International Institute for Applied Systems Analysis
IIASA
www.iiasa.ac.at

Central Limit Theory for Lipschitz Mappings

King, A.J.

IIASA Working Paper

WP-87-127

December 1987

WORKING PAPER

CENTRAL LIMIT THEORY FOR LIPSCHITZ MAPPINGS

Alan J. King

December 1987
WP-87-127

IIASA
International Institute for Applied Systems Analysis

CENTRAL LIMIT THEORY FOR LIPSCHITZ MAPPINGS

Alan J. King

December 1987
WP-87-127

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

FOREWORD

Central limit theorems are derived for mappings that are Lipschitzian at a given point. This theory results from a new perspective on first-order behaviour—the upper pseudo-derivative, the graph of which is the contingent cone to the graph of the mapping at a given point. We adopt the general setting of the convergence in distribution of measures induced by mappings that may be multi-valued on sets of measure zero. By requiring the upper pseudo-derivative to be single-valued a.s., we obtain a central limit theorem under distinctively weaker conditions than classical Fréchet differentiability.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

CENTRAL LIMIT THEORY FOR LIPSCHITZ MAPPINGS

Alan J. King

1. Introduction

This paper addresses the question: if $f : Z \to \mathbb{R}^n$ is a mapping and $\{\mathbf{z}_N\}$ is a sequence of random variables in Z, a Banach space, that satisfies a central limit formula

$$\sqrt{N}(\mathbf{z}_N - \overline{z}) \to \mathfrak{z} \quad \text{in distribution,} \tag{1.1}$$

then under what conditions does there exist a mapping $\varphi(\cdot)$ such that

$$\sqrt{N}(f(\mathbf{z}_N) - f(\overline{z})) \to \varphi(\mathfrak{z}) \quad \text{in distribution?} \tag{1.2}$$

Obviously if f is Fréchet differentiable one has

$$|f(\overline{z} + h) - f(\overline{z}) - f'(\overline{z}; h)| = o(\|h\|), \tag{1.3}$$

where $f'(\overline{z}; \cdot)$ is the linear mapping we call the *derivative*, and then (1.2) follows with limiting distribution $f'(\overline{z}; \mathfrak{z})$. Our goal is to derive conditions yielding (1.2) which are more attuned to the underlying convergence theory and which are applicable to the sorts of mappings that arise naturally in optimization theory.

Consider the following simple example. Let $\{\mathbf{z}_n : n = 1, 2, \ldots\}$ be independent, identically distributed versions of a normal random variable with known mean μ and variance σ^2. For each $N = 1, 2, \ldots$, let $\mathbf{x}_N$ be the solution to the problem:

$$\text{minimize } \frac{1}{N} \sum_{n=1}^{N} |x - \mathbf{z}_n|^2 \quad \text{over all } x \geq \mu. \tag{1.4}$$

The asymptotic distribution of $\{\mathbf{x}_N\}$ is easily computed from the explicit formula

$$\mathbf{x}_N = \max\left\{\mu, \frac{1}{N} \sum_{n=1}^{N} \mathbf{z}_n\right\} := f\left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{z}_n\right); \tag{1.5}$$

it consists of an atom of value $\frac{1}{2}$ at zero combined with the right half of a normal distribution with variance σ^2. The mapping $f(\cdot)$ defined by (1.5) is not differentiable at $z = \mu$. A closer examination reveals that f is *directionally differentiable* there,

$$f'(\mu; h) = \begin{cases} 0 & \text{if } h \leq 0, \\ h & \text{if } h > 0, \end{cases}$$

and the asymptotic distribution is indeed described by $f'(\mu; \mathfrak{z})$, where $\mathfrak{z} \sim N(0, \sigma^2)$ is the asymptotic distribution of the sequence of sample means. In fact it is true that $f'(\mu; \cdot)$ satisfies (1.3), as can be directly verified from the explicit representation (1.5). Directional differentiability is a more natural property of mappings arising from optimization theory—but there are many varieties of directional derivatives, and (1.3) is one of the more restrictive properties.

Our approach to establishing the asymptotic behaviour is not through (1.3), but rather through a fundamental examination of the convergence in distribution of the *difference quotients*

$$\Delta_{\frac{1}{\sqrt{N}}}(\bar{z}; \sqrt{N}(\mathbf{z}_N - \bar{z})) := \frac{f(\bar{z} + \frac{1}{\sqrt{N}} \cdot \sqrt{N}(\mathbf{z}_N - \bar{z})) - f(\bar{z})}{1/\sqrt{N}}. \tag{1.6}$$

We extend the classical theory of convergence in distribution for sequences of such mappings and we are able to characterize the limiting distribution of (1.2), if it exists, as $f^+_{\bar{z}}(\mathfrak{z})$, where $f^+_{\bar{z}}(\cdot)$ is the *upper pseudo-derivative* of the mapping $f(\cdot)$ at $\bar{z}$. This new object is defined to be that mapping whose *graph*, denoted $\operatorname{gph} f^+_{\bar{z}}(\cdot)$, is the *contingent cone* to the graph of the mapping $f(\cdot)$ at $\bar{z}$, i.e.

$$\operatorname{gph} f^+_{\bar{z}}(\cdot) = \limsup_{t \downarrow 0} t^{-1}[\operatorname{gph} f(\cdot) - (\bar{z}, f(\bar{z}))]. \tag{1.7}$$

The limit is to be understood as a limit of *sets* in $Z \times \mathbb{R}^n$ (see (1.11) below). Mappings given by graph limits of the sort described by (1.7) are not necessarily functions. The "value" of $f^+_{\bar{z}}$ at a point $h \in Z$ may contain one point, several points, entire subspaces, or no points at all—in general all one can say is that $f^+_{\bar{z}}$ is a *subset* of $\mathbb{R}^n$. Such mappings are called *multifunctions*; they have long been familiar to students of optimization theory.

Thus new questions are raised: if $f^+_{\bar{z}}$ is not a function then how are we to interpret $f^+_{\bar{z}}(\mathfrak{z})$ as the limit of (1.2) in the sense of convergence of distributions in $\mathbb{R}^n$? To answer this question we must determine when such a multifunction gives rise to a distribution—it turns out that this is the case if and only if the multifunction is single-valued almost surely. This can be interpreted as a *differentiability condition* that f must satisfy in order for (1.2) to hold; it corresponds, as we shall see, to almost sure directional Hadamard differentiability at $\bar{z}$, i.e. the limit

$$\lim_{\substack{t \downarrow 0 \\ h' \to h}} \frac{f(\bar{z} + th') - f(\bar{z})}{t}$$

exists for almost all h. This is a generalization of Fréchet differentiability (1.3) but more importantly it is a generalization which grows naturally from the underlying probability theory.

Having introduced the machinery of multifunctions to analyze the distribution induced by the upper pseudo-derivative, we may as well widen our scope by permitting f itself to be multivalued on sets of measure zero. This additional flexibility is invaluable. Many situations in optimization theory give rise to multifunctions. The solution mapping to a parametric optimization problem is generally multivalued, but under natural regularity conditions turns out to be Lipschitzian (Robinson [**14**]) and single-valued almost everywhere (Rockafellar [**16**]). Therefore we shall adopt the following more general setting: to analyze the asymptotic behaviour of *selections* $\mathbf{x}_N \in \mathbb{R}^n$ such that

$$\mathbf{x}_N \in F(\mathbf{z}_N), \quad N = 1, 2, \ldots, \tag{1.8}$$

where the sequence $\{\mathbf{z}_N\}$ is asymptotically normal (1.1) in the Banach space Z and F is a closed-valued, measurable multifunction that is single-valued and Lipschitzian at $\overline{z} \in \operatorname{int}\operatorname{dom} F$, i.e. we wish to determine the properties of the asymptotic distribution $\mathfrak{X}$, if one exists, for which (with $\overline{x} = F(\overline{z})$) we have

$$\sqrt{N}(\mathbf{x}_N - \overline{x}) \xrightarrow[D]{} \mathfrak{X}. \tag{1.9}$$

Everything developed in this framework applies also to the case where $F = f$, a measurable function that is Lipschitzian at $\overline{z}$; and, as we have noted, the machinery of multifunctions is required even then.

We begin in Section 2 with a review of the standard concepts of measurability for closed-valued multifunctions, using Rockafellar [**15**] as a basic reference, and then establish necessary and sufficient conditions that determine when closed-valued, measurable multifunctions give rise to distributions on the range space $\mathbb{R}^n$—this new theorem depends on certain properties of analytic sets as developed in Meyer [**12**] and the Castaing representation of a closed-valued measurable multifunction. To treat the convergence of the sequence (1.6) we proceed, in Section 3, to explore the fundamentals of convergence of distributions induced by mappings. The crucial insight is achieved through a re-examination of some classical material in Billingsley [**4**], and especially through the analysis of the mysterious exceptional set E that appears in his Theorem 5.5 (attributed to H. Rubin). Finally, in Section 4, we apply this insight to develop the main result of the paper—the identification of the upper pseudo-derivative as the limiting distribution of (1.9) under certain conditions that must be satisfied by F and its upper pseudo-derivative, in particular that F be Lipschitzian at $\overline{z}$ and $F^{+}_{\overline{z},\overline{x}}(\mathfrak{z})$ be single-valued almost surely. To complete our investigation we then show that in case $F = f$, a function, these conditions amount to a Hadamard directional differentiability condition that must hold for almost all directions.

Some of the results presented here are from the author's dissertation [**10**], in which this program was developed in complete detail for the asymptotic analysis of solutions to stochastic optimization problems. There, the central limit theorem of Section 4 was applied to selections $\mathbf{x}_N$ from the mapping

$$F(\mathbf{z}_N) = \{x \in \mathbb{R}^n : 0 \in \mathbf{z}_N(x) + N_X(x)\}$$

where $\mathbf{z}_N(\cdot)$ is the gradient of the objective function

$$\frac{1}{N}\sum_{n=1}^{N} f(\cdot, \mathbf{s}_n)$$

and $N_X(\cdot)$ is the normal cone to the constraint set X. Thus $\{\mathbf{x}_N\}$ is the sequence of solutions to the constrained optimization problems

$$\text{minimize } \frac{1}{N}\sum_{n=1}^{N} f(x, \mathbf{s}_n) \text{ over all } x \in X,$$

and we ask: In what sense does x_N approximate the "true" solution $\bar{x}$ that minimizes $Ef(x, \mathbf{s})$ over all $x \in X$? It is for the analysis of such sequences that the techniques and ideas introduced in the present paper were developed. The asymptotic distribution of the gradient estimates $\{\mathbf{z}_N(\cdot)\}$ is readily computed as a distribution over $C(X : \mathbb{R}^n)$. Then the rapidly developing theory of pseudo-derivatives is applied to the mapping F to achieve, via the main result of the present paper, an explicitly computable description of the asymptotic distribution of the solution estimates $\{\mathbf{x}_N\}$; it turns out, of course, that this distribution generally is not normal just as in the simple constrained least squares example above. We plan to report these results in future papers.

The key role of the upper pseudo-derivative is the aspect in which our theory is "attuned" to the needs of optimization theory. The pseudo-derivative is a powerful new concept in optimization—it is linked to the fundamental circle of ideas centered around the epi-convergence of convex functions (besides [**10**], see also Rockafellar [**18**], [**19**] and [**20**] for more on pseudo-derivatives). Our work here has discovered the importance of the upper pseudo-derivative in matters close to the heart of statistical theory. We anticipate that this surprising coincidence will eventually be viewed as yet another chapter in the exploration of the theory of epi-convergence initiated by Wijsman [**30**], and subsequently developed by others, for example Wets [**29**].

The problems raised by stochastic optimization stimulated this research. As pointed out in [**10**], the theory of maximum likelihood estimation raises similar issues; however such work has almost always emphasized asymptotic *normality*, which in turn relies on differentiability. In maximum likelihood estimation attention has been focussed on the solution as a mapping from the space of empirical distributions topologized by the Prohorov metric; see von Mises [**13**] and the more recent papers of Boos and Serfling [**5**] and Clarke [**6**]. An alternative and more flexible point of view was taken by Huber [**9**]. All of these eventually rely on Fréchet differentiability to establish the asymptotic behaviour. (But we should note that the directional derivative makes a brief appearance in Huber [**8**].) Dupačová and Wets [**7**] applied epiconvergence concepts to obtain consistency and then Huber's approach in [**9**] to obtain asymptotic normality in the stochastic optimization setting—where the role of constraints is emphasized. Constrained maximum likelihood estimation was explored by Aitchison and Silvey [**1**]; again, differentiability was crucial. Finally, Shapiro [**25**] examined the asymptotic behaviour of solution mappings for parametric optimization problems. In each of these areas the results of this paper may be immediately applied to yield conclusions about asymptotic behaviour under conditions where strong differentiability conditions such as (1.3) cannot, or should not, be assumed.

¿From a broader point of view, our work here fits into a tradition of analysis that uses correspondences between the closed-valued measurable multifunctions and certain functions that map into spaces on which distributions may be defined. Artstein [**3**] studied the correspondence $F \leftrightarrow \sigma_F$, where $\sigma_F(z)$ is the support function of $F(z)$:

$$\sigma_F(z)(y) = \sup_{x \in F(z)} x \cdot y.$$

Under this correspondence F induces, via σ_F, a distribution on the space of continuous functions on the unit ball in $\mathbb{R}^n$, and a limit theorem of Weil [**28**] may be used to analyze the asymptotic behaviour. Salinetti and Wets [**24**] developed a comprehensive treatment of convergence in distribution employing the function γ_F, where

$$\gamma_F(z) = \{F(z)\}$$

is to be considered as an element of the power set $2^{\mathbb{R}^n}$, which is equipped with the topology of the Haussdorff metric. The above approaches render conclusions that are indirect, abstract, and difficult to apply to selections; in contrast our approach is specifically designed to apply directly to the study of the asymptotic behaviour of such selections.

A correspondence that allows the treatment of multifunctions that are not single-valued, but which delivers useful information about selections, is given by the one-sided Haussdorff metric

$$h(z \mid F, \bar{z}) = \sup_{x \in F(z)} \operatorname{dist}(x \mid F(\bar{z})).$$

When F is Lipschitzian, $h(\cdot \mid F, \bar{z})$ is a Lipschitz function and can be analyzed within the framework developed here. The asymptotic distribution can be used to approximate the distance of $\mathbf{x}_N$ from $F(\bar{z})$, since

$$\operatorname{dist}(\mathbf{x}_N \mid F(\bar{z})) \leq h(\mathbf{z}_N \mid F, \bar{z}).$$

This approach was suggested to us by Professor R.J-B. Wets; it will be the subject of a future paper.

Let us take the opportunity here to review a notion that will be fundamental in the development to follow. For a sequence $\{B_k\}$ of subsets of a topological space we define

$$\limsup_{k\to\infty} B_k = \{b \mid \exists \text{ subsequence } \{k_n\},\ b_n \in B_{k_n} \text{ with } b_n \to b\} \tag{1.10}$$

$$\liminf_{k\to\infty} B_k = \{b \mid \exists b_k \to b \text{ with } b_k \in B_k \text{ for all sufficiently large } k\}; \tag{1.11}$$

and when these are equal to the same set B, we say B is the "limit", denoted $B = \lim_{k\to\infty} B_k$. These definitions and many properties thereof may be found in Kuratowski [**11**]. See also Salinetti and Wets [**22**] and [**23**]. We shall also need the limit of sets indexed by $[t \downarrow 0]$, as in (1.7). This notion is captured by the general concept of sets indexed by *filters*, introduced in Rockafellar and Wets [**21**]. For our purposes we need only the following characterizations:

$$\limsup_{t\downarrow 0} A_t = \{a : \exists t_n \downarrow 0,\ a_n \in A_{t_n} \text{ with } a_n \to a\}; \tag{1.12}$$

$$\liminf_{t\downarrow 0} A_t = \{a : \forall t_n \downarrow 0,\ \exists a_n \in A_{t_n} \text{ with } a_n \to a\}; \tag{1.13}$$

and we need only note that these are *closed* sets. Details may be found in King [**10**, Ch. 1].

The crucial role of the upper pseudo-derivative in this investigation was discovered following a suggestion by Professor R.T. Rockafellar. In this and many other fruitful speculations, we gratefully acknowledge his contributions.

2. Measurable Multifunctions, Measures Induced by Multifunctions

This section determines when a given multifunction F defined on a measure space $(Z, \mathcal{Z}, P)$ induces a *measure* PF^{-1} on the image space $\mathbb{R}^n$.

Measurability properties of multifunctions taking values in $\mathbb{R}^n$ have been comprehensively treated in Rockafellar [**15**]. We begin by citing some facts from this reference. (Most of the results quoted here can be generalized beyond the finite-dimensional case; see the survey [**27**] by Wagner.) Let Z be a complete, separable metric space and $\mathcal{Z}$ its Borel σ-algebra.

Definition 2.1. A *multifunction*, $F : Z \rightrightarrows \mathbb{R}^n$, is a mapping for which $F(z)$ is in general a (possibly empty) subset of $\mathbb{R}^n$. We define also some associated concepts:

(i) $\operatorname{dom} F = \{z \in Z \,|\, F(z) \neq \emptyset\}$, the *domain* of F;

(ii) $\operatorname{gph} F = \{(z, x) \in Z \times \mathbb{R}^n \,|\, x \in F(z)\}$, the *graph* of F;

(iii) $F^{-1}(A) = \{z \in Z \,|\, F(z) \cap A \neq \emptyset\}$.

We say F is *closed-valued* if $F(z)$ is closed in $\mathbb{R}^n$, and we say F has *closed graph* if $\operatorname{gph} F$ is closed in $Z \times \mathbb{R}^n$.

Definition 2.2. A multifunction F is *measurable* if for all closed subsets $C \subset \mathbb{R}^n$ one has

$$F^{-1}(C) \in \mathcal{Z}.$$

Proposition 2.3. *Suppose F has closed graph. Then F is closed-valued and measurable.*

Proof. That F is closed-valued is trivial. By Rockafellar [**15**; Proposition 1A], F is measurable if and only if $F^{-1}(K) \in \mathcal{Z}$ for all *compact* subsets $K \subset \mathbb{R}^n$. Let K be compact in $\mathbb{R}^n$; we show that $F^{-1}(K)$ is closed in Z. Indeed, define the sequence $\{(z_n, x_n)\}$ with $x_n \in F(z_n) \cap K$, $n = 1, 2, \ldots$, and suppose $z_n \to \bar{z}$. Since the sequence $\{x_n\}$ is contained in K we may suppose, by passing to subsequences if necessary, that $x_n \to \bar{x}$ in K. But $(z_n, x_n) \to (\bar{z}, \bar{x})$ in $\operatorname{gph} F$; it follows therefore that $\bar{x} \in F(\bar{z}) \cap K$, i.e. $\bar{z} \in F^{-1}(K)$. □

The closed-valued measurable multifunctions satisfy a definition of measurability more akin to the usual notions of Borel measurable functions when the measurable space $(Z, \mathcal{Z})$ is *complete*. We shall need only the following specialized result.

Theorem 2.4. *Let P be a σ-finite measure on $(Z, \mathcal{Z})$ and let $\mathcal{Z}_P$ be the σ-algebra generated by all P-measurable subsets of Z (i.e. $\mathcal{Z} \subset \mathcal{Z}_P$ and if $A' \subset A \in \mathcal{Z}$ with $P(A) = 0$ then $A' \in \mathcal{Z}_P$). Suppose F is closed-valued and measurable. Then*

$$F^{-1}(B) \in \mathcal{Z}_P \text{ for all } B \in \mathcal{B};$$

where $\mathcal{B}$ is the σ-algebra of Borel subsets of $\mathbb{R}^n$.

Proof. The σ-algebra $\mathcal{Z}_P$ is complete and $\mathcal{Z} \subset \mathcal{Z}_P$. So $F^{-1}(C)$ belongs to the complete σ-algebra $\mathcal{Z}_P$ for all closed subsets $C \subset \mathbb{R}^n$. The result now follows from Rockafellar [**15**; Thm. 1E]. □

According to this theorem, F is measurable with respect to Borel sets of $\mathbb{R}^n$ whenever the measurable space is complete with respect to some σ-finite measure P. Our interest is in *measures* induced by multifunctions—in which case there is no loss of generality in assuming that $\mathcal{Z}$ is complete relative to P, i.e. that $\mathcal{Z}$ consists of the P-measurable subsets of Z.

The next order of business is to determine when a closed-valued measurable multifunction F gives rise to a measure on $\mathbb{R}^n$. Suppose P is a σ-finite measure on $(Z, \mathcal{Z})$. We define the set-function PF^{-1} on the Borel sets $\mathcal{B}$ by

$$PF^{-1}(B) = P\{z \in Z : z \in F^{-1}(B)\} \quad \text{for all } B \in \mathcal{B}. \tag{2.1}$$

The sets $F^{-1}(B)$ for $B \in \mathcal{B}$ all belong to the class of P-measurable sets, by Theorem 2.4, so this definition makes sense.

Proposition 2.5. *Suppose F is closed-valued and measurable, and let P be a σ-finite measure on $(Z, \mathcal{Z})$. Then PF^{-1} is a measure on $(\mathbb{R}^n, \mathcal{B})$ if and only if*

$$P\{F^{-1}(B) \cap F^{-1}(A)\} = 0 \tag{2.2}$$

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

Proof. It is the requirement of additivity of a measure that necessitates (2.2). Indeed, if PF^{-1} is a measure on $\mathcal{B}$ and $A, B \in \mathcal{B}$ with $A \cap B = \emptyset$ then

$$PF^{-1}(A) + PF^{-1}(B) = PF^{-1}(A \cup B);$$

on the other hand $F^{-1}(A \cup B) = F^{-1}(A) \cup F^{-1}(B)$, and since $F^{-1}(A)$ and $F^{-1}(B)$ are P-measurable then

$$PF^{-1}(A \cup B) = PF^{-1}(A) + PF^{-1}(B) - P\{F^{-1}(A) \cap F^{-1}(B)\}$$

which implies

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

To show sufficiency we must verify that (2.2) implies the set-function PF^{-1} is a measure. Observe that

$$F^{-1}(\emptyset) = \{z : F(z) \cap \emptyset \text{ is nonempty }\} = \emptyset,$$

hence $PF^{-1}(\emptyset) = 0$. It remains to show countable additivity; this follows from an elementary disjointing argument. 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

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

and then

$$PF^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) \tag{2.3}$$

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

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

but this last term is zero by our assumption (2.2). Therefore $P(B_n) = PF^{-1}(A_n)$, $n = 1, 2, \ldots$, and we conclude from this and (2.3) that PF^{-1} is countably additive. □

Let us examine condition (2.2) more closely. Notice that

$$F^{-1}(A) \cap F^{-1}(B) = \{z : F(z) \cap A \neq \emptyset \text{ and } F(z) \cap B \neq \emptyset\}.$$

If A and B are disjoint then any element of this set will be a point where F is not single-valued; hence if F is single-valued except on a set of P-measure zero then condition (2.2) will follow. It turns out that the converse is also true.

Theorem 2.6. *Let P be a σ-finite measure on $(Z, \mathcal{Z})$ and let the multifunction $F : Z \rightrightarrows \mathbb{R}^n$ be closed-valued and measurable. Then PF^{-1} is a measure on $(\mathbb{R}^n, \mathcal{B})$ if and only if*

$$P\{z \in \operatorname{dom} F \mid F(z) \text{ is not single-valued}\} = 0. \tag{2.3}$$

Proof. The preceding remarks established the sufficiency of (2.3). Necessity will follow from the *Castaing representation* for closed-valued measurable multifunctions and from certain properties of $\mathcal{B}$-*analytic* sets. Since F is closed-valued and measurable it follows that $\operatorname{dom} F$ is measurable and that there exists a Castaing representation for F—a countable family $\{x_i\}_{i \in I}$ of measurable functions, $x_i : \operatorname{dom} F \to \mathbb{R}^n$, such that

$$F(z) = \operatorname{cl}\{x_i(z) \mid i \in I\} \quad \text{for all } z \in \operatorname{dom} F; \tag{2.4}$$

cf. Rockafellar [**15**; Thm. 1B]. With such a representation we can characterize the set where F is not single-valued. Define the sequence of sets $M_n \subset \operatorname{dom} F$, $n \geq 2$, by

$$M_n = \bigcup_{k<n} \{z \in \operatorname{dom} F \mid x_n(z) - x_k(z) \neq 0\}; \tag{2.5}$$

these are all elements of $\mathcal{Z}$. From (2.4), F is not single-valued at z if and only if z is an element of M_n for some $n = 2, 3, \ldots$. It follows that the set M of points $z \in \operatorname{dom} \Gamma$ where F is not single-valued is given by

$$M = \bigcup_{n=1}^{\infty} M_n; \tag{2.6}$$

clearly M is a measurable subset of $\operatorname{dom} F$.

To prove the theorem we shall show that if $P(M) > 0$ then there are disjoint sets $B_1, B_2 \in \mathcal{B}$ such that

$$P\{F^{-1}(B_1) \cap F^{-1}(B_2)\} > 0,$$

which by the previous result, Proposition 2.5, will establish that PF^{-1} cannot be a measure on $(\mathbb{R}^n, \mathcal{B})$. Therefore, assume $P(M) > 0$. From (2.6) we then have $P(M_n) > 0$ for some n; and from (2.5) we then have

$$P\{z \in \operatorname{dom} F \mid x_n(z) - x_k(z) \neq 0\} > 0$$

for some $k \leq n$. We can renumber the sequence so that $k = 1$, $n = 2$; hence, without loss of generality,

$$P\{M_2\} > 0.$$

Therefore we have a set M_2 with positive measure and two selections x_1 and x_2 of F such that $x_1(z) \neq x_2(z)$ on M_2. We seek a further subset $N \subset M_2$, of positive measure, which satisfies $x_1(N) \cap x_2(N) = \emptyset$.

To that end, let $\{\varphi_k^1\}$ and $\{\varphi_k^2\}$ be sequences of simple functions that converge pointwise to x_1 and x_2, respectively. By an application of Egorov's Theorem we may suppose that there is a subset $M_2^0 \subset M_2$ with $P(M_2^0) > 0$, for which the convergence of both sequences is uniform on M_2^0 (we may assume without loss of generality that $P(M_2) < \infty$, since P is σ-finite) Passing to subsequences, if necessary, we may suppose that

$$\sup_{z \in M_2^0} |\varphi_k^1(z) - x_1(z)| < 1/k$$

and

$$\sup_{z \in M_2^0} |\varphi_k^2(z) - x_2(z)| < 1/k.$$

Next, note that there must be at least one k for which there exists a subset $N \subset M_2^0$, of positive measure, such that

$$\inf_{z \in N} |\varphi_k^1(z) - \varphi_k^2(z)| > 2/k$$

(since otherwise we would have $\varphi^1 \to \varphi^2$ pointwise, hence $x_1 = x_2$, on M_2^0). On this set N, the simple function φ_k^1 assumes finitely many values. Without loss of generality we may suppose that $\varphi_k^1(z) \equiv f_1$, a constant, on N. Now putting all this together, we have a subset $N \subset M_2$ with $P(N) > 0$ and

$$\sup_{z \in N} |x_1(z) - f_1| < 1/k \text{ but } \inf_{z \in N} |x_2(z) - f_1| > 1/k.$$

By construction, $x_1(N)$ and $x_2(N)$ are disjoint subsets of $\mathbb{R}^n$, furthermore these are $\mathcal{B}$-analytic sets, according to Meyer [**12**, Thm. 13]. Then, by the separation theorem, Meyer [**12**; Thm. 14], there exist disjoint subsets $B_1, B_2 \in \mathcal{B}$ such that

$$B_1 \supset x_1(N) \text{ and } B_2 \supset x_2(N).$$

Now we have

$$F^{-1}(B_1) \cap F^{-1}(B_2) \supset x_1^{-1}x_1(N) \cap x_2^{-1}x_2(N) \supset N,$$

hence

$$P\{F^{-1}(B_1) \cap F^{-1}(B_2)\} \geq P(N) > 0. \qquad \square$$

We record for future reference the following observation made in the proof.

Corollary 2.7. *Let F be closed-valued and measurable. Then the sets*

$$S = \{z \mid F(z) \text{ is single-valued}\}, \text{ and}$$
$$M = \{z \in \operatorname{dom} F \mid F(z) \text{ is not single-valued}\}$$

are measurable subsets of dom F. $\square$

The theorem (and corollary) remain true when $(Z, \mathcal{Z})$ is an arbitrary measurable space and $\mathbb{R}^n$ is replaced by any complete, separable metric space; cf. Wagner [**27**; Thm. 4.2(d)] and the references to Meyer [**12**] cited above.

The importance of this theorem is that it completely characterizes when PF^{-1} can be studied as a *measure* on $(\mathbb{R}^n, \mathcal{B})$ in a manner that is directly verifiable in many applications. Condition (2.3) states that F is "almost" a function with respect to the measure P (or, more graphically speaking, that F is *thin* relative to P). In the following corollary we see that all selections f of F are P-measurable functions that give rise to the same distribution.

Corollary 2.8. *Let P and F be as in Theorem 2.6. Let*

$$f : Z \to \mathbb{R}^n$$

be any selection of F, i.e. $f(z) \in F(z)$ for all $z \in Z$. Then f is P-measurable and

$$Pf^{-1}(A) = PF^{-1}(A), \qquad A \in \mathcal{B}.$$

Proof. We have already noted that

$$M = \{z \in \operatorname{dom} F \mid F \text{ is not single-valued}\}$$

is a measurable subset of $\operatorname{dom} F$. Now

$$f^{-1}(A) = [F^{-1}(A) \cap M^c] \cup [f^{-1}(A) \cap M]$$

since $f = F$ on M^c (complementation is taken with respect to $\operatorname{dom} F$). The first set in this union is P-measurable by 2.4 and the second set is of P-measure zero by assumption. Hence f is P-measurable. Finally,

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

3. Convergence of Distributions Induced by Multifunctions

The starting point for the asymptotic analysis is a thorough re-examination of the weak convergence of the sequence $\{P_k F_k^{-1}\}$ where $\{P_k\}$ are measures on $(Z, \mathcal{Z})$, $P_k \xrightarrow[w]{} P$, and F_k map Z into $\mathbb{R}^n$. Our goal is to rework the classical result, emphasizing the role of the *graphs* of the mappings F_k, and in this way obtain a more precise and illuminating theorem that is directly applicable to the central limit theory presented in the next section.

First we review the fundamental concept of convergence of measures from Billingsley [**4**]. Let Z be a complete, separable metric space and $\mathcal{Z}$ the class of Borel subsets. All measures are assumed to be finite, hence regular [**4**, Thm. 1.1]. We shall need only the definition and the following theorem.

Definition 3.1. Let P, P_k, $k = 1, 2, \ldots$ be finite measures on $(Z, \mathcal{Z})$. We say P_k *converges weakly* to P, $P_k \xrightarrow[w]{} P$, provided

$$\int_Z g(z) P_k(dz) \to \int_Z g(z) P(dz)$$

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

A trivial adjustment to the argument in [**4**; Thm. 2.1] yields the following modification of the Portmanteau Theorem.

Theorem 3.2. *Let P and P_k, $k = 1, 2, \ldots$, be finite measures on Z satisfying $P_k(Z) \to P(Z)$. Then $P_k \xrightarrow[w]{} P$ if and only if*

$$\limsup_{k\to\infty} P_k(C) \le P(C)$$

for all closed $C \subset Z$. □

Now let $\{F_k\}$ be a fixed sequence of closed-valued, measurable multifunctions mapping Z into $\mathbb{R}^n$ and suppose that each F_k is almost surely single-valued relative to a given measure P_k. We ask—if $\{P_k\}$ converges weakly to a measure P then when is it true that $P_k F_k^{-1} \xrightarrow[w]{} PF^{-1}$, i.e. for which F? We begin with a reworking of the classical result (for functions), Billingsley [**4**, Thm. 5.5], generalizing it slightly to cover the multivalued case. Let F be a given closed-valued, measurable multifunction that is almost surely single valued relative to the measure P, and set

$$E = \{z \in Z \,|\, \exists z_n \to z, \exists \text{ subsequence } \{k_n\} \text{ and } \exists x_n \in F_{k_n}(z_n) \text{ such that } \{x_n\} \text{ has no cluster points in } F(z)\}.$$

Theorem 3.3. *Let $P_k \xrightarrow[w]{} P$ and suppose $P_k(\operatorname{dom} F_k) \to P(\operatorname{dom} F)$. If the exceptional set E has P-measure zero, then*

$$P_k F_k^{-1} \xrightarrow[w]{} PF^{-1}.$$

Proof. We shall apply the Portmanteau theorem. Note that $P_k(\operatorname{dom} F_k) = P_k F_k^{-1}(\mathbb{R}^n)$, hence we have $P_k F_k^{-1}(\mathbb{R}^n) \to PF^{-1}(\mathbb{R}^n)$ as required. Let C be an arbitrary closed subset of $\mathbb{R}^n$, we will show that $\limsup_{k\to\infty} P_k F_k^{-1}\{C\} \le PF^{-1}\{C\}$. Let us define the set

$$E_C := [\limsup_{k\to\infty} F_k^{-1}(C)] \setminus F^{-1}(C).$$

Then E_C is a measurable set since $F^{-1}(C)$ is a measurable set and $\limsup F_k^{-1}(C)$ is always a closed set. A more explicit description is

$$\begin{aligned} E_C &= \{z : \exists \{k_n\},\ z_n \to z \text{ with } z_n \in F_{k_n}^{-1}(C)\} \setminus F^{-1}(C) \\ &= \{z : \exists z_k \to z \text{ with } F_k(z_k) \cap C \neq \emptyset \text{ infinitely often but } F(z) \cap C = \emptyset\} \end{aligned}$$

We claim that $E_C \subset E$. Let $z \in Z \setminus E$, and suppose $z_k \to z$. If $F_k(z_k) \cap C = \emptyset$ for all but finitely many k then, vacuously, $z \in Z \setminus E_C$. On the other hand if $F_k(z_k) \cap C \neq \emptyset$ for infinitely many k, we may choose a subsequence $\{x_{k_n}\}$ with $x_{k_n} \in F_{k_n}(z_{k_n}) \cap C$. Since z is not in E and $z_{k_n} \to z$ it follows that $\{x_{k_n}\}$ must have a limit point, say x, with $x \in F(z)$. But C is a closed set and $x_{k_n} \in C$, hence $x \in C$ also. Thus z is not in E_C, proving the claim.

By assumption $P\{E\} = 0$ and since E_C is a measurable subset of E, we have $P\{E_C\} = 0$. Hence

$$P\{\limsup_{k\to\infty} F_k^{-1}(C)\} \le PF^{-1}\{C\}. \tag{3.3}$$

¿From Kuratowski [**11**; 25.IV.8]

$$\limsup_{k\to\infty} F_k^{-1}(C) = \bigcap_{k=1}^{\infty} \text{cl}\left[\bigcup_{\ell\geq k} F_\ell^{-1}(C)\right].$$

For convenience we let $B = \limsup F_k^{-1}(C)$ and

$$B_k = \text{cl}\left[\bigcup_{\ell\geq k} F_\ell^{-1}(C)\right].$$

The sequence of closed sets $\{B_k\}$ decreases to B. Now we argue exactly as in Billingsley, cited above. For any $\varepsilon > 0$ we have for all sufficiently large k that

$$P\{B\} + \varepsilon \geq P\{B_k\}.$$

Since $P_\ell \underset{w}{\longrightarrow} P$ and B_k is closed, Theorem 3.2 yields

$$\limsup_{\ell\to\infty} P_\ell(B_k) \leq P(B_k).$$

Noting that $B_k \supset F_\ell^{-1}(C)$ for all sufficiently large ℓ, we have

$$\limsup_{\ell\to\infty} P_\ell F_\ell^{-1}(C) \leq P\{B\} + \varepsilon$$

for arbitrary $\varepsilon > 0$. This and (3.3) prove the theorem. □

The exceptional set E in this theorem breaks up into two parts: one concerning whether the *graph* of F is large enough, the other concerning local unboundedness of the sequence F_k.

Proposition 3.4. *Define the multifunction G by*

$$\text{gph}\, G = \limsup_{k\to\infty} \text{gph}\, F_k.$$

Then

$$E = \{z : G(z) \setminus F(z) \neq \emptyset\} \cup \{z : \exists z_n \to z,\ x_n \in F_{k_n}(z_n) \text{ with } |x_n| \to +\infty\}.$$

Proof. From the definition of G we have $x \in G(z)$ if and only if there is a subsequence $\{k_n\}$ and a sequence $(z_n, x_n) \in \text{gph}\, F_{k_n}$ with $(z_n, x_n) \to (z, x)$. Now suppose z is a point where there is $x \in G(z)$ but $x \notin F(z)$. Then, trivially, $z \in E$. In the second case if there is a sequence $z_n \to z$ and $x_n \in F_{k_n}(z_n)$ with $|x_n| \to +\infty$ then $\{x_n\}$ has no limit points and, vacuously, $z \in E$. For the other direction suppose $z \in E$, i.e. there are sequences $z_n \to z$ and $x_n \in F_{k_n}(z_n)$ but no limit point of $\{x_n\}$ lies in $F(z)$. If $\{x_n\}$ has no limit points then

$|x_n| \to +\infty$ (since all this takes place in $\mathbb{R}^n$). If $\{x_n\}$ does have a limit point, say x, then $x \in G(z)$ in which case $G(z) \setminus F(z) \neq \emptyset$. □

This decomposition of the mysterious set E is extremely useful in characterizing the appropriate limit mapping F. Clearly any mapping F whose graph *contains* $\limsup \operatorname{gph} F_k$ will suffice, provided it is also single-valued P-a.s.; thus, the limit multifunction is determined by these conditions only up to sets of P-measure zero. We summarize these observations, Proposition 3.4, and Theorem 3.3 in the key result of this section.

Theorem 3.5. *Let P and P_k, $k = 1, 2, \ldots$ be finite measures on a complete, separable metric space Z, and let F and F_k, $k = 1, 2, \ldots$, be closed-valued measurable multifunctions mapping Z into $\mathbb{R}^n$ that are single-valued relative to P and P_k, respectively. Suppose*

$$\operatorname{gph} F \supset \limsup_{k\to\infty} \operatorname{gph} F_k; \tag{3.4}$$

$$\lim_{k\to\infty} P_k(\operatorname{dom} F_k) = P(\operatorname{dom} F); \tag{3.5}$$

and

$$P\{E'\} = 0, \text{ where } E' = \{z : \exists z_n \to z,\ x_n \in F_{k_n}(z_n) \text{ with } |x_n| \to +\infty\}. \tag{3.6}$$

Then $P_k F_k^{-1} \underset{w}{\longrightarrow} P F^{-1}$. □

To aid in the interpretation of condition (3.5) we make the following observation. (The proof is an easy application of the ideas of this section—the reader is encouraged to try an alternative proof based on classical techniques!)

Proposition 3.6. *Let P and P_k, $k = 1, 2, \ldots$, be probability measures on $(Z, \mathcal{Z})$. The following statements are equivalent*

(i) $P_k \underset{w}{\longrightarrow} P$; *and*

(ii) For every sequence $\{C_k\}$ of sets in $\mathcal{Z}$ with

$$P\left(\limsup_{k\to\infty} C_k \cap \limsup_{k\to\infty} C_k^c\right) = 0 \tag{3.7}$$

one has $P_k(C_k) \to P(\limsup_{k\to\infty} C_k)$.

Proof. (i) ⇒ (ii): Let h and h_k be the indicator functions of $\limsup C_k$ and C_k, $k = 1, 2, \ldots$, respectively. Define the multifunction $H : Z \rightrightarrows \mathbb{R}$ by the formula

$$\operatorname{gph} H = \limsup_{k\to\infty} \operatorname{gph} h_k.$$

Since $\operatorname{dom} H = \operatorname{dom} h_k = Z$ and since the images of h_k are uniformly bounded, it follows that $\{h_k\}$ and H satisfy conditions (3.4–6) of Theorem 3.5. Now suppose H is multivalued at z. The only possible values of $H(z)$ are 0 and 1, so $H(z) = \{0,1\}$. Since $1 \in H(z)$, there must exist a subsequence $(z_{k_n}, 1) \in \operatorname{gph} h_{k_n}$ with $z_{k_n} \to z$. Hence $z \in \limsup C_k$. On the other hand, since $0 \in H(z)$, there must exist a subsequence $(z_{k_\ell}, 0) \in \operatorname{gph} h_{k_\ell}$ with $z_{k_\ell} \to z$. Hence $z \in \limsup C_k^c$. Therefore

$$\{z : H(z) \text{ is not single valued}\} = \limsup C_k \cap \limsup C_k^c,$$

which by assumption (3.7) has P-measure zero. Applying Theorem 3.5 yields

$$P_k h_k^{-1} \underset{w}{\longrightarrow} PH^{-1}.$$

It is easily shown that $\operatorname{gph} h \subset \operatorname{gph} H$ (i.e. that h is a selection of H), hence $Ph^{-1} = PH^{-1}$ by Corollary 2.8. It follows that

$$\int h_k dP_k \to \int h dP,$$

which proves (ii).

(ii) $\Rightarrow$ (i). According to Billingsley [**4**; Thm. 2.1], $P_k \underset{w}{\longrightarrow} P$ if and only if $P_k(C) \to P(C)$ for all P-continuity sets C, i.e. for all C such that $P(\operatorname{cl} C \cap \operatorname{cl} C^c) = 0$. Let C be a P-continuity set and let $C_k = C$, all k. From Kuratowski [**11**; 25.IV.6], $\limsup C_k = \operatorname{cl} C$ and $\limsup C_k^c = \operatorname{cl}(C^c)$. Hence the statement (ii) implies that $P_k(C) \to P(\operatorname{cl} C)$ for all P-continuity sets C and, since $P(\operatorname{cl} C) = P(C)$ for all such sets, we conclude that $P_k \underset{w}{\longrightarrow} P$.□

To complete our preparations for the asymptotic theory of the next section we translate Theorem 3.5 into the terminology of random variables in the usual way.

Definition 3.7. Let $\{\mathbf{z}_k\}$ be a sequence of random variables taking values in Z, i.e. each $\mathbf{z}_k$ is a measurable function from a probability space $(\Omega_k, \mathcal{F}_k, \mu_k)$ into $(Z, \mathcal{Z})$. We say $\mathbf{z}_k$ *converges in distribution* to a random variable $\mathbf{z}$ on Z, $\mathbf{z}_k \underset{D}{\longrightarrow} \mathbf{z}$, if the induced measures

$$P_k(A) = \mathcal{P}\{\mathbf{z}_k \in A\} \quad \text{for all } A \in \mathcal{Z},$$

converge weakly to the measure P induced by $\mathbf{z}$, i.e. $P_k \underset{w}{\longrightarrow} P$.

The only possible misunderstanding in the translation will be the meaning attached to $F_k(\mathbf{z}_k)$ and $F(\mathbf{z})$—we do not regard these as *random sets*, but rather as versions of the random variables (in $\mathbb{R}^n$) whose distributions are given by $\mathcal{P}\{\mathbf{z}_k \in F_k^{-1}(\cdot)\}$ and $\mathcal{P}\{\mathbf{z} \in F^{-1}(\cdot)\}$, respectively. ¿From Corollary 2.7, these are distributions if and only if F_k and F are single-valued almost surely relative to the distributions of $\mathbf{z}_k$ and $\mathbf{z}$ respectively. Now, appealing to Corollary 2.8, any selection $\mathbf{x}_k \in F_k(\mathbf{z}_k)$ and $\mathbf{x} \in F(\mathbf{z})$ is a version of $F_k(\mathbf{z}_k)$ and $F(\mathbf{z})$. Thus we have the following corollary to Theorem 3.5.

Corollary 3.8. *Let $\mathbf{z}_k \xrightarrow[D]{} \mathbf{z}$, and F and F_k, $k = 1, 2, \ldots$, be as above. Assume that F satisfies (3.4), that $P(\mathbf{z}_k \in \operatorname{dom} F_k) \to P(\mathbf{z} \in \operatorname{dom} F)$, and that $P\{\mathbf{z} \in E'\} = 0$, where E' is given in (3.6). Then if $\mathbf{x}_k$ is any selection of $F_k(\mathbf{z}_k)$, $k = 1, 2, \ldots$, and $\mathbf{x}$ is any selection of $F(\mathbf{z})$ one has*

$$\mathbf{x}_k \xrightarrow[D]{} \mathbf{x}. \qquad \square$$

4. Pseudo-Derivatives and the Central Limit Theorem for Lipschitz Mappings

The theory of the preceding sections is applied to determine the asymptotic behaviour of selections

$$\mathbf{x}_N \in F(\overline{\mathbf{z}}_N), \qquad N = 1, 2, \ldots,$$

where $\overline{\mathbf{z}}_N = \frac{1}{N}\sum_{n=1}^{N} \mathbf{z}_n$ is the sample mean of the N independent and identically distributed random variables $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N \in Z$.

We assume that Z is a separable Banach space equipped with the Borel sets $\mathcal{Z}$. In this section we are concerned primarily with establishing rather general conditions on the closed-valued measurable mapping F that ensure the existence of a random variable $\mathfrak{X}$ with values in $\mathbb{R}^n$ and a point $\overline{x} \in \mathbb{R}^N$ that satisfy

$$\sqrt{N}[\mathbf{x}_N - \overline{x}] \xrightarrow[D]{} \mathfrak{X},$$

under the assumption that the sample means $\overline{\mathbf{z}}_N$, $N = 1, 2, \ldots$, (we shall henceforth drop the "bar") satisfy a *central limit theorem* in Z—i.e. there exists a (normal) Z-valued random variable $\mathfrak{z}$, with zero mean and covariance equal to $\operatorname{cov} \mathbf{z}_1$, that satisfies

$$\sqrt{N}[\mathbf{z}_N - E\mathbf{z}_1] \xrightarrow[D]{} \mathfrak{z}. \tag{4.1}$$

These notions of normal random variable, expectation, and covariance for Banach spaces are the counterparts of the usual objects in $\mathbb{R}^n$, cf. Araujo and Giné [**2**]. Not all Banach spaces give rise to central limit theorems. In applications the formula (4.1) must be proved for the appropriate Banach space.

For convenience denote $E\mathbf{z}_1$ by $\overline{z}$. We make the following simplifying assumption:

$$F(\overline{z}) = \{\overline{x}\}, \text{ a singleton.} \tag{4.2}$$

Strictly speaking (4.2) is not necessary; however without it the complications are great. One has, somehow, to be able to select $\overline{x} \in F(\overline{z})$ to allow convergence of $\mathbf{x}_N$ to $\overline{x}$ at the appropriate rate when $\mathbf{x}_N$ is not uniquely determined. On the other hand if $\overline{z}$ itself is only an estimate and if $F(\overline{z})$ is single-valued a.s. then (4.2.) may as well be assumed anyway. We shall also assume that F is Lipschitzian at $\overline{z}$ in the following sense due to Robinson [**14**].

Definition 4.1. A multifunction $F : Z \rightrightarrows \mathbb{R}^n$ is said to be *Lipschitzian* with modulus λ at $\bar{z}$ if there exists a neighborhood U of $\bar{z}$ such that

$$F(z) \subset F(\bar{z}) + \lambda \|z - \bar{z}\| B, \qquad \forall z \in U, \tag{4.3}$$

where B is the closed unit ball in $\mathbb{R}^n$. This reduces to the usual definition of Lipschitz behaviour when $F = f$, a function.

We shall apply the theory of the previous section to the sequence of *difference quotients*

$$\Delta_t(\bar{z}; h) = \frac{F(\bar{z} + th) - F(\bar{z})}{t}. \tag{4.4}$$

Recall that when $F = f$, a Fréchet differentiable function, then

$$\sqrt{N}[f(z_n) - f(\bar{z})] \underset{D}{\longrightarrow} \nabla f(\bar{z})\mathfrak{z}. \tag{4.5}$$

Clearly (4.5) is a statement about the convergence of the sequence of distributions

$$\Delta_{\frac{1}{\sqrt{N}}}(\bar{z}; \sqrt{N}(z_N - \bar{z})), \qquad N = 1, 2, \ldots.$$

Therefore, with Corollary 3.8 and condition (3.4) in mind, we make the following definition. For completeness we also define, although we shall not need them, the lower pseudo-derivative and the pseudo-differentiability property.

Definition 4.2. (Rockafellar [20]) The *upper pseudo-derivative*, $F^+_{\bar{z},\bar{x}}(\cdot)$, of a multifunction F mapping Z into $\mathbb{R}^n$, at a point $(\bar{z}, \bar{x})$ in the graph of F (i.e. $\bar{x} \in F(\bar{z})$) is given by the formula

$$\operatorname{gph} F^+_{\bar{z},\bar{x}} = \limsup_{t \downarrow 0} t^{-1}[\operatorname{gph} F - (\bar{z}, \bar{x})]. \tag{4.6}$$

The *lower pseudo-derivative*, $F^-_{\bar{z},\bar{x}}(\cdot)$, is given by

$$\operatorname{gph} F^-_{\bar{z},\bar{x}} = \liminf_{t \downarrow 0} t^{-1}[\operatorname{gph} F - (\bar{z}, \bar{x})], \tag{4.7}$$

and if these are equal then we say F is *pseudo-differentiable* at $(\bar{z}, \bar{x})$ and denote their common limit as $F'_{\bar{z},\bar{x}}(\cdot)$.

It is not necessary that F be single-valued at $\bar{x}$ for these definitions to make sense. In general, one obtains very different pseudo-derivatives for different choices of $\bar{x} \in F(\bar{z})$, therefore our notation must indicate which choice has been made. When F is single-valued at $\bar{z}$ we simply write $F^+_{\bar{z}}$, etc. We note that the limsup and liminf of any collection of sets are closed, hence it follows that $F^+_{\bar{z},\bar{x}}$ and $F^-_{\bar{z},\bar{x}}$ have closed graph and are therefore closed-valued and measurable by Proposition 2.3. We record this as:

Proposition 4.3. *Let $F : Z \rightrightarrows \mathbb{R}^n$ and $(\bar{z},\bar{x}) \in \operatorname{gph} F$. Then the multifunctions $F^+_{\bar{z},\bar{x}}$ and $F^-_{\bar{z},\bar{x}}$ are closed-valued and measurable.* □

If F is Lipschitzian at $\bar{z}$ and $F(\bar{z})$ is a singleton then we can establish an important boundedness property of the difference quotients $\Delta_t(\bar{z};\cdot)$.

Proposition 4.4. *Let $F : Z \rightrightarrows \mathbb{R}^n$ be Lipschitzian at $\bar{z}$, and suppose $F(\bar{z}) = \{\bar{x}\}$. Then there exists a compact set K such that*

$$\Delta_t(\bar{z};h) \subset K$$

for all h with $\|h\| \leq 1$ and all t sufficiently small.

Proof. Let the modulus $\lambda \geq 0$ and neighborhood U of $\bar{z}$ be given as in 4.1. Then

$$F(z) \subset \bar{x} + \lambda\|z-\bar{z}\|B, \qquad \text{all } z \in U$$

where B is the unit ball in $\mathbb{R}^n$. Let h be given and put $z = \bar{z} + th$. Then if $\|h\| \leq 1$ and t is sufficiently small we have $\bar{z} + th \in U$, so

$$\Delta_t(\bar{z};h) = t^{-1}(F(\bar{z}+th) - \bar{x}) \subset \lambda B$$

and B is compact in $\mathbb{R}^n$. The conclusion follows. □

Finally, if $\bar{z} \in \operatorname{int}\operatorname{dom} F$, then we have the following important property.

Proposition 4.5. *Let $F : Z \rightrightarrows \mathbb{R}^n$ be Lipschitzian and single-valued at $\bar{z}$, and suppose $\bar{z} \in \operatorname{int}\operatorname{dom} F$. Then*

(i) $Z = \limsup_{t\downarrow 0} \operatorname{dom} \Delta_t(\bar{z};\cdot) = \operatorname{dom} F^+_{\bar{z}}$; *and*

(ii) $\limsup_{t\downarrow 0}(\operatorname{dom} \Delta_t(\bar{z};\cdot))^c = \emptyset$.

Proof. Note that $h \in \operatorname{dom}\Delta_t(\bar{z};\cdot)$ if and only if $\bar{z}+th \in \operatorname{dom} F$. Hence

$$\operatorname{dom}\Delta_t(\bar{z};\cdot) = t^{-1}[\operatorname{dom} F - \bar{z}].$$

Since $\bar{z} \in \operatorname{int}\operatorname{dom} F$, then for any $s \geq 0$ and all sufficiently small t the set $\operatorname{dom}\Delta_t(\bar{z};\cdot)$ contains sB, where B is the unit ball in Z. From this we obtain (ii) and the first equality in (i). Now let $h \in Z$ be given (without loss of generality, since $\operatorname{gph} F^+_{\bar{z}}$ is a cone, assume $\|h\| = 1$). For all sufficiently small t we have $\bar{z}+th \in \operatorname{dom} F$ (since $\bar{z} \in \operatorname{int}\operatorname{dom} F$) and

$$\Delta_t(\bar{z};h) \subset K$$

for some compact K, by Proposition 4.3. Hence there are $k_t \in \Delta_t(\bar{z};h)$ for all sufficiently small $t \geq 0$, and at least one limit point, say k. By definition this k belongs to $F^+_{\bar{z}}(h)$, i.e. $h \in \operatorname{dom} F^+_{\bar{z}}$, which proves the second equality in (i). □

We are ready to state the main result. We suppose that $\mathbf{z}_N$, $N = 1,2,\ldots$, are random variables in a separable Banach space Z, and that F is a closed-valued measurable multifunction mapping Z into $\mathbb{R}^n$ with $F(\mathbf{z}_N)$ single-valued a.s.; and we put $\bar{z} = E\mathbf{z}_1$.

Theorem 4.6. *Suppose that* $\sqrt{N}[\mathbf{z}_N - \bar{z}] \xrightarrow[D]{} \mathfrak{z}$, *and that the following conditions are satisfied:*

(i) $F(\bar{z}) = \{\bar{x}\}$, *a singleton;*

(ii) F *is Lipschitzian at* $\bar{z}$;

(iii) $\bar{z} \in \operatorname{int}\operatorname{dom} F$; *and*

(iv) $F_{\bar{z}}^{+}(\mathfrak{z})$ *is a.s. single-valued.*

If $\mathbf{x}_N$ *is a selection from* $F(\mathbf{z}_N)$ *and* $\mathfrak{x}$ *a selection from* $F_{\bar{z}}^{+}(\mathfrak{z})$ *then*

$$\sqrt{N}[\mathbf{x}_N - \bar{x}] \xrightarrow[D]{} \mathfrak{x}.$$

Proof. Clearly $\sqrt{N}[\mathbf{x}_N - \bar{x}]$ is a selection from $\Delta_{\frac{1}{\sqrt{N}}}(\bar{z}; \sqrt{N}(\mathbf{z}_N - \bar{z}))$. The conclusion will follow from Corollary 3.8. From (i) and (ii) we have via Proposition 4.4 that the set

$$E' = \{h : \exists h_N \to h \text{ and } t_N \downarrow 0,\ x_N \in \Delta_{t_N}(\bar{z}; h_N) \text{ with } |x_N| \to +\infty\}$$

is empty, hence $P(\mathfrak{z} \in E') = 0$. From (i), (ii) and (iii) and Proposition 4.5 we have

$$\limsup_{N\to\infty} \operatorname{dom} \Delta_{\frac{1}{\sqrt{N}}}(\bar{z}; \cdot) \cap \limsup_{N\to\infty} \operatorname{dom}(\Delta_{\frac{1}{\sqrt{N}}}(\bar{z}; \cdot))^c = \emptyset$$

and

$$\operatorname{dom} F_{\bar{z}}^{+} = \limsup_{N\to\infty} \operatorname{dom} \Delta_{\frac{1}{\sqrt{N}}}(\bar{z}; \cdot),$$

hence by Proposition 3.6

$$P(\sqrt{N}[\mathbf{z}_N - \bar{z}] \in \operatorname{dom} \Delta_{\frac{1}{\sqrt{N}}}(\bar{z}; \cdot)) \to P(\mathfrak{z} \in \operatorname{dom} F_{\bar{z}}^{+}).$$

The condition (iv) assures that $F_{\bar{z}}^{+}(\mathfrak{z})$ induces a distribution on $\mathbb{R}^n$. It remains only to show

$$\operatorname{gph} F_{\bar{z}}^{+} \supset \limsup_{N\to\infty} \operatorname{gph} \Delta_{\frac{1}{\sqrt{N}}}(\bar{z}; \cdot).$$

But this follows trivially from the definition: Let $(h_N, x_N) \in \operatorname{gph} \Delta_{\frac{1}{\sqrt{N}}}(\bar{z}; \cdot)$, $N = 1, 2, \ldots$, with $(h_N, x_N) \to (h, x)$. We have only to show that $x \in F_{\bar{z}}^{+}(h)$. But

$$x_N \in t_N^{-1}(F(\bar{z} + t_N h_N) - \bar{x}),$$

where we set $\sqrt{N} = t_N$, or in other words

$$(h_N, x_N) \in t_N^{-1}[\operatorname{gph} F - (\bar{z}, \bar{x})].$$

Thus, by (1.12) and (4.6), $(h, x) \in \operatorname{gph} F_{\bar{z}}^{+}$. □

In case $F = f$, a measurable function, the conclusions of this theorem may be given a more definite form by analyzing the connections between the pseudo-derivative and ordinary directional derivatives under the conditions (ii) and (iv). Following Rockafellar [**17**], we say f is directionally differentiable at $\bar{z}$ and in the direction h in the *ordinary sense* if the limit

$$f'(\bar{z}; h) = \lim_{t \downarrow 0} \frac{f(\bar{z} + th) - f(\bar{z})}{t} \tag{4.8}$$

exists, and in the *Hadamard sense* if this limit can in fact be taken as

$$\lim_{\substack{h' \to h \\ t \downarrow 0}} \frac{f(\bar{z} + th') - f(\bar{z})}{t}. \tag{4.9}$$

Proposition 4.7. *Suppose that $f : Z \to \mathbb{R}^n$ is Lipschitzian at $\bar{z}$. Then $f_{\bar{z}}^{+}(h)$ is single-valued if and only if $f'(\bar{z}; h)$ exists in the Hadamard sense, and in either case $f_{\bar{z}}^{+}(h) = \{f'(\bar{z}; h)\}$.*

Proof. Suppose $f_{\bar{z}}^{+}(h) = \{k\}$, and let $t_n \downarrow 0$ and $h_n \to h$ be arbitrary. Then, by 4.4, there is a compact set K such that

$$k_n := \frac{f(\bar{z} + t_n h_n) - f(\bar{z})}{t_n} \in K$$

for all n sufficiently large. Hence $\{k_n\}$ has a limit point, say k'. Thus we have a sequence $(h_n, k_n) \to (h, k')$ that satisfies

$$(h_n, k_n) \in t_n^{-1}[\operatorname{gph} f - (\bar{z}, f(\bar{z}))], \qquad n = 1, 2, \ldots,$$

hence, by (1.12) and (4.6), $k' \in f_{\bar{z}}^{+}(h)$; so in fact $k' = k$. Thus the limit in (4.9) exists and is equal to k.

For the other direction we note that (4.9) holds iff for all sequences $t_n \downarrow 0$ and $h_n \to h$ one has

$$\frac{f(\bar{z} + t_n h_n) - f(\bar{z})}{t_n} \to f'(\bar{z}; h).$$

Hence there is only one element in $f_{\bar{z}}^{+}(h)$ and this must be $f'(\bar{z}; h)$. □

It follows as a direct corollary that if $f'(\bar{z}; \mathfrak{z})$ exists a.s. in the Hadamard sense, then the conclusion of Theorem 4.6 holds. We record this as

Corollary 4.8. *Suppose $f : Z \to \mathbb{R}^n$ is measurable and $f'(\bar{z}; \mathfrak{z})$ exists a.s. in the Hadamard sense. Then*

$$\sqrt{N}[f(z_N) - f(\bar{z})] \underset{D}{\longrightarrow} f'(\bar{z}; \mathfrak{z}).$$ □

As we progressively strengthen the differentiability conditions we reach something like (1.3). If $f'(\bar{z}; h)$ exists in the ordinary sense for all h and $f'(\bar{z}; \cdot)$ is continuous then it is well known that (1.3) *implies* (4.9) and if, additionally, Z is finite-dimensional then (1.3), (4.8) and (4.9) are all *equivalent*; see, for example, Shapiro [**26**].

References

[1] J. Aitchison and S.D. Silvey, "Maximum likelihood estimation of parameters subject to restraints", *Ann. Math. Stat.* **29**(1948), 813–828.

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

[3] Z. Artstein, "Limit laws for multifunctions applied to an optimization problem", in *Multifunctions and Integrands*, Lecture Notes in Math. 1091, Springer-Verlag, 1984, pp. 66–79.

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

[5] D.D. Boos and R.J. Serfling, "A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates", *Annals of Statistics* **8**(3)(1980), 618-624.

[6] B.R. Clarke, "Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations", *Annals of Statistics* **11**(4)(1983), 1196–1205.

[7] J. Dupačová and R.J-B. Wets, "On the asymptotic behaviour of constrained estimates and optimal decisions", manuscript (1986).

[8] P.J. Huber, "Robust estimation of a location parameter", *Ann. Math. Stat.* **35**(1964), 73–101.

[9] P.J. Huber, "The behaviour of maximum likelihood estimates under nonstandard conditions", in: *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, University of California Press, Berkeley, 1967.

[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] C. Kuratowski, *Topologie I*, Monographie Matematyczne, Warsawa, 1948.

[12] P.A. Meyer, *Probability and Potentials*, Blaisdell Publ. Co., Massachusetts, 1966.

[13] R. von Mises,"On the asymptotic distribution of differentiable statistical functions", *Ann. Math. Statist.* **18**(1947), 309–348.

[14] S.M. Robinson, "Generalized equations and their solutions, part 1: basic theory", *Math. Prog. Study* **10**(1979),128–141.

[15] R.T. Rockafellar, "Integral functionals, normal integrands and measurable selections", in *Nonlinear Operators and the Calculus of Variations*, Lecture Notes in Math. 543, Springer-Verlag, 1976, pp. 157–207.

[16] R.T. Rockafellar, "Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization", *Math. of O.R.* **6**(3) (1981), 424-436.

[17] R.T. Rockafellar, "Directional differentiability of the optimal value function in a nonlinear programming problem," *Math. Prog. Study* **21**(1984), 213–216.

[18] R.T. Rockafellar, "First and second order pseudo-differentiability in nonlinear programming", (1987) manuscript.

[19] R.T. Rockafellar, "Second order optimality conditions in nonlinear programming obtained by way of pseudo-derivatives", (1987) manuscript.

[20] R.T. Rockafellar, "Generalized second derivatives of convex functions and saddle functions", in preparation.

[21] R.T. Rockafellar and R.J-B. Wets, "Variational systems, an introduction", in *Multifunctions and Integrands*, Lecture Notes in Math. 1091, Springer-Verlag, 1984, pp. 1–54.

[22] G. Salinetti and R.J-B. Wets, "On the convergence of sequences of convex sets in finite dimensions", *SIAM Review* **21**(1)(1979), 18–23.

[23] G. Salinetti and R.J-B. Wets, "On the convergence of closed-valued measurable multifunctions", *Trans. A.M.S.* **266**(1)(1981), 275–189.

[24] 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.

[25] A. Shapiro, "Second order sensitivity analysis and asymptotic theory of parametrized nonlinear programs", *Math. Programming* **33**(1986), 280–299.

[26] A. Shapiro, "On concepts of directional differentiability", manuscript, 1986.

[27] D.H. Wagner, "Survey of measurable selection theorems", *SIAM J. Control and Opt.* **15**(5)(1977), 859–903.

[28] W. Weil, "An application of the central limit theorem for Banach-space-valued random variables to the theory of random sets", *Z. Wahrsch. verw. Geb.* **60**(1982), 203–208.

[29] R.J-B. Wets, "Convergence of convex functions, variational inequalities and convex optimization problems," in *Variational Inequalities and Complementarity Problems* (R.W. Cottle et al, eds.), Wiley, 1980, 375–403.

[30] R.A. Wijsman, "Convergence of sequences of convex sets, II", *Trans. A.M.S.* **123** (1966), 32–45.