WORKING PAPER

CENTRAL LIMIT THEORY
FOR MULTIVALUED MAPPINGS

Alan J. King

October 1987
WP-87-112

IIASA
International Institute
for Applied Systems Analysis

CENTRAL LIMIT THEORY FOR MULTIVALUED MAPPINGS

Alan J. King

October 1987
WP-87-112

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

Foreword

This paper presents new fundamental results concerning central limit behaviour of sequences of random closed sets, modelled as a multivalued mapping of an underlying asymptotically normal sequence of random variables.

The main theorem generalizes the classical result for differentiable functions in a mathematically satisfying way by combining recent developments in convergence theory for random closed sets and recent work in pseudo-differentiability of multifunctions. Potential applications to the asymptotic analysis of solution sets for stochastic and ordinary parametric programs with incomplete information are indicated in the examples.

This paper reports research that was partly performed in the Adaptation and Optimization Project of the System and Decision Sciences Program.

A. Kurzhanski
Chairman
System and Decision Sciences Program

CENTRAL LIMIT THEORY FOR MULTIVALUED MAPPINGS

Alan J. King *

1. Introduction

In an earlier paper [1] we gave conditions that described the asymptotic behaviour of selections from a sequence of random sets in a finite-dimensional Euclidean space X that were single-valued almost surely. The results of the present paper reveal that these conclusions may be derived from a much more general asymptotic result for truly multivalued mappings. The basic approach is the same, however: 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. The principle difference is in the choice of distribution. In [1] we analyzed the distribution induced directly on the image space X of the multivalued mapping — this was possible because of the single-valuedness a.s. assumption. In this paper, we follow Salinetti and Wets [2] in analyzing the distributions induced by the multifunction 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.

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

(1.1) $$\tau_\nu[\mathbf{z}_\nu - \bar{z}] \xrightarrow[\mathcal{D}]{} \mathfrak{z}$$

* International Institute for Applied Systems Analysis, A-2361, Laxenburg, Austria

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

$$(1.2) \qquad \tau_\nu[F(\mathbf{s}_\nu) - \bar{x}] \xrightarrow[D]{} F'_{\bar{z},\bar{x}}(\mathfrak{z}),$$

where $F'_{\bar{z},\bar{x}}$ is a "derivative" of F localized at a given point $\bar{x} \in F(\bar{z})$. (The symbol D under the arrow denotes convergence in distribution.)

The crucial condition turns out to be "semi-differentiability", a concept introduced recently by Rockafellar [8] in his exploration of differentiability concepts for multifunctions. This theory is in its infancy. Nevertheless explicit computations are already possible in some situations. These strong connections between the central limit theory and the theory of pseudo-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.

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

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

where the functions $f_i : \mathbb{R}^n \to \mathbb{R}$ 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 2.6.

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 and we defer our study of asymptotic behaviour until these are better understood. In Shapiro [9] there are some partial results in this direction.

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

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

The pattern is completed if we can establish that the sequence of "sample means" $\{\mathbf{s}_\nu(\cdot)\}$ is asymptotically normal in some suitably generalized sense; cf. King [4] for a treatment of the linear-quadratic case. This is the principle reason why we present our results for multivalued mappings defined in a general metric space Z. This problem was first considered in the setting of maximum likelihood estimation; cf. Huber [7], Aitchison and Silvey [10]. For other recent work in the stochastic optimization literature see Dupačová and Wets [8]. The asymptotic analysis of $\{\mathbf{J}_\nu\}$ based on the fundamentally new results of the present paper will appear shortly.

The reader of this paper is expected to be acquainted with the fundamentals of Kuratowski convergence of closed sets and weak*-convergence of probability measures; see, for example, Salinetti and Wets [11] and Billingsley [6], respectively. A sequence of subsets $\{A_\nu\}$ of a locally compact topological space converges to a subset A *in the Kuratowski sense* if

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

where

$$\liminf A_\nu = \{a \mid a = \lim a_\nu \text{ where } a_\nu \in A_\nu \text{ for all but finitely many } \nu\}$$

$$\limsup A_\nu = \{a \mid a = \lim a_\nu \text{ where } a_\nu \in A_\nu \text{ for infinitely many } \nu\}.$$

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 $f : Z \to \mathbb{R}$.

The main result is developed in Section 2. The rest of the paper is devoted to applying the main result to the asymptotic analysis of *selections* $\mathbf{x}_\nu \in F(\mathbf{z}_\nu)$ when F is almost surely single-valued (Sections 3 and 4), and Hadamard differentiable *functions* when F is actually a function (Section 5).

2. Central Limit Theorem for Multivalued Mappings

We present our major result in this section. 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

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

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 "$\mathbf{F}$".) Following Salinetti and Wets [2], 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 (Kuratowski) 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 [2; Theorem 2.5] demonstrate that a sequence of random closed sets $\{\mathbf{F}_\nu\}$ converges in distribution if and only if the *distance processes* $\{d(\cdot, \mathbf{F}_\nu)\}$ converge as stochastic processes on X. By definition these *stochastic processes* $d(\cdot, \mathbf{F}_\nu)$ *converge to* $d(\cdot, \mathbf{F})$, in notation:

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

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

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

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

(The reader should note that a sequence of *random variables* $\{\mathbf{w}_\nu\}$ defined on a complete separable metric space W converges in distribution to $\mathbf{w}$ if and only if one has weak*-convergence of the measures induced by the $\mathbf{w}_\nu$ on the space W. We may also regard these as *random closed sets* since points are closed in W. But in this view the sequence $\{\mathbf{w}_\nu\}$ converges in distribution if and only if one has weak*-convergence of the distributions induced on the hyperspace $\mathcal{F}(W)$. The two notions are not equivalent. It will always be clear from the context which is being employed.)

Setting the stage for the central limit theorem, let Z be a separable complete metric vector space (*separable Fréchet space*) equipped with its Borel field $\mathcal{B}(Z)$ 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 $\bar{z}$, a sequence of positive numbers $\{\tau_\nu\}$ monotonically decreasing to 0, and a *limit distribution* $\mathfrak{z}$ such that

$$\tau_\nu^{-1}[\mathbf{z}_\nu - \bar{z}] \xrightarrow[\mathcal{D}]{} \mathfrak{z} \tag{2.3}$$

as random variables in Z.

A central limit theorem for the sequence $\{F(\mathbf{z}_\nu)\}$ inevitably rests upon an appropriate definition of first-order behaviour for the multifunction $F : Z \rightrightarrows X$. The theorem given below is based on definitions due to Rockafellar [8]. Fix a point $\bar{z}$ and a point $\bar{x} \in F(\bar{z})$, and define the collection $\{D_t : t > 0\}$ of *difference quotient multifunctions*

$$D_t(z) := t^{-1}[F(\bar{z} + tz) - \bar{x}], \qquad t > 0. \tag{2.4}$$

The multifunction F is said to be *semi-differentiable at $\bar{z}$ relative to $\bar{x}$* if there exists a multifunction $D : Z \rightrightarrows X$ such that for all $z \in Z$,

$$\lim_{\substack{t \downarrow 0 \\ z' \to z}} D_t(z') = D(z) \tag{2.5}$$

taken as a limit of sets (in the Kuratowski sense). If such a property holds then it can be shown that F is *pseudo-differentiable* at $(\bar{z}, \bar{x})$ and D equals the *pseudo-derivative* $F'_{\bar{z},\bar{x}}$, all of which is summarized by the formula

$$\lim_{t \downarrow 0} \operatorname{gph} D_t = \operatorname{gph} F'_{\bar{z},\bar{x}} \tag{2.6}$$

taken as a limit of sets in $Z \times X$. (See the proof of [8; 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 $(\bar{z}, \bar{x})$ in the graph of F and construct there a *tangent cone* to $\operatorname{gph} F$; this cone is then the graph of $F'_{\bar{z},\bar{x}}$. 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 $(\bar{x}, \bar{z})$. Naturally, different choices of tangent cones — e.g. Clarke, intermediate, contingent, etc. — all lead to different derivatives. The choice made in [8] is that $\operatorname{gph} F'_{\bar{z},\bar{x}}$ should equal simultaneously the *contingent* and *intermediate* cones (respectively lim sup and lim inf in (2.6)).

Definition 2.1. Given a measure μ on $(Z, \mathcal{B}(Z))$, the multifunction $F : Z \rightrightarrows X$ is said to be *almost surely semi-differentiable at $\bar{z}$ relative to $\bar{x}$ with respect to μ* if there exists a multifunction $D : Z \rightrightarrows X$ such that (2.5) holds for all points z except possibly those in a set of μ-measure zero. Abusing the notation slightly, we still write $D = F'_{\bar{z},\bar{x}}$ even though the limit (2.5) and not (2.6) is understood here.

This differentiability notion turns out to be exactly what is needed as we see in the following, surprisingly elegant, central limit theorem for multivalued mappings.

Theorem 2.2. *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 and if F is almost surely semi-differentiable at $\bar{z}$ relative to a point $\bar{x} \in F(\bar{z})$ with respect to the measure induced by $\mathfrak{z}$, then $\{F(\mathbf{z}_\nu)\}$ satisfies the generalized central limit formula:*

$$\tau_\nu^{-1}[F(\mathbf{z}_\nu) - \bar{x}] \xrightarrow[\mathcal{D}]{} F'_{\bar{z},\bar{x}}(\mathfrak{z}) \tag{2.7}$$

as random closed sets in X or, equivalently,

$$d(x, \tau_\nu^{-1}[F(\mathbf{z}_\nu) - \bar{x}]) \xrightarrow[\mathcal{D}]{} d(x, F'_{\bar{z},\bar{x}}(\mathfrak{z})), \quad x \in X \tag{2.8}$$

as stochastic processes on X.

Proof. Denote by μ_ν the measures induced on the space Z by the random variables $\tau_\nu^{-1}[\mathbf{z}_\nu - \bar{z}]$ and by μ that induced by $\mathfrak{z}$. The meaning of the formula (2.3) is precisely that μ_ν weak*-converges to μ. Employing the difference quotient notation (2.4), the measures induced on the complete separable metric space $\mathcal{F}(X)$ by the random closed sets on the left side of (2.7) 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 [6 ; 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 $\delta(z) = F'_{\bar{z},\bar{x}}(z)$ for μ-almost all z. This establishes (2.7). That (2.8) is equivalent to (2.7) was shown by Salinetti and Wets [2; Theorem 2.5]. □

Evaluating these distance processes (2.8) 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

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

we obtain the following corollary.

Corollary 2.3. *Under the conditions of Theorem 2.2,*

$$\tau_\nu^{-1} d(\bar{x}, F(\mathbf{z}_\nu)) \xrightarrow[\mathcal{D}]{} d(0, F'_{\bar{z},\bar{x}}(\mathfrak{z})) \tag{2.9}$$

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

Remark 2.4. This corollary leads to an important interpretation of the meaning of the asymptotic distribution $F'_{\bar{z},\bar{x}}(\mathfrak{z})$. It represents the *residual uncertainty* in the estimate $F(\mathbf{z}_\nu)$ relative to $\bar{x} \in F(\bar{z})$. If $\mathbf{x}_\nu \in F(\mathbf{z}_\nu)$ is a measurable selection then clearly

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

so the asymptotic behaviour of $\tau_\nu^{-1}\|\bar{x} - \mathbf{x}_\nu\|$ *cannot be better* than that described in (2.9). If F is convex-valued and $\bar{z} \in \operatorname{int}\operatorname{dom} F$ then it can be shown that there exists a selection $\mathbf{x}_\nu \in F(\mathbf{z}_\nu)$ such that

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

i.e. $\{\mathbf{x}_\nu\}$ in norm converges in distribution to $\bar{x}$. To say more than this about selections seems to require F and $F'_{\bar{z},\bar{x}}$ to be almost surely single-valued — a theme we shall pursue in the rest of this paper.

Example 2.5. A simple counter-example illustrates the semi-differentiability condition. Let $Z = X = \mathbb{R}$ and $F : Z \rightrightarrows \mathbb{R}$ 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 $(\bar{z}, \bar{x}) = (0,0) \in \operatorname{gph} F$. It is easy to see that $F'_{0,0}$ exists in the sense of formula (2.6) with

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

and that the semi-differentiability condition (2.5) 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 $\{\nu^{-1}[\mathbf{z}_\nu - 0]\}$ converges in distribution to the random variable $\mathfrak{z}$ taking the value 0 with probability one. All the conditions of Theorem 2.2 are satisfied except that $\mathfrak{z}$ places nonzero mass on the point at which semidifferentiability fails. Denote by $h_\nu(\cdot)$ the distance function $d(0, \nu^{-1}[F(\cdot) - 0])$ and by $h(\cdot)$ the function $d(0, F'_{0,0}(\cdot))$. If Corollary 2.3 holds then $h_\nu(\mathbf{z}_\nu) \xrightarrow[\mathcal{D}]{} h(\mathfrak{z})$. But 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, and $h(\mathfrak{z}) \in [b, +\infty]$ with probability zero. This contradicts the Portmanteau Theorem [**6**; Theorem 2.1], thus Theorem 2.2 fails for this example.

Example 2.6. 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

$$f_i(z,x) \begin{cases} \leq 0 & i = 1, \ldots, s \\ = 0 & i = s+1, \ldots, m \end{cases} \qquad (2.10)$$

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 $\{s_1, \ldots, s_\nu\}$ and averaging them to form an estimate $\mathbf{s}_\nu$, i.e.

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

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

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

where $\mathfrak{z}$ has a centered Gaussian distribution. If $(\bar{z}, \bar{x})$ is a point where the system (2.10) satisfies the Mangasarian-Fromowitz constraint qualification, then (cf. Rockafellar [**3**; Example 5.5]) the mapping F is semidifferentiable at $\bar{z}$ relative to $\bar{x}$ and, moreover, an explicit formula is obtained for the contingent derivative $F'_{\bar{z},\bar{x}}$, namely for all z the set $F'_{\bar{z},\bar{x}}(z)$ consists of the points x satisfying the linearized system

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

with $I(\bar{z}, \bar{x})$ denoting the inequality constraints of (2.10) active at $(\bar{z}, \bar{x})$. From Theorem 2.2

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

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

$$F'_{\bar{z},\bar{x}}(\mathfrak{z}) = \left\{ x \;\middle|\; \begin{array}{lll} \nabla_x f_i(\bar{z}, \bar{x}) \cdot x + \mathbf{b}_i & \leq 0 & i \in I(\bar{z}, \bar{x}) \\ \nabla_x f_i(\bar{z}, \bar{x}) \cdot x + \mathbf{b}_i & = 0 & i = s+1, \ldots, m \end{array} \right\}.$$

These Gaussian random polyhedra thus play a role in mathematical programming similar to that of Gaussian random variables in classical statistics, since for large ν

$$F(\mathbf{s}_\nu) \approx \bar{x} + \frac{1}{\sqrt{\nu}} F'_{\bar{z},\bar{x}}(\mathfrak{z}),$$

i.e. the distribution of $F(\mathbf{s}_\nu)$ approximates a Gaussian random polyhedron.

3. Convergence in Distribution for Selections

Suppose that a sequence of random closed sets $\{\mathbf{F}_\nu\}$ converges in distribution. What then can be said about the convergence in distribution of arbitrary measurable selections $\mathbf{x}_\nu \in \mathbf{F}_\nu$ as *random variables*? This question was introduced in Remark 2.4; in this section we explore some answers.

As always, the space X is finite-dimensional, linear and normed. We shall find it convenient to refer explicitly to the underlying probability space $(\Omega, \mathcal{A}, \mu)$, therefore in this section we use measurable multifunction and measurable function in place of random set and random variable notations. The ***domain*** of a closed-valued measurable multifunction $F : \Omega \rightrightarrows X$, denoted $\operatorname{dom} F$, is the measurable set

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

A function $x : \Omega \to X$ is called a ***measurable selection of*** F if $x(\omega) \in F(\omega)$ for μ-almost all $\omega \in \operatorname{dom} F$. For such multifunctions there always exists at least one measurable selection; see, for example, Wagner [5]. It is important to note that $\mu(\operatorname{dom} F)$ may be less than one and in this case the measure μx^{-1} induced on X by a measurable selection x of F is *not a probability measure*. This introduces a minor technical difficulty into the very definition of convergence in distribution for sequences $\{x_\nu\}$ of measurable selections, which as the reader recalls is defined to be weak*-convergence of the sequence $\{\mu x_\nu^{-1}\}$ of measures on X.

Lemma 3.1. *A necessary condition for the weak*-convergence for a sequence of finite measures $\{P_\nu\}$ on a complete separable metric space Z is*

$$P_\nu(Z) \to P(Z) \tag{3.1}$$

Furthermore, if (3.1) holds then all the equivalences in the statement of the Portmanteau Theorem [6; Theorem 2.1] hold true for the sequence $\{P_\nu\}$.

Proof. The first statement follows directly from the definition since $P(Z) = \int 1 dP$. For the second statement we must refer to the proof of the cited Portmanteau Theorem. It is only necessary to show, in the notation of the proof, that (iii)⇒(ii). For this everything goes through to conclude that the *linear transformation* of f, i.e. $\alpha f + \beta \in (0,1)$ for $\alpha, \beta \in \mathbb{R}$, satisfies

$$\limsup_\nu \int [\alpha f(z) + \beta] dP_\nu(z) \leq \int [\alpha f(z) + \beta] dP(z).$$

Using (3.1) we infer from this that

$$\limsup_\nu \int f dP_\nu \leq \int f dP,$$

and the rest of the proof follows as written. □

The significance of this lemma is that it allows us to apply all of the main results of weak*-convergence, in [6] for example, that depend on the equivalences in the Portmanteau Theorem but which do not specifically require the measures to be probabilities.

Definition 3.2. A closed-valued measurable multifunction $F : \Omega \rightrightarrows X$ is said to be *μ-almost surely single-valued* if

$$\mu\{\omega \in \operatorname{dom} F \mid F(\omega) \text{ is not a singleton}\} = 0. \tag{3.2}$$

Theorem 3.3. *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*

$$\mu(\operatorname{dom} F_\nu) \to \mu(\operatorname{dom} F), \tag{3.3}$$

and that

$$\mu\{\omega \in \operatorname{dom} F_\nu | F_\nu(\omega) \textit{ is not single-valued}\} \to 0. \tag{3.4}$$

If $x : \Omega \to X$ and $x_\nu : \Omega \to X$ are measurable selections of F and F_ν, respectively, then the sequence $\{x_\nu\}$ converges in distribution to x as random variables in X.

Proof. For convenience denote by P and P_ν the finite measures μx^{-1} and μx_ν^{-1} on X. First note that $P(X) = \mu x^{-1}(X) = \mu(\operatorname{dom} F)$. Thus assumption (3.3) means $P_\nu(X) \to P(X)$, and so Lemma 3.1 applies. Denote by $B(x,\varepsilon)$ the open sphere of radius $\varepsilon > 0$ centered at the pont $x \in X$, i.e.

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

The collection of all sets that are finite intersections of open spheres is a *convergence determining class*; cf. the corollaries to [**6**; 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*. Now note 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

$$\begin{aligned}
&\mu\{\omega \in \Omega \,|\, [d(x_1,F(\omega)),\ldots,d(x_k,F(\omega))] \in \partial \prod_{i=1}^{k}(-\infty,\varepsilon_i)\} \\
&= \mu\{\omega \in \Omega \,|\, [d(x_1,x(\omega)),\ldots,d(x_k,x(\omega))] = [\varepsilon_1,\ldots,\varepsilon_k]\} \\
&\le \sum_{i=1}^{k} P(\partial B(x_i,\varepsilon_i))
\end{aligned}$$

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

$$\begin{aligned}
&\lim_\nu \mu\{\omega \in \Omega \,|\, [d(x_1,F_\nu(\omega)),\ldots,d(x_k,F_\nu(\omega))] \in \prod_{i=1}^{k}(-\infty,\varepsilon_i)\} \\
&\qquad = \mu\{\omega \in \Omega \,|\, [d(x_1,F(\omega)),\ldots,d(x_k,F(\omega))] \in \prod_{i=1}^{k}(-\infty,\varepsilon_i\},
\end{aligned}$$

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

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

Noting that by King [1] the sets S_ν are all measurable, we have

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

Hence by assumption (3.4) and the observation that

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

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 (3.3) in Theorem 3.3 we have the following proposition.

Proposition 3.4. *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 $F : \Omega \rightrightarrows X$. 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

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

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.

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

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

4. Central Limit Theory for Selections

Returning to the setting of Section 2, we let Z be a separable Fréchet space, X a finite-dimensional normed linear space and $F : Z \rightrightarrows X$ a closed-valued measurable multifunction.

Theorem 4.1. *Suppose that the sequence $\{\mathbf{z}_\nu\}$ of random variables in Z satisfies a generalized central limit formula (2.3) with $\bar{z} \in \operatorname{dom} F$, and that the following assumptions hold:*

(4.1) *F is almost surely semi-differentiable at $\bar{z}$ relative to $\bar{x} \in F(\bar{z})$ with respect to the measure induced by $\mathfrak{z}$;*

(4.2) $$F'_{\bar{z},\bar{x}}(\mathfrak{z}) \text{ is almost surely single-valued;}$$

(4.3) $$\Pr\{\mathbf{z}_\nu \in \operatorname{dom} F\} \rightarrow \Pr\{\mathfrak{z} \in \operatorname{dom} F'_{\bar{z},\bar{x}}\}; \text{ and}$$

(4.4) $$\Pr\{F(\mathbf{z}_\nu) \text{ is nonempty and not single-valued}\} \longrightarrow 0.$$

Then for all measurable selections $\mathbf{x}_\nu \in F(\mathbf{z}_\nu)$ and $\mathfrak{X} \in F'_{\bar{z},\bar{x}}(\mathfrak{z})$ one has

(4.4) $$\tau_\nu[\mathbf{x}_\nu - \bar{x}] \xrightarrow[\mathcal{D}]{} \mathfrak{X}$$

as random variables in X.

Proof. In view of assumption (4.1), Theorem 2.2 applies and thus

$$\tau_\nu[F(\mathbf{z}_\nu) - \bar{x}] \xrightarrow[\mathcal{D}]{} F'_{\bar{z},\bar{x}}(\mathfrak{z})$$

as random sets in X. Clearly $\tau_\nu[\mathbf{x}_\nu - \bar{x}]$ is a selection of $\tau_\nu[F(\mathbf{z}_\nu) - \bar{x}]$, $\nu = 1, 2, \ldots$. Assumption (4.3) is the counterpart of (3.3) in this setting, and with assumptions (4.2) and (4.4) the conclusion follows from Theorem 3.3. □

Remark 4.2. This theorem is a far more general version of King [**1**; Theorem 4.6]. Nevertheless it takes some effort to derive that theorem from the present one. We give a brief indication here. If a closed-valued measurable multifunction F satisfies

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

(ii) F is Lipschitzian at $\bar{z}$ (cf. [**1**; Definition 4.1]); and

(iii) $F^+_{\bar{z},\bar{x}}(\mathfrak{z})$ is a.s. single-valued,

where $F^+_{\bar{z},\bar{x}}$ is the *upper pseudo-derivative*, i.e. the mapping whose graph is the contingent cone to $\operatorname{gph} F$ at $(\bar{z}, \bar{x})$ and which therefore satisfies (2.6) with lim sup, then one can show directly that F is almost-surely semi-differentiable at $\bar{z}$ relative to $\bar{x}$ with respect to $\mathfrak{z}$ and that $F^+_{\bar{z},\bar{x}} = F'_{\bar{z},\bar{x}}$ a.s. It is assumed that $F(\mathbf{z}_\nu)$ is single-valued a.s. The only remaining thing to verify is (4.3). This follows from the assumption

(iv) $\bar{z} \in \operatorname{int} \operatorname{dom} F$

which implies that $\operatorname{dom} F'_{\bar{z},\bar{x}} = Z$; thus (4.3) now follows from Proposition 3.4.

5. Hadamard Differentiable Functions

The technique used in the main result, Theorem 2.2, can also be applied to functions. Let both Z and X be separable Fréchet spaces. Following Rockafellar [**12**], we say that a function $f : Z \to X$ is *Hadamard differentiable at* $\bar{z}$ if the limit

$$(5.1) \qquad f'(\bar{z}; z) = \lim_{\substack{t \downarrow 0 \\ z' \to z}} \frac{f(\bar{z} + tz') - f(\bar{z})}{t}$$

exists for all directions z. Paralleling Definition 2.1, we shall say that f is *almost surely Hadamard differentiable at* $\bar{z}$ *with respect to a measure* μ if (5.1) holds for all directions z except possibly those in a set of μ-measure zero.

Theorem 5.1. *Let Z and X be separable Fréchet spaces. Suppose $\{\mathbf{z}_\nu\}$ is a sequence of random variables in Z satisfying a generalized central limit formula (2.3), and suppose also that $f : Z \to X$ is measurable and almost surely Hadamard differentiable with respect to the measure induced by the limit distribution $\mathfrak{z}$. Then*

$$(5.2) \qquad \sqrt{\nu}[f(\mathbf{z}_\nu) - f(\bar{z})] \xrightarrow[\mathcal{D}]{} f'(\bar{z}; \mathfrak{z})$$

as random variables in X.

Proof. As in the proof of Theorem 2.2, apply Billingsley [**6**; Theorem 5.5] to the sequence $\{\mu_\nu \varphi_\nu^{-1}\}$, where φ_ν is the "difference quotient"

$$\varphi_\nu(z) = [f(\bar{z} + \frac{1}{\sqrt{\nu}} z) - f(\bar{z})]\sqrt{\nu}, \qquad \nu = 1, 2, \ldots.$$

The condition that (5.1) holds μ-almost surely is precisely that required to ensure that $\mu_\nu \varphi_\nu^{-1}$ weak*-converges to $\mu f'(\bar{z}; \cdot)^{-1}$, establishing (5.2). □

Acknowledgements

This research incorporates insights based on several important discussions with Professors R.T. Rockafellar and R. J.-B. Wets, for which I am grateful.

References.

[1] A.J. King, "Central limit theory for Lipschitz mappings", manuscript submitted to *Math. of O.R.*, 1987.

[2] 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.

[3] R.T. Rockafellar, "Pseudo-differentiability of set-valued mappings and its applications in optimization", manuscript, 1987.

[4] 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.

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

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

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

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

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

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

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

[12] R.T. Rockafellar, "Directional differentiability of the optimal value function in a nonlinear programming problem", *Math. Prog. Study* **21**(1984), 213–226.