Annals of Operations Research 30 (1991) 169–186

ASYMPTOTIC ANALYSIS OF STOCHASTIC PROGRAMS

Alexander SHAPIRO

Department of Mathematics and Applied Mathematics, University of South Africa, Pretoria, 0001 South Africa

In this paper we discuss a general approach to studying asymptotic properties of statistical estimators in stochastic programming. The approach is based on an extended delta method and appears to be particularly suitable for deriving asymptotics of the optimal value of stochastic programs. Asymptotic analysis of the optimal value will be presented in detail. Asymptotic properties of the corresponding optimal solutions are briefly discussed.

Keywords: Stochastic programming, random sampling, optimal value function, weak convergence, Central Limit Theorem, delta method, Lagrange multipliers, asymptotic distribution, M-estimators.

1. Introduction

Quite often in situations involving uncertainty the considered optimization problem is too complicated to be solved in an exact probabilistic sense and can be approximated at best. In this paper we study stochastic properties of the obtained approximate solutions which are considered as statistical estimators. Suppose that we are interested in solving the mathematical programming problem:

$$(\mathbf{P}_0) \qquad \begin{array}{lll} \text{minimize} & f_0(x), & x \in S, \\ \text{subject to} & f_i(x) = 0, & i = 1, \dots, q, \\ & f_i(x) \leqslant 0, & i = q+1, \dots, r, \end{array}$$

where S is a subset of $\mathbb{R}^k$ and $f_i(x)$ are generally nonlinear, real valued functions. Consider a situation where, for some reason, program $(\mathbf{P}_0)$ is inaccessible and has to be approximated. That is, instead of solving program $(\mathbf{P}_0)$, referred to as the *true* program, we solve a sequence of approximating programs

$$(\hat{\mathbf{P}}_n) \qquad \begin{array}{lll} \text{minimize} & \psi_{0n}(x), & x \in S, \\ \text{subject to} & \psi_{in}(x) = 0, & i = 1, \dots, q, \\ & \psi_{in}(x) \leqslant 0, & i = q+1, \dots, r. \end{array}$$

Here, for every $i = 0, \ldots, r$, $\{\psi_{in}(x)\}$ is a sequence of real valued functions converging in some sense to the corresponding function $f_i(x)$. The calculated optimal value $\hat{\varphi}_n$ and an optimal solution $\hat{x}_n$ of the program $(\hat{\mathbf{P}}_n)$ give approximations to the optimal value φ_0 and the optimal solution x_0 of the true program $(\mathbf{P}_0)$, respectively. We then investigate statistical properties of $\hat{\varphi}_n$ and $\hat{x}_n$ in the cases where the approximating functions ψ_{in}, in some sense, are *stochastic*.

Let us consider the following basic example. Let $(\Omega, \mathscr{F}, P)$ be a probability space and $g_i(x, \omega)$, $i = 0, \ldots, r$, be real valued functions $g_i : \mathbb{R}^k \times \Omega \to \mathbb{R}$. Many problems of stochastic programming can be formulated in the form of program $(\mathbf{P}_0)$ with the corresponding functions f_i given as *expected values*

$$f_i(x) = E_p\{g_i(x, \omega)\} = \int_\Omega g_i(x, \omega) P(\mathrm{d}\omega), \tag{1.1}$$

$i = 0, \ldots, r$ (see, e.g., [9, 11, chapter 1]). In applications the space Ω is usually a finite dimensional vector space and $\mathscr{F}$ is its Borel σ-algebra. However, this will be of a little importance in our considerations and we resort to the general scheme of probability spaces.

Now let $\omega_1, \omega_2, \ldots,$ be a sequence of independent identically distributed (iid) random elements in $(\Omega, \mathscr{F})$ with the common probability distribution P. Then one can estimate $f_i(x)$ by the sample mean function

$$\psi_{in}(x) = n^{-1} \sum_{j=1}^{n} g_i(x, \omega_j). \tag{1.2}$$

Particular cases of the associated problem $(\hat{\mathbf{P}}_n)$ have been studied in statistics over generations. The maximum likelihood and nonlinear regression methods can be formulated in this framework and the M-estimators of Huber [19,21], are just the (unconstrained) minimizers of $\psi_{0n}(x)$. Recently this scheme has also attracted a renewed interest in stochastic programming (e.g., [7,10,23,24]). It also appears naturally in sensitivity analysis and optimization of simulation models and stochastic networks by the Score Function method (see [2,39] and references therein).

The sample mean functions ψ_{in} can be represented as expected values

$$\psi_{in}(x) = E_{P_n}\{g_i(x, \omega)\} \tag{1.3}$$

with respect to the empirical measure

$$P_n = n^{-1} \sum_{j=1}^{n} \delta(\omega_j),$$

where $\delta(\omega)$ denotes the probability measure of mass one at the point ω. By constructing a sequence $\{P_n\}$ of probability measures, not necessarily empirical, it is possible to generate the corresponding functions φ_{in} in the form of the expected values as in (1.3) other than the sample mean functions (cf. [10]).

Another possible extension of the basic scheme (1.1) and (1.2) is to take a composite function of the expected value functions (1.1). For example, one may want to minimize the variance

$$f(x) = \mathrm{Var}\{g(x, \omega)\} = E_P\{g(x, \omega)^2\} - [E_P\{g(x, \omega)\}]^2$$

by minimizing the corresponding sample variance

$$\psi_n(x) = (n-1)^{-1} \sum_{j=1}^{n} \left[g(x, \omega_j) - n^{-1} \sum_{j=1}^{n} g(x, \omega_j) \right]^2$$

subject to some constraints.

Investigation of statistical properties of the estimators $\hat{\varphi}_n$ and $\hat{x}_n$ is the main subject of this paper. In the following section we describe a general approach to deriving asymptotic results based on an extended delta method. This approach seems to be particularly suitable for studying asymptotics of the optimal value $\hat{\varphi}_n$, which will be discussed in detail in section 3. Section 4 is devoted to a discussion of the optimal solutions $\hat{x}_n$ and has mostly a survey character.

By $x \cdot y$ we denote the scalar product of two vectors x, $y \in \mathbb{R}^k$. The gradients $\nabla g(x, \omega)$, $\nabla L(x, \lambda, \mu)$ and the Hessian matrices $\nabla^2 L(x, \lambda, \mu)$, etc., of the considered functions are always taken with respect to the first variable x.

2. General framework

In this section we describe a general approach to asymptotic analysis of the approximating program $(\hat{\mathbf{P}}_n)$. The approach is based on an extension of the (finite dimensional) delta method, which is routinely employed in multivariate statistical analysis (e.g. [32, p. 388]). We work here directly with functions of the programs $(\mathbf{P}_0)$ and $(\hat{\mathbf{P}}_n)$ which are viewed as elements of an appropriate functional space, rather than with the corresponding (empirical) measures and von Mises statistical functionals. This allows to avoid considerable technical difficulties related to the theory of empirical processes. The idea of asymptotic analysis via functional spaces is due to King [23].

Consider the mathematical programming problem

$$(\mathbf{P}_\xi) \qquad \begin{array}{lll} \text{minimize} & \xi_0(x), & x \in S, \\ \text{subject to} & \xi_i(x) = 0, & i = 1, \ldots, q, \\ & \xi_i(x) \leqslant 0, & i = q+1, \ldots, r, \end{array}$$

with the functions ξ_i, $i = 0, \ldots, r$, viewed as *parameters* in a chosen linear (vector) space $\mathscr{X}_i$ of real valued functions defined on the set $S \subset \mathbb{R}^k$. For example one can take $\mathscr{X}_i = C(S)$, $i = 0, \ldots, r$, where $C(S)$ denotes the linear space of continuous bounded functions $\zeta : S \to \mathbb{R}$ endowed with the sup-norm

$$\| \zeta \| = \sup_{x \in S} | \zeta(x) |.$$

Then the optimal value $\bar{\varphi}(\xi)$ and an optimal solution $\bar{x}(\xi)$ of the program $(\mathbf{P}_\xi)$ are considered as functions of the vector $\xi = (\xi_0, \ldots, \xi_r)$ from the Cartesian product space $\mathscr{X} = \mathscr{X}_0 \times \cdots \times \mathscr{X}_r$. (The space $\mathscr{X}$ can be equipped, for example, with the max-norm which is the maximum of the individual norms of the components $\xi_i \in \mathscr{X}_i$, $i = 0, \ldots, r$, of the vector ξ.)

Now suppose that $\mu = (f_0, \ldots, f_r) \in \mathscr{X}$ and that $Y_n = (\psi_{0n}, \ldots, \psi_{rn})$, $n = 1, 2, \ldots$, can be considered as a random element in the normed space $\mathscr{X}$ which is assumed to be equipped with the associated Borel σ-algebra. (For definitions and a basic theory of such notions as random elements and weak convergence in metric spaces, the reader is referred to an exposition book of Billingsley [4].) Then, of course, $\varphi_0 = \bar{\varphi}(\mu)$ and $\hat{\varphi}_n = \bar{\varphi}(Y_n)$ and similarly for the optimal solution. Furthermore, suppose that $n^{1/2}(Y_n - \mu)$ converges in distribution (weakly) to a random element Z. If we can approximate the function $\bar{\varphi}(\xi)$ near the point $\xi = \mu$, say by a linear or more generally positively homogeneous function, then we may hope to derive the corresponding asymptotics of the estimator $\hat{\varphi}_n$ as well. This idea is accomplished by the delta-method a version of which we describe now.

Consider the normed space $\mathscr{X}$ and a mapping $g : \mathscr{X} \to \mathscr{Y}$ from $\mathscr{X}$ to a normed space $\mathscr{Y}$. It is said that g is (Gâteaux) *directionally differentiable* at $\mu \in \mathscr{X}$ if the limit

$$g'_\mu(\zeta) = \lim_{t \to 0^+} \frac{g(\mu + t\zeta) - g(\mu)}{t} \tag{2.1}$$

exists for all $\zeta \in \mathscr{X}$. The directional derivative $g'_\mu(\cdot)$ gives a local approximation of the mapping g at μ in the sense that

$$g(\mu + \zeta) - g(\mu) = g'_\mu(\zeta) + r(\zeta), \tag{2.2}$$

where the remainder $r(\zeta)$ is "small" along any fixed direction ζ. That is, $t^{-1}r(t\zeta) \to 0$ at $t \to 0^+$. Notice that the directional derivative, if it exists, is positively homogeneous, i.e., $g'_\mu(t\zeta) = tg'_\mu(\zeta)$ for all ζ and $t > 0$. If g is directionally differentiable at μ and in addition $g'_\mu(\zeta)$ is linear and continuous in ζ, then g is said to be Gâteaux differentiable at μ.

From (2.2) we have that

$$n^{1/2}[g(Y_n) - g(\mu)] = g'_\mu[n^{1/2}(Y_n - \mu)] + n^{1/2}r(Y_n - \mu). \tag{2.3}$$

Therefore if the remainder term in the right hand side of (2.3) is asymptotically negligible and $g'_\mu(\cdot)$ is continuous, then we may conclude that the left hand side term in (2.3) converges in distribution to $g'_\mu(Z)$. In order to carry this idea through we need a stronger notion of directional differentiability.

It is said that the mapping g is *Hadamard directionally differentiable* at μ if for any sequence $\{\zeta_n\}$ converging to a vector ζ and every sequence $\{t_n\}$ of positive numbers converging to zero, the limit

$$g'_\mu(\zeta) = \lim_{n \to \infty} \frac{g(\mu + t_n\zeta_n) - g(\mu)}{t_n} \tag{2.4}$$

exists. Of course, Hadamard directional differentiability implies Gâteaux directional differentiability and both directional derivatives coincide. Moreover, it follows from Hadamard directional differentiability that the directional derivative $g'_\mu(\zeta)$ is continuous, although possibly nonlinear, in ζ (see, e.g., [3,30,44] for a discussion of various forms of differentiability in normed spaces).

In some situations we shall need a notion of Hadamard directional differentiability *tangentially to a set*. That is, let $\mathscr{K}$ be a subset of $\mathscr{X}$. We say that g is Hadamard directionally differentiable at μ *tangentially* to $\mathscr{K}$ if the limit (2.4) exists for all sequences $\{\zeta_n\}$ converging to ζ, of the form $\zeta_n = t_n^{-1}(\xi_n - \mu)$ where $\xi_n \in \mathscr{K}$ and $t_n \to 0^+$. The obtained directional derivative $g'_{\mu,\mathscr{K}}(\cdot)$ is then defined on the *contingent* (Bouligand) cone $T_\mu(\mathscr{K})$ to $\mathscr{K}$ at μ. The cone $T_\mu(\mathscr{K})$ is formed by the limits of sequences $\zeta_n = t_n^{-1}(\xi_n - \mu)$ with $\xi_n \in \mathscr{K}$ and $t_n \to 0^+$. The contingent cone is always closed and it is not difficult to show that the directional derivative $g'_{\mu,\mathscr{K}}(\cdot)$ is continuous on $T_\mu(\mathscr{K})$.

Gâteaux derivatives, with respect to the probability measures, are employed in statistics as a heuristic device suggesting correct formulas but unfortunately are useless for proving purposes (see [21]). On the other hand, the stronger notion of Fréchet differentiability often is not applicable in many interesting situations. It was suggested by Reeds [33] that the concept of Hadamard differentiability is exactly attuned to statistical applications. Reeds' approach was described and further developed in Fernholz [12]. King [23] put forward an idea of asymptotic analysis via random functions rather than the underlying probability measures. In the following version of the delta method, given in theorem 2.1, we follow Grübel [18], Gill [16] and King [25]. We assume that the normed spaces $\mathscr{X}$ and $\mathscr{Y}$ are endowed with their Borel σ-algebras and $\{Y_n\}$ is a sequence of random elements of $\mathscr{X}$.

THEOREM 2.1

(*δ-method*)

Let $\mathscr{K}$ be a subset of $\mathscr{X}$ and suppose that:

(i) The mapping g is Hadamard directionally differentiable at μ tangentially to $\mathscr{K}$.

(ii) The mapping g is measurable.

(iii) There is a sequence $\tau_n \to \infty$ of positive numbers such that $\tau_n(Y_n - \mu)$ converges in distribution to a random element Z, where the distribution of Z is concentrated on a separable subset of $\mathscr{X}$.

(iv) For all n large enough,

$$\text{Prob}\{Y_n \in \mathscr{K}\} = 1.$$

Then

$$\tau_n[g(Y_n) - g(\mu)] \xrightarrow{D} g'_{\mu,\mathscr{K}}(Z). \tag{2.5}$$

Moreover, if the set $\mathscr{K}$ is convex, then

$$\tau_n[g(Y_n) - g(\mu)] = g'_{\mu,\mathscr{K}}[\tau_n(Y_n - \mu)] + o_p(1). \tag{2.6}$$

Here the superscript D stands for convergence in distribution and $o_p(1)$ denotes a random element of $\mathscr{Y}$ converging in probability to zero.

The result of theorem 2.1 follows easily from the Skorohod–Dudley almost sure representation theorem and the definition of Hadamard directional differentiability [16]. An alternative proof can be derived from an extended continuous mapping theorem [25,46]. For the sake of completeness we give a proof of theorem 2.1 and some additional technical details in the appendix. Notice that it follows from the conditions (iii) and (iv) of theorem 2.1 that the distribution of Z is concentrated on $T_\mu(\mathscr{K})$ and hence the distribution of $g'_{\mu,\mathscr{K}}(Z)$ is well defined. Also it follows from the condition (i) that $g'_{\mu,\mathscr{K}}(\cdot)$ is continuous on $T_\mu(\mathscr{K})$ and hence is measurable.

In applications the asymptotic behavior of Y_n is usually verified by the Central Limit Theorem with $\tau_n = n^{1/2}$. Calculation of Gâteaux directional derivatives does not involve any topological or probabilistic assumptions about the space $\mathscr{X}$ and the random elements Y_n. Often it is carried out by an application of the corresponding results from sensitivity analysis of nonlinear programs [8,9,43]. Then Hadamard directional differentiability can be verified by some additional effort. In this respect we mention that for locally Lipschitz mappings in normed spaces the concepts of Hadamard and Gâteaux directional derivatives are equivalent. This result was noticed by many authors (see, e.g., [44] for relevant references).

3. Asymptotic properties of the optimal value

In this section we investigate asymptotic behavior of the optimal value $\hat{\varphi}_n$ of the program $(\hat{\mathbf{P}}_n)$. Let us start by considering the situation where all constraints of the programs $(\mathbf{P}_0)$ and $(\hat{\mathbf{P}}_n)$ are included in the fixed set S. That is, define the optimal value function as

$$\bar{\varphi}(\xi) = \inf\{\xi(x) : x \in S\}. \tag{3.1}$$

We assume that the set S is *compact* and that ξ belongs to the Banach space $C(S)$ of real valued continuous functions defined on S and equipped with the sup-norm. The function $\bar{\varphi}(\cdot)$ is given by the pointwise minimum of a family of linear functionals on the space $C(S)$ and hence is a concave function. Its Gâteaux directional derivatives can be calculated by methods of convex analysis (e.g. [22, section 4.5.2]). Hadamard directional differentiability then follows, for example by noting that $\bar{\varphi}$ is Lipschitz continuous on $C(S)$. The following result is also related to a theorem of Danskin [6].

THEOREM 3.1

Suppose that the set S is compact and that $\bar{\varphi} : C(S) \to \mathbb{R}$ is the optimal value function defined in (3.1). Then for any $\mu \in C(S)$, $\bar{\varphi}$ is Hadamard directionally differentiable at μ and

$$\bar{\varphi}'_\mu(\zeta) = \min_{x \in S^*(\mu)} \zeta(x), \tag{3.2}$$

where

$$S^*(\mu) = \operatorname{argmin}_{x \in S} \mu(x).$$

In particular, it follows from (3.2) that if the function $\mu(x)$ has a unique minimizer x_0 over the set S, i.e. $S^*(\mu) = \{x_0\}$ is a singleton, then $\bar{\varphi}'_\mu(\zeta) = \zeta(x_0)$ is linear in ζ. Consequently in this case $\bar{\varphi}$ is Hadamard differentiable at μ. Notice that Fréchet differentiability does not follow here.

Theorems 2.1 and 3.1 imply the following result.

THEOREM 3.2

Let S be compact and let $\{\psi_n\}$ be a sequence of random elements in $C(S)$, $f \in C(S)$ and $\varphi_0 = \bar{\varphi}(f)$, $\hat{\varphi}_n = \bar{\varphi}(\psi_n)$, where $\bar{\varphi}$ is the optimal value function defined in (3.1). Suppose that $n^{1/2}(\psi_n - f)$ converges in distribution to a random element Z of $C(S)$. Then

$$n^{1/2}(\hat{\varphi}_n - \varphi_0) \xrightarrow{D} \min_{x \in S^*(f)} Z(x).$$

In particular, if $f(x)$ attains its minimum over S at a unique point x_0, then $n^{1/2}(\hat{\varphi}_n - \varphi_0)$ converges in distribution to $Z(x_0)$.

Now consider the situation where the objective function $f(x)$ of the true program is given as the expected value $E\{g(x, \omega)\}$ of a function $g : \mathbb{R}^k \times \Omega \to \mathbb{R}$. (Unless stated otherwise all probabilistic statements will be given with respect to the probability measure P.) Let $\psi_n(x)$ be the corresponding sample mean function based on a sample of size n. Suppose that:

(a) For every $x \in S$, the function $g(x, \cdot)$ is measurable.

(b) For some point $x_0 \in S$, the expectation $E\{g(x_0, \omega)^2\}$ is finite.

(c) There exists a function $b : \Omega \to \mathbb{R}$ such that $E\{b(\omega)^2\}$ is finite and that

$$|g(x, \omega) - g(y, \omega)| \leqslant b(\omega)\, \|x - y\|$$

for all x, $y \in S$.

Notice that assumption (c) implies that for almost every ω, the function $g(\cdot, \omega)$ is Lipschitz continuous on S. Moreover, since S is compact and, hence, bounded, it also follows from assumption (c) that for every $x \in S$ the function $|g(x, \omega)|$ is dominated by the function $h(\omega) = g(x_0, \omega) + \alpha b(\omega)$, where α is a sufficiently large positive number. Because of assumptions (b) and (c), the function $h(\omega)$ is integrable and hence, by the Lebesgue dominated convergence theorem, the expected value function $f(x)$ is continuous on S. That is $f \in C(S)$. It follows from assumptions (a) and (c) that the sample mean functions ψ_n, $n = 1, 2, \ldots$, can be considered random elements of $C(S)$ (cf. [23, proposition A4],

[24]). Finally, assumptions (b) and (c) imply that the sequence $Z_n = n^{1/2}(\psi_n - f)$ of random elements of $C(S)$ converges in distribution to a random element Z. This result is given in [23, proposition A5] and is based on a Central Limit Theorem in $C(S)$ as it is given in [1, chapter 7]. The random element Z of $C(S)$ is also called a *random function* or a *random process*. For a finite subset $\{x_1, \ldots, x_m\}$ of S, the distribution of the corresponding random vector $(Z(x_1), \ldots, Z(x_m))$ is called a finite dimensional distribution of Z. In the present situation all finite dimensional distributions of Z are multivariate normal, that is, the random element Z is Gaussian. In particular, the sequence $\{Z_n(x_0)\}$ of real valued random variables converges in distribution to normal $N(0, \sigma^2)$ with zero mean and the variance

$$\sigma^2 = E\left\{g(x_0, \omega)^2\right\} - \left[E\{g(x_0, \omega)\}\right]^2. \tag{3.3}$$

The following result is now a consequence of theorem 3.2.

THEOREM 3.3

Let f and ψ_n be the expected value and the sample mean functions, respectively. Suppose that the set S is compact, that the assumptions (a)–(c) hold and that $f(x)$ has a unique minimizer x_0 over S. Then $n^{1/2}(\hat{\varphi}_n - \varphi_0)$ converges in distribution to normal with zero mean and the variance σ^2 given in (3.3).

Theorem 3.1 shows that the first order analysis of $\bar{\varphi}(\cdot)$ cannot distinguish between different feasible sets possessing the same optimal solutions of the true program. That is, if $\hat{\varphi}_n$ and $\tilde{\varphi}_n$ are the minima of the function $\psi_n(x)$ over compact sets S and S', respectively, and f has the same set of minimizers over both sets S and S', then $n^{1/2}(\hat{\varphi}_n - \tilde{\varphi}_n)$ converges in probability to zero. Therefore, for the purpose of distinguishing between S and S', a second order analysis is required. Second order expansions of the optimal value function are closely related to the first order behavior of the corresponding optimal solutions and are derived under more restrictive assumptions ([41,42]).

In the situations where constraints of the program $(\mathbf{P_0})$ have to be approximated, investigation of differentiability properties of the optimal value function $\bar{\varphi}(\xi)$ is considerably more complicated. In all its generality the corresponding deterministic problem is still unsolved. We consider two particularly important cases. First, suppose that the programs $(\mathbf{P_0})$ and $(\hat{\mathbf{P}}_n)$ are subject to the inequality constraints only and that both of them are *convex*. That is, suppose that all functions $f_0, f_1, \ldots, f_r$, and $\psi_{0n}, \psi_{1n}, \ldots, \psi_{rn}$, are real valued and convex on $\mathbb{R}^k$ and that the set S is compact and convex. It follows from the convexity assumption that the functions $f_0, \ldots, f_r$ and $\psi_{0n}, \ldots, \psi_{rn}$, are continuous [38] and hence belong to the space $C(S)$.

In the Cartesian product space $\mathscr{X} = C(S) \times \cdots \times C(S)$ of vectors $\xi = (\xi_0, \ldots, \xi_r)$ consider the set $\mathscr{K}$ formed by vectors ξ such that each component ξ_i, $i = 0, 1, \ldots, r$, of ξ is a convex function on an open neighborhood of S. Convexity of $(\mathbf{P}_0)$ and $(\hat{\mathbf{P}}_n)$ means, of course, that $\mu=(f_0,\dots,f_r)$ and $Y_n=(\psi_{0n},\dots,\psi_{rn})$ belong to the set $\mathscr{K}$. Consider the Lagrangian function

$$L(x,\lambda,\xi)=\xi_0(x)+\sum_{i=1}^{r}\lambda_i\xi_i(x)$$

and denote by $S^*(\xi)$ the set of optimal solutions of the program $(\mathbf{P}_\xi)$. Suppose that the *Slater condition* for the program $(\mathbf{P}_0)$ holds, i.e. there exists a point $x\in S$ such that $f_i(x)<0$ for all $i=1,\dots,r$. Then, to every point $\bar{x}\in S^*(\mu)$ corresponds a vector $\bar{\lambda}=(\bar{\lambda}_1,\dots,\bar{\lambda}_r)$ of Lagrange multipliers such that

$$L(\bar{x},\bar{\lambda},\mu)=\min_{x\in S}L(x,\bar{\lambda},\mu), \tag{3.4}$$

$$\bar{\lambda}_i\geqslant 0 \quad\text{and}\quad \bar{\lambda}_i f_i(\bar{x})=0,\quad i=1,\dots,r, \tag{3.5}$$

(e.g. [22, p. 68]). The above optimality conditions can be also formulated in the subdifferential form. In particular, if $L(\cdot,\bar{\lambda},\mu)$ is differentiable at $\bar{x}$ and $\bar{x}$ is an interior point of S, then (3.4) is equivalent to a more familiar condition

$$\nabla L(\bar{x},\bar{\lambda},\mu)=0. \tag{3.6}$$

It can be shown by the arguments of duality that the set $\Lambda(\mu)$ of Lagrange multipliers satisfying optimality conditions (3.4), (3.5), is the same for all $\bar{x}\in S^*(\mu)$. It follows from the Slater condition that the set $\Lambda(\mu)$ is nonempty and bounded.

The following theorem is an extension of a result due to Golshtein [17, chapter 5, section 3]. Its proof is given in the appendix.

THEOREM 3.4

Suppose that the program $(\mathbf{P}_0)$ is convex, S is compact and the Slater condition for the program $(\mathbf{P}_0)$ holds. Then the optimal value function $\bar{\varphi}(\xi)$ is Hadamard directionally differentiable at $\mu=(f_0,\dots,f_r)$ tangentially to the set $\mathscr{K}$ and for $\zeta\in T_\mu(\mathscr{K})$,

$$\bar{\varphi}'_{\mu,\mathscr{K}}(\zeta)=\min_{x\in S^*(\mu)}\max_{\lambda\in\Lambda(\mu)}L(x,\lambda,\zeta). \tag{3.7}$$

The delta method of theorem 2.1 and theorem 3.4 implies the following result.

THEOREM 3.5

Let $Y_n=(\psi_{0n},\dots,\psi_{rn})$ be a sequence of random elements in the space $\mathscr{Z}=C(S)\times\cdots\times C(S)$. Suppose that $\mu=(f_0,\dots,f_r)\in\mathscr{K}$, S is compact, the Slater condition for the program $(\mathbf{P}_0)$ holds, for all n large enough $Y_n\in\mathscr{K}$ with probability one and that $n^{1/2}(Y_n-\mu)$ converges in distribution to a random element Z. Then

$$n^{1/2}(\hat{\varphi}_n-\varphi_0)\xrightarrow{\text{D}}\min_{x\in S^*(\mu)}\max_{\lambda\in\Lambda(\mu)}L(x,\lambda,Z). \tag{3.8}$$

Consider now the case where f_i are given as the expected value functions and ψ_{in} are their sample mean estimators. Suppose that for almost every ω, the functions $g_i(\cdot, \omega)$, $i = 0, \dots, r$, are convex, that the assumptions (a)–(c) hold for every function g_i, $i = 0, \dots, r$, and that program $(\mathbf{P}_0)$ has a unique optimal solution x_0 and a unique vector λ_0 of Lagrange multipliers. Then we obtain that $n^{1/2}(\hat{\varphi}_n - \varphi_0)$ is asymptotically normal with zero mean and variance

$$\sigma^2 = E\{L(x_0, \lambda_0, g)^2\} - [E\{L(x_0, \lambda_0, g)\}]^2, \tag{3.9}$$

where

$$L(x, \lambda, g) = g_0(x, \omega) + \sum_{i=1}^{r} \lambda_i g_i(x, \omega).$$

In situations where convexity is not presented it has to be replaced by differentiability assumptions. Suppose that the functions f_i, $i = 0, \dots, r$, of the program $(\mathbf{P}_0)$ are continuously differentiable. Consider an optimal solution $\bar{x} \in S^*(\mu)$ and suppose that $\bar{x}$ is an interior point of S. Then, under a constraint qualification, there is a vector $\bar{\lambda}$ such that the first order (Kuhn–Tucker) necessary conditions, given in (3.6) and (3.5), hold. We assume that to every $\bar{x} \in S^*(\mu)$ corresponds a *unique* vector $\bar{\lambda}(\bar{x})$ of Lagrange multipliers. This, in itself, is a constraint qualification (cf. [27]).

Consider the linear space $\mathrm{Lip}(S)$ of real valued, Lipschitz continuous on S, functions $\zeta(x)$. Choose a point $x_0 \in S$ and for $\zeta \in \mathrm{Lip}(S)$ define the norm

$$\|\zeta\| = |\zeta(x_0)| + \sup_{\substack{x, y \in S \\ x \neq y}} \frac{|\zeta(x) - \zeta(y)|}{\|x - y\|}. \tag{3.10}$$

Endowed with this norm, referred to as Lipschitz norm, $\mathrm{Lip}(S)$ becomes a Banach space. Notice that $\mathrm{Lip}(S)$ lies somewhere between the spaces $C(S)$ and $C^1(S)$, that is, $C(S) \subset \mathrm{Lip}(S) \subset C^1(S)$. Recall that $C^1(S)$ is the space of real valued, continuously differentiable on an open neighborhood of S, functions $\zeta(x)$ equipped with the norm

$$\|\zeta\| = \sup_{x \in S} |\zeta(x)| + \sup_{x \in S} \|\nabla\zeta(x)\|. \tag{3.11}$$

The following theorem is closely related to results of Levitin [29], Lempio and Maurer [28] and Gauvin and Dubeau [14]. Its proof is given in the appendix.

THEOREM 3.6

Consider the optimal value function $\bar{\varphi}(\xi)$, $\xi \in \mathscr{X}$, of the program $(\mathbf{P}_\xi)$ with equality and inequality constraints and $\mathscr{X} = \mathrm{Lip}(S) \times \cdots \times \mathrm{Lip}(S)$. Suppose that S is compact, that $\mu = (f_0, \dots, f_r) \in \mathscr{X}$, that the functions f_i, $i = 0, \dots, r$, are continuously differentiable in a neighborhood of the optimal set $S^*(\mu)$, that all points of $S^*(\mu)$ are interior points of S and to every $x \in S^*(\mu)$ corresponds a unique vector $\bar{\lambda}(x)$ of Lagrange multipliers. Then $\bar{\varphi}$ is Hadamard directionally differentiable at μ and

$$\bar{\varphi}'_\mu(\zeta) = \min_{x \in S^*(\mu)} L(x, \bar{\lambda}(x), \zeta). \tag{3.12}$$

Formula (3.12) suggests that for the basic scheme of expected value-sample mean construction and if $S^*(\mu) = \{x_0\}$ and $\lambda_0 = \bar{\lambda}(x_0)$, then $n^{1/2}(\hat{\varphi}_n - \varphi_0)$ is asymptotically normal with zero mean and variance σ^2 given in (3.9). The required regularity conditions, however, are more delicate here since they have to be verified in the Cartesian product of the spaces Lip(S). This problem becomes even more apparent in analysis of optimal solutions and puts serious limitations on applicability of the delta method in nondifferentiable cases. We discuss this further in the next section.

4. Asymptotic properties of optimal solutions

In this section we briefly discuss asymptotic behavior of optimal solutions $\hat{x}_n$ of the program ($\hat{\mathbf{P}}_n$). Analysis of the corresponding optimal solution mapping $\bar{x}(\cdot)$ is more delicate than the first order analysis of $\bar{\varphi}(\cdot)$ and more restrictive assumptions are required. In order to get a general idea about the speed of convergence of $\hat{x}_n$ let us consider the case where the feasible sets of programs ($\mathbf{P}_0$) and ($\hat{\mathbf{P}}_n$) coincide and are given by the set S. Consider the objective function $f(x)$ of the true program and let W be a convex neighborhood of the corresponding optimal set $S^*(f)$. Denote

$$\text{dist}(x, C) = \inf\{ \|x - y\| : y \in C\},$$

the distance from a point x to a set C.

ASSUMPTION A

There exists a positive constant α such that

$$f(x) \geqslant \inf_{x \in S} f(x) + \alpha[\text{dist}(x, S^*(f))]^2 \tag{4.1}$$

for all $x \in S \cap W$.

The above assumption can be ensured by various forms of second order sufficient conditions. Usually such sufficient conditions imply that a considered optimal solution is locally unique and hence the set $S^*(f)$ is discrete.

For a function $\xi : S \to \mathbb{R}$ let $\bar{x}(\xi)$ be a minimizer of $\xi(x)$ over S.

LEMMA 4.1

Suppose that assumption A holds, that the function $\delta(x) = \xi(x) - f(x)$ is Lipschitz continuous on $S \cap W$ and let $\bar{x}(\xi) \in W$. Then

$$\text{dist}(\bar{x}(\xi), S^*(f)) \leqslant \alpha^{-1}\kappa(\delta), \tag{4.2}$$

where

$$\kappa(\delta)=\sup\left\{\frac{|\delta(x)-\delta(y)|}{\|x-y\|}: x\in S^*(f),\ y\in S\cap W,\ x\neq y\right\}. \tag{4.3}$$

Proof

Consider a positive number ϵ and let x_0 be an element of the set $S^*(f)$ such that, for $\bar{x}=\bar{x}(\xi)$,

$$\|\bar{x}-x_0\|\leqslant \operatorname{dist}(\bar{x}, S^*(f))+\epsilon.$$

We have that $\xi(\bar{x})$ is less than or equal to $\xi(x_0)$ and hence

$$f(\bar{x})-f(x_0)=\delta(x_0)-\delta(\bar{x})+\xi(\bar{x})-\xi(x_0)\leqslant\delta(x_0)-\delta(\bar{x})$$
$$\leqslant\kappa(\delta)\,\|x_0-\bar{x}\|.$$

On the other hand, by assumption A,

$$f(\bar{x})-f(x_0)\geqslant\alpha(\|\bar{x}-x_0\|-\epsilon)^2.$$

Putting these two inequalities together and letting $\epsilon\to 0^+$, we obtain (4.2). □

Inequality (4.2) gives a simple bound for the speed of convergence of $\hat{x}_n=\bar{x}(\psi_n)$ to the optimal set of the true program. It depends on the Lipschitz constant of the difference function $\delta_n(x)=\psi_n(x)-f(x)$. If $\delta_n(x)$ is continuously differentiable then, under assumption A, $\hat{x}_n$ converges to $S^*(f)$ at a rate which is of order $O(\kappa_n)$, at most, where κ_n is the supremum of $\|\nabla\delta_n(x)\|$ over a neighborhood of $S^*(f)$. In the differentiable case of expected value-sample mean construction it follows from the Central Limit Theorem, applied to the gradients of the involved functions, that this supremum is of stochastic order $O_p(n^{-1/2})$. Consequently in this case $\hat{x}_n$ converges to $S^*(f)$ at a rate of $O_p(n^{-1/2})$.

In order to apply the delta method to the evaluation of the asymptotic distribution of $\hat{x}_n$, one has to calculate directional derivatives of the corresponding optimal solution mapping $\bar{x}(\cdot)$. There are basically two approaches to differential analysis of $\bar{x}(\cdot)$. In the first method the optimization problem is replaced by equations representing the associated necessary, and occasionally sufficient, optimality conditions. These equations are then investigated by application of an appropriate form of the Implicit Function Theorem. This approach was exploited by many authors and various extensions of the classical Implicit Function Theorem were proposed (see, e.g., [13] and references therein). In this respect we cite the recent work on generalized equations of Robinson [35–37] and King and Rockafellar [26]. In the second approach the optimization problem is approximated by a simpler one, directly [15,40,41].

The calculated directional derivatives can be employed to write in a closed form the asymptotic distribution of $\hat{x}_n$ (cf. [23,24,43]). Unfortunately, it seems that the described delta method is not well suited for proving the obtained asymptotics of $\hat{x}_n$ in nondifferentiable cases. Such nondifferentiable situations appear naturally, for example, at the second stage of stochastic programming with recourse [47]. Then the different techniques should be applied. A basic theory for dealing with nondifferentiable cases was laid down by Huber [20] in his work on statistical theory of M-estimators. An interesting discussion of the involved ideas can be found in [31]. A derivation of asymptotics of $\hat{x}_n$ when the feasible set is fixed is given in [42].

Appendix

The following proof of theorem 2.1 is taken from Gill [16] and is based on the Skorohod–Dudley–Wichura almost sure representation theorem.

THEOREM A1

(Representation theorem)

Suppose $Z_n \xrightarrow{D} Z$ in $\mathscr{Z}$, where Z takes values in a separable subset of $\mathscr{Z}$. Then there exists a sequence Z_n', Z' defined on a single probability space such that $Z_n' \overset{D}{=} Z_n$, for all n, $Z' \overset{D}{=} Z$, and $Z_n' \to Z'$ almost surely.

Here $Z' \overset{D}{=} Z$ means that the probability measures, induced by Z' and Z, coincide.

Proof of theorem 2.1

Consider the sequence $Z_n = \tau_n(Y_n - \mu)$ of random elements in $\mathscr{Z}$. By theorem A1, there exist Z_n', Z', on a single probability space with $Z_n' \overset{D}{=} Z_n$, $Z' \overset{D}{=} Z$, and $Z_n' \to Z'$ almost surely. Now for $Y_n' = \mu + \tau_n^{-1} Z_n$ we have $Y_n' \overset{D}{=} Y_n$ for all n. It follows that for al n large enough, $Y_n' \in \mathscr{K}$ almost surely. Then we obtain from Hadamard differentiability tangentially to $\mathscr{K}$, that

$$\tau_n[g(Y_n') - g(\mu)] \to g'_{\mu,\mathscr{K}}(Z') \text{ a.s.} \tag{1a}$$

Since convergence almost surely implies convergence in distribution and the terms in (1a) have the same distributions as the corresponding terms in (2.5), the asymptotic result (2.5) follows.

If the set $\mathscr{K}$ is convex, then the contingent cone $T_\mu(\mathscr{K})$ coincides with the tangent cone which is given by the topological closure of the set $\cup\{t^{-1}(\mathscr{K} - \mu) : t > 0\}$. In this case $\mathscr{K} \subset \mu + T_\mu(\mathscr{K})$. Therefore $Z_n' \in T_\mu(\mathscr{K})$ almost surely and, by continuity of $g'_{\mu,\mathscr{K}}(\cdot)$ on $T_\mu(\mathscr{K})$,

$$g'_{\mu,\mathscr{K}}(Z_n') \to g'_{\mu,\mathscr{K}}(Z') \text{ a.s.}$$

Together with (1a) this implies that

$$\tau_n[g(Y_n') - g(\mu)] = g'_{\mu,\mathscr{K}}(Z_n') + o_p(1)$$

and hence (2.6) follows.

Notice that it follows from the definition of Z_n' that the distribution of Z', and hence the distribution of Z, is concentrated on $T_\mu(\mathcal{X})$. □

Proof of theorem 3.4

Let v be a point in S such that $f_i(v) < -\epsilon$, $i = 1, \ldots, r$, for some $\epsilon > 0$. Existence of such a point is guaranteed by the Slater condition. It also follows from the Slater condition and compactness of S that the sets $S^*(\mu)$ and $\Lambda(\mu)$ are nonempty and compact. Consider a sequence $\xi_n = (\xi_{0n}, \ldots, \xi_{rn}) \in \mathscr{X}$ converging to μ and a point $\bar{x} \in S^*(\mu)$. Since S is convex, the segment joining v and $\bar{x}$ is contained in S. It also follows from convexity of the functions $f_i(x)$ that

$$f_i[\bar{x} + \alpha(v - \bar{x})] \leqslant (1-\alpha) f_i(\bar{x}) + \alpha f_i(v) \leqslant -\alpha\epsilon$$

for all $\alpha \in [0, 1]$. Since the functions $\xi_{in}(x)$ converge to $f_i(x)$ uniformly on S as $n \to \infty$, we obtain that there exists a sequence $\alpha_n \to 0^+$ such that for all n large enough, $\xi_{in}(v_n) \leqslant 0$, $i = 1, \ldots, r$, where $v_n = \bar{x} + \alpha_n(v - \bar{x})$. It follows that for sufficiently large n, the optimal set $S^*(\xi_n)$ is nonempty and if $x_n \in S^*(\xi_n)$, then the distance from x_n to $S^*(\mu)$ tends to zero.

Now consider a sequence $\{\xi_n\} \subset \mathscr{X}$ of the form $\xi_n = \mu + t_n \zeta_n$ such that $t_n \to 0^+$ and $\{\zeta_n\}$ converges to a vector ζ. Let $\bar{\lambda} \in \Lambda(\mu)$. Then

$$\bar{\varphi}(\xi_n) \geqslant L(x_n, \bar{\lambda}, \xi_n)$$

and for any $\bar{x} \in S^*(\mu)$,

$$\bar{\varphi}(\mu) = L(\bar{x}, \bar{\lambda}, \mu).$$

By the assumption of convexity we have then from the optimality of $\bar{x}$ that

$$L(\bar{x}, \bar{\lambda}, \mu) \leqslant L(x_n, \bar{\lambda}, \mu).$$

Consequently

$$\bar{\varphi}(\xi_n) - \bar{\varphi}(\mu) \geqslant L(x_n, \bar{\lambda}, \xi_n) - L(x_n, \bar{\lambda}, \mu) = t_n L(x_n, \bar{\lambda}, \zeta_n).$$

Since $\bar{\lambda}$ was an arbitrary element of $\Lambda(\mu)$ and because of continuity of the Lagrangian function we obtain that

$$\liminf_{n \to \infty} \frac{\bar{\varphi}(\xi_n) - \bar{\varphi}(\mu)}{t_n} \geqslant \min_{x \in S^*(\mu)} \max_{\lambda \in \Lambda(\mu)} L(x, \lambda, \zeta).$$

On the other hand, consider a point $\bar{x} \in S^*(\mu)$. Let λ_n be a vector of Lagrange multipliers corresponding to a point $x_n \in S^*(\xi_n)$. Notice that by duality the distance from λ_n to $\Lambda(\mu)$ tends to zero as well as the distance from x_n to $S^*(\mu)$. We have that

$$\bar{\varphi}(\xi_n) = L(x_n, \lambda_n, \xi_n)$$

and by the assumption of convexity of $(\mathbf{P}_{\xi_n})$ and optimality of x_n,

$$L(x_n, \lambda_n, \xi_n) \leqslant L(\bar{x}, \lambda_n, \xi_n).$$

Also

$$\bar{\varphi}(\mu) \geqslant L(\bar{x}, \lambda_n, \mu)$$

and hence

$$\bar{\varphi}(\xi_n) - \bar{\varphi}(\mu) \leqslant L(\bar{x}, \lambda_n, \xi_n) - L(\bar{x}, \lambda_n, \mu) = t_n L(\bar{x}, \lambda_n, \zeta_n).$$

Since the point $\bar{x}$ was an arbitrary point of $S^*(\mu)$ and by continuity of the Lagrangian function, it follows that

$$\limsup_{n\to\infty} \frac{\bar{\varphi}(\xi_n) - \bar{\varphi}(\mu)}{t_n} \leqslant \min_{x\in S^*(\mu)} \max_{\lambda\in\Lambda(\mu)} L(x, \lambda, \zeta)$$

and hence the proof is complete.

Proof of theorem 3.6

Consider a point $\bar{x} \in S^*(\mu)$, the corresponding vector $\bar{\lambda} = \bar{\lambda}(\bar{x})$ of Lagrange multipliers, the index set

$$J = \{i : f_i(\bar{x}) = 0,\ i = q+1, \ldots, r\}$$

of active at $\bar{x}$ inequality constraints of the program ($\mathbf{P}_0$), and the index sets

$$J_+ = \{i \in J : \bar{\lambda}_i > 0\} \quad \text{and} \quad J_0 = \{i \in J : \bar{\lambda}_i = 0\}.$$

Existence and uniqueness of $\bar{\lambda}(x)$ is equivalent to the following regularity conditions. The gradient vectors $\nabla f_i(\bar{x})$, $i \in \{1, \ldots, q\} \cup J_+$, are linearly independent and there exists a vector v such that $v \cdot \nabla f_i(\bar{x}) = 0$, $i \in \{1, \ldots, q\} \cup J_+$, and $v \cdot \nabla f_i(\bar{x}) < 0$, $i \in J_0$ [27]. These regularity conditions are the Mangasarian–Fromowitz constraint qualification, at the point $\bar{x}$, for the set defined by constraints $f_i(x) = 0$, $i \in \{1, \ldots, q\} \cup J_+$, and $f_i(x) \leqslant 0$, $i \in J_0$. Let us consider, for $\xi = (\xi_0, \ldots, \xi_r) \in \mathscr{X}$, the set

$$\Phi(\xi) = \{x \in S : \xi_i(x) = 0,\ i \in \{1, \ldots, q\} \cup J_+;\ \xi_i(x) \leqslant 0,\ i \in J_0\}.$$

It then follows that there exists a positive number c such that for every $\xi \in \mathscr{X}$ sufficiently close to μ (in the Lipschitz norm) there exists a point $\bar{v}(\xi) \in \Phi(\xi)$ such that

$$\| \bar{x} - \bar{v}(\xi) \| \leqslant c \left\{ \sum_{i\in\{1,\ldots,q\}\cup J_+} | \xi_i(\bar{x}) | + \sum_{i\in J_0} \xi_i^+(\bar{x}) \right\}, \tag{2a}$$

where α^+ denotes $\max\{0, \alpha\}$. This result follows from an adaptation of the Robinson–Ursescu stability theorem [34,45] to the present nondifferentiable situation. Because of the Mangasarian–Fromowitz constraint qualification, the system of constraints defining the set $\Phi(\xi)$ for $\xi = \mu$ is stable at $x = \bar{x}$. Small Lipschitz perturbations in ξ preserve this stability property [5, theorem 2.1]. A thorough discussion of the involved concept of metric regularity and additional references can be found in [5].

By the definition of Lipschitz norm it follows from (2a) that

$$\| \bar{x} - \bar{v}(\xi) \| \leqslant c \| \xi - \mu \|, \tag{3a}$$

with possibly a different constant c. Consider now a sequence $\xi_n = \mu + t_n \zeta_n$ with $\zeta_n \to \zeta$ and $t_n \to 0^+$. Let $v_n = \bar{v}(\xi_n)$. Because of (3a) we have that $\| \bar{x} - v_n \|$ is of order $O(t_n)$. It follows that for n large enough, the optimal set $S^*(\xi_n)$ is nonempty and if $x_n \in S^*(\xi_n)$, then the distance from x_n to $S^*(\mu)$ tends to zero. Moreover,

$$\bar{\varphi}(\xi_n) \leqslant \xi_{0n}(v_n) = L(v_n, \bar{\lambda}, \xi_n) = L(v_n, \bar{\lambda}, \mu) + t_n L(v_n, \bar{\lambda}, \zeta_n).$$

From the first order necessary conditions and since $\| \bar{x} - v_n \| = O(t_n)$ we obtain that

$$L(v_n, \bar{\lambda}, \mu) = L(\bar{x}, \bar{\lambda}, \mu) + o(t_n).$$

Furthermore, $\bar{\varphi}(\mu)$ is equal to $L(\bar{x}, \bar{\lambda}, \mu)$ and hence

$$\bar{\varphi}(\xi_n) - \bar{\varphi}(\mu) \leqslant t_n L(v_n, \bar{\lambda}, \zeta_n) + o(t_n).$$

Since $\bar{x}$ was an arbitrary point of $S^*(\mu)$ it follows then that

$$\limsup_{n \to \infty} \frac{\bar{\varphi}(\xi_n) - \bar{\varphi}(\mu)}{t_n} \leqslant \min_{x \in S^*(\mu)} L(x, \bar{\lambda}(x), \zeta).$$

On the other hand, consider a sequence $x_n \in S^*(\xi_n)$ converging to a point $\bar{x}$. We have then that $\bar{x} \in S^*(\mu)$. Because of the assumption of uniqueness of $\bar{\lambda}(\bar{x})$, it follows again from the stability theorem, now applied to the system defining $\Phi(\mu)$, that there exist points $u_n \in \Phi(\mu)$ such that $\| u_n - x_n \| = O(t_n)$. Then

$$\bar{\varphi}(\mu) \leqslant f_0(u_n) = L(u_n, \bar{\lambda}(\bar{x}), \mu)$$

and

$$\bar{\varphi}(\xi_n) \geqslant L(x_n, \bar{\lambda}(\bar{x}), \xi_n) = L(x_n, \bar{\lambda}(\bar{x}), \mu) + t_n L(x_n, \bar{\lambda}(\bar{x}), \zeta_n).$$

By the mean value theorem it follows from the first order necessary conditions and continuous differentiability of $L(\cdot, \bar{\lambda}(\bar{x}), \mu)$ that

$$L(x_n, \bar{\lambda}(\bar{x}), \mu) - L(u_n, \bar{\lambda}(\bar{x}), \mu) = o(t_n).$$

Consequently

$$\bar{\varphi}(\xi_n) - \bar{\varphi}(\mu) \geqslant t_n L(x_n, \bar{\lambda}(\bar{x}), \zeta_n) + o(t_n).$$

We obtain that

$$\liminf_{n \to \infty} \frac{\bar{\varphi}(\xi_n) - \bar{\varphi}(\mu)}{t_n} \geqslant \min_{x \in S^*(\mu)} L(x, \bar{\lambda}(x), \zeta)$$

and the proof is complete. □

References

[1] A. Araujo and E. Giné, *The Central Limit Theorem for Real and Banach Valued Random Variables* (Wiley, 1980).

[2] S. Asmussen and R.Y. Rubinstein, The efficiency and heavy traffic properties of the score function method in sensitivity analysis of queueing models, Preprint, Aalborg University, Denmark, and Technion, Haifa, Israel (1989).

[3] V.I. Averbukh and O.G. Smolyanov, The theory of differentiation in linear topological spaces, Russian Math. Surveys 22 (1967) 201–258.

[4] P. Billingsley, *Convergence of Probability Measures* (Wiley, 1968).

[5] R. Cominetti, Metric regularity tangent sets and second order optimality conditions, Appl. Math. Optim. 21 (1990) 265–287.

[6] J.M. Danskin, *The theory of Max-min and Its Applications to Weapons Allocation Problems* (Springer, 1967).

[7] G.B. Dantzig and W. Glynn, Parallel processors for planning under uncertainty, Preprint, Stanford University (1989).

[8] J. Dupačová, Stability in stochastic programming with recourse. Contaminated distribution, Math. Progr. Study 27 (1986) 133–144.

[9] J. Dupačová, Stochastic programming with incomplete information: A survey of results on post optimization and sensitivity analysis, Optimization 18 (1987) 507–532.

[10] J. Dupačová and R.J.-B. Wets, Asymptotic behavior of statistical estimators and optimal solutions for stochastic optimization problems, Ann. Statist. 16 (1988) 1517–1549.

[11] Yu. Ermoliev and R.J.-B. Wets (eds.), *Numerical Techniques for Stochastic Optimization* (Springer, Berlin, 1988).

[12] L.T. Fernholz, Von Mises calculus for statistical functionals, *Lecture Notes in Statistics* (Springer, New York, 1983).

[13] A.V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming* (Academic Press, 1983).

[14] J. Gauvin and F. Dubeau, Differential properties of the marginal function in mathematical programming, Math. Progr. Study 19 (1982) 101–119.

[15] J. Gauvin and R. Janin, Directional behaviour of optimal solutions in nonlinear mathematical programming, Math. Oper. Res. 13 (1988) 629–649.

[16] R.D. Gill, Non- and semiparametric maximum likelihood estimators and the von Mises method (Part I), Scand. J. Statist. 16 (1989) 97–124.

[17] E.G. Golshtein, *Theory of Convex Programming*, Trans. Math. Monographs 26 (American Mathematical Society, Providence, RI, 1972).

[18] R. Grübel, The length of the shorth, Ann. Statist. 16 (1988) 619–628.

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

[20] P.J. Huber, The behaviour of maximum likelihood estimates under nonstandard condition, in: *Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability*, vol. 1 (University of California Press, Berkeley, 1967) pp. 221–233.

[21] P.J. Huber, *Robust Statistics* (Wiley, 1981).

[22] A.D. Ioffe and V.M. Tihomirov, *Theory of Extremal Problems* (North-Holland, 1979).

[23] A.J. King, Asymptotic behavior of solutions in stochastic optimization: nonsmooth analysis and the derivation of non-normal limit distributions, Dissertation, University of Washington (1986).

[24] A.J. King, Asymptotic distributions for solutions in stochastic optimization and generalized *M*-estimation, Working paper, International Institute for Applied Systems Analysis, Laxenburg, Austria (1988).

[25] A.J. King, Generalized delta theorems for multivalued mappings and measurable selections, Math. Oper. Res. 14 (1989) 720–736.
[26] A.J. King and R.T. Rockafellar, Sensitivity analysis for nonsmooth generalized equations, Working paper, IBM Research Division, T.J. Watson Research Center (1989).
[27] J. Kyparisis, On uniqueness of Kuhn–Tucker multipliers in nonlinear programming, Math. Progr. 32 (1985) 242–246.
[28] F. Lempio and H. Maurer, Differential stability in infinite-dimensional nonlinear programming, Appl. Math. Optim. 6 (1980) 139–152.
[29] E.S. Levitin, On differential properties of the optimum value of parametric problems of mathematical programming, Sov. Math. Dokl. 15 (1974) 603–608.
[30] M.Z. Nashed, Differentiability and related properties of nonlinear operators: some aspects of the role of differentials in nonlinear functional analysis, in: *Nonlinear Functional Analysis and Applications*, ed. L.B. Rall (Academic Press, New York, 1971) pp. 103–309.
[31] D. Pollard, Asymptotics via empirical processes, Statist. Sci. 4 (1989) 341–366.
[32] C.R. Rao, *Linear Statistical Inference and Its Applications* (Wiley, 1973).
[33] J.A. Reeds, On the definition of von Mises functionals, Ph.D. thesis, Harvard University (1976).
[34] S.M. Robinson, Stability theory for systems of inequalities, part II: Differentiable nonlinear systems, SIAM J. Numer. Anal. 13 (1976) 497–513.
[35] S.M. Robinson, Generalized equations and their solutions, part 1: basic theory, Math. Progr. Study 10 (1979) 128–141.
[36] S.M. Robinson, Strongly regular generalized equations, Math. Oper. Res. 5 (1980) 43–62.
[37] S.M. Robinson, An implicit function theorem for B-differentiable functions, Working paper, International Institute for Applied Systems Analysis, Laxenburg, Austria (1988).
[38] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, 1970).
[39] R.Y. Rubinstein, Monte Carlo methods for performance evaluation, sensitivity analysis and optimization of stochastic systems, to appear in: *Encyclopedia of Computer Science and Technology*, eds. Kent and Williams (1990).
[40] A. Shapiro, Second-order sensitivity analysis and asymptotic theory of parametrized nonlinear programs, Math. Progr. 33 (1985) 280–299.
[41] A. Shapiro, Gâteaux, Fréchet and Hadamard directional differentiability of functional solutions in stochastic programming, Operations Research and Statistics Mimeograph Series No. 395, Technion, Haifa, Israel (1988).
[42] A. Shapiro, Asymptotic properties of statistical estimators in stochastic programming, Ann. Statist. 17 (1989) 841–858.
[43] A. Shapiro, On differential stability in stochastic programming, Math. Progr. 47 (1990) 107–116.
[44] A. Shapiro, On concepts of directional differentiability, J. Optim. Theory Appl. 66 (1990) 477–487.
[45] C. Ursescu, Multifunctions with convex closed graph, Czech. Math. J. 25 (1975) 438–441.
[46] J.A. Wellner, Discussion of Gill's paper "Non- and semiprametric maximum likelihood estimators and the von Mises method (Part I)", Scand. J. Statist. 16 (1989) 124–127.
[47] R.J.-B. Wets, Stochastic programming: Solution techniques and approximation schemes, in: *Mathematical Programming. The State of the Art*, eds. A. Bachem, M. Grötschel and B. Korte (Springer, Berlin, 1983).