Working Paper

Asymptotic Theory for Solutions in Generalized M-Estimation and Stochastic Programming

Alan J. King† *and R. Tyrrell Rockafellar*‡

WP-90-76
December 1990

IIASA

International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria

Telephone: (0 22 36) 715 21 *0 □ Telex: 079 137 iiasa a □ Telefax: (0 22 36) 71313

Asymptotic Theory for Solutions in Generalized M-Estimation and Stochastic Programming

Alan J. King† *and R. Tyrrell Rockafellar*‡

WP-90-76
December 1990

†IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, U.S.A. Research supported in part by a grant from the National Science Foundation while with the International Institute for Applied Systems Analysis, Laxenburg, Austria.
‡Department of Mathematics, University of Washington, Seattle, WA 98195, U.S.A.

IIASA

International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria

Telephone: (0 22 36) 715 21 *0 □ Telex: 079 137 iiasa a □ Telefax: (0 22 36) 71313

Foreword

New techniques of local sensitivity analysis in nonsmooth optimization are applied to the problem of studying the asymptotic behavior (generally non-normal) for solutions in stochastic optimization, and generalized M-estimation – a reformulation of the traditional maximum likelihood problem that allows the introduction of hard constraints.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

1. Introduction

Many problem formulations in statistics and stochastic optimization generate estimates from data by selecting a "best" or "optimal" point $\mathbf{x}^\nu = x^\nu(\mathbf{s}_1, \ldots, \mathbf{s}_\nu)$, frequently by choosing $\mathbf{x}^\nu$ to solve a *generalized equation* in the form

$$\text{(1.1)} \qquad \text{Choose } x \in \mathbb{R}^n \text{ such that } 0 \in \tfrac{1}{\nu} \sum_{i=1}^{\nu} f(x, \mathbf{s}_i) + N(x),$$

where $g : \mathbb{R}^n \times S \to \mathbb{R}^n$ is a function that is continuous in the first argument and measurable in the second, $\{\mathbf{s}_i\}$ an i.i.d. sequence of random variables in a complete separable metric space S, and $N : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ a multifunction. In stochastic programming, for example, this equation can represent the first-order necessary conditions for the optimization problem

$$\text{(1.2)} \qquad \text{minimize } \tfrac{1}{\nu} \sum_{i=1}^{\nu} h(x, \mathbf{s}_i) \text{ over all } x \in X \subset \mathbb{R}^n,$$

with $f(x, s) = \nabla_x h(x, s)$, the gradient of h at x, and $N(x) = N_X(x)$, the normal cone to X at x in the sense of nonsmooth analysis. In maximum likelihood estimation, the generalized equation (1.1) can represent the so-called "normal equations" by setting $N(x)$ identically equal to the zero vector; this situation represents the case where no "hard" (i.e. *a priori* deterministic) constraints are placed on the maximum likelihood estimator. Introducing the multifunction N into the normal equations is natural for optimization, because it permits the specification of constraints that one knows must be true (e.g. nonnegativity in variance estimation). In this case, solutions to (1.1) could be called *generalized M-estimates.*

We shall study the asymptotics of the sequence of estimates $\{\mathbf{x}^\nu\}$ from the point of view of consistency and central limits. The presence of the multifunction N complicates the asymptotic analysis, but in ways that can be analyzed using the special techniques of this paper. The problem (1.1) can be viewed as a generalized equation in which the first term is perturbed in a neighborhood of the function $E\mathbf{f}(\cdot) := Ef(\cdot, \mathbf{s}_1)$, by replacing it with the approximation $E^\nu \mathbf{f}(\cdot) := \frac{1}{\nu} \sum_{i=1}^{\nu} f(\cdot, \mathbf{s}_i)$, or, equivalently, adding to it the term $E^\nu \mathbf{f} - E\mathbf{f}$. We shall view these functions as elements of the Banach space $\mathcal{C}_n(U)$, the space of $\mathbb{R}^n$-valued functions that are continuous on a (yet to be determined) compact neighborhood U in $\mathbb{R}^n$. The asymptotics of the solutions to (1.1) can then be derived from the asymptotics of the sequence $\{E^\nu \mathbf{f}\}$, as random elements of $\mathcal{C}_n(U)$, and from the sensitivity analysis of the *solution mapping* $J : \mathcal{C}_n(U) \rightrightarrows \mathbb{R}^n$ defined by

$$\text{(1.3)} \qquad J(g) = \{x \in \mathbb{R}^n \mid 0 \in g(x) + N(x)\}$$

as developed by us in [**8**]. Consistency will follow from a sort of local continuity of J called *subinvertibility*, and the central limits from a certain differentiability property of J, employing the generalized delta method of King [**7**].

The asymptotic distributions obtained for solutions to (1.1) will not in general be normally distributed, because the multifunction N may impose restrictions that will affect the support of the asymptotic distribution. In stochastic optimization, constraints are fundamental to modelling practical decision problems and asymptotic normality cannot be assured except under rather special circumstances. This feature requires us to venture outside of the usual route to proving asymptotic results in maximum likelihood estimation, which considers the solution mapping as a functional of the probability measure or distribution function—cf. Clarke [**3**], for example. Our approach has a similar abstract flavor, but by considering the solution mapping as a functional of $E^{\nu}\mathbf{f}$, the analysis is both simplified, because the perturbation is *additive*, and enriched, because we are able to draw upon useful results from nonsmooth analysis.

There has been much activity recently in proving asymptotic theorems for solutions to stochastic programs. An earlier version of the approach we follow here first appeared in King [**7**]. Recently, Dupačová and Wets [**4**] and Shapiro [**11**] have applied a theorem of Huber [**5**] to the problem of determining the central limit behavior of the solutions to (1.2); this technique employs standard finite-dimensional parametric analysis after making assumptions that ensure asymptotic normality of $E^{\nu}\mathbf{f}(\mathbf{x}^{\nu}) - E\mathbf{f}(x^*)$. While the conclusions of this approach are similar to ours at first glance, our assumptions are simpler, less restrictive in practice, and cover a wider class of stochastic programs.

Our study begins with the general theorems concerning consistency and central limits for the generalized M-estimates determined by the sequence of generalized equations (1.1). These results will then be specialized to asymptotic analysis for stochastic programs Pertinent details concerning the asymptotic normality of the sequence $\{E^{\nu}\mathbf{f}\}$ appear in an appendix.

Much of the fundamental material on which our presentation is based has been comprehensively treated in [**7**] and [**8**], which we shall consider as read. Nevertheless, we cannot resist repeating some of the more important ideas. We mention here the underlying topology on which our analysis is based: namely that of the convergence of closed sets in $\mathbb{R}^n$. Let $\{A_\nu\}$ be a sequence of closed subsets of $\mathbb{R}^n$ and define the (closed) sets

$$\liminf_{\nu} A_\nu = \{x = \lim x_\nu \mid x_\nu \in A_\nu \text{ for all but finitely many } \nu\}$$

$$\limsup_{\nu} A_\nu = \{x = \lim x_\nu \mid x_\nu \in A_\nu \text{ for infinitely many } \nu\},$$

then $\{A_\nu\}$ *set-converges to* $A = \lim_\nu A_\nu$, if $A = \liminf A_\nu = \limsup A_\nu$.

2. General Theory

We first investigate consistency of solution sequences to (1.1). There may be more than one cluster point for such a sequence, or there may be none. Though we are able to provide natural assumptions under which unique limits exist, we prefer to study a weaker form of consistency that is more in keeping with our view that these sequences are actually *selections* from the sequence of *random sets* $\{J(E^\nu \mathbf{f})\}$. We review for the convenience of the reader some basic definitions concerning multifunctions.

Let $(Z, \mathcal{A})$ be an arbitrary measurable space. A multifunction $F : Z \rightrightarrows \mathbb{R}^n$ is *measurable* if for all closed subsets C of $\mathbb{R}^n$ the set $F^{-1}(C) := \{z \in Z \,|\, F(z) \cap C \neq \emptyset\}$ belongs to the sigma-algebra $\mathcal{A}$. It is *closed-valued* (or convex, etc.) if F has closed (or convex, etc.) images. If the measurable space is a probability space, we shall refer to a closed-valued measurable multifunction F as a *random closed set* and denote it $\mathbf{F}$. The *domain* of the multifunction F, $\operatorname{dom} F$, is the set of points where its image is nonempty; its *graph* is the set of pairs $\operatorname{gph} F := \{(z, x) \in Z \times \mathbb{R}^n \,|\, x \in F(z)\}$. If Z is a topological space then we say that F is *closed* (or upper semicontinuous) if $\operatorname{gph} F$ is a closed subset of $Z \times \mathbb{R}^n$. It is well-known that a closed multifunction is closed-valued and measurable; cf. Rockafellar [**10**].

Proposition 2.1. *For any compact set U in $\mathbb{R}^n$, let $\mathcal{C}_n(U)$ be made into a measurable space by equipping it with its Borel subsets, and suppose that the multifunction $N : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is closed. Then the solution mapping $J : \mathcal{C}_n(U) \rightrightarrows \mathbb{R}^n$ defined by (1.3) is closed (and therefore, closed-valued and measurable).*

Proof. Let us consider a sequence of pairs $\{(f^\nu, x^\nu)\}$, each an element of $\operatorname{gph} J$, that converges to a pair (f^*, x^*) in $\mathcal{C}_n(U) \times \mathbb{R}^n$. By uniform convergence, $f^\nu(x^\nu) \to f^*(x^*)$. Since N is closed, it follows that $-f^*(x^*) \in N(x^*)$. This implies $x^* \in J(f^*)$, so J is closed. □

To analyze the existence of solutions to generalized equations such as (1.1), we introduced in [**8**] the following notion: a multifunction $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *subinvertible* at a point y^* in $\mathbb{R}^n$ if there are a compact, convex neighborhood V of y^* and a nonempty, compact, convex-valued multifunction $G : V \rightrightarrows \mathbb{R}^n$ such that G is closed, and $G(y) \subset F^{-1}(y)$ for all $y \in V$. The reader may easily verify that maximal monotone operators are subinvertible at every point in the relative interior of their domains, and that multifunctions admitting selections that are continuous on a neighborhood of a given point are also subinvertible there.

Theorem 2.2. *Suppose N is closed and the multifunction $E\mathbf{f} + N$ is subinvertible at 0. Let V be the compact set and G the multifunction as are guaranteed by the definition of the subinvertibility of $E\mathbf{f} + N$, let U be a compact set containing $\cup\{G(y) : y \in V\}$ in its interior, and suppose that*

$$E\{\sup_{x \in U} |f(x, \mathbf{s}_1)|\} < \infty. \tag{2.1}$$

Then, with probability one,

$$\emptyset \neq \limsup_{\nu \to \infty} J(E^\nu \mathbf{f}) \subset J(E\mathbf{f}).$$

Proof. Condition (2.1) implies by the strong law of large numbers that $E^\nu \mathbf{f} \to E\mathbf{f}$ in $\mathcal{C}_n(U)$, with probability one. In the event of such convergence, the subinvertibility of $E\mathbf{f} + N$ implies, by Lemma 2.1 of [**8**], that $U \cap J(E^\nu \mathbf{f})$ is eventually nonempty; this and the compactness of U prove that $\limsup J(E^\nu \mathbf{f}) \neq \emptyset$. Since N is closed, we know by Proposition 2.1 that J is closed, from which we obtain the inclusion $\limsup J(E^\nu \mathbf{f}) \subset J(E\mathbf{f})$. □

Corollary 2.3. *(Consistency.) Under the conditions of Theorem 2.2, if $\{\mathbf{x}^\nu\}$ is a sequence of solutions to (1.1) and if $\mathbf{x}$ is a cluster point of this sequence, then $0 \in E\mathbf{f}(\mathbf{x}) + N(\mathbf{x})$ with probability one.*

Remark. The corollary can be strengthened if there are natural conditions that imply (or if one does not mind imposing conditions that require) that solutions of (1.1) belong to some compact set. In this case, almost all solution sequences will have cluster points.

We next consider the possibility that there is a central limit theorem for $\{\mathbf{x}^\nu\}$, that is, the existence of a random vector $\mathbf{u}$ and a point x^* such that $\{\sqrt{\nu}(\mathbf{x}^\nu - x^*)\}$ converges in distribution to $\mathbf{u}$. This will follow from asymptotic normality of the $E^\nu \mathbf{f}$ and a certain differentiability property of the solution mapping J, which we now briefly review.

For this discussion only, let Z be a Banach space. The *contingent derivative* of a multivalued mapping $F : Z \rightrightarrows \mathbb{R}^n$ at a point $z \in \operatorname{dom} F$ and $x \in F(z)$ is the mapping $DF(z|x)$ whose graph is the *contingent cone* to the graph of F at $(z, x) \in Z \times \mathbb{R}^n$, i.e.

$$\limsup_{t \downarrow 0} t^{-1}[\operatorname{gph} F - (z, x)] = \operatorname{gph} DF(z|x). \tag{2.2}$$

The contingent derivative always exists, because the lim sup of a net of sets always exists; and it is closed because the lim sup is always a closed set. The contingent derivative of the inverse of F is just the inverse of the contingent derivative, and is denoted $DF^{-1}(x|z)$. This definition may be specialized in two directions. If one has lim sup = lim inf in (2.2), then F is called ***proto-differentiable*** at (z, x). A stronger property that is related to differentiability for functions is *semi-differentiability*, which requires that

$$\lim_{\substack{t\downarrow 0 \\ w' \to w}} (F(z + tw') - x)/t = DF(z|x)(w) \tag{2.3}$$

for all directions w in Z. These definitions can be applied to functions, of course. If $f : Z \to \mathbb{R}^n$ has a contingent derivative $Df(z)$, as defined by the graph limit (2.1), that is everywhere single-valued, then f is *B-differentiable* at z and formula (2.3) holds.

For convenient reference, we make a list of the principal assumptions that we shall impose on the function f, random variables $\{\mathbf{s}_i\}$, and multifunction N in order that the solutions to (1.1) obey a central limit theorem. We suppose that a given point x^* belongs to the set $J(E\mathbf{f})$.

Analytical Assumptions for Generalized M-Estimates

M.1 The function $f(\cdot, s) : \mathbb{R}^n \to \mathbb{R}^n$ is continuous for all $s \in S$, and $E\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is B-differentiable on dom N.

M.2 The operator $N : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is closed and proto-differentiable at $(x^*, -E\mathbf{f}(x^*))$.

M.3 The multifunction $E\mathbf{f} + N$ is subinvertible at 0.

Let the compact set U be as defined in Theorem 2.2.

Probabilistic Assumptions

P.1 For all $x \in U$, the function $f(x, \cdot) : S \to \mathbb{R}^n$ is measurable.

P.2 The sequence of random variables $\{\mathbf{s}_i\}$ is independent and identically distributed.

P.3 There is some $a : S \to \mathbb{R}$ with $E|a(\mathbf{s}_1)|^2 < \infty$ and

$$|f(x_1, s) - f(x_2, s)| \le a(s)|x_1 - x_2| \quad \forall x_1, x_2 \in U.$$

P.4 There is some $x \in U$ with $E|f(x, \mathbf{s}_1)|^2 < \infty$.

In the Appendix we show that assumptions P.1–4 imply that the functions $E^\nu \mathbf{f}$ are $\mathcal{C}_n(U)$-valued random variables that satisfy the central limit property

$$\sqrt{\nu}(E^\nu \mathbf{f} - E\mathbf{f}) \xrightarrow[\mathcal{D}]{} \mathbf{w}, \tag{2.4}$$

where $\mathbf{w}$ is a centered, Gaussian $\mathcal{C}_n(U)$-valued random variable with covariance equal to that of $f(\cdot, \mathbf{s}_1)$. (In particular, this means that $\mathbf{w}(x^*)$ is a centered, normally-distributed random vector in $\mathbb{R}^n$ with covariance equal to that of $f(x^*, \mathbf{s}_1)$.)

Theorem 2.4. *(Central Limits.) Suppose the assumptions M.1–3 and P.1–4 hold, and that the random closed set*

$$\{u \in \mathbb{R}^n \mid 0 \in \mathbf{w}(x^*) + D(E\mathbf{f})(x^*)(u) + DN(x^* \mid -E\mathbf{f}(x^*))(u)\} \tag{2.5}$$

is almost surely single-valued. If a sequence $\{\mathbf{x}^\nu\}$ of measurable selections from the solution sets to (1.1) converges almost surely, then it converges to the point x^*, and moreover,

$$\sqrt{\nu}(\mathbf{x}^\nu - x^*) \to \mathbf{u}, \tag{2.6}$$

where $\mathbf{u}$ is any selection from (2.5).

Proof. As shown in the Appendix, the probabilistic assumptions imply that $\sqrt{\nu}(E^\nu\mathbf{f} - E\mathbf{f})$ is asymptotically normal, and as in (2.4) we denote the asymptotic distribution by $\mathbf{w}$. The analytical assumptions M.1–3 and the almost sure single-valuedness of (2.5) imply that x^* is the unique element of $U \cap J(E\mathbf{f})$, that $U \cap J$ is upper Lipschitzian at $E\mathbf{f}$, and that J is semi-differentiable at $(E\mathbf{f}, x^*)$ with contingent derivative

$$DJ(E\mathbf{f}|x^*)(w) = \{u \in \mathbb{R}^n \mid 0 \in w(x^*) + D(E\mathbf{f})(x^*)(u) + DN(x^* \mid -E\mathbf{f}(x^*))(u)\};$$

cf. Theorem 4.1 and Remarks 4.2 and 4.3 of [**8**], noting that (2.5) is precisely $DJ(E\mathbf{f}|x^*)(\mathbf{w})$. Observe that

$$\sqrt{\nu}(\mathbf{x}^\nu - x^*) \in \sqrt{\nu}[J(E^\nu\mathbf{f}) - x^*].$$

The semi-differentiability of J implies by Theorem 3.2 of [**7**] that the sequence of sets on the right side converges in distribution to $DJ(E\mathbf{f}|x^*)(\mathbf{w})$. To obtain from this the convergence in distribution of the selections on the left side to a selection from $DJ(E\mathbf{f}|x^*)(\mathbf{w})$, we can apply Theorem 2.3 of [**7**], provided this sequence is *tight*. But we already know that

$$\sqrt{\nu}|\mathbf{x}^\nu - x^*| \le \lambda\sqrt{\nu}\|E^\nu\mathbf{f} - E\mathbf{f}\|,$$

where λ is the Lipschitz constant for J at $E\mathbf{f}$. Since $\sqrt{\nu}(E^\nu\mathbf{f} - E\mathbf{f})$ is asymptotically normal, it is *a fortiori* tight. This final detail completes the proof. □

3. Asymptotics for Stochastic Programs

We consider the asymptotic behavior of sequences of solutions to a slightly more general version of a stochastic program than mentioned in the introduction, namely

$$\begin{aligned} \text{minimize} \quad & E^\nu \mathbf{h}(x) \\ \text{subject to} \quad & E^\nu \mathbf{g}(x) \in Q^o \\ \text{and} \quad & x \in C, \end{aligned} \tag{3.1}$$

where the set C is a convex polyhedral subset of $\mathbb{R}^n$, the set Q^o is the polar of a convex polyhedral cone in $\mathbb{R}^m$, and for all $s \in S$ the functions $h(\cdot, s) : \mathbb{R}^n \to \mathbb{R}$ and $g(\cdot, s) : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. This form is a mathematically convenient generalization of the usual statement of a nonlinear program with equality and inequality constraints (which can be obtained by setting $Q = \mathbb{R}^{m_1} \times \mathbb{R}_+^{m_2}$); it was originally introduced and studied by Robinson [**9**]. The problems (3.1) are to be regarded as perturbations of the "true" problem

$$\begin{aligned} \text{minimize} \quad & E\mathbf{h}(x) \\ \text{subject to} \quad & E\mathbf{g}(x) \in Q^o \\ \text{and} \quad & x \in C, \end{aligned} \tag{3.2}$$

In [**8**] we provided a second-order sensitivity analysis of this type of nonlinear program. The results of this section are direct consequences of that analysis, together with our results from the preceeding section.

In nonlinear programming, the sensitivity analysis of solutions cannot be separated from the sensitivity analysis of the Lagrange multipliers for the constraints. Our study is no exception. Since in (3.1) we wish to cover the case of estimated constraints $E^\nu \mathbf{g}(x) \in Q^o$, we are forced to consider sequences of Kuhn-Tucker pairs $(\mathbf{x}^\nu, \mathbf{y}^\nu)$ for (3.1) and not only sequences of solutions.

Define the Lagrangian $k(x, y, s) = h(x, s) + y^T g(x, s)$, and let (x^*, y^*) be a Kuhn-Tucker pair for the problem (3.2), i.e. a solution to the Kuhn-Tucker equations

$$\begin{aligned} 0 \in & \nabla E\mathbf{h}(x) + y^T \nabla E\mathbf{g}(x) + N_C(x) \\ 0 \in & -E\mathbf{g}(x) + N_Q(x) \end{aligned} \tag{3.3}$$

By $N_C(x)$ and $T_C(x)$ we denote the normal and tangent cones to the set C at a point x.

Analytical Assumptions for Stochastic Programs

S.1 The Lagrangian $E\mathbf{k}(x,y)$ is twice continuously differentiable, and the *second-order sufficient condition* holds at (x^*,y^*):

$$u^T\nabla^2 E\mathbf{k}(x^*,y^*)u > 0$$

for every vector $u \in T_C(x^*)$ satisfying

$$\nabla E\mathbf{g}(x^*)u \in T_{Q^o}(E\mathbf{g}(x^*))$$

and $\nabla E\mathbf{h}(x^*) = 0$.

S.2 The constraint set $\{x \in C \mid E\mathbf{g}(x) \in Q^o\}$ is *regular* at x^*:

$$0 \in \operatorname{int}[E\mathbf{g}(x^*) + \nabla E\mathbf{g}(x^*)(C - x^*) - Q^o].$$

S.3 The *linear independence condition* holds at x^*, that is, the Jacobian matrix $\nabla E\mathbf{g}(x^*)$ has full rank.

The reader will recall that S.2 is the counterpart in this more general formulation of the Mangasarian-Fromowitz constraint qualification for nonlinear programs in the usual format. The linear independence assumption does not explicitly exclude inactive constraints as in the usual statement of this condition: we simply suppose these are dropped from the problem statement.

To correspond with the setting of the previous section, define the function $f : \mathbb{R}^{n+m} \times S \to \mathbb{R}^{n+m}$ by

$$f(x,y,s) = (\nabla k(x,y,s), -g(x,s)),$$

and note that the Kuhn-Tucker conditions (3.3) correspond to the generalized equation

$$0 \in E\mathbf{f}(x,y) + N_{C\times Q^o}(x,y).$$

Theorem 3.1. *(Consistency.) Suppose that condition (2.1) holds with the function f as above, and that the analytical assumptions S.1–2 hold. If $\{(\mathbf{x}^\nu,\mathbf{y}^\nu)\}$ is a sequence of Kuhn-Tucker pairs for (3.1) and $(\mathbf{x},\mathbf{y})$ is a cluster point of this sequence, then $(\mathbf{x},\mathbf{y})$ is a Kuhn-Tucker pair for (3.2) with probability one.*

Proof. Under the assumptions S.1–2, it was shown in [**8**], Example 2.2, that the multifunction $E\mathbf{f} + N_{C\times Q^o}$ is subinvertible at 0. Now apply Corollary 2.3. □

To obtain an expression for the central limit behavior, we saw in the previous section that it was necessary to consider an associated random generalized equation involving the derivatives of $E\mathbf{f} + N$ and the normal random vector $\mathbf{w}(x^*)$. For stochastic programs the corresponding object is a certain random quadratic program, which we now describe. If the probabilistic assumptions are satisfied for ∇k and g, then from the Appendix we deduce that there exist Gaussian random functions $\mathbf{w}_1$ and $\mathbf{w}_2$ such that

$$\sqrt{\nu}[E^\nu \mathbf{k} - E\mathbf{k}] \xrightarrow[\mathcal{D}]{} \mathbf{w}_1$$

and

$$\sqrt{\nu}[E^\nu \mathbf{g} - E\mathbf{g}] \xrightarrow[\mathcal{D}]{} \mathbf{w}_2$$

Let $\mathbf{c}_1 = \mathbf{w}_1(x^*, y^*)$ and $\mathbf{c}_2 = \mathbf{w}_2(x^*)$, and consider the random quadratic program

$$\begin{aligned} &\text{minimize} && \mathbf{c}_1 u + \tfrac{1}{2} u^T \nabla^2 E\mathbf{k}(x^*, y^*) u \\ &\text{subject to} && \nabla E\mathbf{g}(x^*) u + \mathbf{c}_2 \in [Q']^o \\ &\text{and} && x \in C' \end{aligned} \tag{3.4}$$

where

$$Q' = \{v \in T_Q(y^*) \,|\, v^T E\mathbf{g}(x^*) = 0\}$$

and

$$C' = \{u \in T_C(x^* \,|\, u^T \nabla E\mathbf{k}(x^*, y^*) = 0\}.$$

Theorem 3.2. *Suppose that the probabilistic assumptions P.1–4 are satisfied for ∇k and g and the analytical assumptions S.1–3 hold. If a sequence of Kuhn-Tucker pairs $\{(\mathbf{x}^\nu, \mathbf{y}^\nu)\}$ for the problems (3.1) converges almost surely, then it converges to (x^*, y^*), and moreover,*

$$\sqrt{\nu}[(\mathbf{x}^\nu, \mathbf{y}^\nu)) - (x^*, y^*)] \xrightarrow[\mathcal{D}]{} (\mathbf{u}, \mathbf{v}),$$

where $(\mathbf{u}, \mathbf{v})$ is the Kuhn-Tucker pair for the random quadratic program (3.4).

Proof. In [**8**], Example 6.3, we showed that assumptions S.1–3 imply our assumptions M.1–3 and also that the set (2.5) in our Theorem 2.4 is single-valued, for the corresponding function f as above and multifunction $N_{C\times Q}$. An application of Theorem 2.4 finishes the prof. □

Remark 3.3. Theorem 3.2 resembles standard results in maximum likelihood estimation, except in that we allow constraints to be placed on the estimators. Aitchison and Silvey [**1**] worked out the asymptotic distribution for equality constraints only, which turns out to be asymptotically normal. Their results may be easily derived from ours. Shapiro [**11**] treats asymptotics for stochastic programs by applying a theorem of Huber [**5**], but does not consider estimated constraints.

Remark 3.4. There are interesting parallels to be drawn between our result and those of Huber [**5**] in the unconstrained situation. Our probabilistic assumptions P.1–4 correspond roughly to Huber's assumptions N1, N3(ii) and (iii), and N4, and our monotonicity assumptions correspond practically to Huber's N2 and N3(i). They imply his condition that $\mathbf{x}^\nu \to x^*$ with probability one. Huber's goal is to prove that $\sqrt{\nu}(E\nabla f(x^\nu) - E\nabla f(x^*))$ has the same asymptotic distribution as $\sqrt{\nu}(E^\nu \nabla h(x^*) - E\nabla h(x^*))$; then he can derive the asymptotic distribution of $\sqrt{\nu}(x^\nu - x^*)$ via the classical delta method under the assumption that $E\nabla h(\cdot)$ is Frechét differentiable at x^* with invertible Jacobian H. We achieve the same result, namely that $\sqrt{\nu}(\mathbf{x}^\nu - x^*)$ is asymptotically normal with asymptotic distribution $H^{-1}\mathbf{w}(x^*)$, but under our slightly different assumptions. For a further discussion of asymptotic theory in stochastic programming from Huber's perspective, see Dupačová and Wets [**4**].

Appendix

In this appendix we briefly discuss central limit theory for random variables in $\mathcal{C}_n(U)$, the space of continuous $\mathbb{R}^n$-valued functions on a compact subset $U \subset \mathbb{R}^n$. Further details may be found in Araujo and Giné [**2**], on which this presentation has been based.

For now, let Z be a separable Banach space equipped with its Borel sets $\mathcal{A}$, and let Z^* be the dual space of continuous linear functionals on Z. If $\mathbf{z}$ is a random variable taking values in Z, we say that $\mathbf{z}$ is (Pettis) *integrable* if there is an element $E\mathbf{z} \in Z$ for which $\ell(E\mathbf{z}) = E\{\ell(\mathbf{z})\}$ for all $\ell \in Z^*$, where $E\{\cdot\}$ denotes ordinary expected value. (Clearly, if $Z = \mathcal{C}_n(U)$ then $E\mathbf{z}$ exists if and only if $(E\mathbf{z})(x) = E\{\mathbf{z}(x)\}$ for every $x \in U$.) The *covariance* of $\mathbf{z}$, denoted $\operatorname{cov}\mathbf{z}$ is defined to be the mapping from $Z^* \times Z^*$ into $\mathbb{R}$ given by

$$(\operatorname{cov}\mathbf{z})(\ell_1, \ell_2) = E\{(\ell_1(\mathbf{z}) - \ell_1(E\mathbf{z})][\ell_2(\mathbf{z}) - \ell_2(E\mathbf{z})]\}.$$

A random variable $\mathbf{z}$ taking values in Z will be called *Gaussian* with mean $E\mathbf{z}$ and covariance $\operatorname{cov}\mathbf{z}$ provided that for all $\ell \in Z^*$ the real-valued random variable $\ell(\mathbf{z})$ is normally distributed with mean $\ell(E\mathbf{z})$ and covariance $\operatorname{cov}\ell(\mathbf{z})$.

Let us now return to the specific case at hand, that of the Banach space $\mathcal{C}_n(U)$. The first assertion leading to (4.2) is that the functions $E^\nu\mathbf{f}(\cdot)$ are $\mathcal{C}_n(U)$-valued random variables. This is a consequence of the following proposition.

Proposition A1. *Let $(S, \mathcal{S})$ be a measurable space, and let $g : U \times S \to \mathbb{R}^n$ be continuous in the first argument, $\forall s \in S$, and measurable in the second, $\forall x \in U$. Then the mapping $s \mapsto f(\cdot, s)$ is Borel measurable as a mapping from S into $\mathcal{C}_n(U)$.*

Proof. It suffices to show that for every $\alpha > 0$, the set

$$\{s \mid \sup_{x \in U} |f(s,x)| \leq \alpha\}$$

is a measurable subset of $\mathbb{R}^n$. This follows easily from standard results in the theory of measurable multifunctions; see, for example, Rockafellar [**10**; Theorem 2K]. □

Corollary A2. *$E^\nu\mathbf{f}$ is a $\mathcal{C}_n(U)$-valued random variable for every $\nu = 1, 2, \ldots$.*

The main result is a "well-known" theorem that does not seem to have been published for $\mathcal{C}_n(U)$ with $n \geq 2$. The argument presented here was suggested by Professor R. Pyke.

Theorem A3. *Suppose that $g : U \times S \to \mathbb{R}^n$ satisfies the probabilistic assumptions P.1–4. Then there exists a Gaussian random variable $\mathbf{w}$ taking values in $\mathcal{C}_n(U)$ such that*

$$\sqrt{\nu}(E^\nu\mathbf{f} - E\mathbf{f}) \xrightarrow[\mathcal{D}]{} \mathbf{w},$$

where for all $x \in U$, $\mathbf{w}(x)$ is a normal $N(0, \Sigma(x))$ random variable with covariance $\Sigma(x) =$ $\text{cov}[f(x, \mathbf{s}_1)]$.

Proof. Each $E^\nu \mathbf{f}$ is a vector of continuous functions $(E^\nu \mathbf{f}_1, \ldots, E^\nu \mathbf{f}_n)$. The conditions of the theorem imply that for each $j = 1, \ldots, n$ there is a Gaussian random variable in $\mathcal{C}_n(U)$ with zero mean and covariance equal to $\text{cov}\, \mathbf{f}_j$, which we suggestively call $\mathbf{w}_j$, such that

$$\sqrt{\nu}(E^\nu \mathbf{f}_j - E\mathbf{f}_j) \xrightarrow[\mathcal{D}]{} \mathbf{w}_j;$$

cf. Araujo and Giné [**2**; 7.17]. It follows that the finite-dimensional distributions of $\mathbf{w}^\nu := \sqrt{\nu}(E^\nu \mathbf{f} - E\mathbf{f})$ converge to those of $\mathbf{w}$, i.e. for all finite subsets $\{x_1, \ldots, x_k\} \subset U$ one has

$$(\mathbf{w}^\nu(x_1), \ldots, \mathbf{w}^\nu(x_k)) \xrightarrow[\mathcal{D}]{} (\mathbf{w}(x_1), \ldots, \mathbf{w}(x_k)).$$

This determines the limit $\mathbf{w}$, if it exists, uniquely as that in the statement of the theorem. Thus by Prohorov's Theorem (Billingsley [**5**; 6.1]) it remains only to show that the sequence $\{\mathbf{w}^\nu\}$ is *tight* in $\mathcal{C}_n(U)$, i.e. for each $\varepsilon > 0$ there is a compact set $A \subset \mathcal{C}_n(U)$ such that $\Pr\{\mathbf{w}^\nu \in A\} > 1 - \varepsilon$ for all sufficiently large ν. By adapting the argument of [**5**; 8.2] for $\mathcal{C}_n(U)$ we find that the tightness of $\{\mathbf{w}^\nu\}$ is equivalent to the simultaneous satisfaction of the following two conditions:

(i) There exists $x \in U$ such that for each $\eta > 0$ there is $\alpha \geq 0$ with

$$\Pr\{|\mathbf{w}^\nu(x)| > \alpha\} \geq \eta, \quad \forall \nu \geq 1.$$

(ii) For each positive ε and η there exist $\delta > 0$ and an integer ν_0 such that

$$\Pr\{\sup_{(x-y)<\delta} |\mathbf{w}^\nu(x) - \mathbf{w}^\nu(y)| \geq \varepsilon\} \leq \eta, \quad \forall \nu \geq \nu_0.$$

These conditions follow easily from the tightness of the coordinate sequences $\{\mathbf{w}_j^\nu\}$ for $j = 1, \ldots, n$ since

$$\Pr\{|\mathbf{w}^\nu(x)| > \alpha\} \leq \sum_{j=1}^{n} \Pr\left\{|\mathbf{w}_j^\nu(x)| > \frac{\alpha}{\sqrt{n}}\right\},$$

and similarly for the probability in condition (ii), and hence these can be made as small as one pleases by application of conditions (i) and (ii) to the co-ordinate sequences. Thus $\{\mathbf{w}^\nu\}$ is tight, and the proof is complete. □

References

1. J. Aitchison and S.D. Silvey, "Maximum likelihood estimation of parameters subject to restraints", *Annals of Mathematical Statistics* **29**(1948), 813–828.
2. A. Araujo and E. Giné, *The Central Limit Theorem for Real and Banach Valued Random Variables*, Wiley, 1980.
3. B.R. Clarke, "Nonsmooth analysis and Fréchet differentiability of M-functionals", *Probability Theory and Related Fields* **73** (1986) 197–209.
4. J. Dupačová and R.J-B Wets, "Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems", *Annals of Mathematical Statistics* (1988) 1517–1549.
5. P.J. Huber, "The behavior of maximum likelihood estimates under non-standard conditions", *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics*(1967), 221–233.
6. A.J. King, *Asymptotic Behaviour of Solutions in Stochastic Optimization: Nonsmooth Analysis and the Derivation of Non-normal Limit Distributions*, Ph.D. Dissertation, University of Washington, 1986.
7. A.J. King, "Generalized delta theorems for multivalued mappings and measurable selections", *Mathematics of Operations Research* (1989) (to appear).
8. A.J. King and R.T. Rockafellar, "Sensitivity analysis for nonsmooth generalized equations", Research Report RC14639, International Business Machines Corporation, T.J. Watson Research Center, Yorktown Heights, New York, 1989.
9. S.M. Robinson, "Generalized equations and their solutions, part II: applications to nonlinear programming", *Mathematical Programming Study* **19** (1982) 200–221.
10. R.T. Rockafellar, "Integral functionals, normal integrands and measurable selections", in *Nonlinear Operators and the Calculus of Variations*, Lecture Notes in Mathematics 543, Springer-Verlag, 1976, pp. 157–207.
11. A. Shapiro, "Asymptotic properties of statistical estimators in stochastic programming", manuscript, University of South Africa, Pretoria, South Africa, 1988.