



HAL
open science

On a uniform law of large numbers for random sets and subdifferentials of random functions

Pedro Terán

► **To cite this version:**

Pedro Terán. On a uniform law of large numbers for random sets and subdifferentials of random functions. *Statistics and Probability Letters*, 2009, 78 (1), pp.42. <10.1016/j.spl.2007.05.026>. <hal-00580101>

HAL Id: hal-00580101

<https://hal.science/hal-00580101v1>

Submitted on 26 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Author's Accepted Manuscript

On a uniform law of large numbers for random sets
and subdifferentials of random functions

Pedro Terán

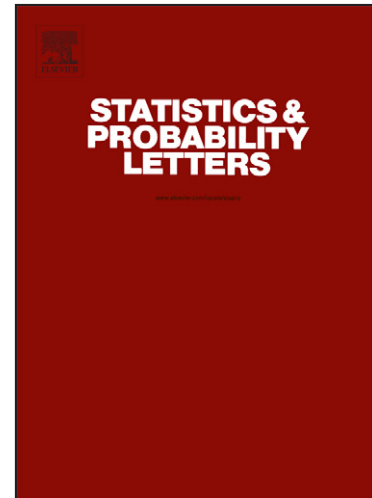
PII: S0167-7152(07)00193-9
DOI: doi:10.1016/j.spl.2007.05.026
Reference: STAPRO 4673

To appear in: *Statistics & Probability Letters*

Received date: 9 February 2007
Accepted date: 23 May 2007

Cite this article as: Pedro Terán, On a uniform law of large numbers for random sets and subdifferentials of random functions, *Statistics & Probability Letters* (2007), doi:10.1016/j.spl.2007.05.026

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/stapro

On a uniform law of large numbers for random sets and subdifferentials of random functions

Pedro Terán

Facultad de Ciencias Económicas y Empresariales

Unidad Docente de Métodos Estadísticos

Grupo Decisión Multicriterio Zaragoza

Universidad de Zaragoza

Gran Vía, 2. E-50005 Zaragoza, Spain

e-mail: `teran@unizar.es`

Phone number: (+34) 976 76 10 00, ext. 4675.

Fax number: (+34) 976 76 17 70

Abstract

In this paper, we strengthen the convergence in a uniform law of large numbers for random upper semicontinuous multifunctions of Shapiro and Xu. The proof is based on an abstract law of large numbers in a metric space endowed with a convex combination operation. Convergence in the Hausdorff metric is obtained, whereas the original result presented a weakened form of convergence of excess functionals. As a consequence, another law of large numbers for subdifferentials of random functions is improved as well.

Keywords: Clarke generalized gradient. Uniform Strong Law of Large Numbers. Upper semicontinuous multifunction. Subdifferential.

1 Introduction

Shapiro and Xu [6] have recently presented a uniform law of large numbers for multifunctions, motivated by consistency analysis of stationary points

in non-convex non-smooth stochastic optimization. Stationary points of the problem

$$\min_x Ef(x, \xi)$$

are approximated by those of

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

but the non-uniqueness of such points requires replacing the ordinary law of large numbers for random variables by a law of large numbers for random sets (or multifunctions) given by subdifferentials of the random functions above. The most salient feature of their result is that the LLN is needed to be uniform over compact sets but the subdifferential (Clarke generalized gradient) at a point does not vary continuously, it is only *upper semicontinuous* in the usual sense of multifunctions.

This leads Shapiro and Xu to working with a one-sided version of the Hausdorff metric (convergence of excess functions). In [6, Theorem 2], they essentially prove that, if X is a random multifunction, upper semicontinuous, then almost surely the average $n^{-1} \sum X_i(x, \cdot)$ is asymptotically covered by the unions of expectations of $X(y, \cdot)$ where y is near x , and the expectation of $X(x, \cdot)$ is asymptotically covered by the unions of averages at points near x , uniformly as x ranges over a compact metric space.

Thus two (asymmetric) convergences are proven, each being a relaxed form of half what it takes to prove Hausdorff convergence. This is to account for X being only u.s.c. Of course, if X is assumed to be continuous then one can do without those relaxations and convergence in the Hausdorff metric is reached [6, Theorem 4].

They remark [6, Remark 2] that continuity might be needed to obtain convergence in the Hausdorff metric. That is eminently reasonable, yet we set out to prove that it is not correct. In fact, the uniform law of large numbers can hold without relaxing Hausdorff convergence if X is only upper semicontinuous, and we will show that even this is not strictly necessary. This leads to an important sharpening of their basic result.

The uniform law of large numbers will be obtained as an application of an abstract law of large numbers in metric spaces due to Molchanov and the present author [7]. That paper presents a set of five axiomatic conditions under which an otherwise arbitrary convex combination operation admits

the construction of an expectation operator satisfying the strong law of large numbers.

In this paper, the carrier space will be that of upper semicontinuous multifunctions on a compact space. Proving the law of large numbers reduces to showing that the natural definition of convex combinations fulfils those five conditions.

2 Convex combination spaces

Let (\mathbf{E}, d) be a metric space endowed with a *convex combination operation* $[\cdot, \cdot]$ yielding a point of \mathbf{E} for each tuple of weights $(\lambda_1, \dots, \lambda_n)$ and points (u_1, \dots, u_n) :

$$(\lambda_1, \dots, \lambda_n, u_1, \dots, u_n) \mapsto [\lambda_1, u_1; \dots; \lambda_n, u_n],$$

where $n \geq 2$, $\lambda_i > 0$, $\sum \lambda_i = 1$, $u_i \in \mathbf{E}$. The notation $[\lambda_1, u_1; \dots; \lambda_n, u_n]$ and its shorthand $[\lambda_i, u_i]_{i=1}^n$ convey the usual meaning of $\lambda_1 u_1 + \dots + \lambda_n u_n$ and $\sum_{i=1}^n \lambda_i u_i$ but \mathbf{E} is not assumed to have an addition.

Terán and Molchanov [7] call \mathbf{E} a *convex combination space* if the following axioms are satisfied:

- (i) (Commutativity) For every permutation σ of $\{1, \dots, n\}$, $[\lambda_i, u_i]_{i=1}^n = [\lambda_{\sigma(i)}, u_{\sigma(i)}]_{i=1}^n$;
- (ii) (Associativity) $[\lambda_i, u_i]_{i=1}^{n+2} = [\lambda_1, u_1; \dots; \lambda_n, u_n; \lambda_{n+1} + \lambda_{n+2}, [\frac{\lambda_{n+1} u_{n+1} + \lambda_{n+2} u_{n+2}}{\lambda_{n+1} + \lambda_{n+2}}; u_{n+2}]]_{j=1}^2$;
- (iii) (Continuity) If $u, v \in \mathbf{E}$ and $\lambda^{(k)} \rightarrow \lambda \in (0, 1)$ as $k \rightarrow \infty$, then

$$[\lambda^{(k)}, u; 1 - \lambda^{(k)}, v] \rightarrow [\lambda, u; 1 - \lambda, v];$$
- (iv) (Negative curvature) For all $u_1, u_2, v_1, v_2 \in \mathbf{E}$ and $\lambda \in (0, 1)$,

$$d([\lambda, u_1; 1 - \lambda, u_2], [\lambda, v_1; 1 - \lambda, v_2]) \leq \lambda d(u_1, v_1) + (1 - \lambda) d(u_2, v_2);$$
- (v) (Convexification) For each $u \in \mathbf{E}$, there exists $\lim_{n \rightarrow \infty} [n^{-1}, u]_{i=1}^n$, which will be denoted by Ku (or $K_{\mathbf{E}}u$ if there is ambiguity).

These axioms are substantially weaker than those defining a normed linear space. It is proven in [7, Sections 4 and 5] that they suffice to construct an expectation operator E satisfying the law of large numbers. E is defined on the space $L_{\mathbf{E}}^1$ of all Borel functions X such that $Ed(X, u)$ is finite for some arbitrary $u \in \mathbf{E}$. Examples, besides the usual operations in a Banach space, include their uplift to compact sets, convolution of measures, some geometric definitions of average in metric spaces, and others (see [7, Section 9]). The aim was to find a unifying framework valid for various non-Banach spaces.

Theorem 1. *Let \mathbf{E} be a separable complete convex combination space. Then \mathbf{E} admits an expectation operator E such that*

- (a) *If $X = \sum_{j=1}^r I_{\Omega_j} u_j$ is a simple function, then $EX = [P(\Omega_j), Ku_j]_{j=1}^r$;*
- (b) *E extends by continuity to all $X \in L_{\mathbf{E}}^1$;*
- (c) *If $X \in L_{\mathbf{E}}^1$ and $\{X_i\}_{i \in \mathbf{N}}$ are pairwise i.i.d. as X , then $[n^{-1}, X_i]_{i=1}^n \rightarrow EX$ almost surely.*

3 Preliminaries

The space \mathbf{R}^d is endowed with the Euclidean norm. Its closed unit ball will be denoted by B , and its open unit ball by B° .

For any closed set $A \subset \mathbf{R}^d$, we denote by $\mathcal{K}(A)$ the family of all non-empty compact subsets of A . The *excess* of A over a set C is defined to be

$$e(A, C) = \inf\{\varepsilon > 0 \mid A \subset C + \varepsilon B\}.$$

The *Hausdorff metric* between them is

$$d_H(A, C) = \max\{e(A, C), e(C, A)\}.$$

Thus the excess is just a one-sided version of the Hausdorff metric. The *norm* of A is defined to be

$$\|A\| = d_H(A, \{0\}) = \sup_{x \in A} |x|.$$

Note that the space $\mathcal{K}(\mathbf{R}^d)$ is an example of convex combination space, having the convex hull mapping co as its convexification operator. In this case, the expectation operator is the convex hull of the Aumann expectation, see [2].

An important property of d_H is

$$d_H(\text{co } A, \text{co } C) \leq d_H(A, C).$$

In fact, the convexification operator of any convex combination space is always non-expansive [7, Proposition 3.6].

A *random closed set* is an *Effros measurable* mapping X from a measurable space (Ω, \mathcal{A}) to \mathcal{F} , i.e. the sets

$$\{\omega \in \Omega \mid X(\omega) \cap U \neq \emptyset\}$$

are measurable for each open $U \subset \mathbf{E}$. Functions from (Ω, \mathcal{A}) to $(\mathcal{K}(\mathbf{R}^d), d_H)$ are random closed sets if and only if they are Borel.

The abbreviation p.i.i.d. stands for ‘pairwise independent and identically distributed’.

4 The space of u.s.c. multifunctions

Let (\mathcal{X}, ρ) denote a compact metric space, and let \mathcal{U} be the space of *upper semicontinuous multifunctions* on \mathcal{X} with non-empty compact values, namely $u \in \mathcal{U}$ if for every $\varepsilon > 0$ and $x \in \mathbf{R}^d$ there is some $\delta > 0$ such that $\rho(x, y) < \delta$ implies $u(y) \subset u(x) + \varepsilon B^o$. Equivalently, the *excess* of $u(y)$ over $u(x)$,

$$e(u(y), u(x)) = \inf\{\varepsilon > 0 \mid u(y) \subset u(x) + \varepsilon B\}$$

goes to 0 as $y \rightarrow x$.

We define naturally the convex combination of several u.s.c. multifunctions u_i to take on the values

$$[\lambda_i, u_i]_{i=1}^n(x) = \sum_{i=1}^n \lambda_i u_i(x).$$

The space \mathcal{U} is endowed with the uniform metric

$$H(u, v) = \sup_{x \in X} d_H(u(x), v(x)).$$

Given an underlying measurable space (Ω, \mathcal{A}) , a *random u.s.c. multifunction* is defined in [6] to mean a mapping $X : \Omega \times \mathcal{X} \rightarrow \mathcal{K}(\mathbf{R}^d)$ such that $X(\cdot, x)$ is measurable for each $x \in \mathcal{X}$. It can be equivalently regarded as a random

element of \mathcal{U} with its cylindrical σ -algebra, and we will not really distinguish between both descriptions. For instance, in this context a ‘Borel function’ will be a random element of \mathcal{U} with the Borel σ -algebra generated by the metric H .

Notice that, even if \mathcal{X} is compact metric and $\mathcal{K}(\mathbf{R}^d)$ is separable, (\mathcal{U}, H) is not separable in general. In order to apply Theorem 1, we will need to impose the assumption that the range (modulo a null set) of the random multifunction is separable. This is equivalent to demanding that the random multifunctions can be approximated by simple \mathcal{U} -valued mappings; under the Continuum Hypothesis or even weaker set-theoretical assumptions, every Borel \mathcal{U} -valued function has that property¹. However, this is not a restriction to our main aim of showing that LLN convergence in the Hausdorff metric can be achieved, with a certain generality, for non-continuous multifunctions.

Theorem 2. *The space (\mathcal{U}, H) is a convex combination space where the convexification operator $K_{\mathcal{U}}$ is given by*

$$(K_{\mathcal{U}}u)(x) = \text{co } u(x).$$

Proof. Set

$$R(u) = \bigcup_{x \in \mathcal{X}} u(x).$$

We begin by proving that $R(u)$ is compact.

(a) $R(u)$ is bounded. By the upper semicontinuity of u , for each $x \in \mathcal{X}$ there exists some $\delta(x) > 0$ such that

$$\rho(x, y) < \delta(x) \implies u(y) \subset u(x) + B^o.$$

Since $\{x + \delta(x)B^o\}_{x \in \mathcal{X}}$ is an open cover of \mathcal{X} , it has a finite subcover $\{x_i + \delta(x_i)B^o\}_{i=1}^k$. Consequently, for each $y \in \mathcal{X}$,

$$u(y) \subset \bigcup_{i=1}^k [u(x_i) + \varepsilon B^o],$$

¹Being a family of functions from \mathcal{X} to $\mathcal{K}(\mathbf{R}^d)$, the space (\mathcal{U}, H) has separability character bounded above by the cardinality of $\mathcal{P}(\mathcal{X} \times \mathcal{K}(\mathbf{R}^d))$. Since both factors are separable metric spaces, that is in turn bounded by $2^{\mathfrak{c}}$. Under the hypothesis that \mathfrak{c} is a measure-free cardinal (much weaker than the Continuum Hypothesis), $2^{\mathfrak{c}}$ is so also and [4, Proposition 438D] or [3, Appendix! C, Theorem 3.2] ensure that every \mathcal{U} -valued Borel function has essentially separable range.

so that $R(u)$ is included in a finite union of bounded sets.

(b) $R(u)$ is closed. Take $z_n \rightarrow z$ with $\{z_n\}_n \subset R(u)$. Then there exist $x_n \in \mathcal{X}$ such that $z_n \in u(x_n)$ for each n . Because \mathcal{X} is compact metric, $\{x_n\}_n$ has a subsequence (still denoted x_n) converging to some $x \in \mathcal{X}$.

Since u is upper semicontinuous, for any given $\varepsilon > 0$ and all sufficiently large n we have

$$z_n \in u(x_n) \subset u(x) + \varepsilon B.$$

But the latter set is closed, so $z \in u(x) + \varepsilon B$. The arbitrariness of ε yields $z \in u(x) \subset R(u)$.

We proceed now to checking properties (i) through (v). Properties (i) and (ii) are immediate. The compactness of $R(u)$ for $u \in \mathcal{U}$ is the key tool in order to prove properties (iii) and (v).

As to (iii), we have to show that

$$\sup_{x \in \mathcal{X}} d_H(\lambda^{(k)} u(x) + (1 - \lambda^{(k)}) v(x), \lambda u(x) + (1 - \lambda) v(x)) \rightarrow 0$$

whenever $\lambda^{(k)} \rightarrow \lambda \in (0, 1)$. To that purpose, it is enough to prove

$$\sup_{A \in \mathcal{K}(R(u)), C \in \mathcal{K}(R(v))} d_H(\lambda^{(k)} A + (1 - \lambda^{(k)}) C, \lambda A + (1 - \lambda) C) \rightarrow 0.$$

Reasoning by contradiction, assume the contrary. Then there exists a subsequence $\{k'\}_k$, some $\varepsilon > 0$ and some sets $A_{k'} \in \mathcal{K}(R(u)), C_{k'} \in \mathcal{K}(R(v))$ such that

$$\varepsilon < d_H(\lambda^{(k')} A_{k'} + (1 - \lambda^{(k')}) C_{k'}, \lambda A_{k'} + (1 - \lambda) C_{k'})$$

for all k .

By the compactness of $R(u)$ and $R(v)$, the family $\mathcal{K}(R(u)) \times \mathcal{K}(R(v))$ is d_H -compact [5, Propositions 1-4-1 and 1-4-4]. Then we take a further subsequence $\{k''\}_k$ such that $A_{k''} \rightarrow A, C_{k''} \rightarrow C$ for some $A \in \mathcal{K}(R(u)), C \in \mathcal{K}(R(v))$. By the triangle inequality,

$$\begin{aligned} \varepsilon &< d_H(\lambda^{(k'')} A_{k''} + (1 - \lambda^{(k'')}) C_{k''}, \lambda^{(k'')} A + (1 - \lambda^{(k'')}) C) + \\ &\quad d_H(\lambda^{(k'')} A + (1 - \lambda^{(k'')}) C, \lambda A + (1 - \lambda) C) + \\ &\quad d_H(\lambda A + (1 - \lambda) C, \lambda A_{k'} + (1 - \lambda) C_{k'}) = (I) + (II) + (III). \end{aligned}$$

But

$$(I) \leq \lambda^{(k'')} d_H(A_{k''}, A) + (1 - \lambda^{(k'')}) d_H(C_{k''}, C) \rightarrow 0,$$

similarly (III) $\rightarrow 0$ and finally

$$\begin{aligned} (II) &\leq d_H(\lambda^{(k'')}A, \lambda A) + d_H((1 - \lambda^{(k'')})C, (1 - \lambda)C) \\ &\leq |\lambda^{(k'')} - \lambda| \cdot \|A\| + |\lambda^{(k'')} - \lambda| \cdot \|C\| \rightarrow 0. \end{aligned}$$

This is a contradiction, so (iii) is proven.

Property (iv), negative curvature of \mathcal{U} , follows easily from that of $\mathcal{K}(\mathbf{R}^d)$:

$$\begin{aligned} &H(\lambda u_1 + (1 - \lambda)u_2, \lambda v_1 + (1 - \lambda)v_2) = \\ &\sup_{x \in \mathcal{X}} d_H(\lambda u_1(x) + (1 - \lambda)u_2(x), \lambda v_1(x) + (1 - \lambda)v_2(x)) \leq \\ &\sup_{x \in \mathcal{X}} (\lambda d_H(u_1(x), v_1(x)) + (1 - \lambda)d_H(u_2(x), v_2(x))) \leq \\ &\leq \lambda H(u_1, v_1) + (1 - \lambda)H(u_2, v_2). \end{aligned}$$

Finally, in order to establish property (v), the existence of $K_{\mathcal{U}}$, we need to prove

$$H(n^{-1} \sum_{i=1}^n u, K_{\mathcal{U}}u) \rightarrow 0.$$

Similar to (iii), it suffices to show that

$$\sup_{A \in \mathcal{K}(R(u))} d_H(n^{-1} \sum_{i=1}^n A, \text{co } A) \rightarrow 0.$$

Assume the contrary; a similar compactness argument yields a value $\varepsilon > 0$, a subsequence $\{n''\}_n$ and sets $A_{n''}, A \in \mathcal{K}(R(u))$ such that

$$d_H(A_{n''}, A) \rightarrow 0$$

and

$$\begin{aligned} \varepsilon &< d_H((n'')^{-1} \sum_{i=1}^{n''} A_{n''}, \text{co } A_{n''}) \leq \\ &d_H((n'')^{-1} \sum_{i=1}^{n''} A_{n''}, (n'')^{-1} \sum_{i=1}^{n''} A) + d_H((n'')^{-1} \sum_{i=1}^{n''} A, \text{co } A) \\ &+ d_H(\text{co } A, \text{co } A_{n''}) = (IV) + (V) + (VI). \end{aligned}$$

Then

$$(IV) \leq n'' \cdot (n'')^{-1} d_H(A_{n''}, A) \rightarrow 0,$$

(V) goes to 0 by [1] and

$$(VI) \leq d_H(A, A_{n''}) \rightarrow 0.$$

A contradiction obtains, so (v) is proven. \square

We will also need the following lemma about \mathcal{U} .

Lemma 3. *The metric space (\mathcal{U}, H) is complete.*

Proof. Since $\mathcal{K}(\mathbf{R}^d)^{\mathcal{X}}$, with the uniform metric, is complete, it suffices to prove that \mathcal{U} is closed in that space. The proof is rather standard. Let $\{u_n\}_n \subset \mathcal{U}$ H -converge to some $u \in \mathcal{K}(\mathbf{R}^d)^{\mathcal{X}}$, and fix $\varepsilon > 0$. Then, $d_H(u_n(y), u(y)) < \varepsilon/3$ for some $n \in \mathbf{N}$ and all $y \in \mathcal{X}$.

Let $x \in \mathcal{X}$. By the upper semicontinuity of u_n there exists $\delta > 0$ such that $\rho(x, y) < \varepsilon$ implies $u_n(x) \subset u_n(y) + (\varepsilon/3)B^o$. We deduce that

$$u(x) \subset u_n(x) + \frac{\varepsilon}{3}B^o \subset u_n(y) + \frac{2\varepsilon}{3}B^o \subset u(y) + \varepsilon B^o.$$

\square

5 Laws of large numbers

It is not hard now to derive the LLN for random u.s.c. multifunctions. We adopt the following notation from [6]: if X_n are p.i.i.d. as X ,

$$S_n(x) = n^{-1} \sum_{i=1}^n X_i(x, \cdot),$$

$$\mathcal{E}(x) = E \operatorname{co} X(x, \cdot).$$

Theorem 4. *Let X be a random u.s.c. multifunction whose range is essentially H -separable, and let $\{X_n\}_n$ be p.i.i.d. as X . If there exists an integrable function κ such that*

$$\|X(x, \omega)\| \leq \kappa(\omega) \text{ for all } (x, \omega) \in \mathcal{X} \times \Omega,$$

then

$$\sup_{x \in \mathcal{X}} d_H(S_n(x), \mathcal{E}(x)) \rightarrow 0$$

almost surely.

Proof. For each $n \in \mathbf{N}$ there exists a null set $N_n \subset \Omega$ such that the image of $\Omega \setminus N_n$ by X_n is H -separable. The H -closure of $\bigcup_n X_n(\Omega \setminus N_n)$, denoted \mathcal{V} , is then a complete (by Lemma 3) separable convex combination space.

Theorems 2 and 1 yield now the following. Let E be the expectation operator in the space \mathcal{V} . Then $H(n^{-1} \sum_{i=1}^n X_i, EX) \rightarrow 0$ almost surely, under the assumption that $X \in L^1_{\mathcal{V}}$, namely $EH(X, u) < \infty$ for some (arbitrary) $u \in \mathcal{U}$.

But taking u to be the null function, we obtain

$$H(X(\omega), 0) = \sup_{x \in \mathcal{X}} \|X(x, \omega)\| \leq \kappa(\omega)$$

so the existence of κ implies that $X \in L^1_{\mathcal{V}}$ indeed holds.

Moreover, one checks routinely, using approximations by simple functions, that $\mathcal{E}(x) = (EX)(x)$ for every $x \in \mathcal{X}$, so convergence

$$H(n^{-1} \sum_{i=1}^n X_i, EX) \rightarrow 0$$

is the same thing as

$$\sup_{x \in \mathcal{X}} d_H(S_n(x), \mathcal{E}(x)) \rightarrow 0.$$

□

This should be compared to [6, Theorem 2], where it is proven (without the separability restriction on the range of X) that, for any $r > 0$,

$$\sup_{x \in \mathcal{X}} e(S_n(x), \bigcup_{\rho(x,y) \leq r} \mathcal{E}(y)) \rightarrow 0$$

and

$$\sup_{x \in \mathcal{X}} e(\mathcal{E}(x), \bigcup_{\rho(x,y) \leq r} S_n(y)) \rightarrow 0.$$

Clearly, those statements become weaker as r increases. The convergence we prove is equivalent to taking $r = 0$ above. Shapiro and Xu [6, Theorem 4]

get to prove convergence in the case $r = 0$ by assuming additionally that X is almost surely continuous. But then its range is separable, because the whole space of continuous functions from \mathcal{X} to the metric space $\mathcal{K}(\mathbf{R}^d)$ is separable. Thus Theorem 4 includes [6, Theorem 4].

Remark 1. From the proof of Theorem 2 one can perceive that upper semi-continuity is used only to ensure that $\bigcup_{x \in \mathcal{X}} u(x)$ is compact for $u \in \mathcal{U}$, therefore it can actually be replaced by the latter condition, which is *not linked to continuity anymore*. Thus continuity properties seem not to be really essential to this problem, contrary to what might seem from Shapiro and Xu's approach.

Remark 2. On the other hand, [6, Corollary 1] considers the point-valued case in which $X(x, \cdot)$ is almost surely a singleton, proving

$$\sup_{x \in \mathcal{X}} e(S_n(x), \mathcal{E}(x)) \rightarrow 0 \text{ a.s.}$$

In this case the excess is the same as the Hausdorff metric, but observe that every u.s.c. singleton-valued function u identifies with the continuous function \tilde{u} given by $u(x) = \{\tilde{u}(x)\}$. But we have just shown the same conclusion by requiring only compact range: now $\bigcup_x u(x)$ is just the range of \tilde{u} .

The application of this result to Clarke generalized gradients of random functions mirrors that in [6]. Since it follows by the same techniques therein, proofs are omitted. The reader is referred to [6, Section 3], on which the following is heavily based, for details.

Let $\{\xi_n : \Omega \rightarrow \mathbf{R}^d\}_n$ be a p.i.i.d. sequence supported on a closed set Ξ . Consider a Carathéodory function $F : \mathbf{R}^k \times \Xi \rightarrow \mathbf{R}$, such that the expected value function $f(x) := EF(x, \xi)$ is finite, and the sample average function $\hat{f}_n(x) = n^{-1} \sum_{i=1}^n F(x, \xi_i)$.

Take \mathcal{X} to be in $\mathcal{K}(\mathbf{R}^d)$ and set $F_\xi = F(\cdot, \xi)$. We assume that F_ξ is Lipschitzian on a neighbourhood of \mathcal{X} , and denote its Clarke generalized gradient at $x \in \mathcal{X}$ by $\partial F_\xi(x)$.

With this notation, we have the following variant of [6, Theorem 5] with strengthened convergence. Note that we are applying Theorem 4 with $\partial F_{\xi_i(\omega)}(x)$ in the role of $X_i(x, \omega)$.

Theorem 5. *With the definitions above, if the following conditions are met:*

(i) *There exists an integrable function $\kappa : \Xi \rightarrow \mathbf{R}$ such that*

$$|F(x, \xi) - F(y, \xi)| \leq \kappa(\xi)|x - y| \quad \text{for all } x, y \in \mathcal{X}, \xi \in \Xi,$$

- (ii) $F(\cdot, \xi)$ is regular at every $x \in \mathcal{X}$, for almost every $\xi \in \Xi$,
- (iii) The range of $\xi \mapsto \partial F_\xi$ is essentially separable;

then

$$\sup_{x \in \mathcal{X}} d_H(\partial \hat{f}_n(x), \partial f(x)) \rightarrow 0$$

almost surely.

Remark 3. This covers [6, Theorem 6] as well.

Acknowledgements: I am grateful to Alexander Shapiro and Huifu Xu for their kind remarks. Research has been partially funded by the Spanish *Ministerio de Educación y Ciencia*, under its research grants MTM2005-02254 and TSI2005-02511, and the *Gobierno de Aragón*, under its research grant PM2004-052. This paper was conceived during a 6-hour stay at the *Reina Sofía* airport in Tenerife.

References

- [1] Z. Artstein, J. C. Hansen (1985). Convexification in limit laws of random sets in Banach spaces. *Ann. Probab.* **13**, 307–309.
- [2] Z. Artstein, R. A. Vitale (1975). A strong law of large numbers for random compact sets. *Ann. Probab.* **3**, 879–882.
- [3] R. M. Dudley (1999). *Uniform central limit theorems*. Cambridge University Press, Cambridge.
- [4] D. H. Fremlin (2002). *Measure theory*. Volume 3. Torres Fremlin, Essex.
- [5] G. Matheron (1975). *Random sets and integral geometry*. Wiley, New York.
- [6] A. Shapiro, H. Xu (2007). Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions. *J. Math. Anal. Appl.* **325**, 1290–1299.
- [7] P. Terán, I. Molchanov (2006). The law of large numbers in a metric space with a convex combination operation. *J. Theoret. Probab.* **19**, 875–898.