The Annals of Statistics
1981, Vol. 9, No. 3, 567–577

MAXIMIZING THE VARIANCE OF M-ESTIMATORS USING THE GENERALIZED METHOD OF MOMENT SPACES

By John R. Collins[1] and Stephen L. Portnoy[2]

University of Calgary and University of Illinois

The problem considered is that of optimizing a function of a finite number of linear functionals over an infinite dimensional convex set S. It is shown that under some reasonably general conditions the method of moment spaces can be used to reduce the problem to one of optimizing over a simple finite dimensional set (generally a set of convex combinations of extreme points of S). The results are applied to finding the maximum asymptotic variance of M-estimators over classes of distributions arising in the theory of robust estimation.

1. Introduction. The method of moment spaces has been used in a wide variety of situations to characterize solutions of variational problems. One problem to which the method is often applicable is that of optimizing a function depending on a finite number of linear functionals over an infinite dimensional convex set, S. In such cases the method of moment spaces will often reduce the problem to one of optimizing over a simple finite dimensional set (generally a set of convex combinations of extreme points of S). Three results which are formal statements of this principle are presented in Section 2. They are simple, but appear to be unpublished. In Section 3 the results are directly applied to finding the maximum asymptotic variance of M-estimators over classes of symmetric distributions arising in the theory of robust estimation. The results extend those of Collins [2], and show that the maximum asymptotic variance can be found by examining contaminating distributions which are convex combinations of at most two pairs of symmetric point masses. Naive intuition might suggest that a single pair of symmetric point masses may be sufficient (e.g., see the conjecture on page 33 of [6]). However, a simple example is presented here showing that consideration of two pair of point masses is necessary.

Section 4 extends the results of Section 2 to some special cases where the maximizing convex combination can be easily identified, and includes an application to a recent result of Efron and Olshen [3].

2. The general results.

Theorem 1. *Let S be a convex compact subset of a locally convex linear topological Hausdorff space, let $A: S \to R^n$ be affine and continuous, and let $T: R^n \to R$ be a function which attains its maximum over $A(S)$ on the boundary of $A(S)$. Then there is a convex combination of at most n extreme points of S at which $T \circ A$ is maximized over S.*

Proof. By the hypotheses on S and A, $A(S)$ is a compact, convex subset of R^n; and so by Carathéodory's Theorem (see [7], page 155 ff) each boundary point of $A(S)$ is a convex combination of at most n extreme points of $A(S)$. The theorem follows since each

Received December, 1977; revised February, 1979.

[1] Research supported by National Research Council of Canada Grant A-4499.

[2] Research supported by National Science Foundation Grant MCS 75-07978, MCS 78-04014 and MCS 80-02340.

AMS 1970 *subject classifications.* Primary 62G35; secondary 62G05.

Key words and phrases. Method of moment spaces, robust estimation, asymptotic variance.

extreme point of $A(S)$ is the image of an extreme point of S ([5], page 132, problem C). □

REMARK 1. If the condition in the hypothesis that T attains its maximum on the boundary of $A(S)$ is replaced by the weaker condition that T attains its maximum at some point in $A(S)$, then the conclusion of the theorem is true with "n" replaced by "$n + 1$." Also the theorem is true with "maximized" replaced by "minimized."

REMARK 2. Theorem 1 is a satisfactory result for the particular applications treated in Sections 3 and 4. We present two more versions of the result which may have useful applications to cases where the strong assumption that the functional A is continuous does not hold. In Theorems 2 and 3, the condition that A is continuous is weakened and replaced by other topological conditions that can be checked in R^n. The proofs are relatively straightforward and will be deleted.

Notation. E^c denotes the convex hull of E, and $\bar{E}$ denotes the closure of E.

THEOREM 2. *Let S be a convex subset of a locally convex linear topological Hausdorff space, and let E be any subset of S such that $S \subset \overline{E^c}$. Let $A:S \to R^n$ be affine and let $T:R^n \to R$ attain its maximum over $\overline{[A(E)]^c}$ on the boundary of $\overline{[A(E)]^c}$. Suppose that $A(E)$ is compact and that $A(S) \subset [A(E)]^c$. Then there is a convex combination of at most n points of E at which $T \circ A$ is maximized over S.*

THEOREM 3. *Theorem 2 remains valid if the hypothesis that $A(S) \subset \overline{[A(E)]^c}$ is replaced by the hypothesis that $T \circ A$ is lower semicontinuous on $\overline{E^c}$.*

REMARK 3. Clearly Theorems 2 and 3 remain valid if the word "maximized" is replaced by "minimized," and in Theorem 3 the words "lower semicontinuous" are replaced by "upper semicontinuous."

REMARK 4. If S is compact and E is the set of extreme points of S, then $S = \overline{E^c}$ (by the Krein-Milman Theorem, page 131 of [5]). If A is continuous and E is closed, then Theorem 1 follows from Theorem 2. However, Theorem 1 is not a special case of Theorem 2, since Theorem 1 is true even when E is not closed and $A(E)$ is not compact.

3. An application: Maximizing the asymptotic variance of M-estimators. Consider the robust estimation model of Huber [4]. Let $X_1, \cdots, X_n$ be i.i.d. random variables with distribution function $F(x - \theta)$, where θ is unknown and F is an unknown member of a specified class of distributions $\mathscr{F}$. An M-estimator of θ is defined as a solution $\hat{\theta}_n$ of an equation of the form

$$\sum_{i=1}^{n} \psi(X_i - \hat{\theta}_n) = 0. \tag{3.1}$$

Under regularity conditions on ψ and F, $\{\hat{\theta}_n\}$ is a consistent sequence of estimators of θ and $n^{1/2}(\hat{\theta}_n - \theta)$ converges in distribution to the normal distribution with mean 0 and variance $V(\psi, F)$, where

$$V(\psi, F) = \int \psi^2 \, dF \Big/ \left[\int \psi' \, dF \right]^2. \tag{3.2}$$

One then considers $\sup\{V(\psi, F) : F \in \mathscr{F}\}$ to be a measure of the robustness of the M-estimator based on ψ.

Collins [2] studied the case of M-estimators for ψ's which vanish off a compact set $[-c, c]$. Consider now the problem of finding $\sup\{V(\psi, F) : F \in \mathscr{F}\}$ for such ψ's. In many interesting examples, such as Andrew's [1] $\psi(x) = \sin(\pi x/c)$ for $|x| \leq c$ and $= 0$ elsewhere, ψ' is continuous everywhere except at $\pm c$. In such cases $V(\psi, F)$ is not a continuous functional, so that Theorem 1 does not apply. One may either: (i) apply Theorem 3; or (ii) by restricting the domain of both ψ and F to $[-c, c]$, define an equivalent problem to which Theorem 1 applies. The latter approach is taken below.

Let $c > 0$ be fixed, and let $\mathscr{F}_c$ be a specified class of (possibly substochastic) distributions F which are symmetric about 0 and have support $[-c, c]$. Assume further that $\mathscr{F}_c$ is convex and compact when endowed with the vague topology (i.e., the weakest topology for which the mapping $F \to \int_{-c}^{c} g\, dF$ is continuous for all continuous $g:[-c, c] \to R$). Let Ψ_c denote the class of functions $\psi:[-c, c] \to R$ satisfying (i) ψ is continuous; (ii) $\psi(x) = -\psi(-x)$ for all $x \in [-c, c]$; (iii) $\psi(x) > 0$ when $0 < x < c$; (iv) $\psi(c) = 0$; and (v) the derivative $\psi':[-c, c] \to R$ is continuous.

Assume that ψ and $\mathscr{F}_c$ satisfy

$$\inf\left\{\int_{-c}^{c} \psi'\, dF : F \in \mathscr{F}_c\right\} > 0. \tag{3.3}$$

The problem is to find $\sup\{V_c(\psi, F) : F \in \mathscr{F}_c\}$ for a given $\psi \in \Psi_c$, where the functional $V_c(\psi, F)$ is defined by:

$$V_c(\psi, F) = \int_{-c}^{c} \psi^2\, dF \Big/ \left[\int_{-c}^{c} \psi'\, dF\right]^2. \tag{3.4}$$

Define $A:\mathscr{F}_c \to R^2$ by

$$A(F) = \left[\int_{-c}^{c} \psi'\, dF, \int_{-c}^{c} \psi^2\, dF\right],$$

and note that A is continuous and affine. Define $T: R^2 \to R$ by $T(x, y) = y/x^2$. Since $A(\mathscr{F}_c)$ is a compact subset of $\{(x, y): x > 0\} \subset R^2$, T must attain its maximum value (call it m_0) over $A(\mathscr{F}_c)$ at some point $(x_0, y_0) \in A(\mathscr{F}_c)$. Since $\{(x, y): T(x, y) = m_0\}$ is a connected set with at least one point in $A(\mathscr{F}_c)$ and also points in the complement of $A(\mathscr{F}_c)$ (in a neighborhood of $(0, 0)$), T must attain m_0 on the boundary of $A(\mathscr{F}_c)$. By Theorem 1, $T \circ A(F)[= V_c(\psi, F)]$ attains its maximum over $\mathscr{F}_c$ at a convex combination of at most two extreme points of $\mathscr{F}_c$.

EXAMPLE 1. (The gross errors model). Let ε be fixed, $0 < \varepsilon < 1$, let $c > 0$ be fixed and let Φ denote the standard normal distribution function. Let $\mathscr{F}_{1,\varepsilon}$ denote the class of distributions of the form $F = (1 - \varepsilon)\Phi + \varepsilon G$ for some symmetric G, with domain restricted to the set $[-c, c]$. The extreme points of $\mathscr{F}_{1,\varepsilon}$ are the d.f.'s of the form

$$(1 - \varepsilon)\Phi + \varepsilon\delta_{\pm x},$$

where $\delta_{\pm x}$ is the d.f. that puts mass ½ at x and mass ½ at $-x$. So if $\psi \in \Psi_c$ and ψ and $\mathscr{F}_{1,\varepsilon}$ satisfy (3.3), then $\sup\{V_c(\psi, F): F \in F_{1,\varepsilon}\}$ is attained at a convex combination of at most two extreme distributions. The problem of computing the supremum is reduced to the problem of finding the triple $(x_1, x_2, \alpha) \in [0, c]^2 \times [0, 1]$ which maximizes

$$\{A_0 + \varepsilon[\alpha\psi^2(x_1) + (1 - \alpha)\psi^2(x_2)]\}/\{B_0 + \varepsilon[\alpha\psi'(x_1) + (1 - \alpha)\psi'(x_2)]\}^2,$$

where $A_0 = (1 - \varepsilon)\int_{-c}^{c} \psi^2\, d\Phi$ and $B_0 = (1 - \varepsilon)\int_{-c}^{c} \psi'\, d\Phi$.

EXAMPLE 2. (The Kolmogorov model). Let ε be fixed, $0 < \varepsilon < 1$, and define $\mathscr{F}_{2,\varepsilon}$ to be the class of restrictions to $[-c, c]$ of distributions F which are symmetric and which satisfy $\sup_x |F(x) - \Phi(x)| \leq \varepsilon$. The extreme points of $\mathscr{F}_{2,\varepsilon}$ are those symmetric distributions which have restrictions to $[0, c]$ of the form

$$F(y) = F_*(y) \qquad 0 < y < x$$
$$= F^*(y) \qquad y \geq x$$

for some $x \geq 0$, where $F_*(y) = \max\{\tfrac{1}{2}, \Phi(y) - \varepsilon\}$ and $F^*(y) = \min\{\Phi(y) + \varepsilon, 1\}$. If $\psi \in \Psi_c$, and ψ and ε are such that (3.3) holds, then $\sup\{V_c(\psi, F): F \in \mathscr{F}_{2,\varepsilon}\}$ is attained at a convex combination of at most two such extreme distributions.

Examples 1 and 2 greatly improve the results of [2], Sections 4 and 5, respectively. In [2], the problem of finding $\sup\{V_c(\psi, F): F \in \mathscr{F}\}$ was approached by obtaining a characterization of the convex set of F's in $\mathscr{F}$ at which the supremum is attained (Theorem 3.1 of [2]). It was shown in [2] that under some quite special conditions on ψ, there is an extreme distribution maximizing $V_c(\psi, F)$. But the general question of the "simplest" possible form of F required to attain sup $V_c(\psi, F)$ was left unanswered in [2].

For all the special ψ's in Ψ_c that have previously been considered in the literature, $\sup\{V_c(\psi, F): F \in \mathscr{F}_{1,\varepsilon}\}$ is attained at an extreme point of $\mathscr{F}_{1,\varepsilon}$. We now give an example of a ψ for which the simplest F maximizing $V(\psi, F)$ is a proper convex combination

$$(1 - \varepsilon)\Phi + \varepsilon[p\delta_{\pm x_1} + (1 - p)\delta_{\pm x_2}],$$

where $0 < p < 1$ and $0 < x_1 < x_2$.

Let $0 < d < a$, let $M > 1$, and define

$$\begin{aligned} \psi_0(x) &= x && 0 \leq x \leq a \\ &= 2a - x && a \leq x \leq 2a - d \\ &= d + M(2a - d - x) && 2a - d \leq x \leq c \\ &= 0 && x \geq c \\ &= -\psi_0(-x) && x \leq 0, \end{aligned}$$

where $c = 2a - d\left(1 - \dfrac{1}{M}\right)$. Note that $\psi_0'(x) = 1$ for $0 < x < a$, $\psi_0'(x) = -1$ for $a < x < 2a - d$, $\psi_0'(x) = -M$ for $2a - d < x < c$. Consider a choice of a, d, M and ε for which ψ_0 satisfies (3.3). Note that no F in $\mathscr{F}$ maximizes

$$\int_{-c}^{c} \psi_0^2 \, dF, \quad \text{but that} \quad \int_{-c}^{c} \psi_0^2 \, dF_n \to \sup \int_{-c}^{c} \psi_0^2 \, dF \text{ as } n \to \infty,$$

where $F_n = (1 - \varepsilon)\Phi + \varepsilon\,\delta_{\pm x_n}$, where $\{x_n\}$ is any sequence such that $x_n \downarrow a$. We express this figuratively as:

$$(1 - \varepsilon)\Phi + \varepsilon\delta_{\pm(a+0)} \text{ "maximizes" } \int_{-c}^{c} \psi_0^2 \, dF.$$

Similarly

$$(1 - \varepsilon)\Phi + \varepsilon\delta_{\pm(2a-d+0)} \text{ "minimizes" } \int_{-c}^{c} \psi_0' \, dF.$$

Finally one can see directly that a "distribution" of form

$$(1 - \varepsilon)\Phi + \varepsilon[p\delta_{\pm(a+0)} + (1 - p)\delta_{\pm(2a-d+0)}] \tag{3.5}$$

"maximizes" $V_c(\psi_0, F)$, for some p, $0 \leq p \leq 1$. For if F^* is any other distribution in $\mathscr{F}_{1,\varepsilon}$, one can strictly increase $V_c(\psi_0, F^*)$ by moving the mass on $(a, 2a - d)$ under F^* to $a + 0$, and the mass on $(2a - d, c]$ to $2a - d + 0$.

For example, let $\varepsilon = 0.1$, $a = 10$, $d = 10^{-10}$ and $M = 6$. Then (3.3) is satisfied:

$$\inf \int_{-c}^{c} \psi' \, dF \doteq 0.3 > 0.$$

Using the notation $V(p)$ to stand for V evaluated at ψ_0 and the convex combination (3.5), one calculates that $V(0) \doteq 10$, $V(1) \doteq 17.031$ and that $V(p)$ attains a maximum value $\doteq 19.608$ when $p \doteq 0.42$.

Now let ψ be a function in Ψ_c which is obtained by smoothing ψ_0 in very small neighborhoods of its discontinuities at $\pm a$ and $\pm(2a - d)$. Then sup $V_c(\psi, F)$ is attained at some F in $\mathscr{F}_{1,\varepsilon}$; and by a continuity argument there is an F maximizing $V_c(\psi, F)$ which is a proper convex combination "very close" to the F which "maximizes" $V_c(\psi_0, F)$.

In Figure 1, $\int_{-c}^{c} \psi' \, dF$ is plotted against $\int_{-c}^{c} \psi^2 \, dF$ for the ψ of the previous paragraph (but the plot is not to scale). The solid line in Figure 1 is $A(E)$, the image of the extreme points $E = \{(1 - \varepsilon)\Phi + \varepsilon\delta_{\pm x} : x \in [0, c]\}$ under $A(F) = (\int_{-c}^{c} \psi' \, dF, \int_{-c}^{c} \psi^2 \, dF)$. The shaded area is $A(\mathscr{F}_{1,\varepsilon})$, which in this case is the convex hull of $A(E)$. For the choice of parameters in the example, $T(x, y) = y/x^2$ is maximized over $A(\mathscr{F}_{1,\varepsilon})$ at a unique point in the interior of the dotted line in the figure (necessarily the image under A of a proper convex combination of two extreme points).

The general method used for finding sup $V_c(\psi, F)$ (including the idea of plotting $A(F)$ in R^2) is outlined in a discussion on pages 32 and 33 of Portnoy [6]. The only basic idea overlooked there was that boundary points of $A(\mathscr{F})$, which correspond to extreme points of $\mathscr{F}$ in the special example of the sine-wave ψ, may sometimes correspond to proper convex combinations of two extreme points.

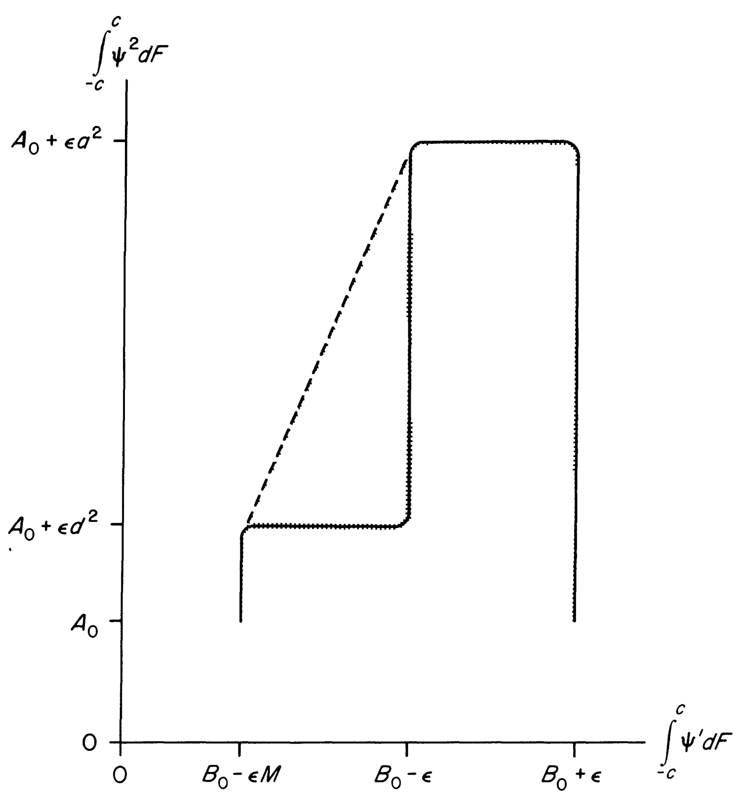

FIG. 1. Graph of $\int_{-c}^{c} \psi' \, dF$ vs. $\int_{-c}^{c} \psi^2 \, dF$ for the ψ of the example.
Notation: $A_0 = (1 - \varepsilon) \int_{-c}^{c} \psi^2 \, d\Phi$ and $B_0 = (1 - \varepsilon) \int_{-c}^{c} \psi' \, d\Phi$.

We conclude this section with another application of Theorem 1. Let ψ be in Ψ_c and suppose that $\inf\{\int_{-c}^{c} \psi' \, dF : F \in \mathscr{F}_{1,\varepsilon}\} > 0$. For a special model of dependence described in Section 1 of [6], an approximate expression for the asymptotic variance of the M-estimator based on ψ is ([6], formula (2.11)):

$$\frac{\int_{-c}^{c} \psi^2(x)\, dF}{\left(\int_{-c}^{c} \psi'(x)\, dF\right)^2} + 4\rho \frac{\int_{-c}^{c} x\psi(x)\, dF}{\int_{-c}^{c} \psi'(x)\, dF}. \tag{3.6}$$

Then (3.6) is maximized over $\mathscr{F}_{1,\varepsilon}$ at a convex combination of at most three extreme points of $\mathscr{F}_{1,\varepsilon}$. To see this, apply Theorem 1 with $A: \mathscr{F}_{1,\varepsilon} \to R^3$ defined by

$$A(F) = \left(\int_{-c}^{c} \psi'(x)\, dF, \int_{-c}^{c} \psi^2(x)\, dF, \int_{-c}^{c} x\psi(x)\, dF\right)$$

and $T: R^3 \to R$ defined by $T(x, y, z) = (y/x^2) + 4\rho z/x$.

4. Identifying the maximizing convex combination. The results of Section 2 reduce an infinite-dimensional problem to the problem of finding the maximizing (or minimizing) convex combination of n extreme points. In application to some special cases when $n = 2$ [$n = 3$], this problem may further be reduced to a simple one-dimensional [two-dimensional] problem. Some examples are given below.

Returning to Example 1 of Section 3, we consider the problem of finding sufficient conditions for the asymptotic variance to be maximized at a single extreme point. For $\psi \in \Psi_c$, the set $A(E)$ in R^2 is the curve $\{(x(t), y(t)) : t \in [0, c]\}$, where $x(t) = (1 - \varepsilon) \int_{-c}^{c} \psi'(u)\, du + \varepsilon\psi'(t)$ and $y(t) = (1 - \varepsilon) \int_{-c}^{c} \psi^2(u)\, du + \varepsilon\psi^2(t)$. A sufficient condition for $T(x, y) = y/x^2$ to attain its maximum over $A(\mathscr{F}_{1,\varepsilon})$ at a point on $A(E)$ is clearly that, on $A(E)$, ψ be twice differentiable and $dy/dx = y'(t)/x'(t) = 2\psi(t)\psi'(t)/\psi''(t)$ be monotone nondecreasing, so that the boundary of $A(\mathscr{F}_{1,\varepsilon})$ is the union of the concave curve $A(E)$ and the line segment connecting $(x(0), y(0))$ to $(x(c), y(c))$. This is condition (4.4) of Collins [2], which is satisfied by Andrew's sine-wave ψ. A plot of $A(\mathscr{F}_{1,\varepsilon})$ in R^2 for the sine-wave ψ is given in Figure 1 of Portnoy [6].

For ψ's in Ψ_c for which $2\psi\psi'/\psi''$ is not monotone nondecreasing, one can sometimes deduce the maximizing convex combination from the behavior of $2\psi\psi'/\psi''$. For example, suppose that $\psi \in \Psi_c$ is defined by $\psi(t) = ct - t^2$ for $0 \le t \le c$, and $\psi(t) = -\psi(-t)$ for $-c \le t \le 0$. Then $2\psi(t)\psi'(t)/\psi''(t) = -t(c-t)(c-2t)$, which is not monotone nondecreasing on $[0, c]$. In fact $(d/dt)[2\psi(t)\psi'(t)/\psi''(t)] = 6t(c - t) - c^2$, which is ≥ 0 in the interval $(c/2) \pm (\sqrt{3}/6)c$ and is ≤ 0 elsewhere in $[0, c]$. From this one easily sees that the line segments connecting $[x(0), y(0)]$ to $[x(c/3), y(c/3)]$ and $[x(2c/3), y(2c/3)]$ to $[x(c), y(c)]$ are on the boundary of the convex hull of $\{(x(t), y(t)) : t \in [0, c]\}$. Thus for ε and c satisfying (3.3), the distribution on $[-c, c]$ maximizing $V_c(\psi, F)$ over $\mathscr{F}_{1,\varepsilon}$ has (depending on the values of ε and c) either the form (i) $(1 - \varepsilon)\Phi + \varepsilon\delta_{\pm x}$ for some $x \in [c/2, 2c/3]$, or (ii) $(1 - \varepsilon)\Phi + p\varepsilon\delta_{\pm 2c/3} + (1 - p)\varepsilon\delta_{\pm c}$ for some $p \in [0, 1]$. In either case one has a simple one-dimensional problem. A graph of $A(\mathscr{F}_{1,\varepsilon})$ in R^2 is presented for this example in Figure 2.

We conclude by investigating a special case motivated by results of Efron and Olshen [3]. In the context of Section 2, if $n = 3$, T is linear, and E is one-dimensional, then there are some simple sufficient conditions under which the boundary point in $A(S)$ can be identified. In particular, the following theorem often provides the answer:

THEOREM 4. *Let $\{(x, f(x), g(x)) : \underline{x} \le x \le \bar{x}\}$ be a curve in R^3 with f and g twice differentiable and f strictly monotonic and either concave or convex. Define for $\underline{x} \le u \le \bar{x}$ and $\underline{x} \le v \le \bar{x}$,*

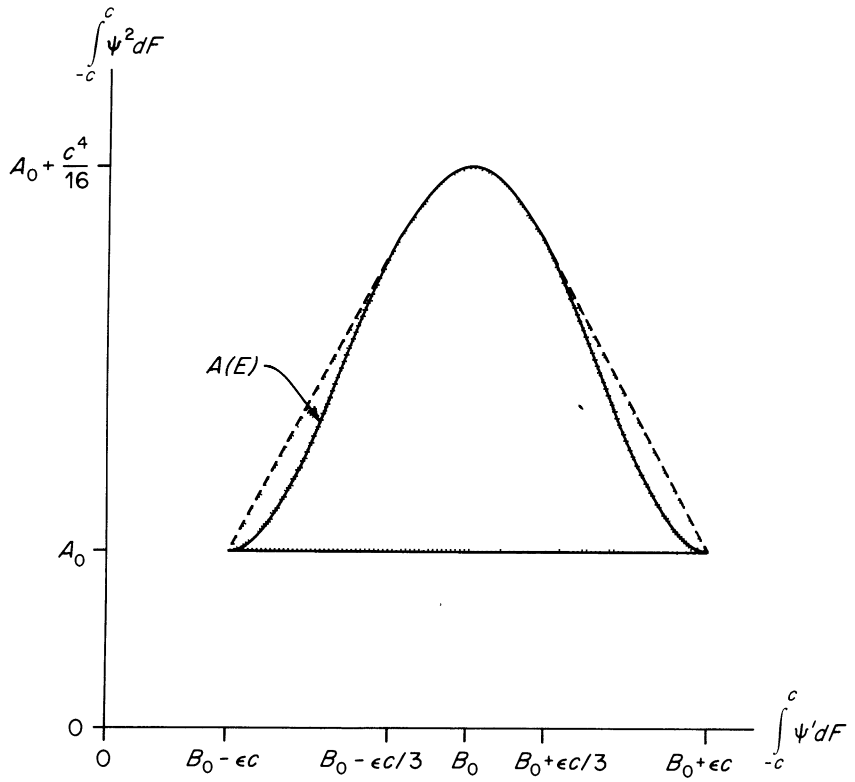

FIG. 2. Graph of $\int_{-c}^{c} \psi' \, dF$ vs. $\int_{-c}^{c} \psi^2 \, dF$ when $\psi(x) = cx - x^2, 0 \le x \le c$
$= cx + x^2, -c \le x \le 0.$

$$V(u, v) = f''(v)\, g''(u) - f''(u) g''(v), \tag{4.1}$$

and assume that $V(u, v)$ has the same sign for all $v \ge u$. Let (x_0, y_0) be in the convex hull of $\{(x, f(x)) : \underline{x} \le x \le \bar{x}\}$ in R^2, and make the following definitions (see Figure 3): let $\alpha = (x_0 - \underline{x})/(\bar{x} - \underline{x})$ and (since f is monotonic) define X_1 and X_2 uniquely so that

$$\begin{aligned} f(X_2) &= (y_0 - (1 - \alpha) f(\underline{x}))/\alpha \\ f(X_1) &= (y_0 - \alpha f(\bar{x}))/(1 - \alpha). \end{aligned} \tag{4.2}$$

Define

$$\begin{aligned} \bar{Z} &= \alpha g(\bar{x}) + (1 - \alpha) g(X_1) \\ \underline{Z} &= \alpha g(X_2) + (1 - \alpha) g(\underline{x}). \end{aligned} \tag{4.3}$$

Let S be the convex hull of $\{(x, f(x), g(x)) : \underline{x} \le x \le \bar{x}\}$ in R^3 and let $S(x_0, y_0) = \{z : (x_0, y_0, z) \in S\}$. Then if $V(u, v)$ (for $v \ge u$) has the same sign as $f''(x)$ (which is either always nonpositive or always nonnegative) then

$$\underline{Z} = \inf S(x_0, y_0) \text{ and } \bar{Z} = \sup S(x_0, y_0). \tag{4.4}$$

If the signs are opposite then $\underline{Z}$ and $\bar{Z}$ are reversed in (4.4).

PROOF. For each $a \in (\underline{x}, \bar{x})$ there is a unique point $b \in (\underline{x}, \bar{x})$ such that (x_0, y_0) lies on the line from $(a, f(a))$ to $(b, f(b))$ in R^2. (Note that $b = b(a)$ is a differentiable function of a.) That is, $f(b)$ satisfies

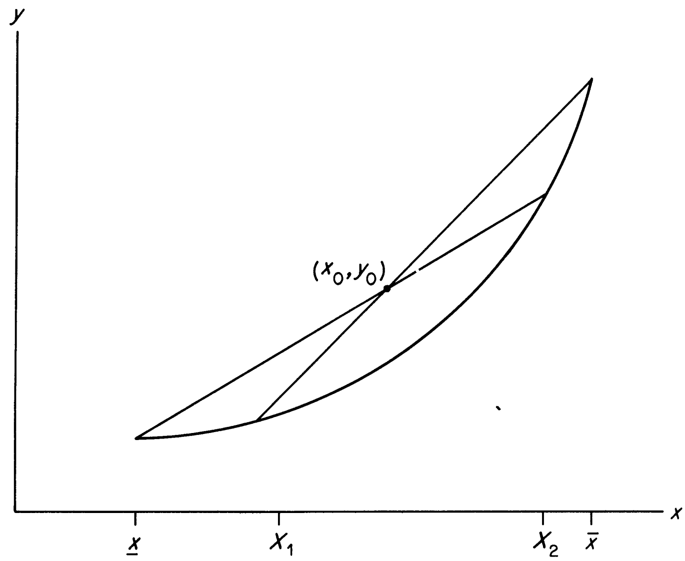

FIG. 3. $y = f(x)$

$$y_0 = p(a)f(b) + (1 - p(a))f(a) \tag{4.5}$$

where

$$p(a) = (x_0 - a)/(b - a). \tag{4.6}$$

Let $z(a)$ be the z coordinate of the point in R^3 vertically above (or below) (x_0, y_0) on the line from $(a, f(a), g(a))$ to $(b, f(b), g(b))$:

$$z(a) = p(a)g(b) + (1 - p(a))g(a). \tag{4.7}$$

We want to show first that $z(a)$ is a monotonic function of a; so differentiating (4.7) yields

$$z'(a) = p'(a)(g(b) - g(a)) + p(a)b'(a)g'(b) + (1 - p(a))g'(a). \tag{4.8}$$

Differentiating (4.5) and (4.6) yields (with some calculation)

$$0 = p'(a)(f(b) - f(a)) + p(a)b'(a)f'(b) + (1 - p(a))f'(a) \tag{4.9}$$

$$p'(a) = -\frac{1}{b - a}(1 - p(a) + p(a)b'(a)). \tag{4.10}$$

From (4.9) and (4.10)

$$0 = -(1 - p(a))A + p(a)b'(a)B \text{ or } p(a)b'(a) = (1 - p(a))A/B \tag{4.11}$$

where

$$A = \frac{f(b) - f(a)}{b - a} - f'(a), \qquad B = f'(b) - \frac{f(b) - f(a)}{b - a}. \tag{4.12}$$

Note that since f is concave or convex, B is always of the same sign (for $\underset{\sim}{x} \le a \le \bar{x}$). From (4.11) and (4.10)

$$p'(a) = -\frac{1 - p(a)}{b - a}\,\frac{A + B}{B}$$

and, hence, from (4.8) and (4.11)

$$
\begin{aligned}
z'(a) &= -(1-p(a))\frac{g(b)-g(a)}{b-a}\frac{A+B}{B} + (1-p(a))\,g'(b)\frac{A}{B} + (1-p(a))g'(a) \\
&= \frac{1-p(a)}{B}\left\{Ag'(b) + Bg'(a) - \frac{g(b)-g(a)}{b-a}(A+B)\right\} \qquad (4.13) \\
&= \frac{1-p(a)}{-B}(BD-AC)
\end{aligned}
$$

where

$$
C = g'(b) - \frac{g(b)-g(a)}{b-a}, \qquad D = \frac{g(b)-g(a)}{b-a} - g'(a).
$$

Writing function differences as integrals,

$$
\begin{aligned}
(b-a)^2(BD-AC) &= \int_a^b\int_a^b [(f'(b)-f'(s))(g'(r)-g'(a)) - (f'(r)-f'(a)) \\
&\qquad \cdot (g'(b)-g'(s))]\,dr\,ds \\
&= \int_a^b\int_a^b\int_a^r\int_s^b [f''(v)g''(u) - f''(u)g''(v)]\,dv\,du\,dr\,ds \\
&= \int_a^b\int_a^b\int_a^r\int_s^b V(u,v)\,dv\,du\,dr\,ds.
\end{aligned}
$$

The inner two integrals are over a lower-right-hand rectangular corner of the rectangle with vertices (a, a) and (b, b) in the $(n-u)$ plane. Since $V(u, v) = -V(v, u)$, the integral of the part of this domain (if any) above the line $u = v$ cancels with a congruent part below the line, leaving an integral over a set with $v \geq u$ where $V(u, v)$ always has the same sign. Since the sign of B is the same as the sign of $f''(x)$ (4.12) and since the sign of $z'(a)$ is the product of the sign of $(-B)$ and the sign of $V(u, v)$ (for $v \geq u$; see (4.13)), $z(a)$ will be decreasing if and only if $f''(x)$ and $V(u, v)$ (for $v \geq u$) have opposite signs. Thus, among points in $S(x_0, y_0)$ which are convex combinations of at most two points along the curve, the desired result follows.

To complete the proof note that by the argument of Theorem 1, every boundary point of S is a convex combination of at most 3 points along the curve. So let U be a convex combination of points P_1, P_2 and P_3 on the curve (with x-coordinates $x_1 < x_2 < x_3$ respectively—see Figure 4 for the projection into the $(x-y)$ plane). Let V be the intersection of $\overline{P_1U}$ with $\overline{P_2P_3}$ and let Q be the point at which the curve meets the vertical plane through P_1 and U. From the preceding argument (with (x_0, y_0) the $(x-y)$ projection of V) there is a point of $\overline{P_1Q}$ vertically below V, if $z'(a)$ is positive (and vertically above V if $z'(a)$ is negative). Thus, since lines $\overline{P_1Q}$ and $\overline{P_1V}$ are in the same plane, P_1Q lies entirely below $\overline{P_1V}$ if $z'(a)$ is positive (and entirely above if $z'(a)$ is negative). Thus, if $z'(a)$ is positive, the z-coordinate of U lies above the z-coordinate of the corresponding convex combination of P_1 and Q; and, hence, the z-coordinate along the vertical line through U is minimized by choosing P_1 to have smallest x-coordinate ($\underline{x}$). If $z'(a)$ is negative, the z-coordinate is maximized by choosing Q to have largest possible x-coordinate ($\bar{x}$). A similar graphical argument (introducing the line $\overline{P_3U}$) completes the proof of the theorem. □

COROLLARY. *Let $\{(x(t), y(t), z(t)) : \underline{t} \leq t \leq \bar{t}\}$ be a curve in R^3 with x, y, and z strictly monotonic and twice differentiable. For r, s, and t in $(\underline{t}, \bar{t})$, let*

$$
F(r) = x'(r)y''(r) - x''(r)y'(r)
$$

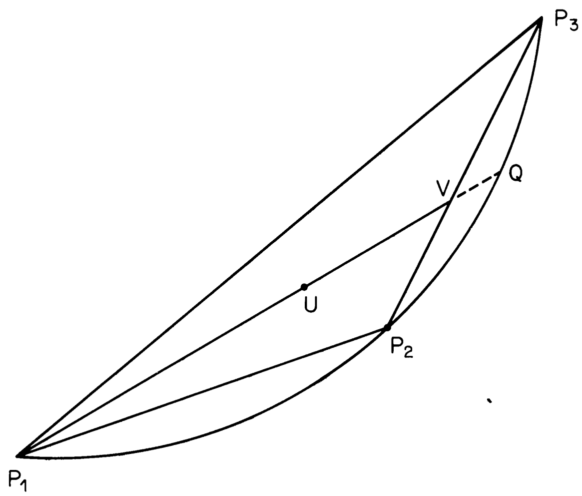

FIG. 4. Three points along the curve.

$$G(r) = x'(r)z''(r) - x''(r)z'(r) \tag{4.14}$$

$$W(s, t) = F(t)G(s) - F(s)G(t).$$

Suppose $F(r)$ has the same sign for all $r \in (\underline{t}, \bar{t})$. Then for any (x_0, y_0) in the convex hull of $\{(x(t), y(t)) : \underline{t} \leq t \leq \bar{t}\}$ in R^2, there are unique points p_1, p_2 in $[0, 1]$ and t_1, t_2 in $\underline{t}, \bar{t})$ satisfying

$$\begin{aligned} p_1x(\underline{t}) + (1 - p_1)x(t_1) &= x_0 \qquad p_2x(t_2) + (1 - p_2)x(\bar{t}) = x_0 \\ p_1y(\underline{t}) + (1 - p_1)y(t_1) &= y_0 \qquad p_2y(t_2) + (1 - p_2)y(\bar{t}) = y_0. \end{aligned} \tag{4.15}$$

Let

$$\underline{Z} = p_1z(\underline{t}) + (1 - p_1)z(t_1) \qquad \bar{Z} = p_2z(t_2) + (1 - p_2)z(\bar{t}). \tag{4.16}$$

Then if $W(s, t)$ has the same sign for all $t \geq s$, the conclusion of Theorem 4 holds with $W(s, t)$ replacing $V(u, v)$ and $F(r)$ replacing $f''(x)$.

PROOF. By monotonicity, f and g can be defined so that $y = f(x)$ and $z = g(x)$. Calculus yields

$$f''(x(t)) = (x'(t)y''(t) - x''(t)y'(t))/(x'(t))^3$$

(and similarly for $g''(x(t))$). Thus, $W(s, t)$ has the same sign as $V(x(s), x(t))$ and $F(t)/x'(t)$ has the same sign as $f''(x(t))$. Thus, whether $x'(t)$ is positive or negative the corollary immediately follows from Theorem 4. □

REMARK 1. Note that the values X_1 and X_2 in the theorem (or p_i and t_i ($i = 1, 2$) in the corollary) depend only on $f(x)$ (or $x(t)$ and $y(t)$). Thus, the z-coordinate is maximized or minimized at the same convex combination of two points of the curve for all functions $g(x)$ (or $z(t)$) satisfying the hypotheses of the theorem (or the corollary).

REMARK 2. The corollary immediately yields a recent result proved quite differently by Efron and Olshen [3]. Let μ be a probability measure on $[0, \infty)$ and let Φ denote the standard normal cdf. For $x > 0$ define

$$H(x) = \int_0^\infty \Phi(hx)\, d\mu(h).$$

Consider the problem of maximizing or minimizing $H(x_3)$ subject to fixed values for $H(x_1)$ and $H(x_2)$ over all probability measures, μ (on $[0, \infty)$). By Theorem 1, since the extreme measures are point masses, this is equivalent to maximizing or minimizing $z(t) \equiv \Phi(x_2 t)$ subject to $x(t) \equiv \Phi(x_1 t) = x_0$ and $y(t) \equiv \Phi(x_2 t) = y_0$. Thus, the corollary applies, and directly computing (4.14) yields

$$F(r) = -c^2 x_1 x_2 (x_2^2 - x_1^2) r \exp\{-\tfrac{1}{2}(x_1^2 + x_2^2) r^2\}$$

$$G(r) = -c^2 x_1 x_3 (x_3^2 - x_1^2) r \exp\{-\tfrac{1}{2}(x_1^2 + x_3^2) r^2\}$$

$$W(s, t) = c^4 x_1^2 x_2 x_3 (x_2^2 - x_1^2)(x_3^2 - x_1^2) st \exp\{-\tfrac{1}{2} x_1^2 (s^2 + t^2)\} \cdot \{\exp[-\tfrac{1}{2}(x_2^2 t^2 + x_3^2 s^2)] - \exp[-\tfrac{1}{2}(x_3^2 t^2 + x_2^2 s^2)]\}.$$

Thus, $F(r)$ has constant sign; and $W(s, t)$ has constant sign for $t \geq s$. Thus, $z(t)$ is maximized or minimized at one of two fixed convex combinations of values of $z(t)$, as described in the proof of the theorem on page 1160 of [3]. (It should be noted that the statement of the theorem incorrectly reverses the inclusion symbols $\in$ and $\notin$ in (1.3) and (1.4), though the proof derives them correctly.)

It is important to note that the theorem (or corollary) provides simple conditions to check when similar results hold for other related mixtures of distributions. For example, to minimize and maximize the Laplace transform Ee^{-sX} for $s_1 \leq s \leq s_2$ subject to $Ee^{-s_1 X} = x_0$ and $Ee^{-s_2 X} = y_0$ (over all distributions for X), let $x(t) = e^{-s_1 t}$, $y(t) = e^{-s_2 t}$, and $z(t) = e^{-st}$. Then $f(x) = x^{s_2/s_1}$ and $g(x) = x^{s/s_1}$; and the conditions for Theorem 4 can be directly checked. Thus, Theorems 1 and 4 immediately yield minimum and maximum values for Ee^{sX}. In particular, for $s_1 = 1$, $s_2 = 2$, $x_0 = \tfrac{1}{2}$ and $y_0 = \tfrac{1}{3}$ (values for the Laplace transform of a negative exponential density), direct computations show that Ee^{-sX} is maximized at $p_1 g(0) + (1 - p_1) g(b_1)$ where $p_1 = \tfrac{1}{4}$ and $b_1 = \tfrac{2}{3}$; and it is minimized at $p_2 g(b_2) + (1 - p_2) g(1)$ where $p_2 = \tfrac{3}{4}$ and $b_2 = \tfrac{1}{3}$. Therefore, for $1 \leq s \leq 2$

$$\tfrac{3}{4} (\tfrac{1}{3})^s + \tfrac{1}{4} \leq Ee^{-sX} \leq \tfrac{3}{4} (\tfrac{2}{3})^s$$

if X satisfies $Ee^{-X} = \tfrac{1}{2}$, $Ee^{-2X} = \tfrac{1}{3}$. Reverse inequalities hold if $0 \leq s < 1$ or $s > 2$.

5. Acknowledgement. The authors would like to thank Dr. Esther Portnoy for her assistance with the proof of Theorem 4, and the referees for their helpful comments.

REFERENCES

[1] ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust Estimates of Location: Survey and Advances.* Princeton Univ. Press.

[2] COLLINS, J. R. (1977). Upper bounds on asymptotic variances of M-estimators of location. *Ann. Statist.* **5** 646–657.

[3] EFRON, B. and OLSHEN, R. A. (1978). How broad is the class of normal scale mixtures? *Ann. Statist.* **6** 1159–1164.

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

[5] KELLY, J. L., NAMIOKA, I., et al. (1963). *Linear Topological Spaces.* Van Nostrand, Princeton.

[6] PORTNOY, S. L. (1977). Robust estimation in dependent situations. *Ann. Statist.* **5** 22–43.

[7] ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton Univ. Press.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF CALGARY
CALGARY, ALBERTA
CANADA T2N IN4.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801