The Annals of Mathematical Statistics
1969, Vol. 40, No. 1, 24–39

ON ROBUST LINEAR ESTIMATORS

BY JOSEPH L. GASTWIRTH[1] AND HERMAN RUBIN[2]

The Johns Hopkins University and Purdue University

1. Introduction. The problem of finding robust estimators for the location parameter of symmetric unimodal distributions has been the subject of much recent research (e.g. [2], [3], [6], [8], [9], [13], [15]). This paper is concerned with finding robust estimators which are linear functions of the ordered observations. Thus, the robust estimators proposed by Hodges and Lehmann [8] and Huber [9] are not considered in detail. The spirit of the present work is similar to a previous paper of one of the authors [7] and is also related to the fundamental work of Tukey [13], [14].

We assume that the density function of the population sampled is a member of a class $\mathfrak{F}$ of densities. For every member f_γ (γ runs through an index set Γ), there is an asymptotically efficient estimator S_γ which is a linear combination of the ordered observations [1], [5], [10]. The asymptotic efficiency of any estimator D for samples from the density f_γ is the reciprocal of the ratio of the asymptotic variance of D to the asymptotic variance of S_γ. Throughout the paper the asymptotic variance of an estimator will mean the variance of its asymptotic normal distribution.

An estimator will be called a maximin efficient estimator within a class C of estimators for a family of densities if it maximizes the minimum asymptotic efficiency over the family $\mathfrak{F}$. In Section 2, we demonstrate that within a large class C of linear estimators, there is a unique maximin efficient linear estimator for general families of densities. Under somewhat more restrictive conditions on the family $\mathfrak{F}$ of possible densities we show that within the class C of linear estimators the Bayes estimators are the minimal complete class. These results are asymptotic generalizations of the work of Birnbaum and Laska [3].

In Section 3 we discuss, in detail, the special case when $\mathfrak{F}$ contains the logistic and double-exponential distributions. The maximin efficient linear estimator (m.e.l.e.) is found and is compared to the best convex combination of the individual optimum linear estimator and also to a Hodges-Lehmann type estimator based on the corresponding maximin rank test [7].

In general, the m.e.l.e. for specific families of densities is quite difficult to compute. It seems appropriate, therefore, to look for a maximin efficient estimator in smaller classes of linear estimators which are easy to use. Two such

Received 1 June 1967; revised 15 May 1968.

[1] Research supported by the National Science Foundation Research Grant No. GP 6225 awarded to the Department of Statistics, The Johns Hopkins University. This paper in whole or in part may be reproduced for any purpose of the University State Government.

[2] Research supported by the Office of Naval Research under contract NONR 2587 (02) awarded to the Department of Statistics, Michigan State University. This paper in whole or in part may be reproduced for any purpose of the United States Government.

families of linear estimators are the trimmed means and linear combinations of a few (r) sample percentiles. Under suitable regularity conditions, a maximin efficient estimator for each of these classes exists (see Section 4). Some numerical examples are also given. A more detailed numerical study of some robust estimators is due to Crow and Siddiqui [6].

2. Maximin efficient linear estimators and admissible linear estimators. The purpose of this section is to prove the existence of an asymptotically maximin efficient linear estimator (a.m.e.l.e.) and determine the family of all admissible linear estimators of a parameter θ when the observations come from one member of a family $\mathfrak{F} = \{f_\gamma\}$ of density functions. In practice, θ is usually a location or scale parameter. We shall restrict ourselves to estimators which are linear in the order statistics, i.e., they are of the form

$$T_n = n^{-1} \sum_{i=1}^{n} w_i x_{(i)}\ , \tag{2.1}$$

where $x_{(1)} \leqq x_{(2)} \leqq \cdots \leqq x_{(n)}$ are the ordered observations and the weights w_i are determined from a measure μ on $(0, 1)$, of variation 1, by setting

$$w_i = n\mu((i-1)n^{-1}, in^{-1}).$$

Bennett [1] and Jung [10] considered linear estimators to be specified by a function w on $(0, 1)$ such that $\int_0^1 w(u)\, du = 1$ and set $w_i = w(i/n + 1)$. When μ is sufficiently smooth both methods lead to asymptotically equivalent linear estimators, however, the present method is more general as it allows us to consider statistics which are linear combinations of a few sample percentiles.

For any cdf F with density f, define the function h by

$$h(u) = f[F^{-1}(u)]. \tag{2.3}$$

When μ is composed of an absolutely continuous part and a finite number of atoms, then under mild restrictions [5], the estimator determined by μ has an asymptotically normal distribution with variance given by

$$n^{-1} \int_0^1 \int_0^1 [(\min(u, v) - uv)/h(u)h(v)]\ d\mu(u)\ d\mu(v). \tag{2.4}$$

This formula holds for more complicated measures μ but the exact conditions are under investigation. Of course, the reciprocal of the asymptotic efficiency of the estimator μ to the optimum estimator is the ratio of (2.4) to the Cramér-Rao lower bound.

In order to apply the results of the theory of games to our problem we must define precisely the strategy spaces involved. Nature's action space is the set of all densities $\{f_\gamma\}$, $\gamma \varepsilon \Gamma$ and each strategy can be thought of as a probability measure $\zeta(\gamma)$ on Γ. If c_γ denotes the Cramér-Rao lower bound for samples from f_γ, then the risk function for the efficiency problem when nature uses ζ and the statistician uses the estimator determined by μ is

$$R(\zeta, \mu) = \int \int \int [(\min(u, v) - uv)/c_\gamma h_\gamma(u)h_\gamma(v)]\, d\mu(u)\ d\mu(v)\ d\zeta(\gamma). \tag{2.5}$$

If this risk function is just the variance of the estimator the c_γ's are omitted in expression (2.5). In both problems the statistician desires to minimize the risk.

The class C of estimators that the statistician will be allowed to use will be those linear estimators which are specified by measures μ of variation one, satisfying

$$\int_0^1 |d\mu(u)|/q(u) \leqq A, \tag{2.6}$$

where q is a positive, continuous function on (0, 1) vanishing at 0 and 1 and A is an arbitrary constant. Unfortunately, the authors have found this condition necessary to ensure that the estimator μ does not place too much weight on the extreme observations. The reason that we require $\int |d\mu| < \infty$ in addition to the condition $\int d\mu = 1$, is due to the existence of density functions whose optimal linear estimators are specified by measures whose total variation is infinite. For example, the best linear estimator for a sample from the density $f(x) = ce^{|x|^{\frac{3}{2}}}$ is given by a measure placing infinite mass at $\frac{1}{2}$ and an infinite negative measure elsewhere. Once q is chosen the set of measures corresponding to the linear estimator is given the weak* topology with respect $q^{-1}\, d\mu$. A sequence $\{\mu_\alpha\}$ of measures converges to μ in this topology, if for every continuous function $g(u)$ such that $g(0) = g(1) = 0$,

$$\int_0^1 gq^{-1}\, d\mu_\alpha \to \int_0^1 gq^{-1}\, d\mu \tag{2.7}$$

as $\alpha \to \infty$.

It is convenient to discuss the case when the risk function is the asymptotic variance of the estimator. The main existence theorem is the following:

THEOREM 2.1. *Let C be the class of measures satisfying Condition* (2.6) *and suppose that C is not empty. For any family $\mathfrak{F} = \{f_\gamma\}$ of continuous densities there is a linear estimator in C which minimizes the maximum asymptotic variance over all the densities in $\mathfrak{F}$.*

PROOF. Since lower semi-continuity and convexity properties of functions are preserved when they are mixed by a probability distribution it suffices to show that each of the functions

$$R(f_\gamma, \mu) = \int_0^1 \int_0^1 [(\min(u, v) - uv)/h_\gamma(u)h_\gamma(v)]\, d\mu(u)\, d\mu(v) \tag{2.8}$$

is convex and lower semi-continuous in μ. Convexity follows from the fact that the variance is a convex function. The kernel $[\min(u, v) - uv]/h_\gamma(u)h_\gamma(v)$ is continuous and positive in the interval (0, 1). If we write $d\mu = q\, d\nu$, then weak* convergence of the measures (μ) in our topology corresponds to the ordinary weak* convergence of $d\nu$. Since

$$\begin{aligned} \liminf_\gamma \int_0^1 \int_0^1 q(u)[(\min(u,v) - uv)/h_\gamma(u)h_\gamma(v)]q(v)\, d\nu_\alpha(u)\, d\nu_\alpha(v) \\ \geqq \int_0^1 \int_0^1 q(u)[(\min(u, v) - uv)/h_\gamma(u)h_\gamma(v)]q(u)\, d\nu(u)\, d\nu(v) \end{aligned} \tag{2.9}$$

as $d\nu_\alpha \to d\nu$ in the weak* topology implies $d\nu_\alpha \times d\nu_\alpha \to d\nu \times d\nu$, the function $R(f_\gamma, \mu)$ is lower-semicontinuous in μ. The existence of a linear estimator achieving the minimax asymptotic variance follows by applying Sion's theorem [11].

If all the densities in $\mathfrak{F}$ are normalized so that they have Fisher information one, then Theorem 2.1 implies that an asymptotically maximin efficient linear estimator (a.m.e.l.e.) exists. In order that it be non-trivial it is necessary to assume that the non-normalized densities have Fisher informations which are bounded away from zero. Rather than state a general theorem we present, in detail, an existence theorem for an a.m.e.l.e. of the location parameter from families of symmetric density functions. Precisely, we have

THEOREM 2.2. *Let $\mathfrak{F}$ be a family of symmetric density functions whose Fisher informations are bounded away from ∞ and such that $f_\gamma(0) \geqq \epsilon$ for some $\epsilon > 0$. Then there is a class C of linear estimators satisfying condition* (2.6) *and an* a.m.e.l.e. *within the class C exists.*

PROOF. The conditions of the theorem imply that the median is an estimator with non-zero relative efficiency for all members of $\mathfrak{F}$. Let μ denote the measure generating the median. Since $\int d\mu = 1$, there is a continuous positive function g such that $\int g\, d\mu < \infty$ and $g(0) = g(1) = \infty$. Taking $q = g^{-1}$, a non-empty class C of estimators satisfying Condition (2.6) exists. The existence of an a.m.e.l.e. within the class C follows by applying Theorem 2.1 to the densities of $\mathfrak{F}$ normalized to have information one.

Theorem 2.2 does not depend critically upon the assumption that the efficiency of the median relative to the best estimator for each member of $\mathfrak{F}$ is bounded away from 0. The existence of any estimator with this property will suffice for our purposes. The choice of the function q also is quite arbitrary. In general, q should be chosen so that the class C of linear estimators is as large as possible.

So far we have shown that the game has a solution and that the statistician has a strategy attaining the value of the game. In order to show that nature also has an optimum strategy (rather than just an ϵ-optimal one) we introduce the following assumption on the family $\mathfrak{F} = \{f_\gamma\}$ or equivalently on the functions $\{h_\gamma(u)\}$.

ASSUMPTION 1. The informations I_γ corresponding to the densities f_γ are bounded away from 0 and ∞, i.e., constants I_1 and I_2 exist such that

$$0 < I_1 < I_\gamma < I_2 \quad \text{for all} \quad \gamma. \tag{2.10}$$

ASSUMPTION 2. For all u in the open interval $(0, 1)$ $h_\gamma(u) > 0$, and the function h^* defined by

$$h^*(u) = \inf_\gamma h_\gamma(u) > 0. \tag{2.11}$$

From Assumptions 1 and 2, we can derive the following:

LEMMA 2.1. *The functions h_γ are equi-continuous on the interval* $(0, 1)$.

PROOF. For any function h_γ and any partition $0 = u_0 < u_1 < \cdots < u_n < u_{n+1} = 1$ of the interval $(0, 1)$,

$$I_\gamma \geqq \textstyle\sum_1^{n+1} [h_\gamma(u_i) - h_\gamma(u_{i-1})]^2(u_i - u_{i-1})^{-1}. \tag{2.12}$$

Thus,

$$|h_\gamma(u_i) - h_\gamma(u_{i-1})| \leqq I_2^{\frac{1}{2}}(u_i - u_{i-1})^{\frac{1}{2}}. \tag{2.13}$$

Since the right side of inequality (2.13) is independent of γ the functions h_γ are equi-continuous. It should be noted that if a density f has finite information, then $\lim_{u\to 0 \text{ or } 1} h(u) = 0$.

We topologize the set of functions h_γ, nature's strategy space, by the compact open topology, i.e., $h_\gamma \to h$ if the convergence is uniform on all compact subsets of (0, 1). This is equivalent to the topology of the pointwise convergence of the h functions. The last condition we require on the h functions is the following:

ASSUMPTION 3. The set $\{h_\gamma\}$ of nature's strategies is compact in the compact-open topology in the space of continuous functions on (0, 1).

The natural choice for the function q is

$$q(u) = [u(1-u)]^{\frac{1}{2}}/h^*(u). \tag{2.14}$$

Clearly, $q(u)$ is continuous and positive in (0, 1). The following lemma shows that $q(u) \to \infty$ as $u \to 0$ or 1. Specifically, we have

LEMMA 2.2. *For any family* $\mathfrak{F}$ *of densities whose* h *functions satisfy Assumptions* 1 *and* 2,

$$\lim_{u\to 0 \text{ or } 1} [u(1-u)]^{\frac{1}{2}}/h^*(u) = \infty. \tag{2.15}$$

PROOF. Since $h^*(u) \leqq h_\gamma(u)$, the lemma will be proved if (2.15) holds for any h_γ (written as h for convenience). In order to show that (2.15) holds as $u \to 0$ suppose to the contrary that $u^{\frac{1}{2}}/h(u)$ does not approach ∞ as u tends to 0. Then there is a decreasing sequence u_i converging to 0 such that

$$h(u_{2i}) > cu_{2i}^{\frac{1}{2}}, \tag{2.16}$$

where c is a positive constant. Since $\lim_{u\to 0 \text{ or } 1} h(u) = 0$, a sequence u_i satisfying (2.16) can be chosen that also satisfies

$$h(u_{2i+1}) < h(u_{2i})/2. \tag{2.17}$$

The information in the fractiles corresponding to the partition u_i of the interval (0, 1) is

$$\sum [h(u_{2i}) - h(u_{2i+1})]^2(u_{2i} - u_{2i+1})^{-1} \geqq \tfrac{1}{4} \sum c^2 u_{2i} u_{2i}^{-1} = \infty. \tag{2.18}$$

This contradicts the first assumption, i.e. all the informations are uniformly bounded by $I_2(<\infty)$. The proof for the case when $u \to 1$ is similar.

If μ is any measure of variation one, then Lemma 2.2 implies that $\int q^{-1}\, d\mu$ is also finite. Then the constant A in Condition (2.6) can be chosen so large that the class C of estimators, where q is given by (2.14) is non empty.

The following theorems tell us under what conditions nature has as optimum strategy when the risk is asymptotic variance of the estimator.

THEOREM 2.3. *If the family* f_γ *of densities satisfies Assumptions* 1, 2 *and* 3 *and if the set* C *of linear estimators satisfying Condition* (2.6), *where* q *is given by* (2.14), *contains a member with finite asymptotic variance for all densities* f_γ, *then the game has a value and nature has a strategy*, ζ, *which attains the maximin asymptotic variance.*

PROOF. Letting $\bar{h}_\gamma = h_\gamma/h^*$, the variance of the estimator based on the measure μ on data from the density f_γ is

$$R(h_\gamma, \mu) = \int_0^1 \int_0^1 f(u, v)(d\mu(u)/q(u))(d\mu(v)/q(v)) \tag{2.19}$$

where $f(u, v) = [\min(u, v) - uv][u(1-u)v(1-v)]^{-\frac{1}{2}}[\bar{h}_\gamma(u)\bar{h}_\gamma(v)]^{-1}$.
The function $\int R(h_{\gamma,\mu})\, d\xi(\gamma)$ is convex in the statistician's strategy space (measures μ) and is linear in nature strategies (the measures ζ over the functions h_γ).
Since nature's strategy space is compact the results follow from the Kneser-Fan minimax theorem [11] if $R(h_\gamma, \mu)$ is continuous in h_γ (or equivalently $\bar{h}_\gamma$) for fixed μ. The function

$$(\min(u, v) - uv)[u(1-u)]^{-\frac{1}{2}}[v(1-v)]^{-\frac{1}{2}} \tag{2.20}$$

is a bounded continuous function on $(0, 1) \times (0, 1)$ and $1/h_\gamma(u) \leqq 1$; therefore, the integrand in expression (2.19) is a bounded function on $(0, 1) \times (0, 1)$. Since $q^{-1}\, d\mu$ is a measure with finite total variation, for any $\epsilon > 0$, there is a compact set $K = K_1 \times K_1$ in $(0, 1) \times (0, 1)$ such that

$$\int \int_{(\overline{K_1 \times K_1})} f(u, v)(d\mu(u)/q(u))(d\mu(v)/q(v)) \tag{2.21}$$

where $(\overline{K_1 \times K_1})$ is the complement of $K_1 \times K_1$, and $f(u, v) = [\min(u, v) - uv] \cdot [u(1-u)]^{-\frac{1}{2}}[v(1-v)]^{-\frac{1}{2}}[\bar{h}_\gamma(u)\bar{h}_\gamma(v)]^{-1}$. Since the topology on the functions h_γ is the compact open topology, the functions in a neighborhood of $h_{\gamma(u)}h_{\gamma(v)}$ are uniformly close to $h_\gamma(u)h_\gamma(v)$ on compact sets. Thus, the integral

$$\int \int_{K_1 \times K_1} f(u, v)(d\mu(u)/q(u))(d\mu(v)/q(v)) \tag{2.22}$$

is a continuous function of $\bar{h}_\gamma$. Since the risk of function (2.19) is the sum of (2.21) and (2.22) it is continuous in h_γ.

REMARK. In the uninteresting case where the class C of linear estimators does not contain a member with finite asymptotic variance for all members of $\mathfrak{F}$ the value of the game is ∞ and nature has an optimum strategy.

So far we have not used the hypothesis that the informations I_γ are bounded away from ∞. This condition is required for the existence of an a.m.e.l.e. The result is a rewording of Theorem 2.3, where the conclusion now is that nature has a strategy which attains the maximin risk. (The risk is the reciprocal of the asymptotic relative efficiency). The proof consists of normalizing the densities to have Fisher information 1 and applying Theorem 2.3.

The remainder of this section will be devoted to showing that the Bayes' strategies form a minimal complete class. It should be mentioned that Birnbaum and Laska [3] were the first to discuss complete class theorems for robust estimators. They dealt with finite sample sizes and their loss function was the variance of the estimator. In the asymptotic case, we show that the Bayes' strategies for the problem form the complete class of linear estimators. Formally, we have

THEOREM 2.4. *If the assumptions of Theorem 2.3 hold, then the Bayes strategies for the statistician form a minimal complete class where the risk function is either* *asymptotic variance of a linear estimator or the reciprocal of the efficiency of a linear estimator to the best possible one.*

PROOF. Since the proofs for both risk functions are similar, consider the case where the risk is asymptotic variance. Since the risk function (2.19) is strictly convex in μ a Bayes strategy is unique and thus admissible. Suppose μ_0 is an admissible strategy. Consider a new game with risk function

$$R^*(h, \mu) = R(h, \mu) - R(h, \mu_0),$$

where $R(h, \mu)$ is given by (2.19). Since the new risk function $R^*(h, \mu_0)$ differs from $R(h, \mu)$ by a constant (depending on h) under the conditions of the theorem the new game satisfies the hypothesis of Theorem 2.1 and 2.3. Hence, the new game has a value and both players have good strategies. By Theorem 5.6.5 of Blackwell and Girshick [4] the Bayes strategies are complete for the first game.

REMARK. Under the conditions of Theorem 2.4 the class of Bayes strategies for the statistician is the same for either loss function. However, any particular strategy μ will usually be Bayes against different strategies of nature for each of the two risk functions.

APPENDIX TO SECTION 2. The conditions of Theorems 2.1 and 2.3 were quite complicated because both strategy spaces were infinite. When nature's strategy space consists of a finite number of density functions with finite Fisher informations, the statement of Theorem 2.1 becomes much simpler. Precisely, we have

THEOREM 2.1*. *Let C be the class of measures μ such that $\int d\mu = 1$ and suppose that there is a member of C with finite asymptotic variance for all the densities $\mathfrak{F} = \{f_1, \cdots, f_n\}$ in nature's strategy space. Then the game has a value and both players have good strategies*, i.e., *there is a linear estimator which minimizes the maximum asymptotic variance over the family $\mathfrak{F}$ and there is a probability measure $\{p_1, \cdots, p_n\}$ on the family $\mathfrak{F}$ so that nature can achieve the maximin asymptotic variance.*

The proof follows from the following minimax theorem of Stein [12]. We have

THEOREM (Stein). *Let X be a finite set, Y an arbitrary set and K a bounded real valued on $X \times Y$. If K is pseudo-convex in the second argument* (y), *then*

$$\sup_\zeta \inf_Y K^*(\zeta, y) = \inf_Y \sup_\zeta K^*(\zeta, y), \tag{2.1*}$$

where ζ is a set in the space of probability measures on X and $K^*(\zeta, y) = \sum_{x \in X} K(x, y)\zeta_x$. Furthermore, the infimum on the right-hand side is attained.

To prove Theorem 2.1* from Stein's theorem note that the existence of an estimator ν in C with finite asymptotic variance for all members of $\mathfrak{F}$ allows us to consider only those members (μ) of C satisfying

$$\sup_\gamma R(f_\gamma, \mu) \leqq A = \sup_\gamma R(f_\gamma, \nu). \tag{2.2*}$$

Then the risk function is a bounded real valued function on $X \times Y$. Since the space C of measures is convex and the risk function is convex the set of estimators satisfying (2.2*) is convex and R is pseudo-convex in y. Thus, Stein's theorem applies. Moreover, the proof that the Bayes estimators are a complete class when $\mathfrak{F}$ is finite is almost identical to the previous proof and will be omitted.

3. An example. Although an a.m.e.l.e. exists for most families $\mathfrak{F}$ it is usually difficult to obtain explicitly. In this section we shall obtain the a.m.e.l.e. when nature presents the statistician with either double-exponential or logistic data.

The a.m.e.l.e. is found in the following manner. For each of nature's strategies $(\lambda_1, \lambda_2)(\lambda_1 + \lambda_2 = 1$, λ_1 is the probability of nature using logistic data) we obtain the statisticians best strategy, i.e. the linear estimator maximizing the efficiency against the Bayes mixture of the two densities. Varying natures strategies in order to minimize this efficiency yields the a.m.e.l.e. Thus, given λ_1 and λ_2 we must find a measure μ minimizing

$$(3.1)\quad \int_0^1 \int_0^1 [\min (u, v) - uv][3\lambda_1\varphi_1(u)\varphi_1(v) + \lambda_2\varphi_2(u)\varphi_2(v)]\, d\mu(u)\, d\mu(v),$$

where $\varphi_i(u) = 1/h_i(u)$ and $h_i(u)$ is defined by (2.3). The factor 3 enters into expression (3.1) because the Cramér-Rao lower bound for the logistic cdf is $\frac{1}{3}n$. Since we are dealing with symmetric densities it suffices to determine μ on $[\frac{1}{2}, 1]$. Thus, we must find a measure μ^* such that $\int_{\frac{1}{2}}^1 d\mu^* = \frac{1}{2}$ which minimizes

$$(3.2)\quad 2\int_{\frac{1}{2}}^1 \int_{\frac{1}{2}}^1 [1 - \max (u, v)][3\lambda_1\varphi_1(u)\varphi_1(v) + \lambda_2\varphi_2(u)\varphi_2(v)]\, d\mu(u)\, d\mu(v).$$

By the calculus of variations it follows that μ^* yields the minimum of expression (3.2) if

$$(3.3)\quad \int_{\frac{1}{2}}^1 [1 - \max (u, v)](3\lambda_1\varphi_1(u)\varphi_1(v) + \lambda_2\varphi_2(u)\varphi_2(v))\, d\mu^*(u) = q$$

where q is a constant. In our example $\varphi_1(u) = 1/u(1 - u)$ and $\varphi_2(u) = (1 - u)^{-1}$. Without loss of generality, we may assume that $\lambda_1 = 1/1 + 3\beta^2$ and $\lambda_2 = \beta^2/1 + 3\beta^2$. For simplicity, we shall omit the factor $(1 + 3\beta^2)^{-1}$ in the computation of μ^*. Finally, it will be convenient to write

$$(3.4)\qquad d\mu^*(u) = u(1 - u)\, d\mu(u).$$

Thus the problem is to find a measure

$$(3.5)\qquad d\nu(u) = c\delta(\tfrac{1}{2}) + f(u)\, du,$$

where $\delta(\frac{1}{2})$ is the measure placing mass 1 at $\frac{1}{2}$ and 0 elsewhere, satisfying (3.3). Differentiating the resulting expression and solving the differential equation obtained for the absolutely continuous part $f(v)$ of the measure ν yields

$$(3.6)\qquad f(v) = K(\beta^2 + \nu^2)^{-2}.$$

Substituting $f(v)$ into the differentiated version of (3.3) allows us to evaluate c and we obtain K from the fact that μ^* is a measure of variation one-half on $(\frac{1}{2}, 1]$. Finally, we obtain that the measure generating the Bayes linear estimator against nature's strategy $(1, \beta)$ is

$$(3.7) \qquad \mu^*(v) = Kv(1-v)/(\beta^2+v^2)^2, \qquad \text{if } 0 < v < 1, \quad v \neq \tfrac{1}{2}$$
$$= \tfrac{1}{2}c\delta(\tfrac{1}{2}), \qquad v = \tfrac{1}{2},$$

where

$$(3.8) \qquad K = [2s(1+\beta^2/4)]^{-1}, \qquad c = Ks/\beta^2$$

and

$$(3.9) \qquad s = (4\beta^2+1) - (2\beta)^{-1}\tan^{-1}(\beta/(2\beta^2+1)).$$

The relative risk is

$$q/(1+3\beta^2),$$

where

$$(3.10) \qquad q = \beta^2[4s(\beta^2+\tfrac{1}{4})]^{-1} + [4s(\beta^2+\tfrac{1}{4})]^{-1}.$$

By varying β we can obtain all of nature's possible strategies. The a.m.e.l.e. achieves a relative efficiency of about 92% when β is about .61. This compares to 94.3% for the Hodges-Lehmann type of estimate based on the maximin rank test [7].

It is interesting to compare the a.m.e.l.e. to a convex combination of the median (M is the a.b.l.u.e. for double exponential data) and the best estimator, L, for logistic data (it is based on the measure $d\mu(u) = 6u(1-u)\,du$). Using formula (2.4) the asymptotic variance (times n) of the best estimator for logistic data when the data is from a double exponential distribution can be shown to equal 1.425. The asymptotic variance (times n) of $pM + qL$ is $1 + .425\, q^2$ when the observations are from double exponential data and $1 + p^2/3$ for logistic data (scaled to have Fisher information 1). The variances will be equal when $p \sim .53$, $q \sim .47$, and the asymptotic efficiency of this estimator is 91.43% relative to M for double exponential data and to L for logistic data. In this example, the a.m.e.l.e. is only a slight improvement over the optimum convex combination of the best estimators. In general, this will not be true as the best estimator for one cdf may not even be consistent for another member of $\mathfrak{F}$.

We should like to mention that Yhap [16] has obtained a general formula for the Bayes' solution for a finite number of densities and has studied contaminated normal distributions in detail. As far as the authors know, there is no example of an estimator which has analytically been demonstrated to be an a.m.e. l.e. for an infinite family of possible densities.

4. Smaller classes of linear estimators. While a maximin efficient linear estimator exists for most families $\mathfrak{F}$ that are possible densities for a set of observations, we have seen that it is usually difficult to find explicitly and that its efficiency is not much larger than that of an estimator based on a few order statistics or a suitably chosen combination of the relevant best estimators. A practicing statistician might prefer to choose a class C of linear estimators that are easy to work with and use the maximin efficient estimator in C. In this section we show that maximin efficient linear estimators of the location parameter for families of symmetric densities exist within the class of trimmed means and within the class of estimators which are linear combination of k sample percentiles.

Let $x_{(1)} \leqq \cdots \leqq x_{(n)}$ be an ordered sample of size n from a symmetric cdf $F(x)$. Since the estimators we discuss are translation invariant we may assume that the location parameter θ equals 0. The α-trimmed mean is defined as

$$T_\alpha = [n(1-2\alpha)]^{-1} \textstyle\sum_{i=r}^{n-r+1} x_{(i)}\,, \tag{4.1}$$

where $r = [\alpha n] + 1$. It is known [2] that the trimmed mean has an asymptotic variance given by

$$v_F(\alpha) = n^{-1}(1-2\alpha)^{-2}[2\alpha B^2 + \textstyle\int_{-B}^{B} t^2\, dF(t)], \tag{4.2}$$

where

$$B = F^{-1}(1-\alpha) \quad \text{and} \quad -B = F^{-1}(\alpha).$$

We now show that under rather weak conditions on the family $\mathfrak{F}$ a maximin efficient trimmed mean exists. Specifically, we have

THEOREM 4.1. *Suppose that all* cdf's *in a family $\mathfrak{F}$ of symmetric* cdf's *have finite Fisher information, satisfy the conditions for the trimmed mean to have asymptotic variance given by* (4.2), *have a Dini derivative at* 0 (*the location parameter*) *and are strictly monotonic for all x in* $(F^{-1}(0, 1))$. *Then, amongst all trimmed means, a maximin efficient one exists.*

PROOF. For each $F \varepsilon \mathfrak{F}$, $v_F(\alpha)$ is a continuous function of α for $0 \leqq \alpha \leqq \frac{1}{2}$. The last assumption is required since it eliminates the possibility of the density being zero in a one-sided interval about $F^{-1}(\alpha)$ or $F^{-1}(1-\alpha)$ which would make $v_F(\alpha)$ discontinuous. When α approaches 0 $v_F(\alpha)$ approaches either $\int_{-\infty}^{\infty} t^2\, dF(t)$ or infinity continuously. When α tends to $\frac{1}{2}$,

$$(\tfrac{1}{2} - \alpha)B^{-1} = \lim_{B\to 0}\, [F(B) - F(0)]B^{-1} = F'(0) \tag{4.3}$$

so that

$$V_F(\alpha) = \tfrac{1}{4}\{F'(0)\}^2. \tag{4.4}$$

Thus, the function

$$V(\alpha) = \sup_F V_F(\alpha)/c_F\,, \tag{4.5}$$

where c_F is the variance of the best estimator for samples from $F(x)$, is a lower semi-continuous function of α and therefore attains a minimum.

REMARKS. If we let c_F be an arbitrary number so that expression (4.5) is a risk, then we can drop the condition that all members of $\mathfrak{F}$ have finite information and also allow $F(x)$ to have a jump at 0 or a Dini derivative there. The proof proceeds in the same manner as before. The extension is of interest when c_F is a scaling parameter so that our objective is to minimize the maximum asymptotic variance of a trimmed mean when all the populations have equal scale. It should be noted that all symmetric unimodal cdf's either have a jump at 0 or a Dini derivative there.

A theory of admissible trimmed means analogous to the theory of Section 2 can be developed, however, we shall just discuss some examples. When $\mathfrak{F}$ consists of the Cauchy and normal distributions the best choice of α is approximately .275 and the ARE of this trimmed mean to the best estimate is 81.78%. A numerical search in the neighborhood of $\alpha = .275$ yielded an efficiency of 81.801%.

Since the Hodges-Lehmann rank type estimator based on the maximin rank test [7] has ARE 82.8% it appears that this trimmed mean is very close to the a.m.e.l.e. Since this trimmed mean has ARE 83.5% compared to the median for double-exponential data, it is the maximin trimmed mean when the double-exponential distribution is added to $\mathfrak{F}$.

An interesting contrast to the situation of the Cauchy and normal distributions is provided by considering $\mathfrak{F}$ to consist of the logistic and double-exponential distributions. The maximin efficient trimmed mean has efficiency of about 82–83% relative to the best estimators. The maximin efficient linear estimator was about 92% as efficient as the b.l.u.e.'s.Thus, restricting oneself to a trimmed mean can lead to a considerable loss of efficiency. Intuitively, this is not surprising. When $\mathfrak{F}$ contains distributions which are not too disparate, the a.m.e.l.e. should be closely approximated by a convex combination of the relevant a.b.l. u.e.'s rather than by a trimmed mean. If $\mathfrak{F}$ contains widely different distributions, then the a.m.e.l.e. will not have components that are a.b.l.u.e.'s for some of the densities. Some sort of weighted average of the middle portion of the ordered sample would be the only reasonable linear estimator to use and the trimmed mean is an average of the middle portion of the data. As an illustration consider the case where $\mathfrak{F}$ contains the Cauchy and normal densities. The a.m. e.l.e. can't behave like the sample mean in the tail as the mean is not consistent for Cauchy data. In fact, the a.b.l.u.e. for Cauchy data puts negative weights in both extreme fourths of the sample so this estimator behaves poorly for normal data. Intuitively, it appears that the a.b.l.u.e. would put positive weight in the middle half of the ordered observations and essentially very little weight elsewhere. It is not too surprising, therefore, that the asymptotically optimum trimmed mean averages only the middle 45% of the data.

Recently [15] Tukey proposed the mid-mean, which averages the middle 50% of the data as a robust estimator for general use. He cited the numerical work of Crow and Siddiqui [6] in support of his suggestion. Since the mid-mean is quite close to the asymptotically optimum trimmed mean when sampling from either normal or Cauchy data, it seems to us that the practitioner would be overly cautious if he used it when he thought his data was nearly normal with some possible contamination.

The second class of robust linear estimators which are computationally convenient are linear combinations of a few sample percentiles. Let $0 = u_0 \leqq u_1 \leqq \cdots u_r \leqq u_{r+1} = 1$ be r-fractiles and let $n_i = [nu_i] + 1$ so that $x(n_i)$ is the sample $100r_i$th percentile. The estimators in this class C are representable in the form

$$S = \sum_{i=1}^{r} w_i x_{(n_i)}, \tag{4.6}$$

where $\sum_1^r w_i = 1$. For any continuous density function f (with corresponding h function given by (2.3)), the asymptotic variance of S is given by

$$\begin{aligned} V(h, \bar{u}, \bar{w}) &= \sum_{i=1}^{r} \sum_{j=1}^{r} [w_i w_j / h_i h_j][\min (u_i, u_j) - u_i u_j], \\ &= \sum_{i=1}^{r} (u_i - u_{i-1})[(1 - u_i)(1 - u_{i-1})]^{-1} \\ &\qquad \cdot \left(\sum_{j=1}^{r} w_j (1 - u_j) h_j^{-1}\right)^2 = \sum_{i=1}^{r} \Delta_i S_i^2, \end{aligned} \tag{4.7}$$

where

$$\Delta_i = (u_i - u_{i-1})[(1 - u_i)(1 - u_{i-1})]^{-1}, \qquad S_i = \sum_{j=i}^{r} w_j (1 - u_j) h_j^{-1}, \tag{4.8}$$

$u_0 = 0$, $h_i = h(u_i)$ and $h_0 = 0$.

We now state the main result of this section

THEOREM 4.2. *If the h functions corresponding to the members of a family $\mathfrak{F}$ of density functions satisfy*

$$h^*(u) = \inf h(u) > 0 \qquad \text{for} \quad u \,\varepsilon\, (0, 1), \tag{4.9}$$

$$[h_i - h_{i-1}]^2 \tau_i^{-1} \to 0, \tag{4.10}$$

where $\tau_i = u_i - u_{i-1} \to 0$ uniformly for all the h functions, and if there is one estimator in the class C of linear estimators using at most r percentiles with bounded variance for all members of $\mathfrak{F}$, then C contains an estimator which attains the minimax variance for samples from any member of the family $\mathfrak{F}$. Moreover, if the densities in $\mathfrak{F}$ satisfy Assumption one of Section 2, then C contains a maximin efficient estimator for $\mathfrak{F}$.

Before presenting the proof we shall show how the assumption that an estimator with uniformly bounded asymptotic variance for all members of $\mathfrak{F}$ exists implies that a robust linear estimator does not place all its weight on the extreme order statistics. We shall treat the smallest order statistics in detail. For any estimator based on r fractiles ($r \leqq m$), the maximum information possible to achieve is

$$I_r = \sum_{i=1}^{r} (h_i - h_{i-1})^2 \tau_i^{-1}. \tag{4.11}$$

Thus, the asymptotic variance of any estimator based on $u_1, \cdots, u_r$ is always greater than or equal to I_r^{-1}. As $V(h, u) \leqq Q$, where Q is the uniform bound of the variance, and since $(h_i - h_{i-1})^2/\tau_i \to 0$ as $\tau_i \to 0$, there is a constant $c > 0$ such that $u_r > c$.

PROOF OF THE THEOREM. By hypothesis there is an estimator $\nu \varepsilon C$ satisfying

$$\sup_h V(h, \nu) \leqq Q \tag{4.12}$$

so it suffices to consider those estimators $\mu \varepsilon C$ such that

$$R(\mu) = \sup_h V(h, \mu) \leqq Q. \tag{4.13}$$

As $R(\mu)$ is a number in the interval [0, Q], there is a sequence $\mu^{(j)} = (\bar{u}^{(j)}, \bar{w}^{(j)})$ of estimators in C, obeying Condition (4.13), such that

$$R(\mu^{(j)}) \to q = \inf R(\mu). \tag{4.14}$$

Since each coordinate $u_i^{(j)}$ of $\mu^{(j)}$ is in [0, 1], there is a subsequence (also written as $\mu^{(j)}$) of estimators which have mu_i's and such that $u_i^{(j)} \to v_i$, $i = 1, \cdots, m$. Although some of the v_i's may be identical, for any $\epsilon > 0$ there is a $\gamma > 0$ such that

$$|u_i^{(j)} - v_i| < \epsilon \tag{4.15}$$

where

$$|v_i - v_k| = 0 \quad \text{or is} > \gamma \tag{4.16}$$

if j is sufficiently large. It will be shown that for any $\eta > 0$ the estimator $\mu^{(j)}$ can be approximated by an estimator $\mu^{*(j)}$ whose u_i's are close to the distinct v_i's, i.e.,

$$V(h, \mu^{*(j)}) \leqq V(h, \mu^{(j)}) + \eta, \tag{4.17}$$

for all h, if j is sufficiently large. Once this has been demonstrated, since the weights $w_i^{*(j)}$ are bounded (the estimator has asymptotic variance $\leqq Q$), there is another subsequence $\mu^{(j)} = (u_i^{(j)})$ converging to $\mu' = (u_i', w_i)$. Since the $\{u_i'\}$ are separated, i.e., $|u_i' - u_{i-1}'| > \gamma/2$, $i = 1, \cdots, r \leqq m$ and since the variance (4.7) is a continuous function of the w_i and u_i, for every h, $V(h, \mu^{*(j)}) \to V(h, \mu')$. Thus for any $\eta > 0$

$$V(h, \mu') \leqq \lim_{j\to\infty} V(h, \mu^{(j)}) + \eta \tag{4.18}$$

or

$$V(h, \mu') \leqq \limsup V(h, \mu^{(j)}) \leqq \limsup R(\mu^{(j)}) = q. \tag{4.19}$$

The definition of q means that $R(\mu') \geqq q$ so that $R(\mu') = q$. The proof will be complete once (4.17) is proved. When there are r distinct separated v_i's, the result follows from the continuity of the variance (4.7). The difficult part consists in showing that if several u_i's are close to one another, then replacing them by one point (or deleting them if they are close to 0) and renormalizing the weights given to the remaining fractiles does not increase the variance by more than η. Consider the case when several u_i's are near 0 and are deleted, i.e., it will be shown that for any $\eta > 0$ and $\gamma > 0$ if $\tau_{j+1} > \gamma$ then there is a $\delta > 0$

such that if $u_1, \cdots, u_j < \delta$, the estimator μ^* formed from the original estimator μ by deleting the fractiles $u_i, \cdots, u_j$ and renormalizing the weights satisfies

$$V(h, \mu^*) \leqq V(h, \mu) + \eta. \tag{4.20}$$

The asymptotic variance of the estimator $\mu = (\bar{u}, \bar{w})$ is given by

$$\textstyle\sum_{i=1}^{j} \Delta_i S_i^2 + \sum_{i=j+1}^{r} \Delta_i S_i^2. \tag{4.21}$$

Now let $S_i = S_{j+1} + S_{ij}$. This defines S_{ij}. Consider the quantity

$$Z = \textstyle\sum_{i=1}^{j} \Delta_i S_{ij}^2 . \tag{4.22}$$

Let $z = \sum_1^j w_i$. If $z \neq 0$, Z is z^2 times the asymptotic variance of the estimator based on the fractiles $u_1, \cdots, u_j$ with weights w_i/z. However, the information in these quantiles is uniformly small if all the u_i are small. Thus, for any $\epsilon > 0$ there is a positive $\delta < \frac{1}{2}$ so that if $u_j < \delta$, then $z^2 \leqq \epsilon Z$. By the Minkowski inequality

$$Z \leqq [(\textstyle\sum_1^j \Delta_i S_i^2)^{\frac{1}{2}} + (u_j/(1 - u_j))^{\frac{1}{2}}|S_{j+1}|]^2 < Q(1 + 2\delta^{\frac{1}{2}}\gamma^{-\frac{1}{2}})^2. \tag{4.23}$$

Therefore, when $u_j < \delta$

$$|z| < \epsilon^{\frac{1}{2}} Q^{\frac{1}{2}}(1 + 2\delta^{\frac{1}{2}}\gamma^{-\frac{1}{2}}). \tag{4.24}$$

Now the estimator given by $(u_{j+1}, \cdots, u_r; w_{j+1}(1 - z)^{-1}, \cdots, w_r(1 - z)^{-1})$ has asymptotic variance

$$(1 - z)^{-2}(\textstyle\sum_{j+1}^{r} \Delta_i S_i^2 + u_j(1 - u_j)^{-1} S_{j+1}^2). \tag{4.25}$$

This can exceed (4.21) by at most

$$2Q(|z| + 2\delta/\gamma)/(1 - |z|)^2, \tag{4.26}$$

which can be made arbitrarily small by choosing δ sufficiently small. A similar argument shows that we may also assume that the u's are bounded away from one.

Now suppose u_i and u_{i+1} are close but $u_{i+2} > u_{i+1} + \gamma/2$. Let us consider the effect on the asymptotic variance if u_{i+1} is deleted and w_i is replaced by $w_i + w_{i+1}$. Since $|S_{i+2}| < (2Q/\gamma)^{\frac{1}{2}}$ and $|S_{i+1}| < (Q/(u_{i+1} - u_i))^{\frac{1}{2}}$,

$$|w_{i+1}/h_{i+1}| < K(Q/(u_{i+1} - u_i))^{\frac{1}{2}}. \tag{4.27}$$

For the new u's and weights w's, which will be denoted by primes, $\Delta_j' = \Delta_j$ for $j \leqq i$, $\Delta_{i+1}' = 0$, $\Delta_{i+2}' = \Delta_{i+2} + (u_{i+1} - u_i)[(1 - u_i)(1 - u_{i+1})]^{-1}$, $S_j' = S_j$ for $j \geqq i + 2$, and for $j \leqq i$

$$\begin{aligned} S_j' - S_j &= w_{i+1}((1 - u_{i+1})h_{i+1}^{-1} - (1 - u_i)h_i^{-1}) \\ &= w_{i+1}(1 - u_i)((h_i - h_{i+1})/h_i h_{i+1}) \\ &\qquad + w_{i+1}((u_i - u_{i+1})/h_{i+1}). \end{aligned} \tag{4.28}$$

We have placed weight $w_i + w_{i+1}$ on u_i and have made the new u_{i+1} equal to the old u_i with no weight. The validity of formula (4.7) is unaffected by these changes. Therefore,

$$(4.29)\quad |S_j' - S_j| < K_1|h_i - h_{i+1}|(u_{i+1} - u_i)^{-\frac{1}{2}} + K_2(u_{i+1} - u_i)^{\frac{1}{2}}, \qquad j \neq i + 1.$$

As u_{i+1} approaches u_i, the first term on the right side of (4.29) approaches zero because of the finiteness of information (Condition (4.10)) and the second term clearly goes to zero. Hence, the difference in the variances of the two estimators tends to zero as u_{i+1} approaches u_i. Thus, in at most finitely many steps we may reduce the problem to the case where the u_i's used are separated by at least $\gamma/2$.

6. Discussion. The existence proofs given in this paper are mainly of theoretical interest and the reader may ask what are their implications for further research which may lead to practical procedures.

The assumptions of the present article were unrealistic in that the statistician had to choose one estimator and could not change it as the data accumulated. Originally, we had hoped that the maximin asymptotic efficiency would be higher so that one might recommend these procedures for moderate sample sizes. Indeed, we have seen that for some families of possible densities a suitably chosen trimmed mean is quite efficient. The numerical work of Crow and Siddiqui [6] also support this conclusion for small sample sizes.

There are several ways one could modify his estimate as the data accumulated. One procedure would be to reduce the family $\mathfrak{F}$ of possible densities for the observations. This could also be combined with the Bayesian approach of Birnbaum and Laska since Yhap [16] has obtained the Bayes estimate for a prior distribution over a finite family of densities. At present the authors are working on estimating the optimum weights for a linear combination of a finite number of percentiles directly from the data. This can be done since these weights only depend on the density function at these percentiles. As the sample gets larger one would use more percentiles. Thus, eventually one would have a fully efficient estimator.

Since procedures, such as the one proposed, would be more complicated to use than a linear function of the ordered observations, it is useful to obtain a lower bound on the maximin efficiency attainable by various simpler types of estimators over families $\mathfrak{F}$ of densities which are likely to arise in practice. The results of [7] applied to Hodges-Lehmann estimators constructed from the maximin rank test for a family $\mathfrak{F}$ of densities gives another lower bound on the maximin efficiency obtainable by using a non-linear function of the ordered observations. In the example given in Section 3, we noted that the maximin efficiency of the Hodges-Lehmann estimator based on the maximin rank test was higher than that of the maximin efficient linear estimator. We conjecture that this is generally true.

REFERENCES

[1] BENNETT, C. A. (1952). Asymptotic properties of ideal linear estimators. Unpublished dissertation. Univ. of Michigan.

[2] BICKEL, P. J. (1965). On some robust estimates of location. *Ann. Math. Statist.* **36** 847–858.

[3] BIRNBAUM, A. and LASKA, E. (1965). Optimal robustness for estimators and tests. Tech. Rep. 334, Courant Inst. of Math. Sciences.

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

[5] CHERNOFF, H., GASTWIRTH, J. L. and JOHNS, M. V., JR. (1967). Asymptotic distribution of linear combinations of order statistics. *Ann. Math. Statist.* **38** 52–72.

[6] CROW, E. L. and SIDDIQUI, M. (1967). Robust estimation of location. *J. Amer. Statist. Assoc.* **62** 353–389.

[7] GASTWIRTH, J. L. (1966). On robust procedures. *J. Amer. Statist. Assoc.* **61** 929–948.

[8] HODGES, J. L., JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **27** 324–335.

[9] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–102.

[10] JUNG, J. (1955). On linear estimates defined by continuous weight function. *Ark. Math. Band* **3** 199–209.

[11] SION, M. (1958). On general minimax theorems. *Pacific J. Math.* **8** 171–179.

[12] STEIN C. (1963). Statistical Decision Theory. Lecture Notes. Stanford Univ.

[13] TUKEY, JOHN W. (1960). A survey of sampling from contaminated distributions. Paper 39 (448–485) in *Contribution to Probability and Statistics* (edited by I. Olkin, et. al.) Stanford Univ. Press.

[14] TUKEY, JOHN W. (1962). The future of data analysis. *Ann. Math. Statist.* **33** 1–67.

[15] TUKEY, JOHN W. (1967). Some perspectives on data analysis. R. A. Fisher Memorial Lecture. December, 1967.

[16] YHAP, E. (1966). Asymptotic optimally robust linear unbiased estimators of location and scale parameters. To be submitted to *Ann. Math. Statist.*