MATHEMATICS OF OPERATIONS RESEARCH
Vol 18, No 4, November 1993
Printed in U S A

ASYMPTOTIC BEHAVIOR OF OPTIMAL SOLUTIONS IN STOCHASTIC PROGRAMMING

ALEXANDER SHAPIRO

Asymptotic behavior of optimal solutions $\hat{x}_n$ of a sequence of stochastic programming problems is studied. Variational and generalized equations approaches are discussed. An expansion of $\hat{x}_n$ in terms of a parametrized mathematical programming problem, depending on a single random vector, is given. When optimal solutions of the parametrized program are directionally differentiable, this expansion leads to a close form expression for the asymptotic distribution of $\hat{x}_n$. Applicability of the involved regularity conditions to nondifferentiable cases, and in particular to stochastic programming with recourse, is discussed.

1. Introduction. In this paper we study asymptotic behavior of optimal solutions of the following stochastic programming problems. Let $(\Omega, \mathscr{F}, P)$ be a probability space and consider a sequence $\psi_n(x) = \psi_n(x, \omega)$, $(x, \omega) \in \Re^m \times \Omega$, of real-valued random functions, a sequence $\gamma_n(x) = \gamma_n(x, \omega)$ of random mappings, $\gamma_n: \Re^m \times \Omega \to \Re^k$, a closed set $S \subset \Re^m$, a closed convex cone $K \subset \Re^k$ and the associated programs

$$(\hat{P}_n) \qquad \begin{array}{ll} \text{minimize} & \psi_n(x), \\ \text{subject to} & \gamma_n(x) \in K, \\ & x \in S. \end{array}$$

We assume that sequences $\{\psi_n(x)\}$ and $\{\gamma_n(x)\}$ converge in a certain sense to (deterministic) functions $f: \Re^m \to \Re$ and $g: \Re^m \to \Re^k$, respectively, and view $(\hat{P}_n)$ as approximations of the "true" or "limiting" program

$$(P_0) \qquad \begin{array}{ll} \text{minimize} & f(x), \\ \text{subject to} & g(x) \in K, \\ & x \in S. \end{array}$$

Our study is motivated by the following basic example. Let $\xi_i = \xi_i(\omega)$, $i = 1, 2, \ldots$, be a sequence (a sample) of independent s-dimensional random vectors with a common probability measure (distribution) F. Consider a mapping $H: \Re^m \times \Re^s \to \Re \times \Re^k$, $H(x, \xi) = (h(x, \xi), G(x, \xi))$, and

$$(1.1) \qquad (\psi_n(x), \gamma_n(x)) = n^{-1} \sum_{i=1}^{n} H(x, \xi_i).$$

We refer to the random functions $\psi_n(x)$ and $\gamma_n(x)$ in the form (1.1), as sample mean functions and to the corresponding program $(\hat{P}_n)$ as the sample mean construction. By the Law of Large Numbers we then have that $(\psi_n(x), \gamma_n(x))$ converge (pointwise)

Received April 3, 1991; revised July 25, 1992.
AMS 1980 subject classification. Primary: 90C15, 90C31
IAOR 1973 subject classification. Main: Programming: Stochastic.
OR/MS Index 1978 subject classification. Primary: 663 Programming/Stochastic
Key words. Stochastic programming, empirical probability measure, asymptotic distribution, M-estimators, parametrized programming, generalized equations.

0364-765X/93/1804/0829/$01.25

with probability one to the expected value mapping

$$(f(x), g(x)) = \int H(x, \xi) F(d\xi), \tag{1.2}$$

which leads to the limiting program of the form (P_0). The above construction can be extended, for example, by considering a composition of several sample mean functions.

Another way of looking at the sample mean construction is to write the involved random functions in the form

$$(\psi_n(x), \gamma_n(x)) = \int H(x, \xi) F_n(d\xi), \tag{1.3}$$

with $F_n = F_n(\cdot, \omega)$ being the empirical probability measure on $\mathfrak{R}^s$ corresponding to the sample $\xi_1, \dots, \xi_n$. It is possible to generate a random mapping in the form (1.3) by considering probability measures F_n (not necessarily empirical) converging in some sense to a "true" probability measure F (cf. Dupačová and Wets 1988, p. 1524).

In the case of the sample mean construction, optimal solutions $\hat{x}_n$ of the program $(\hat{P}_n)$ correspond to the M-estimators which can be considered as extensions of the standard maximum likelihood estimators (see Huber 1981). Statistical properties of M-estimators were studied by Huber 1967. He showed that in the unconstrained case (where $S = \mathfrak{R}^m$ and there are no constraints) under mild regularity conditions the minimizers $\hat{x}_n$ have asymptotically normal distributions. Asymptotic behavior of $\hat{x}_n$ is more involved if the "true" optimal solution of the program (P_0) lies on the boundary of the set S or when there is a number of inequality constraints present in the problem (Dupačová and Wets 1988, King 1986, King and Rockafellar 1993, Shapiro 1989, Shapiro 1990). Asymptotics which were derived in this case suggest that the asymptotic distribution of $\hat{x}_n$ may be nonnormal (King 1986). Necessary and sufficient conditions for the asymptotic normality of $\hat{x}_n$ are discussed in King and Rockafellar 1993, Shapiro 1989, Shapiro 1990.

In this paper we derive an expansion of $\hat{x}_n$ in terms of an associated parametrized mathematical programming problem depending on a single random vector. This has an advantage over previous results in that the involved differentiability assumptions are reduced to a minimum. Problems where the corresponding random functions are not everywhere differentiable appear naturally, for example, at the second stage of stochastic programming with recourse. We discuss applicability of our results in that case in §4. In §2 we consider the situations where the feasible set is fixed, i.e., it is the same for the true and the approximating programs. In §3 we discuss the general case with the set S being convex. In order to deal with this case we reformulate the corresponding optimization problem in a form of generalized equations.

It is convenient to use symbols $o_p(\cdot)$ and $O_p(\cdot)$ as stochastic analogues of $o(\cdot)$ and $O(\cdot)$ (Mann and Wald 1943). In particular, $X_n = o_p(1)$ is another way of writing that the random vectors X_n converge to zero in probability as $n \to \infty$, and $X_n = O_p(1)$ means that X_n, $n = 1, \dots,$ are bounded in probability. The scalar product of two vectors $x, y \in \mathfrak{R}^n$ will be denoted by $x \cdot y$ and by $B(x, r)$ we denote the ball $\{y: \|y - x\| \leqslant r\}$.

2. Asymptotic behavior of the optimal solutions in the case of fixed feasible set. In this section we discuss the case where programs $(\hat{P}_n)$ involve only deterministic constraints. The common feasible set of the true and the approximating programs will be denoted by S. That is, we consider programs

$$(P_0) \qquad \text{minimize } f(x) \quad \text{subject to } x \in S,$$

and

$$(\hat{P}_n) \qquad \text{minimize } \psi_n(x) \quad \text{subject to } x \in S.$$

Asymptotic analysis of optimal solutions of $(\hat{P}_n)$ involves two types of regularity conditions. One set of conditions concerns the deterministic program (P_0) alone while the other set of assumptions is related to the stochastic behavior of the random functions $\psi_n(x)$. We introduce the required "deterministic" conditions first.

ASSUMPTION A1. *Program (P_0) has a unique optimal solution x_0.*

ASSUMPTION A2. *There exist a neighborhood W of x_0 and a positive constant α such that*

$$f(x) \geqslant f(x_0) + \alpha\|x - x_0\|^2 \tag{2.1}$$

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

Assumption A2 will be referred to as *the second-order growth condition* and can be ensured by various forms of second-order sufficient conditions (e.g. Robinson 1982, §2). We shall discuss it later.

In the subsequent analysis we shall make use of the following variational principle. Let $\phi(x)$ be a real valued function and let x^* be an ϵ-optimal solution of the problem of minimization of $\phi(x)$ over S. That is, $\epsilon \geqslant 0$, $x^* \in S$ and

$$\phi(x^*) \leqslant \inf_{x \in S} \phi(x) + \epsilon.$$

Suppose that assumptions A1 and A2 hold and that $x^* \in W$. Then

$$\|x^* - x_0\| \leqslant \alpha^{-1}\kappa + \alpha^{-1/2}\epsilon^{1/2}, \tag{2.2}$$

where

$$\kappa = \sup\left\{\frac{|h(x) - h(x_0)|}{\|x - x_0\|} : x \in S \cap W, x \neq x_0\right\} \tag{2.3}$$

and $h(x) = \phi(x) - f(x)$ (see Shapiro 1991, Lemma 4.1 and Shapiro 1992, Lemma 2.1).

Now let ϵ_n be a sequence of nonnegative random numbers converging in probability to zero and let $\hat{x}_n = \hat{x}_n(\omega)$ be an ϵ_n-optimal solution of $(\hat{P}_n)$. We assume that $\hat{x}_n(\omega)$ are measurable. An interested reader is referred to Rockafellar 1976 and Rockafellar and Wets 1984 for a discussion of existence of measurable selections in optimization problems. We also assume that $\hat{x}_n$ is a consistent estimator of x_0, i.e., that $\hat{x}_n$ converges in probability to x_0 as $n \to \infty$. There are numerous studies where regularity conditions required to ensure consistency of $\hat{x}_n$ are discussed in detail (see, e.g., Dupačová and Wets 1988, Huber 1967 and references therein).

The following "stochastic" assumptions related to the probabilistic structure of the random functions $\psi_n(x)$ will be used.

ASSUMPTION B1. *In a neighborhood of the point x_0 the random functions*

$$\delta_n(x) = \psi_n(x) - f(x), \qquad n = 1,2,\ldots$$

are Lipschitz continuous and are differentiable at x_0 with probability one.

ASSUMPTION B2. *The gradients $\nabla\delta_n(x_0)$ are of stochastic order*

$$\|\nabla\delta_n(x_0)\| = O_p(n^{-1/2}). \tag{2.4}$$

ASSUMPTION B3. *There is a neighborhood W of x_0 such that*

$$\sup\left\{\frac{\|\nabla\delta_n(x) - \nabla\delta_n(x_0)\|}{n^{-1/2} + \|x - x_0\|} : x \in W \setminus E_n\right\} = o_p(1), \tag{2.5}$$

with E_n being the set of those x where $\nabla\delta_n(x)$ fails to exist.

In the case of the sample mean construction it follows by the Central Limit Theorem that, under standard regularity conditions, the random vectors $n^{1/2}\,\nabla\delta_n(x_0)$ converge in distribution to a multivariate normal. This implies that $n^{1/2}\,\nabla\delta_n(x_0)$ are bounded in probability and hence (2.4) follows. Assumption B3 is explicit in Huber 1967. It will be discussed further in §4.

Let us consider now the following optimization problem

$$\text{minimize } f(x) + v \cdot x \quad \text{subject to } x \in S, \tag{P_v}$$

depending on the parameter vector $v \in \Re^m$. Let $\bar{x}(v)$ be an optimal solution of (P_v). Note that for $v = 0$, program (P_v) coincides with the program (P_0) and hence $\bar{x}(0) = x_0$.

ASSUMPTION A3. *There exist a neighborhood W of x_0 and a positive constant α such that for every v in a neighborhood of zero, program (P_v) has an optimal solution $\bar{x} = \bar{x}(v) \in W$ and*

$$f(x) + v \cdot (x - \bar{x}) \geqslant f(\bar{x}) + \alpha\|x - \bar{x}\|^2 \tag{2.6}$$

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

Assumption A3 can be viewed as a uniform version of Assumption A2 applied to the program (P_v). It also can be ensured by various forms of second-order conditions. For example, suppose that $f(x)$ is representable in the form $f(x) = f_1(x) + f_2(x)$ where $f_1(x)$ is a convex function and $f_2(x)$ is a twice continuously differentiable function, that the Hessian matrix $\nabla^2 f_2(x_0)$ is positive definite and that the set S is convex and compact. Then $\bar{x} = \bar{x}(v)$ exists for all v and $\bar{x}(v) \to x_0$ as $v \to 0$. Moreover, by the first order necessary conditions applied to the program (P_v) at $\bar{x}$, we have that there is $z \in \partial f_1(\bar{x})$ such that

$$\{z + \nabla f_2(\bar{x}) + v\} \cdot (x - \bar{x}) \geqslant 0$$

for all $x \in S$. Consequently,

$$f(x) + v \cdot (x - \bar{x}) \geqslant f_1(x) - z \cdot (x - \bar{x}) + f_2(x) - \nabla f_2(\bar{x}) \cdot (x - \bar{x})$$

for every $x \in S$. Since

$$f_1(x) - f_1(\bar{x}) \geqslant z \cdot (x - \bar{x})$$

and

$$f_2(x) = f_2(\bar{x}) + \nabla f_2(\bar{x}) \cdot (x - \bar{x}) + \frac{1}{2}(x - \bar{x}) \cdot \nabla^2 f_2(x^*)(x - \bar{x}),$$

where x^* is a point on the segment joining x and $\bar{x}$, we obtain that for $x \in S$,

$$f(x) + v \cdot (x - \bar{x}) \geqslant f(\bar{x}) + \frac{1}{2}(x - \bar{x}) \cdot \nabla^2 f_2(x^*)(x - \bar{x}).$$

It remains to note that by the arguments of continuity the minimal eigenvalue of the matrix $\nabla^2 f(x^*)$ is greater than a positive constant β for all x^* in a neighborhood of x_0 and hence

$$(x - \bar{x}) \cdot \nabla^2 f_2(x^*)(x - \bar{x}) \geqslant \beta \|x - \bar{x}\|^2.$$

This shows that Assumption A3 holds here. Note that compactness of S was used in the above arguments only in order to ensure existence of $\bar{x}(v) \in W$ for all v sufficiently close to zero.

We can now formulate the main result of this section.

THEOREM 2.1. *Suppose that Assumptions* A1, A3 *and* B1–B3 *hold, that $\hat{x}_n$ converges in probability to x_0 and let $\epsilon_n = o_p(n^{-1})$. Then*

$$\hat{x}_n = \bar{x}(\nabla \delta_n(x_0)) + o_p(n^{-1/2}). \tag{2.7}$$

PROOF. Without loss of generality we can assume that the neighborhood W is an open ball centered at x_0. We first show that

$$\|\hat{x}_n - x_0\| = O_p(n^{-1/2}). \tag{2.8}$$

Since $\hat{x}_n$ converges in probability to x_0 we have that with probability tending to one, $\hat{x}_n \in W$ as $n \to \infty$. Moreover, Assumption A3 implies Assumption A2 and since $\epsilon_n = o_p(n^{-1})$ it follows from the upper bound (2.2) that

$$\|\hat{x}_n - x_0\| \leqslant \alpha^{-1}\kappa_n + o_p(n^{-1/2}), \tag{2.9}$$

where

$$\kappa_n = \sup\left\{\frac{|\delta_n(x) - \delta_n(x_0)|}{\|x - x_0\|} : x \in B(x_0, r_n),\ x \neq x_0\right\}$$

and $r_n = \|\hat{x}_n - x_0\|$. By the Mean Value Theorem for Lipschitz functions (Clarke 1983, p. 41) we have further that $\kappa_n \leqslant \bar{\kappa}_n$ where

$$\bar{\kappa}_n = \sup\{\|\nabla \delta_n(x)\| : x \in B(x_0, r_n) \setminus E_n\}.$$

Furthermore,

$$\bar{\kappa}_n \leqslant \|\nabla\delta_n(x_0)\| + \sup_{x \in B(x_0, r_n) \setminus E_n} \|\nabla\delta_n(x) - \nabla\delta_n(x_0)\|$$

which together with Assumption B3 implies that

$$\bar{\kappa}_n \leqslant \|\nabla\delta_n(x_0)\| + o_p(1)\left(n^{-1/2} + r_n\right).$$

It follows then from (2.9) that

$$r_n \leqslant \alpha^{-1}\|\nabla\delta_n(x_0)\| + o_p(1)\left(n^{-1/2} + r_n\right)$$

and, hence,

$$\left[1 + o_p(1)\right] r_n \leqslant \alpha^{-1}\|\nabla\delta_n(x_0)\| + o_p(n^{-1/2}).$$

This last inequality and Assumption B2 imply (2.8).

Now let us consider $\bar{x}_n = \bar{x}(\nabla\delta_n(x_0))$. By Assumption A3 it follows from (2.2) that for all v in a neighborhood of zero,

$$\|\bar{x}(v) - x_0\| \leqslant \alpha^{-1}\|v\|,$$

which together with Assumption B2 implies

$$\|\bar{x}_n - x_0\| = O_p(n^{-1/2}).$$

Applying again the upper bound (2.2) to the programs $(\hat{P}_n)$ and (P_v), $v = \nabla\delta_n(x_0)$, at the point $\bar{x}_n$ we obtain

$$\|\hat{x}_n - \bar{x}_n\| \leqslant \alpha^{-1}\mu_n + o_p(n^{-1/2}), \tag{2.10}$$

where

$$\mu_n = \sup\left\{\|\nabla\delta_n(x) - \nabla\delta_n(x_0)\| : x \in B(x_0, \rho_n) \setminus E_n\right\}$$

and

$$\rho_n = \max\left\{\|\hat{x}_n - x_0\|, \|\bar{x}_n - x_0\|\right\}.$$

(Note that $\nabla\delta_n(x) - \nabla\delta_n(x_0)$ is the gradient of the difference of the objective functions of the programs $(\hat{P}_n)$ and (P_v) for $v = \nabla\delta_n(x_0)$.) By Assumption B3 we have that

$$\mu_n = o_p(1)\left(n^{-1/2} + \rho_n\right)$$

and since $\rho_n = O_p(n^{-1/2})$ it follows that $\mu_n = o_p(n^{-1/2})$. Together with (2.10) this completes the proof. □

The result of Theorem 2.1 shows that $\hat{x}_n$ is asymptotically equivalent, up to a term of order $o_p(n^{-1/2})$, to the optimal solution $\bar{x}(Z_n)$ of the program (P_{Z_n}) with

$Z_n = \nabla\delta_n(x_0)$. Suppose now that $\bar{x}(v)$ is directionally differentiable at $v = 0$ in the sense of Fréchet; that is,

$$\bar{x}(v) - x_0 = D\bar{x}(v) + o(\|v\|),$$

where $D\bar{x}$: $\Re^m \to \Re^m$ is a positively homogeneous mapping. We then obtain that, under assumptions of Theorem 2.1,

$$\bar{x}(Z_n) - x_0 = D\bar{x}(Z_n) + o_p(n^{-1/2})$$

and hence

$$n^{1/2}(\hat{x}_n - x_0) = D\bar{x}(n^{1/2}Z_n) + o_p(1). \tag{2.11}$$

If we assume further that $n^{1/2}Z_n$ converges in distribution to a random vector Z and that the directional derivative $D\bar{x}(\cdot)$ is continuous, we obtain from (2.11) that $n^{1/2}(\hat{x}_n - x_0)$ converges in distribution to $D\bar{x}(Z)$. In particular, if Z has a multivariate normal distribution and $D\bar{x}(\cdot)$ is linear, then $n^{1/2}(\hat{x}_n - x_0)$ is asymptotically normal.

The program (P_v) can be considered as a parametrized optimization problem depending on the parameter vector v. Directional differentiability properties of optimal solutions of such parametrized programs have been studied in a number of recent publications (Auslender and Cominetti 1990, Gauvin and Janin 1988, King and Rockafellar 1992, Shapiro 1985, Shapiro 1988). The calculated directional derivatives can be employed, via (2.11), in the straightforward way to write explicitly the asymptotic distribution of $\hat{x}_n$ (cf. King and Rockafellar 1993, Shapiro 1989, Shapiro 1990). Expansion (2.7) goes beyond this and can provide information about asymptotic behavior of $\hat{x}_n$ even if $\bar{x}(v)$ is not directionally differentiable. For example, the function $f(x)$ need not be even to be differentiable at x_0.

3. Generalized equations approach. In this section we study constrained problems (P_0) and $(\hat{P}_n)$ formulated in §1. We assume that the functions $f(x)$ and $g(x)$ are continuously differentiable and that the set S is convex. Assuming further that a constraint qualification is satisfied, first-order necessary conditions for the program (P_0) can be written in the form

$$\begin{gathered} 0 \in \nabla_x L(x, y) + N_S(x), \\ g(x) \in K, \qquad y \in K^-,\ y \cdot g(x) = 0, \end{gathered} \tag{3.1}$$

where y is a vector of Lagrange multipliers,

$$L(x, y) = f(x) + y \cdot g(x)$$

is the Lagrangian function,

$$K^- = \{y: y \cdot v \leqslant 0, \text{ for all } v \in K\}$$

is the polar (negative dual) cone of the cone K and $N_S(x)$ is the normal cone to S at x (Maurer and Zowe 1979, Robinson 1976). Following Robinson 1982, we formulate these optimality conditions in a form of generalized equations. That is, we consider generalized equations (P_0^*) given by

$$(P_0^*) \qquad 0 \in G(z) + T(z),$$

where $z = (x, y) \in \Re^m \times \Re^k$,

$$T(z) = N_{S\times K^-}(x, y) = N_S(x) \times N_{K^-}(y) \tag{3.2}$$

and

$$G(z) = \begin{bmatrix} \nabla_x L(x, y) \\ -g(x) \end{bmatrix}. \tag{3.3}$$

Let us consider now stochastic programs $(\hat{P}_n)$. It will be assumed that $\psi_n(x)$ and $\gamma_n(x)$ are locally Lipschitz continuous. The generalized equations associated with $(\hat{P}_n)$ can be written then in the form

$$(\hat{P}_n^*) \qquad 0 \in \Gamma_n(z) + \Upsilon_n + T(z),$$

where

$$\Gamma_n(z) = \begin{bmatrix} \nabla_x \mathscr{L}_n(x, y) \\ -\gamma_n(x) \end{bmatrix} \tag{3.4}$$

and

$$\mathscr{L}_n(x, y) = \psi_n(x) + y \cdot \gamma_n(x).$$

At points where $\mathscr{L}_n(\cdot, y)$ is not differentiable, the gradient $\nabla_x \mathscr{L}_n(x, y)$ is replaced by an element of the generalized gradient $\partial_x \mathscr{L}_n(x, y)$ in the sense of Clarke 1983.

The term Υ_n represents a possible error in the solution of the generalized equations associated with approximating programs $(\hat{P}_n)$. It will be assumed that Υ_n tends in probability to zero at a rate which will be specified later.

We investigate asymptotic properties of solutions $\hat{z}_n = (\hat{x}_n, \hat{y}_n)$ of the generalized equations $(\hat{P}_n^*)$ by employing the particular structure of the mappings $G(z)$, $\Gamma_n(z)$ and the multifunction $T(z)$. Asymptotic behavior of solutions of stochastic generalized equations was studied recently by King and Rockafellar 1993. Our approach here is quite different. In particular it allows us to derive asymptotic results even in situations where the mappings Γ_n are discontinuous.

It will be assumed that the solutions $\hat{z}_n = \hat{z}_n(\omega)$ do exist with probability one and that $\hat{z}_n(\omega)$ are *measurable*. Let us make the following observations about properties of the generalized equations (P_0^*) and $(\hat{P}_n^*)$ which are implied by the particular structure of the involved mappings. Consider the mapping $G(z)$ given in (3.3). Suppose that $f(x)$ and $g(x)$ are twice continuously differentiable. We have then that

$$\nabla G(z) = \begin{bmatrix} \nabla_{xx}^2 L(x, y) & \nabla g(x) \\ -\nabla g(x)^T & 0 \end{bmatrix}.$$

Consequently for $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$,

$$\begin{aligned} &[G(z_1) - G(z_2)] \cdot (z_1 - z_2) \\ &= \int_0^1 (z_1 - z_2) \cdot \nabla G(z_2 + t(z_1 - z_2))(z_1 - z_2)\, dt \\ &= \int_0^1 (x_1 - x_2) \cdot \nabla_{xx}^2 L(z_2 + t(z_1 - z_2))(x_1 - x_2)\, dt. \end{aligned} \tag{3.5}$$

Suppose further that (P_0^*) has a unique solution $z_0 = (x_0, y_0)$ and that the Hessian matrix $\nabla^2_{xx} L(x_0, y_0)$ is positive definite. It follows then by the arguments of continuity that there is a neighborhood W of z_0 and a positive constant α such that the smallest eigenvalue of $\nabla^2_{xx} L(x, y)$ is greater than α for all $z = (x, y) \in W$. Together with (3.5) this implies that

$$[G(z_1) - G(z_2)] \cdot (z_1 - z_2) \geqslant \alpha \|x_1 - x_2\|^2 \tag{3.6}$$

for any $z_1, z_2 \in W$. This motivates us to make the following assumptions.

Assumption A*1. *The generalized equation* (P_0^*) *has a unique solution* $z_0 = (x_0, y_0)$.

Assumption A*2. *There exist a neighborhood* W *of* z_0 *and a constant* $\alpha > 0$ *such that the inequality* (3.6) *holds for any* $z_1, z_2 \in W$.

We also assume monotonicity of the multifunction $T(z)$. Of course, $T(z)$ is monotone if it is given in the form (3.2).

Assumption A*3. *The multifunction* $T(z)$ *is monotone, i.e., for all* z_1, z_2 *and* $v_1 \in T(z_1)$, $v_2 \in T(z_2)$, *one has*

$$(z_1 - z_2) \cdot (v_1 - v_2) \geqslant 0.$$

Consider now the parametric family of generalized equations

$$0 \in G(z) + \xi + T(z) \tag{Q_ξ}$$

depending on the parameter vector $\xi = (\xi_1, \xi_2)$, $\xi_1 \in \Re^m$, $\xi_2 \in \Re^k$. We make the following assumption about (Q_ξ).

Assumption A*4. *For every* ξ *in a neighborhood of zero,* (Q_ξ) *has a solution* $\bar{z}(\xi)$ *such that*

$$\|\bar{z}(\xi) - z_0\| = O(\|\xi\|). \tag{3.7}$$

In the case $T(z)$ and $G(z)$ are given in (3.2) and (3.3), respectively, generalized equations (Q_ξ) represent first order necessary conditions corresponding to the parametrized program

$$(P_\xi) \qquad \begin{aligned} &\text{minimize} && f(x) + \xi_1 \cdot x, \\ &\text{subject to} && g(x) - \xi_2 \in K, \\ &&& x \in S. \end{aligned}$$

(Compare with the parametrized program (P_v) introduced in §2.) In this case existence of $\bar{z}(\xi)$ is implied by the standard arguments of compactness and first order necessary conditions, which hold under various constraint qualifications. For example, suppose for the sake of simplicity that $S = \Re^m$ and assume that the point x_0 is a *regular* point of $g(x)$, with respect to the cone $K_0 = \{v \in K\colon y_0 \cdot v = 0\}$, in the sense of Robinson 1976. That is,

$$0 \in \operatorname{int}\{g(x_0) + \nabla g(x_0)\Re^m - K_0\}. \tag{3.8}$$

Notice that condition (3.8) implies uniqueness of the Lagrange multiplier vector y_0 and that (3.8) is a necessary and sufficient condition for uniqueness of y_0 if the cone

K is polyhedral (Shapiro 1992, Lemma 4.3). Under the constraint qualification (3.8) and Assumption A*2, condition (3.7) holds (Shapiro and Bonnans 1992).

The solution z_0 of (P_0^*) is called *semi-stable* in Bonnans 1990, if for any solution $\bar{z}(\xi)$ (if it exists) of (Q_ξ), condition (3.7) holds. Various characterizations of semi-stability in the case of polyhedral cone K are given in Bonnans 1990. In particular it is shown there that for optimization problems with finite number of constraints, semi-stability is equivalent to standard second-order sufficient conditions. We also remark that semi-stability is weaker than the strong regularity condition of Robinson 1980.

We introduce now some required "stochastic" assumptions. Consider the random mappings

$$\Delta_n(z) = \Gamma_n(z) - G(z). \tag{3.9}$$

ASSUMPTION B*1.

$$\|\Delta_n(z_0)\| = O_p(n^{-1/2}). \tag{3.10}$$

ASSUMPTION B*2. *There is a neighborhood W of z_0 such that*

$$\sup_{z \in W} \frac{\|\Delta_n(z) - \Delta_n(z_0)\|}{n^{-1/2} + \|z - z_0\|} = o_p(1). \tag{3.11}$$

Note that Assumptions B*1 and B*2 are parallel to Assumptions B2 and B3, respectively, which were introduced in §2.

Finally we shall need the following assumption.

ASSUMPTION B*3.

$$\|\hat{y}_n - y_0\| = O_p\big(\|\hat{x}_n - x_0\| + n^{-1/2}\big). \tag{3.12}$$

This assumption requires an explanation. Suppose for the sake of simplicity that $S = \Re^m$, $\Upsilon_n = 0$ and that condition (3.8) holds. It follows then (see Shapiro 1992, Lemma 4.4) that there is a positive constant β such that

$$\begin{aligned} \|\hat{y}_n - y_0\| \leqslant \beta\big(1 + \|\hat{y}_n\|\big)\max\big\{&\|\nabla\psi_n(\hat{x}_n) - \nabla f(x_0)\|, \\ &\|\nabla\gamma_n(\hat{x}_n) - \nabla g(x_0)\|, \|\gamma_n(\hat{x}_n) - g(x_0)\|\big\} \end{aligned} \tag{3.13}$$

for all $\nabla\psi_n(\hat{x}_n)$, $\nabla\gamma_n(\hat{x}_n)$ and $\gamma_n(\hat{x}_n)$ sufficiently close to $\nabla f(x_0)$, $\nabla g(x_0)$ and $g(x_0)$, respectively. Suppose further that Assumptions B2 and B3 of §2 hold and that $\nabla f(x)$ is locally Lipschitz. We have

$$\|\nabla\psi_n(\hat{x}_n) - \nabla f(x_0)\| \leqslant \|\nabla\delta_n(x_0)\| + \|\nabla\delta_n(\hat{x}_n) - \nabla\delta(x_0)\| + \|\nabla f(\hat{x}_n) - \nabla f(x_0)\|$$

and hence the first term inside the maximum in the right-hand side of (3.13) is of stochastic order $O_p(\|\hat{x}_n - x_0\| + n^{-1/2})$. Under similar assumptions for $\gamma_n(x)$ and $\nabla\gamma_n(x)$ we have that two other terms inside the maximum in (3.13) are also of order $O_p(\|\hat{x}_n - x_0\| + n^{-1/2})$. Moreover, it follows from the constraint qualification (3.8) that Lagrange multipliers $\hat{y}_n$ are uniformly bounded for all $\nabla\psi_n(\hat{x}_n)$, $\nabla\gamma_n(\hat{x}_n)$ and $\gamma_n(\hat{x}_n)$ sufficiently close to $\nabla f(x_0)$, $\nabla g(x_0)$ and $g(x_0)$, respectively. Therefore assuming further that $\hat{x}_n$ converges in probability to x_0, we obtain that (3.12) follows from (3.13).

We can formulate now the main result of this section.

THEOREM 3.1. *Suppose that Assumptions* A*1–A*4, B*1–B*3 *hold, that* $\Upsilon_n = o_p(n^{-1/2})$ *and that* $\hat{x}_n$ *converges in probability to* x_0. *Then*

$$\hat{x}_n = \bar{x}_n + o_p(n^{-1/2}), \tag{3.14}$$

where $(\bar{x}_n, \bar{y}_n) = \bar{z}_n = \bar{z}(\zeta_n)$ *is the solution of the generalized equation* (Q_{ζ_n}) *with* $\zeta_n = \Delta_n(z_0)$.

PROOF. Let us show first that

$$\|\hat{z}_n - z_0\| = O_p(n^{-1/2}). \tag{3.15}$$

We have that

$$-G(z_0) \in T(z_0) \quad \text{and} \quad -G(\hat{z}_n) - \Delta_n(\hat{z}_n) - \Upsilon_n \in T(\hat{z}_n).$$

Since $T(z)$ is monotone it follows then that

$$[G(\hat{z}_n) + \Delta_n(\hat{z}_n) + \Upsilon_n - G(z_0)] \cdot (\hat{z}_n - z_0) \leqslant 0$$

and hence

$$[G(\hat{z}_n) - G(z_0)] \cdot (\hat{z}_n - z_0) \leqslant -[\Delta_n(\hat{z}_n) + \Upsilon_n] \cdot (\hat{z}_n - z_0). \tag{3.16}$$

Since $\hat{x}_n$ tends in probability to x_0 and because of Assumption B*3 we have that $\|\hat{z}_n - z_0\| = o_p(1)$. Therefore with probability tending to one as $n \to \infty$, $\hat{z}_n \in W$, and hence by Assumption A*2,

$$[G(\hat{z}_n) - G(z_0)] \cdot (\hat{z}_n - z_0) \geqslant \alpha\|\hat{x}_n - x_0\|^2. \tag{3.17}$$

It follows from (3.16) and (3.17) that

$$\begin{aligned}\alpha\|\hat{x}_n - x_0\|^2 &\leqslant \|\Delta_n(\hat{z}_n) + \Upsilon_n\|\|\hat{z}_n - z_0\| \\ &\leqslant (\|\Delta_n(z_0)\| + \|\Delta_n(\hat{z}_n) - \Delta_n(z_0)\| + \|\Upsilon_n\|)\|\hat{z}_n - z_0\|.\end{aligned}$$

Together with Assumptions B*1, B*2 and since $\Upsilon_n = o_p(n^{-1/2})$, this implies

$$\alpha\|\hat{x}_n - x_0\|^2 \leqslant \{O_p(n^{-1/2}) + o_p(\|\hat{z}_n - z_0\|)\}\|\hat{z}_n - z_0\|.$$

Moreover, because of Assumption B*3 we then have that

$$\alpha\|\hat{x}_n - x_0\|^2 \leqslant \{O_p(n^{-1/2}) + o_p(\|\hat{x}_n - x_0\|)\}O_p(\|\hat{x}_n - x_0\| + n^{-1/2}).$$

Consequently

$$[\alpha + o_p(1)]\|\hat{x}_n - x_0\|^2 \leqslant O_p(n^{-1/2})\|\hat{x}_n - x_0\| + O_p(n^{-1})$$

which implies that

$$\|\hat{x}_n - x_0\| = O_p(n^{-1/2}).$$

By Assumption B*3, (3.15) then follows.

Now again because of the monotonicity of $T(z)$, we have

$$[G(\hat{z}_n) + \Delta_n(\hat{z}_n) + \Upsilon_n - G(\bar{z}_n) - \Delta_n(z_0)] \cdot (\hat{z}_n - \bar{z}_n) \leqslant 0$$

and hence

$$[G(\hat{z}_n) - G(\bar{z}_n)] \cdot (\hat{z}_n - \bar{z}_n) \leqslant -[\Delta_n(\hat{z}_n) + \Upsilon_n - \Delta_n(z_0)] \cdot (\hat{z}_n - \bar{z}_n).$$

Together with Assumption A*2, this implies

$$\alpha\|\hat{x}_n - \bar{x}_n\|^2 \leqslant \|\Delta_n(\hat{z}_n) + \Upsilon_n - \Delta_n(z_0)\|\|\hat{z}_n - \bar{z}_n\|. \tag{3.18}$$

Notice that it follows from (3.7) and (3.10) that $\|\bar{z}_n - z_0\|$ is of order $O_p(n^{-1/2})$ and hence, because of (3.15),

$$\|\hat{z}_n - \bar{z}_n\| = O_p(n^{-1/2}). \tag{3.19}$$

It follows from (3.18), (3.19), Assumption B*2 and $\Upsilon_n = o_p(n^{1/2})$ that

$$\alpha\|\hat{x}_n - \bar{x}_n\|^2 \leqslant [o_p(n^{-1/2}) + o_p(\|\hat{z}_n - z_0\|)]O_p(n^{-1/2})$$

which together with (3.15) implies (3.14). □

The result of Theorem 3.1 means that, under the specified assumptions, the component $\hat{x}_n$ of the solution $\hat{z}_n$ is asymptotically equivalent to the corresponding component $\bar{x}_n$ of the solution $\bar{z}_n$ of

$$0 \in G(z) + \Delta_n(z_0) + T(z). \tag{3.20}$$

This indicates that, asymptotically, all relevant "stochastic" information here is contained in the random vectors $\zeta_n = \Delta_n(z_0)$ while mappings $G(z)$ and $T(z)$ determine "optimization" properties of the considered problems. In the case of the sample mean construction it follows from the Central Limit Theorem that, under standard regularity conditions, $n^{1/2}\zeta_n$ converges in distribution to a multivariate normal.

It follows from Assumptions A*2 and A*3 that the component $\bar{x}(\xi)$ of the solution $\bar{z}(\xi)$ of (Q_ξ) is locally unique, and, under a constraint qualification, coincides with the corresponding optimal solution of (P_ξ). (Notice that this does not mean that the generalized equations $(\hat{P}_n^*)$ have locally unique solutions.) In case the set S and the cone K are polyhedral and the mapping $G(z)$ is continuously differentiable, it is possible to calculate the directional derivatives of $\bar{x}(\xi)$ at $\xi = 0$. This can be done either by an application of the Implicit Function Theorem due to Robinson 1985 to the generalized equations (Q_ξ) or by applying the corresponding results from sensitivity analysis of parametrized programs (Shapiro 1985, Theorem 4.2). These directional derivatives together with the asymptotics of $n^{1/2}\zeta_n$ provide a closed form expression for the asymptotic distribution of $n^{1/2}(\hat{x}_n - x_0)$ (cf. Shapiro 1990).

Finally we remark that our assumptions do not guarantee uniqueness of the component $\bar{y}(\xi)$ as ξ approaches 0. Therefore the asymptotic equivalence (3.14) is only proved for the components $\hat{x}_n$ and $\bar{x}_n$ of the corresponding solutions $\hat{z}_n$ and $\bar{z}_n$. Verification of the asymptotic equivalence between $\hat{y}_n$ and $\bar{y}_n$ requires introduction of further regularity conditions as, for example, strong regularity of z_0 in the sense of Robinson 1980.

4. A discussion of the regularity conditions. In this section we consider the sample mean construction and discuss applicability of the regularity conditions specified in §§2 and 3. In particular we discuss two-stage stochastic programming with fixed recourse. Since regularity assumptions of §3 are similar to the respective assumptions of §2, we mainly discuss the assumptions of §2.

In the case of stochastic programming with fixed recourse, first and second order differentiability properties of the corresponding expected value function $f(x)$ were studied in Kall 1976, Wets 1974 and Römish and Schultz 1989, Wang 1985, respectively. We concentrate here on the "stochastic" assumptions and especially on Assumption B3 (Assumption B*2) which was crucial in deriving the asymptotic result of Theorem 2.1 (Theorem 3.1). (Notice that in the unconstrained case, Assumption B*2 reduces to Assumption B3.)

Unless stated otherwise, all probabilistic statements of this section are made with respect to the probability distribution F of the considered sample and which appears in the calculation of the expected value function $f(x) = \int h(x, \xi)F(d\xi)$. All gradients will be taken with respect to x. Suppose that for almost every ξ the function $h(\cdot, \xi)$ is twice continuously differentiable in a neighborhood W of x_0 and that the families $\{\|\nabla^r h(x, \cdot)\|: x \in W\}$, $r = 1, 2$, are dominated by integrable functions. Then it follows from the Lebesgue dominated convergence theorem that $f(x)$ is twice continuously differentiable on W and that the first and second order derivatives of $f(x)$ can be taken inside the expected value. Furthermore, it follows from the Law of Large Numbers that with probability one, $\nabla^2\psi_n(x)$ tends to $\nabla^2 f(x)$ *uniformly* on any compact subset of W (e.g., [LeCam 1953, Corollary 4.1]). Assuming that W is convex and compact we obtain by the Mean Value Theorem that for any $x \in W$,

$$\|\nabla\delta_n(x) - \nabla\delta_n(x_0)\| \leqslant K_n\|x - x_0\|$$

with

$$K_n = \sup_{x \in W} \|\nabla^2\delta_n(x)\| = o_p(1).$$

Therefore in the above differentiable case, Assumption B3 holds and the term $n^{-1/2}$ in the denominator of the ratio in (2.5) can be omitted.

Let us consider now the following "nondifferentiable" example. Suppose that $S = \mathfrak{R}$ and let $h(x, \xi) = |x - \xi|$, $x, \xi \in \mathfrak{R}$. This example corresponds to (one-dimensional) stochastic programming with simple recourse. Let $F(t)$ be the cumulative probability distribution function of the considered sample and suppose that there exists a unique x_0 such that $F(x_0) = 1/2$. The number x_0 is called the median of F and is given by the minimizer of

$$f(x) = \int |x - t|\, dF(t)$$

over $\mathfrak{R}$. The corresponding minimizer $\hat{x}_n$ of

$$\psi_n(x) = n^{-1} \sum_{i=1}^{n} |x - \xi_i|,$$

is called the sample median. That is, if the sample is arranged in the increasing order $\xi_{(1)} \leqslant \xi_{(2)} \leqslant \cdots \leqslant \xi_{(n)}$ (order statistics), then $\hat{x}_n = \xi_{((n+1)/2)}$ for odd n, and $\hat{x}_n$ can be any number from the interval $[\xi_{(n/2)}, \xi_{(n/2+1)}]$ for even n. Note that for even n, $\hat{x}_n$ is not defined uniquely unless $\xi_{(n/2)} = \xi_{(n/2+1)}$, and for odd n the function $\psi_n(x)$ is not differentiable at $\hat{x}_n$. Note also that if we formulate the associated optimality conditions in the form of equations, the corresponding function $\Gamma_n(x) = \psi_n'(x)$ will be discontinuous at the points $\xi_1, \ldots, \xi_n$. In particular, $\Gamma_n(x)$ will be discontinuous at $\hat{x}_n$ for odd n.

We have here that the difference function $\delta_n(x)$ is differentiable at x and

$$\delta_n'(x) = 2F_n(x) - 2F(x)$$

if the distribution function F and the empirical distribution function F_n are continuous at x. Suppose for a moment that the considered distribution is uniform $(0, 1)$. Then $x_0 = 1/2$ and since (with probability one) F_n is constant in a sufficiently small neighborhood of x_0, we have that

$$\frac{|\delta_n'(x) - \delta_n'(x_0)|}{|x - x_0|} = 2$$

for all x sufficiently close to x_0. Therefore the supremum of $|\delta_n'(x) - \delta_n'(x_0)|/|x - x_0|$ over any neighborhood of x_0 is not less than 2. This demonstrates that the term $n^{-1/2}$ in the denominator of the ratio in (2.5) is essential here. It can be shown that Assumption B3 holds for the uniform distribution and the case of general F can be reduced to the uniform case by the transformation $y = F(x)$ provided F is Lipschitz continuous in a neighborhood of x_0.

In the multivariate case, Assumption B3 was studied by Huber 1967. Let us state some required regularity conditions.

Assumption C1. *There is an integrable function $K(\xi)$ such that*

$$|h(x, \xi) - h(y, \xi)| \leqslant K(\xi)\|x - y\| \tag{4.1}$$

for all x and y in a neighborhood x_0.

Assumption C2. *For each fixed x in a neighborhood of x_0, $h(\cdot, \xi)$ is differentiable at x for almost every ξ.*

Assumption C3. *There are strictly positive numbers c_1, c_2 and d_0 such that for every $d \geqslant 0$ and all x satisfying $\|x - x_0\| + d \leqslant d_0$,*

$$E\{u(x, \xi, d)\} \leqslant c_1 d, \tag{4.2}$$

$$E\{u(x, \xi, d)^2\} \leqslant c_2 d, \tag{4.3}$$

where

$$u(x, \xi, d) = \sup_{\|y - x\| \leqslant d} \|\nabla h(y, \xi) - \nabla h(x, \xi)\|.$$

Assumptions C1 and C2 imply that the expected value function $f(x)$ is differentiable in a neighborhood of x_0 and that the first-order derivatives of $f(x)$ can be taken inside the expected value. Furthermore, it follows from (4.2) that the gradient mapping $\nabla f(x)$ is Lipschitz continuous in a neighborhood of x_0. It is shown in (Huber 1967, Lemma 3) that Assumptions C1–C3 imply Assumption B3.

In fact, instead of the term $\|x - x_0\|$ which appears in the denominator of (2.5), the term $\|\nabla f(x) - \nabla f(x_0)\|$ was used (Huber 1967). However, since $\nabla f(x)$ is Lipschitz continuous near x_0, the condition (2.5) will follow. Note also that it is assumed in (Huber 1967, Assumption (N-3)(i)) that

$$\|\nabla f(x) - \nabla f(x_0)\| \geqslant \alpha\|x - x_0\| \tag{4.4}$$

for some $\alpha > 0$ and all x near x_0. We always can add the quadratic function $q(x) = \alpha(x - x_0) \cdot (x - x_0)$, independent of ξ, to $h(x, \xi)$ without affecting Assumptions C1–C3 and B3. Thus, by choosing a sufficiently large constant α in $q(x)$, we can always ensure condition (4.4). Therefore this condition is unnecessary in the case in which the term $\|\nabla f(x) - \nabla f(x_0)\|$ is replaced by the term $\|x - x_0\|$.

ASSUMPTION C4. *The expectation $E\{\|\nabla h(x_0, \xi)\|^2\}$ is finite.*

By the Central Limit Theorem it follows from Assumption C4 that $n^{1/2}\, \nabla\delta_n(x_0)$ converges in distribution to a multivariate normal, and hence, Assumption B2 follows. We obtain that Assumptions C1–C4 imply Assumptions B1–B3.

Let us finally discuss applicability of Assumptions C1–C4, and especially Assumption C3, in the case of stochastic programming with fixed recourse. That is, consider

$$h(x, \xi) = c \cdot x + Q(x, \xi), \tag{4.5}$$

where

$$Q(x, \xi) = \inf\{q \cdot y \colon Wy = b - Ax,\ y \geqslant 0\} \tag{4.6}$$

and

$$\xi = \xi(\omega) = (b(\omega), A(\omega), q(\omega)).$$

Suppose that for almost every q, the set

$$\Xi(q) = \{\zeta \colon W^T\zeta \leqslant q\} \tag{4.7}$$

is nonempty and bounded. By the arguments of duality (cf. Wets 1974) we can represent $Q(x, \xi)$ in the form

$$Q(x, \xi) = \sup\{(b - Ax) \cdot \zeta \colon \zeta \in \Xi(q)\}. \tag{4.8}$$

It is not difficult then to show that the inequality (4.1) holds with

$$K(\xi) = \|c\| + \|A\| \sup_{\zeta \in \Xi(q)} \|\zeta\|.$$

Assuming, for example, that for almost all q the sets $\Xi(q)$ are uniformly bounded and that $\|A(\omega)\|$ has finite expectation, we obtain that C1 follows.

It also follows from (4.8) that for every ξ, the function $Q(\cdot, \xi)$ is piecewise linear and convex. Its subdifferential is given by

$$\partial Q(x, \xi) = -A\Xi^*(x, \xi),$$

where

$$\Xi^*(x, \xi) = \operatorname{argmax}\{(b - Ax) \cdot \zeta \colon \zeta \in \Xi(q)\}.$$

Consequently $Q(\cdot, \xi)$ is differentiable at x if and only if $\Xi^*(x, \xi)$ is a singleton. Notice that for given Ax and q, $\Xi^*(x, \xi)$ is a singleton for almost every b with respect to the Lebesgue measure. Therefore Assumption C2 will follow if we assume that b is distributed independently of A and q and that the probability distribution of b is absolutely continuous with respect to the Lebesgue measure (i.e., b has a density function).

Let us finally discuss Assumption C3. Consider the set $T(q)$ of such points t that the linear function $\ell(\zeta) = t \cdot \zeta$ has more than one maximizer over $\Xi(q)$. For $t = b - Ax$ this set corresponds to points x where $\Xi^*(x, \xi)$ is not a singleton and hence $Q(x, \xi)$ is not differentiable. Notice that since $\Xi(q)$ is a convex polytope, the set $T(q)$ is formed by a finite number of affine subspaces of the linear space of vectors t. Assume that for almost all q the sets $\Xi(q)$ are uniformly bounded, say by a constant K. We have then that the supremum

$$u(x, \xi, d) = \sup_{\|y-x\| \leqslant d} \|\nabla Q(y, \xi) - \nabla Q(x, \xi)\|$$

is bounded, by $2\|A\|K$, and is nonzero only if the distance from $b - Ax$ to the set $T(q)$ is less than or equal to $\|A\|d$. Therefore we obtain that the inequalities (4.2) and (4.3) hold if b is independent of A and q, $\|A\|^2$ has finite expectation and the following condition is satisfied. For any affine subspace Φ of the linear space of vectors b,

$$\text{(4.9)} \qquad \operatorname{Prob}\{\operatorname{dist}(b(\omega), \Phi) \leqslant d\} = O(d).$$

In turn, condition (4.9) follows if for every nonsingular matrix M the (one-dimensional) marginal distributions of the components of the random vector $Mb(\omega)$ have locally Lipschitzian cumulative distribution functions (or, equivalently, have locally bounded densities). Such distributions were studied in Römisch and Schultz 1989, §2.

We see that, under mild regularity assumptions, the asymptotic results of §2 (of §3) are applicable although the functions of the stochastic programs $(\hat{P}_n)$ (of the stochastic generalized equations $(\hat{P}_n^*)$) are not everywhere differentiable (not everywhere continuous) here.

References

Auslender, A. and Cominetti, R. (1990). First and Second Order Sensitivity Analysis of Nonlinear Programs under Directional Constraint Qualification Conditions. *Optimization* **21** 351–363.

Bonnans, J. F. (1990). Rates of Convergence of Newton Type Methods for Variational Inequalities and Nonlinear Programming, *INRIA*, 78153 Rocquencourt, France (Preprint).

Clarke, F. H. (1983). *Optimization and Nonsmooth Analysis*. Wiley, New York.

Dupačová, J. and Wets, R. J. B. (1988). Asymptotic Behavior of Statistical Estimators and of Optimal Solutions of Stochastic Optimization Problems. *Ann. Statist.* **16** 1517–1549.

Gauvin, J. and Janin, R. (1988). Directional Behaviour of Optimal Solutions in Nonlinear Mathematical Programming. *Math. Oper. Res.* **13** 629–649.

Huber, P. J. (1967). The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. In *Proc. Fifth Berkeley Symp. on Math. Statist. and Probability*, Vol. 1, pp. 221–233, University of California Press, Berkeley, CA.

_____ (1981). *Robust Statistics*. Wiley, New York.

Kall, P. (1976). *Stochastic Linear Programming*. Springer, Berlin.

King, A. J. (1986). Asymptotic Behavior of Solutions in Stochastic Optimization: Nonsmooth Analysis and the Derivation of Nonnormal Limit Distributions. Ph.D. Dissertation, Dept. Applied Mathematics, University of Washington.

_____ and Rockafellar, R. T. (1993). Asymptotic Theory for Solutions in Statistical Estimation and Stochastic Programming. *Math. Oper. Res.* **18** 148–162.

_____ and _____ (1992). Sensitivity Analysis for Nonsmooth Generalized Equations. *Math. Programming* **55** 193–212.

Le Cam, L. (1953). On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes' Estimates. *Univ. California Publ. Statist.* **1** 277–330.

Mann, H. B. and Wald, A. (1943). On Stochastic Limit and Order Relationships. *Ann. Math. Statist.* **14** 217–226.

Maurer, H. and Zowe, J. (1979). First and Second-Order Necessary and Sufficient Optimality Conditions for Infinite-Dimensional Programming Problems. *Math. Programming* **16** 98–110.

Robinson, S. M. (1976). First Order Conditions for General Nonlinear Optimization. *SIAM J. Appl. Math.* **30** 597–607.

______ (1976). Stability Theorems for Systems of Inequalities, Part II: Differentiable Nonlinear Systems. *SIAM J. Numer. Anal.* **13** 497–513.

______ (1980). Strongly Regular Generalized Equations. *Math. Oper. Res.* **5** 43–62.

______ (1982). Generalized Equations and Their Solutions, Part II: Applications to Nonlinear Programming. *Math. Programming Stud.* **19** 200–221.

______ (1985). Implicit B-differentiability in Generalized Equations. Technical Report 2854, Mathematics Research Center, University of Wisconsin-Madison, Madison, WI.

Rockafellar, R. T. (1976). Integral Functionals, Normal Integrands and Measurable Selections. In *Nonlinear Operators and the Calculus of Variations.* Lecture Notes in Math. **543**, pp. 157–207, Springer, Berlin.

______ and Wets. R. J. B. (1984). *Variational Systems, an Introduction. Multifunctions and Integrands.* Lecture Notes in Math. **1091**, pp. 1–54, Springer, New York.

Römisch, W. and Schultz, R. (1989). Stability of Solutions for Stochastic Programs with Complete Recourse Having $C^{1,1}$ Data. Technical Report, Institut für Operations Research, Universität Zürich.

Shapiro, A. (1985). Second-order Sensitivity Analysis and Asymptotic Theory of Parametrized, Nonlinear Programs. *Math. Programming* **33** 280–299.

______ (1988). Sensitivity Analysis of Nonlinear Programs and Differentiability Properties of Metric Projections. *SIAM J. Control Optim.* **26** 628–645.

______ (1989). Asymptotic Properties of Statistical Estimators in Stochastic Programming. *Ann. Statist.* **17** 841–858.

______ (1990). On Differential Stability in Stochastic Programming. *Math. Programming* **47** 107–116.

______ (1991). Asymptotic Analysis of Stochastic Programs. *Ann. Oper. Res.* **30** 169–186.

______ (1992). Perturbation Analysis of Optimization Problems in Banach Spaces. *Numer. Funct. Anal. Optim.* **13** 97–116.

______ and Bonnans, J. F (1992). Sensitivity Analysis of Parametrized Programs under Cone Constraints. *SIAM J. Control Optim.* **30** 1409–1422.

Wang, J. (1985). Distribution Sensitivity Analysis for Stochastic Programs with Complete Recourse. *Math. Programming* **31** 286–297.

Wets, R. J. B. (1974). Stochastic Programs with Fixed Recourse: The Equivalent Deterministic Program. *SIAM Rev.* **16** 309–339.

A. Shapiro: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205