WORKING PAPER

GENERALIZED DELTA THEOREMS FOR MULTIVALUED MAPPINGS AND MEASURABLE SELECTIONS

Alan J. King

July 1988
WP-88-56

IIASA
International Institute
for Applied Systems Analysis

GENERALIZED DELTA THEOREMS FOR MULTIVALUED MAPPINGS AND MEASURABLE SELECTIONS

Alan J. King

July 1988
WP-88-56

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

FOREWORD

The classical delta theorem can be generalized in a mathematically satisfying way to a broad class of multivalued and/or nonsmooth mappings, by examining the convergence in distribution of the sequence of difference quotients from the perspectives of recent developments in convergence theory for random closed sets and new descriptions of first-order behaviour of multivalued mappings. Such a theory opens the way to applications of asymptotic techniques in many areas of mathematical optimization where randomness and uncertainty play a role. Of special importance is the asymptotic convergence of measurable selections of multifunctions when the limit multifunction is single-valued almost surely.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

1 Introduction 1

2 Convergence in Distribution for Measurable Multifunctions and Selections 5

3 Semi-differentiability and the Delta Theorem for Multi-valued Mappings 10

4 Delta Theorem under Single-valuedness Assumptions 15

Appendix 19

Acknowledgements 21

References 22

GENERALIZED DELTA THEOREMS FOR MULTIVALUED MAPPINGS AND MEASURABLE SELECTIONS

*Alan J. King**

Abstract. The classical delta theorem can be generalized in a mathematically satisfying way to a broad class of multivalued and/or nonsmooth mappings, by examining the convergence in distribution of the sequence of difference quotients from the perspectives of recent developments in convergence theory for random closed sets and new descriptions of first-order behaviour of multivalued mappings. Such a theory opens the way to applications of asymptotic techniques in many areas of mathematical optimization where randomness and uncertainty play a role. Of special importance is the asymptotic convergence of measurable selections of multifunctions when the limit multifunction is single-valued almost surely.

Keywords: convergence in distribution, random closed sets, contingent derivatives, asymptotic normality, maximum likelihood estimates, stochastic programs.

* International Institute for Applied Systems Analysis, A-2361, Laxenburg, Austria.
This research was supported in part by grants from the National Sciences and Engineering Research Council of Canada and the National Science Foundation.

1. Introduction

Optimal decision problems must frequently model uncertainties that can be estimated from random samples; however, the possibly nonsmooth and multivalued nature of the mappings that express optimality conditions makes it difficult to apply asymptotic techniques from statistics. The influence of inequality constraints at the boundary of the feasible region and of points of non-differentiability of the objective function may complicate matters considerably. In this paper, we generalize the useful *delta theorem* of Mann and Wald [**14**] into a version for set-valued maps appropriate for applications in optimization. Examples are provided that illustrate (but not exhaustively) the potential of this theory to address asymptotic issues.

We study sequences of random closed sets that have a special form, namely

$$\mathbf{F}_\nu = F(\mathbf{z}_\nu), \qquad \nu = 1, 2, \ldots,$$

where $\{\mathbf{z}_\nu\}$ is a sequence of random variables in a separable Fréchet space Z with known (or knowable) asymptotic behaviour, and $F : Z \rightrightarrows X$ is a closed-valued measurable multifunction. In many applications, as shown by the examples given below, the random closed sets of interest may be described by isolating the stochasticity in an object that can be understood as a random variable $\mathbf{z}_\nu$ and then describing the random closed set as a multivalued but deterministic mapping of this random variable. For systems with this property the asymptotic analysis falls naturally into two pieces: understanding the asymptotic behaviour of the sequence $\{\mathbf{z}_\nu\}$ and describing local properties of the multifunction F. When the sequence of random variables $\{\mathbf{z}_\nu\}$ satisfies an asymptotic formula:

$$\tau_\nu{}^{-1}[\mathbf{z}_\nu - z^*] \xrightarrow[\mathcal{D}]{} \mathfrak{z} \tag{1.1}$$

for some sequence of positive numbers $\{\tau_\nu\}$ decreasing to 0, we prove in the main result of the paper that an analogous formula holds for the random closed sets:

$$\tau_\nu{}^{-1}[F(\mathbf{z}_\nu) - x^*] \xrightarrow[\mathcal{D}]{} F'_{z^*,x^*}(\mathfrak{z}), \tag{1.2}$$

where F'_{z^*,x^*} is a "derivative" of F localized at a given point $x^* \in F(z^*)$. (The symbol $\mathcal{D}$ under the arrow denotes convergence in distribution.) This is the *delta theorem* for multivalued mappings that we are after.

To prove this theorem, we consider the convergence in distribution of the sequence of "difference quotients" and apply some basic results from the theory of convergence of probability measures. We follow Salinetti and Wets [**17**] in analyzing the distributions induced by the multifunctions regarded as a measurable function (random closed set) into the space of closed subsets of X, equipped with the compact, metrizable topology of Kuratowski set convergence. The crucial condition turns out to be "semi-differentiability", a concept introduced recently by Rockafellar [**16**] in his exploration of differentiability concepts for multifunctions. This theory is in its infancy; nevertheless, explicit computations are already possible in some situations as shown by the examples. These strong connections between the delta theorem and the theory of semi-differentiability for multifunctions are a hopeful sign that we are on the threshold of some really useful results concerning the influence of data and statistical approximations in mathematical programming.

Parts of the material in this paper appeared in two earlier working papers: [**11**] and [**12**].

A few examples will help to motivate the formulation of the fundamental problem treated in this paper. In what follows $\{\mathbf{s}_k,\ k = 1, 2, \ldots\}$ is a collection of independent and identically distributed random variables on $\mathbb{R}^d$.

Example 1.1. The set of feasible solutions to a system of smooth constraints depending smoothly on a parameter $z \in \mathbb{R}^d$ may be modelled as a multifunction $F : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$, by

$$F(z) = \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{array}{ll} f_i(z,x) \leq 0, & i = 1, \ldots, s \\ f_i(z,x) = 0, & i = s+1, \ldots, m, \end{array} \right\}$$

where the functions $f_i : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$ are jointly C^1. Suppose that z could be known only through taking a finite sample from the collection $\{\mathbf{s}_k\}$ and forming the sample mean, as might be the case if our knowledge of z came from "noisy" measurements. For each finite sample of size $\nu = 1, 2, \ldots$ we can form the estimated feasible set $F\left(\frac{1}{\nu}\sum_{k=1}^{\nu} \mathbf{s}_k\right)$. If the sequence $\{\mathbf{s}_k\}$ is well behaved then the sequence of sample means is asymptotically normal, i.e. the sample means satisfy (1.1) with $\tau_\nu = \sqrt{\nu}$ and the limit distribution $\mathfrak{z}$ turns out to be normal, or Gaussian. Under reasonable regularity conditions we can study the asymptotic behaviour (1.2) of the sequence of estimated feasibility sets. This will be developed further as Example 3.6.

Example 1.2. We can further ask about optimal solutions to a mathematical programming problem that depends on an estimated parameter. The Kuhn-Tucker conditions can be studied as an extension of the previous example. However, the conditions required to guarantee semi-differentiability are fairly strong. In Section 4, we apply a modified form of the delta theorem to describe the asymptotic behaviour of the solutions to a linearly constrained least squares problem. In Shapiro [**19**] there are similar results for solutions to smoothly constrained mathematical programs.

Example 1.3. This example comes from stochastic optimization. Let us suppose we wish to solve the problem

$$\text{minimize } Ef(x, \mathbf{s}_1) \text{ over all } x \in C,$$

but we can only form approximations to the integral by obtaining samples $\{\mathbf{s}_k\}$. For each sample of size ν we obtain a random solution set

$$\mathbf{J}_\nu = \{x \mid x \text{ minimizes } \frac{1}{\nu}\sum_{k=1}^{\nu} f(x, \mathbf{s}_k) \text{ over all } x \in C\}.$$

We can fit the pattern of the first two examples by observing that the *function*

$$\frac{1}{\nu}\sum_{k=1}^{\nu} f(\cdot, \mathbf{s}_k)$$

is an estimate of the true objective and if J is the solution multifunction

$$J(g) = \{x \mid x \text{ minimizes } g(x) \text{ over } x \in C\},$$

then writing $\mathbf{z}_\nu(\cdot) = \frac{1}{\nu}\sum_{k=1}^{\nu} f(\cdot, \mathbf{s}_k)$ it follows that

$$J(\mathbf{z}_\nu) = \mathbf{J}_\nu.$$

The pattern is completed if we can establish that the sequence of "sample means" $\{\mathbf{z}_\nu(\cdot)\}$ is asymptotically normal in some suitably generalized sense. This is the principle reason why we present our results for multivalued mappings defined in a general metric space Z. Other approaches to the same problem have been explored in the maximum likelihood literature. Asymptotic results have been achieved using "von Mises functionals"; an early reference is Kallianpur [**9**], and some more recent references are Boos and Serfling [**4**] and Clarke [**6**]. The differentiablity notions explored in the present paper may be used in this setting as well, but we shall not consider this matter here. Aitchison and Silvey [**1**] proved asymptotic normality of constrained maximum likelihood estimates. A different view appeared in Huber [**8**], and these ideas have recently been applied to stochastic optimization in Dupačová and Wets [**7**]. The asymptotic analysis of $\{\mathbf{J}_\nu\}$ from the perspective of the present paper appears in King [**13**].

The link between convergence of multifunctions and convergence of selections is an important theme of this paper. This reflects the point of view of the practitioner who may have access only to a single selection of the multifunction and who wishes to draw conclusions based only on this limited knowledge. In some situations, the limit is best regarded as a multifunction that satisfies certain regularity properties—namely, single-valuedness. The nature of the single-valuedness assumption is explored in an appendix.

The principal ideas of convergence in distribution for multifunctions needed in the paper are summarized and some new results concerning convergence of selections are given in Section 2. The main delta theorem appears in Section 3, and an application to feasibility sets of a mathematical program is described. In Section 4, we derive a delta theorem for upper Lipschitzian mappings that are single-valued almost everywhere. This version has proven useful in establishing the asymptotic behaviour of estimates in linear-quadratic stochastic programming (King [**10**]), but given the confines of the present paper, we must content ourselves with a brief derivation of the asymptotic distribution for linearly constrained least squares problems.

The reader of this paper is expected to be acquainted with the fundamentals of convergence of closed sets and weak*-convergence of probability measures; see, for example, Salinetti and Wets [**17**] and Billingsley [**3**], respectively. A sequence of subsets $\{A_\nu\}$ of a locally compact topological space *set-converges* to a subset A if

$$A = \limsup A_\nu = \liminf A_\nu,$$

where

$$\begin{aligned}\liminf A_\nu &= \{a \mid a = \lim a_\nu \text{ where } a_\nu \in A_\nu \text{ for all but finitely many } \nu\}\\ \limsup A_\nu &= \{a \mid a = \lim a_\nu \text{ where } a_\nu \in A_\nu \text{ for infinitely many } \nu\}.\end{aligned}$$

A sequence of probability measures $\{\mu_\nu\}$ on a complete separable metric space Z *weak*-converges* to μ if

$$\lim \int f(z)\mu_\nu(dz) = \int f(z)\mu(dz)$$

for all bounded continuous functions $f : Z \to \mathbb{R}$.

2. Convergence in Distribution for Measurable Multifunctions and Selections

Let X be a finite dimensional linear space equipped with a norm $|\cdot|$. A multivalued map $F : \Omega \rightrightarrows X$ defined on a probability space $(\Omega, \mathcal{A}, \mu)$ whose values are closed subsets of X is said to be a *closed-valued measurable multifunction* if for all closed subsets $C \subset X$ the inverse image

$$F^{-1}(C) := \{\omega \,|\, F(\omega) \cap C \neq \emptyset\}$$

belongs to $\mathcal{A}$. (In parallel with the measurable function/random variable dualism, when the probability space Ω is unspecified we shall call such a mapping a *random closed set* and use the boldface notation "**F**".) Following Salinetti and Wets [**18**], we observe that the mapping F may be identified with a Borel *measurable function* $\varphi : \Omega \to \mathcal{F}(X)$ from Ω into the hyperspace $\mathcal{F}(X)$ of all closed subsets of X equipped with the topology consistent with convergence of sets. This space $\mathcal{F}(X)$ so topologized is in particular compact, separable, and metrizable. Every closed-valued measurable multifunction thus induces a regular probability measure $\mu\varphi^{-1}$ on the Borel field of $\mathcal{F}(X)$. *Convergence in distribution* of a sequence $\{F_\nu\}$ of such mappings, written $F_\nu \xrightarrow[\mathcal{D}]{} F$, is then defined to be the weak*-convergence of the measures $\mu\varphi_\nu^{-1}$ to $\mu\varphi^{-1}$ induced on $\mathcal{F}(X)$.

An important feature of this definition is that it turns out to be equivalent to convergence of certain stochastic processes on X, in the sense of convergence in distribution of the finite-dimensional sections. Each subset $C \subset X$ may be associated in a unique way with the *distance function* $d(\cdot, C) : X \to \bar{\mathbb{R}}_+$ given by

$$d(x, C) := \inf_{y \in C} |x - y|, \tag{2.1}$$

where $\bar{R}_+$ is the space of nonnegative reals made compact with the inclusion of the point at infinity. Relying on the fact that a sequence of closed sets converges in $\mathcal{F}(X)$ if and only if the sequence of distance functions converges pointwise, Salinetti and Wets [**18**; Theorem 2.5] demonstrate that a sequence of closed-valued measurable multifunctions $\{F_\nu\}$ converges in distribution if and only if the *distance processes* $\{d(\cdot, F_\nu)\}$ converge as stochastic processes on X. By definition these stochastic processes $d(\cdot, F_\nu)$ *converge to* $d(\cdot, F)$, in notation:

$$d(x, F_\nu) \xrightarrow[\mathcal{D}]{} d(x, F), \qquad x \in X,$$

if and only if for all finite collections $\{x_1, \ldots, x_k\}$ of points in X one has

$$[d(x_1, F_\nu), \ldots, d(x_k, F_\nu)] \xrightarrow[\mathcal{D}]{} [d(x_1, F), \ldots, d(x_k, F)] \tag{2.2}$$

as random variables in $\bar{\mathbb{R}}_+^{(k)}$. This characterization plays an important role in computations.

The reader should note that a sequence of measurable functions $\{f_\nu\}$, $f_\nu : \Omega \to X$ converges in distribution to f if and only if one has weak*-convergence of the measures induced by the f_ν on the space X. We may also regard these as closed-valued measurable multifunctions since points are closed in X. But in this view the sequence $\{f_\nu\}$ converges in distribution if and only if one has weak*-convergence of the distributions induced on the hyperspace $\mathcal{F}(X)$. Clearly, the weak*-convergence of $\{f_\nu\}$ as functions into X implies that of the distributions induced on $\mathcal{F}(X)$; however, a moment's reflection will convince the reader that the reverse implication does not hold without further assumptions. A related issue is the convergence of selections of a converging sequence $\{F_\nu\}$ of almost surely single-valued multifunctions—the study of which will occupy the rest of this section.

We first make a few preliminary definitions. The *domain* of a closed-valued measurable multifunction $F : \Omega \rightrightarrows X$, denoted by $\operatorname{dom} F$, is the measurable set

$$\operatorname{dom} F = \{\omega \in \Omega \,|\, F(\omega) \neq \emptyset\}.$$

A function $f : \Omega \to X$ is called a *measurable selection of* F if f is measurable and $f(\omega) \in F(\omega)$ for μ-almost all $\omega \in \operatorname{dom} F$. There always exists at least one measurable selection of a nonempty closed-valued measurable multifunction; see, for example, Castaing and Valadier [**5**]. It is important to note that $\mu(\operatorname{dom} F)$ may be less than one and in this case the measure μf^{-1} induced on X by a measurable selection f of F is *not a probability measure.* This introduces a minor technical difficulty into the very definition of convergence in distribution for sequences $\{f_\nu\}$ of measurable selections, which as the reader recalls is defined to be weak*-convergence of the sequence $\{\mu f_\nu^{-1}\}$ of measures on X. A trivial modification to the proof of the Portmanteau Theorem, Billingsley [**3**; Theorem 2.1], yields the following.

Lemma 2.1. *A necessary condition for the weak*-convergence of a sequence of finite measures* $\{P_\nu\}$ *on a complete separable metric space* Z *is*

$$P_\nu(Z) \to P(Z) \tag{2.1}$$

Furthermore, if (3.1) holds then all the equivalences in the statement of the Portmanteau Theorem hold true for the sequence $\{P_\nu\}$. □

The significance of this lemma is that it allows us to apply all of the main results of weak*-convergence, in [**3**] for example, that depend on the equivalences in the Portmanteau Theorem but which do not specifically require the measures to be probabilities.

The key concept that permits us to draw conclusions about convergence in distribution for selections is that of single-valuedness of the limit mapping. We first make the definition.

Definition 2.2. A closed-valued measurable multifunction $F : \Omega \rightrightarrows X$ is said to be *μ-almost surely single-valued* if

$$\mu\{\omega \in \operatorname{dom} F \,|\, F(\omega) \text{ is not a singleton}\} = 0. \tag{2.4}$$

In the Appendix we present a study of μ-almost surely single-valued multifunctions. It is shown there that for such a multifunction one can interpret μF^{-1} or a measure on the space X itself and, moreover, that any selection f of F gives rise to the same distribution, i.e. $\mu F^{-1} = \mu f^{-1}$ on the Borel sets of X. We shall not make use of this interpretation, however, since we prefer to retain the distinction between measures on $\mathcal{F}(X)$ induced by multifunctions and those on X induced by selections.

We now present a pair of results concerning the convergence in distribution of a sequence of measurable selections from a converging sequence of measurable multifunctions. The first theorem makes use of the concept of *tightness*: a sequence of measurable functions $\{f_\nu\}$ from the probability space $(\Omega, \mathcal{A}, \mu)$ into X is *tight* if for all $\varepsilon > 0$ there is a compact set A of X such that $\mu\{f_\nu \in A\} \geq 1-\varepsilon$. The second theorem supposes that the converging multifunctions become more and more single-valued.

Theorem 2.3. *Let the closed-valued measurable multifunctions $F_\nu : \Omega \rightrightarrows X$ converge in distribution to F. Suppose that F is μ-almost surely single-valued and that*

$$\mu(\operatorname{dom} F_\nu) \to \mu(\operatorname{dom} F) \tag{2.5}$$

If $f_\nu : \Omega \to X$ and $f : \Omega \to X$ are measurable selections of F_ν and F, respectively, and if the sequence $\{f_\nu\}$ is tight, then the f_ν converge in distribution to f.

Proof. First, note that $\mu f_\nu^{-1}(X) = \mu(\operatorname{dom} F_\nu)$. Thus assumption (2.5) means $\mu f_\nu^{-1}(X) \to \mu f^{-1}(X)$, and so Lemma 2.1 applies. Let A be an arbitrary closed subset of X; we plan to show that

$$\limsup_\nu \mu f_\nu^{-1}(A) \leq \mu f^{-1}(A).$$

Fix $\varepsilon > 0$ and let $K \subset X$ be such that $\mu f_\nu^{-1}(K) \geq 1 - \varepsilon$ for every ν, then

$$\mu f_\nu^{-1}(A) - \varepsilon \leq \mu f_\nu^{-1}(K \cap A).$$

Identify with each F_ν and F the maps $\varphi_\nu : \Omega \to \mathcal{F}(X)$ and $\varphi : \Omega \to \mathcal{F}(X)$, and define for any set $D \subset X$ the collection

$$\mathcal{F}_D := \{C \in \mathcal{F}(X) \,|\, C \cap D \neq \emptyset\}.$$

Then, clearly,

$$\mu f_\nu^{-1}(K \cap A) \le \mu \varphi_\nu^{-1}(\mathcal{F}_{K\cap A}),$$

and furthermore $\mathcal{F}_{K\cap A}$ is closed in $\mathcal{F}(X)$ — since $K \cap A$ is compact — hence

$$\limsup_\nu \mu \varphi_\nu^{-1}(\mathcal{F}_{K\cap A}) \le \mu \varphi^{-1}(\mathcal{F}_{K\cap A}).$$

It follows that

$$\limsup_\nu \mu f_\nu^{-1}(A) - \varepsilon \le \mu \varphi^{-1}(\mathcal{F}_{K\cap A}).$$

To complete the proof note that $\mu\varphi^{-1}(\mathcal{F}_A) = \mu f^{-1}(A)$, by the single-valuedness of F (cf. the appendix), and that $\varepsilon > 0$ was arbitrary. □

Theorem 2.4. *Suppose that the closed-valued measurable multifunctions $F_\nu : \Omega \rightrightarrows X$, $\nu = 1, 2, \ldots$, converge in distribution to the closed-valued measurable multifunction $F : \Omega \rightrightarrows X$. Suppose, moreover, that F is μ-almost surely single-valued, that (2.5) holds, and that*

$$\mu\{\omega \in \operatorname{dom} F_\nu | F_\nu(\omega) \text{ is not single-valued}\} \to 0. \tag{2.6}$$

If $f : \Omega \to X$ and $f_\nu : \Omega \to X$ are measurable selections of F and F_ν, respectively, then the sequence $\{f_\nu\}$ converges in distribution to f.

Proof. For convenience denote by P and P_ν the finite measures μf^{-1} and μf_ν^{-1} on X. First note that $P(X) = \mu f^{-1}(X) = \mu(\operatorname{dom} F)$. Thus assumption (2.5) means $P_\nu(X) \to P(X)$, and so Lemma 2.1 applies. Denote by $B(x, \varepsilon)$ the open sphere of radius $\varepsilon > 0$ centered at the point $x \in X$, i.e.

$$B(x, \varepsilon) = \{y \in X \,\big|\, |y - x| < \varepsilon\}.$$

The collection of all sets that are finite intersections of open spheres is a *convergence determining class*; cf. the corollaries to [**3**; Theorem 2.2]. Let A be a member of this class, i.e.

$$A = \bigcap_{i=1}^{k} B(x_i, \varepsilon_i).$$

We may suppose without loss of generality that the $B(x_i, \varepsilon_i)$ are *P-continuity sets*; this implies that $\prod_{i=1}^{k}(-\infty, \varepsilon_i)$ is a continuity set for the random vector

$$\omega \mapsto [d(x_1, F(\omega)), \ldots, d(x_k, F(\omega))],$$

since

$$\mu\{\omega \in \Omega \mid [d(x_1, F(\omega)), \dots, d(x_k, F(\omega))] \in \partial \prod_{i=1}^{k}(-\infty, \varepsilon_i)\}$$
$$= \mu\{\omega \in \Omega \mid [d(x_1, x(\omega)), \dots, d(x_k, f(\omega))] = [\varepsilon_1, \dots, \varepsilon_k]\}$$
$$\leq \sum_{i=1}^{k} P(\partial B(x_i, \varepsilon_i))$$

which is zero. The convergence of the processes $d(\cdot, F_\nu(\omega))$ to $d(\cdot, F(\omega))$ — cf. equation (2.2) — and the Portmanteau Theorem imply

$$\lim_{\nu} \mu\{\omega \in \Omega \mid [d(x_1, F_\nu(\omega)), \dots, d(x_k, F_\nu(\omega))] \in \prod_{i=1}^{k}(-\infty, \varepsilon_i)\}$$
$$= \mu\{\omega \in \Omega \mid [d(x_1, F(\omega)), \dots, d(x_k, F(\omega))] \in \prod_{i=1}^{k}(-\infty, \varepsilon_i\},$$

and this latter set is equal to $P(A)$ since F is μ-almost surely single-valued. Define the sets S_ν, $\nu = 1, 2, \dots$, by

$$S_\nu = \{\omega \in \operatorname{dom} F_\nu \mid F_\nu(\omega) \text{ is a singleton}\}.$$

Noting that by the Appendix the sets S_ν are all measurable, we have

$$\begin{aligned} P_\nu(A) = {} & \mu\{\omega \in \operatorname{dom} F_\nu \mid d(x_i, F_\nu(\omega)) < \varepsilon_i, \quad i = 1, \dots, k\} \\ & + \mu\{\omega \in \operatorname{dom} F_\nu \setminus S_\nu \mid d(x_i, x_\nu(\omega)) < \varepsilon_i, \quad i = 1, \dots, k\} \\ & - \mu\{\omega \in \operatorname{dom} F_\nu \setminus S_\nu \mid d(x_i, F_\nu(\omega)) < \varepsilon_i, \quad i = 1, \dots, k\} \end{aligned}$$

Hence by assumption (2.6) and the observation that

$$\begin{aligned} & \mu\{\omega \in \operatorname{dom} F_\nu \mid d(x_i, F_\nu(\omega)) < \varepsilon_i, \quad i = 1, \dots, k\} \\ & \qquad = \mu\{\omega \in \Omega \mid d(x_i, F_\nu(\omega)) < \varepsilon_i, \quad i = 1, \dots, k\}, \end{aligned}$$

we have $P_\nu(A) \to P(A)$. Since A was an arbitrary member of a convergence determining class it follows that P_ν weak*-converges to P and the proof is complete. □

To assist in the verification of condition (2.5) in Theorem 2.3 we have the following proposition.

Proposition 2.5. *Suppose that the closed-valued measurable multifunctions $F_\nu : \Omega \rightrightarrows X$, $\nu = 1, 2, \dots$, converge in distribution to the closed-valued measurable multifunction $F : \Omega \rightrightarrows X$. If $\mu(\operatorname{dom} F) = 1$, then*

$$\mu(\operatorname{dom} F_\nu) \to \mu(\operatorname{dom} F).$$

Proof. Since the F_ν converge in distribution to F, the random variables $\omega \mapsto d(0, F_\nu(\omega))$ must converge in distribution to the random variable $\omega \mapsto d(0, F(\omega))$; see equation (2.2). Now

$$\mu(\operatorname{dom} F_\nu) = \mu\{\omega \in \Omega \,|\, d(0, F_\nu(\omega)) < \infty\},$$

and thus by the Portmanteau Theorem $\mu(\operatorname{dom} F_\nu) \to \mu(\operatorname{dom} F)$, provided $\bar{\mathbb{R}}_+$ is a continuity set for the random variable $\omega \mapsto d(0, F(\omega))$, i.e., provided

$$\mu\{\omega \in \Omega \,|\, d(0, F(\omega)) = \infty\} = 0,$$

which is indeed the case by our assumption that $\mu(\operatorname{dom} F) = 1$. □

3. Semi-differentiability and the Delta Theorem for Multi-valued Mappings

The main result is presented in this section. Let Z be a separable complete metric vector space (*separable Fréchet space*) equipped with its Borel field $\mathcal{Z}$, let X be a finite-dimensional Euclidean space, and let the map $F : Z \rightrightarrows X$ be closed-valued and measurable. On the space Z define a sequence $\{\mathbf{z}_\nu\}$ of random variables. Trivially, each $F(\mathbf{z}_\nu)$ is a random closed set in X. Our interest here is in the possibility of describing the asymptotic behaviour of this sequence of random closed sets, when the sequence $\{\mathbf{z}_\nu\}$ of random variables satisfies a *generalized central limit formula*: there are a point z^*, a sequence of positive numbers $\{\tau_\nu\}$ monotonically decreasing to 0, and a *limit distribution* $\mathfrak{z}$ such that

$$\tau_\nu^{-1}[\mathbf{z}_\nu - z^*] \xrightarrow[\mathcal{D}]{} \mathfrak{z} \tag{3.1}$$

as random variables in Z.

A delta theorem for the sequence $\{F(\mathbf{z}_\nu)\}$ inevitably rests upon an appropriate definition of first-order behaviour for the multifunction $F : Z \rightrightarrows X$. Fix a point z^* and a point $x^* \in F(z^*)$, and define the collection $\{D_t : t > 0\}$ of *difference quotient multifunctions*

$$D_t(z) := t^{-1}[F(z^* + tz) - x^*], \qquad t > 0. \tag{3.2}$$

The *contingent derivative* of F at the point (z^*, x^*), denoted $F^+_{z^*,x^*}$, is the mapping whose graph is the *contingent cone* to the graph of F at (z^*, x^*), given by the formula

$$\limsup_{t \downarrow 0} \operatorname{gph} D_t = \operatorname{gph} F^+_{z^*,x^*} \tag{3.3}$$

This derivative was introduced by Aubin [**2**]. It always exists, but it may be very difficult to use unless more information is available. Further regularity conditions on the contingent derivative were introduced in Rockafellar [**16**]. If one actually has lim sup = lim inf in (3.3), i.e.

$$\lim_{t\downarrow 0} \operatorname{gph} D_t = \operatorname{gph} F'_{z^*,x^*} \tag{3.4}$$

then F is said to be *proto-differentiable*, and the notation F' is used for the *proto-derivative*. The multifunction F is said to be *semi-differentiable at* z^* *relative to* x^* if there exists a multifunction $D : Z \rightrightarrows X$ such that for all $z \in Z$,

$$\lim_{\substack{t\downarrow 0\\ z'\to z}} D_t(z') = D(z) \tag{3.5}$$

taken as a limit of sets in X. If such a property holds then it can be shown that F is proto-differentiable at (z^*, x^*) and that the limit mapping D equals the proto-derivative F'_{z^*,x^*}. (See the proof of [**16**; Theorem 3.2] which generalizes to this infinite dimensional setting.)

The underlying philosophy of this differentiability notion is best considered from the geometric point of view. Take a point (z^*, x^*) in the graph of F and construct there a *tangent cone* to $\operatorname{gph} F$; this cone is then the graph of F'_{z^*,x^*}. The picture is the exact analogue of that for differentiable functions (going back to the original ideas of Fermat) viewing the graph of the derivative as the hyperplane in $Z \times X$ tangent to the graph of the function at (x^*, z^*). Naturally, different choices of tangent cones — e.g. Clarke, intermediate, contingent, etc. — all lead to different derivatives. The choice made in (3.4) is that $\operatorname{gph} F'_{z^*,x^*}$ should equal simultaneously the *contingent* and *intermediate* cones (respectively lim sup and lim inf in (3.4)).

In Section 4 we explore properties of the contingent derivative when F is single-valued at z^*; in this case one writes $F^+_{z^*}$. In all cases, the contingent derivative has *closed graph*, since the sets in (3.3) and (3.4) are closed in $Z \times X$ and are therefore *closed-valued and measurable.*

Semi-differentiability is a stronger property than proto-differentiability. When Z is finite-dimensional it can be shown that proto-differentiability plus a certain Lipschitz property (pseudo-Lipschitzian) imply semi-differentiability [**16**; 4.3]. We shall explore other connections between Lipschitz properties and semi-differentiability in Section 4. The crucial property in our present undertaking is a slightly modified definition of semi-differentiability.

Definition 3.1. Given a measure μ on $(Z, \mathcal{B}(Z))$, the multifunction $F : Z \rightrightarrows X$ is said to be *almost surely semi-differentiable at* z^* *relative to* x^* *with respect to* μ if there exists a multifunction $D : Z \rightrightarrows X$ such that (3.5) holds for all points z except possibly those in a set of μ-measure zero. Abusing the notation slightly, we still write $D = F'_{z^*,x^*}$ even though the limit (3.5) and not (3.4) is understood here. Note also that F'_{z^*,x^*} in this less restrictive definition is closed-valued and measurable, since it is the pointwise limit of a sequence of closed-valued and measurable multifunctions.

This differentiability notion turns out to be exactly what is needed, as we see in the following delta theorem for multivalued mappings. This theorem, the main result of the paper, opens the way to applications of asymptotic theory to situations in optimization theory that must be modelled by nonsmooth, multivalued mappings. Following the theorem, we explore its implications in several directions. First, we interpret the meaning of the asymptotic distribution as a statement concerning errors due to sampling. Second, we present an example that illustrates, although in a rather artificial way, the necessity of the semi-differentiability property. Finally, we develop an immediate application of the theorem to the asymptotic properties of feasibility sets in mathematical programming, when some parameters must be estimated from samples.

Theorem 3.2. *Let Z be a separable Fréchet space and X a finite dimensional normed linear space, and suppose $F : Z \rightrightarrows X$ is closed-valued and measurable. If the sequence of random variables $\{\mathbf{z}_\nu\}$ satisfies a generalized central limit formula, with limit z^* and limit distribution $\mathfrak{z}$, and if F is almost surely semi-differentiable at z^* relative to a point $x^* \in F(z^*)$ with respect to the measure induced by $\mathfrak{z}$, then $\{F(\mathbf{z}_\nu)\}$ satisfies the generalized central limit formula:*

$$\tau_\nu^{-1}[F(\mathbf{z}_\nu) - x^*] \xrightarrow[\mathcal{D}]{} F'_{z^*,x^*}(\mathfrak{z}) \tag{3.6}$$

as random closed sets in X or, equivalently,

$$d(x, \tau_\nu^{-1}[F(\mathbf{z}_\nu) - x^*]) \xrightarrow[\mathcal{D}]{} d(x, F'_{z^*,x^*}(\mathfrak{z})), \quad x \in X \tag{3.7}$$

as stochastic processes on X.

Proof. Denote by μ_ν the measures induced on the space Z by the random variables $\tau_\nu^{-1}[\mathbf{z}_\nu - z^*]$ and by μ that induced by $\mathfrak{z}$. The meaning of the generalized central limit formula (3.1) is precisely that μ_ν weak*-converges to μ. Employing the difference quotient notation (3.2), the measures induced on the complete separable metric space $\mathcal{F}(X)$ by the random closed sets on the left side of (3.6) may be represented as $\mu_\nu \delta_{\tau_\nu}^{-1}$, where $\delta_{\tau_\nu} : Z \to \mathcal{F}(X)$ is the function identified with D_{τ_ν}. By Billingsley [**3** ; Theorem 5.5] the sequence $\{\mu_\nu \delta_{\tau_\nu}^{-1}\}$ weak*-converges to $\mu\delta^{-1}$ if the set of points z for which $\lim \delta_{\tau_\nu}(z_\nu) = \delta(z)$ fails to hold for some sequence $\{z_\nu\}$ approaching z has μ-measure zero. This is precisely what is meant by almost sure semi-differentiability with respect to μ; hence the condition is satisfied if $\delta(z) = F'_{z^*,x^*}(z)$ for μ-almost all z. This establishes (3.6). That (3.7) is equivalent to (3.6) was shown by Salinetti and Wets [**18**; Theorem 2.5]. □

Evaluating these distance processes (3.7) at $x = 0$ gives a converging sequence of random variables in $\bar{\mathbb{R}}_+$; and, noting that for any subset $C \subset X$ the linearity of the norm implies

$$d(0, t^{-1}[C - x^*]) = t^{-1}d(x, C), \qquad t > 0,$$

we obtain the following corollary.

Corollary 3.3. *Under the conditions of Theorem 3.2,*

$$\tau_\nu^{-1} d(x^*, F(\mathbf{z}_\nu)) \xrightarrow[\mathcal{D}]{} d(0, F'_{z^*,x^*}(\mathfrak{z})) \tag{3.8}$$

as random variables in $\bar{\mathbb{R}}_+$. □.

Remark 3.4. This corollary leads to an important interpretation of the meaning of the asymptotic distribution $F'_{z^*,x^*}(\mathfrak{z})$. It represents the *residual uncertainty* in the estimate $F(\mathbf{z}_\nu)$ relative to $x^* \in F(z^*)$. If $\mathbf{x}_\nu \in F(\mathbf{z}_\nu)$ is a measurable selection then clearly

$$\tau_\nu^{-1}|x^* - \mathbf{x}_\nu| \geq \tau_\nu^{-1} d(x^*, F(\mathbf{z}_\nu))$$

so the asymptotic behaviour of $\tau_\nu^{-1}|x^* - \mathbf{x}_\nu|$ *cannot be better* than that described in (3.8). If F is convex-valued and $z^* \in \operatorname{int} \operatorname{dom} F$ then it can be shown that there exists a selection $\mathbf{x}_\nu \in F(\mathbf{z}_\nu)$ such that

$$|x^* - \mathbf{x}_\nu| = d(x^*, F(\mathbf{z}_\nu)),$$

i.e. $\{\mathbf{x}_\nu\}$ in norm converges in distribution to x^*. To say more than this about selections seems to require F and F'_{z^*,x^*} to be almost surely single-valued.

Example 3.5. A simple example illustrates the necessity of the semi-differentiability condition. Let $Z = X = \mathbb{R}$ and let $F : Z \rightrightarrows \mathbb{R}$ be the subgradient of the absolute value function,

$$F(z) = \begin{cases} +1 & \text{if } z > 0, \\ [-1, +1] & \text{if } z = 0, \\ -1 & \text{if } z < 0. \end{cases}$$

Choose $(z^*, x^*) = (0, 0) \in \operatorname{gph} F$. It is easy to see that $F'_{0,0}$ exists in the sense of formula (3.4) with

$$F'_{0,0}(z) = \begin{cases} \mathbb{R} & \text{if } z = 0, \\ \emptyset & \text{otherwise,} \end{cases}$$

and that the semi-differentiability condition (3.5) holds for every point $z \neq 0$ but fails at $z = 0$. For each $\nu = 1, 2, \ldots$ let $\mathbf{z}_\nu$ be the "random variable" taking the value ν^{-2} with probability one, then the sequence $\{\nu[\mathbf{z}_\nu - 0]\}$ converges in distribution to the random variable $\mathfrak{z}$ taking the value 0 with probability one. All the conditions of Theorem 3.2 are satisfied except that $\mathfrak{z}$ places nonzero mass on the point at which semi-differentiability fails. Denote by $h_\nu(\cdot)$ the distance function $d(0, \nu[F(\cdot) - 0])$ and by $h(\cdot)$ the function $d(0, F'_{0,0}(\cdot))$. If Corollary 3.3 holds then $h_\nu(\mathbf{z}_\nu) \xrightarrow[\mathcal{D}]{} h(\mathfrak{z})$. For any closed interval $[b, +\infty]$ in $\bar{\mathbb{R}}_+$ we have $h_\nu(\mathbf{z}_\nu) \in [b, +\infty]$ with probability one for all sufficiently large ν, but $h(\mathfrak{z}) \in [b, +\infty]$ with probability zero. This contradicts the Portmanteau Theorem [**3**; Theorem 2.1], thus Theorem 3.2 fails for this example.

Example 3.6. An immediate application reveals the computational potential of the theorem in mathematical programming. Let $Z = \mathbb{R}^d$ and $X = \mathbb{R}^n$, and define $F(z)$ to be the set of all $x \in \mathbb{R}^n$ satisfying

$$f_i(z, x) \begin{cases} \leq 0 & i = 1, \ldots, s \\ = 0 & i = s + 1, \ldots, m \end{cases} \tag{3.9}$$

where $f_i : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable for $i = 1, \ldots, m$. Suppose that the parameter z is known only in a statistical sense by making repeated observations $\{\mathbf{s}_1, \ldots, \mathbf{s}_\nu\}$ and averaging them to form an estimate $\mathbf{z}_\nu$, i.e.

$$\mathbf{z}_\nu = \frac{1}{\nu} \sum_{i=1}^{\nu} \mathbf{s}_\nu.$$

Under easily satisfied conditions the $\mathbf{z}_\nu$ obey a *central limit formula*

$$\sqrt{\nu}[\mathbf{z}_\nu - z^*] \xrightarrow[\mathcal{D}]{} \mathfrak{z},$$

where $\mathfrak{z}$ has a centered Gaussian distribution. If (z^*, x^*) is a point where the system (3.9) satisfies the Mangasarian-Fromowitz constraint qualification, then (cf. Rockafellar [**3**; Example 5.5]) the mapping F is semi-differentiable at z^* relative to x^* and, moreover, an explicit formula is obtained for the proto-derivative F'_{z^*,x^*}, namely for all z the set $F'_{z^*,x^*}(z)$ consists of the points x satisfying the linearized system

$$\nabla_z f_i(z^*, x^*) \cdot z + \nabla_x f_i(z^*, x^*) \cdot x \begin{cases} \leq 0 & \text{for all } i \in I(z^*, x^*), \\ = 0 & \text{for } i = s + 1, \ldots, m, \end{cases}$$

with $I(z^*, x^*)$ denoting the inequality constraints of (3.9) active at (z^*, x^*). From Theorem 3.2,

$$\sqrt{\nu}[F(\mathbf{z}_\nu) - x^*] \xrightarrow[\mathcal{D}]{} F'_{z^*,x^*}(\mathfrak{z}).$$

The limit distribution $F'_{z^*,x^*}(\mathfrak{z})$ is a *Gaussian random polyhedron*: letting $\mathbf{b}_i$ denote the (Gaussian) random variable $-\nabla_z f_i(z^*, x^*) \cdot \mathfrak{z}$ for $i = 1, \ldots, m$, we have

$$F'_{z^*,x^*}(\mathfrak{z}) = \left\{ x \,\middle|\, \begin{array}{lll} \nabla_x f_i(z^*, x^*) \cdot x & \leq \mathbf{b}_i & i \in I(z^*, x^*) \\ \nabla_x f_i(z^*, x^*) \cdot x & = \mathbf{b}_i & i = s+1, \ldots, m \end{array} \right\},$$

a set defined by a set of linear constraints with random right hand sides that indicates how the random closed set $F(\mathbf{z}_\nu)$ approximates the limit set $F(z^*)$ near the point x^*.

4. Delta Theorem under Single-valuedness Assumptions

The assumption of single-valuedness of a limit mapping is a kind of regularity condition that allows other conditions to be simplified. In this section we explore the concept of semi-differentiability for a specialized class of mappings. An application to constrained least squares estimation shows the potential of this theory.

The space Z, in this section only, will be assumed to be a Banach space with norm $\|\cdot\|$. As usual, X is finite-dimensional and Euclidean with norm $|\cdot|$. Following Robinson [**15**], we say that a multifunction $F : Z \rightrightarrows X$ is *locally upper Lipschitzian* at z^* if there is a modulus $\lambda \geq 0$ and a neighborhood U of z^* such that

$$F(z) \subset F(z^*) + \lambda \|z - z^*\| B, \tag{4.1}$$

where B is a unit ball in X.

Proposition 4.1. *Let $F : Z \rightrightarrows X$ be locally upper Lipschitzian and single-valued at a point $z^* \in \operatorname{int}(\operatorname{dom} F)$, with $F(z^*) = \{x^*\}$. If the contingent derivative $F^+_{z^*}$ is almost-surely single-valued with respect to a given probability measure P on Z, then F is P-a.s. semi-differentiable at z^* relative to x^* and, moreover, $F^+_{z^*} = F'_{z^*,x^*}$, P-a.s.*

Proof. Let D_t be the difference quotient multifunction

$$D_t(z) = t^{-1}[F(z^* + tz) - x^*], \qquad t > 0.$$

Let z be a point where $F^+_{z^*,x^*}$ is single-valued. We want to show that

$$\lim_{\substack{t \downarrow 0 \\ z' \to z}} D_t(z') = F^+_{z^*}(z). \tag{4.2}$$

Since $z^* \in \operatorname{int}(\operatorname{dom} F)$ it follows that $D_{t_\nu}(z_\nu)$ is eventually nonempty for any sequence $t_\nu \downarrow 0$ and $z_\nu \to z$. The locally upper Lipschitzian property also implies that eventually

$$D_{t_\nu}(z_\nu) \subset \lambda \|z\| B. \tag{4.3}$$

Thus any sequence $x_\nu \in D_{t_\nu}(z_\nu)$ is eventually bounded, has limit points, and these limit points must be in $F^+_{z^*}(z)$ by definition of the contingent derivative (3.3). Now, by our assumption, $F^+_{z^*}(z)$ is a singleton, say $\{x\}$, and we have in fact shown that $\lim x_\nu = x$ for all sequences $x_\nu \in D_{t_\nu}(z_\nu)$, all $t_\nu \downarrow 0$ and $z_\nu \to z$. This proves (4.2), and the conclusions of the proposition follow. □

Note that we do not require F itself to be P-a.s. single-valued, just $F^+_{z^*}$. This is mostly a vacuous generality since F could not be too far from being P-a.s. single-valued in small neighborhoods of z^*, but it affords a useful flexibility in some applications. An immediate consequence of the above is a delta theorem for selections of F. We first deal with a technical lemma.

Lemma 4.2. *Let $F : Z \rightrightarrows X$ be locally upper Lipschitzian and single-valued at a point $z^* \in \operatorname{int}(\operatorname{dom} F)$. Then*

$$Z = \operatorname{dom} F^+_{z^*}. \tag{4.4}$$

Proof. Let $z \in Z$ be given. For all sufficiently small t we have $z^* + tz \in \operatorname{dom} F$, since $z^* \in \operatorname{int}(\operatorname{dom} F)$, and $D_t(z) \subset \lambda\|z\|B$, since F is locally upper Lipschitzian. Hence there are points $x_t \in D_t(z)$ for all sufficiently small t and at least one limit point of $\{x_t\}$, which by definition must belong to $F^+_{z^*}(z)$. Therefore $z \in \operatorname{dom} F^+_{z^*}$, proving (4.3). □

Theorem 4.3. *Let $\{\mathbf{z}_\nu\}$ be a sequence of random variables in Z satisfying a generalized central limit theorem with limit z^* and asymptotic distribution $\mathfrak{z}$. Let $F : Z \rightrightarrows X$ be a closed-valued measurable multifunction that is locally upper Lipschitzian and single-valued at z^*, with $F(z^*) = \{x^*\}$. Suppose further that:*

$$z^* \in \operatorname{int}(\operatorname{dom} F); \tag{4.5}$$

and

$$F^+_{z^*}(\mathfrak{z}) \text{ is a.s. single-valued;} \tag{4.6}$$

Then for all measurable selections $\mathbf{x}_\nu \in F(\mathbf{z}_\nu)$ and $\mathfrak{x} \in F^+_{z^*,x^*}(\mathfrak{z})$ one has

$$\tau_\nu^{-1}[\mathbf{x}_\nu - x^*] \xrightarrow[\mathcal{D}]{} \mathfrak{x} \tag{4.7}$$

as random variables in X.

Proof. From assumptions (4.5) and (4.6) and Proposition 4.1, it follows that F is a.s. semi-differentiable at z^* relative to x^* with respect to the measure induced by $\mathfrak{z}$. Hence

Theorem 3.2 applies and $\tau_\nu^{-1}(F(\mathbf{z}_\nu)-x^*]\xrightarrow[\mathcal{D}]{}F^+_{z^*}(\mathfrak{z})$. The conclusion will follow from Theorem 2.3. From (4.3) we have that eventually $\tau_\nu^{-1}|\mathbf{x}_\nu - x^*| \le \lambda\tau_\nu^{-1}\|\mathbf{z}_\nu - z^*\|$, and this latter sequence is tight. The only thing that remains is to show the counterpart of (2.5). Let μ be the probability measure corresponding to $\mathfrak{z}$. By Lemma 4.2, we have $\mu(\operatorname{dom} F^+_{z^*}) = 1$; therefore, (2.5) follows from Proposition 2.5. □

Example 4.4. The statement of Theorem 4.3 is particularly useful in investigations of limit behaviour of solution sequences in statistical estimation and stochastic programming, as described above in Example 1.3. There is no space here to give a full treatment; we merely indicate the possibilities by an illustrative example (that can certainly be treated by existing parametric theories — for example, Shapiro [**19**]). Consider the constrained least squares problem

$$(\mathrm{Q}_\nu)\qquad \begin{array}{ll} \text{minimize} & \displaystyle\frac{1}{\nu}\sum_{i=1}^{\nu}\frac{1}{2}|x-\mathbf{z}_i|^2 \\ \text{subject to} & x\in C \end{array}$$

where $\{\mathbf{z}_\nu\}$ is an independent and identically distributed sequence of random variables in $\mathbb{R}^n$ that satisfies a central limit theorem, and where C is a polyhedral convex subset of $\mathbb{R}^n$. (C is given by a finite number of linear inequalities.) The optimal solution $\mathbf{x}_\nu$ to (Q_ν) is unique and satisfies

$$0\in \mathbf{x}_\nu - \frac{1}{\nu}\sum_{i=1}^{\nu}\mathbf{z}_i + N_C(\mathbf{x}_\nu),$$

where $N_C(\cdot)$ is the normal cone multifunction of convex analysis. The mapping $F(z)$ given by

$$F(z)=\{x\,|\,0\in x-z+N_C(x)\}$$

is locally upper Lipschitzian by Robinson [**15**], and it clearly has $\operatorname{dom} F = \mathbb{R}^n$. Furthermore, it is proto-differentiable by Rockafellar [**16**; 5.6], and we have, for any $(z^*,x^*)\in \operatorname{gph} F$,

$$F^+_{z^*}(z)=\{\xi\in C'_{z^*}\,|\,0\in\xi-z+N_{C'_{z^*}}(\xi)\}\tag{4.8}$$

where

$$C'_{z^*}=\{\xi\in T_C(x^*)\,|\,\xi{\cdot}[x^*-z^*]=0\}$$

and $T_C(\cdot)$ is the tangent cone of convex analysis. It is easy to see that $x\in F^+_{z^*}(z)$ if and only if x is the solution to

$$\begin{array}{ll} \text{minimize} & \displaystyle\frac{1}{2}|x-z|^2 \\ \text{subject to} & x\in T_C(x^*) \\ & [x^*-z^*]{\cdot}x=0, \end{array}$$

therefore $F^{+}_{z^*}$ is single-valued everywhere.

From Theorem 4.3, it follows that the sequence of least squares estimates satisfy the asymptotic formula

$$\sqrt{\nu}[\mathbf{x}_\nu - x^*] \xrightarrow[\mathcal{D}]{} \mathfrak{x},$$

where x^* is the solution to the "true" problem

$$\text{(Q)} \qquad \begin{array}{rl} \text{minimize} & \frac{1}{2}E|x - \mathbf{z}_1|^2 \\ \text{subject to} & x \in C \end{array}$$

and where the asymptotic law $\mathfrak{x}$ is the solution to a *random quadratic program*

$$\text{(Q}'\text{)} \qquad \begin{array}{rl} \text{minimize} & \frac{1}{2}|x - \mathfrak{z}|^2 \\ \text{subject to} & x \in T_C(x^*) \\ & [x^* - E\mathbf{z}_1]\cdot x = 0, \end{array}$$

with $\mathfrak{z} \sim N(0, \operatorname{cov} \mathbf{z}_1)$. This result is a precursor to an asymptotic theory for solutions to convex stochastic optimization problems along the lines sketched in Example 1.3.

Appendix

In this appendix, we record some specifics about measures induced by a closed-valued measurable multifunction from a given probability space $(\Omega, \mathcal{A}, P)$ into a complete, separable metric space X with Borel sets B, when the multifunction is almost surely single-valued. Our interest is in when such a multifunction can be interpreted as inducing a measure on X itself.

Recall that a closed-valued multifunction $F : \Omega \rightrightarrows X$ is *measurable* if $F^{-1}(C) \in \mathcal{A}$ for all closed subsets $C \subset X$. Using Castaing and Valadier [**5**; III.30], this implies that $F^{-1}(B) \in \mathcal{A}$ for all Borel subsets $B \in \mathcal{B}$ provided we assume $\mathcal{A}$ is *complete* relative to the measure P (includes all sets of P-measure zero). Thus PF^{-1} is a *set-function* on $\mathcal{B}$. Our first result determines when this set function is a *measure* on $\mathcal{B}$.

Proposition A.1. *Suppose $F : \Omega \rightrightarrows X$ is closed-valued and measurable. Then PF^{-1} is a measure on $(X, \mathcal{B})$ if and only if*

$$P\{F^{-1}(B) \cap F^{-1}(A)\} = 0 \tag{A.1}$$

for every $A, B \in \mathcal{B}$ with $A \cap B = \emptyset$.

Proof. It is the requirement of additivity of a measure that necessitates (A.1). Indeed, it is trivial that if $A \cap B = \emptyset$, then $PF^{-1}(A \cup B) = PF^{-1}(A) + PF^{-1}(B)$ if and only if condition (A.1) holds. To complete the proof we need only demonstrate that (A.1) implies countable additivity of PF^{-1}. Let A_n, $n = 1, 2, \ldots$, be a sequence of pairwise disjoint sets in $\mathcal{B}$. Define B_n, $n = 1, 2, \ldots$, by

$$\begin{aligned} B_1 &= F^{-1}(A_1), \\ B_2 &= F^{-1}(A_2) \setminus B_1, \text{ etc.,} \end{aligned}$$

and then

$$PF^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n)$$

by the countable additivity of P. Now note that $B_n \subset F^{-1}(A_n)$ for every n, and furthermore that

$$F^{-1}(A_n) = B_n \cup [F^{-1}(A_n) \cap B_{n-1}] \subset B_n \cup [F^{-1}(A_n) \cap F^{-1}(A_{n-1})].$$

Hence

$$P(B_n) \le PF^{-1}(A_n) \le P(B_n) + P\{F^{-1}(A_n) \cap F^{-1}(A_{n-1})\},$$

but this last term is zero by our assumption (A.1). Therefore $P(B_n) = PF^{-1}(A_n)$, $n = 1, 2, \ldots$, and we conclude from this that PF^{-1} is countably additive. $\square$

If a multifunction $F : \Omega \rightrightarrows X$ is single-valued almost surely, then condition (A.1) will be satisfied—since when A and B are disjoint, any element of $F^{-1}(A) \cap F^{-1}(B)$ will be a point where F is not single-valued. Thus PF^{-1} is a measure on $\mathcal{B}$ when F is a.s. single-valued. It turns out that the converse is also true, but first let us review a fundamental concept. If $F : \Omega \rightrightarrows X$ is closed-valued and measurable, then it is well-known that there exists a *Castaing representation* for F — a countable family $\{x_i : i = 1, 2, \ldots\}$ of measurable functions, $x_i : \operatorname{dom} F \to X$, such that

$$F(\omega) = \operatorname{cl}\{x_i(\omega) \,|\, i = 1, 2, \ldots\} \text{ for all } \omega \in \operatorname{dom} F;$$

cf. [**5** ; III.30]. This representation of F is very useful in measure theoretic arguments concerning set-valued maps.

Theorem A.2. *Let $F : \Omega \rightrightarrows X$ be closed-valued and measurable. Then PF^{-1} is a measure on $(X, \mathcal{B})$ if and only if*

$$P\{\omega \in \operatorname{dom} F \,|\, F(\omega) \text{ is not single-valued}\} = 0. \tag{A.2}$$

Proof. It remains only to show the necessity of (A.2). Let us characterize the set $M \subset \Omega$ where F is not single-valued as follows. Let Q be the rationals in $\mathbb{R}$, and let D be a countable dense subset of X. Denote by $B(d; r)$ the closed ball with center d and radius r in X. Then

$$M = \bigcup_{d \in D} \bigcup_{r \in Q} \bigcup_{i=1}^{\infty} \bigcup_{j \geq i} \{\omega \in \operatorname{dom} F \,|\, x_i(\omega) \in B(d; r) \text{ and } x_j(\omega) \notin B(d; r)\},$$

where $\{x_i\}$ is the Castaing representation for F. It follows that M is a measurable subset of Ω. Furthermore, $P(M) > 0$ if and only if there is some quadruple (d, r, i, j) with

$$P\{\omega \in \operatorname{dom} F \,|\, x_i(\omega) \in B(d; r) \text{ and } x_j(\omega) \notin B(d; r)\} > 0.$$

If this is the case, then clearly

$$P\{F^{-1}(B(d; r)) \cap F^{-1}(B^c(d; r))\} > 0,$$

which contradicts (A.1). Thus $P(M) = 0$ and the proof is complete. $\square$

The existence of a Castaing representation is crucial in establishing that the set M of points where F is not single-valued is measurable. When this set has P-measure zero, then it is clear that the x_i differ from F only on the set M and hence they all induce the same measure Px_i^{-1}. In fact, if f is any selection of F we have

$$f^{-1}(A) = [F^{-1}(A) \cap M^c] \cup [f^{-1}(A) \cap M], \quad A \in \mathcal{B},$$

since $f = F$ on M^c (complementation is taken relative to $\operatorname{dom} F$) and M is measurable. Thus f is measurable, and

$$Pf^{-1}(A) = P\{F^{-1}(A) \cap M^c\} = PF^{-1}(A);$$

and we have proved the following corollary.

Corollary A.3. *Let $F : \Omega \rightrightarrows X$ be a closed-valued measurable multifunction satisfying condition (A.2). Then any selection f of F is P-measurable and $Pf^{-1} = PF^{-1}$.* □

Acknowledgements.

The author is indebted to Professors R.T. Rockafellar and R. J-B Wets for the many conversations that sparked crucial ideas in the evolution of this research. Thanks also are due to the patient referees for their careful reading of two versions of this paper and for their useful comments: in particular, the proof of Theorem A.2 was suggested by an anonymous referee.

References.

[1] J. Aitchison and S.D. Silvey, "Maximum likelihood estimation of parameters subject to restraints", *Ann. Math. Stat.* **20**(1948), 813–828.

[2] J-P. Aubin, "Lipschitz behavior of solutions to convex minimization problems",*Mathematics of Operations Research* **9** (1984) 97–102.

[3] P. Billingsley, *Convergence of Probability Measures*, Wiley, 1968.

[4] D.D. Boos and R.J. Serfiling, "A note on differentials and the CLT and LIL for statistical functionals, with application to M-estimates", *Annals of Statistics* **8** (1980) 618–624.

[5] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Springer-Verlag Lecture Notes in Mathematics no. 580, 1977.

[6] B.R. Clarke, "Nonsmooth analysis and Fréchet differentiability of M-functionals", *Probab. Th. Rel. Fields* **73** (1986) 197–209.

[7] J. Dupačová and R.J-B. Wets, "On the asymptotic behaviour of constrained estimates and optimal decisions", to appear in *Ann. of Math. Stat.* (1988).

[8] P.J. Huber, *Robust Statistics*, Wiley, 1981.

[9] G. Kallianpur, "Von Mises functionals and maximum likelihood estimation", *Sankhya Ser. A* **25** (1963) 149–158.

[10] A.J. King, *Asymptotic Behaviour of Solutions in Stochastic Optimization: Nonsmooth Analysis and the Derivation of Non-Normal Limit Distributions*, Dissertation, University of Washington, 1986.

[11] A.J. King, "Central limit theory for multivalued mappings", Working Paper WP-87-112, International Institute for Applied Systems Analysis, 1987.

[12] A.J. King, "Central limit theory for Lipschitz mappings", Working Paper WP-87-127, International Institute for Applied Systems Analysis, 1987.

[13] A.J. King, "Asymptotic distributions for solutions in stochastice optimization and generalized M-estimation", Working Paper WP-88- , International Institute for Applied Systems Analysis, 1988.

[14] H.B. Mann and A. Wald, "On stochastic limit and order relationships", *Ann. of Math. Stat.* **14** (1943) 217–226.

[15] S.M. Robinson, "Some continuity properties of polyhedral multifunctions", *Math. Prog. Study* **14**(1981), 206–214.

[16] R.T. Rockafellar, "Proto-differentiability of set-valued mappings and its applications in optimization", to appear in *Ann. Inst. H. Poincare: Analyse Non Lineaire*, 1988.

[17] G. Salinetti and R.J-B. Wets, "On the convergence of closed-valued measurable multifunctions", *Trans. A.M.S.* **266**(1)(1981), 275–289.

[18] G. Salinetti and R.J-B. Wets, "On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima", *Math. of O.R.* **11**(3)(1986), 385–419.

[19] A. Shapiro, "Second order sensitivity analysis and asymptotic theory of parametrized nonlinear programs", *Math. Prog.* **33**(1985).