The Annals of Statistics
1986, Vol. 14, No. 1, 174–187

ASYMPTOTICS FOR CONFIGURAL LOCATION ESTIMATORS[1]

BY STEPHAN MORGENTHALER

Yale University

This paper examines the asymptotic properties of compromise estimators. By this we mean an estimation method which compromises between a finite number of sampling situations in a small sample optimal way. We develop the asymptotic theory of such estimators in the location problem and show that under a specific choice of a pair of sampling situations the compromise estimator is asymptotically robust in Huber's sense.

1. Introduction. Configural polysampling denotes a method of estimation which is geared to small sample sizes and produces "robust" methods [see Pregibon and Tukey (1981)]. There are important differences to the robustness philosophy as developed by Huber (1964). Since in small samples the distributions of estimators are quite intractable, one has to rely on numerical methods in order to evaluate even relatively simple performance summaries such as the mean-square-error. This holds true except in some simple cases—such as the Gaussian sampling model—where a few expectations can be evaluated in closed form. In this connection, it is important for the statistical community to realize that numerical methods are perfectly acceptable. They do, however, limit the number of sampling situations we can take into consideration. This is in contrast to an asymptotic approach, where, for simple models, an infinity of sampling situations can be considered simultaneously [Huber (1964)].

Pitman (1939), for example, solves the small sample problem for a single sampling situation in a location and scale setting. In this paper we will show what happens if Pitman's method is extended to two sampling situations with known scale. And we will address the question of the asymptotic performance of such estimators.

An asymptotic analysis is the simplest way to learn something about the behavior of an estimator in a variety of sampling situations. But it only gives a partial answer and we should not forget the more important approach based on performing "experiments" for small sample sizes. This paper, however, will restrict attention to asymptotic discussions.

In Section 2 we will introduce the idea of compromise estimators and discuss their optimality properties. Section 3 contains the corresponding asymptotic theory. As an example we define a compromise estimator which is asymptotically everywhere at least as good as Huber's minimax estimator.

Received April 1984; revised June 1985.
[1] The work reported here was supported in part by ARO Grant DAAG29-84-K-0207.
AMS 1980 *subject classifications*. Primary 62F35; secondary 62F12.
Key words and phrases. Robust estimation, conditional inference, equivariance, asymptotics.

2. Configural estimators.

2.1. Pitman's estimator. Let $x_1, x_2, \ldots, x_n$ be n independent observations from a symmetric distribution $F(x - \mu)$ where $1 - F(x) = F(-x)$ for all x. We also assume that $F(x) \neq 0$ or 1 for any finite x and furthermore that $F(\ )$ has density $f(\ )$ with respect to Lebesgue measure.

We restrict attention to symmetric sampling situations in order to avoid the issue of what "parameter" we try to estimate. Symmetry of the underlying distribution allows us to define a target, namely μ = *center of symmetry*. Furthermore, we will not get into any discussions if later on we allow for two—or many—different sampling situations. The center of symmetry is well defined for all symmetric shapes which means the estimation of μ is a well defined problem for a large class of sampling situations.

The solution Pitman gives is

$$T_F(x_1, \ldots, x_n) = -\frac{\int_{-\infty}^{\infty} r \prod_{i=1}^{n} f(x_i + r)\, dr}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f(x_i + r)\, dr} \tag{2.1}$$

[see Pitman (1939)]. This estimator has the smallest mean-square-error among all location equivariant estimators. Location equivariance is a reasonable restriction on a location estimator $T(\ )$. It means that

$$T(x_1 + r, \ldots, x_n + r) = T(x_1, \ldots, x_n) + r, \qquad r \in R, \tag{2.2}$$

i.e., the estimator changes in the same way as the sample.

REMARKS. (1) The most revealing way of deriving (2.1) is through the concept of "configurations." By this notion, we mean the pattern of the points in the (ordered) sample, as specified, for example, by the gaps between the observations. It is easily seen that this is an ancillary statistic. The Pitman estimator then is chosen such that *conditioned* on the configuration the estimate is unbiased. Since the conditional variance cannot be affected by the choice of the estimate (under equivariance), this has to produce the smallest mean-square-error.

(2) The conditions on $f(\ )$ such that (2.1) exists are discussed in Pitman (1939).

Formula (2.1) produces an estimator T_F of the center of symmetry μ no matter what the underlying sampling situation. It, therefore, need not be so that the x_i's are sampled from $F(x - \mu)$.

Let us introduce $G(x - \mu)$—again $G(x) = 1 - G(-x)$ for all x's—as the sampling situation for $x_1, \ldots, x_n$. This is a new way of looking at the Pitman estimator T_F and it, of course, immediately lets us see the optimality property in a new light. If, e.g., $F = \Phi$ and G = Cauchy, we are looking at the behavior of the arithmetic mean under Cauchy sampling. If we are open minded about the assumptions we base our inference on, we have to admit that in small samples we cannot, with any reasonable precision, know what the underlying sampling situation is nor should we attempt to make inferences about it. Huber (1964) formalizes the idea of a robust method as a procedure which "behaves well" in the neighborhood of a parametric model. Huber, therefore, would allow $G(\ )$ to be chosen somewhere near $F(\ )$ and he modifies T_F in such a way that the behavior of the new estimate is acceptable for all allowed $G(\ )$s. This leads us away from considering estimates like T_F which are optimized at a single "point." Since—in small samples—we will never be able to tell at which "point" we are, it ought to be obvious that single-point-optimization is a bad strategy.

2.2. Compromise estimators. Let us now consider the case where $x_1, \dots, x_n$ is a sample from either $F_1(x - \mu)$ or $F_2(x - \mu)$, where F_1 and F_2 satisfy all the constraints of F (see the beginning of Section 2.1). We are now interested in location equivariant estimators which optimize at two "points," namely F_1 and F_2, simultaneously. This is obviously impossible. However, decision theory teaches us that estimates of the form

$$T_{F_1, F_2, \alpha}(x_1, \dots, x_n) = -\frac{\int r\{\alpha\prod_{i=1}^n f_1(x_i + r) + (1 - \alpha)\prod_{i=1}^n f_2(x_i + r)\}\, dr}{\int\{\alpha\prod_{i=1}^n f_1(x_i + r) + (1 - \alpha)\prod_{i=1}^n f_2(x_i + r)\}\, dr} \tag{2.3}$$

$(0 < \alpha < 1)$ are bioptimal in the sense that they cannot be improved in both sampling situations F_1 and F_2 simultaneously [see Ferguson (1967)].

REMARKS. (1) We can also write

$$\begin{aligned} T_{F_1, F_2, \alpha}(x_1, \dots, x_n) &= \alpha w_{F_1}(x_1, \dots, x_n) T_{F_1}(x_1, \dots, x_n) \\ &\quad + (1 - \alpha) w_{F_2}(x_1, \dots, x_n) T_{F_2}(x_1, \dots, x_n), \end{aligned} \tag{2.4}$$

where

$$w_{F_k}(x_1, \dots, x_n) = \frac{\int\prod_{i=1}^n f_k(x_i + r)\, dr}{\int\{\alpha\prod_{i=1}^n f_1(x_i + r) + (1 - \alpha)\prod_{i=1}^n f_2(x_i + r)\}\, dr}$$

$(k = 1, 2)$ and $T_{F_k}(\ )$ is defined in (2.1). We, therefore, can interpret the family of bioptimal estimators as a weighted mean of the single-situation optimal estimators. Note, however, that the weights are "adaptive," they depend on the sample values. Of course, any equivariant estimator can be represented as a weighted mean of the single-situation optimal estimators. What matters here is the simplicity and form of the weights together with their small sample optimality property.

(2) It is clear from (2.3) that $T_{F_1, F_2, 0} = T_{F_2}$ and $T_{F_1, F_2, 1} = T_{F_1}$.

(3) The picture which helps us most in understanding the compromise estimators is shown in Figure 2.1.

Note that since we only consider location equivariant estimators the risk in any given situation does not depend on the parameter value μ [see Ferguson (1967)]. The bioptimal or compromise estimators are the ones which lie on the convex boundary curve.

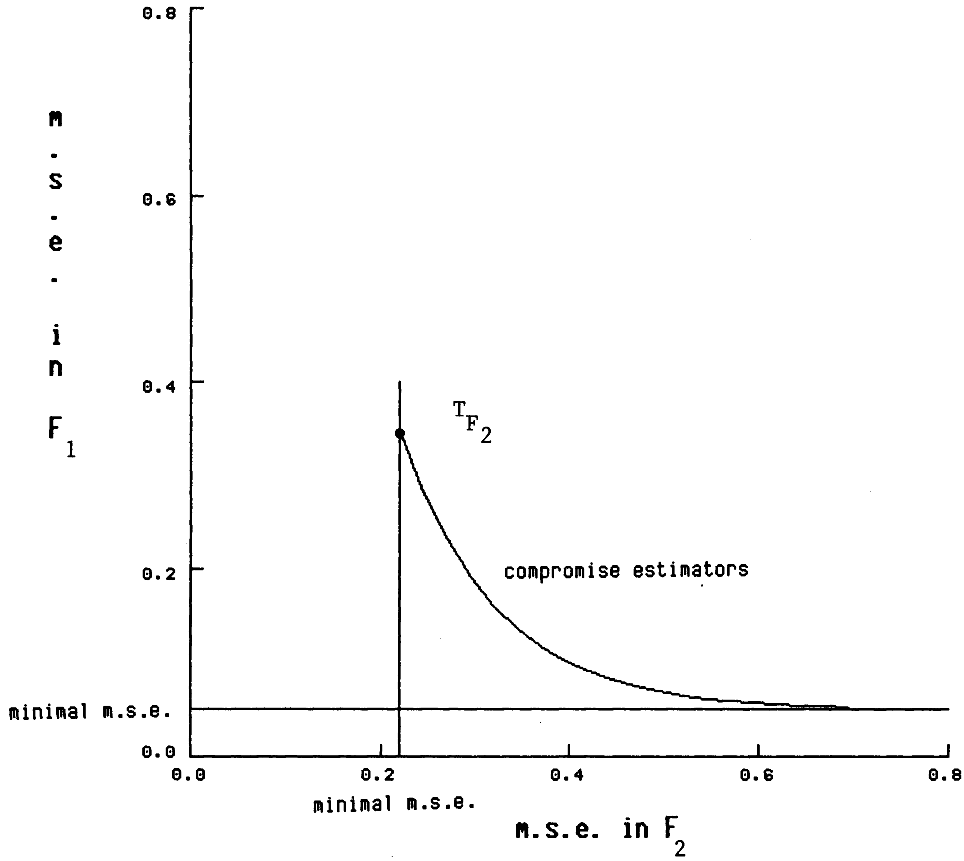

FIG. 2.1. *Plot of the mean-square-errors.*

(4) A Bayesian interpretation of the estimator (2.3) is possible. In that framework, $(\alpha, 1-\alpha)$ is a prior distribution on the set of underlying sampling shapes.

(5) In order to implement (2.3) in an actual application, formula (2.4) has some interesting interpretations. Pregibon and Tukey (1981) derive the formulas from the point of view of sampling in the case of unknown scale. This leads to the consideration of different weights w_{F_1} and w_{F_2}.

The choice of the two compromising distributions F_1 and F_2 is of importance in applications of the technique. In many applications it is traditional to consider $F_1 = \Phi$, the Gaussian shape. The choice of F_2 is somewhat related to the choice of the contamination parameter ε in Huber's model. F_2 will influence two aspects [see (2.4)]:

(i) the "relative weights" w_{F_1} and w_{F_2};

(ii) the "other" optimal estimator T_{F_2}.

These two aspects have an interpretation in the theory of M estimators. The first is connected with the choice of tuning constants, as for example, k in Huber's ψ_k

function ($\psi_k(x) = \max(-k, \min(k, x))$), and the second with the shape of the ψ function. From small sample studies, we know for example that a redescending ψ function is advantageous—it costs little and buys a lot. If we want such a behavior then F_2 has to be chosen as a heavy-tailed counterpart to the Gaussian, for example, a distribution with Pareto tails.

3. The asymptotic behavior of compromise estimators. In this section we are going to explore what happens to compromise estimators [see (2.3) or (2.4)] if we sample from a distribution $G(\ )$ and let the sample size n grow. We will see that the weights w_{F_1} and w_{F_2} usually tend to $(0,1)$ or $(1,0)$, respectively. A compromise estimator for large sample sizes, therefore, will be close to either the optimal estimate under F_1 or the optimal estimate under F_2. This is a reasonable behavior since the "information" about the sampling situation $G(\ )$ grows as the sample size gets large. The distinction between F_1 and F_2 is, therefore, more and more estimable. In a few words then, we can say that compromise estimators exhibit an adaptive behavior with the relative weights w_{F_1} and w_{F_2} [see (2.4)] guiding the adaptation.

3.1. The asymptotic behavior of the relative weights. Suppose $x_1, \dots, x_n$ is a sample of size n from $G(x - \mu)$. We assume that $G(\ )$ is symmetric around 0. The relative weights are defined as

$$(3.1)\quad w_{F_k}(x_1, \dots, x_n) = \frac{\int \prod_{i=1}^n f_k(x_i + r)\, dr}{\int \{\alpha \prod_{i=1}^n f_1(x_i + r) + (1-\alpha) \prod_{i=1}^n f_2(x_i + r)\}\, dr}$$

($k = 1$ or 2), where the notation is the same as in (2.4).

The following lemma treats an overly nice case, a model distribution with Pareto tails for example is not covered.

LEMMA 3.1. *Let us assume that both* $-\log f_1$ *and* $-\log f_2$ *are convex. We also require that the first two derivatives exist and that the second derivatives be bounded*:

$$\frac{d}{dx} \log f_1(x), \qquad \frac{d}{dx} \log f_2(x)$$

exist and

$$\frac{d^2}{dx^2} \log f_1(x), \qquad \frac{d^2}{dx^2} \log f_2(x)$$

exist and are bounded from below.

Finally, with regard to the modeling distributions F_1 *and* F_2, *we need*

$$\frac{d^2}{dx^2} \log f_1(x) < 0 \quad \text{and} \quad \frac{d^2}{dx^2} \log f_2(x) < 0 \quad \text{for } x \in [-\delta, \delta] \quad \text{for some } \delta > 0.$$

Let us furthermore assume that G *is such that the functions*

$$A^1(r) = \int \log f_1(x + r)\, dG(x)$$

and

$$A^2(r) = \int \log f_2(x+r)\,dG(x)$$

exist for all r, achieve a unique maximum at r = 0, and are such that the second derivative at r = 0 exists and is bounded. If

$$\int \log f_1(x)\,dG(x) > \int \log f_2(x)\,dG(x) \tag{3.2}$$

it follows that

$$\frac{w_{F_2}(x_1,\dots,x_n)}{w_{F_1}(x_1,\dots,x_n)} \to 0 \quad a.s.$$

PROOF. Let $X_1, X_2, \dots$ denote a sequence of i.i.d. random variables with common distribution G. From (3.1) we have

$$\frac{w_{F_2}(X_1,\dots,X_n)}{w_{F_1}(X_1,\dots,X_n)} = \frac{\int\prod_{i=1}^n f_2(X_i+r)\,dr}{\int\prod_{i=1}^n f_1(X_i+r)\,dr}.$$

Now

$$\begin{aligned} I(X_1,\dots,X_n) &= \int \prod_{i=1}^n f(X_i+r)\,dr \\ &= \int \exp\left(n\left(\frac{1}{n}\sum_{i=1}^n \log(f(X_i+r))\right)\right) dr \\ &= \int \exp(nA_n(r))\,dr, \end{aligned}$$

where

$$A_n(r) = \frac{1}{n}\sum_{i=1}^n \log(f(X_i+r))$$

and f stands for either f_1 or f_2.

(1) Due to the strong law of large numbers we have

$$A_n(r) \to A(r) \quad \text{a.s. for all } r. \tag{3.3}$$

Convergence is uniformly in r for $r \in [-\delta, \delta]$ because of our convexity assumptions. For the same reason, the function $A_n(r)$ is maximized in some interval. Let R_0^n denote the center of that interval of maximal points. For n large enough the maxima is unique since $-\log f$ is strictly convex in $[-\delta, \delta]$. It follows that

$$R_0^n \to 0 \quad \text{a.s.}$$

(2) Let us define

$$I_{m,n}(X_1,\dots,X_n) = \int \exp(mA_n(r))\,dr.$$

The integral $I_{m,n}$ allows an asymptotic expansion as $m \to \infty$. For large values of

m, we get

$$I_{m,n}(X_1,\ldots,X_n) \sim \int \exp\left(m\left[A_n(R_0^n) - \tfrac{1}{2}|A_n''(R_0^n)|(r - R_0^n)^2\right]\right) dr \tag{3.4}$$
$$\sim \exp(mA_n(R_0^n))\left(\frac{2\pi}{m}\right)^{1/2}(|A_n''(R_0^n)|)^{-1/2}.$$

The theory of asymptotic expansions in this simple case is treated, for example, in Chapter 4 of deBruijn (1981). Convergence of $I_{m,n}(X_1,\ldots,X_n)$ takes place in the sense of real functions. The values of $X_1,\ldots,X_n$ do not matter. For large enough n, the convergence in (3.4) is uniform in n. This follows from the error formula in deBruijn (1981, page 64) and the bound on the second derivative of $A_n(r)$ provided by the lower bound on the second derivative of $\log f$. In short, the argument goes as follows. The asymptotic expansion (3.4) works uniformly with respect to n because the function $A_n(r)$ can be approximated by a quadratic near its maxima uniformly in n.

(3) If we blend the probability structure which underlies the sequence $X_1, X_2,\ldots$ with the asymptotic approximation (3.4), we get

$$\frac{1}{n}\log I(X_1,\ldots,X_n) \to A(0) = \int \log f(x)\,dG(x) \quad \text{a.s.}$$

We therefore conclude from

$$\int \log f_1(x)\,dG(x) > \int \log f_2(x)\,dG(x)$$

that

$$\frac{1}{n}\log I_1(X_1,\ldots,X_n) - \frac{1}{n}\log I_2(X_1,\ldots,X_n)$$
$$= -\left(\frac{1}{n}\right)\log\left(\frac{w_{F_2}(X_1,\ldots,X_n)}{w_{F_1}(X_1,\ldots,X_n)}\right) \to \text{constant} > 0 \quad \text{a.s.,}$$

where $I_1(X_1,\ldots,X_n)$, $I_2(X_1,\ldots,X_n)$ refer to $f = f_1$ and $f = f_2$, respectively. From this last statement the assertion of the lemma follows immediately.

However, Lemma 3.1 is not strong enough for our purpose. If the underlying distribution G does not have a first moment for example, then $A^1(r)$ and $A^2(r)$ will be $-\infty$ for all r. Under stronger assumptions on F_1 and F_2, we can prove the following.

LEMMA 3.2. *Let G be such that its second moment is infinite. The assumptions on F_1 and F_2 from Lemma* 3.1 *still hold. Furthermore we assume that the ratio $d(x) = f_2(x)/f_1(x)$ is bounded and satisfies* $(\log d(x))^{-1} = O(x^{-2})$ *for large* $|x|$, i.e., $\log d(x)$ *tends to* $-\infty$ *like a quadratic or faster.*

It then follows that

$$\frac{w_{F_2}(x_1,\ldots,x_n)}{w_{F_1}(x_1,\ldots,x_n)} \to 0 \quad a.s.$$

PROOF. Let $X_1, X_2, \ldots$ denote a sequence of i.i.d. random variables with common distribution G. Then we have

$$\frac{\int \prod_{i=1}^n f_2(X_i + r)\, dr}{\int \prod_{i=1}^n f_1(X_i + r)\, dr} = \frac{\int \prod_{i=1}^n d(X_i + r) \prod_{i=1}^n f_1(X_i + r)\, dr}{\int \prod_{i=1}^n f_1(X_i + r)\, dr}$$

$$\leq \max_r \prod_{i=1}^n d(X_i + r).$$

The lemma is proved if we can show that $\max \prod_{i=1}^n d(X_i + r) \to 0$ a.s. But this is a consequence of $\sum_{i=1}^n \log d(X_i + r) \to -\infty$ a.s. which follows from our assumptions via the strong law of large numbers.

REMARKS. (1) The asymptotic expansion (3.4) shows how closely the maximum likelihood estimator is connected to the Pitman estimator. Note that $|A''(0)|$ is equal to the Fisher information if $G = F$, i.e., the sampling distribution and the modeling distribution are the same. We will see below that the maximum likelihood estimator is indeed asymptotically equivalent to the Pitman estimator.

(2) It is reasonable to believe that Lemma 3.1 holds in greater generality. The convexity conditions on the $-\log$ densities are probably not needed and could be replaced by suitable assumptions on F_1 and F_2 close to the origin.

COROLLARY 3.1. *Under the assumptions of the Lemma* 3.1 *or Lemma* 3.2, *the compromise estimator* $T_{F_1, F_2, \alpha}$ $(\alpha = 0)$ *is asymptotically equivalent to the Pitman estimator* T_{F_1}.

PROOF. Apply the lemmas to formula (2.4).

REMARKS. (1) Corollary 3.1 states that with increasing sample size the compromise estimator will pick either one of the two single-situation-optimal estimates depending on (3.2).

We therefore expect that

$$\int \log f_1(x)\, dG(x) - \int \log f_1(x)\, dG(x) = \int \log\left(\frac{f_1(x)}{f_2(x)} \right) dG(x) \tag{3.5}$$

is a quantity which decides whether the sampling situation G is "closer" to the modeling situation F_1 or the modeling situation F_2.

The quantity (3.5) is closely related to the Kullback–Leibler mean information for discrimination [Kullback and Leibler (1951)]. Their formula is

$$I(1{:}\,2) = \int \log\left(\frac{f_1(x)}{f_2(x)} \right) f_1(x)\, dx,$$

where $I(1{:}\ 2)$ is the mean information for discrimination per observation from sampling situation F_1.

(2) The asymptotic behavior of the compromise estimators (2.3) does not depend on α [unless (3.5) = 0].

(3) More results about Pitman estimators can be found in Johns (1979) and Klaassen (1981). Easton (1984) has proved the results given in Section 3.1 for the more general case of unknown scale.

3.2. Asymptotics of the Pitman estimator. In order to get asymptotic efficiencies for the compromise estimators we need to know more about the asymptotic behavior of the Pitman estimators T_{F_1} and T_{F_2}. Port and Stone (1974) provide the information in the case where the sampling situation and the modeling situation are identical. In our more general setup we can argue the following way:

$$T_F(x_1,\dots,x_n) = -\frac{\int r\exp(nA_n(r))\,dr}{\int \exp(nA_n(r))\,dr},$$

where

$$A_n(r) = \frac{1}{n}\sum_{i=1}^{n}\log f(x_i+r) \qquad \left(f(\ ) = \frac{d}{dx}F(\ )\right).$$

If we expand the numerator asymptotically we get

$$T_F(x_1,\dots,x_n) \sim -\frac{\exp(nA_n(r_0^n))\int r\exp\left(-\frac{1}{2}nA_n''(r_0^n)(r-r_0^n)^2\right)}{\int \exp(nA_n(r))\,dr}$$

$$\sim -r_0^n,$$

where r_0^n maximizes $A_n(r)$ (see deBruijn, page 66). We therefore showed that asymptotically the Pitman estimator and the maximum likelihood estimator $(-r_0^n)$ agree. This agreement is good enough—namely $T_F(X_1,\dots,X_n) + r_0^n = o_p(n^{-1/2})$—to conclude that the asymptotic distributions are the same. Huber (1967) then provides the necessary results.

3.3. Huber's contamination model: An example. To illustrate the use of the theory we developed, let us look at the compromise estimators based on the two modeling densities

$$f_1(x) = \phi(x) = \frac{1}{(2\pi)^{1/2}}\exp\left(-\tfrac{1}{2}x^2\right),$$

$$\begin{aligned} f_2(x) &= (1-\varepsilon)\phi(x) && \text{if } |x| < k,\\ &= \frac{(1-\varepsilon)}{(2\pi)^{1/2}}\exp\left(\frac{k^2}{2} - k|x|\right) && \text{otherwise,} \end{aligned}$$

where k is such that $(2\phi(k)/k) - 2\Phi(-k) = \varepsilon/(1-\varepsilon)$. The alternative density is, of course, the least favorable choice in the class of distributions $\{(1-\varepsilon)\Phi(\ ) + \varepsilon H(\ ) : H(\ ) \text{ symmetric}\}$ [see Huber (1964)].

The asymptotic variance of an estimator compromising between these two symmetric situations [see (2.3)] will be equal to either of the asymptotic variances of the Pitman estimators,

$$T_{F_1} = \text{arithmetic mean}$$

or

$$T_{F_2} = \text{Pitman estimator for the least favorable distribution.}$$

If we sample from distribution $G(\ )$, we have for these asymptotic variances ($\mu_G = \int x\, dG(x)$)

$$\text{as.}\operatorname{var}_G(T_{F_1}) = \int (x - \mu_G)^2\, dG(x),$$

$$\text{as.}\operatorname{var}_G(T_{F_2}) = \frac{\int (\psi_k(x - \mu_G))^2\, dG(x)}{\left(\int \psi_k'(x - \mu_G)\, dG(x)\right)^2},$$

where

$$\psi_k(x) = -f_2'(x)/f_2(x) = \max(-k, \min(k, x)).$$

In his 1964 paper, Huber shows that the M-estimator based on $\psi_k(\ )$ is asymptotically minimax for sampling situations chosen from the contamination class. Since T_{F_2} has the same asymptotic behavior as this M-estimator, the same claim can be made for T_{F_2}. The following proposition explains the asymptotic behavior of the compromise estimator [see (2.3)]. In order to be able to prove it, the contamination class needs to be reduced a bit.

PROPOSITION 3.1. *Let* $G(x) = (1 - \varepsilon)\Phi(x) + \varepsilon H(x)$, *where* $H(x) + H(-x) = 1$ *for all* x*'s and* $H(\ )$ *puts all its mass outside the interval* $[-k, k]$, *but is otherwise arbitrary. Furthermore, assume that* $0 \le \varepsilon \le 0.5$. *Then*

$$\text{as.}\operatorname{var}_G(\textit{compromise estimator}) \le \text{as.}\operatorname{var}_G(\textit{Huber's minimax estimator}).$$

PROOF. From Lemma 3.1 and Lemma 3.2, we know that

$$\begin{aligned}
&\int \log f_1(x)\, dG(x) - \int \log f_2(x)\, dG(x)\\
&\quad = \int \left\{\log \frac{1}{(2\pi)^{1/2}} - \tfrac{1}{2}x^2\right\} dG(x) - \int \log\left(\frac{1-\varepsilon}{(2\pi)^{1/2}}\right) dG(x)\\
&\qquad - \int_{-k}^{k} -\tfrac{1}{2}x^2\, dG(x) - 2\int_k^{\infty}\left(\frac{k^2}{k} - k|x|\right) dG(x)\\
&\quad = -\log(1-\varepsilon) + 2\int_k^{\infty}\left\{k|x| - \frac{k^2}{2} - \frac{x^2}{2}\right\} dG(x)
\end{aligned} \tag{3.6}$$

is the quantity which decides about the asymptotic variance of the compromise estimator. Note that we made use of the symmetry of the sampling distribution G in the derivation of (3.6). If (3.6) is positive, the compromise estimators will behave asymptotically like the arithmetic mean, otherwise like the Huber estimator. All that remains to be considered, therefore, is the case where (3.6) is positive (or zero) because in the other case the assertion of the proposition is trivial.

First, note that (3.6) can only be positive if G has finite variance. Using our assumptions about $G = (1 - \varepsilon)\Phi + \varepsilon H$ stated in the proposition, (3.6) can be written as

$$
\begin{aligned}
& -\log(1-\varepsilon) + 2(1-\varepsilon)\int_k^\infty \left(k|x| - \frac{k^2}{2} - \frac{x^2}{2} \right) \phi(x)\,dx \\
& \quad + \varepsilon \int \left(k|x| - \frac{k^2}{2} - \frac{x^2}{2} \right) dH(x) \\
& = -\log(1-\varepsilon) - (1-\varepsilon)\int_k^\infty (x-k)^2\, d\Phi(x) \qquad (3.7) \\
& \quad - \varepsilon \int_k^\infty (x-k)^2\, dH(x) \\
& > -\log(1-\varepsilon) - (1-\varepsilon)\left[-k\phi(k) + \Phi(-k)(1+k^2)\right] \\
& \quad - \frac{\varepsilon}{2}k^2 - \frac{\varepsilon}{2}\sigma_H^2,
\end{aligned}
$$

where $\sigma_H^2 = \int x^2\,dH(x)$ is the variance of the contaminating distribution.

A comparison of the asymptotic variances of the sample mean and Huber's estimator is not hard. We have

$$\text{as. var}_G(\text{sample mean}) = (1-\varepsilon) + \varepsilon\sigma_H^2, \tag{3.8}$$

$$
\begin{aligned}
\text{as. var}_G(\text{Huber estimator}) &= \frac{\int (\psi_k(x))^2\, dG(x)}{\left(\int \psi_k'(x)\, dG(x)\right)^2} \\
&= \frac{\int_{-k}^{k} x^2\, dG(x) + 2\int_k^\infty k^2\, dG(x)}{\left(\int \psi_k'(x)\, dG(x)\right)^2} \qquad (3.9) \\
&= \frac{(1-\varepsilon)\int_{-k}^{k} x^2\, d\Phi(x) + k^2\varepsilon + 2k^2(1-\varepsilon)\Phi(-k)}{(1-\varepsilon)^2(\Phi(k) - \Phi(-k))^2}.
\end{aligned}
$$

In this last formula we have again used all our knowledge about the sampling situation $G(\ )$.

What remains to be shown is

$$\text{nonnegativeness in } (3.7) \to 3.8 \le (3.9).$$

But

$$
\begin{aligned}
(3.7) \ge 0 \to \varepsilon\sigma_H^2 \le & -2\log(1-\varepsilon) + 2(1-\varepsilon)k\phi(k) \\
& -2(1-\varepsilon)\Phi(-k)(1+k^2) - \varepsilon k^2
\end{aligned}
$$

and, therefore, we have

$$
\begin{aligned}
(3.8) = (1-\varepsilon) + \varepsilon\sigma_H^2 \le (1-\varepsilon) + \log\left(\frac{1}{1-\varepsilon}\right)^2 + 2(1-\varepsilon)k\phi(k) \\
-2(1-\varepsilon)\Phi(-k)(1+k^2) - \varepsilon k^2.
\end{aligned}
$$

Using the equation linking ε and k,

$$\frac{2\phi(k)}{k} - 2\Phi(-k) = \frac{\varepsilon}{(1-\varepsilon)},$$

we can simplify and get

$$(3.8) \le \log\left(\frac{1}{1-\varepsilon}\right)^2 + (1-\varepsilon)k^2\frac{\varepsilon}{1-\varepsilon} + (1-\varepsilon) - 2(1-\varepsilon)\Phi(-k) - \varepsilon k^2,$$

$$(3.8) \le \log\left(\frac{1}{1-\varepsilon}\right)^2 + (1-\varepsilon)(1 - 2\Phi(-k)).$$

Along the same line of thought, we can simplify (3.9) to get

$$(3.9) = \frac{1}{(1-\varepsilon)(1-2\Phi(-k))}.$$

Putting all these results together, we finally have

$$\begin{aligned}(3.9) &> 1 + (1-\varepsilon)\frac{2\phi(k)}{k}\\ &\ge (1-\varepsilon)(1-2\Phi(-k)) + (1-\varepsilon)\frac{4\phi(k)}{k}\\ &\ge (1-\varepsilon)(1-2\Phi(-k)) + \log\left(\frac{1}{1-\varepsilon}\right)^2 \ge (3.8)\end{aligned}$$

if only we show that

$$(1-\varepsilon)\frac{4\phi(k)}{k} \ge \log\left(\frac{1}{1-\varepsilon}\right)^2 \tag{3.10}$$

holds. This last inequality is only true for ε small enough, e.g., $\varepsilon \le 0.5$. For such ε values we have

$$\log\left(\frac{1}{1-\varepsilon}\right)^2 \le 3\varepsilon, \qquad 0 \le \varepsilon \le 0.5,$$

and (3.10) is therefore proved if we show that

$$(1-\varepsilon)\frac{4\phi(k)}{k} \ge 3\varepsilon$$

$$\begin{aligned}&\leftrightarrow \frac{4\phi(k)}{3k} \ge \frac{\varepsilon}{1-\varepsilon} = 2\frac{\phi(k)}{k} - 2\Phi(-k)\\ &\leftrightarrow 2k\Phi(-k) \ge \tfrac{2}{3}\phi(k)\\ &\leftrightarrow 3k\Phi(-k) \ge \phi(k) \quad \text{for } k \in [0.436, \infty).\end{aligned} \tag{3.11}$$

This last inequality (3.11), which is equivalent to (3.10), does indeed hold and is left for the reader to check.

Proposition 3.1 is now proved for all the cases where (3.6) is strictly positive. Some care is needed if (3.6) is zero. Then the compromise estimator is asymptotically a convex linear combination of T_{F_1} and T_{F_2}, but since the asymptotic variance of T_{F_1} is lower than the asymptotic variance of T_{F_2}, the compromise estimator will have an asymptotic variance below the asymptotic variance of T_{F_2}.

REMARKS. (1) We have identified a class of sampling situations G, namely those where (3.6) is positive, for which the mean is a more efficient estimator than Huber's minimax estimator. It would be of interest to show how big this class is and also to check whether it contains all sampling situations for which the sample mean is asymptotically better than Huber's minimax estimator.

4. Discussion. This paper deals with estimators which compromise between different "shapes." This idea, as we have seen, produces robust estimators. If we compromise between the Gaussian and Huber's least favorable distribution, we have a family of estimators (for different values of α) which dominate Huber's minimax M-estimator asymptotically.

Several points need to be clarified, however. The idea of compromising is different from the usual asymptotic robustness theory as developed by Huber (1964) and (1981). There, the compromising takes place in a neighborhood of the "central" model, whereas in our approach the different shapes need not be close together. A neighborhood model is in fact only a first step toward robust/resistant techniques for small sample sizes. For samples of size 5, we would advise to compromise between the Gaussian and something like the slash (= distribution of a ratio of a Gaussian over an independent uniform) rather than using the only moderately tailed least favorable distribution.

The intention of this paper is *not* to show that we should use a compromise between the Gaussian and the least favorable distribution, but rather to let people know of the merits of compromise estimators in a language which many statisticians are used to, namely asymptotics.

Results found through small sample experiments are of greater importance. It is clear, for example, that the situations (or shapes) we compromise ought to change with the sample size. The amount of "information" in the sample grows with the sample size. Not only are we able to estimate "parameters" with less variability, we also gain insight into the underlying shape. Compromise estimators use this knowledge in an optimal way and with our choice of the shapes we can fine-tune the procedure. Important choices have to be made in that respect and more (probably experimental) research for small sample sizes is needed. Subject-matter knowledge might prove useful in this connection.

The extension of Pitman's ideas to more than one shape provides us with a tool to find meaningful small sample methods of the robust/resistant kind. In order to make the asymptotics simple, we did not deal with the scale parameter. In actual applications, the inclusion of this additional parameter is, however, no problem [see Bell and Morgenthaler (1981) for an example].

Acknowledgment. Many thanks to the referees for a careful reading of the manuscript, and for pointing out several mistakes.

REFERENCES

BELL-KRYSTINIK, K. K. and MORGENTHALER, S. (1981). Comparison of the bioptimal curve with curves for two robust estimates. Technical Report 195, Series 2, Department of Statistics, Princeton University.

DEBRUIJN, N. G. (1981). *Asymptotic Methods in Analysis*. Dover, New York.

EASTON, G. (1984). Personal communication.

FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic, New York.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1**. *Univ. California Press.*

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

JOHNS, M. V. (1979). Robust Pitman-like estimators. In *Robustness in Statistics*. (R. L. Launer and G. N. Wilkinson, eds.). Academic, New York.

KLAASSEN, C. A. J. (1981). *Statistical Performance of Location Estimators*. Mathematical Centre Tracts, No. 133, Amsterdam.

KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22** 79–86.

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

PORT, S. C. and STONE, C. J. (1974). Fisher information and the Pitman estimator of a location parameter. *Ann. Statist.* **2** 225–247.

PREGIBON, D. and TUKEY, J. W. (1981). Assessing the behavior of robust estimates of location in small samples: introduction to configural polysampling. Technical Report 185, Series 2, Department of Statistics, Princeton University.

DEPARTMENT OF STATISTICS
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06520