ON THE ADMISSIBILITY OF INVARIANT ESTIMATORS OF ONE OR MORE LOCATION PARAMETERS[1]

By Lawrence David Brown

Cornell University, Ithaca

TABLE OF CONTENTS

0. Introduction. Many statistical estimation problems possess certain natural symmetries. The location parameter estimation problem is an important example. It is symmetric, or, to use the usual terminology, invariant with respect to translations of the sample space. This strongly suggests that the statistician should use an estimation procedure which also has the property of being invariant.

Until recently it seemed reasonable to expect that the best invariant estimator is a "good" estimator. In particular, it seemed reasonable to expect that it is admissible—that is, that no other estimator has a risk which is never larger and is sometimes smaller than the risk of the best invariant estimator. However, Stein (1956) gave an example of a very simple problem in which the best invariant estimator is inadmissible. Previously several authors had proven admissibility in different problems, and there has been much research in the area since then. Several references are contained in the bibliography of this paper.

Received 17 December 1964; revised 12 January 1966.

[1] The preparation of this manuscript was supported in part by the Office of Naval Research contract number Nonr-266(04). This manuscript is a revised version of the author's thesis presented to Cornell University in partial fulfillment of the requirements for a Ph.D. degree.

The published results of C. Stein [Stein (1956), (1959), (1960)] are limited to the case where the loss function is the square of the distance between the estimate and the true parameter value. With minor restrictions on the probability densities involved he proved admissibility of the best invariant estimator if the location parameter is one or two dimensional, and he proved its inadmissibility if the location parameter is of dimension three or greater.

Primarily because of its mathematical simplicity, the loss function considered by Stein is the one traditionally used in such problems; see, for instance, Cramér (1946), p. 179, p. 473. However, there are many other loss functions which it is quite reasonable to use; for example, the distance between the estimate and the true value. It is natural to ask whether the dichotomy observed by Stein, with dividing line between two and three dimensions, depends on the choice of the loss function.

At the same time the question of admissibility can be generalized from a fixed sample size problem to a sequential problem. Such a generalization has been occasionally considered in the past but results have been obtained only in very special cases.

In this paper we consider both the fixed sample size and sequential problems for a one-dimensional or a three or higher dimensional location parameter, and make very few restrictions on the distribution of the random variables involved or on the loss function. A paper dealing in similar generality with the two dimensional case is now in preparation. (If the reader desires, the results in the three dimensional case may be read after Section 1.1, which contains the necessary definitions. Section three of the paper is logically independent of Section two.)

It is shown here that for the most part the dichotomy between three or more dimensional cases and fewer than three dimensional cases persists. There are a few minor exceptions; see Sections 2.4 and 3.2 of this paper, and also the proof in Farrell (1964) that if the best invariant estimator is not uniquely determined (almost everywhere) then it is inadmissible.

It is tempting to regard these exceptions as being very minor and to conjecture that admissibility of the best invariant estimator in any problem is primarily a property of the sample space and the group of symmetries which acts upon it. However, one should bear in mind that, at least when nuisance parameters are present, the situation is not so simple. Blackwell (1951) and Stein (1960), p. 375, point out an example of a problem in which the best invariant estimator is admissible when its form would be independent of the value of the nuisance parameters if these values were known, but in which it is inadmissible in some cases where its form if the value of the nuisance parameters were known is dependent on the value of those parameters. We have an example (which will appear elsewhere) of an estimation problem in which the form of the best invariant estimator remains the same independently of any knowledge concerning the value of the nuisance parameters, and in which that estimator is sometimes admissible and sometimes inadmissible, depending on the variance-covariance matrix of the observed random variables (which is assumed to be known).

The main theorems of this paper are Theorems 2.1.1, 3.1.1, 3.3.1, and 3.3.2. We have tried to state the assumptions of these theorems in as widely applicable and as weak a form as is consistent with their being meaningful and manageable. Where it is necessary, lemmas have been stated and proved which interpret the assumptions of the main theorems in a wide variety of standard statistical situations. If the reader wishes, these lemmas (in Sections 2.2, 2.3, and 3.2) may be read before the proofs of the main theorems.

Stating the theorems in this way, of course, emphasizes the uniformity of the dichotomy mentioned previously. More importantly, we hope it will enable the essential nature of the theorems and their proofs to be more easily and more clearly understood.

1. Preliminaries. 1.1 *General notation.* We first give the notation for the general problem which is considered in Sections 2 and 3. In Section 1.2 it is shown how this general problem is related to other more specialized problems which are also considered in various parts of Chapters 2 and 3; and some simple observations are made concerning the formulation of the general problem.

Let $H = E_m \times E_m \times \cdots$ (E_m denotes m-dimensional Euclidean space). Let $\mathcal{A}_i$ be the σ-field of subsets of H which are Lebesgue cylinder sets on the ith m-dimensional co-ordinate subspace of H. Let $\mathcal{A}^{(n)} = \mathcal{B}(\bigcup_{i=1}^{n} \mathcal{A}_i)$ where $\mathcal{B}(\mathcal{C})$ denotes the smallest σ field containing $\mathcal{C}$. Let $X, Y_1, Y_2, \cdots$ be a random variable on $H, \mathcal{A}^{(\infty)}$ such that X (which takes values in E_m) is measurable with respect to $\mathcal{A}_1$, and Y_i (which also takes values in E_m) is measurable with respect to $\mathcal{A}_{i+1}$, $i = 1, 2, \cdots$. Denote the values taken on by $X, Y_1, Y_2, \cdots$ as $x, y_1, y_2, \cdots$.

Let $Y = Y_1, Y_2, \cdots$ and $\theta \varepsilon E_m$. If $S \varepsilon \mathcal{A}^{(n)}$ assume

$$\Pr\{(X, Y) \varepsilon S\} = \int \int_S p_n(x - \theta, y)\, dx\mu(dy), \tag{1.1.1}$$

where $\int_{E_m} p_n(x, y)\, dx = 1$ for almost all $y(\mu)$. In (1.1.1) μ is a probability measure with respect to the field $\mathcal{B}(\bigcup_{i=2}^{\infty} \mathcal{A}_i)$, $p_n(x, y)$, the conditional probability density of X given the value $y_1, y_2, \cdots, y_{n-1}$ of Y is measurable with respect to $\mathcal{A}^{(n)}$ and $\int dx$ represents integration with respect to Lebesgue measure. Thus we assume throughout that *the conditional distribution of* X given $Y_1, Y_2, \cdots, Y_{n-1}$ *is absolutely continuous with respect to Lebesgue measure* a.e. (μ). It is not assumed that the distribution of X given $Y_1, Y_2, \cdots$ is absolutely continuous with respect to Lebesgue measure. Nevertheless that distribution will be treated in some ways as if it were absolutely continuous. The reason this can be done is contained in the following paragraph.

Let $\pi(S, Y)$ denote the conditional probability that $X \varepsilon S$ given Y, and $\theta = 0$. Let $g: H \to E_1$. g will be called *finitely measurable* (with respect to $\pi \cdot \mu$) if for any $\epsilon > 0$ there is an $n < \infty$ and an $\mathcal{A}^{(n)}$ measurable set $S \subset H$ such that $\int \pi(S, y)\mu(dy) > 1 - \epsilon$ and g restricted to $S(g \mid S)$ is measurable with respect to $\mathcal{A}^{(n)} \mid S$. If $g \geqq 0$ is finitely measurable then there exists $S_1 \subset S_2 \subset \cdots \subset H$ such that $\int \pi(S_n, y)\mu(dy) \to 1$ as $n \to \infty$ and $g \mid S_n$ is measurable with respect to $\mathcal{A}^{(n)} \mid S_n$. Thus if $g \geqq 0$ is finitely measurable define

$$(1.1.2)\quad \int\int g(x, y)p(x, y)\, dx\mu(dy) = \lim_{n\to\infty} \int\int_{S_n} g(x, y)p_n(x, y)\, dx\mu(dy)$$
$$= \int\int g(x, y)\pi(dx, y)\mu(dy) \leqq \infty,$$

where $\{S_n\}$ is defined above. (It can be verified that the limit in (1.1.2) always exists.) It is easy to check using (1.1.2) that, so long as the function g is finitely integrable, the notational device, p, may be treated as a probability density when one wishes to make a change of variable (as in (2.1.6)) or break up the region of integration in (1.1.2) into several parts (as in (2.1.7)). This restriction to finitely measurable functions for g will cause little difficulty in what follows. If $A \subset \mathcal{A}^{(\infty)}$, it is called finitely measurable if its indicator function is finitely measurable.

In (1.1.1) $\theta \varepsilon E_m$ is an unknown parameter. The statistical problem to be considered in this paper is the estimation of θ when the observations $X, Y_1, Y_2, \cdots$ become available sequentially, one after the other.

An *estimation procedure*, δ, consists of two measurable functions, an *estimator* ϵ and a *stopping rule* σ. $\sigma = n$ is interpreted to mean that the sequential sampling stops after observing the n variables $X, Y_1, Y_2, \cdots Y_{n-1}$. The estimate ϵ of θ is then made. Thus $\epsilon: H \rightarrow E_m$ and $\sigma: H \rightarrow J^*$ ($J^* = \{0, 1, 2, \cdots, \infty\}$). The intuitive requirement that both the stopping rule and the estimate depend only on past observations is the motivation for the following assumption: For all Lebesgue measurable $A \subset E_m$,

$$(1.1.3)\qquad \{x, y: \sigma(x, y) = n, \epsilon(x, y)\, \varepsilon\, A\}\ \varepsilon\ \mathcal{A}^{(n)}.$$

If for some $x, y, \sigma(x, y) = 0$, (1.1.3) is interpreted to mean that, for all x, y, $\sigma(x, y) = 0$ and that $\epsilon(x, y)$ is a constant. (See Section 1.3 for comments concerning the extension of this definition to allow the use of "randomized" estimation procedures.)

The loss functions considered in this paper may depend on three things: the difference between ϵ and the true value of θ, σ, and the values $y_1, y_2, \cdots$. Formally, a *loss function* is a Borel measurable function $W: E_m \times J^* \times (E_m \times E_m \times \cdots) \rightarrow E_1^*$ ($E_1^* = E_1 \cup \{\infty\}$). It will always be assumed that $W(t, n, y)$ is $\mathcal{A}^{(n)}$ measurable (thus if σ stops with probability one the function $W(\epsilon(x, y), \sigma(x, y), y)$ is finitely measurable). Assume also,

$$(1.1.4)\qquad 0 \leqq W \leqq \infty \quad\text{and}\quad W(t, \infty, y) = \infty.$$

Note that there are very few practical statistical problems in which the dependence on y actually occurs. For that reason we shall often pay special attention to the case where W is independent of y.

The *risk*, $R(\theta, \delta)$, of a procedure δ having $\sigma < \infty$ a.e. $(\pi \cdot \mu)$ when θ is the true parameter value is

$$(1.1.5)\quad R(\theta, \delta) = \int\int W(\epsilon(x, y) - \theta, \sigma(x, y), y)p(x - \theta, y)\, dx\mu(dy).$$

Since we will usually write $\gamma(x, y) = \epsilon(x, y) - x$ (and even $\delta = (\gamma, \sigma)$ when no confusion can occur) (1.1.5) often appears in the form

(1.1.5′) $R(\theta, \delta) = \int\int W(z + \gamma(z + \theta, y), \sigma(z + \theta, y), y)p(z, y)\, dz\mu(dy)$.

Since ϵ and σ are measurable and W is Borel measurable the integrands in (1.1.5) and (1.1.5′) are measurable. The integrals in (1.1.5) and (1.1.5′) are well defined (see (1.1.2)) only if $\sigma < \infty$ $(\pi\cdot\mu)$, however, if $\sigma \not< \infty$ a.e. then R (if we define it properly and use (1.1.4)) is infinite and so the procedure with the stopping rule σ is of no interest to us. Thus we define $R(\theta, \delta) = \infty$ if the integrand in (1.1.5) is not finitely integrable.

In accordance with the usual terminology an estimation procedure is (translation) *invariant* if

$$\epsilon(x, y) = x + \gamma(y), \qquad \sigma(x, y) = \sigma(y), \tag{1.1.6}$$

(i.e. σ and γ do not depend on x. Observe that this definition and (1.1.3) imply $\sigma \geqq 1$ for all invariant procedures). Note from (1.1.5′) that if δ is invariant $R(\theta, \delta) = R(\delta)$ (say) is independent of θ.

Let $\mathcal{I}$ be the class of invariant estimation procedures. Unless otherwise stated it is assumed throughout that there exists at least one procedure $\delta_0 \varepsilon \mathcal{I}$ such that

$$R_0 = R(\delta_0) = \inf_{\delta\varepsilon\mathcal{I}} R(\delta). \tag{1.1.7}$$

The procedure δ_0 is called a *best invariant estimator* (best inv est). Note that δ_0 is not necessarily uniquely determined. The symbols R_0 and δ_0 $(= (\epsilon_0, \sigma_0)$, or as will sometimes be written, $(\gamma_0, \sigma_0))$ are always used as in (1.1.7). (See Section 1.2 and Lemma 2.2.1 for some results concerning the existence of δ_0.)

Unless otherwise stated it is assumed that

$$R_0 < \infty. \tag{1.1.8}$$

An estimation procedure is called *admissible* (or admissible within the set $\mathcal{S}$, respectively) if for any procedure δ' ($\delta' \varepsilon \mathcal{S}$, respectively)

$$R(\theta, \delta') \leqq R(\theta, \delta) \quad \text{for all} \quad \theta \varepsilon E_m \qquad \text{implies} \quad R(\theta, \delta') = R(\theta, \delta) \quad \text{for all} \quad \theta \varepsilon E_m. \tag{1.1.9}$$

An estimation procedure δ is called *almost admissible* (with respect to Lebesgue measure) if

$$R(\theta, \delta') \leqq R(\theta, \delta) \quad \text{for all} \quad \theta \varepsilon E_m \qquad \text{implies} \quad R(\theta, \delta') = R(\theta, \delta) \quad \text{for almost all} \quad \theta \varepsilon E_m(d\theta). \tag{1.1.10}$$

The definitions stated above are sufficiently general to allow consideration of problems which have a fixed-sample-size, and are not at all sequential. Thus, a *fixed-sample-size* problem is one in which

$$W(t, j, y) = W(t, y), \quad j = n, \qquad = \infty, \quad j \neq n, \tag{1.1.11}$$

and where (for simplicity) $W(t, y)$ is $\mathfrak{A}^{(n)}$ measurable. If W is given by (1.1.11) then for every (sequential) δ for which $\sigma \not\equiv n$ there is a δ' with $\sigma \equiv n$ such that $R(\theta, \delta') \leqq R(\theta, \delta)$ for all θ. Thus in looking for admissible procedures it is enough to look only at procedures which always observe $X, Y_1, \cdots, Y_{n-1}$ and stop. For this reason it will be assumed that $\sigma \equiv n$ in fixed-sample-size problems, and σ will not be mentioned in the notation. In this way δ consists merely of an estimator ϵ (or γ, where $\epsilon(x, y) = x + \gamma(x, y)$), and ϵ (or γ) is $\mathfrak{A}^{(n)}$ measurable. Thus in the fixed-sample-size-case variables $Y_n, Y_{n+1}, \cdots$ may be completely ignored, and this is the approach we shall take. Let us also point out that the restriction that Y_i take its values in E_m (rather than some other space) is not essential for the argument in any part of this paper.

1.2. *Comments concerning the general problem; specializations and generalizations.* This section begins with a description of a statistical problem which will be referred to in this paper (for want of a better name) as the *special-sequential-problem.* It is shown that this problem is truly a special case of the general problem described in Section 1.1.

Suppose the experimenter may observe the independent identically distributed random variables $X_1, X_2, X_3, \cdots (X_i \varepsilon E_m)$ in sequence. Each X_i has the probability density $f(x - \theta)$ with respect to Lebesgue measure on E_m. After observing $X_1, \cdots, X_n$ the experimenter may choose either to stop and make an estimate (based on $x_1, \cdots, x_n$) of the unknown parameter, $\theta \varepsilon E_m$, or he may choose to observe X_{n+1}, If he stops and makes the estimate ϵ, his loss is $W(\epsilon - \theta, n)$. A (non-randomized) translation invariant estimation procedure in this case is one in which the stopping rule depends only on the maximal invariant statistic, $(x_2 - x_1, x_3 - x_1, \cdots)$, and the estimate is of the form $\epsilon(x_1, x_2, \cdots) = x_1 + \gamma(x_2 - x_1, x_3 - x_1, \cdots)$ (certain requirements analogous to (1.1.3) must also be satisfied by the stopping rule, σ, and by ϵ).

To exhibit the equivalence of this problem to that formulated in Section 1.1, let $X = X_1, Y_i = X_{i+1} - X_1$. Let

$$\begin{aligned} p_n(x, y) &= f(x)f(y_1 + x)f(y_2 + x) \cdots f(y_{n-1} + x) \\ &\qquad /\textstyle\int f(x)f(y_1 + x)f(y_2 + x) \cdots f(y_{n-1} + x)\, dx; \qquad (1.2.1) \\ \mu_n(dy) &= (\textstyle\int f(x)f(y_1 + x)f(y_2 + x) \cdots f(y_{n-1} + x)dx) \prod_{i=1}^{n-1} dy_i\,. \end{aligned}$$

It is easily checked that if θ is the true value of the parameter and $\mu(S) = \lim_{n\to\infty} \mu_n(S)$ then $p_n(x - \theta, y)$ satisfies (1.1.1). An invariant procedure in this X, Y, problem is of the form (1.1.6) and the risk is of the form (1.1.5). The risk of a procedure in the special problem is equal to the risk of the corresponding procedure in the general problem. Thus the special problem falls into the framework of the general problem of Section 1.1 where X, Y and $p_n(x, y)$ are defined as above.

We now turn to a different matter. In the fixed sample-size-problem the transformation $X \to X + g(Y)$ (g measurable) does not change the essential nature of the problem in any way. That is; the risk of the estimator $\epsilon(x, y)$ on the basis of X, Y having density $p(x - \theta, y)$ equals the risk of the estimator $\epsilon(x' - g(y), y)$

$= \epsilon'(x', y)$ for θ on the basis of X', Y (where $x' = x + g(y)$) having density $p(x' - g(y) - \theta, y) = p'(x' - \theta, y)$. The above can easily be checked using (1.1.5). If $g(y) = \gamma_0(y)$ then in the transformed problem $\epsilon_0'(x', y) = \epsilon_0(x' - g(y), y) = x' - \gamma_0(y) + \gamma_0(y) = x'$. Thus in the transformed problem a best inv est is given by X', i.e., $\gamma_0'(y) = 0$. The above shows that there is no loss of generality in assuming that $\gamma_0(y) \equiv 0$ in every fixed-sample-size problem. We shall often assume in fixed-sample-size problems that $\gamma_0 = 0$ (throughout Section 3 this assumption will be made).

The reader will notice that although the essential nature of a fixed-sample-size problem remains unchanged by the transformation $X \to X + g(Y)$, some of the assumptions in the following chapters about finiteness of moments are satisfied only for certain choices of $g(Y)$. (Often $g(y) = \gamma_0(y)$ is a good choice.) For example it may be that while the assumptions (2.1.2) or (2.1.3) of Theorem 2.1.1 are not satisfied for some problem they will be satisfied for a problem equivalent in the above sense, and thus δ_0 will be admissible for the original problem as well as for the problem in its transformed form.

In the general sequential problem the variety of transformations which leave the problem unchanged in the above sense is limited by the restriction (1.1.3). In general, the only such transformations are $x \to x + k$ where k is some constant, $k \varepsilon E_m$. The same remarks as above apply: δ_0 may sometimes be proved admissible using Theorem 2.1.1 by applying that theorem to the transformed problem.

We will conclude this section with two remarks concerning possible generalizations of the theorems of this paper.

First, a more general method of sequential sampling is possible. Namely, after observing $X, Y_1, Y_2, \cdots, Y_{n-1}$ the experimenter may stop sampling and make an estimate or he may decide to continue sampling by observing $Y_n, Y_{n+1}, \cdots, Y_{n+k}$ all at once. The choice of k is up to the experimenter, and may depend on $X, Y_1, \cdots, Y_{n-1}$. In this case the sampling rule σ contains the information on how to continue sampling, as well as on when to stop. This can most easily be accomplished by letting $\sigma: E_m \times E_m \times \cdots \to J_1^* \times J_1^* \times \cdots$ where, for example $\sigma(x, y) = 1, 2, 0, 0, 0, \cdots$ means that first X was observed, and then Y_1, Y_2 were observed together, and then an estimate was made. Certain requirements generalizing (1.1.3) must be made on ϵ and σ. See Kiefer (1957) for a complete definition of this type of problem. The theorems and lemmas of Sections 2.1, 2.2, and 2.3 remain valid with several minor changes (mostly in the notation). For instance, in Lemma 2.3.4 it should be assumed that the cost of any group of observations is greater than $c > 0$, instead of merely assuming that each observation costs more than $c > 0$.

A second possible generalization is to drop the requirement that the distribution of X given $Y_1, Y_2, \cdots, Y_n$ be absolutely continuous with respect to Lebesgue measure (in this case the density $p(x, y)\, dx$ should be replaced by the distribution $dP(x, y)$. The results of Section 3 are almost unaltered by this change—see Corollary 3.1.2. On the other hand if this requirement of absolute continuity is dropped, it is in general at most possible to conclude almost admissibility of

δ_0 in Theorem 2.1.1. In fact, subject to some minor additional regularity conditions the method of proof used in Theorem 2.1.1 can actually be extended to prove almost admissibility of δ_0. The way to proceed is to everywhere replace $p(z, y)\, dz$ by $d_z\, P(z, y)$.

1.3. *Randomized estimators.* It should be noted that in the formulation of the problem in Section 1.1 no explicit provision was made to allow the use of randomized estimators; that is, estimators which depend on the value of Z, a random variable (independent of X, Y) having uniform distribution on $(0, 1)$, as well as on the sequentially observed values of X, Y. (For rigorous definitions of all the concepts involved in a randomized problem see Kiefer (1957).) Randomization after observing X can easily be introduced into the formulation of Section 1.1. If the original problem consisted of sampling from $X, Y_1, Y_2, \cdots$ and has loss function W, construct a new (*randomization-allowing*) problem which consists of sampling from $X, Y_1', Y_2', \cdots$ where $Y_1' = Z$, $Y_{i+1}' = Y_i$, $i = 1, 2, \cdots$, and which has loss function $W'(t, n, y') = W(t, n-1, (y_2', y_3', \cdots))$ for $n \geqq 2$ and $W'(t, 1, y') = W(t, 1, (y_2', y_3', \cdots))$. In this new problem, Z may be sampled without cost after the observation of X. An estimator for the new problem is a randomized estimator for the original problem, but one in which randomization is not introduced until after X has been observed.

It may easily be checked that for any randomized invariant estimator, δ_r, there is a non-randomized invariant estimator δ such that $R(\delta) \leqq R(\delta_r)$, (see Blackwell and Girshick (1954), p. 312). It follows that (when the terms used for the randomized problem are properly defined) if a randomized best inv est exists, then also a non-randomized best inv est exists. Thus the introduction of randomization does not change the existence of a best inv est.

It can also be easily checked that the hypotheses of Theorem 2.1.1 are satisfied for some problem if and only if they are satisfied for the randomization allowing problem as defined previously. A best inv est is the same for the two problems. Thus if the hypotheses of Theorem 2.1.1 are satisfied the best inv est is admissible even among estimators allowing randomization after the observation of X.

In unusual cases the best inv est may be admissible among non-randomized estimators (or among randomized estimators which observe at least X), but may not be admissible among the class of *all* randomized estimators. The following is an example:

Let, for $n = 1, 2, \cdots$,

$$p_n(x, y) = 1, \qquad -\tfrac{1}{2} < x < \tfrac{1}{2}, \qquad = 0, \qquad \text{otherwise,} \tag{1.3.1}$$

and let $W(t, n, y) = W_1(t) + n$ where

$$\begin{aligned} W_1(t) &= 0, && |t| \leqq \tfrac{1}{2}, \\ &= \tfrac{3}{2}, && \tfrac{1}{2} < |t| < \tfrac{3}{2}, \\ &= \tfrac{1}{2}, && |t| \geqq \tfrac{3}{2}. \end{aligned} \tag{1.3.2}$$

It can easily be checked that the assumptions of Theorem 2.1.1 are satisfied so that the best inv est is admissible among the class of estimators allowing randomization after the observation of X. The best inv est has risk $R_0 = 1$. The procedure, δ_1, which makes the estimates -3, 0, 3 each with probability $\frac{1}{3}$, respectively without observing the value of X has risk

$$
\begin{aligned}
R(\theta, \delta_1) &= \tfrac{1}{2}, & |\theta| &\geqq \tfrac{9}{2}, \\
&= \tfrac{5}{6}, & \tfrac{7}{2} &< |\theta| < \tfrac{9}{2}, \\
&= \tfrac{1}{3}, & \tfrac{5}{2} &\leqq |\theta| \leqq \tfrac{7}{2}, \\
&= \tfrac{5}{6}, & \tfrac{3}{2} &< |\theta| < \tfrac{5}{2}, \\
&= \tfrac{1}{2}, & |\theta| &= \tfrac{3}{2}, \\
&= \tfrac{5}{6}. & \tfrac{1}{2} &< |\theta| < \tfrac{3}{2}, \\
&= \tfrac{1}{3}, & 0 &\leqq |\theta| < \tfrac{1}{2}.
\end{aligned}
\tag{1.3.3}
$$

Thus $R(\theta, \delta_1) < R_0$ so that δ_1 is a better procedure than δ_0.

We leave it to the reader to check that if $\sup W_1(t) = \infty$ and $R_0 < \infty$, or if $\liminf_{t\to\infty} W_i(t) > R_0$ or $\liminf_{t\to-\infty} W_1(t) > R_0$; then no randomized procedure, δ, for which $\sigma = 0$ with positive probability, can be better than δ_0 (in the sense that $R(\theta, \delta) \leqq R_0$ for all θ). Thus in these cases if the hypotheses of Theorem 2.1.1 are satisfied δ_0 is admissible among the class of all randomized estimators.

In fixed-sample-size problems the above difficulty is, of course, not present. There the randomization allowing problem as defined at the beginning of this section allows all randomized estimators for consideration.

2. Estimation of real location parameters.

2.1. *Admissibility of the best invariant estimator for $m = 1$.* The general result given in this section depends in part on the following two assumptions:

(2.1.1) $\qquad R(\delta_i) \to R_0, \ \delta_i \in \mathcal{G}, \qquad i = 1, 2, \cdots,$ implies

(a) $\epsilon_i(x, y) \to x + \gamma_0(y)$ (or $\gamma_i(y) \to \gamma_0(y)$) in measure (with respect to μ),

and

(b) $\sigma_i(y) \to \sigma_0(y)$ in measure (μ),

and

$$
\begin{aligned}
\text{(2.1.2)} \quad \int_0^\infty d\lambda \{\sup\nolimits_{\delta=(\gamma,\sigma)\in\mathcal{G}} \int \mu(dy) \int_{-\lambda}^{\lambda} [W(x + \gamma_0(y), \sigma_0(y), y) \\
- W(x + \gamma(y), \sigma(y), y)] p(x, y)\, dx\} < \infty.
\end{aligned}
$$

(The quantity in braces in (2.1.2) is the difference of two monotone functions, hence is measurable ($d\lambda$).)

Neither of these conditions is very restrictive, but neither is vacuous. Some more easily verified assumptions which imply (2.1.1) and (2.1.2) are given in

Sections 2.2 and 2.3 respectively. The reader may read these lemmas before Theorem 2.1.1 if he desires.

The statement and proof of Theorem 2.1.1 partially parallels the statement and proof of a much more specialized result given by Blackwell (1951). Where Blackwell used finite discrete probability distributions and finite summations we use probability densities and integrals over infinite regions. As is usual in such cases Blackwell's convergence has of necessity been replaced by convergence a.e. or convergence in measure. This replacement causes difficulties that were, of course, not present in Blackwell's proof. In addition the replacement of finite sums by integrals over the real line has required the assumption that certain moments of p exist. The existence of these moments is used in the proof to show that certain error terms which occur are $O(1)$ or $o(1)$ (as $L \to \infty$), whichever is necessary for the argument.

In Section 2.4 some examples will be given to show that at least some of the moment conditions assumed in Theorem 2.3.1 are necessary in order for the best invariant estimator to be admissible. Let $\mathcal{S}_1$ be the set of estimation procedures which take at least one observation.

THEOREM 2.1.1. *In the problem defined in Section* 1.1, *let* $m = 1$ (i.e. X *a real random variable*), $R_0 < \infty$. *Suppose assumptions* (2.1.1) *and* (2.1.2) *are satisfied, and*

$$(2.1.3) \qquad \int \mu(dy) \int |xW(x + \gamma_0(y), \sigma_0(y), y)| \cdot p(x, y)\, dx < \infty.$$

Then the best invariant estimation procedure, δ_0, *is admissible in the set* $\mathcal{S}_1$. *If* $\sup_t \inf_y W(t, 0, y) > R_0$ *then* δ_0 *is admissible.*

PROOF. Suppose there exists an estimator δ, $\delta \in \mathcal{S}_1$, such that $R(\theta, \delta) \leqq R_0$. We shall show that δ is equivalent to δ_0, i.e.,

$$(2.1.4) \quad \int \mu(dy) \int [|\epsilon_0(x, y) - \epsilon(x, y)| + |\sigma_0(y) - \sigma(x, y)|]\, dx = 0.$$

For a δ satisfying $R(\theta, \delta) \leqq R_0$,

$$(2.1.5) \quad \int_{-L}^{L} d\theta \int \mu(dy) \int [W(\epsilon_0(x, y) - \theta, \sigma_0(y), y) - W(\epsilon(x, y) - \theta, \sigma(x, y), y)]p(x - \theta, y)\, dx \geqq 0.$$

Since $W \geqq 0$ and $R(\theta, \delta) \leqq R_0$ the order of integration in (2.1.5) may be changed. Suppose this is done so that the θ integration is performed first, $x - z$ is substituted for θ, and $\epsilon(x, y)$ is replaced by $x + \gamma(x, y)$ and $\epsilon_0(x, y)$ by $x + \gamma_0(y)$. Then (2.1.5) becomes

$$(2.1.6) \quad \int \mu(dy) \int_{-\infty}^{\infty} dx \int_{x-L}^{x+L} [W(z + \gamma_0(y), \sigma_0(y), y) - W(z + \gamma(x, y), \sigma(x, y), y)]p(z, y)\, dz \geqq 0.$$

(Since $W \geqq 0$ and $R(\theta, \delta) \leqq R_0$ the order of integration in (2.1.6) is immaterial. When it is convenient that order will be changed without further justification.)

The range of integration in (2.1.6) can be broken up as

$$(2.1.7)\quad \int \mu(dy) \int dx \int_{x-L}^{x+L} dz = \int \mu(dy)\{\int_{-L/2}^{L/2} dx \int_{-L/2}^{L/2} dz + \int_{-3L/2}^{-L/2} dx \int_{-L/2}^{L+x} dz + \int_{L/2}^{3L/2} dx \int_{x-L}^{L/2} dz + \int_{L/2}^{\infty} dz \int_{z-L}^{z+L} dx + \int_{-\infty}^{-L/2} dz \int_{z-L}^{z+L} dx\}.$$

(Note the change of order of integration in the last two terms.)

Using (2.1.3),

$$\begin{aligned} &\int \mu(dy)\{\int_{-\infty}^{-L/2} dz \int_{z-L}^{z+L} dx + \int_{L/2}^{\infty} dz \int_{z-L}^{z+L} dx\} \\ &\quad [W(z + \gamma_0(y), \sigma_0(y), y) - W(z + \gamma(x, y), \sigma(x, y), y)]p(z, y) \\ &\leqq \int \mu(dy)\{\int_{-\infty}^{-L/2} dz \int_{z-L}^{z+L} dx + \int_{L/2}^{\infty} dz \int_{z-L}^{z+L} dx\} \\ (2.1.8)\qquad &\quad \cdot W(z + \gamma_0(y), \sigma_0(y), y)p(z, y) \\ &= 2L \int \mu(dy) \int_{|z|>L/2} W(z + \gamma_0(y), \sigma_0(y), y)p(z, y)\, dz \\ &\leqq 4 \int \mu(dy) \int_{|z|>L/2} |z| W(z + \gamma_0(y), \sigma_0(y), y)p(z, y)\, dz \\ &\to 0 \quad \text{as} \quad L \to \infty. \end{aligned}$$

For convenience, let

$$\begin{aligned} (2.1.9)\quad \omega(z, \gamma_0, \sigma_0, \gamma, \sigma, y) &= W(z + \gamma_0, \sigma_0, y) - W(z + \gamma, \sigma, y); \\ \bar{\omega}(z, x, y) &= \omega(z, \gamma_0(y), \sigma_0(y), \gamma(x, y), \sigma(x, y), y). \end{aligned}$$

If $\sup_t \inf_y W(t, 0, y) > R_0 \geqq R(\theta, \delta)$ then $\sigma \geqq 1$ (see after (1.1.3)). Otherwise $\delta \in S_1$ by assumption. Hence we may assume $\delta \in \mathcal{S}_1$.

For any fixed x_0 define the estimation procedure δ_{x_0} by the estimator and stopping rule $x + \epsilon(x_0, y) - x_0$, $\sigma(x_0, y)$ considered as functions of y for that fixed x_0. It is easily checked, using the fact that $\sigma \geqq 1$, that δ_{x_0} is an invariant estimation procedure; i.e., $\delta_{x_0} \in \mathcal{G}$. Using (2.1.2) and (2.1.3),

$$\begin{aligned} &\int \mu(dy) \int_{-3L/2}^{-L/2} dx \int_{-L/2}^{L+x} \bar{\omega}(z, x, y)p(z, y)\, dz \\ &= \int \mu(dy)\{\int_{-L/2}^{0} dx \int_{-L/2}^{x} dz + \int_{0}^{L/2} dx \int_{-L/2}^{-x} dz + \int_{0}^{L/2} dx \int_{-x}^{x} dz\} \\ &\quad \cdot \omega(z, \gamma_0(y), \sigma_0(y), \gamma(x - L, y), \sigma(x - L, y), y)p(z, y)\, dz \\ &\leqq 2 \int \mu(dy) \int_{-L/2}^{0} |z| W(z + \gamma_0(y), \sigma_0(y), y)p(z, y)\, dz \\ (2.1.10)\qquad &\quad + \int_0^{L/2} dx \int \mu(dy) \int_{-x}^{x} \omega(z, \gamma_0(y), \sigma_0(y), \gamma(x - L, y), \\ &\quad \sigma(x - L, y), y)p(z, y)\, dz \\ &\leqq 2 \int \mu(dy) \int_{-\infty}^{\infty} |z| W(z + \gamma_0(y), \sigma_0(y), y)p(z, y)\, dz \\ &\quad + \int_0^{\infty} dx\{\sup_{\delta \in \mathcal{G}} \int \mu(dy) \\ &\quad \int_{-x}^{x} \omega(z, \gamma_0(y), \sigma_0(y), \gamma(y), \sigma(y), y)p(z, y)dz\} \\ &\leqq c_1 \qquad (c_1 \text{ is some constant, independent of } L). \end{aligned}$$

A similar result holds for the third term on the right in (2.1.7). Combining the inequality (2.1.6) with (2.1.7), (2.1.8), and (2.1.10) there is a constant c_2 ($\geqq 2c_1$) such that

$$\int \mu(dy) \int_{-L/2}^{L/2} dx \int_{-L/2}^{L/2} \bar{\omega}(z, x, y)p(z, y)\, dz \geqq -c_2 . \tag{2.1.11}$$

Since the integral in (2.1.2) is finite there exists a sequence $\lambda_i \to \infty$ such that the quantity in braces ($\{\ \}$) in that integral tends to 0 like $o(\lambda_i^{-1})$ (as $i \to \infty$). Hence if k is any fixed integer,

$$\begin{aligned} &\liminf_{i\to\infty} \int_{-\lambda_i}^{\lambda_i} dx \int \mu(dy) \int_{-\lambda_i}^{\lambda_i} \bar{\omega}(z, x, y)p(z, y)\, dz \\ &\leqq \liminf_{i\to\infty} \{\int_{-\lambda_k}^{\lambda_k} dx \int \mu(dy) \int_{-\lambda_i}^{\lambda_i} \bar{\omega}(z, x, y)p(z, y)\, dz \\ &\quad + \int_{-\lambda_i}^{\lambda_i} dx \{\sup_{\delta' \varepsilon \mathscr{J}} \int \mu(dy) \int_{-\lambda_i}^{\lambda_i} \omega(z, \gamma_0(y), \sigma_0(y), \gamma'(y), \sigma'(y), y) \\ &\qquad \cdot p(z, y)\, dz\}\} \\ &= \liminf_{i\to\infty} \int_{-\lambda_k}^{\lambda_k} dx \int \mu(dy) \int_{-\lambda_i}^{\lambda_i} \bar{\omega}(z, x, y)p(z, y)\, dz \\ &= \int_{-\lambda_k}^{\lambda_k} dx \int \mu(dy) \int_{-\infty}^{\infty} \bar{\omega}(z, x, y)p(z, y)\, dz. \end{aligned} \tag{2.1.12}$$

(Again we have used the fact that for all fixed values of x, $\delta_x(y) \,\varepsilon\, \mathscr{J}$.) Since the choice of k in (2.1.12) is arbitrary,

$$\begin{aligned} &\liminf_{\lambda\to\infty} \int_{-\lambda}^{\lambda} dx \int \mu(dy) \int_{-\lambda}^{\lambda} \bar{\omega}(z, x, y)p(z, y)\, dz \\ &\qquad \leqq \int_{-\infty}^{\infty} dx \{\int \mu(dy) \int_{-\infty}^{\infty} \bar{\omega}(z, x, y)p(z, y)\, dz\}. \end{aligned} \tag{2.1.13}$$

The braces in (2.1.13) have been added to emphasize that the order of integration on the right of (2.1.13) cannot be changed.

The bulk of the remainder of the proof is devoted to establishing the result in Equation (2.1.24). First we derive the result in (2.1.18), and then this result and Lemma 2.1.1 are applied to the second and third terms on the right of the integral decomposition in (2.1.7), in order to prove (2.1.24).

For any fixed $\alpha > 0$, $A > 0$ define

$$\begin{aligned} S(L) = \{x, y\colon -L - A < x < -L + A, &\text{ and either} \\ |\gamma(x, y) - \gamma_0(y)| > \alpha &\text{ or } \sigma(x, y) \neq \sigma_0(y)\}. \end{aligned} \tag{2.1.14}$$

Let $\beta_1 > 0$ be given. Just for this paragraph define $T(L) = \{x\colon \mu\{y\colon (x, y) \,\varepsilon\, S(L)\} > \beta_1/4A\}$. We note that if for some L, $\iint_{S(L)} \mu(dy)\, dx > \beta_1$, then

$$2 \int_{T(L)} dx \geqq \beta_1 . \tag{2.1.15}$$

Let $T = \bigcup_{L>0} T(L)$. It follows from the definition of $T(L)$ and T that there does *not* exist a sequence $\{x_i\}$ such that $x_i \,\varepsilon\, T$ and $\delta_{x_i} \to \delta_0$ (i.e. $\gamma(x_i, y) \to \gamma_0(y)$ and $\sigma(x_i, y) \to \sigma_0(y)$). Hence, using (2.1.1) there is a $\beta_2 > 0$ such that $x \,\varepsilon\, T$ implies

$$\int \mu(dy) \int_{-\infty}^{\infty} \omega(z, \gamma_0(y), \sigma_0(y), \gamma(x, y), \sigma(x, y), y)p(z, y)\, dz \leqq -\beta_2 . \tag{2.1.16}$$

Using (2.1.11), (2.1.13), and (2.1.15),

$$(2.1.17)\qquad -c_2 \leqq \int_{-\infty}^{\infty} dx\{\int \mu(dy) \int \omega(z, \gamma_0(y), \sigma_0(y), \gamma(x, y), \sigma(x, y), y)\, p(z, y)\, dz\} \leqq -\beta_2 \int_T dx.$$

Thus $\int_T dx \leqq c_2/\beta_2$. This implies the existence of an L^* such that $|L| > L^*$ implies $\int_{T(L)} dx < \beta_1/2$ which, using (2.1.15), implies $\int\int_{S(L)} \mu(dy)\, dx \leqq \beta_1$. Hence,

$$(2.1.18)\qquad \int\int_{S(L)} \mu(dy)\, dx \to 0 \quad \text{as} \quad L \to \infty.$$

At this point the following simple lemma is needed.

LEMMA 2.1.1. *Suppose f, g are Lebesgue measurable, $g \geqq 0$, and $\int g(z)\, dz < \infty$. Then $a_i \to a$, implies $f(z + a_i) \to f(z + a)$ in measure (with respect to the measure generated by $\int g(z)\, dz$). If, furthermore, $f \geqq 0$ and $\int f(z + a)g(z)\, dz < \infty$, then*

$$(2.1.19)\qquad \limsup_{a_i \to a} (\sup_Q \int_Q [f(z + a) - f(z + a_i)]g(z)\, dz) \leqq 0.$$

PROOF. If $R \subset (-\infty, \infty)$ define $R^c = \{x: x \not\in R\}$ and $R - b = \{x: x + b \in R\}$. Using Lusin's theorem and absolute continuity, for any $\alpha > 0$ there is a (closed, bounded) set R such that $\int_{R^c} g(z)\, dz < \alpha$, f is uniformly continuous on R, and for all i, $\int_{R^c + (a_i - a)} g(z)\, dz < \alpha$. If i is chosen sufficiently large, $|a_i - a|$ will be small enough so that $|f(z + a_i) - f(z + a)| < \alpha$ for all

$$z \in S = \{z: (z + a) \in R, \text{ and } (z + a_i) \in R\}.$$

$\int_{S^c} g(z)\, dz < 2\alpha$. Thus $f(z + a_i) \to f(z + a)$ in measure. (2.1.19) follows from Fatou's lemma, and positivity and integrability of f. This completes the proof of the lemma.

Using (2.1.19) for almost all $y(\mu)$,

$$(2.1.20)\qquad \limsup_{\varphi \to \gamma_0(y)} (\sup_{\lambda_1, \lambda_2} \int_{\lambda_1}^{\lambda_2} \omega(z, \gamma_0(y), \sigma_0(y), \varphi, \sigma_0(y), y)p(z, y)\, dz) \leqq 0.$$

Using (2.1.18) and (2.1.20) (for the inequality giving the 0 in the last step of (2.1.21)) and $R_0 < \infty$ (for bounded convergence),

$$\begin{aligned}
&\limsup_{L\to\infty} \int \mu(dy) \int_{-3L/2}^{-L/2} dx \int_{-L/2}^{L+x} \bar{\omega}(z, x, y)\cdot p(z, y)\, dz \\
&\quad = \limsup_{L\to\infty} \int \mu(dy)\{\int_{-L-A}^{-L+A} dx \int_{-L/2}^{L+x} dz + \int_{-3L/2}^{-L-A} dx \int_{-L/2}^{L+x} dz \\
&\qquad + \int_{-L+A}^{-L/2} dx \int_{-L-x}^{L+x} dz + \int_{-L+A}^{-L/2} dx \int_{-L/2}^{-L-x} dz\}\, \bar{\omega}(z, x, y)p(z, y)\, dz \\
&\quad \leqq \limsup_{L\to\infty} \int_{-A}^{A} dx \int \mu(dy) \int_{-L/2}^{x} \\
(2.1.21)\qquad &\qquad \cdot\omega(z, \gamma_0(y), \sigma_0(y), \gamma(x - L, y), \sigma(x - L, y), y)p(z, y)\, dz \\
&\qquad + \limsup_{L\to\infty} \int \mu(dy)\ \{\int_{-L/2}^{-A} dz \int_{z-L}^{-A-L} dx + \int_{-L+A}^{-L/2} dx \int_{-L-x}^{L+x} dz \\
&\qquad + \int_{-L/2}^{-A} dz \int_{A-L}^{-z-L} dx\}\ \bar{\omega}(z, x, y)p(z, y)\, dz \\
&\quad \leqq 0 + 2 \lim_{L\to\infty} \int \mu(dy) \int_{-L/2}^{-A} (|z| + A)W(z + \gamma_0(y), \\
&\qquad \cdot\sigma_0(y), y)p(z, y)dz + \lim_{L\to\infty} \int_{A}^{L/2} dx\{\sup_{\delta' \in \mathcal{G}} \int \mu(dy) \\
&\qquad \cdot\int_{-x}^{x} \omega(z, \gamma_0(y), \sigma_0(y), \gamma'(y), \sigma'(y), y)p(z, y)\, dz\}.
\end{aligned}$$

The choice of A in (2.1.21) is arbitrary. Using (2.1.2) and (2.1.3) the last two terms on the right of (2.1.21) may be made as small as desired by choosing A sufficiently large. Hence

(2.1.22) $$\limsup_{L\to\infty} \int \mu(dy) \int_{-3L/2}^{-L/2} dx \int_{-L/2}^{L+x} \bar{\omega}(z, x, y)p(z, y)\, dz \leqq 0.$$

Similarly,

(2.1.23) $$\limsup_{L\to\infty} \int \mu(dy) \int_{L/2}^{3L/2} dx \int_{x-L}^{L/2} \bar{\omega}(z, x, y)p(z, y)\, dz \leqq 0.$$

Substituting (2.1.22), (2.1.23), and (2.1.8) into (2.1.7) yields

(2.1.24) $$\liminf_{L\to\infty} \int \mu(dy) \int_{-L/2}^{L/2} dx \int_{-L/2}^{L/2} \bar{\omega}(z, x, y)p(z, y)\, dz \geqq 0.$$

Since δ_0 is a best inv est the quantity in braces in (2.1.13) is non-positive. Thus (2.1.13) and (2.1.24) together imply

(2.1.25) $$\int \mu(dy) \int_{-\infty}^{\infty} \bar{\omega}(z, x, y)p(z, y)\, dz = 0 \qquad \text{a.e.} \qquad (dx).$$

(2.1.1) implies that the best inv est $(\gamma_0(y), \sigma_0(y))$ is uniquely determined a.e. (μ). This fact and (2.1.25) prove that for almost all x (dx), $\gamma_0(y) = \gamma(x, y)$ and $\sigma_0(y) = \sigma(x, y)$ a.e. (μ). This proves (2.1.4). Finally, (2.1.4) implies that $R(\theta, \delta) = R(\theta, \delta_0)$. Thus δ_0 is admissible.

This completes the proof of the theorem.

2.2. *Some results concerning assumption* (2.1.1). For the fixed-sample-size problem Farrell (1964) has shown (under some mild restrictions) that if the best inv est (given by $\gamma_0(y)$, $\sigma_0(y)$) is not uniquely determined a.e. (μ) then it is not admissible. Assumption (2.1.1) clearly implies that the best inv est is unique a.e. (μ). It will be shown in this section that (2.1.1) is in fact only a little stronger than uniqueness of the best invariant estimator in the general sequential problem and almost the same as uniqueness in the more specialized fixed-sample-size problem (see Corollary 2.2.2).

For the general problem as defined in Section 1.1, if δ is any invariant procedure (and $\epsilon(x, y) = x + \gamma(y)$), let

(2.2.1) $$R(\delta \mid y) = \int W(z + \gamma(y), \sigma(y), y)p(z, y)\, dz.$$

It is convenient to abuse this notation somewhat as follows: If v is a real number then we let $R((v, \sigma')|y) = R(\delta \mid y)$ where $\delta = (\gamma, \sigma')$ with $\sigma = \sigma'$ and $\gamma(y) = v$. As a convention, if $\varnothing$ is an empty set of real numbers, let $\inf \varnothing = \infty$.

LEMMA 2.2.1. *Let $\sigma'(\cdot)$ be a fixed invariant stopping rule. Suppose there exists a procedure $\delta = (\gamma, \sigma') \in \mathcal{I}$ such that $R(\delta) < \infty$. Suppose $p_n(\cdot, \cdot)$ and $W(\cdot, n, \cdot)$ are measurable with respect to the field of Borel subsets of $E_n (= \{x, y_1, y_2, \cdots, y_{n-1}\})$. Define $\gamma_0'(y)$ by*

(2.2.2) $$\gamma_0'(y) = \inf \{v: R((v, \sigma') \mid y) = \inf_t R((t, \sigma') \mid y)\}.$$

Suppose $|\gamma_0'(y)| < \infty$ a.e. (μ).

Then $\gamma_0'(y)$ is finitely (Borel) measurable, $\delta_0' = (\gamma_0', \sigma') \in \mathcal{I}$, and $R(\delta_0') = \inf_{\delta=(\gamma,\sigma')\in\mathcal{I}} R(\delta)$.

δ_0' *is the unique* (a.e. (μ)) *estimator of the form* $(\gamma, \sigma') \varepsilon \mathcal{G}$ *having risk* $R(\delta_0')$ *if and only if for almost all* y *the set in braces on the right of* (2.2.2) *contains only one point.*

Suppose δ_0' *is unique* (a.e. (μ)) *as above. Suppose also that*

$$(2.2.3) \qquad \liminf_{v \to \pm\infty} R((v, \sigma') \mid y) > \inf_t R((t, \sigma') \mid y) \text{ a.e. } (\mu).$$

Let $\delta_k' = (\gamma_k', \sigma') \varepsilon \mathcal{G}$ *be any sequence of estimators such that* $R(\delta_k') \to R(\delta_0')$. *Then* $\gamma_k' \to \gamma_0'$ *in measure* (μ) *as* $k \to \infty$.

PROOF. Consider the function

$$(2.2.4) \qquad f_n(v, y) = \int W(z + v, n, y) p_n(z, y)\, dz = R((v, \sigma') \mid y)$$

defined on $\Sigma(n) = [y : \sigma'(y) = n\}$. Since $W(z + v, n, y)p_n(z, y)$ is a Borel measurable function of $z, v, y_1, y_2, \cdots, y_{n-1}$ for $y \varepsilon \Sigma(n)$ it follows that $f_n(v, y)$ is well defined for almost all v, $y \varepsilon \Sigma(n)$ and is a Borel measurable function of $v, y_1, \cdots, y_{n-1}$.

It follows directly from Lemma 2.1.1 that $f_n(v, y)$ is lower semi-continuous in v for each fixed $y_1, \cdots, y_{n-1}$. Hence for each fixed $y_1, \cdots, y_{n-1}$, $T(y) = \{v : f(v, y) = \inf_t f(t, y)\}$ is a closed set.

We now use a theorem of Novikoff, see Arsenin and Liapunov (1955), p. 80 or Hahn (1948), p. 389–391. This theorem, slightly generalized, states that if S is a Borel set in $E_1 \times E_k$ such that the section of S at each $y \varepsilon E_k$ is closed, then the projection of S on E_k is Borel measurable.

Let $g_n(y) = \inf_v f_n(v, y)$. Let $\mathfrak{U}_a = \{v, y : f_n(v, y) \leqq a\}$. Since f_n in v for each fixed y is lower- semi-continuous the cross sections of $\mathfrak{U}_a$ at each $y \varepsilon \Sigma(n)$ are closed. Since f_n is Borel, $\mathfrak{U}_a$ is a Borel set. Hence $\{y: \exists v;\ (v, y) \varepsilon \mathfrak{U}_a\} = \{y : g_n(y) \leqq a\}$ is a Borel set, and g_n is a Borel measurable function.

Now, let $T = \bigcup_{y \varepsilon \Sigma(n)} T(y)$. Also, let $S_a = \{T\} \cap \{(-\infty, a] \times \Sigma(n)\}$. It follows from the previous paragraph that $T = \{v, y : f_n - g_n = 0\}$ is Borel. Hence for each a, S_a is a Borel set in $E_1 \times E_{n-1}$ and the section of S_a at each $y \varepsilon E_{n-1}$ is closed. Thus the projection of S_a is Borel measurable. But, using (2.2.2,) this projection is precisely $\{y: y \varepsilon \Sigma(n), \gamma_0'(y) \leqq a\}$. Hence $\gamma_0'(\cdot)$ is Borel measurable on $\Sigma(n)$ for each n. Since $\sigma' < \infty$ with probability 1, γ_0' is finitely measurable. Clearly, also, $\delta_0' = (\gamma_0', \sigma')$ satisfies (1.1.3), and $\delta_0' \varepsilon \mathcal{G}$.

For the second part of the lemma, suppose for some n the set on the right of (2.2.2) contains more than one point for a set of positive μ measure in $\Sigma(n)$. Since $\gamma_0'(y)$ is measurable, the sets $Q_j = \{(v, y) : y \varepsilon \Sigma(n), v \geqq \gamma_0'(y) + 1/j\}$ are measurable in E_n. For some $j < \infty$ the set Q_j has a projection (on $\{y\}$) of positive μ measure. For this j define $\bar{\gamma}_0(y) = \gamma_0'(y)$ if $y \not\varepsilon Q_j$, $= \inf \{v : v \geqq \gamma_0'(y) + 1/j, f(v, y) = \inf_t f(t, y)\}$ if $y \varepsilon Q_j$. Using the methods of the first part $\bar{\delta}_0 = (\bar{\gamma}_0, \sigma') \varepsilon \mathcal{G}$ and $R(\bar{\delta}_0) = R(\delta_0')$. This proves the "only if" assertion; the "if" assertion is obvious.

For the third part of the lemma, suppose $(\delta_k', \sigma') \varepsilon \mathcal{G}$ and $R(\delta_k') \to R(\delta_0')$.

Then $R(\delta_k' \mid y) \geqq R(\delta_0' \mid y)$. Since $\int [R(\delta_k' \mid y) - R(\delta_0' \mid y)]\mu(dy) \to 0$ it follows that $R(\delta_k' \mid y) \to R(\delta_0' \mid y)$ in measure (μ). Finally since δ_0' is unique a.e. (μ) as above, since $R((t, \sigma') \mid y)$ is lower-semi-continuous in t for $-\infty < t < \infty$, and since, by assumption, $\lim \inf_{t\to\pm\infty} R((t, \sigma') \mid y) > R(\delta_0' \mid y)$ it follows that $\gamma_k' \to \gamma_0'$ in measure (μ) as $k \to \infty$. This completes the proof of the lemma.

Suppose for each k, either $W(t, k, y) = \infty$ or

$$\lim \inf_{t\to\pm\infty} W(t, k, y) \geqq W(v, k, y) \quad \text{for all } v \text{ a.e. } (\mu) \tag{2.2.5}$$

and that for almost all y strict inequality holds in (2.2.5) for a set of values of v having positive Lebesgue measure. Then it is clear that γ_0' as defined in (2.2.2) satisfies $|\gamma_0'(y)| < \infty$ a.e. (μ), and that (2.2.3) is satisfied. Thus the following lemma for the fixed-sample-size case follows immediately from Lemma 2.2.1.

LEMMA 2.2.2. *In the fixed sample size case suppose there is a* $\delta \varepsilon \mathcal{J}$ *such that* $R(\delta) < \infty$. *Suppose* W *and* p *are Borel measurable functions and* $\lim \inf_{t\to\pm\infty} W(t, y) \geqq W(v, y)$ *for all* v a.e. (μ) *and*

$$\int [\lim \inf_{t\to\pm\infty} W(t, y) - W(v, y)]\, dv > 0 \quad \text{a.e.} \quad (\mu).$$

Then a best invariant estimator, say δ_0, *exists. Assumption* (2.1.1) *is satisfied if and only if this estimator is uniquely determined* a.e. (μ).

We now turn to the general sequential case.

LEMMA 2.2.3. *Suppose for some* $\delta \varepsilon \mathcal{J}$, $R(\delta) < \infty$. *Suppose for every fixed* $j \varepsilon J$, W *and* p *are Borel measurable, and*

$$\lim \inf_{t\to\pm\infty} W(t, j, y) \geqq W(t, j, y) \quad \text{a.e.} \quad (\mu) \tag{2.2.6}$$

with strict inequality holding for a set of values of t *having positive measure* (a.e. (μ) *and for all* $j < \infty$). *Suppose also that*

$$\lim \inf_{j\to\infty} \inf_t W(t, j, y) = \infty \quad \text{a.e.} \quad (\mu) \tag{2.2.7}$$

and (as in (1.1.4)) $W(t, \infty, y) = \infty$. *Then there exists at least one best invariant estimator. Furthermore, assumption* (2.1.1) *is satisfied if and only if the best invariant estimator is uniquely determined* a.e. (μ).

PROOF. As a technical device we introduce the possibility of randomized invariant estimators. It is convenient here to use the notation of Section 1.3 with one slight change. Here, we let $Y_0' = Z$ (instead of $Y_1' = Z$) be a real random variable uniformly distributed on $(0, 1)$, measurable with respect to a field labeled $\mathcal{A}_0$. Z is observed without cost after observing X and before Y_1. To be consistent in our notation Y_i is still measurable with respect to $\mathcal{A}_{i+1}$, and $\sigma(z, y) = n \geqq 2$ means (as in the non-randomized case) that the sampling stops after observing $X, Z, Y_1, \cdots, Y_{n-1} = X, Y_0', Y_1', \cdots, Y_{n-1}'$.

Also $\mathcal{A}^{(i)} = \mathcal{B}(\bigcup_{k=0}^{i} \mathcal{A}_i)$. By convention, $\sigma(z, y) = 1$ will mean (for this lemma only) that sampling stops after observing X *and* Z. It is never advantageous to stop after observing X but before observing Z, and that possibility is not allowed for in this notation.

For each $y = y_1, y_2, \cdots$ and any $\sigma \varepsilon \mathcal{J}$ let

$$\tau_k(y, \sigma) = \Pr\{z : \sigma(z, y) = k\} = \int_{\{z:\sigma(z,y)=k\}} dz. \tag{2.2.8}$$

In words, $\tau_k(y, \sigma)$ is the probability of stopping after observing $y_1, y_2, \cdots, y_{k-1}$ when using the rule $\sigma \varepsilon \mathcal{G}$. (Clearly, τ_k is measurable with respect to $\mathcal{A}^{(k)}$.)

Using (2.2.6) and Lemma 2.2.1 it follows that for any fixed stopping rule $\sigma' \varepsilon \mathcal{G}$ there exists at least one procedure $\delta \varepsilon \mathcal{G}$ such that $R(\delta) = \inf_{\delta=(\gamma,\sigma')\varepsilon\mathcal{G}} R(\delta)$. Call this estimation procedure $\delta_{\sigma'}$, and $\delta_{\sigma'} = (\gamma_{\sigma'}, \sigma')$. As in Section 1.3, we may (and will) assume that $\gamma_{\sigma'}$ is non-randomized, i.e. is not a function of z given the value of σ'.

Let $\gamma^{(k)}$ be a best estimate corresponding to the stopping rule $\sigma(y) \equiv k$ as above. And let $\delta^{(k)} = (\gamma^{(k)}, k)$. Then for any $\sigma \varepsilon \mathcal{G}$,

$$R(\gamma_\sigma, \sigma) = \sum_{k=1}^{\infty} \int \tau_k(y, \sigma) R(\delta^{(k)} \mid y)\mu(dy). \tag{2.2.9}$$

Now, let $\delta_i = (\gamma_i, \sigma_i)$ be any sequence of estimators in $\mathcal{G}$ such that $R(\delta_i) \to \inf_{\delta\varepsilon\mathcal{G}} R(\delta)$. Since the functions $\tau_k(y, \sigma_i)$, $k = 1, 2, \cdots$, are a countable set of bounded functions of y, there is a subsequence $\{i'\} \subset \{i\}$ such that the functions $\tau_k(y, \sigma_{i'})$ converge weakly (μ) to a limit, say $\tau_k(y)$, $k = 1, 2, \cdots$. These functions τ_k are measurable with respect to $\mathcal{A}^{(k)}$, hence they determine a stopping rule, say σ^*. The fact that $R(\delta) < \infty$ for some $\delta \varepsilon \mathcal{G}$ and (2.2.7) guarantee that $\sigma^* < \infty$ a.e. (μ). Clearly then $\sigma^* \varepsilon \mathcal{G}$. Associated with σ^* is a procedure $\delta^* = (\gamma_{\sigma^*}, \sigma^*)$. It remains to be shown that $R(\delta^*) = \inf_{\delta\varepsilon\mathcal{G}} R(\delta)$.

We note first that for any fixed k,

$$\int \tau_k(y, \sigma_{i'}) R(\delta^{(k)} \mid y)\mu(dy) \to \int \tau_k(y) R(\delta^{(k)} \mid y)\mu(dy) \quad \text{as} \quad i' \to \infty. \tag{2.2.10}$$

If $\int R(\delta^{(k)} \mid y)\mu(dy) < \infty$ this follows from the definition of weak convergence. If this integral is infinite the result (2.2.10) is not quite standard, though very easy to prove. Note that for any positive number t, $\int \inf (R(\delta^{(k)} \mid y), t)\mu(dy) < \infty$. Hence, for each fixed $t > 0$ as $i' \to \infty$,

$$\int \tau_k(y, \sigma_{i'}) \inf (R(\delta^{(k)} \mid y), t)\mu(dy) \to \int \tau_k(y) \inf (R(\delta^{(k)} \mid y), t)\mu(dy).$$

In addition,

$$\lim_{t\to\infty} \int \tau_k(y, \sigma_{i'}) \inf (R(\delta^{(k)} \mid y), t)\mu(dy) = \int \tau_k(y, \sigma_{i'}) R(\delta^{(k)} \mid y)\mu(dy)$$

and

$$\lim_{t\to\infty} \int \tau_k(y) \inf (R(\delta^{(k)} \mid y), t)\mu(dy) = \int \tau_k(y)(R(\delta^{(k)} \mid y)\mu(dy) \leqq \infty.$$

These equations prove that (2.2.10) is valid.

Using (2.2.9) and (2.2.10)

$$\begin{aligned} \lim_{i'\to\infty} R(\gamma_{\sigma_{i'}}, \sigma_{i'}) &= \lim_{i'\to\infty}{}' \sum_{k=1}^{\infty} \int \tau_k(y, \sigma_{i'}) R(\delta^{(k)} \mid y)\mu(dy) \\ &\geqq \sum_{k=1}^{\infty} \int \tau_k(y) R(\delta^{(k)} \mid y)\mu(dy) \\ &= R(\delta^*). \end{aligned} \tag{2.2.11}$$

Since

$$\begin{aligned} \inf_{\delta\varepsilon\mathcal{G}} R(\delta) &= \lim_{i'\to\infty} R(\delta_{i'}) \geqq \lim_{i'\to\infty} R(\delta_{\sigma'}) \\ &\geqq R(\delta^*) \geqq \inf_{\delta\varepsilon\mathcal{G}} R(\delta), \end{aligned}$$

it follows that $R(\delta^*) = \inf_{\delta \varepsilon \mathcal{J}} R(\delta) = R_0$. Hence there exists at least one best invariant randomized estimator. Then by the reasoning of Section 1.3 there exists at least one best invariant non-randomized estimator, say δ_0. Thus δ_0 is a best invariant estimator for the original (non-randomized) problem.

To prove the second assertion of the lemma, suppose that δ_0 is uniquely determined a.e. (μ) in the original problem. Then δ^* is uniquely determined a.e. in the randomized problem; and the functions $\tau_k(y, \sigma^*)$ are 0-1 functions a.e., for $k = 1, 2, \cdots$. That is, $\tau_k(y, \sigma^*) = 1$ if $\sigma_0(y) = k$, and $= 0$ otherwise, a.e. (μ). Let $\delta_i = (\gamma_i, \sigma_i)$ be any sequence of procedures (in the randomized problem) with $R(\delta_i) \to R_0$. Then there is a subsequence $\{i'\} \subset \{i\}$ such that $\tau_k(y, \sigma_{i'}) \to \tau_k(y, \sigma^*)$ weakly.

We assert that $\tau_k(y, \sigma_{i'}) \to \tau_k(y, \sigma^*)$ in measure (μ). Note first that $0 \leqq \tau_k(y, \sigma_i) \leqq 1$. From the weak convergence

$$\int_{\{y:\tau_k(y,\sigma^*)=1\}} [\tau_k(y, \sigma^*) - \tau_k(y, \sigma_{i'})]\mu(dy) \to 0. \tag{2.2.12}$$

Since the integrand of (2.2.12) is non-negative, it follows that $\tau_k(y, \sigma_{i'}) \to 1$ in measure (μ) on the set where $\tau_k(y, \sigma^*) = 1$. Similarly for the set where $\tau_k(y, \sigma^*) = 0$. The assertion of this paragraph is thus verified. It follows immediately that $\tau_k(y, \sigma_i) \to \tau_k(y, \sigma^*)$ in measure (μ), and then that $\sigma_i \to \sigma^*$ in measure.

It is now almost immediate that $\gamma_i \to \gamma^*$ in measure (μ). Thus assumption (2.1.1) is satisfied, at least in the randomized problem. This implies that assumption (2.1.1) is also satisfied in the original non-randomized problem.

This completes the proof of the lemma.

2.3. *Some lemmas concerning Assumption* 2.1.2. The lemmas in this section describe conditions that imply the validity of assumption (2.1.2), i.e.

$$\int_0^\infty d\lambda\{\sup_{\delta\varepsilon\mathcal{J}} \int \mu(dy) \int_{-\lambda}^{\lambda} [W(x + \gamma_0(y), \sigma_0(y), y) - W(x + \gamma(y), \sigma(y), y)]p(x, y)\, dx\} < \infty.$$

In the fixed-sample-size case it is shown that (2.1.2) is very nearly implied by the moment condition, (2.1.3), of Theorem 2.1.1. Apart from some particular sequential cases it seems much more difficult to prove that (2.1.2) is valid in the sequential case than in the fixed-sample-size case. We have been able to state conditions which can be easily interpreted in the case of the special sequential estimation problem described in Section 1.2.

In order that the results may be easily compared all the lemmas are collected together in the first part of this section. Their proofs are all given in the last part of this section.

Throughout this section (as in all of Section 2), we assume the general setup of Section 1.1 with dimension $m = 1$.

The first lemma is quite obvious. It merely states formally the fact that (2.1.2) depends only on the tails (in x) of $p(x, y)$.

LEMMA 2.3.1. *Suppose there exists a* $\lambda_0 < \infty$ *such that*

$$\int \mu(dy) \int_{-\lambda_0}^{\lambda_0} p(x, y)\, dx = 1. \tag{2.3.1}$$

Then assumption (2.1.2) *is satisfied.*

The next lemma applies to the sequential case when the sample size of any procedure having finite risk is bounded by, say, $N < \infty$. In particular this lemma applies in the fixed sample size case. Conditions (2.3.2) and (2.3.3) of the lemma are closely related to (2.1.3).

LEMMA 2.3.2. *Suppose there exists an $N < \infty$ such that $W(t, n, y) \equiv \infty$ for $n > N$. Suppose there is a number $0 \leqq \beta < \infty$ such that either for all t, n, y $(n \leqq N) W(t, n, y) < \beta$ and*

$$\int \mu(dy) \int |x|\, p(x, y)\, dx < \infty; \tag{2.3.2}$$

or, that for each n, y $(n \leqq N) W(t, n, y)$ is non-increasing in t for $t < -\beta$ and non-decreasing in t for $t > \beta$, and

$$\int \mu(dy) \int |x \sup_{|\tau|<2|x|+\beta} W(\tau, n, y)|\, p(x, y)\, dx < \infty. \tag{2.3.3}$$

Then (2.1.2) is satisfied.

Recall from Section 1.2 that in the fixed-sample-size case a best invariant estimator on the basis of X, Y is admissible if and only if a best invariant estimator on the basis of the transformed variables $X' = X + \tau(Y)$, $Y' = Y$ is admissible. Thus it may be that (2.3.2) (and even (2.1.2)) fail to be satisfied for X, Y; but that the problem may be transformed, and (2.3.2) (and (2.1.2)) will be satisfied in the transformed problem. (Generally speaking such a transformation will not affect the validity of any of the other hypotheses of Corollary (2.2.2).) These remarks apply also to (2.1.3) in Theorem 2.1.1 and to Lemma 2.3.3.

Note also that if W is uniformly continuous, (2.3.3) is implied by the condition $\int \mu(dy) \int x^2 p(x, y)\, dx < \infty$.

The next lemma applies to two important fixed-sample-size cases: confidence interval estimation and convex loss functions.

LEMMA 2.3.3. *Suppose the problem is of fixed-sample-size type, and suppose $\gamma_0(x) = 0$.*

(a) *If there exist $\alpha, \beta < \infty$ such that $W(t, y) \equiv \alpha$ for $|t| > \beta$ and $\sup_t W(t, y) = \alpha$ then (2.1.2) is satisfied.*

Or,

(b) *if $W(t, y) = W(t)$ is a convex function of t and there is a $\xi_0 > 0$ such that*

$$\int \mu(dy) \int |x^2 W'(x + \xi)|\, p(x, y)\, dx < \infty \tag{2.3.4}$$

for all ξ such that $|\xi| < \xi_0$. Then (2.1.2) is satisfied $(W'(x) = (d/dx)W(x))$.

In Lemma 2.3.3 we have assumed for convenience in the proof that $\gamma_0(x) = 0$, which may always be achieved by a change of variables. If this assumption is removed then the condition (2.3.4) must be replaced by

$$\int \mu(dy) \int (x^2 + \gamma_0^2(y))|W'(x + \gamma_0(y) + \xi)|\, p(x, y)\, dx < \infty, \tag{2.3.4'}$$

and the statement of condition (a) must also be altered. We leave the necessary modifications to the interested reader.

The remaining two lemmas (whose hypotheses can be made slightly weaker) deal with the true sequential problem, where the sample size may be unbounded. They deal only with the frequently treated case where

$$W(t, n, y) = W_1(t) + W_2(n), \qquad W_1 \geqq 0. \tag{2.3.5}$$

The second lemma is a (non-trivial) specialization of the first to the special sequential problem (see Section 1.2).

Let $\mathcal{G}_n$ denote the set of invariant estimates which depend only on the first n-observations $(x, y_1, \cdots, y_{n-1})$ i.e. $\epsilon(x, y) \varepsilon \mathcal{G}_n$ if and only if $\epsilon(x, y) = \epsilon(x, y_1, y_2, \cdots, y_{n-1}) = x + \gamma(y_1, \cdots, y_{n-1})$. Recall from (1.1.1) that $p_n(x, y)$ is the conditional density of x given $y_1, \cdots, y_{n-1}$ taken as a measurable function on $\mathfrak{A}^{(n)}$.

LEMMA 2.3.4. *Let W satisfy* (2.3.5). *Suppose for some $c > 0$, $W_2(n) - W_2(n-1) \geqq c$, $n = 0, 1, 2, \cdots$, and $\sup_t W_1(t) > c$, and, for some k, $1 \leqq k < \infty$, $W_2(n) = O(n^k)$. Assume for some $\alpha > 0$;*

$$\int \mu(dy) \int |x|^{1+\alpha} p(x, y)\, dx < \infty. \tag{2.3.6}$$

Assume there is a sequence γ_n, $\gamma_n \varepsilon \mathcal{G}_n$, $n = 1, 2, \cdots$, such that

$$\mu\{y: \int W_1(x + \gamma_n(y)) p(x, y)\, dx > c/2\} = O(n^{-k(1+2/\alpha)+2}) \quad \text{as} \quad n \to \infty. \tag{2.3.7}$$

Suppose either (a) *W_1 is bounded; or* (b) *$W_1(t)$ is non-increasing for $\gamma < -\beta$, non-decreasing for $\gamma > \beta$, and*

$$\int \mu(dy) \int |x \sup_{|\tau|<2|x|+\beta} W_1(\tau)|\, p(x, y)\, dx < \infty. \tag{2.3.3$'$}$$

Then (2.1.2) *is satisfied.*

It should be remarked that in (2.3.7) the constant $c/2$ has been chosen for convenience in the proof. Any other constant less than c could replace $c/2$ in (2.3.7).

A condition analogous to (2.3.4$'$) may be substituted for conditions (a) or (b) in the above lemma. Certain sections of the proof must then be altered. Since it occurs only rarely that (2.3.4$'$) is satisfied when (2.3.3$'$) is not, we leave it to the interested reader to prove this additional result.

Recall that in the special sequential problem the independent identically distributed random variables $X_1, X_2, \cdots$ are observed. $p(x, y)$, given by (1.2.1), is computed from f, the probability density function of X_1. Let the probability distribution function of X_1 when $\theta = 0$ be F.

LEMMA 2.3.5. *In the special sequential problem, suppose W is given by* (2.3.5) *where $W_2(n+1) - W_2(n) \geqq c$, and suppose $W_1(t) < d < c/2$ on an interval $\zeta_1 < t < \zeta_2$ such that $F(\zeta_2) > F(\zeta_1)$, and $W_2(n) = O(n^k)$.*

Assume for some $\alpha > 0$,

$$\int |x|^{1+\alpha} f(x)\, dx < \infty. \tag{2.3.6$'$}$$

Suppose either (a) *W_1 is bounded; or* (b) *$W_1(t)$ is non-increasing for $t < -\beta$, non-decreasing for $t > \beta$, and*

$$\int |x| \sup_{|\tau|<2|x|+\beta} W_1(\tau) f(x)\, dx < \infty. \tag{2.3.3$''$}$$

Then (2.1.2) *is satisfied.*

As in Lemma 2.3.4 any constant less than c an replace $c/2$ in the above lemma.

Note that the hypotheses of this lemma are non-sequential and fairly easy to verify in any particular case.

The proofs of these lemmas will now be given. As in Section 2.1, let

$$\omega(z, \gamma_0, \sigma_0, \gamma, \sigma, y) = W(z + \gamma_0, \sigma_0, y) - W(z + \gamma, \sigma, y) \tag{2.3.8}$$

and we shall also often omit writing the argument y of $(\gamma(y), \sigma(y)) \varepsilon \mathcal{G}$. Also, let

$$I = \int_0^\infty d\lambda\{\sup_{\delta\varepsilon\mathcal{G}} \int \mu(dy) \int_{-\lambda}^{\lambda} \omega(x, \gamma_0(y), \sigma_0(y), \gamma(y), \sigma(y), y)p(x, y)\, dx\}. \tag{2.3.9}$$

Condition (2.1.2), then, is that $I < \infty$.

The technique of "truncating" estimators (as used in (2.3.15)) is central to the proof of several of these lemmas. The "truncated" estimator is at least as good as the original estimator and the "tail integral" of its risk is finite (as in (2.3.17)), whereas the tail integral of the risk of the original estimator need not be finite.

Proof of Lemma 2.3.1. As a consequence of (2.3.1) for $\lambda > \lambda_0$,

$$\sup_{\delta\varepsilon\mathcal{G}} \int\int_{-\lambda}^{\lambda} \omega(x, \gamma_0, \sigma_0, \gamma, \sigma, y)p(x, y)\, dx\mu(dy) = 0. \tag{2.3.10}$$

Hence,

$$I \leqq \int_0^{\lambda_0} d\lambda \int \mu(dy) \int W(x + \gamma_0(y), \sigma_0(y), y)p(x, y)\, dx \leqq \lambda_0 R_0 < \infty. \tag{2.3.11}$$

Proof of Lemma 2.3.2. Since δ_0 is the best invariant estimator

$$\int \mu(dy) \int_{-\infty}^{\infty} \omega(x, \gamma_0, \sigma_0, \gamma, \sigma, y)p(x, y)\, dx \leqq 0. \tag{2.3.12}$$

Hence,

$$\int \mu(dy) \int_{-\lambda}^{\lambda} \omega(x, \gamma_0, \sigma_0, \gamma, \sigma, y)p(x, y)\, dx \leqq -\int \mu(dy) \int_{|x|>\lambda} \omega(x, \gamma_0, \sigma_0, \gamma, \sigma, y)p(x, y)\, dx. \tag{2.3.13}$$

If $W < \beta$ then using (2.3.2),

$$\begin{aligned} I &\leqq -\int d\lambda \int \mu(dy) \int_{|x|>\lambda} \omega(x, \gamma_0, \sigma_0, \gamma, \sigma, y)p(x, y)\, dx \\ &\leqq \beta \int d\lambda \int \mu(dy) \int_{|x|>\lambda} p(x, y) dx \\ &= \beta \int \mu(dy) \int |x|\, p(x, y)\, dx < \infty. \end{aligned} \tag{2.3.14}$$

Suppose, on the other hand, W satisfies the monotonicity assumptions in Lemma 2.3.2. If $\delta' \varepsilon \mathcal{G}$ is any estimation procedure, then the procedure $\hat{\delta}'$ given by $\hat{\sigma}' = \sigma'$ and

$$\begin{aligned} \hat{\gamma}'(y) &= -\lambda - \beta, & \gamma'(y) &< -\lambda - \beta, \\ &= \gamma'(y), & |\gamma'(y)| &\leqq \lambda + \beta, \\ &= \lambda + \beta, & \gamma'(y) &> \lambda + \beta, \end{aligned} \tag{2.3.15}$$

satisfies

$$\int_{-\lambda}^{\lambda} W(x + \hat{\gamma}', \hat{\sigma}', y)p(x, y)\, dx \leqq \int_{-\lambda}^{\lambda} W(x + \gamma', \sigma', y)p(x, y)\, dx. \tag{2.3.16}$$

Hence, it is sufficient to consider $\sup_{\delta \varepsilon \mathfrak{L}}$ of the appropriate integral in the definition of I, where $\mathfrak{L} = \{\delta : \delta = (\delta, \sigma) \varepsilon \mathfrak{g}, |\gamma| \leqq \lambda + \beta\}$. If this is done then using (2.3.13) and (2.3.3)

$$\begin{aligned} I &\leqq \sup_{\delta \varepsilon \mathfrak{L}} \{\textstyle\int_0^\infty d\lambda \int \mu(dy) \int_{|x|>\lambda} W(x + \gamma(y), \sigma(y), y)p(x, y)\, dx\} \\ &\leqq \textstyle\int_0^\infty d\lambda \int_{|x|>\lambda} \sup_{|\tau(y)| \leqq \lambda+\beta} W(x + \tau(y), \sigma(y), y)p(x, y)\, dx \\ &\leqq \textstyle\sum_{n=1}^N \int \mu(dy) \int_{-\infty}^{\infty} p(x, y)\, dx \int_0^{|x|} \sup_{|\tau|<2x+\beta} W(\tau, n, y)\, d\lambda \\ &= \textstyle\sum_{n=1}^N \int \mu(dy) \int |x \sup_{|\tau|<2|x|+\beta} W(\tau, n, y)|\, p(x, y)\, dx < \infty. \end{aligned} \tag{2.3.17}$$

This concludes the proof of Lemma 2.3.2.

Proof of Lemma 2.3.3. (a) If $W(t, y) \equiv \alpha$ for $|t| > \beta$ and $\sup_t W(t, y) = \alpha$ then for $\lambda > \beta$ and any $\delta \varepsilon \mathfrak{g}$,

$$\int \mu(dy) \int_{|x|>\lambda} W(x + \gamma(y), y)p(x, y)\, dx \leqq \int \mu(dy) \int_{|x|>\lambda} W(x, y)p(x, y)\, dx.$$

Since $\gamma_0(y) \equiv 0$ is the best invariant estimator it follows using the above equation that for $\lambda > \beta$,

$$\int \mu(dy) \int_{-\lambda}^{\lambda} W(x + \gamma(y), y)p(x, y)\, dx \geqq \int \mu(dy) \int_{-\lambda}^{\lambda} W(x, y)p(x, y)\, dx.$$

(Otherwise $\gamma(y)$ would be a better estimator than γ_0.) Hence

$$\sup_{\delta \varepsilon \mathfrak{g}} \int_{-\lambda}^{\lambda} \omega(x, \gamma_0, \sigma_0, \gamma, \sigma, y)p(x, y)\, dx = 0. \tag{2.3.18}$$

Thus as in (2.3.11), $I \leqq \beta R_0$.

(b) In the case of convex W, using (2.3.4),

$$0 = (d/d\xi) \int W(x + \xi)p(x, y)\, dx \,|_{\xi=0} = \int W'(x)p(x, y)\, dx. \tag{2.3.19}$$

Since W is convex $W(x) - W(x + \gamma) \leqq -\gamma W'(x)$. By assumption W assumes its minimum at some point, say β. Then reasoning as in Lemma 2.3.2 the supremum on the left of (2.3.20) (below) occurs for a γ with $|\gamma| < \lambda + \beta$. Define $\mathfrak{B}_\lambda = \{\gamma : |\gamma| < \lambda + \beta\}$. Then, using (2.3.19),

$$\begin{aligned} &\textstyle\int d\lambda \{\sup_\gamma \int \mu(dy) \int_{-\lambda}^{\lambda} (W(x) - W(x + \gamma(y)))p(x, y)\, dx\} \\ &\quad\leqq \textstyle\int d\lambda \{\sup_{\gamma \varepsilon \mathfrak{B}_\lambda} \int \mu(dy) \int_{-\lambda}^{\lambda} -\gamma(y)W'(x)p(x, y)\, dx\} \\ &\quad\leqq \textstyle\int d\lambda \{\sup_{\gamma \varepsilon \mathfrak{B}_\lambda} \int \mu(dy) \int_{|x|>\lambda} \gamma(y)W'(x)p(x, y)\, dx\}. \end{aligned} \tag{2.3.20}$$

Thus to continue,

$$\begin{aligned} I &\leqq \textstyle\int \mu(dy) \int |W'(x)|\, p(x, y)\, dx \int_0^{|x|} (\lambda + |\beta|)\, d\lambda \\ &\leqq \textstyle\int \mu(d_y) \int ((x^2/2) + |\beta x|)|W'(x)|\, p(x, y)\, dx < \infty. \end{aligned} \tag{2.3.21}$$

This completes the proof of Lemma 2.3.3.

Proof of Lemma 2.3.4. Let $\delta^* = (\gamma^*, \sigma^*)$ be the estimation procedure (invariant) which stops the first time $\int W_1(x + \gamma_n(y))p_n(x, y)\, dx \leqq c/2$ and makes the estimate γ_n when it stops. As a consequence of (2.3.7), $\Pr\{\sigma^* \geqq n\} = O(n^{-k(1+2/\alpha)+2})$ so that

$$\int (W_2(\sigma^*(y)))^{1+2/\alpha}\mu(dy) < \infty. \tag{2.3.22}$$

(In particular, σ^* stops with probability one and δ^* has finite risk.)

It follows from the definition of σ^* and the property $W_2(n) - W_2(n-1) \geqq c$ that $\sigma_0(y) \leqq \sigma^*(y)$ a.e. (μ). (If $y_1, \cdots, y_n$ with $\sigma^*(y_1, \cdots, y_n) = n$ has been observed, then the expected value of W_1 given $y_1, y_2, \cdots, y_n$, and hence the maximum amount the expectation of W_1 could be decreased by further sampling, is less than the cost of taking one more observation.)

In the bounded sample size case of Lemma 2.3.2b the first step of the proof was to construct for any given $\delta \varepsilon \mathcal{g}$ another (truncated) estimator $\hat{\delta} = (\hat{\gamma}, \sigma)$ which gives a value for the integral in (2.3.16) which is not larger than the value given by δ. The condition (2.1.2) then only needed to be verified for the set of all possible $\hat{\delta}$ rather than for the set of all δ in $\mathcal{g}$. In the sequential case the analog of $\hat{\delta}$ is a procedure called δ^+ which may be different from δ in both its estimator and its stopping rule. After constructing δ^+ in order to prove that (2.1.2) is satisfied we have found it necessary to introduce as an analytical device a variant of δ^+, called $\bar{\delta}^+$, and defined with the aid of the construction of δ^* above. The heart of this verification is the steps (2.3.29)–(2.3.35), which conclude the proof of the lemma.

It seems desirable at this point to use the theory of stochastic processes. As before, let $H = E_1 \times E_1 \times \cdots$ and let $\mathcal{A}^{(\infty)}$ be the usual σ-field on H. Let $\hat{\mathcal{A}}_n$ be the σ-field on H induced by the random variables $Y_1, Y_2, \cdots, Y_{n-1}$. Then $E(X \mid \hat{\mathcal{A}}_n)$, $n = 1, 2, \cdots, \infty$, is a martingale (see (2.3.6)). Let

$$J_\lambda(X) = 1 \quad \text{if} \quad |X| \leqq \lambda$$
$$= 0 \quad \text{if} \quad |X| > \lambda, \tag{2.3.23}$$

then $E(J_\lambda \mid \hat{\mathcal{A}}_n)$ and $E(1 - J_\lambda \mid \hat{\mathcal{A}}_n)$, $n = 1, 2, \cdots, \infty$, are also martingales.

We assert that

$$\begin{aligned} \inf_{\delta\varepsilon\mathfrak{D}_\lambda} \int \mu(dy) \int_{-\lambda}^{\lambda} W(x + \gamma(y), \sigma(y))p(x, y)\,dx \\ = \inf_{\delta\varepsilon\mathcal{g}} \int \mu(dy) \int_{-\lambda}^{\lambda} W(x + \gamma(y), \sigma(y))p(x, y)\,dx, \end{aligned} \tag{2.3.24}$$

where $\mathfrak{D}_\lambda$ is defined by the following: Let

$$\begin{aligned} \mathfrak{D}_\lambda' = \{\delta : \delta = (\gamma, \sigma) \varepsilon \mathcal{g}, \sigma \leqq n \quad &\text{on the} \\ \hat{\mathcal{A}}_n\text{-measurable set where both} \\ E\{\textstyle\int_{-\lambda}^{\lambda} p_n(x, y)\,dx \mid \hat{\mathcal{A}}_n\} > \tfrac{1}{2} \quad &\text{and} \\ \textstyle\int_{-\infty}^{\infty} W_1(x + \gamma_n(y))p_n(x, y)\,dx < \tfrac{1}{2}c\} \end{aligned} \tag{2.3.25}$$

(γ_n is as previously defined). Then define

$$\mathfrak{D}_\lambda = \{\delta : \delta \varepsilon \mathfrak{D}_\lambda', |\gamma(y)| < \lambda + \beta\} \tag{2.3.25a}$$

where in case (a) we take $\beta = \infty$. (Thus in case (a) $\mathfrak{D}_\lambda = \mathfrak{D}_\lambda'$.) To prove this assertion, consider any $\delta = (\gamma, \sigma) \varepsilon \mathfrak{g}$. Let $\delta^+ = (\gamma^+, \sigma^+)$ be constructed from δ by the following truncation process: If $\sigma(y) = n$ $(y = y_1, y_2, \cdots)$ but for some $n' < n$, $\int W_1(x + \gamma_{n'}(y))p_n(x, y)\, dx < \frac{1}{2}c$ and at y the function $E\{\int_{-\lambda}^{\lambda} p_n(x, y)\, dx \mid \hat{\mathfrak{a}}_{n'}\}$ is $> \frac{1}{2}$ let $\sigma^+(y)$ be the least such value of n', and let $\gamma^+ = \hat{\gamma}_{\sigma^+(y)}(y)$ ($\hat{\gamma}$ is defined in (2.3.15). Otherwise let $\sigma^+(y) = \sigma(y)$ and $\gamma^+(y) = \gamma(y)$. It is an easy matter to verify that $\delta^+ \varepsilon \mathfrak{g}$, (i.e. that δ^+ satisfies 1.1.3 a.e. (μ)), and hence $\delta^+ \varepsilon \mathfrak{D}_\lambda$. Now,

$$\begin{aligned}
&\int \mu(dy) \int_{-\lambda}^{\lambda} [W(x + \gamma(y), \sigma(y)) - W(x + \gamma^+(y), \sigma^+(y))]p(x, y)\, dx \\
&\quad = \int_{\{y: \delta^+(y) \neq \delta(y)\}} \mu(dy) \int_{-\lambda}^{\lambda} [W_1(x + \gamma(y)) - W_1(x + \gamma^+(y)) \\
&\qquad + W_2(\sigma(y)) - W_2(\sigma^+(y))]p(x, y)\, dx \qquad (2.3.26)\\
&\quad \geqq \int_{\{y: \delta^+(y) \neq \delta(y)\}} \mu(dy) \int_{-\lambda}^{\lambda} [-W_1(x + \gamma^+(y)) + c]p(x, y)\, dx \\
&\quad \geqq -\tfrac{1}{2}c + c \int_{\{y: \delta^+(y) \neq \delta(y)\}} \mu(dy) \int_{-\lambda}^{\lambda} p(x, y)\, dx \\
&\quad \geqq -\tfrac{1}{2}c + \tfrac{1}{2}c = 0.
\end{aligned}$$

Since for every $\delta \varepsilon \mathfrak{g}$ there is a $\delta^+ \varepsilon \mathfrak{D}_\lambda$ satisfying (2.3.26), the assertion (2.3.24) is proved.

If $\delta = (\gamma, \sigma) \varepsilon \mathfrak{g}$ define $\tilde{\delta} = (\tilde{\gamma}, \tilde{\sigma})$ by

$$\begin{aligned}
\tilde{\sigma}(y) &= \inf (\sigma(y), \sigma^*(y)), \\
\tilde{\gamma}(y) &= \gamma(y) \qquad \text{if} \quad \tilde{\sigma}(y) = \sigma(y) \qquad (2.3.27)\\
&= \gamma^*(y) \qquad \text{if} \quad \tilde{\sigma}(y) < \sigma(y).
\end{aligned}$$

It can be easily checked that $\tilde{\delta} \varepsilon \mathfrak{g}$. Note that if $\delta \varepsilon \mathfrak{D}_\lambda$, then (definition)

$$K = \{y: \delta(y) \neq \tilde{\delta}(y)\} \subset \{y: \exists n \ni E\{J_\lambda \mid \hat{\mathfrak{a}}_n\}(y) \leqq \tfrac{1}{2}\} \qquad (2.3.28)$$

where $E\{J_\lambda \mid \hat{\mathfrak{a}}_n\}(y)$ denotes the value of the $\hat{\mathfrak{a}}_n$-measurable function $E\{J_\lambda \mid \hat{\mathfrak{a}}_n\}$ at the point y.

If $\gamma^+, \sigma^+ \varepsilon \mathfrak{D}_\lambda$, using (2.3.13),

$$\begin{aligned}
&\int \mu(dy) \int_{-\lambda}^{\lambda} [W(x + \gamma_0(y), \sigma_0(y)) - W(x + \gamma^+(y), \sigma^+(y))]p(x, y)\, dx \\
&\quad = \int \mu(dy) \int_{-\lambda}^{\lambda} [W(x + \gamma_0, \sigma_0) - W(x + \tilde{\gamma}^+, \tilde{\sigma}^+) + W(x + \tilde{\gamma}^+, \tilde{\sigma}^+) \\
&\qquad - W(x + \gamma^+, \sigma^+)]p(x, y)\, dx \\
&\quad \leqq \int \mu(dy) \int_{|x| > \lambda} W(x + \tilde{\gamma}^+, \tilde{\sigma}^+)p(x, y)\, dx \\
&\qquad + \int_K \mu(dy) \int_{-\lambda}^{\lambda} W(x + \gamma^*, \sigma^*)p(x, y)\, dx \qquad (2.3.29)\\
&\quad = \int_{K^c} \mu(dy) \int_{|x| > \lambda} W_1(x + \tilde{\gamma}^+(y))p(x, y)\, dx \\
&\qquad + \int \mu(dy) \int_{|x| > \lambda} W_2(\tilde{\sigma}^+(y))p(x, y)\, dx \\
&\qquad + \int_K \mu(dy) \int_{-\infty}^{\infty} W_1(x + \gamma^*(y))p(x, y)\, dx \\
&\qquad + \int_K \mu(dy) \int_{-\lambda}^{\lambda} W_2(\sigma^*(y))p(x, y)\, dx,
\end{aligned}$$

where $K^c = \{y: y \not\in K\}$. (Note that $(\tilde{\delta})^+ = \widetilde{(\delta^+)}$, hence we have written merely $\tilde{\delta}^+$.)

Using (2.3.27), Hölder's inequality, (2.3.22), and (2.3.6), there is a b_1 such that

$$\begin{aligned}
&\int \mu(dy) \int_{|x|>\lambda} W_2(\tilde{\sigma}^+(y))p(x, y)\, dx \\
(2.3.30) \qquad &\leqq \int \mu(dy) \int_{|x|>\lambda} W_2(\sigma^*(y))p(x, y)\, dx \\
&\leqq [\int (W_2(\sigma^*(y)))^{1+2/\alpha}\mu(dy)]^{\alpha/(2+\alpha)}[\int \mu(dy) \int_{|x|>\lambda} p(x, y)\, dx]^{2/(2+\alpha)} \\
&\leqq b_1(1/\lambda)^{(2+2\alpha)/(2+\alpha)}.
\end{aligned}$$

Since $E\{1 - J_\lambda \mid \hat{\mathfrak{a}}_n\}$ is a martingale, the martingale stopping theorem implies

$$\begin{aligned}
(2.3.31) \quad \mu\{y: \min_n E\{J_\lambda \mid \hat{\mathfrak{a}}_n\}(y) < \tfrac{1}{2}\} &= \mu\{y: \max_n E\{1 - J_\lambda \mid \hat{\mathfrak{a}}_n\}(y) > \tfrac{1}{2}\} \\
&\leqq 2 \int \mu(dy) \int_{|x|>\lambda} p(x, y)\, dx.
\end{aligned}$$

See Doob (1953), p. 314. This fact together with the definition of γ^*, (2.3.28), and (2.3.6) implies there is a $b_2 > 0$ such that

$$\begin{aligned}
\int_K \mu(dy) \int W_1(x + \gamma^*(y))p(x, y)\, dx &\leqq (c/2) \int_K \mu(dy) \\
(2.3.32) \qquad &\leqq c \int \mu(dy) \int_{|x|>\lambda} p(x, y)\, dx \\
&\leqq b_2(1/\lambda)^{1+\alpha}.
\end{aligned}$$

Finally, there is a b_3 such that

$$\begin{aligned}
&\int_K \mu(dy) \int_{-\lambda}^{\lambda} W_2(\sigma^*(y))p(x, y)\, dx \\
(2.3.34) \qquad &\leqq \int_K W_2(\sigma^*(y))\mu(dy) \\
&\leqq [\int (W_2(\sigma^*(y)))^{1+2/\alpha}\mu(dy)]^{\alpha/(2+\alpha)}[\int \mu(dy) \int_{|x|>\lambda} p(x, y)\, dx]^{2/(2+\alpha)} \\
&\leqq b_3(1/\lambda)^{(2+2\alpha)/(2+\alpha)}.
\end{aligned}$$

Combining the remarks around (2.3.25a) with (2.3.6), and (2.3.3′) as in the proof of Lemma 2.3.2 to get the bound b_4 for the first term of the last expression of (2.3.29), and using Equations (2.3.29)–(2.3.33)

$$\begin{aligned}
I &= \int_0^\infty d\lambda \sup_{\delta\varepsilon\mathcal{J}} \{\int \mu(dy) \int_{-\lambda}^{\lambda} [W(z + \gamma_0(y), \sigma_0(y)) \\
&\qquad - W(z + \gamma(y), \sigma(y))]p(x, y)\, dx\} \\
&= \int_0^\infty d\lambda \sup_{\delta\varepsilon\mathfrak{D}_\lambda} \{\int \mu(dy) \int_{-\lambda}^{\lambda} [W(z + \gamma_0(y), \sigma_0(y)) \\
(2.3.34) \qquad &\qquad - W(z + \gamma(y), \sigma(y))]p(x, y)\, dx\} \\
&\leqq b_4 + b_1 \int_1^\infty (1/\lambda)^{(2+2\alpha)/(2+\alpha)}\, d\lambda + b_2 \int_1^\infty (1/\lambda)^{1+\alpha}\, d\lambda \\
&\qquad + b_3 \int_1^\infty (1/\lambda)^{(2+2\alpha)/(2+\alpha)}\, d\lambda + R_0 \\
&< \infty.
\end{aligned}$$

This completes the proof of Lemma 2.3.4.

PROOF OF LEMMA 2.3.5. Let $\bar{W}_1(t) = \max\{W_1(t), d\}$ where (as in the statement of the lemma) $W_1(t) < d < c/2$ on $\zeta_1 < t < \zeta_2$. There is a ζ_3 such that $F(\zeta_1) < F(\zeta_3) < F(\zeta_2)$ where $F(\zeta_3) = r/s = p$ (r, s integers), a rational number. If $y_1, y_2, \cdots, y_n$, $n = vs$, v an integer, is a set of numbers, define $\pi(y_1, \cdots, y_n)$ to be the vrth largest of the y_i ($i = 1, 2, \cdots, n = vs$). Thus when, for example, $p = r/s = \frac{1}{2}$, $\pi = (y_1, \cdots, y_n)$ is (almost) the median of $y_1, \cdots, y_n$. For each integer n, let $\langle n\rangle$ denote the largest integer k such that $k \leqq n$ and $k = vs$ for some integer v. If $y = y_1, y_2, \cdots$, let $\pi_n(y) = \pi(y_1, y_2, \cdots, y_{\langle n\rangle})$.

(2.3.7) will certainly be satisfied if for any κ, $1 \leqq \kappa < \infty$,

$$\mu\{y: \textstyle\int \bar{W}_1(x + \pi_n(y))p_n(x, y)\, dx_1 > c/2\} = O(n^{-\kappa}) \tag{2.3.35}$$

where $p_n(x, y)$ is determined by f and the nature of the special sequential problem (see (1.2.1)). Since $y_i = x_{i+1} - x$, $x + \pi_n(y) = \pi_n(x_2, x_3, \cdots)$. Thus, using the definition (1.2.1) of $p(x, y)$ and μ,

$$\begin{aligned} n^\kappa \textstyle\int \mu(dy) \int [\bar{W}_1(x + \pi_n(y)) - d]p_n(x, y)\, dx \\ = n^\kappa \textstyle\int \mu(dy) \int [\bar{W}_1(\pi_n(x_2, x_3, \cdots)) - d]p_n(x, y)\, dx \\ = n^\kappa \textstyle\int [\bar{W}_1(\pi_n(x_2, x_3, \cdots)) - d] \prod_{i=2}^{\langle n\rangle+1} f(x_i)\, dx_i . \end{aligned} \tag{2.3.36}$$

It is easy to compute the distribution of π_n. If this distribution is substituted (2.3.36) becomes

$$n^\kappa \textstyle\int [\bar{W}_1(\pi) - d]\binom{\langle n\rangle-1}{(r/s)\langle n\rangle-1}\langle n\rangle(F(\pi))^{p\langle n\rangle-1}(1 - F(\pi))^{(1-p)\langle n\rangle} f(\pi)\, d\pi. \tag{2.3.37}$$

Using $\bar{W}_1(t) = d$ for $\zeta_1 < t < \zeta_2$ and Stirling's formula the expression (2.3.37) is the order of

$$\begin{aligned} O\{n^\kappa \cdot \langle n\rangle^{\frac{1}{2}} p^{-p\langle n\rangle}(1 - p)^{-(1-p)\langle n\rangle} \\ \cdot \textstyle\int [\bar{W}_1(\pi) - d](F(\pi))^{p\langle n\rangle}(1 - F(\pi))^{(1-p)\langle n\rangle} f(\pi)\, d\pi\} \\ = O\{n^\kappa \cdot n^{\frac{1}{2}}(p)^{-pn}(1 - p)^{-(1-p)n} \max_{\zeta_1 \leqq t \leqq \zeta_2} (F(t))^{pn}(1 - F(t))^{(1-p)n} \\ \cdot \textstyle\int_{\pi \leqq \zeta_1 \text{ or } \pi \geqq \zeta_2} [W_1(\pi) - d]f(\pi)\, d\pi\} \qquad \text{as } n \to \infty. \end{aligned} \tag{2.3.38}$$

(2.3.3″) proves that

$$\textstyle\int W_1(x)f(x)\, dx < \infty. \tag{2.3.39}$$

Applying (2.3.39) to (2.3.38) and simplifying yields

$$n^\kappa \textstyle\int \mu(dy) \int [\bar{W}_1(x + \pi_n(y)) - d]\, p_n(x, y)\, dx \to 0 \qquad \text{as} \qquad n \to \infty.$$

Thus (Chebyshev),

$$\mu\{y: \bar{W}_1(x + \pi_n(y)) - d \geqq c/2 - d\} = o(n^{-\kappa}) \qquad \text{as} \qquad n \to \infty, \tag{2.3.40}$$

which proves that (2.3.35) and thus (2.3.7) are satisfied.

It can be easily checked that the remaining conditions of Lemma 2.3.5 are equivalent to the remaining conditions of Lemma 2.3.4. Thus Lemma 2.3.4 may be applied to prove that (2.1.2) is satisfied. This completes the proof of Lemma 2.3.5.

2.4. *A result concerning the necessity of the conditions in Theorem* 1.1.1. The hypotheses of Theorem 2.1.1 include explicitly (in (2.1.3)) and implicitly (in (2.1.2)) certain conditions on the moments of p. In the fixed-sample-size case if W is independent of y, $W(x)$ is convex, and $W(x) \sim c|x|^k$ for some k, $1 \leqq k < \infty$, as $x \to \infty$ then these conditions will all be satisfied if

$$\int \mu(dy) \int |x^\alpha| W(x) p(x, y)\, dx < \infty \tag{2.4.1}$$

for $\alpha = 1$ (see Lemma 2.3.3b). It is of some interest to determine whether the condition $\alpha = 1$ is the best that can be obtained under such circumstances; and even further whether, perhaps, the best inv est is in general admissible if (2.4.1) holds only for $\alpha = 0$.

It is trivially true that the best invariant estimator is inadmissible if (2.4.1) is not valid for $\alpha = 0$. If (2.4.1) is not true for $\alpha = 0$ then $R_0 = \infty$ and the estimator $\delta(x, y) \equiv 0$ is better than the best inv est.

We have been able to show by an example that if $W(x) = |x|$ the condition $\alpha = 1$ is the weakest condition of the type (2.4.1) under which it can be concluded that the best inv est, if unique, is always admissible. More generally we have shown that for $W(x) = |x|^k$, $k \geqq 1$, one must at least assume (2.4.1) with $\alpha \geqq k/(2^k - 1) > 0$ in order to conclude that the best inv est is always admissible. (Note that $k/(2^k - 1) = 1$ if $k = 1$.) We do not know whether if the number of observations from a set of independent identically distributed variables is $n > 1$, the best inv est is always admissible if (2.4.1) is valid for some α, $1 > \alpha \geqq k/(2^k - 1)$. These results are collected in the following theorem in which the random variable Y is real-valued, and the notation p_n does not refer to the previous definition (1.1.1). p_n as used in this section is defined by (2.4.3) below.

THEOREM 2.4.1. *In the fixed-sample-size case suppose the loss function is* $W(x) = |x|^k$, $k > 1$. *Let* α *be any number satisfying* $0 < \alpha < k/(2^k - 1)$. *Then there is a probability density such that the best invariant estimator* $\delta(x, y) = x$ *is uniquely determined, but is inadmissible. Furthermore*

$$\int \mu(dy) \int |x|^\alpha W(x) p(x, y)\, dx < \infty. \tag{2.4.2}$$

PROOF. If $k > 1$, let

$$\begin{aligned} p_n(x, 1) &= n/4 \quad &&\text{if} \quad 1 - 1/n \leqq x \leqq 1 + 1/n \\ &= n/4 \quad &&\text{if} \quad -1 - 1/n \leqq x \leqq -1 + 1/n \\ &= 0 \quad &&\text{otherwise} \end{aligned} \tag{2.4.3}$$

and let $p_n(x, y) = (1/y)p_n(x/y, 1)$. Let $P_0(x, y)$ be the probability distribution which assigns probability $\frac{1}{2}$ to each of the points $+y$ and $-y$. Let $\mu(dy) = ((\beta + k)/y^{1+\beta+k})\, dy$, $1 \leqq y < \infty$, $\alpha < \beta < k/(2^k - 1)$. The best invariant estimator is $\delta_0(x, y) = x$ uniquely a.e. (μ) for all $n = 0, 1, 2, \cdots$. Let

$$\begin{aligned}(2.4.4)\qquad \delta(x, y) &= 0 \quad \text{if} \quad -y \leqq x \leqq y \\ &= x \quad \text{otherwise}\end{aligned}$$

and $\gamma(x, y) = \delta(x, y) - x$.

The difference in the risks of these two estimators for a given value of $n > 0$ is

$$\begin{aligned}(2.4.5)\quad & R^{(n)}(\theta, \delta_0) - R^{(n)}(\theta, \delta) \\ &= \int_1^\infty ((\beta + k)/y^{1+\beta+k})\, dy\{\int(|x - \theta|^k - |\delta(x, y) - \theta|^k)p_n(x - \theta, y)\, dx\} \\ &= \int_1^\infty ((\beta + k)/y^{1+\beta+k})\, dy\{\int(|z|^k - |z + \gamma(z + \theta, y)|^k)p_n(z, y)\, dz\} \\ &= \int_1^\infty ((\beta + k)/y^{1+\beta+k})\, dy\{y^k\int(|z|^k - |\gamma(z + \theta/y, 1) + z|^k)p_n(z, 1)\, dz\} \\ &= \int_1^\infty ((\beta + k)/y^{1+\beta})\, dy\int_{|z+\theta/y|\leqq 1}(|z|^k - |-\theta/y|^k)p_n(z, 1)\, dz.\end{aligned}$$

For $n = 1, 2, \cdots$ and $|\varphi| < \infty$ define

$$\begin{aligned}(2.4.6)\quad g(n, \varphi) = & \int_{|z+\varphi|\leqq 1}(|z|^k - |\varphi|^k)p_n(z, 1)\, dz \\ & - \int_{|z+\varphi|\leqq 1}(|z|^k - |\varphi|^k)\, dP_0(z, 1).\end{aligned}$$

Note that for each fixed $\varphi \neq 0, \pm 2$, $g(n, \varphi) \to 0$ as $n \to \infty$, $g(n, 0) \to -\frac{1}{2}$ as $n \to \infty$, $g(n, \pm 2) \to -\frac{1}{4}(1 - 2^k))$. Also note that $g(n, \varphi) = 0$ for $|\varphi| > 3$ and all $n = 1, 2, \cdots$; and $|g(n, \varphi)| < 2\cdot 3^k$ for $|\varphi| \leqq 3$ and all $n = 1, 2, \cdots$. Hence

$$\begin{aligned}(2.4.7)\quad & \int_1^\infty ((\beta + k)/y^{1+\beta})\, dy\int_{|(z+\theta)/y|\leqq 1}(|z|^k - |\theta/y|^k)p_n(z, 1)\, dz \\ &= \int_1^\infty ((\beta + k)/y^{1+\beta})\, dy\int_{|(z+\theta)/y|\leqq 1}(|z|^k - |\theta/y|^k)\, dP_0(z, 1) \\ &\qquad + \int_1^\infty ((\beta + k)/y^{1+\beta})g(n, \theta/y)\, dy \\ &= \int_1^\infty ((\beta + k)/y^{1+\beta})\, dy\int_{|(z+\theta)/y|\leqq 1}(|z|^k - |\theta/y|^k)\, dP_0(z, 1) \\ &\qquad + \int_{\max(\frac{1}{3}, 1/\theta)}^\infty ((\beta + k)/\theta^\beta\eta^{1+\beta})g(n, 1/\eta)\, d\eta. \\ &= \begin{cases}\frac{1}{2}((\beta + k)/\beta) + o(1) & \text{if} \quad \theta = 0 \\ \frac{1}{2}((\beta + k)/\beta - |\theta|^k) + o(1) & \text{if} \quad 0 < |\theta| < 2 \\ \int_{|\theta|_2}^\infty ((\beta + k)/y^{1+\beta})\cdot\frac{1}{2}(1 - |\theta/y|^k)\, dy + o(1/|\theta|^\beta) & \\ = (2^{\beta-1}/|\theta|^\beta)((\beta + k)/\beta - 2^k) + o(1/|\theta|^\beta) & \text{if} \quad |\theta| \geqq 2\end{cases} \\ &\qquad\qquad \text{as} \quad n \to \infty.\end{aligned}$$

$\beta < k/(2^k - 1)$ implies $(\beta + k)/\beta - 2^k > 0$. The $o(1)$ terms in (2.4.7) are valid uniformly in θ. Hence (2.4.7) proves that for n sufficiently large $R^{(n)}(\theta, \delta_0) - R^{(n)}(\theta, \delta) > 0$ for all θ. Hence δ_0 is inadmissible. Since $\alpha < \beta$, (2.4.2) is satisfied. This proves the theorem for the case $k > 1$.

If $k = 1$, let

$$\begin{aligned}p_n(x, 1) &= n^2/(2 + 4n) && \text{if} \quad 1 - 1/n \leqq x \leqq 1 + 1/n, \\ (2.4.8)\qquad & && \text{or if} \quad -1/n^2 < x < 1/n^2, \\ & && \text{or if} \quad -1 - 1/n \leqq x \leqq -1 + 1/n \\ &= 0 && \text{otherwise.}\end{aligned}$$

This modification is necessary in order that the best invariant estimate be uniquely determined. The remainder of the computation is very similar to that given previously. It again turns out that δ as defined by (2.4.4) is a better procedure than δ_0. This completes the proof of the theorem.

2.5. *Special results for W convex and $n = 1$.* In the preceding section it was shown that, in the general fixed-sample-size case, some moment condition more than merely $R_0 < \infty$ ($\alpha = 0$ in (2.4.1)) must be assumed in order to prove that the best invariant estimator is admissible. In this section it is shown that there is a special case in which the assumption $R_0 < \infty$ is sufficient (and necessary) to insure that the best invariant estimator is admissible. This special case is the fixed-sample-size problem with sample size $n = 1$ (only the variable X is observed) and with convex loss function, W.

As explained in Section 1.2 there is no loss of generality in assuming that the best invariant estimator is $\delta_0(x) = x$, and we shall do so throughout this section.

Note that if $W(x)$ is convex then $W'(x)$ exists a.e. If $W(x) \sim c|x|^k$ as $x \to \pm\infty$, $k \geqq 1$, is convex then $xW'(x) = O(W(x))$ and $W(x) = O(xW'(x))$ as $x \to \pm\infty$. Therefore in at least this case the following condition is equivalent to $R_0 < \infty$:

$$\int |xW'(x)|p(x)\,dx < \infty. \tag{2.5.1}$$

If for some $a > 0$,

$$\int |W'(x \pm a)|p(x)\,dx < \infty \tag{2.5.2}$$

as is, again, the case if $W(x) \sim c|x|^k$, $k \geqq 1$, and $R_0 < \infty$ then W can be differentiated under the integral sign and (since $\delta_0(x) = x$)

$$\int W'(x)p(x)\,dx = 0 \tag{2.5.3}$$

(see, for example, Farrell (1964)).

For $W(x) = x^2$ the following result coincides with a result of Stein (1959) specialized to $n = 1$. (Note that if $W(x) = x^2$ and $R_0 < \infty$ then (2.5.1) and (2.5.3) are satisfied.)

THEOREM 2.5.1. *Suppose $n = 1$; and W is strictly convex, satisfies* (2.5.1) *and* (2.5.3), *and* $\lim_{x\to\pm\infty} W(x) = \infty$; *and* $\delta_0(x) = x$, *and* $R_0 < \infty$. *Then the best invariant estimator is admissible.*

PROOF. Farrell (1964) has shown that δ_0 is admissible if (and only if) it is almost admissible (with respect to Lebesgue measure on θ). Hence it is sufficient to prove that δ_0 is almost admissible.

Now, assume that δ_0 is not almost admissible. Then there is a procedure, call it δ, such that

$$R(\theta, \delta) \leqq R(\theta, \delta_0) < \infty; \qquad \int [R(\theta, \delta_0) - R(\theta, \delta)]\,d\theta > 0. \tag{2.5.4}$$

Define $\gamma(x) = \delta(x) - x$ and

$$\gamma^*(x) = \int \gamma(x + \xi)p(\xi - \alpha)\,d\xi \tag{2.5.5}$$

where $\alpha = -\int \xi p(\xi)\,d\xi \neq \pm\infty$ (since $R_0 < \infty$). Also define

$$\begin{aligned} \delta^*(x) &= \int \delta(x + \xi)p(\xi - \alpha)\, d\xi \\ &= \int [x + \xi + \gamma(x + \xi)]p(\xi - \alpha)\, d\xi \\ &= x + \gamma^*(x). \end{aligned} \tag{2.5.6}$$

Since W is convex

$$\begin{aligned} R(\theta, \delta^*) &= \int W(\delta^*(x) - \theta)p(x - \theta)\, dx \\ &= \int W(\int (x + \gamma(x + \xi) - \theta)p(\xi - \alpha)\, d\xi)p(x - \theta)\, dx \\ &\leqq \int\int W(x + \gamma(x + \xi) - \theta)p(\xi - \alpha)p(x - \theta)\, d\xi\, dx \\ &= \int\int W(z + \gamma(z + \theta + \xi))p(z)p(\xi - \alpha)\, dz\, d\xi \\ &= \int R(\theta + \xi, \delta)p(\xi - \alpha)\, d\xi \leqq R_0 . \end{aligned} \tag{2.5.7}$$

Thus δ^* is at least as good as δ_0. Furthermore, using (2.5.4),

$$\begin{aligned} \int [R_0 - R(\theta, \delta^*)]\, d\theta &\geqq \int [R_0 - R(\theta + \xi, \delta)]p(\xi - \alpha)\, d\xi\, d\theta \\ &= \int \{\int [R_0 - R(\theta + \xi, \delta)]\, d\theta\} p(\xi - \alpha)\, d\xi \\ &> 0. \end{aligned} \tag{2.5.8}$$

Also,

$$\begin{aligned} W(\gamma^*(x) - \alpha) &= W(\int [\gamma(x + \xi) - \alpha]p(\xi - \alpha)\, d\xi) \\ &= W(\int [\gamma(x + \xi) + \xi - \alpha]p(\xi - \alpha)\, d\xi) \\ &\leqq \int W(\xi - \alpha + \gamma(x + \xi))p(\xi - \alpha)\, d\xi \\ &= \int W(z + \gamma(z + x + \alpha))p(z)\, dz \\ &= R(x + \alpha, \delta) \leqq R_0 . \end{aligned} \tag{2.5.9}$$

Since $\lim_{x\to\pm\infty} W(x) = \infty$, (2.5.9) implies that there is a bound $B < \infty$ such that $|\gamma^*(x)| < B$.

As in (2.3.20),

$$\int_{|z|<L} [W(z) - W(z + \gamma(x))]p(z)\, dz \leqq \int_{|z|<L} \gamma(x)W'(z)p(z)\, dz \tag{2.5.10}$$

uniformly in x as $L \to \infty$. Also as in (2.3.20), using (2.5.3),

$$\begin{aligned} \int_0^\infty d\lambda \sup_{\{\gamma:|\gamma|\leqq B\}} \{\int_{-\lambda}^{\lambda} [W(z) - W(z + \gamma)]p(z)\, dz \\ &\leqq \int_0^\infty d\lambda \sup_{\{\gamma:|\gamma|\leqq B\}} \{\int_{-\lambda}^{\lambda} \gamma W'(z)p(z)\, dz \\ &= B|\int_0^\infty d\lambda \int_{-\lambda}^{\lambda} W'(z)p(z)\, dz| \\ &= +B|\int_0^\infty d\lambda \int_{|z|>\lambda} W'(z)p(z)\, dz| \\ &\leqq B\int_0^\infty d\lambda \int_{|z|>\lambda} |W'(z)|p(z)\, dz \\ &= B\int_0^\infty |zW'(z)|p(z)\, dz < \infty . \end{aligned} \tag{2.5.11}$$

(2.5.11), of course, verifies that (2.1.2) is satisfied when the supremum is taken only over those γ such that $|\gamma(x)| < B$. The remainder of the proof of Theorem 2.5.1 is analogous step-by-step with the proof of Theorem 2.1.1. In view of the preceding, the proof may be begun as in Theorem 2.1.1 by assuming $\delta_0(x) = x$ and using a δ such that $R(\theta, \delta) \leqq R_0$, *and*, in addition such that $|\delta(x) - x| \leqq B$. Having done this, it follows that (2.5.11) may be used in the proof in place of (2.1.2). Also (2.5.10) and (2.5.1) may be used in place of (2.1.3). For example, in the analog of (2.1.8) to show that

$$\{\int_{-\infty}^{-L/2} dz \int_{z-L}^{z+L} dx + \int_{L/2}^{\infty} dz \int_{z-L}^{z+L} dx\}\bar{\omega}(z, x, y)p(z) \to 0$$

use (2.5.10) to show that this is no larger than

$$\{\int_{-\infty}^{-L/2} dz \int_{z-L}^{z+L} dx + \int_{L/2}^{\infty} dz \int_{z-L}^{z+L} dx\}BW'(z)p(z)$$

and then use (2.5.1) to complete the argument. As another example, the step after the second $\leqq$ in the analog of (2.1.10) should read $2 \int_{-L/2}^{0} BW'(z)p(z)\, dz + \int_0^{L/2} dx \int_{-x}^{x} \omega(z, \gamma_0, 1, \gamma(x - L), 1, y)p(z)\, dz$. Use (2.5.1) to show the first term is bounded, and (2.5.11) to show the third term is bounded. Some similar changes must also be made in the analog of steps (2.1.12), (2.1.21), (2.1.22), and (2.1.23).

It should be noted that strict convexity of W rather than just convexity has been used only to pass from the conclusion of almost admissibility to admissibility and to establish uniqueness of the best invariant estimator. Hence

COROLLARY 2.5.1. *Suppose $n = 1$; and W is convex, satisfies* (2.5.1) *and* (2.5.3), *and* $\lim_{x\pm\infty} W(x) = \infty$; *and suppose $\delta_0(x) = x$ is the unique best invariant estimator, and $R_0 < \infty$. Then δ_0 is almost admissible.*

PROOF. The proof is the same as the proof of the theorem from Equation (2.5.4) to the end.

2.6. *Summary of Section* 2. In Section 2.1, we proved a theorem stating that, subject to certain assumptions, the best invariant estimator of a location parameter is admissible. In Sections 2.2 and 2.3 we examined in detail the assumptions of Theorem 2.1.1 and showed that in the fixed-sample-size case these assumptions are very weak; but, on the other hand, they are not trivially satisfied. Lemma 2.3.4 indicates that this is probably also the case in the general sequential problem, although easily interpretable conditions are given there only for the special sequential problem (see Lemma 2.3.5).

Let us examine here how the results of Section 2 relate to similar theorems given by other authors.

In the fixed-sample-size case there are several previous results. Stein (1959) considered the case where $W(t) = t^2$. The condition he gave for the best invariant estimator (which, without loss of generality, is assumed to be $\delta_0(x) = x$) to be admissible is

$$\int \mu(dy)[\int x^2 p(x, y)\, dx]^{\frac{3}{2}} < \infty. \tag{2.6.1}$$

Using Theorem 2.1.1 and Lemmas 2.2.1 and (2.3.3) we have the result that for

$W(t) = t^2$ and $n > 1$ the best invariant estimator ($\delta_0(x) = x$) is admissible if

$$\int \mu(dy) \int x^3 p(x, y)\, dx < \infty; \tag{2.6.2}$$

and, using Theorem 2.5.1, if $n = 1$ the best invariant estimator is admissible so long as

$$\int x^2 p(x)\, dx < \infty. \tag{2.6.3}$$

It is easy to see (using the Cauchy-Schwartz inequality) that when $W(t) = t^2$, Stein's result contains ours when $n > 1$, and the two theorems are identical for $n = 1$.

R. Farrell (1964) proved a theorem in which he extended Stein's results to other types of loss functions (and also proved admissibility of many estimators other than the best invariant estimator). He considered only loss functions satisfying the monotonicity condition (1.2.2), and which were either bounded, uniformly continuous, convex, or part bounded and part convex. His results are stronger than ours only if W is uniformly continuous, or part bounded-part convex and the support of μ is one point ($n = 1$). For example, in the uniformly continuous case he requires

$$\int |x|^p W(x) p(x)\, dx < \infty, \tag{2.6.4}$$

while we require

$$\int |x|^p \sup_{\tau \leqq 2x} W(\tau) p(x)\, dx < \infty. \tag{2.6.5}$$

In the part bounded-part convex case when $n = 1$ his theorem is also slightly stronger. If the support of μ is more than one point our results are always stronger than his (since in that case his theorem assumes that $p(\cdot, \cdot)$ has compact support).

Karlin (1958) and Fox and Rubin (1964), among others, have also treated the one-dimensional fixed-sample-size problem. Their results are weaker than ours if X, Y has a density. (Fox and Rubin also treat the non-absolutely continuous case—see the discussion in Section 1.2.)

There seem to be no previous (non-trivial, correct) theorems concerning admissibility in the general sequential case. Part of the reason for this may be found by comparing our method of proof with the method used by Blyth (1951), Stein (1959), Farrell (1964), and others. Their proofs seem to require that one know the best invariant estimator and that one be able to determine Bayes estimators for a sequence of *a priori* distributions. At least enough about the various estimators must be determined so that it can be computed that the difference between their Bayes risks tends to 0 sufficiently fast. (This is often done by proving that the Bayes estimator tends sufficiently fast pointwise to the best invariant estimator.) For the method of proof used in this paper no Bayes procedures are computed, and it is not necessary to know the best invariant estimator or its risk. (In fact, these are usually not known in the general sequential problem.) Of course, Blyth's method is much more general than the method used here. Our method cannot even be used without a significant modification to prove admissibility of the best invariant estimator in the case $m = 2$ (see, Brown (1965)).

It should be noted here that our method can also be used to prove admissibility of suitable generalized Bayes estimators of a location parameter (other than the best invariant estimator) although we have not pursued that topic in this paper.

3. Results for $m < 3$.

3.1. *Inadmissibility of invariant estimates for convex loss functions and dimension* $m \geqq 3$. In the previous section it was shown that the best invariant estimator of a location parameter is usually admissible in dimension $m = 1$. An interesting phenomenon is that the situation is quite different in three or more dimensions. The case $m = 2$ will be treated in a separate paper (Brown (1965)).

Only the fixed-sample-size problem (as defined in Section 1.1) will be considered in Sections 3.1 to 3.3. The sequential case is discussed in Section 3.4.

Stein (1956), (1960) considered the case where the loss function is $W(t) = \|t\|^2$. He stated that if for each $y_1, \cdots, y_{n-1}$ among a set of values of $Y_1, \cdots, Y_{n-1}$ having positive μ measure the co-ordinates $X_1, X_2, \cdots, X_m$ have non-singular variance-covariance matrix, then the best invariant estimator of θ is inadmissible. (Certain conditions concerning the finitemess of moments of X are also assumed.)

In this section and in Section 3.3, we will show that inadmissibility of the best inv est when $m \geqq 3$ is a quite general phenomenon, not limited to the case discussed above.

In this section the case where W is convex will be considered. In Section 3.3 this result will be extended to other loss functions. The two types of loss functions will be considered separately because the regularity conditions in the convex case are much simpler, and the assumptions are more nearly necessary (as well as sufficient). Thus we hope that this theorem will give more of an intuition into the situation than would Theorem 3.3.1 or 3.3.2.

Although there are a few techniques which have been used successfully to prove admissibility of an estimate, there is essentially only one technique which has been used to prove that an estimator is inadmissible; namely, explicitly exhibit a better estimator. This is what I will do here. The estimator in all cases will be a generalization of the estimator $\delta(x) = (1 - b/(a + \|x\|^2))x$ which was used by Stein.

In what follows wherever possible capital letters will be used to denote matrices or vectors and lower case letters will denote scalars. (Exceptions to this will be the symbols W, R, and δ which we have previously defined.) No distinction will be made in the notation between random variables and their values.

In what follows vectors are understood to be column vectors except where otherwise stated. If M is a matrix (or vector) we denote its transpose by M^{τ}.

If $W(T, Y)$ is any scalar function of the vector $T = (t_1, t_2, \cdots, t_m)^{\tau}$ and of

Y (the argument Y will often be omitted), define the *row* vector

$$W'(T, Y) = (w_1'(T, Y), w_2'(T, Y), \cdots, w_m'(T, Y)) \tag{3.1.1}$$

where

$$\begin{aligned} w_i'(T, Y) &= (\partial/\partial t_i)W(T, Y) \qquad \text{if} \quad (\partial/\partial t_i)W(T, Y) \text{ exists} \\ &= 0 \qquad\qquad\qquad\quad \text{otherwise.} \end{aligned}$$

In the applications $(\partial/\partial t_i)W(T, Y)$ exists almost everywhere, and it really does not matter how $W'(T, Y)$ is defined when $(\partial/\partial t_i)W(T, Y)$ does not exist. The value 0 has been chosen for convenience.

Let $g(T)$ be a scalar function of t such that g is of bounded variation in t_1 for each fixed $(t_2, t_3, \cdots, t_m)$. For each $(t_2, \cdots, t_m)$ let $d_1g(T)$ denote the Lebesgue-Stieltjes measure generated by g considered as a function of the real variable t_1. More precisely, for each $(t_2, \cdots, t_m)$, $d_1g(T)$ is the unique measure on the Lebesgue measureable sets of the real line such that $g(T) = \int_{-\infty}^{t_1} d_1g(T)$ almost everywhere (dt_1). $d_kg(T)$, $k = 2, \cdots, m$, is similarly defined. Note that if $g(T)$ is absolutely continuous as a function of t_1 then $d_kg(T) = [(\partial/\partial t_k)g(T)]\, dt_k$.

Let T_k^* denote $(t_1, \cdots, t_{k-1}, t_{k+1}, \cdots, t_m)^\tau$.

We shall use the following simple result concerning change of variables and order of integration in the proof of Theorem 3.1.1.

LEMMA 3.1.1. *Let $f(T) (= f(t, T_1^*)$ be a probability density on E_m. Let $g(t, T_1^*)$ be a non-decreasing function of t for each T_1^*. Let*

$$\int dT_1^* \int f(t + z, T_1^*)\, d_1g(t, T_1^*) = r(z) < \infty \text{ a.e.}(dz) \qquad \text{for} \quad z \,\varepsilon\, [0, a],\ a > 0. \tag{3.1.2}$$

Then $r(z)$ is Lebesgue measurable, non-negative, and, for $a > 0$,

$$0 \leqq \int dT_1^* \int [g(t, T_1^*) - g(t - a, T_1^*)]f(t, T_1^*)\, dt = \int_0^a r(z)\, dz \leqq \infty. \tag{3.1.3}$$

PROOF. Using the definition of $d_1g(T)$ and Fubini's theorem,

$$\begin{aligned} \int dT_1^* \int [g(t, T_1^*) &- g(t - a, T_1^*)]f(t, T_1^*)\, dt \\ &= \int dT_1^* \int f(t, T_1^*)\, dt \int_{t-a}^{t} d_1g(\alpha, T_1^*) \\ &= \int dT_1^* \int_{-\infty}^{\infty} d_1g(\alpha, T_1^*) \int_{\alpha}^{a+\alpha} f(t, T_1^*)\, dt \\ &= \int dT_1^* \int_{-\infty}^{\infty} d_1g(\alpha, T_1^*) \int_0^a f(z + \alpha, T_1^*)\, dz \\ &= \int_0^a dz \int dT_1^* \int_{-\infty}^{\infty} f(z + \alpha, T_1^*)\, d_1g(\alpha, T_1^*) \\ &= \int_0^a r(z)\, dz. \end{aligned} \tag{3.1.4}$$

This proves the lemma.

In Section 3.2 the implications of the assumptions in Theorem 3.1.1 will be discussed. (In particular, see Lemmas 3.2.1 and 3.2.2.) Some important special cases of Theorem 3.1.1 will also be considered. In Section 3.3 a result similar to Theorem 3.1, but without the requirement that W be convex, will be proved.

Let U_i be the unit vector in the ith co-ordinate direction. Let I be the $(m \times m)$ identity matrix.

For any measurable function g, let $E_{\theta;Y}(g) = \int g(X, Y)p(X - \theta, Y)\,dX$ where $Y = Y_1, Y_2, \cdots, Y_{n-1}$. By convention, if $n = 1$, $p(X - \theta, Y) = p(X - \theta)$ for all Y.

We will assume that the best inv est is $\epsilon_0(X, Y) = X$. This is no loss of generality (see Section 1.2). We also assume throughout that $R_0 < \infty$.

THEOREM 3.1.1. *Consider the fixed-sample-size problem with $m \geqq 3$, $W(T, Y)$ convex in T for each Y, $\epsilon_0(X, Y) = X$, and $R_0 < \infty$. Suppose there is a set S of values of Y such that $\mu(S) > 0$, satisfying the following conditions:*

(i) *For each $Y \varepsilon S$ there is a $\gamma > 0$ and c_1 such that $|\beta| < \gamma$ implies*

$$E_{0Y}\{\|X\|^4\|W'(X + \beta U_i, Y)\|\} < c_1 \tag{3.1.5}$$

for $1 \leqq i \leqq m$.

(ii) *For each $Y \varepsilon S$ there is a $c_2 > 0$ such that*

$$\int\int p(T + \beta U_i, Y)(d_i w_i'(T, Y))\,dT_i^* < c_2 \tag{3.1.6}$$

for $0 < \beta < \gamma$ and $1 \leqq i \leqq m$.

(iii) *For each $Y \varepsilon S$,*

$$E_{0Y}(XW'(X, Y)) \tag{3.1.7}$$

is a non-singular matrix $(m \times m)$.

Define the estimator ϵ by

$$\begin{aligned} \epsilon(X, Y) &= [I - A(Y)/(a(Y) + \|X\|^2)]X && \text{if } Y \varepsilon S \\ &= X && \text{if } Y \not\varepsilon S \end{aligned} \tag{3.1.8}$$

where $A(Y) = (1/b(Y))E_{0Y}(XW'(X))^{-1}$. Then there exist functions $a(Y) > 0$, $b(Y) > 0$ such that $R(\theta, \epsilon) < R(\theta, \epsilon_0)$ for all θ. (i.e. *ϵ is better than ϵ_0; and, hence, ϵ_0 is inadmissible.*)

PROOF. Pick any value $\tilde{Y} \varepsilon S$. To shorten the notation let $E = E_{0\tilde{Y}}$, and $W(T) = W(T, \tilde{Y}$ (and $A = A(\tilde{Y})$. Let J_i be the $m \times m$ matrix with entries $(J_i)_{jk} = \delta_{ijk}(=1$ only when $i = j = k)$.

Since W is convex

$$\begin{aligned} d &= E_{\theta,\tilde{Y}}(W(\epsilon_0(X, \tilde{Y}) - \theta)) - E_{\theta,\tilde{Y}}(W(\epsilon(X, \tilde{Y}) - \theta)) \\ &= E_{\theta,\tilde{Y}}\{W(X - \theta) - W([I - A/(a + \|X\|^2)]X - \theta)\} \\ &= E\{W(X) - W(X - A\cdot(X + \theta)/(a + \|X + \theta\|^2))\} \\ &\geqq E\{W(X) - (1/m) \\ &\qquad \cdot \textstyle\sum_{i=1}^m W(X - mJ_iA\cdot(X + \theta)/(a + \|X + \theta\|^2))\} \\ &\geqq E\{\textstyle\sum_{i=1}^m W'(X - mJ_iA\cdot(X + \theta)/(a + \|X + \theta\|^2)) \\ &\qquad \cdot J_iA\cdot(X + \theta)/(a + \|X + \theta\|^2)\} \\ &= E\{\textstyle\sum_{i=1}^m d_i\} \end{aligned} \tag{3.1.9}$$

where d and d_i are defined by the expressions given above (in (3.1.9)). Note that in a terminology similar to that of Section 2.2, $d = R(\epsilon_0 \mid Y) - R(\theta, \epsilon \mid Y)$. The second inequality above follows from the fact that the directional derivative along the line connecting X to $X - mJ_iA(X + \theta)/(a + \|X + \theta\|^2)$ is nondecreasing.

For convenience, let $J_iA = B_i$.

Notice the equality

$$\begin{aligned}
&1/(a + \|X + \theta\|^2) \\
&= 1/[(a + \textstyle\sum \theta_i^2) + (\sum 2x_i\theta_i + \sum x_i^2)] \\
&= 1/(a + \textstyle\sum \theta_i^2) - (\sum 2x_i\theta_i + \sum x_i^2)\,(a + \sum \theta_i^2)^{-2} \\
&\quad + (\textstyle\sum 2x_i\theta_i + \sum x_i^2)^2/(a + \sum \theta_i^2)^2(a + \sum \theta_i^2 + \sum 2x_i\theta_i \\
&\quad + \textstyle\sum x_i^2).
\end{aligned} \tag{3.1.10}$$

Using this,

$$\begin{aligned}
Ed_i &= E\{W'(X - mB_i(X + \theta)/(a + \|X + \theta\|^2)) \\
&\qquad B_i(X + \theta)/(a + \|X + \theta\|^2)\} \\
&= E\{W'(X - mB_i(X + \theta)/(a + \|X + \theta\|^2)) \\
&\qquad B_i(X + \theta)/(a + \|\theta\|^2) \\
&\qquad - (\textstyle\sum (2x_i\theta_i + x_i^2)) \\
&\qquad W'(X - mB_i(X + \theta)/(a + \|X + \theta\|^2)) \\
&\qquad B_i(X + \theta)/(a + \|\theta\|^2)^2 + (\textstyle\sum (2x_i\theta_i + x_i^2))^2 \\
&\qquad W'(X - mB_i(X + \theta)/(a + \|X + \theta\|^2)) \\
&\qquad B_i(X + \theta)/(a + \|\theta\|^2)^2(a + \|X + \theta\|^2)\}.
\end{aligned} \tag{3.1.11}$$

Notice that

$$\begin{aligned}
\theta_j/(a + \|\theta\|^2) &= O(a^{-\frac{1}{2}}), \\
(X + \theta)/(a + \|X + \theta\|^2) &= O(a^{-\frac{1}{2}}), \\
\theta_j\theta_k/(a + \|\theta\|^2) &= O(1),
\end{aligned} \tag{3.1.12}$$

where here as throughout the proof the o and O are uniform in θ as $a \to \infty$.

Using (3.1.5) and (3.1.12), if b is chosen sufficiently large the expectation of the third term on the right in (3.1.11) and all of the second term except that part appearing below in (3.1.13) is $o(1/(a + \|\theta\|^2))$ uniformly in θ as $a \to \infty$. (3.1.11) becomes

$$\begin{aligned}
Ed_i &= E\{W'(X - mB_i(X + \theta)/(a + \|X + \theta\|^2)) \\
&\qquad B_i(X + \theta)/(a + \|\theta\|^2) - (\textstyle\sum 2x_i\theta_i) \\
&\qquad W'(X - mB_i(X + \theta)/(a + \|X + \theta\|^2))B_i\theta/(a + \|\theta\|^2)^2\} \\
&\qquad + o(1/(a + \|\theta\|^2)).
\end{aligned} \tag{3.1.13}$$

Since $\int \|W'(X + \beta U_i)\| p(X, \tilde{Y})\, dX < \infty$,

$$(3.1.14)\quad (\partial/\partial\beta) \int W(X + \beta U_i) p(X, \tilde{Y})\, dX = \int (\partial/\partial\beta) W(X + \beta U_i) p(X, \tilde{Y})\, dX$$

exists for all $|\beta| < \gamma$. Then, since $\epsilon_0(X, Y) = X$,

$$(3.1.15)\quad \int W'(X) p(X, \tilde{Y})\, dX = 0.$$

Let $A^* = E(XW'(X))^{-1} = (a_{ij}^*)$ in order to define $\alpha = \sup |a_{ij}^*|$. Let $\sigma = \sup (\|\theta\|, a^{\frac{1}{2}})$. Note $A = A^*/b$. A simple maximization yields

$$(mB_i(X + \theta)/(a + \|X + \theta\|^2))_i \leqq m\alpha/2ba^{\frac{1}{2}}$$

for all X. Also, $(mB_i(X + \theta)/(a + \|X + \theta\|^2))_j = 0$ if $j \neq i$. If $\|X\| < \sigma/2$,

$$(mB_i(X + \theta)/(a + \|X + \theta\|^2))_i < \min (m\alpha/2ba^{\frac{1}{2}}, 2m\alpha/b\|\theta\|) \leqq 2m\alpha/b\sigma.$$

If we now choose a so large that $2m\alpha/b\sigma < \gamma$ then, since $w_i'(T)$ is non-decreasing in t_i,

$$\begin{aligned} &E\{W'(X - mB_i(X + \theta)/(a + \|X + \theta\|^2))\} \\ &\geqq \int_{\|X\| < \sigma/2} w_i'(X - (2m\alpha/\sigma b) U_i) p(X, \tilde{Y})\, dX \\ (3.1.16)\qquad &\quad + \int_{\|X\| \geqq \sigma/2} w_i'(X - (m\alpha/2ba^{\frac{1}{2}}) U_i) p(X, \tilde{Y})\, dX \\ &\geqq \int [w_i'(X - (2m\alpha/b\sigma) U_i) - w_i'(X)] p(X, \tilde{Y})\, dX \\ &\quad - \int_{\|X\| \geqq \sigma/2} w_i'(X - (2m\alpha/b\sigma) U_i) p(X, \tilde{Y})\, dX \\ &\quad + \int_{\|X\| \geqq \sigma/2} w_i'(X - (m\alpha/2ba^{\frac{1}{2}}) U_i) p(X, \tilde{Y})\, dX \end{aligned}$$

using (3.1.15). Since $E\{\|X\| \|W'(X - \beta U_i)\|\} < \infty$ for $\beta < \gamma$, the last two terms on the right of (3.1.16) tend to 0 as $a \to \infty$, uniformly in θ. Using the above and (3.1.6) and Lemma 3.1.1,

$$\begin{aligned} &E\{W'(X - mB_i(X + \theta)/(a + \|X + \theta\|^2)) B_i\theta\} \\ (3.1.17)\qquad &\geqq -c_2(2m\alpha/b\sigma) \cdot B_i\theta + o(1) \\ &\geqq -c_2(2m\alpha/b\sigma) \cdot \alpha\sigma/b + o(1) = -2c_2 m\alpha^2/b^2 + o(1) \\ &= -k/b^2 + o(1) \qquad \text{where} \qquad k = 2c_2 m\alpha^2. \end{aligned}$$

From (3.1.5) and (3.1.12) (or from (3.1.16))

$$(3.1.18)\quad E\{w_i'(X - mB_i(X + \theta)/(a + \|X + \theta\|^2)) x_j\} \to E\{w_i'(X) x_j\}$$

uniformly in θ, i, j as $a \to \infty$.

Applying (3.1.16) and (3.1.17) to (3.1.13) yields

$$\begin{aligned} (3.1.19)\quad Ed_i \geqq\ &E\{(W'(X)B_iX - k/b^2)/(a + \|\theta\|^2) \\ &- W'(X)B_i\theta(\textstyle\sum 2x_i\theta_i)/(a + \|\theta\|^2)^2\} + o(1/(a + \|\theta\|^2)). \end{aligned}$$

Now note that if $C = (c_{ij})$ is any $(m \times m)$ matrix

$$E\{W'(X)ACX\} = E\,(\text{trace}\,(XW'(X)AC)) = 1/b\ \text{trace}\ C. \tag{3.1.20}$$

Hence,

$$\begin{aligned} d = E\{\textstyle\sum d_i\} &\geqq E\{(W'(X)AX - mk/b^2)/(a + \|\theta\|^2) \\ &\quad - W'(X)A\theta(\textstyle\sum 2x_i\theta_i)/(a + \|\theta\|^2)^2\} + o(1/(a + \|\theta\|^2)) \\ &\geqq (1/(a + \|\theta\|^2))^2 E\{a(W'(X)AX - mk/b^2) \\ &\quad + W'(X)A(\|\theta\|^2 I - 2\theta\cdot\theta^T)X - mk\|\theta\|^2/b^2\} \\ &\quad + o(1/(a + \|\theta\|^2)) \\ &= (1/(a + \|\theta\|^2))^2[a(m/b - mk/b^2) \\ &\quad + 1/b\ \text{trace}\ (\|\theta\|^2 I - 2\theta\cdot\theta^T) - mk\|\theta\|^2/b^2] \\ &\quad + o(1/(a + \|\theta\|^2)). \end{aligned} \tag{3.1.21}$$

If b and then a are chosen sufficiently large,

$$\begin{aligned} d &\geqq (1/(a + \|\theta\|^2))^2[a(m/b - mk/b^2) \\ &\quad + (1/b)(m - 2)\|\theta\|^2 - mk\|\theta\|^2/b^2] + o(1/(a + \|\theta\|^2)) \\ &\geqq \tfrac{1}{2}(1/b)/(a + \|\theta\|^2) > 0. \end{aligned} \tag{3.1.22}$$

This proves that for each $Y \varepsilon S$ there are constants $a(Y) > 0$, $b(Y) > 0$ such that $d(Y) > 0$. Now,

$$\begin{aligned} &R(\theta, \epsilon_0) - R(\theta, \epsilon) \\ &= \textstyle\int \mu(dY) \int [W(\epsilon_0(X, Y) - \theta, Y) \\ &\quad - W(\epsilon(X, Y) - \theta,\ Y)]p(X - \theta, Y)\,dX \\ &= \textstyle\int_S d(Y)\mu(dY) > 0. \end{aligned} \tag{3.1.23}$$

This completes the proof of the theorem.

Note that the dimension, n, plays its decisive role only in the last step, (3.1.22), of the proof!

It is important to notice that the choice of a and b in Theorem 3.1.1 depends on w and f only through certain constants which appear in Theorem 3.1.1, and in no other way. A weak result of this type is specifically stated below.

COROLLARY 3.1.1. *Suppose the hypotheses of Theorem* 3.1.1 *are satisfied. Let*

$$\sigma_k(Y) = \sup_{\{|\beta|<\gamma, 1\leqq i\leqq m\}} E_{0Y}\{\|X\|^k W'(X - \beta U_i)\}.$$

Define γ, c_2, α *as in* (3.1.5), (3.1.6), *and* (3.1.16) *respectively. Then there is a* $b(Y)$ *depending only on* c_2, m, *and* α; *and an* $a(Y)$ *depending only on* γ, α, b, *and* σ_k, $k = 1, 2, 3, 4$, *such that the estimator* δ *given by* (3.1.8) *for these values of* a *and* b *satisfies* $R(\theta, \delta) < R(\theta, \delta_0)$ *for all* θ.

PROOF. From (3.1.22) it can easily be seen that it is sufficient to choose b

large enough so that $(m - 2)/b - mk/b^2 > 3/4b$. Thus the choice of b only depends on m and k; or m, c_2, and α. After b has been chosen, a need be chosen only large enough so that the term $o(1/(a + \|\theta\|^2))$ in (3.1.22) is less than $(1/4b)(1/(a + \|\theta\|^2))$. This $o(\quad)$ term is made up of contributions from the error terms in (3.1.13) and (3.1.17). It can be checked that the size of these error terms depend on at most the constants γ, α/b, and σ_k, $k = 1, 2, 3, 4$. (It is not clear that the error terms really depend on all the σ_k.) This completes the proof of the corollary.

It would be of interest to determine exactly how small a and b can be chosen. We have not done this.

Although we have not usually considered in this paper random variables which do not possess densities, Theorem 3.1.1 generalizes so easily to some such cases that we will state the following corollary:

COROLLARY 3.1.2. *Suppose the distribution of X given Y is $P(X - \theta, Y)$, θ unknown; X, $\theta \varepsilon E_m$, $m \geqq 3$. Suppose $\epsilon_0(X, Y) = X$ is a best invariant estimator. Let $W(T, Y)$ be convex in T for each Y. Suppose* (3.1.5) *and* (3.1.7) *are satisfied. Replace* (3.1.6) *by the assumption that $w_i'(T, Y)$ is absolutely continuous as a function of t_i for all i and T_i^*, and, for each $Y \varepsilon S$,*

$$\int (\partial/\partial t_i) w_i'(T, Y)\, d_i P(T + \beta U_i, Y) < c_2, \qquad 1 \leqq i \leqq m. \tag{3.1.24}$$

Then there is an a and b such that δ defined by (3.1.8) *satisfies $R(\theta, \delta) < R(\theta, \delta_0)$ for all θ.*

PROOF. For the most part the proof of the corollary is the same as the proof of the theorem, with the sole exception that $p(X - \beta U_i, \tilde{Y})\, dx_i$ must be replaced wherever it occurs by $d_i P(X - \beta U_i, \tilde{Y})$. In making the transition from (3.1.16) to (3.1.17), however, one must do a little more. It can easily be verified that there is a version of Lemma 3.1.1 which is appropriate for making this transition if the condition (3.1.6) is replaced by (3.1.24). When this is done, the proof of the corollary is complete.

3.2. *Examples and lemmas concerning the assumptions of Theorem* 3.1.1. There are four major assumptions in Theorem 3.1.1 whose meaning and/or significance should be discussed. These are the assumptions that W is convex, and the assumptions (3.1.5), (3.1.6) and (3.1.7). Section 3.3 will be primarily concerned with the first of these. Some of the results of this section can also be generalized to non-convex W.

Of the remaining three assumptions, (3.1.5), that is

$$\int \|X\|^4 \|W'(X - \beta U_i)\| p(X, Y)\, dX \tag{3.2.1}$$

is the simplest and probably the least significant. By being more careful in the transition from (3.1.11) to (3.1.13) the $\|X\|^4$ can be replaced by an $\|X\|^3$. It seems reasonable to conjecture that δ_0 remains inadmissible even if (3.1.5) is not satisfied, and perhaps even that an estimator of the form of δ, (3.1.8), remains better than δ_0. The heuristic argument that δ_0 is less likely to be admissible as p becomes more dispersed (as in Section 2.4) would seem to support this conjecture.

Some assumptions at least resembling (3.1.6) and (3.1.7) are necessary. We will show this by examples later in this section. First, however, it will be shown that these assumptions are satisfied in many statistical situations. In most of this section we are interested only in the situation for a fixed $Y \varepsilon S$; we therefore write only $W(T)$, not $W(T, Y)$.

If each w_i' is a differentiable function of t_i then $d_i w_i'(T) = (\partial/\partial t_i) w_i'(T)\, dt_i$. In general w_i' will be differentiable with respect to t_i only almost everywhere and the measure $d_i w_i'(T)$ will have a singular part, say $\sigma_i(dt_i, T_i^*)$ (where $T_i^* = (t_1, \cdots, t_{i-1}, t_{i+1}, \cdots, t_n)$) concentrated on a set of measure 0. The following lemma gives a sufficient (but not necessary) condition for assumption (3.1.6) to be satisfied.

LEMMA 3.2.1. *Let $d_i w_i'(T) = (\partial/\partial t_i) w_i'(T) dt_i + \sigma_i(dt_i, T_i^*)$ where σ_i is a singular measure. Suppose there exists c_3 such that*

$$\int p(T + \beta U_i, Y)(\partial/\partial t_i) w_i'(T)\, dt_i\, dT^* < c_3 \tag{3.2.2}$$

for all $0 < \beta < \gamma$, $1 \leqq i \leqq m$. Suppose either $\sigma_i(E_m) = 0$ or $p(\cdot, Y)$ is bounded and there is a c_4 such that $\int \sigma_i(dt_i, T^*) \leqq c_4$ for each T_i^*, $1 \leqq i \leqq m$. Then assumption (3.1.6) *is satisfied,* i.e. *there is a c_2 such that*

$$\int p(T + \beta U_i, Y)\, d_i w_i'(T)\, dT^* < c_2 \tag{3.2.3}$$

for $0 < \beta < \gamma$, $1 \leqq i \leqq m$.

PROOF. Substitute $(\partial/\partial t_i) w_i(T)\, dt_i + \sigma_i(dt_i, T^*)$ for $d_i w_i'(T)$ in (3.2.3). Use the linearity of the integral, and choose $c_2 = c_3 + c_4 \sup_T p(T, Y)$ if $\sigma_i(E_n) > 0$, or $c_2 = c_3$ if $\sigma_i(E_n) = 0$. Then (3.2.3) will be satisfied.

The condition (3.1.7) (that is, $E_{0Y}(XW'(X))$ non-singular) appears to be much deeper than (3.1.5) or (3.1.6). It is connected with the dimension of the random variable X, but it also is dependent on the relation between X and the loss function W. For the statistical situations which are described in the following lemma, we have succeeded in proving that (3.1.7) is satisfied. In these situations the conditions on W are such that (3.1.7) is satisfied as long as X given Y is a true (non-degenerate) m-dimensional random variable. In general, it is not sufficient to assume merely that X is non-degenerate; see Examples 3.2.3 and 3.2.4.

In the lemma below non-degeneracy is automatically implied by the assumption that X given Y has a probability density with respect to Lebesgue measure on E_m. The results of the lemma can be immediately generalized to the case where X does not possess a probability density so long as the distribution of X given Y is non-degenerate (i.e. is not supported on any set of Lebesgue measure 0 in E_m).

The expectation operator E_{0Y} is defined above Theorem 3.1.1.

LEMMA 3.2.2. *Suppose $\epsilon_0(X, Y) = X$, $R_0 < \infty$, and*

$$\int \|X\| \|W'(X - \beta U_i)\| p(X, Y)\, dX < c < \infty \tag{3.2.4}$$

for all $|\beta| < \gamma$ ($\gamma > 0$), $1 \leqq i \leqq m$. Suppose either

(a) $W(T) = g((T - T_0)'A(T - T_0))$ *where* T_0 *is a fixed vector;* A *is symmetric positive definite; and* $g(\tau)$ *is a non-negative function, differentiable for almost all* $\tau > 0$, $g'(\tau) \geqq 0$, *and for all* $Y \varepsilon S$ (*with* $\mu(S) > 0$).

$$(3.2.5) \qquad \int_V p(X, Y)\, dX > 0$$

where $V = \{T: g'((T - T_0)'A(T - T_0)) > 0\}$;
(*In particular, if* $g'(\tau) > 0$ *for almost all* τ, (3.2.5), *is satisfied.*)
or

(b) $W(T) = \sum_{i=1}^m h_i(t_i)$ *where each* h_i *is a convex function whose derivative* $h_i'(\tau)$ (*which exists* a.e.) *is zero for at most one value of* τ *and*

$$p(T, Y) = \prod_{i=1}^m p_i(t_i, Y).$$

PROOF. As before, let $E(\cdot) = E_{0Y}(\cdot)$ for a fixed $Y \varepsilon S$. The assumption (3.2.4) guarantees that $E(XW'(X))$ exists, and also (as in (3.1.15)) that $E(W'(X)) = 0$.

For case (a), note first that

$$E(XW'(X)) = E((X - T_0)W'(X)).$$

Also, note that

$$W'(X) = 2(X - T_0)'A \cdot g'((X - T_0)'A(X - T_0)).$$

Hence,

$$(3.2.6) \quad E(XW'(X)) = 2E\{(X - T_0)(X - T_0)'g'((X - T_0)'A(X - T_0))\}A.$$

Since the matrix in braces on the right of (3.2.6) is positive semi-definite, it follows that its expectation must also be positive semi-definite.

According to (3.2.5) the vectors of V span the space E_m. Thus the expectation on the right of (3.2.6) must be strictly positive definite. Hence its product with A is non-singular; which is the desired result.

For case (b), since $E(W'(X)) = 0$, $E(h_j'(x_j)) = 0$. Thus $E(x_i h_j'(x_j)) = 0$ if $i \neq j$.

Define $\hat{t}_i$ by $h_j'(\hat{t}_i) = 0$. By assumption, $\hat{t}_i$ is uniquely determined. Now,

$$(3.2.7) \qquad E(x_i h_i'(x_i)) = E((x_i - \hat{t}_i)h_i'(x_i)).$$

The integrand on the right of (3.2.7) is positive almost everywhere. Hence, $E(x_i h_i'(x_i)) > 0$, $i = 1, 2, \cdots, m$. Therefore $E(XW'(X))$ is a diagonal matrix with positive terms on the diagonal, hence non-singular.

This completes the proof of Lemma 3.2.2.

We shall now give several examples. The first few of these will exhibit commonly occurring situations in which Theorem 3.1.1 is valid. The remaining examples will give some situations in which the conclusion of Theorem 3.1.1 is false, either because (3.1.7) is not satisfied or because (3.1.6) is not satisfied, EXAMPLE 3.2.1. If W is "squared loss," i.e. $W(T) = \|T\|^2 = \sum t_i^2$, then the appropriate form of Stein's estimator, (3.1.8), is better than the best invariant estimator if the moment condition (3.1.5) is satisfied. This result was, of course, considered by Stein (1956), (1960). Note also that it is not necessary for the distribution of X given Y to possess a density; it is only necessary that it be non-degenerate. (A minor modification of Lemma 3.2.2 will prove this last fact.)

EXAMPLE 3.2.2. Some other reasonable loss functions (or, "measures of distance") are $W(T) = \|T\|$, $W(T) = \sum |t_i|$. (While we have not been able to verify in general for the last of these that (3.1.7) is satisfied, Lemma 3.2.2 will suffice for the following situation.) Suppose the co-ordinates $x_1, x_2, \cdots, x_m$ of X are independent identically distributed random variables with probability density f. Suppose f is bounded. (Some assumption of this type is necessary in order that (3.1.6) be satisfied.) And suppose $\int y^4 f(y)\, dy < \infty$ (so that (3.1.5) is satisfied). Then if W is any one of the two forms listed above there is an appropriate version of Stein's estimator which is better than the best invariant estimator.

The next two examples exhibit situations in which the conclusion of Theorem 3.1.1 is invalid because (3.1.7) is not satisfied. In a sense they are both trivial exceptions to the theorem, but we do not know of any significantly different situations in which (3.1.7) is not satisfied.

EXAMPLE 3.2.3. Let $m \geqq 3$. Suppose $W(T) = \omega(t_i)$ for some i or, more generally, suppose there is an orthogonal transformation B such that if $U = BX$, $W(X) = \omega(u_1)$. Then the rank of $E_{0Y}(XW'(X))$ is one, and the problem becomes a one dimensional estimation problem—namely that of observing $(u_i + \psi_i, Y)$, $i = 1, 2, \cdots, m$, and estimating ψ_1 ($\psi = U\theta$) with loss function ω. The variables $u_2 + \psi_2, \cdots, u_n + \psi_m$ in this case are "nuisance" variables. The question of admissibility of δ_0 in such a case is probably not easy to settle, see Stein (1960). However, if u_1 is independent of $(u_2, \cdots, u_m)$ then δ_0 can often be shown to be admissible as follows:

Suppose δ is better than δ_0. Then (by translation) there is a δ' which is better than δ_0 and has smaller risk for some ψ_1 and $(\psi_2, \cdots, \psi_m) = (0, \cdots, 0)$. Since $u_2, \cdots, u_m$ are independent of u_1, δ' is a randomized estimator of ψ_1 (given $(\psi_2, \cdots, \psi_m) = (0, \cdots, 0)$). It follows from the discussion in Section 1.3 that if the hypotheses of Theorem 2.1.1 are satisfied, δ' cannot be better than δ_0, a contradiction. Similar remarks are valid if $W(T) = \omega(t_i, t_j)$, etc., using Brown (1965).

To summarize, suppose there is some co-ordinate system in which the following is true. $W(T) = \omega(t_i)$ and, for fixed i and Y, X_i is independent of X_k, $k \neq i$. Suppose further that the hypotheses of Theorem 2.1.1 are satisfied for the problem where $\theta_k = 0$, $k \neq i$, and θ_i is unknown. Then the best invariant estimator δ_0 is admissible.

In the previous example one might say that the structure of W reduced the problem to one dimension. If the restriction that X have a probability density in E_m is removed similar examples can be constructed where X is essentially one-dimensional, and the best inv est is admissible. In the next example neither W nor X are by themselves essentially one-dimensional, yet the best inv est is easily seen to be admissible.

EXAMPLE 3.2.4. Suppose $m \geqq 3$, $n = 1$ (i.e. only X is observed), and

$$\begin{aligned} W(T) &= 0, & \|T\| &\leqq 1, \\ &> 0, & \|T\| &> 1; \\ p(X) &> 0, & \|X\| &\leqq 1, \\ &= 0, & \|X\| &> 1. \end{aligned} \tag{3.2.8}$$

Then it is easy to see that $R(\theta, \delta_0) \equiv 0$. Hence the best inv est is admissible. In addition, if δ is any estimator different from δ_0 on a set of positive measure then $R(\theta, \delta) > 0$ for some θ. Hence δ_0 is the (essentially) unique best inv est even though $m \geqq 3$.

The situations of Examples 3.2.3 and 3.2.4 can be combined to give slightly more general examples.

In the following example only assumption (3.1.6) of Theorem 3.1.1 is not satisfied, and the conclusion of that theorem is false. In this example we do not know whether the best inv est is admissible.

EXAMPLE 3.2.5. Let $m = 3$, $n = 1$, and

$$\begin{aligned} p(X) &= 1/8\pi\|X\|^{5/2}, & \|X\| &< 1, \\ &= 0, & \|X\| &> 1, \end{aligned} \tag{3.2.9}$$

and let $W(T) = \|T\|$. Since both p and W are spherically symmetric, it can be shown as in Stein (1960) that if δ_0 is inadmissible, then there is a spherically symmetric estimator which is better than δ_0. In particular, if δ' is a version of Stein's estimator, (3.1.8), which is better than δ_0, then there is an estimator δ of the form

$$\delta(X) = X(1 - b/(a + \|X\|^2)) \tag{3.2.10}$$

which is also better than δ_0. ((3.2.10) is the estimator originally considered by Stein.) It can be computed that assumption (3.1.6) is not satisfied in this example (although (3.1.5) and (3.1.7) are), and that

$$R(\theta, \delta_0) - R(\theta, \delta) = -8\pi\|\theta\|^{\frac{1}{2}}b/(a + \|\theta\|^2) + O(1/(a + \|\theta\|^2)) \qquad \text{as} \quad \theta \to \infty.$$

Hence the conclusion of Theorem 3.1.1 is false in this example; that is, no version of Stein's estimator, (3.1.8), is better than the best inv est.

Although we do not know whether the best inv est is admissible in the preceding example, it is possible (using $W(T) = \|T\|$ and an X such that $\Pr\{X = 0\} > 0$) to construct an example where $m = 3$, assumptions (3.1.5) and (3.1.7) are satisfied and yet the best inv est is admissible. We shall not give the details here.

It is possible to construct analogues of Example 3.2.5 for the other loss function given in Example 3.2.2.

3.3. *Inadmissibility of invariant estimates for non-convex loss functions and dimension* $m \geqq 3$. Here it is again the case that subject to certain assumptions there is a version of Stein's estimator which is always strictly better than the best invariant estimator. Both the statement and proof of Theorem 3.3.1 are very similar to the statement and proof of Theorem 3.1.1. In view of this similarity, a few of the details have been omitted from the proof of Theorem 3.3.1. Theorem 3.3.2—which is designed primarily to treat the case of discontinuous W—introduces a few new complications.

In the theorems below we have not attempted to state the regularity conditions in their weakest possible form, but have left them in a fairly weak form which has the advantage of being easy to state. (It is probable that in some of the conditions the phrase, "twice continuously differentiable" may be replaced by, "possessing a derivative having bounded variation"; and $[(\partial^2/\partial t_j \partial t_k)W(T)]dt_j$ may be replaced, as in (3.1.6), by $d_j w_k'(T)$—see above Lemma 3.1.1 for the definition of $d_j w_k'(T)$.)

It will be convenient to write $(\partial^2/\partial t_j \partial t_k)W(T, Y) = w_{jk}''(T, Y)$, and $W''(T, Y) = (w_{jk}''(T, Y))$, an $m \times m$ matrix. As in Section 3.1, observe the convention that if $n = 1$, $p(X - \theta, Y) = p(X - \theta)$ for all Y. E_{0Y} and $W'(T)$ are defined above Theorem 3.1.1.

THEOREM 3.3.1. *Consider a fixed-sample-size problem with* $m \geqq 3$, $\epsilon_0(X, Y) = X$, *and* $R_0 < \infty$, *in which all the second derivatives* w_{jk}'' *of* W *exist and are continuous in* T. *Suppose there is a set* S *of values of* Y *with* $\mu(S) > 0$ *satisfying the following*:

(i) *For all* $Y \varepsilon S$,

$$E_{0Y}\{\|X\|^4\|W'(X, Y)\|\} < \infty; \tag{3.3.1}$$

(ii) *For all* $Y \varepsilon S$ *there is a* $\gamma > 0$ *and a* c *such that*

$$E_{0Y}\{\|X\|^2|w_{jk}''(X + G(X), Y)|\} < c < \infty \tag{3.3.2}$$

for all functions $G(X)$ *such that* $\|G(X)\| < \gamma$, $1 \leqq j, k \leqq m$.

(iii) *For each* $Y \varepsilon S$,

$$E_{0Y}(XW'(X, Y)) \tag{3.3.3}$$

is a non-singular $(m \times m)$ *matrix.*

Define the estimator ϵ *(as in Theorem* 3.1.1*) by*

$$\begin{aligned} \epsilon(X, Y) &= (I - A(Y)/a(Y)\|X\|^2)X \qquad &\text{if } Y \varepsilon S \\ &= X &\text{if } Y \not\varepsilon S, \end{aligned} \tag{3.3.4}$$

where $A(Y) = (1/b(Y))E_{0Y}(XW'(X, Y))^{-1}$. *Then there exist functions* $a(Y)$, $b(Y)$ *such that* $R(\theta, \epsilon) < R(\theta, \epsilon_0)$ *for all* θ.

PROOF. Let $\tilde{Y} \varepsilon S$, $E(\cdot) = E_{0\tilde{Y}}(\cdot)$, and $W(T) = W(T, \tilde{Y})$ and $A = A(\tilde{Y})$. Using Taylor's theorem in m dimensions,

$$\begin{aligned} d &= E_{\theta,\tilde{Y}}\{W(\epsilon_0(X, \tilde{Y}) - \theta) - W(\epsilon(X, Y) - \theta)\} \\ &= E\{W(X) - W(X - A\cdot(X + \theta)/(a + \|X + \theta\|^2))\} \\ &= E\{W'(X)\cdot A\cdot(X + \theta)/(a + \|X + \theta\|^2) \\ &\quad - \tfrac{1}{2}(A\cdot(X + \theta)/(a + \|X + \theta\|^2))^\tau W''(X + G(X))(A\cdot(X \\ &\quad + \theta)/(a + \|X + \theta\|^2))\} \end{aligned} \tag{3.3.5}$$

where $\|G(X)\| \leqq \|A\cdot(X + \theta)/(a + \|X + \theta\|^2)\| \leqq k_1/a^{\frac{1}{2}}$ for some $k_1 < \infty$. (τ is the matrix, or vector, transpose operator.)

In view of (3.3.1), (3.3.2), and the fact that $\epsilon_0(X, Y) = X$, $E(W'(X)) = 0$. Using this, (3.3.1), and the equality (3.1.10) as in Theorem 3.1.1, the first term on the right of (3.3.5) becomes

$$E\{W'(X)AX/(a + \|\theta\|^2) - (\textstyle\sum 2x_i\theta_i)W'(X)A\theta/(a + \|\theta\|^2)\} + o(1/(a + \|\theta\|^2)). \tag{3.3.6}$$

In order to treat the second term in the expectation on the right of (3.3.5), use the inequality

$$\begin{aligned} (1/(a + \|X + \theta\|))^2 &\geqq (1/(a + \|\theta\|^2))^2 \\ &\quad - 2(\textstyle\sum x_i\theta_i + \sum x_i^2)/(a + \|\theta\|^2)^2(a + \|X + \theta\|^2). \end{aligned} \tag{3.3.7}$$

If this inequality is applied to that term and orders of magnitude are computed using (3.3.2) and the (trivial) inequalities (3.1.12) the resulting term is greater than

$$-(n^2\alpha^2\beta c/b^2)/(a + \|\theta\|^2) + o(1/(a + \|\theta\|^2)) \tag{3.3.8}$$

where $\alpha = \sup_{j,k}(E\{XW'(X)\}^{-1})_{jk}$ and $\beta = \sup_{j,k} E\{w''_{jk}(X)\}$. Hence

$$\begin{aligned} d &\geqq E\{W'(X)AX/(a + \|\theta\|^2) - (\textstyle\sum 2x_i\theta_i)W'(X)A\theta)/(a + \|\theta\|^2)^2\} \\ &\quad - (k^2/b^2)/(a + \|\theta\|^2) + o(1/(a + \|\theta\|^2)). \end{aligned} \tag{3.3.9}$$

As in (3.1.21) and (3.1.22) if a and b are chosen sufficiently large

$$d \geqq \tfrac{1}{2}(1/b)/(a + \|\theta\|^2) > 0 \tag{3.3.10}$$

for all θ. This implies that if $a(Y)$ and $b(Y)$ are chosen sufficiently large $R(\theta, \epsilon) < R(\theta, \epsilon_0)$.

This completes the proof of the theorem.

There are some important cases in which the hypotheses of Theorem 3.3.1 are not satisfied. The most notable of these is the case of fixed-size confidence set estimation. In that case W is a 0-1 function, and hence is not even continuous, let alone twice differentiable. The following theorem is designed to treat that case. By using a type of duality argument, the regularity conditions of Theorem 3.3.1 on W are shifted to p and so become in Theorem 3.3.2 analogous conditions on p. As was previously the case, the regularity conditions in Theorem 3.3.2 are not the weakest possible.

Define $p'(\cdot, Y)$ and $p''(\cdot, Y)$ from $p(\cdot, Y)$ in the same way as $W'(\cdot)$ and $W''(\cdot)$ were defined from W. (p' and p'' are row-vector and matrix-valued functions respectively.)

THEOREM 3.3.2. *Consider the fixed-sample-size problem with* $m \geqq 3$, $\epsilon_0(X, Y) = X$, *and* $R_0 < \infty$. *Suppose there is a set* S *with* $\mu(S) > 0$ *such that*

(i) *for each* $Y \varepsilon S$ *all the second partial derivatives* $p''_{jk}(\cdot, Y)$ *exist and are continuous;*

(ii) *for each* $Y \varepsilon S$,

$$\int W(X, Y)\|X\|^4\|p'(X, Y)\|\, dX < \infty; \tag{3.3.11}$$

(iii) *for each* $Y \varepsilon S$ *there is a* $\gamma > 0$ *and a* $c > 0$ *such that*

$$\int W(X, Y)\|X\|^2|p''_{jk}(X + G(X), Y)|\, dX < c \tag{3.3.12}$$

for all $G(X)$ *such that* $\|G(X)\| < \gamma$, $1 \leqq j, k \leqq m$;

(iv) *for each* $Y \varepsilon S$ *the* $m \times m$ *matrix*

$$M(Y) = \int W(X, Y)\{Xp'(X, Y)\}\, dX - E_{0Y}(W(X, Y)) \tag{3.3.13}$$

(*which exists as a consequence of* (3.3.11)) *is non-singular.*

Define ϵ *by*

$$\begin{aligned} \epsilon(X, Y) &= (I + B(Y)/(a(Y) + \|X\|^2))X && \text{if } Y \varepsilon S \\ &= X && \text{if } Y \not\varepsilon S \end{aligned} \tag{3.3.14}$$

where $B(Y) = (1/b(Y))M(Y)^{-1}$. *Then there exist functions* $a(Y)$, $b(Y)$ *such that* $R(\theta, \epsilon) < R(\theta, \epsilon_0)$ *for all* θ.

NOTE. The reader may check, using integration by parts, that $M(Y) = -E_{0Y}(XW'(X))$ if W is smooth enough. In that case (3.3.14) is the same as (3.3.4).

PROOF. Let $\tilde{Y} \varepsilon S$, $B = B(\tilde{Y})$ etc., and

$$Z = X + B\cdot(X + \theta)/(a + \|X + \theta\|^2) = H(X). \tag{3.3.15}$$

It is not hard to check that for all a sufficiently large (1) H is 1-1 and continuous, (2) H^{-1} is continuous, (3) all partial derivatives of H and H^{-1} exist and are continuous. ($a > n^2(\sup b_{ij})$ is large enough.) Thus

$$\begin{aligned} \int \hat{W}(X + B\cdot(X + \theta)/(a + \|X + \theta\|^2))p(X, \tilde{Y})\, dX \\ = \int \hat{W}(Z)p(H^{-1}(Z), \tilde{Y})|\det J_H(z)|^{-1}\, dZ \end{aligned} \tag{3.3.16}$$

where $J_H(z)$ is the Jacobian of the transformation H, and $\hat{W}(X) = W(X) - E_{0\tilde{Y}}(W(X))$.

Recall that $B \cdot (X + \theta)/(a + \|X + \theta\|^2) = O(a^{-\frac{1}{2}})$. Hence

$$H^{-1}(Z) = X = Z - B \cdot (X + \theta)/(a + \|X + \theta\|^2)$$
$$= Z - B \cdot (Z + \theta)/(a + \|Z + \theta\|^2) \tag{3.3.17}$$
$$+ O(a^{-\frac{1}{2}}/(a + \|Z + \theta\|^2)).$$

It can also be computed that

$$\det J_H(Z) = 1 + O(1/(a + \|Z + \theta\|^2)). \tag{3.3.18}$$

Thus, using the above,

$$\int \hat{W}(X + B(X + \theta)/(a + \|X + \theta\|^2))p(X, \tilde{Y})\, dX$$
$$= (1 + O(1/a)) \int [\hat{W}(Z)p(Z - B \cdot (Z + \theta)/(a + \|Z + \theta\|^2)), \tilde{Y}) \tag{3.3.19}$$
$$+ O(a^{-\frac{1}{2}}/(a + \|Z + \theta\|^2))\|p'(Z, Y)\|$$
$$+ O(1/a(a + \|Z + \theta\|^2))\|p''(Z + G(Z), Y)\|]\, dZ.$$

Using (3.3.11) and (3.3.12) as in the proof of Theorem 3.3.1 verifies that the last two terms in (3.3.19) are $O(1/(a + \|\theta\|^2))$. We shall skip the details. This gives finally

$$d = \int W(X)p(X, \tilde{Y})\, dX$$
$$- \int W(X + B \cdot (X + \theta)/(a + \|X + \theta\|^2))p(X, \tilde{Y})\, dX$$
$$= \int \hat{W}(Z)\{p(Z, \tilde{Y}) - p(Z - B \cdot (Z + \theta)/(a + \|Z + \theta\|^2), \tilde{Y})\}dZ$$
$$+ O(1/a) \int \hat{W}(Z)p(Z - B \cdot (Z + \theta)/(a + \|Z + \theta\|^2), \tilde{Y})dZ \tag{3.3.20}$$
$$+ o(1/(a + \|\theta\|^2))$$
$$= \{\int [p(Z, \tilde{Y}) - p(Z - B \cdot (Z + \theta)/(a + \|Z + \theta\|^2), \tilde{Y})]\hat{W}(Z)\, dZ\}$$
$$\cdot (1 - O(1/a)) + o(1/(a + \|\theta\|^2)).$$

The integral on the right of (3.3.20) is the same as (3.3.5) with the roles of $\hat{W}$ and $p(\cdot, Y)$ interchanged. Also the conditions of Theorem 3.3.2 are those of (3.3.1) with the roles of $\hat{W}$ and $p(\cdot, Y)$ interchanged. (It is of course not true that $W \geqq 0$ or that $\int |\hat{W}(z)|\, dZ < \infty$, but it can easily be checked that in the proof of Theorem 3.3.1 only $\int |W(Z)|\, p(Z, Y)\, dZ < \infty$ and the fact that p is bounded below has been used; $\int p(X, Y)\, dX < \infty$ has not been used!) Similarly the estimator (3.3.14) is the same as the estimator (3.3.4) with the roles of p and $\hat{W}$ interchanged. Thus just as in the proof of Theorem 3.3.1 for $a(Y)$ and $b(Y)$ chosen sufficiently large $d \geqq \frac{1}{2}(1/b)/(a + \|\theta\|^2) > 0$. This then implies $R(\theta, \epsilon) < R(\theta, \epsilon_0)$ if $a(Y)$ and $b(Y)$ are properly chosen.

This completes the proof of the theorem.

3.4. *Summary, and remarks concerning the general sequential case.* In Sections 3.1 and 3.3 we have shown that for $m \geqq 3$ the best invariant estimator is inadmissible, subject to certain assumptions. The assumptions regarding convergence of moments and smoothness (i.e. (3.1.5) and (3.1.6), or (3.3.1) and (3.3.2)) can be easily interpreted, and are of a fairly mild nature. It appears likely that the remaining assumption (concerning non-singularity of the matrix (3.1.7) or of (3.3.3)) is usually, or always, satisfied if the problem is a non-degenerate problem in $m \geqq 3$ dimensions. However, we have only been able to prove that this non-singularity condition is satisfied in certain important special cases (see Lemma 3.2.2). The results given in this paper generalize those of Stein (1956), and Farrell (private communication).

While the previous theorems deal only with the fixed-sample-size case, the results can also be applied to the general sequential problem. Suppose the best invariant estimator is $\delta_0 = (\gamma_0, \sigma_0)$. Define $X = X$, $\dot{Y} = Y_1, Y_2 \cdots, Y_{\sigma_0(Y)-1}$ (where $Y = (Y_1, Y_2, \cdots))$ and $\dot{W}(t, y) = W(t, \sigma_0(y), y)$. If for a set of values of Y having positive μ measure the hypotheses of Theorems 3.1.1, 3.3.1, or 3.3.2 are satisfied then δ_0 is inadmissible. Further details are included in the following theorem, which is the sequential analog of Theorem 3.1.1. Similar analogs of Theorems 3.3.1 and 3.3.2 are also valid.

Given $\delta_0 = (\gamma_0, \sigma_0)$, define

$$\dot{E}_{\theta,Y}(f) = \int f(X, Y) p_{\sigma_0(Y)}(X - \theta, Y)\, dX \tag{3.4.1}$$

where p_n is defined in (1.1.1). Thus $\dot{E}_{\theta,Y}(f)$ is the expected value of f when θ is the true parameter value given that $\dot{Y} = Y_1, Y_2, \cdots, Y_{\sigma_0(y)-1}$. Let $\mathfrak{A}_\sigma$ denote the sub σ-field of $\mathfrak{A}^\infty$ induced by $\dot{Y}$.

THEOREM 3.4.1. *Suppose $\dot{W}(T, Y)$ is convex in T for each $\dot{Y}$, $m \geqq 3$, and $R_0 < \infty$. Suppose there is a set $S \varepsilon \mathfrak{A}_\sigma$ of values of $\dot{Y}$ such that $\mu(S) > 0$ satisfying:*

(i) *For each $\dot{Y} \varepsilon S$ there is a $\gamma > 0$ and c_1 such that $|\beta| < \gamma$ implies*

$$E_{0,\dot{Y}}\{\|X - \gamma_0(\dot{Y})\|^4 \|W'(X - \gamma_0(\dot{Y}) + \beta U_i, \dot{Y})\|\} < c_1 \tag{3.4.2}$$

for $1 \leqq i \leqq m$.

(ii) *For each $\dot{Y} \varepsilon S$ there is a $c_2 > 0$ such that*

$$\iint p_{\sigma_0(\dot{Y})}(T + \beta U_i, \dot{Y})(d_i \dot{w}_i'(T - \gamma_0(\dot{Y}), \dot{Y})\, dT_i^* < c_2 \tag{3.4.3}$$

for $0 < \beta < \gamma$ and $1 \leqq i \leqq m$.

(iii) *For each $\dot{Y} \varepsilon S$,*

$$E_{0,\dot{Y}}\{(X - \gamma_0(\dot{Y}))W'(X - \gamma_0(\dot{Y}), \dot{Y})\} \tag{3.4.4}$$

is a non-singular matrix ($m \times m$).

Then δ_0 is inadmissible.

PROOF. Using Theorem 3.1.1 for each $\dot{Y} \varepsilon S$ there is an estimator $\epsilon(X, \dot{Y})$ such that $\dot{E}_{\theta,\dot{Y}}\{W(\epsilon(X, \dot{Y}) - \theta, \dot{Y})\} < E_{\theta,\dot{Y}}\{W(\epsilon_0(X, \dot{Y}) - \theta, \dot{Y})\}$ for all θ. Since the conditions (3.4.2), (3.4.3), and (3.4.4) depend only on $Y_1, \cdots, Y_{\sigma_0(y)}$ and since $S \varepsilon \mathfrak{A}_\sigma$, if we let $\delta(x, y) = (\epsilon(x, y), \sigma_0(y))$ if $y \varepsilon S$ and $\delta(x, y) = (\epsilon_0(y), \sigma_0(y))$ otherwise, then $\delta(x, y)$ is an estimation procedure (satisfying (1.1.3)); and δ is better than δ_0. This completes the proof of the theorem.

This theorem indicates that even in sequential cases δ_0 is usually inadmissible if $m \geqq 3$. In specific cases, it is usually difficult to verify whether the hypotheses of Theorem 3.4.1 are satisfied. However in the case of square error this is not true. The following corollary gives the result.

COROLLARY 3.4.1. *In the special sequential problem, suppose* $W(T, n) = \|T\|^2 + W_2(n)$ *where* $W_2(n)$ *increases to* ∞ *as* $n \to \infty$. *Suppose*

$$\int \|T\|^6 f(T)\, dT < \infty. \tag{3.4.5}$$

Then δ_0 *is inadmissible.*

PROOF. Let $S = \{\dot{Y}\}$, $\dot{W}(T, Y) = W(T, \sigma_0(Y))$. Using Lemma 3.2.1, condition (3.4.3) is satisfied. Using (3.4.5), and the Martingale stopping theorem, Doob (1953), p. 300, (3.4.2) is satisfied. Then, using Lemma 3.2.2, (3.4.4) is satisfied. This completes the proof of the corollary.

Acknowledgment. I would like to express my deepest gratitude to Professor J. Kiefer for his very helpful interest throughout the preparation of this manuscript. I would also like to thank Professor R. Farrell for giving an earlier version of this paper a very careful reading, for finding one major error and for making many other suggestions which have improved the final product.

REFERENCES

BLACKWELL, D. (1951). On the translation parameter problem for discrete variables. *Ann. Math. Statist.* **22** 393–399.

BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

BLYTH, C. R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Statist.* **22** 22–42.

BROWN, L. D. (1965). On the admissibility of invariant estimators of two dimensional location parameters. Mimeographed notes, Birkbeck College, London.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

FARRELL, R. (1964). Estimators of a location parameter in the absolutely continuous case. *Ann. Math. Statist.* **35** 949–959.

FOX, M. and RUBIN, H. (1964). Admissibility of quantile estimates of a single location parameter. *Ann. Math. Statist.* **35** 1019–1031.

HAHN, H. (1948). *Reelle Funktionen.* Chelsea, New York.

HODGES, J. L., JR., and LEHMANN, E. L. (1951). Some applications of the Cramér-Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 13–22. Univ. of California Press.

KARLIN, S. (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist.* **29** 406–436.

KATZ, M. W. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Ann. Math. Statist.* **32** 136–142.

KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28** 573–601.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

LI͡APUNOV, A. A. and ARSENIN, W. J. (1955). *Arbeiten zur Deskriptiven Mengenlehre.* Berlin, Deutscher Verlag von Wissenschaften.

PITMAN, E. J. G. (1939). The estimation of location and scale parameters of a continuous population of any given form. *Biometrika* **30** 391–421.

STEIN, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Statist.* **26** 518–522.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 197–206. Univ. of California Press.

STEIN, C. (1959). The admissibility of Pitman's estimator for a single location parameter. *Ann. Math. Statist.* **30** 970–979.

STEIN, C. and JAMES, W. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379. Univ. California Press.

WALD, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *Ann. Math. Statist.* **10** 299–326.