Theorem 3.1) is that there is an essentially unique loss function for which the best invariant estimator is always unbiased Because of this, any general principle to the effect that lacking other information one should always use an unbiased estimator in scale problems is equivalent to a principle that in such situations one should always use this unique loss function. It follows from this, as a kind of corollary to the inadmissibility theorems mentioned earlier, that one may say (if one wishes!) that the unbiased estimator of any power of the scale parameter is inadmissible. Theorem 3.3 looks at a different property of this unique loss function.

The main inadmissibility theorems of the paper are contained in Sections four and five.

2. Definitions, assumptions, and remarks on differentiation. Let S, T, Z be random variables taking the values s, t, z in $E^1 \times E^1 \times E^m$, with $s > 0$. ($m = 0$ is possible, in which case we will drop the symbol z from the notation.) We suppose (if $m \geq 1$) that Z has a probability distribution $\eta(\cdot)$; and for almost all z, S and T have a conditional probability density with respect to Lebesque measure, given by

$$f_{\mu,\sigma}(s, t \mid z) = \sigma^{-2} f_{01}(s/\sigma, (t - \mu)/\sigma \mid z)$$

$$= \sigma^{-2} f(s/\sigma, (t - \mu)/\sigma \mid z).$$

(If $m = 0$, S, T have the density (not conditional) $f_{\mu,\sigma}(s, t) = \sigma^{-2} f(s/\sigma, (t - \mu)/\sigma)$.) μ and σ are unknown parameters. (When, as in the last line of the above, subscripts are omitted; it is assumed that $\mu = 0$, $\sigma = 1$.) We will often use the conditional density, which can easily be shown to exist:

$$f_\sigma(s \mid z) = \sigma^{-1} f_1(s/\sigma \mid z) = \sigma^{-1} f(s/\sigma \mid z).$$

($f_\sigma(s \mid z)$ does not depend on μ.)

If X_1, X_2, $\cdots$, X_n, $n \geq 2$, are independent identically distributed real-valued random variables with density $\sigma^{-1} g((x - \mu)/\sigma)$ with respect to Lebesgue measure then the variables

$$t = \bar{x} = n^{-1} \sum_{i=1}^{n} x_i,$$

(2.1)
$$s = [(n - 1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2]^{\frac{1}{2}},$$

$$z_i = (x_{i+2} - \bar{x})/s, \qquad i = 1, \cdots, n - 2 = m,$$

satisfy the previous definition of a location and scale problem. (In Theorem 3.2 we use a very slightly different but equivalent definition.)

An *estimator* of σ^k is a function $\delta(s, t, z)$. k is a fixed, known, constant, $k > 0$. It is *invariant* (or, *fully invariant*) if $\delta(s, t, z) = \varphi(z)s^k$. (If $m = 0$, φ is a constant.) It is *scale invariant* if it is of the form $\delta(s, t, z) = \pi(t/s; z)s^k$. Note that invariant estimators are also scale invariant. (Randomized estimators may be accommodated, as in Brown (1966), Section 1.3.)

Throughout the paper we will deal only with non-negative *invariant loss functions*—that is, with loss functions of the form $L(\delta/\sigma^k)$ where $L \geq 0$. We will also *assume throughout* that $L(\cdot)$ is of bounded variation on every interval (a, b) with $0 < a < b < \infty$, and that L is right continuous. (Since L is of bounded variation, there is always a right continuous version of L.)

The *risk function* is $R(\delta; \mu, \sigma) = E_{\mu,\sigma}(L(\delta/\sigma_k))$. If δ is scale invariant R is a function only of δ and the ratio μ/σ. Hence to prove that one scale invariant estimator is better than another it is sufficient (and necessary) to consider only the case $\sigma = 1$. For this reason we will need to consider only the case $\sigma = 1$ in the proofs of the theorems in sections 3, 4, and 5.

Throughout the paper we *assume* that there exists an essentially unique measurable *best invariant estimator* described by $\varphi_0(\cdot)$ which satisfies:

$$\int L(\varphi_0(z)s^k/\sigma^k) f_\sigma(s \mid z) \, ds = R(\varphi_0(z)s^k \mid z) = \inf_\varphi R(\varphi s^k \mid z).$$

(The above expression defines $R(\cdot\,|\,z)$, which, to be formally precise should be written $R((\varphi s^k)\,|\,z)$; since the argument is really the function (φs^k).) Actually, the basic inadmissibility theorems of the paper remain valid with several minor modifications if φ_0 exists, but is not essentially unique.

We will also have to assume that the expression

$$R(\varphi s^k\,|\,z) = \int L(\varphi s^k) f(s\,|\,z)\,ds$$

can be differentiated in some sense with respect to φ under the integral sign in a neighborhood of $\varphi_0(z)$. Since we want this assumption to be as weak as possible, we will now briefly discuss the problem of differentiating under the integral.

For $-\infty \leqq a < b \leqq \infty$ let

$$_0C_0(a,b) = \{h(\cdot)\,|\,h(\cdot) \text{ continuous on } (a,b); \exists\, a < \alpha < \beta < b \text{ such that}$$

$$(2.2) \qquad a \leqq x \leqq \alpha \text{ or } \beta \leqq x \leqq b \text{ implies } h(x) = 0\},$$

$$C_0(a,b) = \{h(\cdot)\,|\,h(\cdot) \text{ bounded and continuous on } (a,b); \exists\, \beta < b \text{ such}$$

$$\text{that } \beta \leqq x \leqq b \text{ implies } h(x) = 0\}$$

$(_0C(a,b)$ and $C(a,b)$ can be similarly defined.)

Let $h(\cdot)\,\varepsilon\,_0C_0(-\infty,\infty)$ and suppose λ is a real function of bounded variation on every bounded interval of $(-\infty,\infty)$. In Brown (1966) we proved the following simple fact:[2]

$$(2.3) \quad (\partial/\partial\psi)\int h(v)\lambda(v+\psi)\,dv = \int h(v)\,d_v\lambda(v+\psi) = \int h(v-\psi)\,d\lambda(v).$$

It follows that the derivative (2.3) is a continuous function of ψ.

Now, suppose $g\,\varepsilon\,_0C_0(0,\infty)$. By assumption $L(\cdot)$ is of bounded variation on every interval (a,b) where $0 < a < b < \infty$. If we let $v = \ln s^k$ and $\psi = \ln\varphi$, and use (2.3)

$$(\partial/\partial\varphi)\int g(s)L(\varphi s^k)\,ds = (\partial/\partial\varphi)\int g(e^{v/k})L(e^{\psi+v})(e^{v/k}/k)\,dv$$

$$(2.4) \qquad\qquad = e^{-\psi}(\partial/\partial\psi)\int h(v)\lambda(\psi+v)\,dv$$

$$= e^{-\psi}\int h(v)\,d_v\lambda(v+\psi)$$

where $h(v) = e^{v/k}g(e^{v/k})/k$ and $\lambda(y) = L(e^y)$. It follows from (2.4) that for each $\varphi > 0$, $(\partial/\partial\varphi)\int g(s)L(\varphi s_k)\,ds$ is a linear functional on $_0C_0(0,\infty)$. Thus for each $\varphi > 0$ there is a unique measure $\nu_\varphi(\cdot)$ on $(0,\infty)$ such that

$$(2.5) \qquad\qquad (\partial/\partial\varphi)\int g(s)L(\varphi s^k)\,ds = \int g(s)\nu_\varphi(ds).$$

If we substitute s and φ into the right side of (2.4) that expression becomes $\int g(s)(s/\varphi k)\,d_vL(\varphi s^k)$. It is therefore reasonable to write ν_φ symbolically as $\nu_\varphi(ds) = (s/\varphi k)\,d_{\ln s^k}L(\varphi s^k)$. In particular, if L is absolutely continuous then $\nu_\varphi(ds) = s^k L'(\varphi s^k)\,ds$ and (2.5) is the usual formula

$$(2.6) \qquad\qquad (\partial/\partial\varphi)\int g(s)L(\varphi s^k)\,ds = \int g(s)s^k L'(\varphi s_k)\,ds.$$

---

[2] We suspect this result is in all standard real variable texts, but we have never found an explicit proof of it in such a text.

With this definition (2.5) of differentiation under the integral sign and of ν_φ we can state the following assumption, which is to hold throughout this paper.

ASSUMPTION A. There is a function $\epsilon(z) > 0$ such that for almost all z, $|\varphi - \varphi_0(z)| < \epsilon(z)$ implies

$$(2.7) \qquad (\partial/\partial\varphi) \int L(\varphi s^k) f(s \mid z) \, ds = \int f(s \mid z) \nu_\varphi(ds) \neq \pm\infty.$$

If L is of bounded variation on every interval $0 < a < b < \infty$ and $f(\cdot \mid z) \varepsilon {}_0 C_0(0, \infty)$ (which is rarely the case in practice) then Assumption A is automatically satisfied. Thus for continuous $f(\cdot \mid z)$ the assumption is a condition only on the behavior of $L(\cdot)$ and of $f(\cdot \mid z)$ near 0 and near ∞. (Near 0 it is usually enough just for f to be bounded since ν_φ will usually satisfy (2.8).) But Assumption A may also be satisfied if $f(\cdot \mid z)$ is not continuous (roughly speaking so long as "uncontinuities" of f and of L do not "coincide"). Using standard methods one can give various sufficient conditions for Assumption A to be satisfied and thus show that in practice it is very mild assumption. We will not pursue the question further here, except to note that all the examples of Section 3 satisfy the assumption.

Let us point out one further property which is needed later. First we need the following:

DEFINITION. L is *bowl-shaped* (respectively, *strictly bowl-shaped*) if there is a number a_L, $0 < a_L < \infty$, such that $L(y)$ is non-increasing (respectively, strictly decreasing) for $y \leqq a_L$ and non-decreasing (respectively, strictly increasing) for $y \geqq a_L$.

LEMMA 2.1. *Suppose L is bowl-shaped, and $g(s, \mu)$ is a bounded non-negative function, continuous on $(0, \infty) \times (\alpha, \beta)$. Assume that for some $a > 0$ and $\varphi_1 > 0$*

$$(2.8) \qquad \int_0^a |\nu_{\varphi_1}(ds)| < \infty.$$

Then, if $\mu_0 \varepsilon (\alpha, \beta)$,

$$(2.9) \qquad \liminf_{\varphi\to\varphi_1,\mu\to\mu_0} \int g(s, \mu)\nu_\varphi(ds) \geqq \int g(s, \mu_0)\nu_{\varphi_1}(ds).$$

PROOF. This result follows easily from the representation (2.4), (2.5). We have

$$\int g(s, \mu)\nu_\varphi(ds) = \int h(v - \psi, \mu) \, d\lambda(v)$$

where $h(v, \mu) = (e^{v/k}/k)g(e^{v/k}, \mu)$. Thus $h(\cdot, \cdot)$ is non-negative and continuous. $\lambda(v) = L(e^v)$ is also bowl-shaped with its minimum at $\ln a_L$. Using (2.8), and (2.4)

$$(2.10) \qquad \int_0^{\ln a_L} e^{v/k} \, d\lambda(v) > -\infty.$$

It then follows, using the boundedness of h and Fatou's theorem that

$$(2.11) \qquad \liminf_{\psi\to\psi_1,\mu\to\mu_0} \int h(v - \psi, \mu) \, d\lambda(v) \geqq \int h(v - \psi_1, \mu) \, d\lambda(v)$$

where $\psi_1 = \ln \varphi_1$. This verifies the inequality in (2.9) and proves the lemma.

The assumption (2.8) will occur so frequently that we will state it formally as

ASSUMPTION B. For some $a > 0$, and $\varphi_1 > 0$

$$(2.12) \qquad \int_0^a |\nu_{\varphi_1}(ds)| < \infty.$$

As in (2.10), this implies $\int_0^a |\nu_\varphi(ds)| < \infty$ for all $\varphi > 0$, $a > 0$. We will specifically state when this assumption is being used. As with Assumption A, this assumption is also very mild. In particular, it is satisfied by the three loss functions discussed in Section 3. Note that if Assumption B is satisfied and $f(\cdot \mid z) \, \varepsilon \, C_0(0, \infty)$ then Assumption A is also satisfied.

If we let $f(s \mid z) = g(s, \mu)$ the above lemma shows in particular that if Assumption A is satisfied, and $f(\cdot \mid z)$ is a bounded continuous function and Assumption B is satisfied then $(\partial/\partial\varphi)R(\varphi s^k \mid z)$ is lower semi-continuous in φ at $\varphi_0(z)$.

Actually, if L is absolutely continuous (so that ν_φ is given by (2.6)), and the Assumptions A and B are satisfied then $f(\cdot \mid z)$ bounded implies $(\partial/\partial\varphi)R(\varphi s^k \mid z)$ is lower semi-continuous at $\varphi_0(z)$. It is not necessary in this case that $f(\cdot \mid z)$ be continuous. This result follows from the fact that with h and φ as above $h(v - \psi) \to h(v - \psi_0)$ in measure (Lebesgue measure) as $\psi \to \psi_0$ (see, for example, Wiener (1933), p. 13). The desired lower semi-continuity then follows using the representation (2.6) of ν_φ and Fatou's theorem in an expression like (2.11).

3. Relationships among best invariant estimators for certain invariant loss functions; unbiased estimation.
We have already mentioned one common invariant loss function, namely squared error, which is given by the formula

$$L_s(y) = (y - 1)^2 \qquad (y = \delta/\sigma^k).$$

(Often, in the literature, squared error is defined as $(\delta - \sigma^k)^2$ which is σ^k times our function. However, these two forms are equivalent for consideration of questions of admissibility.)

The best invariant estimator for this problem is the solution to the equation

$$(3.1) \qquad \int s^k(\varphi(z)s^k - 1)f(s \mid z) \, ds = 0.$$

That is

$$(3.2) \qquad \varphi(z) = \int s^k f(s \mid z) \, ds / \int s^{2k} f(s \mid z) \, ds.$$

Here we have made use of Assumption A to differentiate under the integral sign. For later use let us *denote* this function φ by $\varphi_s(\cdot)$ and $\delta_s(\cdot) = \varphi_s(\cdot)s^k$.

As we mentioned in the introduction many people prefer the use, in theoretical assumption problems, of the unbiased estimator rather than the use of δ_s. The *best unbiased estimator* is defined by

$$\varphi_u(z) = (\int s^k f(s \mid z) \, ds)^{-1} \qquad \text{and} \qquad \delta_u(\cdot) = \varphi_u(\cdot)s^k$$

Define the loss function $L_u(y)$ by

$$(3.3) \qquad L_u(y) = y - 1 - \ln y.$$

Note that $L_u(y)$ is strictly convex in y. Hence the best invariant estimator for L_u, if it exists, is unique. Also, note that if L is given by (3.3) then for any φ, $0 < \varphi < \infty$, ν_φ satisfies Assumption B.

The next two theorems make precise the statement, referred to in the introduction, that the use of an unbiased estimator is essentially equivalent to the use of the loss function L_u.

The following theorem applies equally well to the problem of estimating a scale parameter when the location parameter is known, but to be consistent with the notation of the rest of the paper, we state and prove it in the situation of Section 2.

THEOREM 3.1. *Let $L(\cdot)$ be any non-constant invariant loss function for the problem of estimating σ^k, $k > 0$. Suppose the best invariant estimator $\delta_0 = \varphi_0(\cdot)s^k$ for L satisfies $\varphi_0(z) = \varphi_u(z)$ a.e. (η) for all densities such that $f(\cdot \mid z) \varepsilon_0 C_0(0, \infty)$ a.e. (η). Then there are constants $\alpha > 0$, $\beta \geqq 0$ such that $L(y) = \alpha L_u(y) + \beta$.*

Conversely, suppose $L(y) = \alpha L_u(y) + \beta(\alpha > 0, \beta \geqq 0)$ and $f(\cdot \mid \cdot)$ satisfies Assumption A. Then $\varphi_0(z) = \varphi_u(z)$ a.e. (η).

PROOF. The converse is obvious. Using Assumption A

$$(3.4) \qquad (\partial/\partial\varphi)R(\varphi s^k \mid z) = \int s^k(1 - (\varphi s^k)^{-1})f(s \mid z)\,ds.$$

Hence $R(\varphi s^k \mid z)$ is uniquely minimized by choosing $\varphi = (\int s^k f(s \mid z)\,ds)^{-1} = \varphi_u$.

For the main result we suppose there exists an invariant loss function $L(\cdot)$ with $\varphi_0 = \varphi_u$ a.e. (η). For any probability density $g(\cdot) \varepsilon_0 C_0(0, \infty)$ satisfying $\int (s^k - 1)g(s)\,ds = 0$, $\varphi_u = 1$ and hence $\int g(s)\nu_1(ds) = 0$.

If $h \varepsilon_0 C_0(0, \infty)$ satisfies $\int (s^k - 1)h(s)\,ds = 0$ then h can be written as $h = \alpha_1 g_1 - \alpha_2 g_2$ where $\alpha_i \geqq 0$, and g_i are probability densities in $_0 C_0(0, \infty)$ satisfying $\int (s^k - 1)g_i(s)\,ds = 0$; $i = 1, 2$. (In general g_1 and g_2 will not be merely the positive and negative part of h.) Hence, for all $h \varepsilon_0 C_0(0, \infty)$

$$(3.5) \qquad \int (s^k - 1)h(s)\,ds = 0 \quad \text{implies} \quad \int h(s)\nu_1(ds) = 0.$$

As is well known, (3.5) implies that $\nu_1(ds)$ is equivalent to $\alpha(s^k - 1)\,ds$ on $(0, \infty)$ for some constant α; i.e. that for any interval $0 < a < b < \infty$, $\int_a^b \nu_1(ds) = \alpha \int_a^b (s^k - 1)\,ds$. Looking at the derivation (2.4) of ν_1 we see that this implies $d_v L(e^v)$ is equivalent to $\alpha(e^v - 1)\,dv$. This in turn implies that $L(y) = \alpha(y - \ln y - 1) + \beta$ for all but an at most countable set of points. The assumptions $L \geqq 0$, L non-constant, imply $\alpha > 0$, $\beta \geqq 0$. The assumption that L is right continuous gives equality everywhere. This completes the proof of the theorem.

In the main assertion of Theorem 3.1 we required that the best invariant estimator for L satisfies $\varphi_0 = \varphi_u$ (a.e. (η)) for all densities in a rather large family of densities (namely, $f(\cdot \mid z) \varepsilon_0 C_0(0, \infty)$). It should be clear that the conclusion will be valid even if the family of densities is much smaller; for example, as long as it contains enough densities so that their closure spans the subspace of $_0 C_0(0, \infty)$ consisting of all densities satisfying $\int (s^k - 1)g(s)\,ds = 0$.

We have tried to produce a small family of densities containing the densities of the χ^2 variables for which the conclusion of Theorem 3.1 is valid, however we have not succeeded in producing an adequate characterization of such a family. We offer instead the following theorem as an example of the kind of theorem which may be obtained.

THEOREM 3.2. *Let X_1, X_2 be independent random variables each having the*

density $\sigma^{-1}g((x - \mu)/\sigma)$; μ, σ unknown. (For convenience) let

$$(3.6) \qquad\qquad s = |x_2 - x_1|.$$

Suppose $L(\,\cdot\,)$ is a non-constant invariant loss function for the problem of estimating σ^k. Suppose the best invariant estimator $\delta_0 = \varphi_0 s^k$ for L satisfies $\varphi_0 = \varphi_\mu$ for all bounded densities g having compact support. Then there are constants $\alpha > 0$, $\beta \geqq 0$ such that $L(y) = \alpha L_u(y) + \beta$.

Conversely, suppose $L(y) = \alpha L_u(y) + \beta (\alpha > 0, \beta \geqq 0)$ and g is as above. Then $\varphi_0 = \varphi_u$.

PROOF. The converse follows from the converse of Theorem 3.1 using the fact $f(\,\cdot\,)$ (the density of the random variable S defined in (3.6)) satisfies Assumption A with L_u, since L_u satisfies Assumption B and g is bounded and has compact support.

Let

$$_c\Delta_d(x) = x - c, \qquad c \leqq x \leqq (c + d)/2,$$
$$= d - x, \qquad (c + d)/2 \leqq x \leqq d,$$
$$= 0, \qquad\qquad \text{otherwise.}$$

We claim that if $0 < \epsilon < c_2 - c_1$, $c_2 > 2 \times 2^{1/k}$

$$(3.7) \qquad \gamma_1\,_{c_1}\Delta_{(c_1+\epsilon)} + \gamma_2\,_{c_2}\Delta_{(c_2+\epsilon)} = q_{c_1,c_2}$$

satisfies $\int q(s)(s^k - 1)\,ds = 0$ then q is in the linear subspace of $_0C_0(0, \infty)$ spanned by those densities $\{f(\,\cdot\,)\}$ of S which also satsify $\int f(s)(s^k - 1)\,ds = 0$

To see this, let

$$\chi_{(c,d)}(x) = 1, \qquad c \leqq x \leqq d,$$
$$= 0, \qquad \text{otherwise,}$$

then define a probability density g by $g(y) = \sum_{i=3}^{5} \gamma_i \chi_{c_i,d_i}(y)$ where $c_3 = -\epsilon/2$, $d_3 = 0$, $c_4 = c_1$, $d_4 = c_1 + \epsilon/2$, $c_5 = c_2$, $d_5 = c_2 + \epsilon/2$ and γ_i, $i = 3, 4, 5$, determined subject to the two restrictions that (1) g is a probability density and (2) the density of S determined by g satisfies $\int f(s)(s^k - 1)\,ds = 0$, and hence $\int f(s)\nu_1(ds) = 0$ since $\varphi_0 = \varphi_u$. It can be shown by direct (but messy) calculations that this can be done. (It is necessary here that c_2 be sufficiently large; $c_2 > 2 \times 2^{1/k}$ is enough.) In the process of doing this one computes the probability density of S as

$$(3.8) \qquad f = \left(\sum_{i=3}^{5} \gamma_i\right)_0\Delta_{\epsilon/2} + 2 \sum \sum_{3 \leqq j < i \leqq 5} \gamma_i \gamma_j\,_{(c_j-d_i)}\Delta_{(d_j-c_i)}.$$

Note that $_{(c_4-d_3)}\Delta_{(d_4-c_3)} = {}_{c_1}\Delta_{(c_1+\epsilon)}$ and $_{(c_5-d_3)}\Delta_{(d_5-c_3)} = {}_{c_2}\Delta_{(c_2+\epsilon)}$.

A further computation shows that one can write functions of the form q_{c_1,c_2} (3.7) with $0 < c_1 < c_2 - \epsilon$, $c_2 > 2 \times 2^{1/k}$ and $\int q(s)(s^k - 1)\,ds = 0$ as finite linear combinations of densities of the form (3.8). Going one step further, one can then write any function of the form q_{c_1,c_2} with $0 < c_1 < c_2 - \epsilon$ (and $\int q(s)(s^k - 1)\,ds = 0$) as a finite linear combination of densities of the form

(3.8). These functions, q_{c_1,c_2} , are clearly dense in the uniform norm in the subspace of functions in $_0C_0\,(a,\mathrm{b})$ satisfying $\int h(s)(s^k - 1)\,ds = 0$. It follows that ν_1 satisfies the statement (3.5) in the proof of Theorem 3.1.

Thus, just as in Theorem 3.1 it now follows that $L(y) = \alpha L_u(y) + \beta, \alpha > 0$, $\beta \geq 0$. This completes the proof of the theorem.

These two theorems should not be interpreted as an argument that L_u is inherently a more reasonable loss function for this problem than squared error (L_s). Rather, the loss function to be used in any particular situation should depend on as accurate an assessment as possible of the actual losses which will be incurred in that situation. We feel, however, that theorems such as the above may aid in the choice of which loss function to use in some general situations.

In this vein, there is one more loss function (among the many intuitively reasonable ones) which we would like to discuss briefly. For each given value of σ the squared error loss function gives equal importance to equal differences between δ and σ, i.e. to equal values of $|\delta - \sigma|$. Since a scale parameter is being estimated it might seem reasonable to instead weigh equal proportional errors equally; i.e. to let L be a (bowl shaped) function of $|\log \delta/\sigma|$. A mathematically convenient loss function of this type is given by

$$(3.9) \qquad L_l(y) = (\log y)^2.$$

$\delta_l(z) = \varphi_l(z)s^k$ is the corresponding best invariant estimator.

It can be argued heuristically that the squared error loss function gives relatively little importance to errors of the type $\delta < \sigma$ and that, therefore, the factor φ_s will be "small." Comparatively, δ_l appears to give much more relative weight to errors of the type $\delta < \sigma$ and one would expect the multiplier φ_l to be large. The following theorem proves that these statements are correct, and also shows that L_u lies in between L_s and L_l in this regard. Part of this theorem was proven by Pitman (1938).

THEOREM 3.3. *Suppose Assumption A is satisfied for the three loss functions* L_s, L_u , *and* L_l . *Then* $\varphi_s(z) < \varphi_u(z) < \varphi_l(z)$ *a.e.* (η).

PROOF. Using Assumption A, $\varphi_s(z)$ is the solution to the equation $\int s^k(\varphi s^k - 1)f(s \,|\, z)\,ds = 0$. Since s^k and $\varphi s^k - 1$ are both strictly increasing functions of s

$$0 = \int s^k(\varphi_s(z)s^k - 1)f(s \,|\, z)\,ds$$
$$> \int s^k f(s \,|\, z)\,ds \int (\varphi_s(z)s^k - 1)f(s \,|\, z)\,ds.$$

Hence $\int (\varphi_s(z)s^k - 1)f(s \,|\, z)\,ds < 0$ which implies $\varphi_s(z) < \varphi_u(z)$.

Again using Assumption A, and this time using Jensen's inequality,

$$0 = (\varphi_l(z))^{-1}\int (\log \varphi_l(z) + \log s^k)f(s \,|\, z)\,ds$$
$$< (\varphi_l(z))^{-1}(\log \varphi_l(z) + \log \int s^k f(s \,|\, z)\,ds).$$

This implies

$$\varphi_u(z) = (\int s^k f(s \,|\, z)\,ds)^{-1} < \varphi_l(z).$$

This completes the proof of the theorem.

4. Inadmissibility results for normal random variables, and generalizations.
If $X_1, \cdots, X_n$ are independent normal random variables with mean μ and
variance σ^2 then $t = \bar{x}$ and s as defined in (2.1) are sufficient statistics. Hence
their distribution does not depend on z, and neither does the best invariant esti-
mator. However this fact does not play a role in the methods of this section.
What is crucial for the results of this section is the fact that T and S are inde-
pendent (conditionally on z).

Theorem 4.1 contains the condition that

$$(4.1) \qquad\qquad f_{0,1}(s, t \mid z) = f(s \mid z)h(t \mid z)$$

where, of course, both $f(\cdot \mid z)$ and $h(\cdot \mid z)$ are probability densities. Therefore
this theorem applies to the normal distribution case mentioned above. (The
other assumptions of the theorem are easily seen to be also satisfied in the normal
distribution case so long as $L(y)$ does not tend to infinity extremely fast as
$y \to 0$ or $y \to \infty$.)

As mentioned in the introduction the scale invariant estimators which we will
prove to be better than the best invariant estimator are of the form

$$(4.2) \qquad\qquad \delta_1(\cdot) = \theta(z)s^k \quad \text{if} \quad |t/s| < K(z)$$
$$= \varphi_0(z)s^k \quad \text{if} \quad |t/s| \geqq K(z)$$

where (as always) $\varphi_0(z)\,s^k$ is the best invariant estimator of σ^k, and $\theta(z) < \varphi_0(z)$.

We will have to consider various conditional densities; and so we define (for
example)

$$(4.3) \qquad\qquad f(\cdot \mid |t/s| < K, \mu, z)$$

to be the conditional probability density of S given $|t/s| < K$, and $Z = z$, for
the given value of μ (and for $\sigma = 1$). For any scale invariant δ we also define

$$(4.4) \qquad R(\delta; \mu \mid |t/s| < K, z) = E_{\mu,1}(L(\delta(\cdot)/\sigma^k) \mid |t/s| < K, Z = z).$$

It should be clear already that what we will do is to prove that for each given
value of Z

$$(4.5) \qquad R(\delta_1 ; \mu \mid z) = R(\delta_1 ; \mu \mid |t/s| < \infty, z) \leqq R(\delta_0 ; \mu \mid z)$$

with strict inequality for an appropriate set of values of μ, which in our case will
include a neighborhood of $\mu = 0$. (The first equality of (4.5) defines $R(\delta_1 ; \mu \mid z)$.)
Since we only need to prove (4.5) separately for each given value of Z it will be
convenient in the remainder of this section to assume that a value z of Z has
been given. We will make this *assumption* throughout this section except in the
statement of Theorem 4.1, and hence we will not write the given value of Z in
expressions like (4.1)–(4.5). For example, the expression (4.1) becomes
$f_{0,1}(s, t) = f(s)h(t)$.

The main idea of the proof is, roughly speaking, to show that
$f(s \mid |t/s| < K, \mu, z)$ has a uniform (in μ) strict monotone likelihood ratio property

with respect to $f(s \mid z)$. In order to get the necessary uniformity some assumption must be made concerning the density h (see Example 4.1). The following is stronger than necessary, but convenient.

ASSUMPTION (4.6). For the given value of Z there are constants $b_1 > 0$, $b_2 > 0$ such that

$$h(t) > b_2 \quad \text{for} \quad |t| \leqq b_1 \,,$$

(4.6) $\qquad h(t) \quad \text{is non-decreasing when} \quad t < -b_1 \,,$

$$h(t) \quad \text{is non-increasing when} \quad t > b_1 \,.$$

Let

(4.7) $$F(s) = \int_0^s f(y) \, dy,$$

(4.8) $$H(t) = \int_{-\infty}^t h(y) \, dy.$$

(Again, all these expressions depend on the given value of Z.)

We can now state the main theorem of this section.

THEOREM 4.1. *Suppose L is an invariant bowl-shaped loss function satisfying Assumption B. Suppose there is a set $\mathcal{Z}$ of values of Z such that $\eta(\mathcal{Z}) > 0$ and for each $z \, \varepsilon \, \mathcal{Z}$:*

(a) S and T are conditionally independent (i.e. (4.1) is valid), with $f(\cdot \mid z)$ a bounded continuous function,

(b) h satisfies the assumption (4.6), and

(c) Assumption A is satisfied.

Let $K(z)$ be any measurable function such that for $z \, \varepsilon \, \mathcal{Z}$, $K(z) = 0$, and for each $z \, \varepsilon \, \mathcal{Z}$, $0 < K(z) < \infty$, satisfies

(4.9) $$\int_0^{b_1(z)/K(z)} f(s \mid z) |\nu_{\varphi_0}(ds)| > 0, \qquad \varphi_0 = \varphi_0(z).$$

Then there is a measurable $\theta(z) < \varphi_0(z)$ such that for each $z \, \varepsilon \, \mathcal{Z}$

(4.10) $$R(\delta_1 \, ; \mu \mid z) \leqq R(\delta_0 \, ; \mu \mid z)$$

with strict inequality for all μ such that $\mathrm{Pr}_{\mu,1}\{|t/s| < K(z)|z\} > 0$. ($\delta_1$ is defined in (4.2).)

Hence

(4.11) $$R(\delta_1 \, ; \mu, \sigma) \leqq R(\delta_0 \, ; \mu, \sigma)$$

with strict inequality for all values of μ, σ such that $|\mu/\sigma|$ lies in a sufficiently small neighborhood of the origin. (Hence δ_0 is inadmissible.)

NOTE. (4.9) will certainly be satisfied if L is strictly bowl-shaped (so that $|\nu_{\varphi_0}(\cdot)|$ is strictly positive) and for the given $z \, \varepsilon \, \mathcal{Z}$ $F(b_1/K) > 0$.

PROOF. First of all, note that for $z \, \varepsilon \, \mathcal{Z}$ the condition $\mathrm{Pr}_{\mu,1}\{|t/s| < K(z)|Z\} > 0$ will certainly be satisfied for all μ for which $h(-\mu) > 0$; and hence, using Assumption 4.6, for all μ in a neighborhood of the origin. Thus (4.11) follows immediately from (4.10), and the fact that $R(\delta_i \, ; \mu, \sigma)$, $i = 0, 1$, depends only on the ratio μ/σ, and the fact that $\eta(\mathcal{Z}) > 0$.

In the remainder we have only to prove (4.10). As noted in the beginning of this section we choose a value $z \, \varepsilon \, \mathsf{Z}$, and will prove (4.10) for this fixed value. Also we will consider only values of μ such that $\mathrm{Pr}_{\mu,1}\{|t/s| < K\} > 0$. (For values of μ for which $\mathrm{Pr}_{\mu,1}\{|t/s| < K\} = 0$, $\delta_1 = \delta_0$ with probability one and hence there is equality in (4.10).)

For such values of μ

$$(4.12) \quad f(s \,|\, |t/s| < K, \mu) = c(K, \mu)(H(Ks - \mu) - H(-Ks - \mu))f(s).$$

Define

$$(4.13) \quad r(s, \mu) = K[h(Ks - \mu) + h(-Ks - \mu)][H(Ks - \mu) - H(-Ks - \mu)]^{-1}$$

$$= (\partial/\partial s) \log (H(Ks - \mu) - H(-Ks - \mu)).$$

Using the Assumption (4.6), note that $r(s, \mu)$ satisfies:
 (a) if $|Ks - \mu| < b_1$ or $|-Ks - \mu| < b_1$ then $r(s, \mu) > Kb_2$
or
 (b) if $(Ks - \mu) < -b_1$ or $(-Ks - \mu) > b_1$ then (using the "monotone tails" p operty of h)

$$r(s, \mu) \geqq K[(2Ks)^{-1}(H(Ks - \mu) - H(-Ks - \mu))]$$

$$\cdot [(H(Ks - \mu) - H(-Ks - \mu)]^{-1} = (2s)^{-1}.$$

For $0 < s < b_1/K$ the situation $-Ks - \mu < -b_1$ and $Ks - \mu > b_1$ is impossible, and therefore either (a) or (b) above is valid.

Define

$$(4.14) \quad r_0(s) = \min (Kb_2, (2s)^{-1}), \quad 0 < s < b_1/K,$$

$$= 0, \quad\quad\quad\quad\quad s \geqq b_1/K.$$

It follows from (a), (b) above that for $0 < s < b_1/K$, $\inf_\mu r(s, \mu) \geqq r_0(s) > 0$. Let

$$(4.15) \quad\quad\quad\quad r^*(s) = \exp \{\int_0^s r_0(y) \, dy\}$$

so that $r_0(s) = (\partial/\partial s) \ln r^*(s)$. Finally define the probability density f^* by

$$(4.16) \quad\quad\quad\quad f^*(s) = cr^*(s)f(s)$$

where c is the appropriate constant.

Since r^* is bounded it follows that f^* is bounded. Since L is bowl shaped $\nu_{\varphi_0}(\cdot)$ is a non-positive measure on $(0, a_0)$ and a non-negative measure on (a_0, ∞) where $a_0 = (a_L/\varphi_0)^{1/k}$. ($a_L$ is defined above Lemma 2.1.)

Since $r^*(s)$ is non-decreasing, f^* has monotone likelihood ratio with respect to f. In addition, f^* has strict monotone likelihood ratio with respect to f on $(0, b_1/K)$. We may use these likelihood ratio properties and condition (4.9) to obtain strict inequality in the equation

$$(4.17) \quad\quad\quad \int f^*(s)\nu_{\varphi_0}(ds) > \int f(s)\nu_{\varphi_0}(ds) = 0.$$

Since f^* is bounded and continuous and L satisfies Assumption B, Lemma 2.1 applies; hence $\int f^*(s)\nu_\varphi(ds)$ is lower semi-continuous in φ at φ_0. It follows from (4.17) that there is a number $\theta < \varphi_0$ such that $|\theta - \varphi_0| < \epsilon$ (with ϵ as in Assumption A) and such that $\theta \leqq \varphi \leqq \varphi_0$ implies

$$(4.18) \qquad \int f^*(s)\nu_\varphi(ds) > 0.$$

Finally we note from the construction of f^* that each density $f(\cdot \mid |t/s| < K, \mu)$ has non-decreasing monotone likelihood ratio with respect to f^*. In addition, using the expression (4.12) and the fact that H is continuous and bounded the densities $f(\cdot \mid |t/s| < K, \mu)$ satisfy Assumption A for $|\varphi - \varphi_0| < \epsilon$ (ϵ as above).

Using these properties of $f(\cdot \mid \cdot, \cdot)$ and (4.18), for $\theta \leqq \varphi \leqq \varphi_0$

$$(4.19) \quad (\partial/\partial\psi)R(\psi s^k; \mu \mid |t/s| < K)|_{\psi=\varphi} = \int f(s \mid |t/s| < K, \mu)\nu_\varphi(ds) > 0.$$

It follows from (4.19) that

$$R(\theta s^k; \mu \mid |t/s| < K) < R(\varphi_0 s^k \mid |t/s| < K).$$

This verifies (4.10) for the chosen value of Z; and completes the proof of the theorem.

The assumption that $f(\cdot \mid z)$ be continuous and bounded for $z \varepsilon Z$ is needed in Theorem 4.1(a) only so that $\int f^*(s)\nu_\varphi(ds)$ be lower semi-continuous at $\varphi_0(z)$. If L is absolutely continuous then it follows from the remark at the end of Section 2 that f^* bounded is sufficient for $\int f^*(s)\nu_\varphi(ds)$ to have this property. Hence if L is absolutely continuous the assumption (a) of the theorem may be weakened to require only that $f(\cdot \mid z)$ is bounded.

We conclude this section with an example showing that if an estimator like δ_1 is to be better than δ_0 then some assumption must be made concerning the shape of the density h (although the Assumption 4.6 is clearly stronger than necessary). We do not know whether the δ_0 of this example is admissible within the class of scale invariant estimators, nor whether it is admissible.

EXAMPLE 4.1. Let $f(s) = 1, 0 \leqq s \leqq 1, = 0$ otherwise. Let

$$h(t) = 3 \times 2^{-i} \quad 2^i - 2^{-i-2} < |t| < 2^i + 2^{-i-2}, i = 1, 2, \cdots,$$

$$= 0 \qquad \text{otherwise.}$$

(h places a small amount of probability increasingly concentrated (as $t \to \pm \infty$) near a very sparse sequence of points $- \pm 2, \pm 4, \pm 8, \pm 16, \cdots$).

Using the expression (4.12) it is clear that for any fixed $K < \infty$

$$\lim_{i\to\infty} f(s \mid |t/s| < K, \pm 2^i) = f(s).$$

Hence, if L is any loss function such that φ_0 is uniquely determined and δ_1 is any estimator of the form (4.2) with $\theta < \varphi_0$ then

$$\lim_{i\to\infty} R(\delta_1; \pm 2^i \mid |t/s| < K) = R(\theta s^k) > R(\varphi_0 s^k).$$

Hence δ_1 is not better than δ_0.

The fact that f has compact support seems to be important in this particular

example. However there do exist analagous (though messier) examples where f does not have compact support. What seems to be necessary is roughly that the tails of h look somewhat like those in the example and contain much more probability than the tail of f (with the ratio tending swiftly to ∞ as $|t| = s \to \infty$).

5. Inadmissibility results in the general case with compact support. In this section we prove inadmissibility theorems analogous to Theorem 4.1, but without the condition that the random variables S and T be conditionally independent. As we showed in Example 4.1 even when S and T are conditionally independent awkward things can happen if the tails of the distribution of T behave badly. This is even more true when S and T need not be conditionally independent. Therefore in this section we are able to deal only with the case where the distribution of T has no tails at all.

We assume here exists a $B < \infty$ such that $|t| > B$ implies

$$(5.1) \qquad\qquad f_{0,1}(s, t) = 0.$$

At the end of this section we briefly discuss some aspects of the situation when Assumption 5.1 is not satisfied.

We assume throughout this section, as in the bulk of Section 4, that we are dealing only with a fixed value of Z, and thus the value of Z will not be displayed in the notation except in Corollary 5.1. It should be clear from the remarks in Section 4 and from Theorem 4.1 how to deal with the case where Z is not fixed.

The careful reader will note that several of the assumptions in the following theorem are stronger than the corresponding assumptions in Theorem 4.1, and that the conclusion of the theorem is slightly weaker than the conclusion of that theorem.

We remind the reader that (for a fixed value of z)

$$\delta_1 = \theta s^k \quad \text{if} \quad |t/s| < K$$
$$ = \varphi_0 s^k \quad \text{if} \quad |t/s| \geqq K.$$

The intuitive idea which lies behind the proof of the following theorem is that when K is large the line $|t/s| = K$ is (locally in t) almost horizontal (away from the point $s = 0$). Because of the compact support of f; it follows that for all μ sufficiently large ($|\mu| > U$) the region $|t/s| < K$ is practically the same as the region $s > |\mu|/K$; and this approximate identity of regions becomes better and better as $K \to \infty$ *uniformly in* μ (for $|\mu| > U$).

Conditionally, given $s > |\mu|/K$ for $|\mu| > U$, it is clear that the estimator $\varphi_0 s^k$ is not admissible; rather, there is a better estimator of the form θs^k with $\theta < \varphi_0$. (To use the language of the preceding section the distributions of S given $s > |\mu|/K$, $|\mu| > U$, have a uniform monotone likelihood ratio property with respect to the unconditional distribution of S.)

The first part of the proof of Theorem 5.1 is devoted to showing essentially that when K is sufficiently large one need not worry about values of μ with $|\mu| < U$ (U is defined in (5.8)). Then the second part of the proof verifies in the appropri-

ate precise sense that the region $|t/s| < K$ is practically the same as the region $s > |\mu|/K$ for $|\mu| > U$.

THEOREM 5.1. *Suppose L is strictly bowl-shaped, is continuous at its minimum value (a_L) and satisfies Assumption B. Suppose the condition (5.1) is satisfied. Suppose $f(x)$ is bounded, continuous, satisfies Assumption A and satisfies*

$$(5.2) \qquad \int_0^\xi f(s)\, ds > 0 \quad \text{for all} \quad \xi > 0.$$

Suppose also that for each fixed $K > 0$ the densities $f(s \mid |t/s| < K, \mu)$ are jointly continuous in s, μ for all s, μ such that

$$\Pr_{\mu,1}\{|t/s| < K\} > 0.$$

Then there is a $K_0 < \infty$ such that $K > K_0$ implies there is a $\theta < \varphi_0$ (θ depends on K) such that $R(\delta_1 ; \mu) \leq R(\delta_0 ; \mu)$ with strict inequality whenever

$$(5.3) \qquad \Pr_{\mu,1}\{s, t{:}|t/s| < K\} > 0.$$

PROOF. Again we need consider only values of K, μ such that (5.3) is satisfied. If $|\varphi - \varphi_0| < \epsilon$, then the conditions on $f(\cdot)$ and $f(\cdot \mid |t/s| < K, \mu)$ guarantee that the latter satisfies Assumption A. Hence

$$(5.4) \quad (\partial/\partial\varphi)R(\varphi s^k; \mu \mid |t/s| < K) = \int f(s \mid |t/s| < K, \mu)\nu_\varphi(ds) = D(\varphi, K, \mu)$$

where (5.4) defines $D(\cdot, \cdot, \cdot)$.

Let $a_0 = (a_L/\varphi_0)^{1/k}$, so that $L(\varphi_0 s^k)$ attains its minimum value when $s = a_0$. We first prove that

$$(5.5) \quad \text{For each} \ \mu, \qquad K > K_0(\mu) = (B + |\mu|)/a_0 \Rightarrow D(\varphi_0, K, \mu) > 0$$

(for all K, μ satisfying (5.3)).

$|t/s| \geq K > K_0(\mu)$ implies $s < a_0$. Since L is strictly bowl shaped, $\nu_{\varphi_0}(\cdot)$ is a strictly negative measure on this region. Hence for $K > K_0(\mu)$

$$(5.6) \qquad \int f(s \mid |t/s| \geq K, \mu)\nu_{\varphi_0}(ds) < 0.$$

Condition (5.2) guarantees that $\Pr_{\mu,1}\{|t/s| \geq K\} > 0$ so that (5.6) makes sense. Then,

$$
\begin{aligned}
(5.7) \quad 0 &= D(\varphi_0, \infty, \mu) \\
&= \Pr_{\mu,1}\{|t/s| \geq K\} \int f(s \mid |t/s| \geq K, \mu)\nu_{\varphi_0}(ds) \\
&\quad + \Pr_{\mu,1}\{|t/s| < K\}D(\varphi_0, K, \mu).
\end{aligned}
$$

(5.6) and (5.7) prove that (5.5) is valid.

We next prove

$$(5.8) \quad \exists U \quad \text{such that} \quad |\mu| > U \Rightarrow D(\varphi_0, K, \mu) > 0 \quad \text{for all} \quad K > 0,$$

and μ satisfying 5.3.

First, choose $a' > a_0$ such that for some $\alpha > 0$, $\int_{a'}^\infty f(s)\nu_{\varphi_0}(ds) > \alpha > 0$. Then choose $a'' < a_0$ such that $\int_{a''}^{a_0} f(s)\nu_{\varphi_0}(ds) > -\alpha/2$. That a' and a'' exists is a

consequence of the Lebesgue regularity of ν_{φ_0} and the fact that ν_{φ_0} is non-atomic at a_0 (since L is continuous at a_L). Choose

$$(5.9) \qquad U = \max \left(B(a' + a_0)/(a' - a_0); B(a_0 + a'')/(a_0 - a'') \right).$$

A little algebra shows that for any K, $|\mu| > U > B$ implies either

$$(5.10a) \qquad\qquad (|\mu| - B)/a_0 > K \qquad \text{or}$$

$$(5.10b) \qquad\qquad (|\mu| + B)/a_0 < K \qquad \text{or}$$

$$(5.10c) \qquad (|\mu| - B)/a'' > K \quad \text{and} \quad (|\mu| + B)/a' < K.$$

Note that (5.3) is automatically satisfied if (5.10b) or (5.10c) are true.

If (5.10a) is true, then (using (5.1)) with probability one

$$K < (|\mu| - B)/a_0 \leqq |t|/a_0 .$$

Hence, with probability one, the condition $|t/s| < K$ implies $s > a_0$ in this case. $\nu_{\varphi_0}(\,\cdot\,)$ is a strictly positive measure on this region; hence (5.8) is true in this case.

If (5.10b) is true then (using (5.1)) with probability one $|t/s| \geqq K$ implies $s < a_0$. As in the verification of (5.5) it follows that (5.8) is also true in this case.

Finally, if (5.10c) is true then

$$f(s \mid |t/s| < K, \mu) = c \int_{|t/s| < K} f_{0,1}(s, t - \mu)\, dt$$

$$(5.11) \qquad\qquad \begin{cases} = cf(s), & \text{if } s > a', \\ \leqq cf(s), & \text{if } a'' \leqq s \leqq a', \\ = 0, & \text{if } s < a''. \end{cases}$$

(where c is the appropriate normalizing constant). Hence, using the definition of a_0, a', a'',

$$(5.12) \qquad D(\varphi_0, K, \mu) \geqq c \int_{a'}^{\infty} f(s)\nu_{\varphi_0}(ds) + c \int_{a''}^{a_0} f(s)\nu_{\varphi_0}(ds)$$

$$\geqq c(\alpha - \alpha/2) > 0.$$

(5.12) proves that (5.8) is also true in this case. Hence we have proved the statement (5.8).

Let

$$K_0 = K_0(U) = \sup_{|\mu| \leqq U} K_0(\mu).$$

Using (5.5) and (5.8) for all $K > K_0$ and all μ such that μ, K satisfy (5.3), $D(\varphi_0, K, \mu) > 0$.

Choose any $K > K_0$.

We claim there is a number θ, $\varphi_0 - \epsilon < \theta < \varphi_0$, such that $\theta \leqq \psi \leqq \varphi_0$ implies $D(\psi, K, \mu) > 0$ for all μ such that (5.3) is satisfied. Suppose not.

Then there is a sequence ψ_i, μ_i such that (5.3) is satisfied, $\psi_i \nearrow \varphi_0$ and $D(\psi_i, K, \mu_i) \leqq 0$.

Reasoning as in the verification of (5.8) in the case (5.10a), if $(|\mu_i| - B)/a_i > K$ where $a_i = (a_L/\psi_i)^{1/k}$ and (5.3) is satisfied then $D(\psi_i, K, \mu_i) > 0$.

It follows that $|\mu_i| < Ka_i + B \to Ka_0 + B$. By taking subsequences we may assume $\mu_i \to \mu_0$ where $|\mu_0| \leqq Ka_0 + B$. Thus $(|\mu_0| - B)/a_0 \leqq K$, and it follows (since $K > K_0(U)$) that $|\mu_0|$ satisfies (5.10b) or (5.10c) and hence that μ_0, K satisfy (5.3). The results (5.5), (5.8) thus establish that $D(\varphi_0, K, \mu_0) > 0$.

Lemma 2.1 now applies here (with $f(s \mid |t/s| < K, \mu)$ substituted for $g(s, \mu)$ in expression 2.9). Thus

$$\liminf_{\psi_i \to \varphi_0, \mu_i \to \mu_0} D(\psi_i, K, \mu_i) \geqq D(\varphi_0, K, \mu_0).$$

But this implies $D(\varphi_0, K, \mu_0) \leqq 0$ which is a contradiction. Thus we have established our claim, namely that there is a $\theta < \varphi_0$ such that $\theta \leqq \psi \leqq \varphi_0$ implies $D(\psi, K, \mu) > 0$ for all μ such that (5.3) is satisfied. It follows that for these values of μ, $R(\delta_1 ; \mu \mid |t/s| < K) < R(\delta_0 ; \mu \mid |t/s| < K)$. The main assertion of the theorem follows immediately; and the proof of the theorem is complete.

The condition in Theorem 5.1 that $f(s \mid |t/s| < K, \mu)$ be jointly continuous in s, μ is undoubtedly stronger than is necessary for the conclusion of the theorem to be valid. (To weaken the condition and retain the same type of proof would require that the hypotheses of Lemma 2.1 be weakened. This can probably be done.) Nevertheless the hypotheses of Theorem 5.1 are weak enough to apply to most practical situations in which the compact support condition (5.1) applies. The following corollary gives the most important such application.

COROLLARY 5.1. *Let $X_1, \cdots, X_n$ be independent, real-valued random variables each with density $\sigma^{-1} g((x - \mu)/\sigma)$ with respect to Lebesgue measure. Let $t = \bar{x}$, s, z be defined as in (2.1). Suppose $g(\cdot)$ is bounded and has compact support (i.e. $g(y) = 0$ for $|y| > B'$). Suppose L is strictly bowl-shaped, is continuous at its minimum value and satisfies Assumption B.*

Then there is a measurable function $K_0(z) < \infty$ such that for any measurable $K(z) > K_0(z)$ there is a measurable $\theta(z) < \varphi_0(z)$ such that $R(\delta_1 ; \mu \mid |t/s| < K, z) \leqq R(\delta_0 ; \mu \mid |t/s| < K, z)$ with strict inequality for all μ in a neighborhood of $\mu = 0$.

PROOF. The proof consists in verifying that the assumptions of the corollary imply that for each fixed value of z the assumptions of the theorem are satisfied. The hardest condition to check is that $f(s \mid |t/s| < K, \mu, z)$ is jointly continuous in s, μ for each fixed K, z; but this can be verified directly by writing down the explicit expression for f and using elementary properties of the Lebesgue integral. We omit the details.

After this verification the conclusion of the corollary then follows from a direct application of the theorem.

In Example 4.1 we gave one example of a situation in which the compact support assumption (5.1) was not valid and in which the specific conclusion of our Theorems 4.1 and 5.1 was false. If S and T need not be conditionally independent

it is much easier to give such counter-examples. Consider the following:

EXAMPLE 5.1. Let

$$f_{0,1}(s, t) = 1 \quad \text{if} \quad 0 \leqq s \leqq (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}t^2}$$

$$= 0 \quad \text{otherwise.}$$

Note that for any $K < \infty$

$$\sup \{s : |t/s| < K, 0 \leqq s \leqq (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}(t-\mu)^2}\} \to 0 \quad \text{as} \quad \mu \to \pm\infty.$$

Hence if L is any strictly bowl-shaped loss function then for $|\mu|$ sufficiently large and $\varphi \leqq \varphi_0$, $D(\varphi, K, \mu) < 0$ where D is as defined in (5.4). It follows that no estimator like δ_1 can be better than δ_0.

In spite of these examples it seems reasonable to conjecture that a result like Corollary 5.1 should be true without the assumption that $g(\cdot)$ has compact support. Simply the fact that S and T result from independent random variables $X_1, \cdots, X_n$ imposes certain kinds of restrictions on the tails (in t) of the density $f(s, t \mid z)$ which seem to preclude examples like Examples 4.1 and 5.1. (Some preliminary investigations have indicated to us that in the proof of such an inadmissibility result it may be more convenient to define $t = x_1$ and $s = |x_2 - x_1|$ rather than $t = \bar{x}$, etc.)

We have only been able to prove a result like Corollary 5.1 with non-compact support in the special case where $n = 2$, g is symmetric and unimodal, and $K = 2^{-\frac{1}{2}}$ (when $K = 2^{-\frac{1}{2}}$, $\{(x_1, x_2) : |t/s| < K\} = \{(x_1, x_2) : x_1 < 0 < x_2$ or $x_1 > 0 > x_2\}$). Since our proof of this result seems to make strong use of all these special assumptions we do not take space to give it here.

6. Numerical results for estimating the variance of a normal population. In order to get some idea as to the improvement which one may expect from the use of estimators better than the usual one we computed the risks in a few special cases of an estimator much like those developed in Section 4. It should be emphasized again that the scale invariant estimators we have used are themselves inadmissible, and we do not at all know how much the risks of these estimators may be decreased by the use of admissible estimators.

We deal with the case where $X_1, \cdots, X_n$ are independent normal random variables with unknown mean μ and variance σ^2, and t and s are as defined in (2.1). Then T is normal with mean μ and variance $1/n$ and $(n - 1)s^2$ is χ^2 with $n - 1$ degrees of freedom. Rather than using the exact proof in Section 4, it is easier to note that in this case

$$(6.1) \quad f(s \mid |t/s| < K, \mu) = c(K, \mu)(H(Ks - \mu) - H(-Ks - \mu))f(s)$$

has strict monotone likelihood ratio with respect to $f(s \mid |t/s| < K, 0)$ (H and f are as in (4.1), (4.8)). This fact follows directly from well known likelihood properties of the normal distribution function. (Note that $H(x) = \Phi(n^{\frac{1}{2}}x)$). Hence for any appropriately differentiable, convex loss function L and for any $K > 0$

the θ in the definition of δ_1 (4.2) may be chosen as the unique solution of

$$(6.2) \qquad \int s^k W'(\theta s^k)(H(Ks) - H(-Ks))f(s)\, ds = 0.$$

Formula (6.1) can then be used to compute the risk of δ_1 and this risk can be compared with the risk of δ_0.

All the numerical results in this section are for the problem of estimating the variance, i.e. $k = 2$.

For the squared error loss function we computed the risks of δ_1 for $n = 2$ and $n = 10$ for several values of K.

For $n = 2$, $\varphi_0 = \frac{1}{3}$ and $R(\delta_0 ; \mu, 1) = \frac{2}{3}$. The best choice of K overall that we found was $K = .5$. (This choice was best among values of $K = 0(.1)1.0$.) For $K = .5$, θ as in (6.2) was $\theta = .2794(6)$ and the risks of δ_1 were

μ	$=$	0	.50	1.0	1.5
$R(\delta_1 ; \mu, 1)$	$=$	.6595	.6569	.6562	.6579

This represents a maximum improvement of $(.6667 - .6562)/.6667 = 1.6\%$ over the usual estimator at $|\mu|/\sigma = 1$, and an improvement of 1.1% at $\mu = 0$.

For $N = 10$ we found $K = .2$ as a reasonable value of K. Then $\varphi_0 = 9/11 = .81818$ while $\theta = .7602(3)$. $R(\delta_0 ; \mu, 1) = 2/11 = .181818$ while

$\mu =$	0	.5	1.0
$R(\delta_1 ; \mu, 1) =$	.1796	.1804	.181

Thus the maximum improvement we found was $(.1818 - .1796)/.1818 = 1.2\%$ at $\mu = 0$. As one would expect, the improvement of δ_1 over δ_0 falls away to zero faster in this case than in the case $K = 2$.

We also carried out some computations for the same problem when the loss function is L_μ (defined in (3.3)), and when $n = 2$. In this case, of course, $\varphi_0 = 1$. When $K = .7$, $\theta = .60949$. (Note that $\mathrm{Pr}_{0,1}\{|t/s| < .7\} = .49+$. Hence δ_0 and δ_1 are very different—almost $\frac{1}{2}$ of the time δ_1 is only 61% as large as δ_0.) $R(\delta_0 ; \mu, 1) = .924$ and

μ	$=$	0	.5	1.0	1.5
$R(\delta_1 ; \mu, 1)$	$=$	.852	.846	.842	.860

The largest improvement here is $(.924 - .842)/.924 = 8.9\%$, for $\mu = 1$. For $\mu = 0$ the improvement is 7.8%. (The choice $K = .5$ will give an improvement of 8.4% at $\mu = 0$, but only 7.5% at $\mu = 1$.)

REFERENCES

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1135.

BROWN, L. D. (1967). The conditional level of Student's *t*-test. *Ann. Math. Statist.* **38** 1068–1071.

FARRELL, R. (1964). Estimators of a location parameter in the absolutely continuous case. *Ann. Math. Statist.* **35** 949–999.

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.

PITMAN, E. J. G. (1938). The estimation of location and scale parameters of a continuous population of any given form. *Biometrika* **30** 391–421.

WIENER, N. (1933). *The Fourier Integral and Certain of Its Applications*. Cambridge Press, London.