The Annals of Statistics
1987, Vol. 15, No. 3, 1269–1292

ASYMPTOTIC BEHAVIOUR OF S-ESTIMATES OF MULTIVARIATE LOCATION PARAMETERS AND DISPERSION MATRICES

BY P. L. DAVIES

Universität Gesamthochschule Essen

It is shown under appropriate conditions that Rousseeuw's minimum volume estimator and other S-estimators of multivariate location and dispersion parameters are consistent. Under certain differentiability conditions the estimates are asymptotically normally distributed with a norming factor of $n^{1/2}$.

1. Introduction. Let $(X^{(\nu)})_1^\infty$ be a sequence of independently and identically distributed random variables with values in R^k, $k \geq 2$. We suppose that the $X^{(\nu)}$ have a common density function of the form

$$|\Sigma|^{-1/2} f\left((x-\mu)^t \Sigma^{-1}(x-\mu)\right), \tag{1}$$

where t denotes transposition and

(a) $\mu \in R^k$,

(b) $\Sigma \in \mathrm{PDS}(k)$, the set of all positive-definite symmetric (PDS) $k \times k$ matrices,

(c) $|\Sigma|$ denotes the determinant of Σ and

(d) $f: [0, \infty) \to [0, \infty)$ satisfies

$$\int_0^\infty f(r) r^{k/2-1}\, dr = \Gamma(k/2)/\pi^{k/2}.$$

Condition (d) is simply a normalization which guarantees that (1) integrates to 1. The k-dimensional normal distribution may be obtained by setting $f(r) = (2\pi)^{-k/2}\exp(-\frac{1}{2}r)$.

Here we consider the problem of obtaining affine equivariant estimates of μ and Σ with high breakdown points. This problem is discussed in Chapter 8 of Huber (1981) and Chapter 5 of Hampel, Ronchetti, Rousseeuw and Stahel (1986). Maronna (1976) has shown that M-estimators have a breakdown point of at most $(k+1)^{-1}$ and Donoho (1982) gives a list of other affine equivariant estimators and shows that they also have breakdown points of at most $(k+1)^{-1}$.

An affine equivariant estimator with an asymptotic breakdown point of 0.5 was given by Stahel (1981) and Donoho (1982). This is discussed in Donoho and Huber (1983) and Section 5.5c of Hampel, Ronchetti, Rousseeuw and Stahel (1986), and may be briefly described as follows. For each point $X^{(j)}$, $1 \leq j \leq n$, of the sample of size n the remaining $n-1$ points are projected onto the rays

Received August 1985; revised December 1986.

AMS 1980 *subject classifications*. Primary 62F10, 62E20; secondary 62H99.

Key words and phrases. S-estimators, multivariate location and dispersion parameters, consistency, asymptotic normality, finite sample breakdown points, random search.

through $X^{(j)}$. That ray for which $X^{(j)}$ is in some sense most outlying is determined and $X^{(j)}$ is then assigned a weight ω_j which is a decreasing function of its outlyingness. The weights ω_j are to be chosen, as is possible, to be affine invariant. Robust estimates of the mean and the covariance matrix may then be obtained by taking the empirical mean and covariance matrix of the weighted sample.

Rousseeuw (1986) [also see Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 303] introduced two affine equivariant estimators of the location parameter and showed that they have a finite sample breakdown point ε_n^* given by

$$\varepsilon_n^* = \left(\left[\frac{n}{2}\right] - k + 1\right)\Big/ n, \tag{2}$$

where $[a]$ denotes the largest integer less than or equal to a. His estimators may be described as follows. His first estimator, the MVE estimator, is defined to be the centre of the minimum volume ellipsoid which covers $[n/2] + 1$ the data points. The second estimator, the MCD estimator, is defined to be the mean of those $[n/2] + 1$ data points whose empirical covariance matrix has the smallest determinant. For $k = 1$, Rousseeuw showed that both estimators were consistent and that the MCD estimator is asymptotically normally distributed with a norming factor of $n^{1/2}$. The norming factor for the MVE estimator is $n^{1/3}$ and the limiting distribution is not normal.

Rousseeuw's MVE estimator is an S-estimator in the sense of Rousseeuw and Yohai (1984). By modifying the MVE estimator by using a smooth ρ-function we extend Rousseeuw's result for $k = 1$ to general k and obtain consistency and asymptotic normality with a norming factor of $n^{1/2}$. The situation is comparable to that in robust regression where the least median of squares estimator of Rousseeuw (1984) has a norming factor of $n^{1/3}$, whereas an S-estimator with a smooth ρ-function has one of $n^{1/2}$.

We define S-estimators of μ and Σ as follows. We denote by $\kappa\colon R_+ \to [0,1]$ a nonincreasing left continuous function with the properties

$$\kappa(0) = 1, \tag{3}$$

$$\kappa \text{ is continuous at } 0, \tag{4}$$

$$\kappa(u) > 0, \qquad 0 \le u < c, \tag{5}$$

and

$$\kappa(u) = 0, \qquad u > c, \tag{6}$$

for some $c > 0$.

Our estimators $\mu^{(n)}, \Sigma^{(n)}$ of μ and Σ are defined to be a solution (α^*, A^*) of the following minimization problem which we denote by $\mathscr{P}_n$.

Choose $\alpha \in R^k$ and $A \in \mathrm{PDS}(k)$ so as to minimize $|A|$ subject to

$$\frac{1}{n}\sum_1^n \kappa\Big((X^{(\nu)} - \alpha)^t A^{-1}(X^{(\nu)} - \alpha)\Big) \ge 1 - \varepsilon, \tag{7}$$

where

$$1 - \varepsilon = E\Big(\kappa\big((X-\mu)^t\Sigma^{-1}(X-\mu)\big)\Big) \tag{8}$$
$$= \frac{\pi^{k/2}}{\Gamma(k/2)}\int_0^\infty \kappa(r) f(r) r^{k/2-1}\,dr.$$

We shall show later that for $n(1-\varepsilon) \ge k+1$, $\mathscr{P}_n$ has almost surely at least one solution (α^*, A^*) with $|A^*| > 0$. The MVE estimator of Rousseeuw may be obtained by setting $\kappa(u) = 1$, $0 \le u \le c$, and $\kappa(u) = 0$, $u > c$, and then choosing c such that $1 - \varepsilon = ([n/2]+1)/n$.

We note that if we transform the data by setting $Y^{(\nu)} = BX^{(\nu)} + \gamma$, $1 \le \nu \le n$, where $\gamma \in R^k$ and B is a nonsingular $k \times k$-matrix, then $(B\alpha^* + \gamma, BA^*B^t)$ is a solution of $\mathscr{P}_n$ for the Y-data for every solution (α^*, A^*) of $\mathscr{P}_n$ for the X-data. In this sense the S-estimator is, indeed, affine equivariant.

If κ were taken to be continuous then we could replace (7) by

$$\frac{1}{n}\sum_1^n \kappa\big((X^{(\nu)}-\alpha)^t A^{-1}(X^{(\nu)}-\alpha)\big) = 1 - \varepsilon. \tag{9}$$

This would, however, exclude the MVE estimator of Rousseeuw.

In order to obtain consistent estimates of μ and Σ it is necessary to impose restrictions on f. We shall say that two real-valued nonincreasing functions ξ and g defined on a common nondegenerate interval I have a common point of decrease d if d is an interior point of I and

$$\xi(u) > \xi(d) > \xi(v) \quad \text{and} \quad g(u) > g(d) > g(v),$$

for all u, v in I with $u < d < v$. We assume

$$f \text{ is nonincreasing} \tag{10}$$

and

$$\kappa \text{ and } f \text{ have at least one common point of decrease } d_0 > 0. \tag{11}$$

In particular, this implies that $\kappa(u)f(u) > 0$ for $0 \le u < d_0$ so that $1 - \varepsilon$ in (8) is strictly positive.

Notation will be introduced as necessary, but we will adapt the following conventions: $c, c_0, c_1, \dots, d_0, d_1, \dots, e_0, e_1, \dots, r, s, u$ and v will denote real numbers and $\alpha, \beta, \gamma, \mu, x$ and y points in R^k with components α_i through to y_i, $1 \le i \le k$. A and Σ will denote positive definite symmetric $k \times k$ matrices and we will denote by $\Sigma^{1/2}$ the symmetric positive definite $k \times k$ matrix which satisfies $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$. Λ and Δ will denote positive definite $k \times k$ diagonal matrices with diagonal elements $(\lambda_j)_1^k$ and $(\delta_j)_1^k$, respectively. These will also be regarded as points in $R_+^k = [0, \infty)^k$ and denoted by λ and δ, respectively. We write $\Lambda = \operatorname{diag}(\lambda)$ and $\Delta = \operatorname{diag}(\delta)$. The identity matrix will be denoted by I_k and Lebesgue measure in R^k by $m^{(k)}$. $\|\ \|$ will denote the usual Euclidean norm in R^k and $\{x: \cdots\}$ or $\{\cdots\}$ will denote the set of points x with the property $\cdots$ as well as the indicator function of this set.

The paper is organized as follows. In Section 2 we show that the S-estimator is well defined and strongly consistent. Under additional assumptions on κ we show that the S-estimator is asymptotically normally distributed and give the limiting distribution. It is to be noted that the MVE of Rousseeuw is not covered by this result. In Section 4 we discuss the finite sample breakdown point and show that an appropriate choice of κ and ε leads to the highest possible breakdown point for affine equivariant estimators. Finally, in Section 5 we consider the problem of calculating the MVE estimator of Rousseeuw and give the results of a simple simulation.

2. Consistency. Let P_n denote the empirical measure induced by the sample $(X^{(1)},\dots, X^{(n)})$. Then we may write (7) in the form

$$\int \kappa\big((x-\alpha)^t A^{-1}(x-\alpha)\big)\, dP_n \geq 1-\varepsilon, \tag{12}$$

which in the limit as n tends to infinity yields

$$|\Sigma|^{-1/2}\int \kappa\big((x-\alpha)^t A^{-1}(x-\alpha)\big) f\big((x-\mu)^t\Sigma^{-1}(x-\mu)\big)\, dx \geq 1-\varepsilon. \tag{13}$$

We denote the problem of choosing $\alpha \in R^k$ and $A \in \mathrm{PDS}(k)$ so as to minimize $|A|$, subject to (13), by $\mathscr{P}$. A typical consistency proof now involves showing that $\mathscr{P}$ has the unique solution $(\alpha^*, A^*) = (\mu, \Sigma)$ and then using a uniform strong law for the empirical measures P_n to show that any solutions $(\mu^{(n)}, \Sigma^{(n)})$ of $\mathscr{P}_n$ satisfy $\lim_{n\to\infty}(\mu^{(n)}, \Sigma^{(n)}) = (\mu, \Sigma)$ almost surely [see Chapter II of Pollard (1984)]. We therefore turn to the problem of showing that $\mathscr{P}$ has the unique solution $(\alpha^*, A^*) = (\mu, \Sigma)$.

On transforming the integral in (13), using the transformation $x \to \mu + \Sigma^{1/2}x$, we obtain

$$\int \kappa\big((x-\gamma)^t \Sigma^{1/2}A^{-1}\Sigma^{1/2}(x-\gamma)\big) f(x^t x)\, dx \geq 1-\varepsilon, \tag{14}$$

with $\gamma = \Sigma^{-1/2}(\alpha - \mu)$.

Let O be a $k \times k$ orthogonal matrix such that $O^t\Sigma^{-1/2}A\Sigma^{-1/2}O = \Lambda = \mathrm{diag}(\lambda)$. We now transform the integral in (14) using the transformation $x \to Ox$ to obtain

$$\int \kappa\big((x-\beta)^t \Lambda^{-1}(x-\beta)\big) f(x^t x)\, dx \geq 1-\varepsilon, \tag{15}$$

with $\beta = O^t\Sigma^{-1/2}(\alpha-\mu)$.

Let $\mathscr{P}_T$ denote the problem of choosing $\beta \in R^k$ and a positive definite diagonal matrix Λ so as to minimize $|\Lambda|$ subject to (15). From the preceding, it follows that any solution (α^*, A^*) of $\mathscr{P}$ is of the form $(\mu + \Sigma^{1/2}O\beta^*, \Sigma^{1/2}O\Lambda^*O^t\Sigma^{1/2})$, with O an orthogonal matrix and (β^*, λ^*) a solution of $\mathscr{P}_T$. Thus, if $\mathscr{P}_T$ has the unique solution $(0, I_k)$, it follows that $\mathscr{P}$ has the unique solution (μ, Σ). We therefore consider the problem $\mathscr{P}_T$.

For $(\beta, \lambda) \in R^k \times R^k_+$, we define the function $H\colon R^k \times R^k_+ \to R$ by

$$H(\beta, \lambda) = \begin{cases} \int \kappa\big((x-\beta)^t \Lambda^{-1}(x-\beta)\big) f(x^t x)\, dx, & \text{if } |\Lambda| > 0, \\ 0, & \text{if } |\Lambda| = 0. \end{cases}$$

LEMMA 1. *$H\colon R^k \times R^k_+ \to R$ is continuous.*

PROOF. Let $\{d_j\colon j \in J\}$, denote the set of discontinuity points of κ. J is at most countable and may be empty. We write for $j \in J$ and $\lambda \in R^k_+$ with $\prod_1^k \lambda_i > 0$,

$$N_j(\beta, \lambda) = \big\{x\colon (x-\beta)^t \Lambda^{-1}(x-\beta) = d_j\big\}$$

and

$$N(\beta, \lambda) = \bigcup_{j \in J} N_j(\beta, \lambda).$$

As $m^{(k)}(N_j(\beta, \lambda)) = 0$, we have $m^{(k)}(N(\beta, \lambda)) = 0$. Let $(\beta, \lambda) \in R^k \times R^k_+$ with $\prod_1^k(\lambda_i > 0$ and let $(\lambda^{(n)}, \beta^{(n)}) \in R^k \times R^k_+$, $1 \le n < \infty$, satisfy $\lim_{n\to\infty}(\beta^{(n)}, \lambda^{(n)}) = (\beta, \lambda)$. Then

$$\lim_{n\to\infty} \kappa\big((x-\beta^{(n)})^t \Lambda^{(n)-1}(x-\beta^{(n)})\big) = \kappa\big((x-\beta)^t \Lambda^{-1}(x-\beta)\big), \tag{16}$$

for all $x \notin N(\beta, \lambda)$ and, hence, for almost all x. A straightforward application of dominated convergence now yields $\lim_{n\to\infty} H(\beta^{(n)}, \lambda^{(n)}) = H(\beta, \lambda)$.

If $\prod_1^k \lambda_i = 0$ and, say $\lambda_1 = 0$, then $\lim_{n\to\infty} \lambda_1^{(n)} = 0$ and again the result follows from dominated convergence. □

Suppose (β^*, λ^*) is a solution of $\mathscr{P}_T$ with

$$\int \kappa\big((x-\beta^*)^t \Lambda^{*-1}(x-\beta^*)\big) f(x^t x)\, dx > 1 - \varepsilon.$$

Then, from Lemma 1 and the fact that κ is nonincreasing with $\lim_{u\to\infty}\kappa(u) = 0$, it follows that for some $\eta > 0$,

$$\int \kappa\big((1+\eta)(x-\beta^*)^t \Lambda^{*-1}(x+\beta^*)\big) f(x^t x)\, dx = 1 - \varepsilon$$

and, hence, $(\alpha^*, (1+\eta)^{-1}\Lambda^*)$ satisfies (15) with $|(1+\eta)^{-1}\Lambda^*| < |\Lambda^*|$, a contradiction. Thus, any solution (β^*, λ^*) of $\mathscr{P}_T$ must satisfy $H(\beta^*, \lambda^*) = 1 - \varepsilon$ and we may, therefore, reduce the problem $\mathscr{P}_T$ to that of choosing $(\beta, \lambda) \in R^k \times R^k_+$ so as to minimize $\prod_1^k \lambda_j$ subject to

$$\int \kappa\big((x-\beta)^t \Lambda^{-1}(x-\beta)\big) f(x^t x)\, dx = 1 - \varepsilon. \tag{17}$$

We now show that $\mathscr{P}_T$ has at least one solution (β^*, Λ^*). We require

LEMMA 2. *Let* $(\beta, \lambda) \in R^k \times R^k_+$ *and let*

$$A_c(\beta, \lambda) = \int_{(x-\beta)^t \Lambda^{-1}(x-\beta) \le c} f(x^t x)\, dx.$$

Then there exist constants c_1, c_2 *and* c_3, *independent of* β *and* λ *such that*

(i) $A_c(\beta, \lambda) \le c_2 |\Lambda|^{1/2}$,
(ii) $A_c(\beta, \lambda) \le c_3 |\Lambda|^{1/2} f(\frac{1}{4}\beta_j^2)$ *if* $|\beta_j| \ge 2\sqrt{\lambda_j c}$ *and*
(iii) $A_c(\beta, \lambda) \le c_4 \lambda_j^{1/2}$ *for all* j, $1 \le j \le k$.

PROOF. As f is nonincreasing we have

$$A_c(\beta, \lambda) \le f(0) \int_{(x-\beta)^t \Lambda^{-1}(x-\beta) \le c} dx = c_1 |\Lambda|^{1/2},$$

proving the first of the inequalities.

To prove the second, we note that if $|\beta_j| > 2\sqrt{\lambda_j c}$ then $f(x^t x) \le f(\frac{1}{4}\beta_j^2)$ on $\{x\colon (x-\beta)^t \Lambda^{-1}(x-\beta) \le c\}$. The result now follows as before.

Finally, to prove the third inequality we note that

$$\begin{aligned} A_c(\beta, \lambda) &\le \int_{|x_j - \beta_j| \le \sqrt{\lambda_j c}} f\left(x_j^2 + \sum_{\substack{1 \\ i \ne j}}^{k} x_i^2 \right) dx_1 \cdots dx_k \\ &\le 2\sqrt{\lambda_j c} \int f\left(\sum_{i=1}^{k-1} x_i^2 \right) dx_1 \cdots dx_{k-1} \\ &= c_3 \sqrt{\lambda_j}. \end{aligned}$$

□

LEMMA 3. $\mathscr{P}_T$ *has at least one solution* (β^*, λ^*) *and the minimum satisfies*

$$0 \le \Pi^* = \prod_1^k \lambda_j^* \le 1. \tag{18}$$

PROOF. As $(0, I_k)$ satisfies (17), it is sufficient to consider minimizing $\prod_1^k \lambda_j$ on the set

$$\mathscr{S} = \left\{ (\beta, \lambda)\colon (\beta, \lambda) \text{ satisfies (17) and } \prod_1^k \lambda_j \le 1 \right\}. \tag{19}$$

As $(0, I_k)$ belongs to $\mathscr{S}$, $\mathscr{S}$ is not empty. From (11), (15), (3) and Lemma 2(iii) we obtain

$$1 - \varepsilon \le H(\beta, \lambda) \le A_c(\beta, \lambda) \le \lambda_j^{1/2},$$

which, in turn, implies

$$0 < c_4 \le \lambda_j \le c_5 < \infty \tag{20}$$

on $\mathscr{S}$.

From (20) we conclude $\prod_1^k \lambda_j \ge c_6 > 0$ on $\mathscr{S}$ and this, together with the second inequality of Lemma 2 and the fact that $\lim_{r \to \infty} f(r) = 0$, implies

$\|\beta\| \le c_7 < \infty$ on $\mathscr{S}$. Thus, $\mathscr{S}$ is a bounded set. However, Lemma 1 implies that it is a closed set and, hence, it is compact. As $\prod_1^k \lambda_j$ is a continuous function of (β, λ) it attains its minimum at some point (β^*, λ^*) in $\mathscr{S}$ and, thus, $\mathscr{S}_T$ has at least one solution. (18) follows from (19). □

Consider now the problem of maximizing $H(\beta, \lambda)$ subject to $\prod_1^k \lambda_j = \Pi^* = \prod_1^k \lambda_j^*$, where (β^*, λ^*) is a solution of $\mathscr{P}_T$. We denote this problem by $\mathscr{P}_T'$. If there exists a $(\tilde{\beta}, \tilde{\lambda})$ with $H(\tilde{\beta}, \tilde{\lambda}) > H(\beta^*, \lambda^*)$ and $\prod_1^k \tilde{\lambda}_j = \Pi^* > 0$, it follows from the continuity of H and the fact that H is a nondecreasing function of λ that there exists a $\tilde{\lambda}^*$ such that $H(\tilde{\beta}, \tilde{\lambda}^*) = H(\beta^*, \lambda^*) = 1 - \varepsilon$ and $|\tilde{\Lambda}^*| < |\Lambda^*| = \Pi^*$. This, however, contradicts the fact that (β^*, λ^*) is a solution of $\mathscr{P}_T$. Thus, (β^*, λ^*) is also a solution of $\mathscr{P}_T'$ and it now suffices to show that $\mathscr{P}_T'$ has the unique solution $(0, I_k)$. To this end we first show that if (β^*, λ^*) is a solution of $\mathscr{P}_T'$, then so is $(0, \lambda^*)$. This follows from the next lemma.

LEMMA 4. *Let $(\beta, \lambda) \in R^k \times R_+^k$ with $|\Lambda| > 0$ and ξ and g: $R_+ \to R_+$ be nonincreasing functions such that*

$$\int g(x^t x)\, dx < \infty.$$

Then

$$\int \xi\big((x-\beta)^t \Lambda^{-1}(x-\beta)\big) g(x^t x)\, dx \le \int \xi(x^t \Lambda^{-1} x) g(x^t x)\, dx. \tag{21}$$

Furthermore, if $\Lambda = I_k$ and ξ and g have at least one common point of decrease, the inequality is strict unless $\beta = 0$.

PROOF. For fixed $(x_j)_2^k$ and $(\beta_j)_2^k$ we define $\tilde{\xi}$ and $\tilde{g}$ by

$$\tilde{\xi}(u) = \xi\left(u + \sum_2^k (x_j - \beta_j)^2/\lambda_j\right)$$

and

$$\tilde{g}(u) = g\left(u + \sum_2^k (x_j - \beta_j)^2\right).$$

Then $\tilde{\xi}$ and $\tilde{g}$ are nonincreasing functions and, hence,

$$\left(\tilde{\xi}\big((x_1-\beta_1)^2/\lambda_1\big) - \tilde{\xi}\big(x_1^2/\lambda_1\big)\right)\left(\tilde{g}\big((x_1-\beta_1)^2\big) - \tilde{g}\big(x_1^2\big)\right) \ge 0$$

for all x_1. On integrating this inequality with respect to x_1 we obtain

$$\int \tilde{\xi}\big(x_1^2/\lambda_1\big)\tilde{g}\big((x_1-\beta_1)^2\big)\, dx_1 + \int \tilde{\xi}\big((x_1-\beta_1)^2/\lambda_1\big)\tilde{g}\big(x_1^2\big)\, dx_1$$

$$\le 2\int \tilde{\xi}\big(x_1^2/\lambda_1\big)\tilde{g}\big(x_1^2\big)\, dx_1.$$

On using the transformation $x_1 \to -x_1 + \beta$ in the first integral, we conclude

$$\int \tilde{\xi}\big((x_1-\beta_1)^2/\lambda_1\big)\tilde{g}\big(x_1^2\big)\, dx_1 \le \int \tilde{\xi}\big(x_1^2/\lambda_1\big)\tilde{g}\big(x_1^2\big)\, dx_1,$$

which together with Fubini's theorem implies

$$\int \xi\big((x-\beta)^t\Lambda^{-1}(x-\beta)\big)g(x^tx)\,dx \le \int \xi\big((x-\beta^{(1)})^t\Lambda^{-1}(x-\beta^{(1)})\big)g(x^tx)\,dx,$$

where $\beta_1^{(1)} = 0$ and $\beta_j^{(1)} = \beta_j$, $2 \le j \le k$. The first claim of the lemma now follows on applying the same argument to the remaining $k-1$ components.

If $\Lambda = I_k$ we have, directly,

$$(22)\quad \Big(\xi\big((x-\beta)^t(x-\beta)\big) - \xi(x^tx)\Big)\Big(g\big((x-\beta)^t(x-\beta)\big) - g(x^tx)\Big) \ge 0,$$

for all x. Suppose $\beta \ne 0$, let $u_0 > 0$ be a common point of decrease of ξ and g and choose $u > 0$ such that

$$u^2\beta^t\beta < u_0 < (1+u)^2\beta^t\beta.$$

On setting $x(u) = -u\beta$ we see that $x^tx < u_0 < (x-\beta)^t(x-\beta)$ in some neighbourhood of $x(u)$. This implies that strict inequality in (22) holds on a set of positive measure, which yields strict inequality in (21). □

Lemma 4 shows that $H(\beta, \lambda) \le H(0, \lambda)$ and, hence, we can provisionally restrict our search for a solution of $\mathscr{P}_T'$ to the problem of maximizing $H(0,\lambda)$ subject to $\Pi_1^k\lambda_j = \Pi^* > 0$. It turns out that this is a problem of majorization as treated in Marshall and Olkin (1979).

If we define $\tilde{H}$: $R^k \to R$ by

$$(23)\qquad \tilde{H}(x_1,\dots,x_k) = H(0, (\exp(x_1),\dots,\exp(x_k))),$$

we see that we must maximize $\tilde{H}$ subject to $\Sigma_1^k x_j = \log \Pi^*$. From

$$H(0,\lambda) = \int \kappa(x^t\Lambda^{-1}x)f(x^tx)\,dx,$$

it follows that $H(0,\lambda)$ is a symmetric function of $\lambda_1,\dots,\lambda_k$ and, hence, $\tilde{H}$ is a symmetric function of $x_1,\dots,x_k$. We shall now show that $\tilde{H}$ is Schur-convex in the following sense. Let $\Lambda = \operatorname{diag}(\lambda)$ and $\Delta = \operatorname{diag}(\delta)$ be two diagonal matrices with $|\Lambda| = |\Delta|$. We write $\lambda \precsim \delta$ if $\Pi_{j=1}^l\lambda_{(j)} \le \Pi_{j=1}^l\delta_{(j)}$, $1 \le l \le k$, where $\lambda_{(1)} \le \cdots \le \lambda_{(k)}$ and $\delta_{(1)} \le \cdots \le \delta_{(k)}$ denote $\lambda_1,\dots,\lambda_k$ and $\delta_1,\dots,\delta_k$ in increasing order. This may be interpreted as the ellipsoid given by $\{x\colon x^t\Lambda x \le 1\}$ being less disperse than that given by $\{x\colon x^t\Delta x \le 1\}$. Then the Schur-convexity of $\tilde{H}$ is equivalent to $H(0,\lambda) \le H(0,\delta)$.

It follows from Marshall and Olkin [(1979), Section A.5, page 58] that it is sufficient to show that for fixed $x_3,\dots,x_k$ the function $\tilde{H}(x_1, x_2, x_3,\dots,x_k)$ as a function of (x_1, x_2) is Schur-convex. We, therefore, restrict ourselves to $k = 2$ for the next few lemmas.

LEMMA 5. *Let* $\Lambda = \operatorname{diag}(u, u^{-1})$ *and* $\Delta = \operatorname{diag}(v, v^{-1})$ *with* $0 < u < v \le 1$ *be two diagonal matrices. Then* $x^t\Lambda x \le x^t\Delta x$ *if and only if* $(1+uv)/(u+v)x^t\Delta x \le x^tx$.

PROOF. We have $x^t\Lambda x \le x^t\Delta x$ if and only if $v^{-1}(v-u)x_2^2 \le u(v-u)x_1^2$. As $v > u$ this implies

(24) $$x^t\Lambda x \le x^t\Delta x \quad \text{if and only if} \quad x^t\Delta x \le (u+v)x_1^2.$$

Now $x^tx = vx^t\Delta x + (1-v^2)x_1^2$ and consequently $x_1^2 = (x^tx - vx^t\Delta x)/(1-v^2)$.

On using this in the second inequality of (24) we obtain the claim of the lemma. □

LEMMA 6. *Let $\Lambda = \text{diag}(u, u^{-1})$ be a diagonal matrix and ξ and g: $R_+ \to R_+$ nonincreasing functions satisfying $\int g(x^tx)\,dx < \infty$. Then*

(25) $$\int \xi(x^t\Lambda x)g(x^tx)\,dx \le \int \xi(x^tx)g(x^tx)\,dx$$

with strict inequality if $u \ne 1$ and ξ and g have at least one common point of decrease.

PROOF. As ξ and g are nonincreasing we have for all x,

(26) $$\big(\xi(x^t\Lambda x) - \xi(x^tx)\big)\big(g(x^t\Lambda x) - g(x^tx)\big) \ge 0.$$

On integrating we obtain

$$\int \xi(x^t\Lambda x)g(x^tx)\,dx + \int \xi(x^tx)g(x^t\Lambda x)\,dx$$
$$\le \int \xi(x^t\Lambda x)g(x^t\Lambda x)\,dx + \int \xi(x^tx)g(x^tx)\,dx.$$

If we now transform the second and third integrals, using the transformation $x \to \Lambda^{-1/2}x$, we obtain

$$\int \xi(x^t\Lambda x)g(x^tx)\,dx + \int \xi(x^t\Lambda^{-1}x)g(x^tx)\,dx \le 2\int \xi(x^tx)g(x^tx)\,dx.$$

Now, the first two integrals are, in fact, equal as may be seen by considering the transformation $(x_1, x_2) \to (x_2, x_1)$. This gives the desired inequality.

To prove the second part we may suppose without loss of generality that $0 < u < 1$. Then on writing $u = (1+2\eta)^{-1}$ with $\eta > 0$ and on setting $x(\eta)((d_1(1+\eta))^{1/2}, 0)$ where $d_1 > 0$ is the common point of decrease we see that $x(\eta)^t\Lambda x(\eta) < d_1 < x(\eta)^tx(\eta)$. Thus, $x^t\Lambda x < d_1 < x^tx$ in some neighbourhood of $x(\eta)$, which implies that strict inequality in (26) holds on a set of positive Lebesgue measure. This, in turn, implies strict inequality in (25). □

LEMMA 7. *Let $\Lambda = \text{diag}(u, u^{-1})$ and $\Delta = \text{diag}(v, v^{-1})$ with $0 < u \le v \le 1$ and let ξ and g: $R_+ \to R_+$ be nonincreasing and satisfy $\int g(x^tx)\,dx < \infty$. Then*

(27) $$\int \xi(x^t\Lambda x)g(x^tx)\,dx \le \int \xi(x^t\Delta x)g(x^tx)\,dx.$$

PROOF. We need only consider the case $u < v$. From Lemma 5 and the fact that ξ and g are nonincreasing we obtain

$$\left(\xi(x^t\Lambda x) - \xi(x^t\Delta x)\right)\left(g\left(\frac{1+uv}{u+v}x^t\Delta x\right) - g(x^tx)\right) \ge 0.$$

On integrating we deduce

$$\begin{aligned}&\int\left(\xi(x^t\Lambda x) - \xi(x^t\Delta x)\right)g(x^tx)\,dx\\ &\quad\le \int\left(\xi(x^t\Lambda x) - \xi(x^t\Delta x)\right)g\left(\frac{1+uv}{u+v}x^t\Delta x\right)dx\\ &\quad= \int\left(\xi(x^t\Delta^{-1/2}\Lambda\Delta^{-1/2}x) - \xi(x^tx)\right)g\left(\frac{1+uv}{u+v}x^tx\right)dx\\ &\quad\le 0,\end{aligned}$$

where we have made the transformation $x \to \Delta^{-1/2}x$ and then applied Lemma 6. □

We can now prove that $\tilde{H}$ of (23) is Schur-convex.

LEMMA 8. *Let* $\Lambda = \operatorname{diag}(\lambda)$ *and* $\Delta = \operatorname{diag}(\delta)$ *be positive definite diagonal matrices with* $\lambda \lesssim \delta$. *Then* $H(0,\lambda) \le H(0,\delta)$.

PROOF. As mentioned above $H(0,\lambda)$ is a symmetric function of λ and, hence, it is sufficient to prove $H(0,\lambda) \le H(0,\tilde{\lambda})$, where $\lambda_1 \le \tilde{\lambda}_1 \le \tilde{\lambda}_2 \le \lambda_2$, $\lambda_1\lambda_2 = \tilde{\lambda}_1\tilde{\lambda}_2$ and $\tilde{\lambda}_j = \lambda_j$, $3 \le j \le k$. We write $\eta = \lambda_1\lambda_2 = \tilde{\lambda}_1\tilde{\lambda}_2$, $u = \sqrt{\lambda_1/\lambda_2}$ and $v = \sqrt{\tilde{\lambda}_1/\tilde{\lambda}_2}$. Then

$$\begin{aligned}&\iint\kappa\left(\lambda_1x_1^2 + \lambda_2x_2^2 + \sum_{j=3}^k\lambda_jx_j^2\right)g\left(x_1^2 + x_2^2 + \sum_{j=3}^k x_j^2\right)dx_1\,dx_2\\ &\quad= \iint\tilde{\kappa}\left(ux_1^2 + u^{-1}x_2^2\right)\tilde{g}\left(x_1^2 + x_2^2\right)dx_1\,dx_2\\ &\quad\le \iint\tilde{\kappa}\left(vx_1^2 + v^{-1}x_2^2\right)\tilde{g}\left(x_1^2 + x_2^2\right)dx_1\,dx_2\\ &\quad= \iint\kappa\left(\tilde{\lambda}_1x_1^2 + \tilde{\lambda}_2x_2^2 + \sum_{j=3}^k\tilde{\lambda}_jx_j^2\right)g\left(x_1^2 + x_2^2 + \sum_{j=3}^k\lambda_jx_j^2\right)dx_1\,dx_2,\end{aligned}$$

where we have written

$$\tilde{\kappa}(s) = \kappa\left(\eta^{1/2}s + \sum_{j=3}^k x_j^2\right),$$

$$\tilde{g}(s) = g\left(s + \sum_{j=3}^k x_j^2\right)$$

and then applied Lemma 7. On integrating the above inequality with respect to $x_3, \dots, x_k$ we obtain the desired result. □

We can now prove

THEOREM 1. *The problem of choosing* $\alpha \in R^k$ *and* $A \in \mathrm{PDS}(k)$ *so as to minimize* $|A|$ *subject to*

$$|\Sigma|^{1/2} \int \kappa\big((x-\alpha)^t A^{-1}(x-\alpha)\big) f\big((x-\mu)^t \Sigma^{-1}(x-\mu)\big)\, dx = 1 - \varepsilon$$

has the unique solution $(\alpha^*, A^*) = (\mu, \Sigma)$.

PROOF. Let (β^*, λ^*) be a solution of $\mathscr{P}_T$ as in Lemma 3. Then,

$$1 - \varepsilon = \int \kappa\big((x-\beta^*)^t \Lambda^{-1}(x-\beta^*)\big) f(x^t x)\, dx$$

and (β^*, λ^*) is a solution of $\mathscr{P}'_T$.

However, $\lambda^* \lesssim \Pi^{*1/k} 1'_k$ with $1_k = (1, \dots, 1)$ and as $H(\beta^*, \lambda^*) = H(0, \lambda^*) \le H(0, \Pi^{*1/k} 1_k)$, it follows that $(0, \Pi^{*1/k} 1_k)$ is also a solution of $\mathscr{P}'_T$. This implies

$$1 - \varepsilon = \int \kappa(\Pi^{*1/k} x^t x) f(x^t x)\, dx. \tag{28}$$

If $\Pi^{*1/k} < 1$, then on a set of positive k-dimensional Lebesgue measure we have $\Pi^{*1/k} x^t x < d_0 < x^t x$ and $f(x^t x) > 0$, where d_0 is a common point of decrease of κ and f. This implies

$$\int \kappa(\Pi^{*1/k} x^t x) f(x^t x)\, dx < \int \kappa(x^t x) f(x^t x)\, dx = 1 - \varepsilon,$$

contradicting (27). Thus we must have $\Pi^* = 1$ and, hence, $(0, I_k)$ is a solution of $\mathscr{P}'_T$.

Let $(0, \lambda^*)$ be a solution of $\mathscr{P}'_T$ with $\lambda^* \neq 1_k$ but $\Pi^* = \Pi_1^k \lambda_j^* = 1$. Then $\lambda^* \lesssim 1_k$ and we may suppose, without loss of generality, that $\lambda_1 < 1$. We define $\tilde{\lambda}^*$ by $\tilde{\lambda}^* = (\lambda_1, \lambda_1^{-1}, 1, \dots, 1)$ so that $\lambda^* \lesssim \tilde{\lambda}^* \lesssim 1_k$.

For $x_3, \dots, x_k$ fixed we define $\tilde{\kappa}(u) = \kappa(u + \Sigma_{j=3}^k x_j^2)$ and $\tilde{f}(u) = f(u + \Sigma_{j=3}^k x_j^2)$. If $\Sigma_{j=3}^k x_j^2 < d_0$, with d_0 a common point of decrease of κ and f, then $\tilde{\kappa}$ and $\tilde{f}$ have a common point of decrease. This implies, by Lemma 6,

$$\begin{aligned} &\iint \tilde{\kappa}\big(\lambda_1 x_1^2 + \lambda_1^{-1} x_2^2\big) \tilde{f}\big(x_1^2 + x_2^2\big)\, dx_1\, dx_2 \\ &\quad < \iint \tilde{\kappa}\big(x_1^2 + x_2^2\big) \tilde{f}\big(x_1^2 + x_2^2\big)\, dx_1\, dx_2 \end{aligned} \tag{29}$$

on a set of positive $(k-2)$-dimensional Lebesgue measure. On integrating (28) with respect to $x_3, \dots, x_k$ we obtain $H(0, \tilde{\lambda}^*) < H(0, 1_k)$, which implies that $(0, \lambda^*)$ is not a solution of $\mathscr{P}'_T$.

We have now shown that every solution of $\mathscr{P}'_T$ is of the form $(\beta^*, 1_k)$. If $\beta^* \neq 0$ we may use the second part of Lemma 4 to conclude that $H(\beta^*, 1_k) <$

$H(0, 1_k)$. It follows, therefore, that $\mathscr{P}_T'$ has the unique solution $(0, 1_k)$. As every solution of $\mathscr{P}_T$ is a solution of $\mathscr{P}_T'$, we may conclude that $\mathscr{P}_T$ also has the unique solution $(0, 1_k)$, which in turn implies that $\mathscr{P}$ has the unique solution (μ, Σ) as was to be proved. □

THEOREM 2. *If* $k + 1 \le n(1 - \varepsilon)$ *then the problem* $\mathscr{P}_n$ *has at least one solution with probability* 1.

PROOF. From (3), (4) and (6) it follows that

$$\lim_{v \to \infty} \frac{1}{n} \sum_1^n \kappa\left(v^{-1} X^{(\nu)t} X^{(\nu)}\right) = 1$$

and

$$\lim_{v \to 0} \frac{1}{n} \sum_1^n \kappa\left(v^{-1} X^{(\nu)t} X^{(\nu)}\right) = 0,$$

almost surely. Thus, there exists a v_0, $0 < v_0 < \infty$, such that

$$\frac{1}{n} \sum_1^n \kappa\left(v_0^{-1} X^{(\nu)t} X^{(\nu)}\right) \ge 1 - \varepsilon$$

and we define

$$\mathscr{S}_n = \left\{(\alpha, A): \alpha \in R^k, A \in \mathrm{PDS}(k), (\alpha, A) \text{ satisfies (7) with } |A| \le v_0^k + 1\right\}.$$

We next show that $\mathscr{S}_n$ is contained in a compact subset of $R^{k/2(k+1)+k}$. For (α, A) in $\mathscr{S}_n$ we have

$$\frac{1}{n} \sum_1^n \kappa\left(\left(X^{(\nu)} - \alpha\right)^t A^{-1}\left(X^{(\nu)} - \alpha\right)\right) \ge 1 - \varepsilon$$

and, hence,

$$(30) \qquad \left(X^{(\nu)} - \alpha\right)^t A^{-1}\left(X^{(\nu)} - \alpha\right) \le c$$

for at least $n(1 - \varepsilon) \ge k + 1$ of the $X^{(\nu)}$'s. The ellipsoid, defined by

$$(31) \qquad \left\{x: (x - \alpha)^t A^{-1}(x - \alpha) \le c\right\},$$

is convex and, hence, contains the convex hull of at least $k + 1$ of the $X^{(\nu)}$'s.

However, as the $X^{(\nu)}$'s are independent and have a density function, it follows that, with probability 1, no $k + 1$ $X^{(\nu)}$'s lie on a $(k - 1)$-dimensional hyperplane. In particular, the convex hull of any $k + 1$ of the $X^{(\nu)}$'s contains an open sphere of positive radius with probability 1. For each subset of size $k + 1$ of the $X^{(\nu)}$'s we choose such a sphere and denote by U_1, $0 < U_1 < \infty$, the minimum of the radii. Then the ellipsoid (31) contains an open sphere of radius U_1 and, hence, we may conclude

$$\{x: \|x - \alpha\| < U_1\} \subset \left\{x: (x - \alpha)^t A^{-1}(x - \alpha) \le c\right\}.$$

Let $a_{(1)}$ denote the smallest eigenvalue of A. Then $a_{(1)}^{-1}$ is the largest eigenvalue of A^{-1} and we have [Rao (1973), page 62],

$$a_{(1)}^{-1} = \max_{\|x-\alpha\|=\frac{1}{2}U_1} \frac{(x-\alpha)^t A^{-1}(x-\alpha)}{\left(\frac{1}{2}U_1\right)^2} \le \frac{4c}{U_1^2},$$

so that $a_{(1)} \ge \frac{1}{4}U_1^2/c = U_2 > 0$. This together with $|A| \le v_0^k + 1$ yields

$$0 < U_2 \le a_{(j)} \le U_3 < \infty \tag{32}$$

for the eigenvalues $(a_{(j)})_1^k$ of any A for which (α, A) lies in $\mathscr{S}_n$ for some $\alpha \in R^k$. For such a pair (α, A) we have

$$\frac{1}{n}\sum_1^n \kappa\left((X^{(\nu)}-\alpha)^t A^{-1}(X^{(\nu)}-\alpha)\right) \le \frac{1}{n}\sum_1^n \kappa\left(U_3^{-1}(X^{(\nu)}-\alpha)^t(X^{(\nu)}-\alpha)\right)$$

and as

$$\lim_{|\alpha|\to\infty} \frac{1}{n}\sum_1^n \kappa\left(U_3^{-1}(X^{(\nu)}-\alpha)^t(X^{(\nu)}-\alpha)\right) = 0,$$

it follows that (32) and

$$\|\alpha\| \le U_4 \tag{33}$$

hold for any pair (α, A) in $\mathscr{S}_n$. This proves that $\mathscr{S}_n$ is contained in a compact subset of $R^{k+k/2(k+1)}$.

We define

$$D_n = \inf\left\{|A|: (\alpha, A) \in \mathscr{S}_n \text{ for some } \alpha \in R^k\right\}.$$

To show that $\mathscr{P}_n$ has a solution, it is necessary to show that there exists a pair (α^*, A^*) with $A^* \in \mathrm{PDS}(k)$ which satisfies (7) and $|A^*| = D_n$.

As $\mathscr{S}_n$ is contained in a compact set it follows that we can find a convergent subsequence $((\alpha^{(m)}, A^{(m)}))_1^\infty$ in $\mathscr{S}_n$ with

$$\lim_{m\to\infty} |A^{(m)}| = D_n$$

and

$$\alpha^* = \lim_{m\to\infty} \alpha^{(m)}, \qquad A^* = \lim_{m\to\infty} A^{(m)}. \tag{34}$$

It is clear that $A^* \in \mathrm{PDS}(k)$.

From (34) it follows that there exists a sequence $(\eta_m)_1^\infty$ satisfying $\eta_m > 0$, $1 \le m < \infty$, $\lim_{m\to\infty}\eta_m = 0$ and

$$(1-\eta_m)^{-1}(X^{(\nu)}-\alpha^*)^t A^{*-1}(X^{(\nu)}-\alpha^*) \le (X^{(\nu)}-\alpha^{(m)})^t A^{(m)^{-1}}(X^{(\nu)}-\alpha^{(m)})$$

for $1 \le \nu \le n$.

We have

$$\frac{1}{n}\sum_1^n \kappa\Big((1+\eta_m)^{-1}(X^{(\nu)}-\alpha^*)^t A^{*-1}(X^{(\nu)}-\alpha^*)\Big)$$
$$\geq \frac{1}{n}\sum_1^n \kappa\Big((X^{(\nu)}-\alpha^{(m)})^t A^{(m)^{-1}}(X^{(\nu)}-\alpha^{(m)})\Big)$$
$$\geq 1-\varepsilon,$$

so that $(\alpha^*, (1+\eta_n)A^*)$ satisfies (7) for all m. Using the fact that κ is continuous on the left we may deduce

$$1-\varepsilon \leq \lim_{m\to\infty}\frac{1}{n}\sum_1^n \kappa\Big((1+\eta_m)^{-1}(X^{(\nu)}-\alpha^*)^t A^{*-1}(X^{(\nu)}-\alpha^*)\Big)$$
$$= \frac{1}{n}\sum_1^n \kappa\Big((X^{(\nu)}-\alpha^*)^t A^{*-1}(X^{(\nu)}-\alpha^*)\Big),$$

which implies that (α^*, A^*) satisfies (7) and is, therefore, a solution of $\mathscr{P}_n$. □

We can now prove

Theorem 3. *For $n \geq (k+1)(1-\varepsilon)^{-1}$ let $(\mu^{(n)}, \Sigma^{(n)})$ be a solution of $\mathscr{P}_n$. Then*

$$\lim_{n\to\infty}\mu^{(n)} = \mu \quad \textit{and} \quad \lim_{n\to\infty}\Sigma^{(n)} = \Sigma,$$

almost surely.

Proof. By using the transformation $X^{(\nu)} \to \Sigma^{1/2}X^{(\nu)} + \mu$ we see that it is sufficient to consider the case $\mu = 0$ and $\Sigma = I_k$.

Let P denote the measure on R^k with density function $f(x'x)$. Then according to Rao (1962) [also see Chapter 1 of Pollard (1984)], we have

$$\lim_{n\to\infty}\sup_{\substack{C\subset R^k\\ C \text{ convex}}}|P_n(C)-P(C)| = 0,$$

almost surely. On writing

$$E_n = \Big\{x\colon (x-\mu^{(n)})^t\Sigma^{(n)^{-1}}(x-\mu^{(n)}) \leq c\Big\},$$

we may conclude that $P(E_n) \geq \frac{1}{2}(1-\varepsilon)$ for all n sufficiently large, almost surely. This, together with Lemma 2, implies that the eigenvalues of the $\Sigma^{(n)}$ are bounded away from zero and the $\|\mu^{(n)}\|$ are bounded above, almost surely.

The strong law of large numbers implies

$$\text{(35)} \qquad \lim_{n\to\infty}\frac{1}{n}\sum_1^n \kappa\Big((1+\eta)^{-1}X^{(\nu)t}X^{(\nu)}\Big) = \int\kappa\Big((1+\eta)^{-1}x^tx\Big)f(x^tx)\,dx,$$

almost surely, for each $\eta > -1$. The right-hand side of (35) is a nondecreasing function of η. Furthermore, as κ and f have at least one common point of decrease, this function is strictly increasing at $\eta = 0$. Thus, by the strong law of large numbers

$$\lim_{n\to\infty} \frac{1}{n}\sum_{1}^{n} \kappa\left((1+\eta)^{-1}X^{(\nu)t}X^{(\nu)}\right) > 1 - \varepsilon,$$

almost surely, for each $\eta > 0$. In particular, we have for $\eta > 0$,

$$\frac{1}{n}\sum_{1}^{n} \kappa\left((1+\eta)^{-1}X^{(\nu)t}X^{(\nu)}\right) \geq 1 - \varepsilon$$

for all n sufficiently large, almost surely.

This implies that $(0, (1+\eta)I_k)$ satisfies (7) for all n sufficiently large and, hence, as $\Sigma^{(n)}$ is a solution of $\mathscr{P}_n$, $|\Sigma^{(n)}| \leq |(1+\eta)I_k| = (1+\eta)^k$ for all n sufficiently large. As $\eta > 0$ may be taken to be arbitrarily small we may, therefore, conclude

$$\limsup_{n\to\infty} |\Sigma^{(n)}| \leq 1, \tag{36}$$

almost surely.

We have already shown that the eigenvalues of the $\Sigma^{(n)}$ are bounded away from zero and this, together with (36), shows that they are also bounded above. As the $\|\mu^{(n)}\|$ are bounded above it follows that, with probability 1, the sequence $((\mu^{(n)}, \Sigma^{(n)}))_1^\infty$ lies in a compact subset of $R^{k+k/2(k+1)}$.

Let $(\mu^{(n_j)}, \Sigma^{(n_j)})_{j=1}^\infty$ be a convergent sequence with

$$\lim_{j\to\infty} \mu^{(n_j)} = \tilde{\mu}, \qquad \lim_{j\to\infty} \Sigma^{(n_j)} = \tilde{\Sigma},$$

almost surely. On writing

$$\tilde{\kappa}_j(x) = \kappa\left(\left(x - \mu^{(n_j)}\right)^t \Sigma^{(n_j)^{-1}}\left(x - \mu^{(n_j)}\right)\right)$$

and

$$\tilde{\kappa}(x) = \kappa\left((x - \tilde{\mu})^t \tilde{\Sigma}^{-1}(x - \tilde{\mu})\right),$$

we see that

$$\lim_{j\to\infty} \tilde{\kappa}_j(x^{(j)}) = \tilde{\kappa}(x) \tag{37}$$

for any sequence $(x^{(j)})_1^\infty$ with $\lim_{j\to\infty} x^{(j)} = x$ and any x such that $(x - \tilde{\mu})^t\tilde{\Sigma}^{-1}(x - \tilde{\mu})$ is a point of continuity of κ. As κ has at most a countable number of discontinuity points, we see that (37) holds for almost all x with respect to Lebesgue measure and, hence, with respect to P. We now apply Theorem 5.5 of Billingsley (1968). Let g: $R_+ \to R_+$ be defined by $g(u) = u$, $0 \leq u \leq 1$, and $g(u) = 1$ for $u \geq 1$. As the sequence of measures $(P_n)_1^\infty$ converges weakly to P it follows from this theorem that

$$\lim_{j\to\infty} \int g\left(\tilde{\kappa}_j(x)\right) dP_{n_j}(x) = \int g\left(\tilde{\kappa}(x)\right) dP$$

and, hence, as $0 \le \tilde{\kappa}_j(x),\ \tilde{\kappa}(x) \le 1$ for all x,

$$1 - \varepsilon \le \lim_{j\to\infty} \frac{1}{n_j} \sum_1^{n_j} \kappa\Big(\big(X^{(\nu)} - \mu^{(n_j)}\big)^t \Sigma^{(n_j)^{-1}} \big(X^{(\nu)} - \mu^{(n_j)}\big)\Big)$$

$$= \int \kappa\Big((x - \tilde{\mu})^t \tilde{\Sigma}^{-1} (x - \tilde{\mu})\Big) f(x^t x)\, dx.$$

Thus, $(\tilde{\mu}, \tilde{\Sigma})$ satisfies (13) with $\mu = 0$ and $\Sigma = I_k$. However, for this choice of μ and Σ, the problem $\mathscr{P}$ has the unique solution $(0, I_k)$ by Theorem 1. From this we deduce $|I_k| \le |\tilde{\Sigma}|$, which together with (38), implies $|\tilde{\Sigma}| = 1$. Thus, $(\tilde{\mu}, \tilde{\Sigma})$ is also a solution and Theorem 1 implies $(\tilde{\mu}, \tilde{\Sigma}) = (0, I_k)$. We have, therefore, shown that every convergent subsequence of $((\mu^{(n)}, \Sigma^{(n)}))_1^\infty$ converges almost surely to $(0, I_k)$. As the sequence $((\mu^{(n)}, \Sigma^{(n)}))_1^\infty$ is contained in a compact set, this can only be the case if the sequence itself converges to $(0, I_k)$ which proves the theorem. □

3. Asymptotic normality. In addition to the assumptions already made about κ, we now assume that κ has a continuous third derivative. For $i = 1$ and 2 we denote the ith derivative of κ by $\kappa^{(i)}$.

In order to state our results we require the following constants:

$$e_0 = (1 - \varepsilon),$$

$$e_1 = \frac{d(k)}{2k} \int_0^\infty \big(\kappa^{(1)}(r)\big)^2 f(r) r^{k/2}\, dr,$$

$$e_2 = \frac{d(k)}{2k} \int_0^\infty \big(k\kappa^{(1)}(r) + 2r\kappa^{(1)}(r)\big) f(r) r^{k/2-1}\, dr,$$

$$e_3 = \frac{d(k)}{2k} \int_0^\infty \kappa^{(1)}(r) f(r) r^{k/2}\, dr,$$

$$e_4 = \frac{d(k)}{2k(k+2)} \int_0^\infty \big(\kappa^{(2)}(r)\big) f(r) r^{k/2+1}\, dr,$$

$$e_5 = \frac{d(k)}{2(e_3 + 2e_4)^2 k(k+2)} \int_0^\infty \big(\kappa^{(1)}(r)\big)^2 f(r) r^{k/2+1}\, dr,$$

$$e_6 = -\frac{2}{k} e_5 + \frac{1}{k^2 e_3^2}\left(d(k) \int_0^\infty \kappa(r)^2 f(r) r^{k/2-1}\, dr - e_0^2\right),$$

where $d(k) = \pi^{k/2}/\Gamma(k/2)$.

THEOREM 4. *Under the above assumptions on f and κ we have*

$$\big(\sqrt{n}\,(\mu^{(n)} - \mu), \sqrt{n}\,(\Sigma^{(n)} - \Sigma)\big) \Rightarrow (Z, z),$$

where Z and z are independently distributed normal random variables with zero *mean and*

$$E(ZZ^t) = \frac{e_1}{e_2^2}\Sigma, \tag{38}$$

$$E(z_{ij}z_{pq}) = e_6\Sigma_{ij}\Sigma_{pq} + e_5(\Sigma_{ip}\Sigma_{jq} + \Sigma_{iq}\Sigma_{jp}). \tag{39}$$

PROOF. It is sufficient to consider the case $\mu = 0$ and $\Sigma = I_k$. On introducing the Langrange multiplier η_n and writing $\Sigma^{(n)} = I_k + B^{(n)}$ we see that $\mu^{(n)}$ and $B^{(n)}$ must satisfy

$$\left(\frac{\partial}{\partial\alpha}, \frac{\partial}{\partial B}, \frac{\partial}{\partial\eta}\right)F_n(\alpha, B, \eta)|_{\mu^{(n)}, B^{(n)}, \eta_n} = 0,$$

where

$$F_n(\alpha, B, \eta) = |I_k + B| + \eta\left(\sum_{\nu=1}^{n} \kappa\left((X^{(\nu)} - \alpha)^t(I_k + B)^{-1}(X^{(\nu)} - \alpha)\right) - ne_0\right).$$

The theorem may now be proved as follows. The partial derivatives are replaced by their Taylor expansions of order 2, as higher order terms may be neglected because of the consistency of the estimators. It is then a straightforward matter of applying as appropriate, either the strong law of large numbers or the central limit theorem, to those terms remaining and then solving the resulting system of linear equations. □

It was pointed out by one of the referees that (38) and (39) correspond to the general form of the covariance matrices for affine equivariant estimators.

4. Breakdown point. Consider a sample $(X^{(1)},\dots,X^{(n)}) = \mathbf{X}$ with $(1-\varepsilon)n \geq k + 1$ and in general position, that is, no more than k points of the sample lie on any $(k-1)$-dimensional hyperplane. For our estimators $\mu^{(n)}$ and $\Sigma^{(n)}$ based on $\mathbf{X}$ we set

$$T(\mathbf{X}) = \|\mu^{(n)}\| + \sum_{1}^{k}\left(\sigma_j^{(n)} + \sigma_j^{(n)^{-1}}\right), \tag{40}$$

where $\sigma_1^{(n)},\dots,\sigma_k^{(n)}$ are the eigenvalues of $\Sigma^{(n)}$.

The finite sample breakdown point ε_n^* may now be defined by

$$\varepsilon_n^* = \min\left\{l\colon \sup_{\mathbf{X}(l)} T(\mathbf{X}(l)) = \infty\right\}\Big/ n,$$

where $\mathbf{X}(l)$ is obtained from $\mathbf{X}$ by varying at most l of the points $X^{(1)},\dots,X^{(n)}$.

The definition of breakdown point here corresponds to that called ε-replacement in Donoho and Huber (1983). We include the breakdown of the estimate of the dispersion matrix, which will usually lead to a lower breakdown point than obtained by restricting the definition to the breakdown of the location estimate. This is discussed in Section 2.5 of Donoho and Huber (1983).

In order to obtain the finite sample breakdown point we require the following lemma, which relates the volume of an ellipsoid which covers a sphere to the distance between their respective centres.

LEMMA 9. *Let* $\alpha \in R^k$ *and* $A \in \mathrm{PDS}(k)$ *with smallest eigenvalue* $a_{(1)}$. *Suppose that the ellipsoid* $\{x: (x-\alpha)^t A^{-1}(x-\alpha) \le c\}$ *contains the open sphere* $\{x: \|x-\beta\| < \rho\}$ *with* $\beta \in R^k$ *and* $\rho > 0$. *Then there exist positive numbers* c_9 *and* c_{10} *depending only on* c, β *and* ρ *such that*

$$a_{(1)} \ge c_9 > 0 \tag{41}$$

and

$$|A|^{1/2} \ge c_{10}(\|\alpha\| - \|\beta\|). \tag{42}$$

PROOF. Let $x = \beta + y$ with $\|y\| = \rho/2$. Then

$$(\beta + y - \alpha)^t A^{-1}(\beta + y - \alpha) \le c$$

and

$$(\beta - y - \alpha)^t A^{-1}(\beta - y - \alpha) \le c.$$

On adding these two inequalities we obtain

$$(\beta - \alpha)^t A^{-1}(\beta - \alpha) + y^t A^{-1} y \le c$$

and, hence, as A is positive definite,

$$y^t A^{-1} y \le c.$$

This implies [see, for example, Rao (1973), page 62]

$$\begin{aligned} a_{(1)}^{-1} &= \sup_{\|y\|=\rho/2} \frac{y^t A^{-1} y}{(\rho/2)^2} \\ &\le 4c/\rho^2, \end{aligned}$$

which yields the first inequality with $c_8 = \rho^2/4c$.

To prove the second we note that the largest eigenvalue $a_{(k)}$ of A satisfies the inequality [Rao (1973), page 62]

$$\frac{1}{a_{(k)}} \le \frac{(\beta-\alpha)^t A^{-1}(\beta-\alpha)}{\|\beta-\alpha\|^2} \le \frac{c}{\|\beta-\alpha\|^2}$$

and, hence, $a_{(k)} \ge \|\beta - \alpha\|^2/c$. As $|A| \ge a_{(1)}^{k-1} 0 a_{(k)}$ we obtain

$$|A| \ge c_8^{k-1}\|\beta - \alpha\|^2/c,$$

which implies (42). □

We now revert to the assumptions of Section 1 and drop the assumptions concerning the differentiability of κ. However, the results and proofs of this section are simplified if we assume that for some $\eta > 0$,

$$\kappa(u) = 1, \qquad 0 \le u < \eta. \tag{43}$$

Under this additional assumption we now have

THEOREM 5. *If the sample* $(X^{(1)}, \dots, X^{(n)})$ *with* $n(1-2\varepsilon) \ge k+1$ *is in general position, then*

$$\varepsilon_n^* = ([n\varepsilon]+1)/n.$$

PROOF. The proof follows the lines of that given in Rousseeuw (1986).

We first note that for the existence of a solution of the problem $\mathscr{P}_n$ for a sample $\mathbf{X} = (X^{(1)}, \dots, X^{(n)})$, it is only necessary that any subsample of size $n(1-\varepsilon)$ should contain at least $k+1$ points whose simplex contains a nonempty sphere. This follows from the proof of Theorem 2 and the condition is certainly met if $\mathbf{X}$ is in general position.

From (43) it follows that there exists a $v_0 = v_0(\mathbf{X}) > 0$ such that

$$\kappa\left(v_0^{-1} X^{(\nu)t} X^{(\nu)}\right) = 1, \tag{44}$$

for $1 \le \nu \le n$. We set $V = |v_0 I_k| = v_0^k$.

Consider now a subsample S of $\mathbf{X}$ of size $k+1$. From (41) of Lemma 9 it follows that there exists an $r_s > 0$ such that, for all $\alpha \in R^k$ with $\|\alpha\| > r_s$ and $A \in \mathrm{PDS}(k)$ for which

$$(X^{(\nu)} - \alpha)'^t A^{-1}(X^{(\nu)} - \alpha) \le c, \tag{45}$$

for all $X^{(\nu)} \in S$, we have $|A| > V$. By taking the maximum over all subsamples of size $k+1$ we obtain an $r > 0$ such that, if $\|\alpha\| > r$ and (45) is satisfied for all $X^{(\nu)}$ in some subsample of size $k+1$, then $|A| > V$.

Another similar application of Lemma 9 shows that there exists a $c_{11} > 0$ such that the smallest eigenvalue $a_{(1)}$ of any A satisfying (44) satisfies

$$0 < c_{11} \le a_{(1)}. \tag{46}$$

Let $\mathbf{Y}$ denote a sample of size n obtained from $\mathbf{X}$ by replacing $[n\varepsilon]$ of the $X^{(\nu)}$'s. Then any subsample of $\mathbf{Y}$ of size at least $n(1-\varepsilon)$ contains at least $n(1-\varepsilon) - [n\varepsilon] \ge n(1-2\varepsilon) \ge k+1$ of the $X^{(\nu)}$'s and, hence, the problem $\mathscr{P}_n$ for the sample $\mathbf{Y}$ has at least one solution, which we denote by $(\alpha_{\mathbf{Y}}^*, A_{\mathbf{Y}}^*)$. As $(Y^{(\nu)} - \alpha_{\mathbf{Y}}^*)^t A_{\mathbf{Y}}^{*^{-1}} (Y^{(\nu)} - \alpha_{\mathbf{Y}}^*) \le c$ for at least $n(1-\varepsilon)$ of the $Y^{(\nu)}$'s we see that for at least $k+1$ of the $X^{(\nu)}$'s, we have $(X^{(\nu)} - \alpha_{\mathbf{Y}}^*)^t A_{\mathbf{Y}}^{*^{-1}} (X^{(\nu)} - \alpha_{\mathbf{Y}}^*) \le c$. If now $\|\alpha_{\mathbf{Y}}^*\| > r$ with r as defined previously, it follows that $|A_{\mathbf{Y}}^*| > V$. However, $\mathbf{Y}$ contains at least $n - [n\varepsilon] \ge n(1-\varepsilon) X^{(\nu)}$'s and, hence,

$$\frac{1}{n} \sum_1^n \kappa\left(v_0^{-1} Y^{(\nu)t} Y^{(\nu)}\right) \ge 1 - \varepsilon.$$

As $(\alpha_{\mathbf{Y}}^*, A_{\mathbf{Y}}^*)$ is a solution of $\mathscr{P}_n$ for $\mathbf{Y}$ we deduce $|A_{\mathbf{Y}}^*| \le |v_0 I_k| = V$ and obtain a contradiction. Thus,

$$\|\alpha_{\mathbf{Y}}^*\| \le r \tag{47}$$

for any solution $(\alpha_{\mathbf{Y}}^*, A_{\mathbf{Y}}^*)$. From $|A_{\mathbf{Y}}^*| \le V$ and (46) we see that there exists a c_{12}, $0 < c_{12} < \infty$, such that

$$0 < c_{11} \le a_{(i)} \le c_{12}, \qquad 1 \le i \le k, \tag{48}$$

for the eigenvalues $a_{(1)},\dots,a_{(k)}$ of any solution A_Y^*. This together with (47) shows that the breakdown point ε_n^* satisfies $\varepsilon_n^* \ge ([n\varepsilon]+1)/n$.

We now prove the opposite inequality. Let $\mathbf{Y}$ now denote a sample obtained from $\mathbf{X}$ by replacing $[n\varepsilon]+1$ of the $X^{(\nu)}$'s. Let (α_Y, A_Y) now denote a pair for which

$$\frac{1}{n}\sum_1^n \kappa\Big(\big(Y^{(\nu)}-\alpha_Y\big)^t A_Y^{-1}\big(Y^{(\nu)}-\alpha_Y\big)\Big) \ge 1-\varepsilon.$$

Then $(Y^{(\nu)}-\alpha_Y)^t A_Y^{-1}(Y^{(\nu)}-\alpha_Y) \le c$ for at least $n-[n\varepsilon]$ of the $Y^{(\nu)}$'s. Thus, for at least $n-[n\varepsilon]+[n\varepsilon]+1-n\ge 1$ of the replacements, say $Y^{(1)}$, we must have

$$\big(Y^{(1)}-\alpha_Y\big)^t A_Y^{-1}\big(Y^{(1)}-\alpha_Y\big) \le c$$

and, therefore,

$$\|Y^{(1)}-\alpha_Y\|^2 \le c a_{(k)},$$

where $a_{(k)}$ denote the largest eigenvalue of A_Y. As $Y^{(1)}$ may be varied at will, this implies that the S-estimator breaks down and, hence, $\varepsilon_n^* \le ([n\varepsilon]+1)/n$. □

COROLLARY 1.

$$\lim_{n\to\infty} \varepsilon_n^* = \varepsilon.$$

COROLLARY 2. *If* $\varepsilon = 1/2-(k+1)/2n$ *then*

$$\varepsilon_n^* = \left[\frac{n-k+1}{2}\right]\Big/ n. \tag{49}$$

It is perhaps worth noting that the ε_n^* of (49) is higher than that given by Rousseeuw (1986) for his minimum volume estimator, namely $([n/2]-k+1)/n$. This may be obtained by setting $\varepsilon = ([n/2]+1)/n$. As this ε does not fulfill the condition $n(1-2\varepsilon)\ge k+1$ of Theorem 5, we have increased the breakdown point by imposing a slightly more stringent condition on the sample size.

In Donoho and Huber (1983) and Huber (1985) the breakdown point in the ε-contamination model of the Donoho–Stahel estimator described in Section 1 is shown to be $(n-2k+1)/(2n-2k+1)$. The breakdown point in the ε-contamination model for a sample of size n is defined to be $m/(m+n)$, where m is the smallest number of additional arbitrary sample points which, when adjoined to the initial sample, cause the estimator to break down. In order to reinterpret this result in terms of the ε-replacement model we note that the smallest m satisfies

$$\frac{m}{n+m} = \frac{n-2k+1}{2n-2k+1},$$

giving $m=n-2k+1$. Thus, the combined sample has a size of $n=n+m=2n-2k+1$ of which $m=n-2k+1=[N/2]-k+1$ points are bad. This implies a breakdown point for the combined sample in the ε-replacement model of $([N/2]-k+1)/N$.

This is again lower than that given by Corollary 2 and a natural question is whether the result of Corollary 2 is the best possible.

THEOREM 6. *Suppose $n \geq k + 1$, and the sample is in general position. Then the breakdown point of any affine equivariant estimator of the location and dispersion parameters is at most* $[(n - k + 1)/2]/n$.

PROOF. Consider a sample $\mathbf{X}$ of size n and a replacement sample $\mathbf{X}'$ with $[(n - k + 1)/2]$ bad points. We have $n - [(n - k + 1)/2] \geq k$ good points. We choose k of these points and consider the hyperplane determined by them. As the estimator is affine equivariant, we may suppose that this hyperplane is the $(k - 1)$-dimensional subspace $\{x\colon x_k = 0\}$. There remain $n' = n - [(n - k + 1)/2] - k$ good points which we denote by $X^{(1)}, \dots, X^{(n')}$. Now $n' \leq [(n - k + 1)/2]$ and, hence, we can choose n' of the bad points and place them at the points $X'^{(1)}, \dots, X'^{(n')}$ with $X_i'^{(\nu)} = X_i^{(\nu)}$, $1 \leq i \leq k - 1$, $1 \leq \nu \leq n'$, and $X_k'^{(\nu)} = uX_k^{(\nu)}$, $1 \leq \nu \leq n'$, with $u > 0$. We denote this sample by $\mathbf{X}'(u)$. Let $\Lambda(u) = \operatorname{diag}(1, \dots, 1, u^{-1})$ and set $\mathbf{X}''(u) = \Lambda(u)\mathbf{X}'(u)$. Then $\mathbf{X}''(u)$ is also obtained from $\mathbf{X}$ by altering at most $[(n - k + 1)/2]$ points. As our estimator is affine equivariant the estimators of the dispersion matrices $\Sigma(u)$ and $\Sigma''(u)$ obtained from $\mathbf{X}'(u)$ and $\mathbf{X}''(u)$ are related by $\Sigma''(u) = \Lambda(u)\Sigma'(u)\Lambda(u)$ and, consequently, $|\Sigma''(u)| = u^{-2}|\Sigma'(u)|$.

Thus as u tends to zero, one or another of the determinants must tend to zero or infinity, causing the estimator to breakdown. This proves the theorem. □

5. An example. In order to calculate Rousseeuw's minimum volume ellipsoid exactly, it is necessary to consider all $\binom{n}{[n/2] + 1}$ subsamples of the sample and then calculate the minimum covering ellipsoid for each subsample. This latter problem is certainly not easy and it would seem that an exact algorithm is available only for $k = 2$ [see Silverman and Titterington (1980)]. However, even if this problem could be solved, the large number of subsets is an insuperable barrier moderate for even values of n.

In spite of this, the situation is not quite hopeless. The minimum volume property can be used as a criterion for judging the goodness of ellipsoids. If it were possible to provide a large number of suitable candidates, then one could choose the best one according to the minimum volume criterion. One way of finding suitable candidates is to use a random search procedure. Using a random number generator, $q \geq k + 1$ sample points are chosen from the n points in the sample. A similar procedure is suggested in Hampel, Ronchetti, Rousseeuw and Stahel [(1986), page 302] for calculating the Donoho–Stahel estimator. Leroy and Rousseeuw (1984) also used a random search for calculating the best regression in the sense of least median of squares. The idea is due to Siegel (1982).

If we denote the empirical mean and variance-covariance matrix of these points by μ^q and Σ^q, then

$$E(\mu^q) = \mu$$

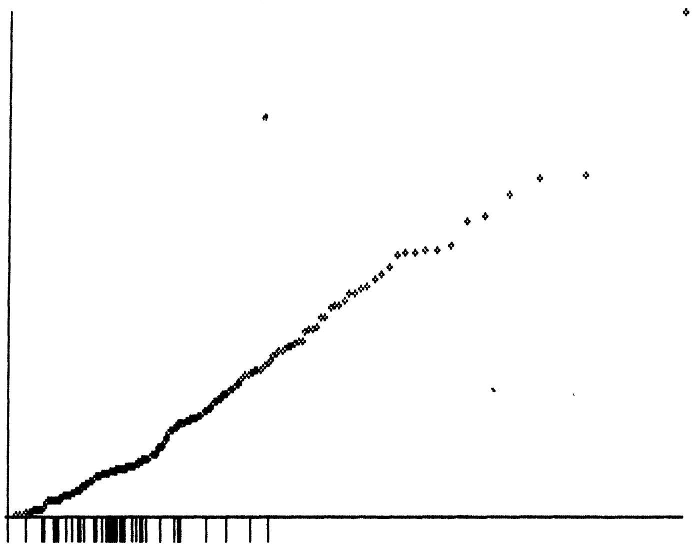

FIG. 1. *Standard χ^2 plot with 5 degrees of freedom.*

and

$$E(\Sigma^q) = \frac{q-1}{q}\Sigma,$$

as long as the points in the sample are distributed according to the assumed theoretical distribution. The ellipsoid $\{x: (x - \mu^q)^t(\Sigma^q)^{-1}(x - \mu^q) \le 1\}$ can now be blown up until it covers, say, $[(n - k + 1)/2]$ of the sample points and the volume calculated. This process is repeated a large number, N, of times and the best ellipsoid then determined.

In order to test the procedure we decided to generate random variables according to a critical distribution for M-estimators, i.e., a distribution which causes M-estimators to breakdown. Such a distribution is described in Hampel, Ronchetti, Rousseeuw and Stahel [(1986), page 297]. In their notation we took $m = 5$ and F_0 to be the standard normal distribution in 5 dimensions. We then generated 200 independent random variables according to the distribution

$$G_k(\cdot) = 0.65F_0(\cdot) + 0.35H_*(\cdot/s_k) \tag{50}$$

for different values of s_k. The distribution G_k of (50) may be interpreted as having 35% bad observations distributed according to $H_*(\cdot/s_k)$.

Figure 1 shows the χ^2-plot based on the mean and covariance matrix of the whole sample. The H_*-observations are denoted by the bars below the x-axis. In this example we set $s_k = 50$. Figure 2 shows the χ^2-plot for the same data, this time using the mean and covariance matrix obtained after 1000 random searches. Increasing the value of s_k has no effect on the minimum volume estimator.

One interesting effect occurs when s_k is reduced somewhat so that the H_*-observations lie closer to the origin. Figure 3 shows the χ^2-plot for the data

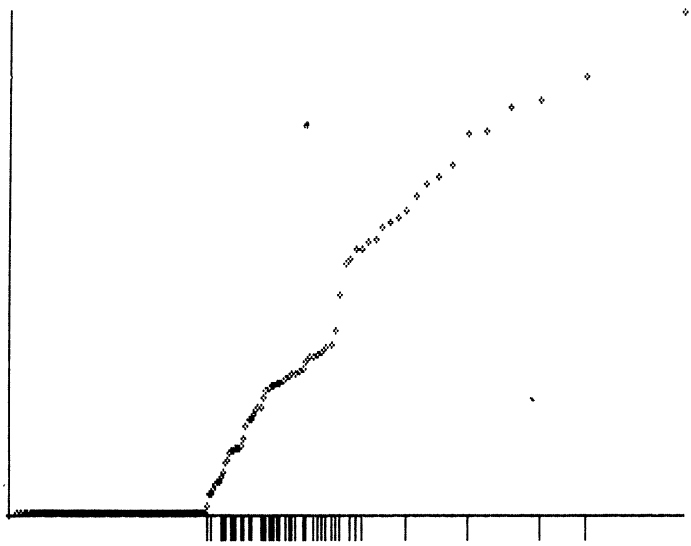

FIG. 2. *Robust* χ^2 *plot with* 5 *degrees of freedom.*

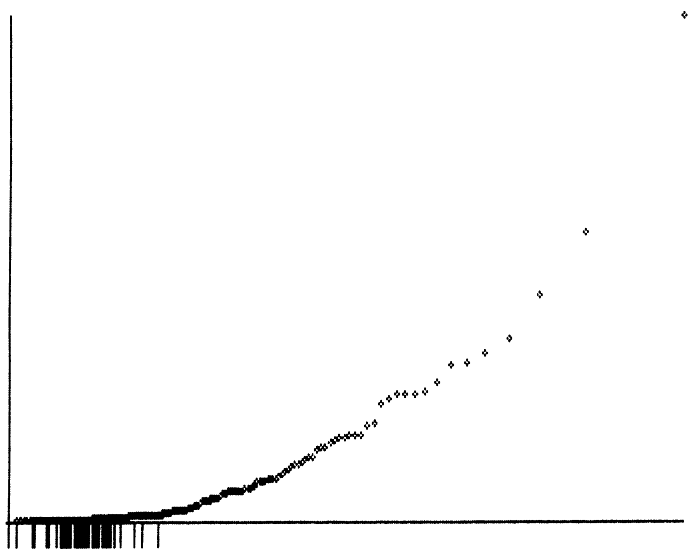

FIG. 3. *Robust* χ^2 *plot with* 5 *degrees of freedom.*

with $s_k = 40$ using the minimum volume estimator obtained from a random search. It can be seen that the minimum volume estimator has been greatly affected by the H_*-observations. This behaviour does not contradict Theorem 5. The $T(\mathbf{X})$ of (40) remains bounded for all values of s_k, $0 \leq s_k \leq \infty$, although, of course, particular values may be very large. The reason for the behaviour is clear. The minimum volume estimator derives from the ellipsoid with the smallest volume which covers $98 = [(200 - 5 + 1)/2]$ observations. The 70 H_*-observations lie on a one-dimensional hyperplane. As they also lie close to the main body of the data, it is possible to find 28 F_0-observations close to the 70 H_*-observations such that the resulting 98 observations may be covered by an ellipsoid with a small volume.

Acknowledgments. My debt to Peter Rousseeuw is obvious. I would like to thank Jörg van Hoorn for checking my calculations and also the referees who found a gross error in the first version of the paper and whose other comments lead to considerable improvements.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Harvard Univ.

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, Calif.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

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

HUBER, P. J. (1985). Projection pursuit. *Ann. Statist.* **13** 435–475.

LEROY, A. and ROUSSEEUW, P. J. (1984). PROGRESS: A program for robust regression. Technical Report CS00 TW/201, Center for Statistics and Operations Research, VUB Brussels.

MARONNA, R. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

RAO, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.

ROUSSEEUW, P. J. (1986). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications* (W. Grossman, G. Pflug, I. Vincze and W. Wertz, eds.) 283–297. Reidel, Dordrecht.

ROUSSEEUW, P. and YOHAI, V. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 256–272. Springer, New York.

SIEGEL, A. F. (1982). Robust regression using repeated medians. *Biometrika* **69** 242–244.

SILVERMAN, B. W. and TITTERINGTON, D. M. (1980). Minimum covering ellipses. *SIAM J. Sci. Statist. Comput.* **1** 401–409.

STAHEL, W. (1981). Breakdown of covariance estimators. Research Report 31, Fachgruppe für Statistik, ETH Zürich.

FACHBEREICH 6 MATHEMATIK
UNIVERSITÄT GESAMTHOCHSCHULE ESSEN
POSTFACH 103 764
4300 ESSEN 1
WEST GERMANY