

ASYMPTOTIC PROPERTIES OF STATISTICAL ESTIMATORS IN STOCHASTIC PROGRAMMING

BY ALEXANDER SHAPIRO

University of South Africa

The aim of this article is to investigate the asymptotic behaviour of estimators of the optimal value and optimal solutions of a stochastic program. These estimators are closely related to the M -estimators introduced by Huber (1964). The parameter set of feasible solutions is supposed to be defined by a number of equality and inequality constraints. It will be shown that in the presence of inequality constraints the estimators are not asymptotically normal in general. Maximum likelihood and robust regression methods will be discussed as examples.

1. Introduction. In this article we study the following stochastic programming problem. Let $(\mathcal{Z}, \mathcal{F}, P)$ be a probability space. Consider a function $f: \mathcal{Z} \times \mathbb{R}^k \rightarrow \mathbb{R}$, a set $S \subset \mathbb{R}^k$ and the associated stochastic programming problem,

$$(\mathcal{P}_0) \quad \text{minimize } \varphi(v) \text{ subject to } v \in S,$$

where φ is the expected value of f ,

$$\varphi(v) = E\{f(z, v)\}$$

with respect to the probability measure P . Let z_1, \dots, z_n be a sample of independent random variables with values in \mathcal{Z} having the common probability distribution P and consider the mathematical programming problem,

$$(\mathcal{P}_n) \quad \text{minimize } \psi_n(v) \text{ subject to } v \in S,$$

where

$$(1.1) \quad \psi_n(v) = n^{-1} \sum_{i=1}^n f(z_i, v).$$

The aim of this article is to study the asymptotic behaviour of the optimal value

$$\vartheta_n = \inf\{\psi_n(v): v \in S\}$$

and a corresponding optimal solution \bar{v}_n of the program (\mathcal{P}_n) as the sample size n tends to infinity.

There is a substantial statistical literature dealing with various cases of the problem formulated above. The classical maximum likelihood approach can be considered in the present framework if the objective function f is taken to be minus the logarithm of a probability density function. For the case where the set S is given by *equality* constraints a relevant asymptotic theory of the

Received November 1987; revised March 1988.

AMS 1980 *subject classifications*. 62F12, 90C15.

Key words and phrases. Stochastic programming, M -estimators, inequality constraints, asymptotic normality, cone approximation, optimality conditions, Lagrange multipliers.

corresponding maximum likelihood estimators is presented in Aitchison and Silvey (1958). Closely related to the program (\mathcal{P}_n) is the method of M -estimators introduced in Huber (1964) and discussed at length in Huber (1981). Asymptotic normality of M -estimators was proved in Huber (1967) essentially under the assumption that the set S is open and consequently there are no restrictions on local variations of \bar{v}_n . For more recent work in this direction see, for example, Boos and Serfling (1980), Tsybakov (1981), Clarke (1983) and Fernholz (1983). It is also relevant to mention some work on maximum likelihood estimation of misspecified models [see, e.g., White (1982) and references therein].

In this article we consider the situation where the set S is defined by a number of equality and *inequality* constraints. Introduction of inequality constraints has the disastrous consequence that then estimators \bar{v}_n are *not asymptotically normal* in general. This is probably why inequality constraints have been avoided in statistical literature. However, in stochastic programming inequality constraints are most relevant. An extensive discussion, examples, references and motivation for developing asymptotic theory involving inequality constraints can be found in King (1986) and Dupačová and Wets (1988).

We proceed as follows. In the next section we introduce required regularity conditions and discuss their immediate implications. The main results are formulated and proved in Section 3. These results are influenced by some recent advances in a (deterministic) perturbation theory of nonlinear programs [Shapiro (1988a)]. Section 4 is devoted to applications and examples.

We denote by ϑ_0 the optimal value $\vartheta_0 = \inf\{\varphi(v): v \in S\}$ of the program (\mathcal{P}_0) and by \mathcal{S}_n the set of optimal solutions of the program (\mathcal{P}_n) . The gradients, for example, $\nabla f(z, v)$, and the Hessian matrices, $\nabla^2 f(z, v) = \{\partial^2 f(z, v)/\partial v_i \partial v_j\}$, are always taken with respect to the parameter vector v . We make use of the concept of generalized gradients of locally Lipschitz functions [Clarke (1983)]. That is let $h: \mathbb{R}^k \rightarrow \mathbb{R}$ be a locally Lipschitz function. Then the generalized gradient of h at v , denoted by $\partial h(v)$, is the convex hull of all limits of the form $\lim \nabla h(v_i)$, where $v_i \rightarrow v$ and h is differentiable at v_i . (By Rademacher's theorem the set of points where a locally Lipschitz function fails to be differentiable has Lebesgue measure zero.)

2. Regularity conditions and preliminary discussion. In this section we introduce some required regularity conditions and discuss their immediate consequences.

ASSUMPTION A.1. For almost every z , $f(z, v)$ is a continuous function of v and for all $v \in S$, $f(z, v)$ is a measurable function of z .

ASSUMPTION A.2. The family $\{f(z, v)\}$, $v \in S$, is uniformly integrable, that is,

$$\lim_{c \rightarrow \infty} \sup_{v \in S} \int_{\{z: |f(z, v)| \geq c\}} |f(z, v)| P(dz) = 0.$$

Notice that assumptions A.1 and A.2 imply that the expected value function φ is continuous on S .

ASSUMPTION A.3. Program (\mathcal{P}_0) has a unique optimal solution v_0 .

If Assumptions A.1–A.3 hold and the set S is compact, then almost surely the optimal set \mathcal{S}_n is nonempty and any selection $\bar{v}_n \in \mathcal{S}_n$ converges a.s. (or in probability) to v_0 . Here the compactness assumption can be replaced by some other regularity conditions. For a detailed discussion of such consistency results see Wald (1949), Le Cam (1953), Huber (1967), Bahadur (1967) and Dupačová and Wets (1988). We also assume that the considered selection $\bar{v}_n \in \mathcal{S}_n$ is measurable. The existence of such measurable selections was proved, under very general conditions in Pfanzagl (1969), Section 1, and Jennrich (1969), Lemma 2. [See also Rockafellar and Wets (1984) for a thorough discussion of measurability in stochastic programming.]

Let \mathcal{N} be a convex neighbourhood of v_0 .

ASSUMPTION A.4. For almost every z , $f(z, \cdot)$ is Lipschitz continuous on \mathcal{N} . That is, there exists a positive constant $K(z)$ such that

$$|f(z, w) - f(z, v)| \leq K(z)\|w - v\|$$

for all $v, w \in \mathcal{N}$.

ASSUMPTION A.5. For each fixed $v \in \mathcal{N}$, $f(z, \cdot)$ is continuously differentiable at v for almost every z .

ASSUMPTION A.6. The family $\{\nabla f(z, v)\}$, $v \in \mathcal{N}$, is uniformly integrable.

PROPOSITION 2.1. Suppose that assumptions A.1, A.2 and A.4–A.6 hold. Then φ is continuously differentiable on \mathcal{N} and

$$(2.1) \quad \nabla \varphi(v) = E\{\nabla f(z, v)\}.$$

PROOF. First consider the case of $k = 1$, that is, v is a scalar. Because of Assumption A.4 we have that for a.e. z and $w, v \in \mathcal{N}$,

$$f(z, w) - f(z, v) = \int_v^w f'_t(z, t) dt.$$

Now

$$\begin{aligned} \varphi(w) - \varphi(v) &= E\{f(z, w) - f(z, v)\} = \int \int_v^w f'_t(z, t) dt dP \\ &= \int_v^w \int f'_t(z, t) dP dt = \int_v^w E\{f'_t(z, t)\} dt. \end{aligned}$$

Notice that the order of integration can be interchanged because of Assumption

A.6. It follows that

$$\inf_{t \in [v, w]} E\{f'_t(z, t)\} \leq \frac{\varphi(w) - \varphi(v)}{w - v} \leq \sup_{t \in [v, w]} E\{f'_t(z, t)\}.$$

Moreover, Assumptions A.5 and A.6 imply that $E\{f'_t(z, t)\}$ is continuous in t and hence

$$\lim_{w \rightarrow v} \frac{\varphi(w) - \varphi(v)}{w - v} = E\{f'(z, v)\}.$$

In the general case of $k \geq 1$, we obtain that the partial derivatives $\partial\varphi/\partial v_i$, $i = 1, \dots, k$, do exist and (2.1) holds. It follows from Assumptions A.5 and A.6 that $\partial\varphi/\partial v_i$ are continuous and hence φ is continuously differentiable on \mathcal{N} . \square

Consider a sequence z_1, z_2, \dots of i.i.d. random variables and the corresponding function $\psi_n(v)$ defined in (1.1). The following result is a consequence of the strong law of large numbers.

PROPOSITION 2.2. *Suppose that Assumptions A.1, A.2 and A.4–A.6 hold and let \mathcal{U} be a compact subset of \mathcal{N} . Then for almost every i.i.d. sequence $\{z_i\}$, $\partial\psi_n(v)$ converges to $\nabla\varphi(v)$ uniformly on \mathcal{U} , that is,*

$$\lim_{n \rightarrow \infty} \sup_{v \in \mathcal{U}} \sup_{u \in \partial\psi_n(v)} \|u - \nabla\varphi(v)\| = 0.$$

PROOF. Since for any set $B \subset \mathbb{R}^k$ and a point x ,

$$\sup\{\|u - x\| : u \in \text{conv}(B)\} = \sup\{\|u - x\| : u \in B\},$$

it suffices to prove that a.s. $\nabla\psi_n(v)$ converges to $\nabla\varphi(v)$ uniformly for all such $v \in \mathcal{U}$ that $\nabla\psi_n(v)$ does exist. Consider a point $v \in \mathcal{U}$ and a sequence \mathcal{W}_k of neighbourhoods of v shrinking to $\{v\}$. By Assumption A.6,

$$\lim_{k \rightarrow \infty} E\left\{\sup_{w \in \mathcal{W}_k} \|\nabla f(z, w) - \nabla f(z, v)\|\right\} = E\left\{\lim_{k \rightarrow \infty} \sup_{w \in \mathcal{W}_k} \|\nabla f(z, w) - \nabla f(z, v)\|\right\}$$

and because of Assumption A.5 the last limit is zero. Moreover, we have that

$$(2.2) \quad \sup_{w \in \mathcal{W}_k} \|\nabla\psi_n(w) - \nabla\psi_n(v)\| \leq n^{-1} \sum_{i=1}^n \sup_{w \in \mathcal{W}_k} \|\nabla f(z_i, w) - \nabla f(z_i, v)\|$$

and by the strong law of large numbers the right-hand side of (2.2) converges a.s. to

$$E\left\{\sup_{w \in \mathcal{W}_k} \|\nabla f(z, w) - \nabla f(z, v)\|\right\}.$$

It follows that for any $\varepsilon > 0$ there exists a neighbourhood \mathcal{V} of v such that a.s. for sufficiently large n ,

$$\sup_{w \in \mathcal{V}} \|\nabla\psi_n(w) - \nabla\psi_n(v)\| < \varepsilon.$$

Since \mathcal{U} is compact, there exists a finite number of points $v_1, \dots, v_m \in \mathcal{U}$ and corresponding neighbourhoods $\mathcal{V}_1, \dots, \mathcal{V}_m$ covering \mathcal{U} such that a.s. for sufficiently large n ,

$$\sup_{w \in \mathcal{V}_j} \|\nabla \psi_n(w) - \nabla \psi_n(v_j)\| < \varepsilon, \quad j = 1, \dots, m.$$

Furthermore, by Proposition 2.1, $\nabla \varphi(v)$ is continuous on \mathcal{N} and hence the neighbourhoods $\mathcal{V}_1, \dots, \mathcal{V}_m$ can be chosen in such a way that

$$\sup_{w \in \mathcal{V}_j} \|\nabla \varphi(w) - \nabla \varphi(v_j)\| < \varepsilon, \quad j = 1, \dots, m.$$

Now by the strong law of large numbers

$$\|\nabla \psi_n(v_j) - \nabla \varphi(v_j)\| < \varepsilon, \quad j = 1, \dots, m,$$

a.s. for sufficiently large n and hence

$$\|\nabla \psi_n(v) - \nabla \varphi(v)\| < 3\varepsilon$$

for all $v \in \mathcal{U}$ where $\nabla \psi_n(v)$ exists. \square

Now let us consider the program (\mathcal{P}_0) . It will be assumed that in a neighbourhood of v_0 the set S is defined by a finite number of equality and inequality constraints.

ASSUMPTION B.1. There exist a neighbourhood \mathcal{N} of v_0 and functions $g_i(v)$ such that

$$S \cap \mathcal{N} = \{v \in \mathcal{N} : g_i(v) = 0, i \in I; g_i(v) \leq 0, i \in J\},$$

where I and J are finite index sets and $g_i(v_0) = 0$ for all $i \in J$.

ASSUMPTION B.2. The functions g_i , $i \in I \cup J$, are twice continuously differentiable in a neighbourhood of v_0 .

It will be assumed that the following Mangasarian–Fromovitz (1967) constraint qualification is satisfied at the point v_0 .

ASSUMPTION B.3 (MF-condition). (i) The gradient vectors $\nabla g_i(v_0)$, $i \in I$, are linearly independent.

(ii) There exists a vector w such that

$$\begin{aligned} w' \nabla g_i(v_0) &= 0, & i \in I, \\ w' \nabla g_i(v_0) &< 0, & i \in J. \end{aligned}$$

Consider the Lagrangian function

$$l(v, \lambda) = \varphi(v) + \sum_{i \in I \cup J} \lambda_i g_i(v),$$

associated with the program (\mathcal{P}_0) , and let Λ_0 be the corresponding set of Lagrange multipliers satisfying the first-order (Kuhn–Tucker) necessary

conditions. That is, $\lambda \in \Lambda_0$ iff

$$\nabla l(v_0, \lambda) = 0$$

and $\lambda_i \geq 0$, $i \in J$. Under Assumption B.3 the set Λ_0 is nonempty (first-order necessary conditions) and bounded [Gauvin (1977)]. Moreover, by the definition Λ_0 is a closed convex polytope and hence is the convex hull of the finite set \mathcal{E}_0 of its extreme points.

We now introduce second-order sufficient conditions for the program (\mathcal{P}_0) .

ASSUMPTION B.4. The function φ is twice continuously differentiable in a neighbourhood of v_0 .

ASSUMPTION B.5 (Second-order sufficient conditions). For all nonzero $w \in C$,

$$(2.3) \quad \max_{\lambda \in \Lambda_0} w' \nabla^2 l(v_0, \lambda) w > 0,$$

where C is the cone of critical directions

$$C = \{w: w' \nabla g_i(v_0) = 0, i \in I; w' \nabla g_i(v_0) \leq 0, i \in J; w' \nabla \varphi(v_0) \leq 0\}.$$

Under the constraint qualification of Assumption B.3, Assumption B.5 is the standard second-order sufficient condition for the program (\mathcal{P}_0) [Hestenes (1975) and Ioffe (1979)]. It implies that if the first-order necessary conditions are satisfied at v_0 , then v_0 is a strict local minimizer of φ over S . The corresponding second-order necessary conditions are obtained by replacing the strict inequality sign in (2.3) with the sign "greater than or equal to."

For a given vector λ of Lagrange multipliers consider the set

$$(2.4) \quad S_+(\lambda) = \{v: g_i(v) = 0, i \in I \cup J_+(\lambda)\},$$

where

$$J_+(\lambda) = \{i \in J: \lambda_i > 0\}.$$

ASSUMPTION C.1. There is a neighbourhood \mathcal{N} of v_0 such that for every $\lambda \in \mathcal{E}_0$, the variable

$$(2.5) \quad \sup_{v \in S_+(\lambda) \cap \mathcal{N}} \frac{\|\nabla \psi_n(v) - \nabla \varphi(v) - \nabla \psi_n(v_0) + \nabla \varphi(v_0)\|}{n^{-1/2} + \|v - v_0\|}$$

tends to 0 in probability as $n \rightarrow \infty$.

Assumption C.1 is not immediately obvious and requires an explanation. Notice that the supremum in (2.5) is taken over such v that $\nabla \psi_n(v)$ does exist. It is a deep result due to Huber (1967) that the following regularity conditions imply Assumption C.1. Notice that for every $\lambda \in \mathcal{E}_0$, the gradient vectors $\nabla g_i(v_0)$, $i \in I \cup J_+(\lambda)$, are linearly independent and hence $S_+(\lambda)$ is a smooth manifold in a neighbourhood of v_0 . The tangent space to this manifold at v_0 is

given by the linear space

$$M(\lambda) = \{w: w' \nabla g_i(v_0) = 0, i \in I \cup J_+(\lambda)\}.$$

ASSUMPTION C.2. For every $\lambda \in \mathcal{E}_0$ the Hessian matrix $\nabla^2 l(v_0, \lambda)$ is positive definite on $M(\lambda)$, that is,

$$w' \nabla^2 l(v_0, \lambda) w > 0$$

for all nonzero $w \in M(\lambda)$.

ASSUMPTION C.3. There are strictly positive numbers c_1 , c_2 and d_0 such that for every $d \geq 0$ and all v satisfying $\|v - v_0\| + d \leq d_0$,

$$(i) \quad E\{u(z, v, d)\} \leq c_1 d,$$

$$(ii) \quad E\{u(z, v, d)^2\} \leq c_2 d,$$

where

$$u(z, v, d) = \sup_{\|w - v\| \leq d} \|\nabla f(z, w) - \nabla f(z, v)\|.$$

Consider the Lagrangian

$$L(z, v, \lambda) = f(z, v) + \sum_{i \in I \cup J} \lambda_i g_i(v).$$

Clearly, $E\{L(z, v, \lambda)\} = l(v, \lambda)$ and

$$\psi_n(v) - \varphi(v) = n^{-1} \sum_{j=1}^n L(z_j, v, \lambda) - l(v, \lambda).$$

It follows from Assumption C.2 and continuity of $\nabla^2 l(v, \lambda)$ that there exists a positive number α such that

$$\|\nabla l(v, \lambda)\| \geq \alpha \|v - v_0\|$$

for all $v \in S_+(\lambda) \cap \mathcal{N}$, $\lambda \in \mathcal{E}_0$. Moreover, because of Assumptions B.2 and C.3 (i),

$$\|\nabla l(v, \lambda)\| \leq b \|v - v_0\|$$

for some $b > 0$. Therefore in the present situation, it does not make any difference if the term $\|v - v_0\|$ in the denominator of the ratio (2.5) is replaced by $\|\nabla l(v, \lambda)\|$. Then we obtain from Lemma 3 of Huber (1967), applied to the function $L(z, v, \lambda)$ on the manifold $S_+(\lambda)$, that *Assumption C.1 is implied by Assumptions B.2, C.2 and C.3.*

It should be mentioned that for very $\lambda \in \mathcal{E}_0$, the linear space $M(\lambda)$ contains the critical cone C . Therefore Assumption C.2 gives conditions which are stronger than the second-order sufficient conditions of Assumption B.5.

In cases where $f(z, \cdot)$ is twice differentiable, Assumption C.1 can be ensured by the following relatively simple conditions.

ASSUMPTION C.4. There is a neighbourhood \mathcal{N} of v_0 such that

- (i) For almost every z , $f(z, \cdot)$ is twice continuously differentiable on \mathcal{N} .
- (ii) The family $\{\nabla^2 f(z, v)\}$, $v \in \mathcal{N}$, is uniformly integrable.

Assumption C.4 implies that $\varphi(v)$ is twice continuously differentiable on \mathcal{N} (Assumption B.4) and that the second-order derivatives of φ can be taken inside the expected value. Moreover, assuming that \mathcal{N} is compact, it then follows from the strong law of large numbers that with probability 1, $\nabla^2 \psi_n(v)$ converges to $\nabla^2 \varphi(v)$ uniformly on \mathcal{N} [see, e.g., Le Cam (1953), Corollary 4.1, and Jennrich (1969), Theorem 2]. By the mean value theorem this implies that

$$\sup_{v \in \mathcal{N}} \frac{\|\nabla \psi_n(v) - \nabla \psi_n(v_0) - \nabla \varphi(v) + \nabla \varphi(v_0)\|}{\|v - v_0\|} \rightarrow 0 \quad \text{a.s.}$$

and hence Assumption C.1 follows.

ASSUMPTION D. The expectation $E\{\|\nabla f(z, v_0)\|^2\}$ is finite.

3. Asymptotic results. In this section we derive asymptotic expansions of the optimal value ϑ_n and an optimal solution $\bar{v}_n \in \mathcal{S}_n$. Put

$$\zeta_n = \nabla \psi_n(v_0) - \nabla \varphi(v_0)$$

and

$$q(w) = \max\{w' \nabla^2 l(v_0, \lambda) w : \lambda \in \Lambda_0\}.$$

Notice that the set Λ_0 in the definition of $q(w)$ can be replaced by the set \mathcal{E}_0 of its extreme point (assuming that Λ_0 is bounded) and hence $q(w)$ is the pointwise maximum of a finite number of quadratic functions. We now formulate the main results of this article.

THEOREM 3.1. *Suppose that Assumptions A.1–A.6, B.1–B.5, C.1 and D hold and that there exists a measurable selection $\bar{v}_n \in \mathcal{S}_n$ converging in probability to v_0 . Then*

$$(3.1) \quad \vartheta_n = \psi_n(v_0) + \min_{w \in C} \{w' \zeta_n + \tfrac{1}{2} q(w)\} + o_p(n^{-1}).$$

Under the second-order sufficient conditions of Assumption B.5, the set of minimizers of the function $w' \zeta + \tfrac{1}{2} q(w)$ over the critical cone C is nonempty and compact for all ζ . Suppose now that this function has a *unique* minimizer $\bar{w}(\zeta)$ over C . Such uniqueness can be ensured, for example, by the strong form of second-order sufficient conditions given in Assumption C.2.

THEOREM 3.2. *Suppose that Assumptions A.1–A.6, B.1–B.5, C.1 and D hold and for all ζ the function $w' \zeta + \tfrac{1}{2} q(w)$ has a unique minimizer $\bar{w}(\zeta)$ over C . Let*

$\bar{v}_n \in \mathcal{S}_n$ be a measurable selection converging in probability to v_0 . Then

$$(3.2) \quad \|\bar{v}_n - v_0 - \bar{\omega}(\xi_n)\| = o_p(n^{-1/2}).$$

By the central limit theorem, under Assumption D, $n^{1/2}\xi_n$ converges in distribution to $N(0, \Psi)$ with

$$(3.3) \quad \Psi = E\{[\nabla f(z, v_0) - \nabla \varphi(v_0)][\nabla f(z, v_0) - \nabla \varphi(v_0)]'\}.$$

Also the minimizer $\bar{\omega}(\xi)$ is a continuous, positively homogeneous function of ξ and hence $n^{1/2}\bar{\omega}(\xi) = \bar{\omega}(n^{1/2}\xi)$. Therefore the following asymptotic result is an immediate consequence of Theorem 3.2.

THEOREM 3.3. *Suppose that the Assumptions of Theorem 3.2 hold. Then $n^{1/2}(\bar{v}_n - v_0)$ converges in distribution to $\bar{\omega}(y)$, where y is a normal vector variable with mean 0 and covariance matrix Ψ given in (3.3).*

In the remainder of this section we concentrate on proofs of Theorems 3.1 and 3.2. Our first step will be in reducing the program (\mathcal{P}_n) to a more convenient one. Consider the Lagrangian

$$l_n(v, \lambda) = n^{-1} \sum_{j=1}^n L(z_j, v, \lambda) = \psi_n(v) + \sum_{i \in I \cup J} \lambda_i g_i(v)$$

of the program (\mathcal{P}_n) and the pointwise maximum function

$$\Phi_n(v) = \max\{l_n(v, \lambda) : \lambda \in \Lambda_0\}.$$

Also consider the restricted feasible set

$$\bar{S} = \cup\{S(\lambda) : \lambda \in \mathcal{E}_0\},$$

where

$$S(\lambda) = S_+(\lambda) \cap \{v : g_i(v) \leq 0, i \in J_0(\lambda)\},$$

$S_+(\lambda)$ is defined in (2.4) and $J_0(\lambda) = \{i \in J : \lambda_i = 0\}$. Clearly, in a neighbourhood of v_0 , the set \bar{S} is a subset of S .

Near the point v_0 first-order necessary conditions for the program (\mathcal{P}_n) can be written in the form: There exists a vector $\bar{\lambda}_n$ of Lagrange multipliers such that

$$0 \in \partial l_n(\bar{v}_n, \bar{\lambda}_n) = \partial \psi_n(\bar{v}_n) + \sum_{i \in I \cup J} \bar{\lambda}_{i,n} \nabla g_i(\bar{v}_n)$$

[Clarke (1983), Section 6.1]. Then by employing Proposition 2.2, under Assumptions A.1–A.6 and B.1–B.3 the following results can be proved essentially along the same lines as Theorem 2.1 in Shapiro (1988a).

LEMMA 3.1. *There exists a neighbourhood \mathcal{N} of v_0 such that if $\bar{v}_n \in \mathcal{N}$, then a.s. \bar{v}_n is a minimizer of $\Phi_n(v)$ over \bar{S} and*

$$\vartheta_n = \min\{\Phi_n(v) : v \in \bar{S}\}.$$

Another required result is that the restricted set \bar{S} is approximated at v_0 by the critical cone C in the sense of the following definition.

DEFINITION. We say that a set $S \subset \mathbb{R}^k$ is approximated at $v_0 \in S$ by a closed cone C , called the approximating cone, if

$$\inf_{w \in C} \|(v - v_0) - w\| = o(\|v - v_0\|), \quad v \in S,$$

and

$$\inf_{v \in S} \|(v - v_0) - w\| = o(\|w\|), \quad w \in C.$$

The concept of cone approximation goes back to Chernoff (1954). Various properties, equivalent definitions and applications of cone approximations are given in Shapiro (1985, 1987) and Rockafellar (1987).

LEMMA 3.2 [Shapiro (1988a), Lemmas 2.3 and 2.4]. *If the functions g_i , $i \in I \cup J$, are continuously differentiable near v_0 and the MF-condition (Assumption B.3) holds, then the restricted set \bar{S} is approximated at v_0 by the critical cone C .*

LEMMA 3.3. *Suppose that the Assumptions of Theorem 3.1 hold. Then $n^{1/2}(\bar{v}_n - v_0)$ is bounded in probability.*

PROOF. We have that

$$\begin{aligned} \Phi_n(v) &= \max\{\psi_n(v) - \varphi(v) + l(v, \lambda) : \lambda \in \mathcal{E}_0\} \\ &= \psi_n(v) - \varphi(v) + \varphi(v_0) + \frac{1}{2}q(v - v_0) + o(\|v - v_0\|^2). \end{aligned}$$

By Lemma 3.1, with probability tending to 1, $\vartheta_n = \Phi_n(\bar{v}_n)$ and hence

$$\vartheta_n - \vartheta_0 = \psi_n(\bar{v}_n) - \varphi(\bar{v}_n) + \frac{1}{2}q(\bar{v}_n - v_0) + o(\|\bar{v}_n - v_0\|^2).$$

Applying the mean value theorem for locally Lipschitz functions [Clarke (1983), page 41],

$$\psi_n(\bar{v}_n) - \varphi(\bar{v}_n) = \psi_n(v_0) - \varphi(v_0) + (\bar{v}_n - v_0)' \zeta_n^*,$$

where

$$\zeta_n^* \in \partial \psi_n(v_n^*) - \nabla \varphi(v_n^*)$$

and v_n^* is a point on the segment joining \bar{v}_n and v_0 . It follows that

$$\begin{aligned} (3.4) \quad \vartheta_n - \vartheta_0 &= \psi_n(v_0) - \varphi(v_0) + (\bar{v}_n - v_0)' \zeta_n^* + \frac{1}{2}q(\bar{v}_n - v_0) \\ &\quad + o(\|\bar{v}_n - v_0\|^2). \end{aligned}$$

On the other hand,

$$\vartheta_n - \vartheta_0 = \psi_n(\bar{v}_n) - \varphi(v_0) \leq \psi_n(v_0) - \varphi(v_0),$$

which together with (3.4) implies that

$$(3.5) \quad (\bar{v}_n - v_0)' \zeta_n^* + \frac{1}{2}q(\bar{v}_n - v_0) + o(\|\bar{v}_n - v_0\|^2) \leq 0.$$

Since with probability tending to 1 we have that $\bar{v}_n \in \bar{S}$ (Lemma 3.1), that the set \bar{S} is approximated at v_0 by the critical cone C (Lemma 3.2) and because of the second-order sufficient conditions (Assumption B.5) it follows that $q(\bar{v}_n - v_0)$ is greater than $\varepsilon \|\bar{v}_n - v_0\|^2$ for some $\varepsilon > 0$. It follows then from (3.5) that

$$(\varepsilon/2) \|\bar{v}_n - v_0\|^2 \leq -(\bar{v}_n - v_0)' \zeta_n^*$$

and hence

$$(3.6) \quad \|\bar{v}_n - v_0\| \leq 2\varepsilon^{-1}(\|\zeta_n\| + \|\zeta_n^* - \zeta_n\|).$$

Now since $\bar{v}_n \in S_+(\lambda)$ for some $\lambda \in \mathcal{E}_0$, it follows from Assumption C.1 that

$$\|\zeta_n^* - \zeta_n\| = o_p(n^{-1/2} + \|\bar{v}_n - v_0\|)$$

and from Assumption D that $\|\zeta_n\|$ is $O_p(n^{-1/2})$. This together with the inequality (3.6) completes the proof. \square

PROOF OF THEOREM 3.1. Consider (3.4). Because of Assumption C.1 and by Lemma 3.3 we have that $\|\zeta_n^* - \zeta_n\|$ is $o_p(n^{-1/2})$. Since $\|\bar{v}_n - v_0\|$ is $O_p(n^{-1/2})$ we then obtain that

$$(3.7) \quad \begin{aligned} \vartheta_n - \vartheta_0 &= \psi_n(v_0) - \varphi(v_0) + (\bar{v}_n - v_0)' \zeta_n \\ &\quad + \tfrac{1}{2}q(\bar{v}_n - v_0) + o_p(n^{-1}). \end{aligned}$$

Now since \bar{S} is approximated by C , there exists a point $w_n \in C$ such that

$$\|\bar{v}_n - v_0 - w_n\| = o(\|\bar{v}_n - v_0\|).$$

Then

$$q(\bar{v}_n - v_0) = q(w_n) + o(\|\bar{v}_n - v_0\|^2)$$

and since $\|\bar{v}_n - v_0\|$ and $\|\zeta_n\|$ are $O_p(n^{-1/2})$ we obtain from (3.7) that

$$\vartheta_n - \vartheta_0 = \psi_n(v_0) - \varphi(v_0) + w_n' \zeta_n + \tfrac{1}{2}q(w_n) + o_p(n^{-1}).$$

It follows that

$$(3.8) \quad \vartheta_n - \vartheta_0 \geq \psi_n(v_0) - \varphi(v_0) + \min_{w \in C} \{w' \zeta_n + \tfrac{1}{2}q(w)\} + o_p(n^{-1}).$$

The other inequality, which is obtained from (3.8) by inverting the inequality sign, can be proved in a similar way by considering a minimizer \bar{w}_n of the function $w' \zeta_n + \tfrac{1}{2}q(w)$ over C . \square

PROOF OF THEOREM 3.2. From (3.1) and (3.7) we have

$$(\bar{v}_n - v_0)' \zeta_n + \tfrac{1}{2}q(\bar{v}_n - v_0) = \min_{w \in C} \{w' \zeta_n + \tfrac{1}{2}q(w)\} + o_p(n^{-1})$$

and hence

$$(3.9) \quad \begin{aligned} &n^{1/2}(\bar{v}_n - v_0)'(n^{1/2}\zeta_n) + \tfrac{1}{2}q[n^{1/2}(\bar{v}_n - v_0)] \\ &= \min_{w \in C} \{w'(n^{1/2}\zeta_n) + \tfrac{1}{2}q(w)\} + o_p(1). \end{aligned}$$

Notice that the multiplier $n^{1/2}$ of w in the right-hand side of (3.9) was absorbed

into the cone C . Now since $n^{1/2}(\bar{v}_n - v_0)$ and $n^{1/2}\zeta_n$ are bounded in probability it will be sufficient to show that if u is an accumulation point of $n^{1/2}(\bar{v}_n - v_0)$ and ξ is an accumulation point of $n^{1/2}\zeta_n$, then with probability arbitrarily close to 1, $u = \bar{\omega}(\xi)$. Because \bar{S} is approximated by C and $\|\bar{v}_n - v_0\|$ is $O_p(n^{-1/2})$, there exists $w_n \in C$ such that $\|\bar{v}_n - v_0 - w_n\|$ is $o_p(n^{-1/2})$. Then with probability arbitrarily close to 1 we obtain that $u \in C$ and the equality $u = \bar{\omega}(\xi)$ follows by passing to the limit in (3.9) and using uniqueness of the minimizer $\bar{\omega}(\xi)$. \square

4. Discussion, applications and examples. In practical applications the required covariance matrix Ψ , the function $q(v)$ and the cone C are associated with the unknown value v_0 of the parameter vector and should be estimated. Notice that Ψ can be written in the equivalent form

$$(4.1) \quad \Psi = E\left\{[\nabla L(z, v_0, \lambda)][\nabla L(z, v_0, \lambda)]'\right\}, \quad \lambda \in \Lambda_0.$$

Let $\{\bar{v}_n\}$, $\bar{v}_n \in \mathcal{S}_n$, be a sequence of optimal solutions converging in probability to v_0 and let $\{\bar{\lambda}_n\}$ be a sequence of the corresponding vectors of Lagrange multipliers. Then the distance from $\bar{\lambda}_n$ to Λ_0 tends in probability to 0 and Ψ can be consistently estimated by

$$\Psi = n^{-1} \sum_{i=1}^n [\nabla L(z_i, \bar{v}_n, \bar{\lambda}_n)][\nabla L(z_i, \bar{v}_n, \bar{\lambda}_n)]'.$$

Estimation of the function $q(v)$ involves an estimation of the whole set Λ_0 of Lagrange multipliers of the program (\mathcal{P}_0) , which can be difficult or even impossible to obtain. In two particularly important cases $q(v)$ becomes a quadratic function. That is, if $\Lambda_0 = \{\lambda_0\}$ is a singleton or if the constraint functions g_i , $i \in I \cup J$, are linear, then $q(v) = v'Hv$ with $H = \nabla^2 l(v_0, \lambda_0)$ or $H = \nabla^2 \varphi(v_0)$, respectively. In both cases, under suitable regularity conditions, the required Hessian matrix H can be consistently estimated.

The critical cone C depends on the index set J of the inequality constraints active at the point v_0 . Determination of this index set may create a certain problem which needs further investigation.

Now let us suppose that the gradient vectors $\nabla g_i(v_0)$, $i \in I \cup J$, are linearly independent (*the linear independence condition*). Then $\Lambda_0 = \{\lambda_0\}$ is a singleton and the critical cone C can be written in the form

$$(4.2) \quad C = \{w: w' \nabla g_i(v_0) = 0, i \in I \cup J_+(\lambda_0); \\ w' \nabla g_i(v_0) \leq 0; i \in J_0(\lambda_0)\}.$$

Here the strong second-order sufficient conditions of Assumption C.2 mean that the Hessian matrix $H = \nabla^2 l(v_0, \lambda_0)$ is positive definite on the linear space $M(\lambda_0)$. Since $\|\bar{v}_n - v_0\|$ is $O_p(n^{-1/2})$ and by Assumption C.1, we have that

$$\nabla \psi_n(\bar{v}_n) - \nabla \varphi(\bar{v}_n) - \zeta_n = o_p(n^{-1/2}).$$

Then

$$\nabla \psi_n(\bar{v}_n) + \sum_{i \in I \cup J} \lambda_{0,i} \nabla g_i(\bar{v}_n) - \nabla l(\bar{v}_n, \lambda_0) - \zeta_n = o_p(n^{-1/2})$$

and hence

$$\sum_{i \in I \cup J} (\bar{\lambda}_{n,i} - \lambda_{0,i}) \nabla g_i(\bar{v}_n) + \zeta_n + H(\bar{v}_n - v_0) = o_p(n^{-1/2}).$$

It follows that under the Assumptions of Theorem 3.2, with Assumptions B.3 and B.5 replaced by the linear independence condition and Assumption C.2,

$$(4.3) \quad n^{1/2}(\bar{\lambda}_n - \lambda_0) = \bar{\alpha}(n^{1/2}\zeta_n) + o_p(1),$$

where $\bar{\alpha}(\zeta)$ denotes the vector of Lagrange multipliers corresponding to the minimizer $\bar{w}(\zeta)$ of the quadratic function $w'\zeta + \frac{1}{2}w'Hw$ over the cone C given by the linear constraints in the form (4.2). Notice that the uniqueness of $\bar{w}(\zeta)$ and $\bar{\alpha}(\zeta)$ is ensured by the linear independence condition and Assumption C.2.

Even under the linear independence condition and the strong second-order sufficient conditions the minimizer $\bar{w}(\zeta)$ is not linear in ζ and hence $n^{1/2}(\bar{v}_n - v_0)$ is not asymptotically normal, unless the critical cone C is a linear space. This is ensured by the following condition.

STRICT COMPLEMENTARITY CONDITION. All Lagrange multipliers $\lambda_{0,i}$, $i \in J$, associated with the inequality constraints of the program (\mathcal{P}_0) are positive.

In the case of strict complementarity the index set $J_0(\lambda_0)$ is empty and the critical cone C , given in (4.2), becomes a linear space. Consider the Jacobian matrix G whose columns are the gradient vectors $\nabla g_i(v_0)$, $i \in I \cup J$. Then $\bar{w} = \bar{w}(\zeta)$ and $\bar{\alpha} = \bar{\alpha}(\zeta)$ are solutions of the equations

$$(4.4) \quad \begin{bmatrix} H & G \\ G' & 0 \end{bmatrix} \begin{bmatrix} w \\ \alpha \end{bmatrix} = \begin{bmatrix} -\zeta \\ 0 \end{bmatrix}.$$

Notice that because of the second-order sufficient conditions (here Assumptions B.5 and C.2 coincide) the matrix in the left-hand side of (4.4) is nonsingular.

THEOREM 4.1. *Suppose that Assumptions A.1–A.6, B.1, B.2, B.4, B.5, C.1, D and the linear independence and strict complementarity conditions hold. Let $\{\bar{v}_n\}$ be a sequence of optimal solutions converging in probability to v_0 and $\{\bar{\lambda}_n\}$ be the corresponding sequence of Lagrange multipliers. Then $n^{1/2}(\bar{v}_n - v_0, \bar{\lambda}_n - \lambda_0)$ converges in distribution to a normal with mean 0 and the covariance matrix*

$$(4.5) \quad \Gamma = \begin{bmatrix} H & G \\ G' & 0 \end{bmatrix}^{-1} \begin{bmatrix} \Psi & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} H & G \\ G' & 0 \end{bmatrix}^{-1}.$$

As an example we now discuss applications of the developed theory to the maximum likelihood (ML) and robust regression methods. Consider a sample

x_1, \dots, x_n of i.i.d. vector observations with a common probability density function $f(x, \theta)$. The parameter vector θ is supposed to vary over a set S and it will be assumed that its true value θ_0 lies on the boundary of S . That is, in a neighbourhood of θ_0 the set S is defined by a number of equality and inequality constraints given in Assumption B.1 (with v replaced by θ). Minimization of minus the log-likelihood function over the set S leads to the ML-estimator $\hat{\theta}_n$. Under standard regularity conditions, in particular the identification condition, we have that θ_0 is the unique unconstrained minimizer of the expected value function $E\{-\log f(x, \theta)\}$ and hence the corresponding Lagrange multipliers are all zeros. The matrices $\Psi = E\{[\nabla \log f(x, \theta_0)][\nabla \log f(x, \theta_0)]'\}$ and $H = E\{-\nabla^2 \log f(x, \theta_0)\}$ are equal to each other and represent Fisher's information matrix. We obtain that $n^{1/2}(\hat{\theta}_n - \theta_0)$ is asymptotically distributed as the minimizer of the quadratic function $w'y + \frac{1}{2}w'Hw$ over C , where y is $N(0, H)$ and

$$(4.6) \quad C = \{w: w' \nabla g_i(\theta_0) = 0, i \in I; w' \nabla g_i(\theta_0) \leq 0, i \in J\}$$

gives the cone approximation of S at θ_0 .

If the information matrix H is nonsingular, then this result can be formulated in a slightly different form. Consider $\xi = -H^{-1}y$. Then ξ is $N(0, H^{-1})$ and since

$$2w'y + w'Hw = (\xi - w)'H(\xi - w) - \xi'H\xi,$$

we obtain that the asymptotic distribution of $n^{1/2}(\hat{\theta}_n - \theta_0)$ is the same as the distribution of the minimizer $\bar{w}(\xi)$ of $(\xi - w)'H(\xi - w)$ over C [cf. Chernoff (1954)]. If the linear independence condition also holds, then $n^{1/2}\bar{\lambda}_n$ is asymptotically distributed as vector $\bar{\alpha}(\xi)$ of Lagrange multipliers corresponding to the minimizer $\bar{w}(\xi)$. It is not difficult to verify that $\bar{w}(\xi)$ and $\xi - \bar{w}(\xi)$ are independent. It follows that the gradient of $(\xi - w)'H(\xi - w)$, with respect to w at $w = \bar{w}(\xi)$, is independent of $\bar{w}(\xi)$. Consequently, $\bar{w}(\xi)$ and $\bar{\alpha}(\xi)$ are independent and hence $n^{1/2}(\hat{\theta}_n - \theta_0)$ and $n^{1/2}\bar{\lambda}_n$ are *asymptotically independent*.

Here the minimizer $\bar{w}(\xi)$ represents the orthogonal projection $\bar{w}(\xi) = \text{proj}(\xi, C)$ of ξ onto C with respect to the scalar product $(u, w) = u'Hw$. This orthogonal projection can be decomposed into two parts as follows. Consider the linear space

$$L = \{w: w' \nabla g_i(\theta_0) = 0, i \in I \cup J\},$$

its orthogonal complement

$$L^\perp = \{u: (u, w) = 0 \text{ for all } w \in L\}$$

and the cone $C^* = C \cap L^\perp$. Since $L \subset C$, we have that C is representable as the direct sum of L and C^* and hence

$$(4.7) \quad \text{proj}(\xi, C) = \text{proj}(\xi, L) + \text{proj}(\xi, C^*)$$

[see, e.g., Stoer and Witzgall (1970)]. It can be calculated that

$$\text{proj}(\xi, L) = [I - H^{-1}G(G'H^{-1}G)^{-1}G']\xi,$$

where G is the Jacobian matrix whose columns are the gradient vectors $\nabla g_i(\theta_0)$, $i \in I \cup J$. It follows that the first term in the right-hand side of (4.7) is normal with mean 0 and the covariance matrix

$$(4.8) \quad P = H^{-1} - H^{-1}G(G'H^{-1}G)^{-1}G'H^{-1}.$$

Moreover, $\text{proj}(\xi, C^*)$ is equal to $\text{proj}(\tau, C^*)$, where $\tau = \text{proj}(\xi, L^\perp)$. It can be verified that $\eta_1 = \text{proj}(\xi, L)$ and τ are independent and hence the two terms in the right-hand side of (4.7) are independent. We obtain the following result. If the information matrix H is nonsingular, then $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges in distribution to the sum $\eta_1 + \eta_2$ of two independent variables, η_1 is normal with mean 0 and the covariance matrix P and $\eta_2 = \text{proj}(\xi, C^*)$. In this decomposition η_2 represents the "nonnormal" part of the distribution with the cone C^* giving the "nonlinear part" of C . For example, if there is only one active at θ_0 inequality constraint, then the cone C^* degenerates into a ray (half-line).

In the situation where the set S is defined by equality constraints only, we obtain that $n^{1/2}(\hat{\theta}_n - \theta_0, \bar{\lambda}_n)$ is asymptotically normal with mean 0 and the covariance matrix Γ given in (4.5). Since here $\Psi = H$, the matrix Γ can be represented in the form [cf. Aitchison and Silvey (1958), Section 5]

$$\Gamma = \begin{bmatrix} P & 0 \\ 0 & R \end{bmatrix},$$

where P is given in (4.8) and $R = (G'H^{-1}G)^{-1}$.

As a second example we discuss the robust regression method of Huber (1973). Consider the (generally nonlinear) model

$$(4.9) \quad y_i = h(x_i, \theta_0) + \varepsilon_i, \quad i = 1, \dots, n,$$

where $h: \mathbb{R}^p \times \mathbb{R}^k \rightarrow \mathbb{R}$ is a known function and $z_i = (x_i, y_i)$, $i = 1, \dots, n$, are i.i.d. observations. It will be assumed that x_i and the errors ε_i are independent and that θ_0 lies on the boundary of the permissible parameter set S . That is, S is defined near θ_0 by constraints as in Assumption B.1. Let $\rho(t)$ be a real-valued function of the scalar t and let $\hat{\theta}_n$ be a minimizer of the function

$$(4.10) \quad \psi_n(\theta) = n^{-1} \sum_{i=1}^n \rho(y_i - h(x_i, \theta))$$

over S . Standard examples of the function ρ are $\rho(t) = t^2$ and $\rho(t) = |t|$ corresponding to the least-squares and least absolute deviations methods, respectively. Some other examples of ρ are given in Huber (1973, 1981). Typically, ρ is a continuous, convex, piecewise differentiable function.

Consider the expected value function $\varphi(\theta) = E\{\rho(y - h(x, \theta))\}$. Under suitable regularity conditions we have that

$$\nabla \varphi(\theta) = E\{\nabla \rho(y - h(x, \theta))\} = -E\{\phi(y - h(x, \theta)) \nabla h(x, \theta)\},$$

with $\phi = \rho'$, and, in particular,

$$\nabla \varphi(\theta_0) = -E\{\phi(\varepsilon)\}E\{\nabla h(x, \theta_0)\}.$$

We assume that $E\{\phi(\varepsilon)\}$ is 0, and hence $\nabla \varphi(\theta_0) = 0$, and that θ_0 is a locally unique minimizer of φ over S . It follows that the Lagrange multipliers corresponding to θ_0 are all zeros. The covariance matrix Ψ is given here by $\Psi = \sigma^2 \Omega$, where $\sigma^2 = E\{\phi(\varepsilon)^2\}$ and

$$\Omega = E\{[\nabla h(x, \theta_0)][\nabla h(x, \theta_0)]'\}.$$

If ρ is twice continuously differentiable, then under suitable regularity conditions the Hessian matrix $H = \nabla^2 \varphi(\theta_0)$ exists and $H = \kappa \Omega$, where $\kappa = E\{\rho''(\varepsilon)\}$. Then assuming that $\kappa > 0$ and Ω is nonsingular, we obtain that $n^{1/2}(\hat{\theta}_n - \theta_0)$ is asymptotically distributed as the minimizer $\bar{w}(\xi)$ of $(\xi - w)' \Omega (\xi - w)$ over C , where ξ is $N(0, \kappa^{-2} \sigma^2 \Omega^{-1})$ and C is the approximating cone of S at θ_0 . Under the linear independence condition, $n^{1/2} \bar{\lambda}_n$ is asymptotically distributed as the corresponding vector $\bar{\alpha}(\xi)$ of Lagrange multipliers and $n^{1/2}(\hat{\theta}_n - \theta_0)$ and $n^{1/2} \bar{\lambda}_n$ are asymptotically independent.

In situations where ρ is nondifferentiable, the existence of $H = \nabla^2 \varphi(\theta_0)$ can often be ensured by smoothness conditions on the distribution of ε . Consider, for example, the least absolute deviations method $\rho(t) = |t|$. Suppose that the distribution function F of ε has median 0, that F is continuous and has continuous and positive density f at zero, that $h(x, \cdot)$ is twice continuously differentiable for a.e. x , that $\|\nabla h(x, \theta)\|^2$ and $\|\nabla^2 h(x, \theta)\|^2$ are dominated by integrable functions for all θ in a neighbourhood of θ_0 and that the matrix Ω is nonsingular. Then $H = \kappa \Omega$, where $\kappa = 2f(0)$, and $\Psi = \Omega$. Moreover, the above conditions imply Assumption C.3 and hence Assumption C.1 follows. Therefore $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges in distribution to the minimizer $\bar{w}(\xi)$ of $(\xi - w)' \Omega (\xi - w)$ over C , where ξ is $N(0, [2f(0)]^{-2} \Omega^{-1})$. For the unconstrained case and linear models this result is due to Bassett and Koenker (1978).

Finally, suppose that we are interested in testing a null hypothesis $H_0: \theta \in S_0$ against an alternative $H_1: \theta \in S_1$ for the model (4.9). Suppose that the true value θ_0 is a boundary point of S_0 and (or) S_1 , and consider the test statistic

$$T = 2n \left[\min_{\theta \in S_0} \psi_n(\theta) - \min_{\theta \in S_1} \psi_n(\theta) \right],$$

where $\psi_n(\theta)$ is defined in (4.10). It follows from the result of Theorem 3.1 that under the Assumptions specified above, the test statistic T is asymptotically distributed as

$$U = \min_{w \in C_0} (\zeta - w)' \Omega (\zeta - w) - \min_{w \in C_1} (\zeta - w)' \Omega (\zeta - w),$$

where ζ is $N(0, \kappa^{-1} \sigma^2 \Omega^{-1})$ and C_0 and C_1 are the approximating cones of S_0 and S_1 at θ_0 , respectively. Suppose that the cones C_0 and C_1 are convex, C_0 is contained in C_1 and either C_0 or C_1 is a linear space. Then $\kappa \sigma^{-2} U$ and hence asymptotically $\kappa \sigma^{-2} T$ are distributed as a mixture of central chi-square distribu-

tions [see, e.g., Shapiro (1988b) for details and a discussion of the so-called chi-bar-squared distributions]. Notice that here the correction multiplier $\kappa\sigma^{-2}$ is the same as in the unconstrained case [cf. Schrader and Hettmansperger (1980) and Koenker and Bassett (1982)].

REFERENCES

- AITCHISON, J. and SILVEY, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29** 813–828.
- BAHADUR, R. R. (1967). Rates of convergence of estimators and test statistics. *Ann. Math. Statist.* **38** 303–324.
- BASSETT, G. and KOENKER, R. (1978). Asymptotic theory of least absolute error regression. *J. Amer. Statist. Assoc.* **73** 618–622.
- BOOS, D. D. and SERFLING, R. J. (1980). A note on differentials and the CLT and LIL for statistical functions, with applications to M -estimates. *Ann. Statist.* **8** 618–624.
- CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25** 573–578.
- CLARKE, B. R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist.* **11** 1196–1205.
- CLARKE, F. H. (1983). *Optimization and Nonsmooth Analysis*. Wiley, New York.
- DUPAČOVÁ, J. and WETS, R. (1988). Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16** 1517–1549.
- FERNHOLZ, L. T. (1983). *Von Mises Calculus for Statistical Functionals. Lecture Notes in Statist.* **19**. Springer, New York.
- GAUVIN, J. (1977). A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Math. Programming* **12** 136–138.
- HESTENES, M. R. (1975). *Optimization Theory—The Finite Dimensional Case*. Wiley, New York.
- HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.
- HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press.
- HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.
- HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.
- IOFFE, A. D. (1979). Necessary and sufficient conditions for a local minimum. 3. Second order conditions and augmented duality. *SIAM J. Control Optim.* **17** 266–288.
- JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643.
- KING, A. J. (1986). Asymptotic behavior of solutions in stochastic optimization: Nonsmooth analysis and the derivation of non-normal limit distributions. Ph.D. dissertation, Dept. Applied Mathematics, Univ. Washington.
- KOENKER, R. and BASSETT, G. (1982). Tests of linear hypotheses and l_1 estimation. *Econometrica* **50** 1577–1583.
- LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Publ. Statist.* **1** 277–330.
- MANGASARIAN, O. L. and FROMOVITZ, S. (1967). The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.* **7** 37–47.
- PFANZAGL, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika* **14** 249–272.
- ROCKAFELLAR, R. T. (1987). Pseudo-differentiability of set-valued mappings and its applications in optimization. Unpublished.
- ROCKAFELLAR, R. T. and WETS, R. (1984). Variational systems, an introduction. *Multifunctions and Integrands. Lecture Notes in Math.* **1091** 1–54. Springer, New York.
- SCHRADER, R. M. and HETTMANSPERGER, T. P. (1980). Robust analysis of variance based on a likelihood ratio criterion. *Biometrika* **67** 93–101.

- SHAPIRO, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72** 133–144.
- SHAPIRO, A. (1987). On differentiability of metric projections in \mathbb{R}^n , 1. Boundary case. *Proc. Amer. Math. Soc.* **99** 123–128.
- SHAPIRO, A. (1988a). Sensitivity analysis of nonlinear programs and differentiability properties of metric projections. *SIAM J. Control Optim.* **26** 628–645.
- SHAPIRO, A. (1988b). Towards a unified theory of inequality constrained testing in multivariate analysis. *Internat. Statist. Rev.* **56** 49–62.
- STOER, J. and WITZGALL, C. (1970). *Convexity and Optimization in Finite Dimensions* 2. Springer, Berlin.
- TSYBAKOV, A. B. (1981). A method of minimization of the empirical risk in identification problems. *Automat. Remote Control* **9** 1196–1203.
- WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.
- WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25.

DEPARTMENT OF MATHEMATICS
AND APPLIED MATHEMATICS
UNIVERSITY OF SOUTH AFRICA
P.O. Box 392
PRETORIA
0001 SOUTH AFRICA