ASYMPTOTIC BEHAVIOR OF STATISTICAL ESTIMATORS AND OF OPTIMAL SOLUTIONS OF STOCHASTIC OPTIMIZATION PROBLEMS, II

Jitka Dupačová
*Roger J.-B. Wets**
*Supported in part by a grant of the National Science Foundation.

January 1987
WP-87-9

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

This paper supplements the results of a new statistical approach to the problem of incomplete information in stochastic programming. The tools of nondifferentiable optimization used here, help to prove the consistency and asymptotic normality of (approximate) optimal solutions without unnatural smoothness assumptions. This allows the theory to take into account the presence of contraints.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

ASYMPTOTIC BEHAVIOR OF STATISTICAL ESTIMATORS AND OF OPTIMAL SOLUTIONS OF STOCHASTIC OPTIMIZATION PROBLEMS, II

Jitka Dupačová [1] *and Roger J.-B. Wets* [2]

[1]Mathematical Statistics, Charles University, Prague
[2]Mathematics, University of California, Davis

INTRODUCTION

These results complement those of Dupačová and Wets (1986). We use the same notation and identical set-up, the reader is thus referred to that article where he shall find definitions and the consistency results. We even continue the numbering of sections and equations, so we start with Section 4.

4 ASYMPTOTICS, CONVERGENCE RATES

In Section 3 of Dupačová and Wets (1986) we exhibited sufficient conditions for the convergence with probability 1 of the estimators $\{x^{\nu}: Z \to R^n, \nu = 1, \ldots\}$ to x^*, the optimal solution of the limit problem. Here we go one step further and analyze the rate of convergence in probabilistic terms. The argumentation is related to that of Huber (1967), adapted to fit the more general class of problems under consideration; this was already the pattern followed by Solis and Wets (1981), in the unconstrained case and by Dupačová (1983a, 1983b, 1984) for stochastic programs with recourse under special assumptions. We extend the results of Huber (1967) in a number of directions: (i) we allow for constraints, (ii) the probability measures converging to P are not necessarily the empirical measures, and (iii) there are no differentiability assumptions on the likelihood (criterion) function (in terms of Huber's set-up, this would correspond to the case when his function Ψ is not uniquely determined, see Section 3 of Huber (1967)).

One way to look at the results of this section is to view them as providing limiting conditions under which one may be able to obtain asymptotic normality. Note that when there are constraints, one should usually not expect the asymptotic distribution to be Gaussian. This, in turn, allows us to obtain certain probabilistic estimates for the convergence "rates". To approximate the distribution of x^ν, to obtain confidence intervals for example, we need an assertion that a suitably normalized sequence converges in distribution to a *nondegenerate* random vector. The normalizing coefficients need not be unique but they suggest a rate of convergence. Following Lehmann (1983) we shall say that the sequence $x^\nu - x^*$ *goes to 0 with the rate of convergence* $1/k_\nu$ if $k_\nu \to \infty$ as $\nu \to \infty$ and if there is a continuous distribution function H such that

$$P\{k_\nu \|x^\nu - x^*\| \le a\} \to H(a) \text{ as } \nu \to \infty .$$

We begin by a quick review of the main definitions and results that provides us with a good notion for the subgradients of not necessarily differentiable functions. Any assumption of differentiability of $f(\cdot, \xi)$, would be inappropriate and would for one reason or another eliminate from the domain of applicability all the examples mentioned in Section 2. To handle the lack of differentiability, we rely on the theory of subdifferentiability developed to handle nonsmooth functions, see Clarke (1983), Rockafellar (1983), Aubin and Ekeland (1984).

The *contingent derivative* of a lower semicontinuous function $h : R^n \to (-\infty, +\infty]$ at x, a point at which h is finite, with respect to the direction y is

$$h'(x; y) := \text{epi-}\liminf_{t \downarrow 0} \frac{h(x + ty) - h(x)}{t}$$

using the convention $\infty - \infty = \infty$. It is not difficult to see that h′ is always well defined with values in the extended reals. If $x \notin \text{dom}\, h$, then $h'(x; \cdot) = \infty$, otherwise

$$h'(x; y) = \liminf_{\substack{y' \to y \\ t \downarrow 0}} \frac{h(x + ty') - h(x)}{t}$$

The *(upper)epi derivative of* h *at* x, where h is finite, *in direction* y, is the epi-limit superior of the collection $\{h'(x'; \cdot), x' \in R^n\}$ at x, i.e.

$$h^\uparrow(x; \cdot) := \text{epi-}\limsup_{x' \to x} h'(x'; \cdot)$$

$$h^\uparrow(x; y) = \inf_{\{y' \to y\}} \limsup_{\{x' \to x\}} h'(x'; y')$$

where by writing $\{x' \to x\}$ and $\{y' \to y\}$ we mean that the infimum must be taken with respect to all nets – or equivalently here sequences – converging to x and y, see Aubin and Ekeland (1984), Chapter 7, Section 3.

It is remarkable that if h is proper, and $x \in \text{dom } h$, the function $y \mapsto h^{\dagger}(x; \cdot)$ is sublinear and l.sc. [Theorems 1 and 2, Rockafellar (1980)]. Moreover, if h is Lipschitzian around x, then $h^{\dagger}(x; \cdot)$ is everywhere finite (and hence continuous); in particular if h is continuously differentiable at x then $h^{\dagger}(x; y)$ is the directional derivative of h in direction y, and if h is convex in a neighborhood of x, then

$$h^{\dagger}(x; y) = \lim_{t \downarrow 0} \frac{h(x + ty) - h(x)}{t}$$

is the one-sided directional derivative in direction y. The sublinearity and lower semicontinuity of $h^{\dagger}(x; \cdot)$ makes it possible to define the notion of a subgradient of h at x, by exploiting the fact that there is a one-to-one correspondence between the proper lower semicontinuous, sublinear functions g and the nonempty closed convex subsets C of R^n, given by

$$g(y) = \sup_{v \in C} v \cdot y \quad \text{for all } y \in R^n \ ,$$

and

$$C = \{v \in R^n | v \cdot y \leq g(y) \quad \text{for all } y \in R^n\}$$

see Rockafellar (1970). Assuming that $h^{\dagger}(x; \cdot)$ is proper, let $\partial h(x)$ be the nonempty closed convex set such that for all y,

$$h^{\dagger}(x; y) = \sup_{v \in \partial h(x)} v \cdot y \ .$$

Every vector in $v \in \partial h(x)$ is a ***subgradient*** of h at x. If h is smooth (continuously differentiable) then

$$\partial h(x) = \{\nabla h(x), \quad \text{the gradient of h at } x\} \ ;$$

if h is convex, then

$$\partial h(x) = \{v | h(x + y) \geq h(x) + v \cdot y \quad \text{for all } y \in R^n\}$$

is the usual definition of the subgradients of a convex function. More generally if h is locally Lipschitz at x, then

$$\partial h(x) = \text{co}\{v = \lim_{x' \to x} \nabla h(x') | h \text{ is smooth at } x'\} \ .$$

For the proofs of these preceding assertions and further details, consult Rockafellar (1981) and Aubin and Ekeland (1984).

Before we return to the problem at hand, we state the results about the additivity of subgradients that are relevant to our analysis, we begin with a general result that shows that the derivatives and subgradient functions of the random l.sc. function f and the expectation functionals $E^{\nu}f$ and Ef have the appropriate measurability properties.

THEOREM 4.1 *Suppose* $h: R^n \times \Xi \to \overline{R}$ *is a random lower semicontinuous function. Then, so are its contingent derivative and its (upper) epi-derivative. Moreover, for all* $x \in R^n$, $\xi \mapsto \partial h(x, \xi)$ *is a random closed convex set.*

PROOF Theorem of Salinetti and Wets (1981) tells us that the lim sup and lim inf of sequences of random closed sets (closed-valued measurable multifunctions) are random closed sets. Since the epigraphs of the epi-lim sup and epi-lim inf are respectively the lim inf and lim sup of the corresponding sequence of epigraphs (see for example, Section 2 of Dolecki, Salinetti and Wets (1983)), the assertion about the derivatives follows from their definitions and property (3.4) of random lower semicontinuous functions. Since $h^{\uparrow}(x; \cdot, \xi)$ is sublinear, it follows that its conjugate – another random l.sc. function, Rockafellar (1976) – is the indicator of the random closed convex set $\xi \mapsto \partial f(x, \xi)$. □

Our interest in subdifferential theory is conditioned by the fact that for a very large class of functions (with values in the extended reals), we can characterize optimality in terms of a differential inclusion, a point x^0 that minimizes the proper l.sc. function on R^n, must necessarily satisfy

$$0 \in \partial h(x^0) \ ,$$

if h is convex this is also a sufficient condition. There is a subdifferential calculus, but for our purposes the following results about the subdifferentials on sums of l.sc. functions is all we need. We say that a l.sc. function is ***subdifferentially regular at*** x if $h'(x; \cdot) = h^{\uparrow}(x; \cdot)$. If h is convex or subsmooth on a neighborhood of x, thus in particular if h is C^1 at x, it is subdifferentially regular at x; h is ***subsmooth*** on a neighborhood V of x, if for all $y \in V$

$$h(y) = \max_{t \in T} \varphi_t(y)$$

where T is a compact topological space, each φ_t is of class C^1, and both $\varphi_t(x)$ and $\nabla_x \varphi_t(x)$ are continuous with respect to (t, x). If h is subsmooth on an open set U, it is also locally Lipschitz on U, Clarke (1975).

LEMMA 4.2 Rockafellar (1979) *Suppose* h_1 *and* h_2 *are l.sc. functions on* R^n *and* x *a point at which both* h_1 *and* h_2 *are finite. Suppose that* $\text{dom}\, h_1(x;\cdot)$ *is nonempty and* h_2 *is locally Lipschitz at* x. *Then*

$$\partial(h_1 + h_2)(x) \subset \partial h_1(x) + \partial h_2(x) .$$

Moreover equality holds if h_1 *and* h_2 *are subdifferentially regular at* x.

LEMMA 4.3 Clarke (1983) *Let* U *be an open subset of* R^n, *and suppose* $h : U \times \Xi \to R$ *is measurable with respect to* ξ *and there exist a summable function* β *such that for all* x^0, x^1 *in* U *and* $\xi \in \Xi$

$$|h(x^0, \xi) - h(x^1, \xi)| \leq \beta(\xi)\|x^0 - x^1\| .$$

Suppose moreover that for some $\bar{x} \in U$, $Eh(\bar{x})$ *is finite. Then* Eh *is finite and Lipschitz on* U, *and for all* x *in* U,

$$\partial Eh(x) \subset E\{\partial h(x, \xi)\} = \int \partial h(x, \xi) P(d\xi) .$$

Moreover, equality holds whenever $h(\cdot, \xi)$ *is a.s. subdifferentially regular at* x, *in which case also* Eh *is subdifferentially regular at* x.

Theorem 4.1 shows that $\xi \mapsto \partial h(x, \xi)$ is a random (nonempty) closed set; it is easy to verify that under the assumptions of Lemma 4.3, h is a random l.sc. function on $U \times \Xi$. In fact for all ξ, $\partial h(x, \xi)$ is a compact subset of R^n, see Proposition 2.1.2 of Clarke (1983). The integral of a random closed set Γ defined on Ξ (with values in the closed subsets of R^n) is

$$\int \Gamma(\xi) P(d\xi) := \{x = \int s(\xi) P(d\xi) | s(\xi) \in \Gamma(\xi) \quad a.s., \ s \in L^1\} ,$$

see Aumann (1965). If P is absolutely continuous, and Γ is integrably bounded (the function $\xi \mapsto \sup \{\|x\| \,|\, \|x\| \in \Gamma(\xi)\}$ is summable), then $\int \Gamma(\xi) P(d\xi) = \int \text{co}\, \Gamma(\xi) P(d\xi)$ is convex, where $\text{co}\, \Gamma(\xi)$ is the convex hull of ξ. If Γ is uniformly bounded then $\int \Gamma(\xi) P(d\xi)$ is a compact subset of R^n.

We shall be working with the same set-up as in Section 3, but with a somewhat more restricted class of random l.sc. functions. Instead of Assumption 3.4, we shall be using the following one:

ASSUMPTION 4.4 *The function* $f: R^n \times \Xi \to (-\infty, \infty]$ *is of the following type:*

$$f(x, \xi) = f_0(x, \xi) + \Psi_S(x)$$

where Ψ_S *is the indicator function of the closed nonempty set* $S \subset R^n$, *i.e.,*

$$\Psi_S(x) = 0 \quad \text{if } x \in S, \quad \text{and } = \infty \text{ otherwise} \ ,$$

and f_0 *is a finite valued function on* $R^n \times \Xi$, *with*

$$\xi \mapsto f_0(x, \xi) \text{ relatively continuous on } \Xi \ ,$$

for all $x \in S$, *and any open set* U *that contains* S, *the function*

$$x \to f_0(x, \xi) \text{ is locally Lipschitz}$$

for all $\xi \in \Xi$, *and such that to any bounded open set* V *there corresponds a* P-*summable function* β *such that for any pair* x^0, x^1 *in* V*:*

$$|f_0(x^0, \xi) - f_0(x^1, \xi)| \leq \beta(\xi) \cdot |x^0 - x^1| \ . \tag{4.1}$$

The only condition of Assumption 3.4 that does not appear explicitly in Assumption 4.4, either in exactly the same form or in a stronger form, is the lower semicontinuity of $f(\cdot, \xi)$ on R^n for all ξ in Ξ. But that is an immediate consequence of the fact that $f_0(\cdot, \xi)$ is locally Lipschitz and S is closed. Thus, f is a proper random lower semicontinuous function, and so is also f_0. Moreover all the results and the observations of Section 3 are immediately applicable to both f and f_0, as well as to the corresponding expectation functionals. Of course these functions will now have Lipschitz properties that we shall exploit in our analysis. In the convex case it might be possible to work with weaker restrictions on the function f by relying on finer results about the additivity of subgradients, see Rockafellar and Wets (1982). Combining the results of Section 3, with those about subgradients of random l.sc. functions, in particular Lemma 4.3, we can show that:

LEMMA 4.5 *Under Assumptions* 4.4 *and* 3.5, *we have that* μ-*a.s.* Ef *and* $\{E^\nu f, \nu = 1, \dots\}$ *are proper lower semicontinuous functions that are locally Lipschitz on* S. *Moreover we always have*

$$\partial E f_0(x) \subset E\{\partial f_0(x, \xi)\} = \int_\Xi f_0(x, \xi) P(d\xi) \ ,$$

and for $\nu = 1, \dots,$

$$\partial E^{\nu} f_0(x, \zeta) \subset \int_{\Xi} f_0(x, \xi) P^{\nu}(d\xi, \zeta) \quad \mu\text{-}a.s,$$

with equality if for all ξ, $f_0(\cdot, \xi)$ *is subdifferentially regular at* x. *Moreover, if* $x \in S$

$$\partial Ef(x) \subset \partial Ef_0(x) + \partial \Psi_S(x) \ ,$$

and for $\nu = 1, \ldots,$

$$\partial E^{\nu} f(x, \zeta) \subset \partial E^{\nu} f_0(x, \zeta) + \partial \Psi_S(x) \quad \mu\text{-}a.s. \ ,$$

with equality if Ψ_S *and for all* ξ, $f_0(\cdot, \xi)$ *are subdifferentially regular at* x.

REMARK 4.6 If $x \in S$, $\partial \Psi_S(x)$ is the polar of the tangent cone $T_S(x)$ to S at x, Clarke (1975). If S is a differentiable manifold, then $\partial \Psi_S(x)$ is the orthogonal complement of the tangent space at x and, of course, Ψ_S is differentially regular at x. This is also the case when S is locally convex at x, or if x belongs to the boundary of S and this boundary is locally a differentiable manifold. More generally, Ψ_S is subdifferentially regular at x, if the tangent cone to S at x, has the following representation

$$T_S(x) = \{y \mid \exists \, \lambda_k \downarrow 0, \; y^k \longrightarrow y \; . \text{ with } \; x + \lambda_k y^k \in S\} \ .$$

So far, we have limited our assumptions to certain continuity properties of the function f with respect to x and ξ. In order to derive the asymptotic behavior we need to impose some additional conditions about the way the information collected from the samples is included in the approximating probability measures P^{ν}, in particular on how it affects the subgradients of the functions $E^{\nu}f$. Let us introduce the following notation: $u_0(x, \xi)$ will always denote an element of $\partial f_0(x, \xi)$ and $v_S(x)$ an element of $\partial \Psi_S(x)$. In view of Theorem 4.1 and Lemma 4.5 if $x \in S$, we always have that $v(x) \in \partial Ef(x)$ implies the existence of $v_S(x) \in \partial \Psi_S(x)$ and $u_0(x, \cdot)$ measurable with $u_0(x, \xi) \in \partial f_0(x, \xi)$ P-*a.s.* such that

$$v(x) = v_0(x) + v_S(x) = E\{u_0(x, \xi)\} + v_S(x)$$

Moreover similar formulas hold μ-*a.s.* if the integration is with respect to $P^{\nu}(\cdot, \zeta)$ instead of P. If the functions $f_0(\cdot, \xi)$, as well as Ψ_S, are a.s. subdifferentially regular, then a type of converse statement also holds. We have that

$$0 \in \partial Ef(x^*)$$

implies the existence of $v_s^* \in \partial\Psi_s(x^*)$ and of a random function $u_0(x^*, \cdot)$ from Ξ to R^n with $u_0(x^*, \cdot) \in \partial f_0(x^*, \xi)$ P-*a.s.* such that

$$0 = E\{u_0(x^*, \xi)\} + v_s(x^*) \ . \tag{4.2}$$

Similarly,

$$0 \in \partial E^\nu f(x^\nu) \ ,$$

means that there exist $v_s(x^\nu) \in \partial\Psi_s(x^\nu)$, and a random function $u_0(x^\nu, \cdot)$ from Ξ to R^n with $u_0(x^\nu, \cdot) \in \partial f_0(x^\nu, \cdot)$ P^ν-*a.s.* such that

$$\begin{aligned} 0 &= v_0^\nu(x^\nu) + v_s(x^\nu) \\ &= E^\nu\{u_0(x^\nu, \xi)\} + v_s(x^\nu) \ . \end{aligned} \tag{4.3}$$

ASSUMPTION 4.7 Statistical Information. *The probability measures* $\{P^\nu, \nu = 1, \ldots\}$ *are such that for some* $v^\nu \in \partial E^\nu f(x^*, \zeta)$ *and* $v \in \partial Ef(x^\nu(\zeta))$

(i) $\sqrt{\nu}[v^\nu(x^*, \zeta) + v(x^\nu(\zeta))]$ *converges to* 0 *in probability;*

(ii) $\sqrt{\nu}[v_s(x^\nu(\zeta)) - v_s(x^*)]$ *converges to* 0 *in probability;*

(iii) $v^\nu(x^*, \zeta)$ *is asymptotically Gaussian with distribution function* $N(0, \Sigma_1)$ *where* Σ_1 *is the covariance matrix.*

Moreover

(iv) Ef_0 *is twice continuously differentiable at* x^* *with nonsingular Hessian* H.

Before we proceed with the main result of this section, let us examine some of the implications of these assumptions. The assumption that Ef_0 is of class C^2 is of course rather restrictive, but without it it maybe hard to obtain asymptotic normality; a more general class of limiting distributions (piecewise normal) for constrained problems has recently been identified by King and Rockafellar (1986). Note that this does not require that f_0 be of class C^2.

The assumption that $\sqrt{\nu}[v_s(x^\nu(\zeta)) - v_s(x^*)]$ converges in probability to 0, essentially means that the convergence of x^ν to x^* is "smooth". Of course, it will be satisfied if x^* belongs to the interior of the set S of constraints, in which case $v_s(x^*)$ and μ-*a.s.* $v_s(x^\nu(\zeta))$ are zero for ν sufficiently large. It will also be trivially satisfied if the binding constraints are linear and, x^* and μ-*a.s.* $x^\nu(\zeta)$, belong to the linear variety spanned by these constraints. In fact, we can expect this condition to be satisfied unless the vector x^* is a boundary point at which the boundary has high curvature, in particular at point at which the boundry is not smooth.

The condition about asymptotic normality of the subgradients $v^\nu(x^*)$ is best understood in the following context. Suppose condition (ii) is satisfied, in fact let us assume that $v_s(x^*) = v_s(x^\nu(\zeta))$ a.s. And suppose also that P^ν is the empirical distribution. Then $\|v^\nu(x^*, \zeta)\|$ records the error of the estimate of the subgradients of Ef at x^*; note that $0 \in \partial Ef(x^*)$.

The first condition yields an estimate for the errors of the subgradients of $E^\nu f$ at x^* and Ef at $x^\nu(\zeta)$. The assumption is that enough information is collected so as to guarantee a certain convergence rate to 0. This is a crucial assumption and after the statement of the theorem will return to this condition and give sufficient conditions that imply it.

THEOREM 4.8 *Under Assumptions* 4.4, 3.5 *and* 4.7, $\sqrt{\nu}(x^\nu(\cdot) - x^*)$ *is asymptotically normal with distribution* $N(0, \Sigma)$ *where* $\Sigma = H^{-1}\Sigma_1(H^{-1})^T$.

PROOF Since Ef_0 is assumed to be C^2, and $x^\nu(\cdot)$ converges to x^*, for ν sufficiently large,

$$\nabla Ef_0(x^\nu) - \nabla Ef_0(x^*) = H(x^\nu - x^*) + o(\|x^\nu - x^*\|) \quad \mu\text{-}a.s.$$

Now, since $v(x^*) = 0$,

$$\sqrt{\nu}(\nabla Ef_0(x^\nu) - \nabla Ef_0(x^*)) = \sqrt{\nu}[v(x^\nu) + v^\nu(x^*)] - \sqrt{\nu}\, v^\nu(x^*)$$

$$+ \sqrt{\nu}[v_s(x^*) - v_s(x^\nu)]$$

By Assumption 4.7 the first term converges to zero in probability, the second one converges in distribution to $N(0, \Sigma_1)$ and the third one converges in probability to zero. Hence $\sqrt{\nu}[\nabla Ef_0(x^\nu) - \nabla Ef_0(x^*)]$ converges in distribution to $N(0, \Sigma_1)$ (Slutsky's Theorem). This is then also the asymptotic distribution of $\sqrt{\nu}H(x^\nu - x^*)$. The result now follows by the nonsingularity of the matrix H. □

The remainder of this section, is devoted to recording certain conditions that will yield condition (i) of Assumption 4.7. In view of Markov's inequality it would suffice to control the variance of $\|v^\nu(x^*) + v(x^\nu)\|$ to obtain the desired convergence. More generally we have the following:

LEMMA 4.9 *Suppose that* $E_\mu\{v^\nu(x^*, \zeta)\} = 0$, *that*

$$E_\mu\{\|v_0^\nu(x^*, \zeta) - v_0(x^*)\|^2\} \le \beta^2/\nu$$

and that

$$\frac{\|v^\nu(x^*, \zeta) + v(x^\nu(\zeta))\|}{\nu^{-1/2} + \|v(x^\nu(\zeta))\|} \quad \textit{converges to 0 in probability } (\mu) \ .$$

Then, under Assumptions 4.4 *and* 3.5, *for any (measurable) selections* $v^\nu(x^*, \cdot)$ *with*

$$v^\nu(x^*, \zeta) \in \partial E^\nu f(x^*, \zeta) \quad \mu\text{-a.s.} \ ,$$

such that μ*-a.s.* $v(x^*) = 0$, *the random vector*

$$\sqrt{\nu}\,[v^\nu(x^*, \zeta) + v(x^\nu(\zeta))]$$

converges to 0 *in probability as* ν *goes to* ∞.

PROOF We need to show that to any $\varepsilon > 0$, there corresponds ν_ε such that for all $\nu \ge \nu_\varepsilon$,

$$\mu[|v^\nu(x^*) + v(x^\nu)| \ge \nu^{-1/2}\delta_\varepsilon] \le \varepsilon$$

where δ_ε goes to zero as ε goes to zero.

Chebychev's inequality and the assumptions of the Theorem imply that for all α,

$$\mu[\|v^\nu(x^*, \zeta)\| > \alpha\nu^{-1/2}] \le \nu\alpha^{-2}E_\mu\|v^\nu(x^*, \zeta)\|^2 \le \nu(\beta/\alpha)^2 \ .$$

And hence with $\alpha^2 = 2\beta^2/\varepsilon$, we have

$$\mu[\|v^\nu(x^*)\| > \nu^{-1/2}\beta\sqrt{2}/\sqrt{\varepsilon}] \le \varepsilon/2 \ .$$

This, in conjunction with the last one of our assumptions, i.e.,

$$\mu[\|v^\nu(x^*) + v(x^\nu)\| \ge \varepsilon(\nu^{-1/2} + \|v(x^\nu)\|)] \le \varepsilon/2 \ , \tag{4.4}$$

implies that the events

$$\|v^\nu(x^*) + v(x^\nu)\| < \varepsilon(\nu^{-1/2} + \|v(x^\nu)\|) \quad \text{and} \quad \|v^\nu(x^*)\| \le \nu^{-1/2}\beta\sqrt{2/\varepsilon} \ ,$$

have probability (μ) at least $1 - \varepsilon$. Thus for ε small,

$$\mu[\|v(x^\nu)\| \le \nu^{-1/2}(\beta + \varepsilon)/(1 - \alpha)] > 1 - \varepsilon \ ,$$

since $\|v(x^\nu)\| \le \|v^\nu(x^*) + v(x^\nu)\| + \|v^\nu(x^*)\|$. This, together with (4.4), gives

$$\mu[\|v^\nu(x^*) + v(x^\nu)\| < \nu^{-1/2}\varepsilon(1 + (\beta + \varepsilon)/(1 - \varepsilon))] > 1 - \varepsilon$$

and this yields the desired expression with $\delta_\varepsilon = \varepsilon(1 + (\beta + \varepsilon)/(1 - \varepsilon))$. □

It is easy to see why the condition $E_\mu\{v^\nu(x^*, \zeta)\} = 0$ would be satisfied when the P^ν are providing moment estimates that are at least as good as the empirical distributions. The same holds for the second assumption in Lemma 4.9, there is a reduction in the variance estimate that is at least as significant as that which would be attained by using the empirical distribution. Finally, the last assumption of Lemma 4.9 means that we can allow for a certain slack in the convergence in probability of $\sqrt{\nu}\|v^\nu(x^*) + v(x^\nu)\|$ to zero. In the Appendix, we give a derivation of this condition by using assumptions that are related to those used by Huber (1967).

5 ASYMPTOTIC LAGRANGIANS

The results of Sections 3 and 4 can be extended to Lagrangians by relying on the theory of epi/hypo-convergence for saddle functions, Attouch and Wets (1983a). This gives us not just asymptotic properties for the sequence $\{x^\nu, \nu = 1, \dots\}$ of optimal solutions but also for the associated Lagrange multipliers.

We now introduce an explicit representation of the constraints in the formulation of the problem:

$$\text{minimize } z = E\{f_0(x, \xi)\} \tag{5.1}$$

$$\text{subject to } \quad f_i(x) \le 0 \ , \quad i = 1, \dots, s,$$

$$f_i(x) = 0 \ , \quad i = s + 1, \dots, m$$

$$x \in X \subset R^n$$

where for $i = 1, \dots, m$, the f_i are finite-valued continuous functions, f_0 is a finite-valued random l.sc. function, and X is a closed subset of R^n. When instead of P, we use P^ν then the objective function is modified and becomes

$$E^{\nu}f_0(x) = \int_{\Xi} f_0(x, \xi) P^{\nu}(d\xi) \ .$$

The (standard) associated Lagrangians are

$$L(x, y) = \begin{cases} Ef_0(x) + \sum_{i=1}^{m} y_i f_i(x) & \text{if } x \in X \ , \text{ and } y_i \geq 0 \ , \text{ for } i = 1, \ldots, s \ , \\ \infty \quad \text{if } x \notin X \ , \\ -\infty \quad \text{otherwise} \ . \end{cases}$$

and

$$L^{\nu}(x, y) = \begin{cases} E^{\nu}f_0(x) + \sum_{i=1}^{m} y_i f_i(x) & \text{if } x \in X \ , \text{ and } y_i \geq 0 \ , \text{ for } i = 1, \ldots, s \ , \\ \infty \quad \text{if } x \notin X \ , \\ -\infty \quad \text{otherwise} \ . \end{cases}$$

Consistency can be studied in the same framework as that described at the beginning of Section 3. The Lagrangians L^{ν} are then also dependent of ζ. Suppose that f_0 satisfies the conditions of Assumption 3.4; Note that some of these conditions are automatically satisfied since f_0 is a finite-valued random l.sc. function. Suppose also that the $\{P^{\nu}, \nu = 1, \ldots\}$ satisfy Assumption 3.5 with f_0 replacing f (in the asymptotic negligibility condition), then it follows from Lemma 3.6 that μ-a.s. the Lagrangians L^{ν} are finite-valued random l.sc. functions on $(R^n \times (R_+^s \times R^{m-s})) \times Z$; on the complement all functions L^{ν} are $-\infty$. This is all we need to guarantee the required measurability properties, in particular we have that

$$((x, y), \zeta) \mapsto L^{\nu}(x, y, \zeta) \quad \text{is } \mathcal{B}^{n+m} \otimes \mathcal{A} - \text{measurable} \ .$$

DEFINTION 5.1 *The sequence of functions* $\{h^{\nu}: R^n \times R^m \rightarrow [-\infty, \infty], \ \nu = 1, \ldots\}$ *epi/hypo-converges to* $h: R^n \times R^m \rightarrow [-\infty, \infty]$ *if for all* (x, y) *we have*

(i) *for every subsequence* $\{h^{\nu_k}, k = 1, \ldots\}$ *and sequence* $\{x^k\}_{k=1}^{\infty}$ *converging to* x, *there exists a sequence* $\{y^k\}_{k=1}^{\infty}$ *converging to* y *such that*

$$h(x, y) \leq \liminf_{k \rightarrow \infty} h^{\nu_k}(x^k, y^k) \ ,$$

and

(ii) *for every subsequence* $\{h^{\nu_k}, k = 1, \ldots\}$ *and sequence* $\{\hat{y}^k\}_{k=1}^{\infty}$ *converging to* y, *there exists a sequence* $\{\hat{x}^k\}_{k=1}^{\infty}$ *converging to* x *such that*

$$h(x, y) \geq \limsup_{k \to \infty} h^{\nu_k}(\hat{x}^k, \hat{y}^k) \ .$$

This type of convergence of bivariate functions was introduced by Attouch and Wets (1983a) in order to study the convergence of saddle points; in Attouch and Wets (1983b) it is argued that it actually is the weakest type of convergence that will guarantee the convergence of saddle points.

THEOREM 5.2 Consistency. *From* Assumptions 3.4 *and* 3.5, *with* f *replaced by* f_0, *it follows that there exists* $Z_0 \in F$ *with* $\mu(Z \setminus Z_0) = 0$ *such that*

$$L = \text{epi/hypo} - \lim_{\nu \to \infty} L^{\nu} \qquad \mu\text{—}a.s.$$

and hence:

(i) *for all* $\zeta \in Z_0$, *any cluster pont* $(\hat{x}, \hat{y})$ *of any sequence* $\{(x^{\nu}, y^{\nu}),\ \nu = 1,..\}$, *with* (x^{ν}, y^{ν}) *a saddle point of* $L^{\nu}(\cdot, \cdot, \zeta)$, *is a saddle point of* L;

(ii) *if* D *is a compact subset of* $R^n \times R^m$ *that meets for all* ν, *or at least for some subsequence, the set of saddle points of* $L^{\nu}(\cdot, \cdot, \zeta)$ *for some* $\zeta \in Z_0$, *then there exist* (x^{ν}, y^{ν}) *saddle points of* $L^{\nu}(\cdot, \cdot, \zeta)$ *for* $\nu = 1, \ldots$ *that have at least one cluster point;*

(iii) *moreover, if the preceding condition is satisfied for all* $\zeta \in Z_0$, *and* L *has a unique saddle point, then there exists a sequence*

$$\{(x^{\nu}, y^{\nu}) : Z_0 \to R^n \times R^m,\ \nu = 1, \ldots\}$$

of F^{ν}*- measurable functions that for all* $\zeta \in Z_0$, *determine saddle point of the* L^{ν}, *and converge to the saddle point of* L.

We note that sufficient condition for the existence of saddle points are provided by the condition introduced in Proposition 3.10 (with f the essential objective function of problem (5.1)), in conjunction with the Mangasarian-Fromovitz constraint qualification.

ASYMPTOTIC NORMALITY 5.3 The techniques of Section 4 can also be used to obtain asymptotic normality results. However, there is not yet a good concept of subdifferentiability for bivariate functions, except in the convex case (Rockafellar (1964)), and in the differentiable case, of course. With $\partial L (\partial L^{\nu}$ resp.) the set of subgradients of the Lagrangians in the convex or differentiable case, the condition that (x^*, y^*) is a saddle point of L can be expressed as

$$0 \in \partial L(x^*, y^*) \ ,$$

and $0 \in \partial L^\nu(x^\nu, y^\nu, \zeta)$ in the case of L^ν. For example, in the convex case when all the functions $\{f_i, i = 0, 1, \ldots, m\}$ are differentiable and $X = R^n$, this condition is equivalent to:

$$0 = E\{\nabla f_0(x^*, \xi)\} + \textstyle\sum_{i=1}^m y_i^* f_i^*(x^*) \ ,$$

$$0 \geq f_i(x^*) \ , \qquad i = 1, \ldots, s \ ,$$

$$0 = f_i(x^*) \ , \qquad i = s+1, \ldots, m \ ,$$

$$0 = y_i^* f_i(x^*) \ , \quad y_i^* \geq 0 \ , \qquad i = 1, \ldots, s \ ,$$

and similarly for L^ν.

It is easy to see that when Assumptions 4.4 and 3.5 hold (with f_0 instead of f), as well as Assumption 4.7, but this time with v^ν and v subgradients of L^ν and L respectively, and $S = X \times (R_+^s \times R^{m-s})$, then by the same argument as in the proof of Theorem 4.8, we obtain:

$$\sqrt{\nu}\,(x^\nu(\cdot) - x^*, y^\nu(\cdot) - y^*) \quad \text{is asymptoticaly normal} \ .$$

For an application to the above results to the case of linearly restricted L_1-regression (2.3) see Dupačová (1987).

APPENDIX

We shall show that the assumption

$$\frac{\|v^\nu(x^*) + v(x^\nu)\|}{\nu^{-1/2} + \|v(x^\nu)\|} \text{converges in probability to } 0 \ ,$$

of Lemma 4.9 follows from a series of sufficient conditions similar to those of Huber (1967) by a slight modification of the paving technique of the same paper. The main difference is due to the fact that the probability measures $P^\nu(\cdot, \zeta)$ are not necessarily the empirical ones so that the expectation $E_\mu E^\nu f(x, \zeta) = \int_Z E^\nu f(x, \zeta)\, \mu(d\zeta)$ need not be equal to $Ef(x)$, etc. and that subgradients are used instead of gradients.

ASSUMPTION A.1 *There is* $d_0 > 0$, $a > 0$ *such that for all* $x \in N(x^*) = \{x : \|x - x^*\| < d_0\}$ *and for an arbitrary* $v(x) \in \partial Ef(x)$

$$\|v(x) - v(x^*)\| \geq a\|x - x^*\| .$$

ASSUMPTION A.2 *For any measurable selection* $u_0(x, \cdot)$ *such that* $u_0(x, \xi) \in \partial f_0(x, \xi)$ P-*a.s. denote*

$$\bar{u}(x, \xi, d) = \sup_{\|x - y\| \leq d} \|u_0(x, \xi) - u_0(y, \xi)\|$$

and assume

(i) *for all* $0 < d \leq d_0$, $x \in N(x^*)$ *there is* $M_1 > 0$ *such that both*

$$E\{\bar{u}(x, \xi, d)\} \leq M_1 d$$

and

$$E_\mu E^\nu\{\bar{u}(x, \xi, d)\} \leq M_1 d$$

(ii) *for all* $0 < d \leq d_0$, $x \in N(x^*)$ *there is* $\bar{M}_2 > 0$ *and* $\bar{\alpha} \in (1/2, 1]$ *such that*

$$\mathrm{var}_\mu E^\nu\{\bar{u}(x, \xi, d)\} \leq \bar{M}_2 d \nu^{-\bar{\alpha}} .$$

ASSUMPTION A.3 *For all* $x \in N(x^*)$, *for any measurable selection* $v_0^\nu(x) \in \partial E^\nu f_0(x)$ *with* $v_0^\nu(x^*) \in \partial E^\nu f_0(x^*)$ μ-*a.s. and for any* $v_0(x) \in \partial Ef_0(x)$ *with* $v_0(x^*) \in \partial E^\nu f_0(x^*)$ *there is* $M_2 > 0$ *and* $\alpha \in (1/2, 1]$ *such that*

$$E_\mu\|v_0^\nu(x) - v_0^\nu(x^*) - v_0(x) + v_0(x^*)\|^2 \leq M_2\|x - x^*\|\nu^{-\alpha}$$

LEMMA A.4 *Under Assumptions* A.1, A.2, A.3

$$\sup_{x \in N(x^*)} \frac{\|v^\nu(x^*) - v^\nu(x) - v(x^*) + v(x)\|}{\nu^{-1/2} + \|v(x^*) - v(x)\|} \longrightarrow 0$$

in μ-*probability as* $\nu \longrightarrow \infty$.

PROOF Put $Z^\nu(x, x') = \dfrac{\|v^\nu(x') - v^\nu(x) - v(x') + v(x)\|}{\nu^{-1/2}\|v(x') - v(x)\|}$

$$U^\nu(x, d) = E^\nu\{\bar{u}(x, \xi, d)\} + E\{\bar{u}(x, \xi, d)\}$$

$$W^\nu(x) = v_o^\nu(x) - v_0^\nu(x^*) - v_0(x) + v_o(x^*) \ .$$

Using (4.2) and (4.3), we can write

$$Z^\nu(x, x') = \frac{\|v_o^\nu(x') - v_0^\nu(x) - v_0(x') + v_0(x)\|}{\nu^{-1/2} + \|v(x') - v(x)\|}$$

and

$$\mu\{\zeta : \sup_{x \in N(x^*)} Z^\nu(x, x^*) \geq \varepsilon\} \leq \mu\left\{\zeta : \sup_{x \in N(x^*)} \|W^\nu(x)\| \geq \frac{\varepsilon}{\sqrt{\nu}}\right\}$$

$$\leq \frac{M_2 d_o}{\nu^\alpha} \cdot \frac{\nu}{\varepsilon^2} = \frac{M_2 d_o}{\varepsilon^2} \nu^{1-\alpha}$$

according to Chebychev inequality and Assumption A.3. This estimate, however, does not yield the assertion of the Lemma.

As in Huber (1967) we cover $N(x^*)$ by shrinking neighborhoods whose size decreases and whose number does not increase to rapidly as $\nu \longrightarrow \infty$.

Let γ be such that $\frac{1}{2} < \gamma < \min(\alpha, \bar{\alpha})$. Put $N_{d_o} = N(x^*)$ and denote by

$$N_{d_0 \nu^{-\gamma}} = \{x : \|x - x^*\| \leq d_0 \nu^{-\gamma}\} \ .$$

By the same argument as above

$$\mu\left\{\zeta : \sup_{x \in N_{d_0 \nu^{-\gamma}}} Z^\nu(x, x^*) \geq \varepsilon\right\} \leq \frac{M_2 d_0}{\varepsilon^2} \nu^{1-\alpha-\gamma} = C_0(\varepsilon) \nu^{1-\alpha-\gamma} \ . \tag{A.1}$$

The area $N_{d_o} \setminus N_{d_0 \nu^{-\gamma}}$ will be covered by finitely many nonoverlapping "borders" of the form

$$N_{(k)} = \{x : d_0 \nu^{-(k+1)\delta} < \|x - x^*\| \leq d_0 \nu^{-k\delta}\} \ , \quad k = 0, \ldots, K_\nu - 1$$

where

$$K_\nu \delta \geq \gamma > (K_\nu - 1)\delta \tag{A.2}$$

and for each ν, δ is fixed in such a way that

$$1 - \nu^{-\delta} = \frac{1}{M_0} \tag{A.3}$$

with $M_0 \geq 2$ an integer to be defined later. As a result,

$$\delta = \frac{\log M_0 - \log (M_0 - 1)}{\log \nu} \quad .$$

To simplify the notation we shall put

$$d_k = d_0 \nu^{-k\delta} \quad , \quad k = 1, \ldots, K_\nu \ .$$

As the next step, we shall cover each of "borders" $N_{(k)}$ by nonoverlapping neighborhoods of an equal volume with centers x′ such that

$$\|x^* - x'\| = \frac{1}{2}(d_k + d_{k+1}) = \frac{1}{2} d_0 \nu^{-k\delta}[1 + \nu^{-\delta}] = d_{(k)}$$

and diameters

$$2d_{(k)} = d_k - d_{k+1} = d_0 \nu^{-k\delta}[1 - \nu^{-\delta}] \ .$$

Their number will not exceed

$$\left(\frac{2d_k}{2d_{(k)}}\right)^n = \left(\frac{2}{1 - \nu^{-\delta}}\right)^n = \left(2M_0\right)^n = M \ .$$

Using (A.3), we have

$$\frac{1}{2} \le \nu^{-\delta} \quad \text{and} \quad 3\nu^{-\delta} \ge 1 + \nu^{-\delta} \ge \nu^{-\delta} \ . \tag{A.4}$$

Let N be any of the neighborhoods of the covering $N_{(k)}$, i.e., $N = \{x : \|x - x'\| \le d_{(k)}\}$. We have according to Assumption A.1

$$\sup_{x \in N} Z(x, x^*) \le \sup_{x \in N} \frac{\|v_0^\nu(x) - v_0^\nu(x^*) - v_0(x) + v_0(x^*)\|}{\nu^{-1/2} + a d_0 \nu^{-(k+1)\delta}}$$

$$\le \frac{\|W^\nu(x')\|}{a d_0 \nu^{-(k+1)\delta}} + \frac{U_\nu(x', d_{(k)})}{a d_0 \nu^{-(k+1)\delta}} \ .$$

Using Assumption A.3, Chebychev inequality and (A.4)

$$\mu\{\zeta : \|W^\nu(x')\| \ge a d_0 \nu^{-(k+1)\delta} \varepsilon\} \le \frac{M_2 d_{(k)}}{\varepsilon^2 \nu^\alpha a^2 d_0^2 \nu^{-2(k+1)\delta}} \tag{A.5}$$

$$\le C_1(\varepsilon) \nu^{(k+1)\delta - \alpha} \ .$$

Similarly, according to Assumption A.2 (ii) and Chebychev inequality

$$\mu\{\zeta : U^{\nu}(x', d_{(k)}) \geq \varepsilon a d_0 \nu^{-(k+1)\delta}\}$$

$$\leq \mu\{\zeta : |E^{\nu}\{\bar{u}(x', \xi, d)\} - E_{\mu}E^{\nu}\{\bar{u}(x', \xi, d)\}| \geq \varepsilon a d_0 \nu^{-(k+1)\delta}$$

$$- E\{\bar{u}(x', \xi, d)\} - E_{\mu}E^{\nu}\{\bar{u}(x', \xi, d)\}\}$$

$$\leq \frac{\operatorname{var}_{\mu} E^{\nu}\{\bar{u}(x', \xi, d)\}}{\eta^2} \leq \frac{\bar{M}_2 d_{(k)}}{\nu^{\bar{\alpha}}\eta^2}$$

where

$$\eta = \varepsilon a d_0 \nu^{-(k+1)\delta} - E\{\bar{u}(x', \xi, d)\} - E_{\mu}E^{\nu}\{\bar{u}(x', \xi, d)\} \tag{A.6}$$

$$\geq \varepsilon a d_0 \nu^{-(k+1)\delta} - 2M_1 d_{(k)} = d_0 \nu^{-(k+1)\delta}[\varepsilon a - M_1(1 - \nu^{-\delta}) \cdot \nu^{-\delta}]$$

$$\geq d_0 \nu^{-(k+1)\delta}\left[\varepsilon a - \frac{M_1}{2M_0}\right]$$

according to Assumption A.2 (i), (A.3) and (A.4). For $M_0 > \dfrac{M_1}{2\varepsilon a}$ the lower bound in (A.6) is nontrivial and we have that

$$\mu\{\zeta : U^{\nu}(x', d_{(k)}) \geq \varepsilon a d_0 \nu^{-(k+1)\delta}\} \leq \frac{\bar{M}_2 d_0 \nu^{-k\delta}[1 - \nu^{-\delta}] \cdot \frac{1}{2}}{d_0^2 \nu^{-2(k+1)\delta}\left[\varepsilon a - \frac{M_1}{2M_0}\right]^2 \nu^{\bar{\alpha}}} \tag{A.7}$$

$$\leq \frac{\bar{M}_2}{d_0} \frac{\nu^{-\bar{\alpha}+(k+1)\delta}}{M_0\left[\varepsilon a - \frac{2M_1}{M_0}\right]^2} = C_2(\varepsilon) \cdot \nu^{-\bar{\alpha}+(k+1)\delta} .$$

Finally, according to (A.1), (A.5) and (A.7)

$$\mu\{\zeta : \sup_{x \in N(x^*)} Z^{\nu}(x, x^*) \geq 2\varepsilon\} \leq \mu\{\zeta : \sup_{x \in N_{d_{K_\nu}}} Z^{\nu}(x, x^*) \geq \varepsilon\}$$

$$+ \sum_{k=0}^{K_\nu - 1} \mu\{\zeta : \sup_{x \in N \subset N_{(k)}} Z^{\nu}(x, x^*) \geq \varepsilon\} \cdot M$$

$$\leq C_0(\varepsilon)\nu^{1-\alpha-K_\nu\delta} + M \sum_{k=0}^{K_\nu - 1} \mu\{\zeta : \|W^{\nu}(x')\| \geq \varepsilon a d_0 \nu^{-(k+1)\delta}\}$$

$$+ M \sum_{k=0}^{K_\nu - 1} \mu\{\zeta : U^{\nu}(x', d_{(k)}) \geq \varepsilon a d_0 \nu^{-(k+1)\delta}\}$$

$$\leq C_0(\varepsilon)\nu^{1-\alpha-K_\nu\delta} + M \sum_{k=0}^{K_\nu - 1} C_1(\varepsilon)\nu^{(k+1)\delta-\alpha} + M \sum_{k=0}^{K_\nu - 1} C_2(\varepsilon)\nu^{(k+1)\delta-\bar{\alpha}}$$

$$= C_0(\varepsilon)\nu^{1-\alpha-K_\nu\delta} + M(C_1(\varepsilon)\nu^{-\alpha} + C_2(\varepsilon)\nu^{-\bar{\alpha}}) \sum_{k=0}^{K_\nu-1} \nu^{(k+1)\delta}$$

$$= C_0(\varepsilon)\nu^{1-\alpha-K_\nu\delta} + MC_1(\varepsilon)\nu^{K_\nu\delta-\alpha} + MC_2(\varepsilon)\nu^{K_\nu\delta-\bar{\alpha}} .$$

In addition, for ν large enough, $1 - K_\nu\delta - \alpha < 0$ and $K_\nu\delta - \alpha < 0$, $K_\nu\delta - \bar{\alpha} < 0$ due to our choice of γ and (A.2).

Summarizing: for an arbitrary $\varepsilon > 0$, $1/2 < \gamma < \min(\alpha, \bar{\alpha})$ it is possible to bound the probability

$$\mu\{\zeta: \sup_{x \in N(x^*)} Z_\nu(x, x^*) \geq \varepsilon\}$$

from above by an expression which converges to zero as $\nu \longrightarrow \infty$. □

REFERENCES

Attouch, H. and R. Wets (1983a): A convergence theory for saddle functions, *Transactions Amer. Math. Soc.*, 280, 1–41.

Attouch, H. and R. Wets (1983b): A convergence for bivariate functions aimed at the convergence of saddle values, in *Mathematical Theories of Optimization*, eds. J. Cecconi and T. Zolezzi, Springer Verlag Lecture Notes in Mathematics 979, 1–42.

Aubin, J.-P. and I. Ekeland (1984): *Applied Nonlinear Analysis*, Wiley Interscience, New York.

Aumann, R.J. (1965): Integrals of set-valued functions, *J. Mathematical Analysis and Applications* 12, 1–12.

Clarke, R. (1975): Generalized gradients and applications, *Transaction American Mathematical Society*, 205, 247–262.

Clarke, R. (1983): *Optimization and Nonsmooth Analysis*, Wiley Interscience, New York.

Dolecki, S., G. Salinetti and R. Wets (1983): Convergence of functions: equisemicontinuity. *Trans. Amer. Math. Soc.* 276, 409–429.

Dupačová, J. (1983a): Stability in stochastic programming with recourse. *Acta Univ. Carol.-Math et Phys.* **24**, 23–34.

Dupačová, J. (1983b): The problem of stability in stochastic programming (in Czech). Dissertation for the Doctor of Sciences degree, Faculty of Mathematics and Phyics, Charles University, Prague.

Dupačová, J. (1984): On asymptotic normality of inequality constrained optimal decisions. In: *Proc. 3-rd Prague conference on asymptotic statistics*, eds. P. Mandl and M. Hušková. Elsevier, Amsterdam, p. 249–257.

Dupačová, J. (1987): Asymptotic properties of restricted L_1-estimates. To appear in Proc. of the 1st International Conference on Statistical Data Analysis based on the L_1 norm and Related Methods, Neuchatel, 31 August – 4 September 1987.

Dupačová, J. and R. Wets (1986): Asymptotic behavior of statistical estimators and of optimal solutions for stochastic optimization problems, IIASA WP-86-41, Laxenburg, Austria.

Huber, P. (1967): The behavior of maximum likelihood estimates under nonstandard conditions, *Proc. Fifth Berkeley Symp. Math. Stat. Prob.* 1, University of California Press, Berkeley, 221–233.

King, A. and R.T. Rockafellar (1986): Non-normal asymptotic behavior of solution estimates in linear-quadratic stochastic optimization, Manuscript, Univ. Washington, Seattle.

Lehmann, E.L. (1983): *Theory of point estimation*. Wiley, Interscience, New York.

Rockafellar, R.T. (1964): Minimax theorems and conjugate saddle functions, *Matematica Scandinavica*, 14, 151–173.

Rockafellar, R.T. (1970): *Convex Analysis*, Princeton University Press, Princeton.

Rockafellar, R.T. (1976): Integral functionals, normal integrands and measurable multifunctions, in *Nonlinear Operators and the Calculus of Variations*, eds. J. Gossez and L. Waelbroeck, Springer Verlag Lecture Notes in Mathematics 543, Berlin, 157–207.

Rockafellar, R.T. (1979): Directionally Lipschitzian functions and subdifferential calculus, *Proceedings London Mathematical Society* 39, 331–355.

Rockafellar, R.T. (1980): Generalized directional derivatives and subgradients of nonconvex functions, *Canadian J. Mathematics* 32, 257–280.

Rockafellar, R.T. (1981): *The Theory of Subgradients and its Applications: Convex and Nonconvex Functions*, Halderman Verlag, Berlin.

Rockafellar, R.T. (1983): Generalized subgradients in mathematical programming, in *Mathematical Programming: The State-of-the-Art*, Springer Verlag, Berlin, 368–390.

Rockafellar, R.T. and R. Wets (1982), On the interchange of subdifferentiation and conditional expectation for convex functionals, *Stochastics* 7, 173–182.

Solis, R. and R. Wets (1981): A statistical view of stochastic programming. Tech. report, Univ. Kentucky.