IMA Annual Program Year Workshop

Computing with Uncertainty:
Mathematical Modeling, Numerical Approximation
and Large Scale Optimization of Complex Systems with Uncertainty

October 18-22, 2010

Stochastic models and application to the convergence and approximation of optimization problems

Christian L. HESS

Université Paris Dauphine

email address: Christian.Hess@dauphine.fr

personal website: http://www.snook.fr/christian.hess/

Stochastic models and application to the convergence and approximation of optimization problems

Contents

N.B. Only Sections 1 to 6 were presented during the oral presentation.

1 Introduction

Given $f : X \to \overline{\mathbf{R}}$ consider the optimization problem

$$(\mathcal{P}) \begin{cases} \text{minimize} & f(x) \\ x \in X \end{cases}$$

Assume that $\tilde{x}$ is a solution of $(\mathcal{P})$, i.e. $\tilde{x}$ is a minimizer

$$f(\tilde{x}) = \inf_{x \in X} f(x)$$

Approximation approach

Try to approximate problem $(\mathcal{P})$ by a sequence of more tractable optimization problems.

For each $n \geq 1$ consider the function $f_n : X \to$ and the optimization problem

$$(\mathcal{P}_n) \begin{cases} \text{minimize} & f_n(x) \\ x \in X \end{cases}$$

Look for a notion of convergence (not too strong) such that

$$f_n \to f$$

implies

$$\inf_{x \in X} f_n \to \inf_{x \in X} f$$

and if x_n is a solution of $(\mathcal{P}_n)$ for all $n \geq 1$

$$x_n \to \tilde{x}$$

Remarks

- The space X is often assumed to be a normed linear space. It may be finite dimensional or infinite dimensional.

- The approximating problems $(\mathcal{P}_n)$ may be deterministic or stochastic

In the stochastic case, functions f_n also depend on a random parameter $\omega \in \Omega$.

2 Preliminaries

X = a finite dimensional Euclidean space, whose norm is denoted by $\|.\|$ and scalar product by $< .,. >$

X^* = the dual space (identified with X)

Almost all results presented remain valid in a more general setting with suitable modifications.

2^X = the set of all subsets of X

Two subspaces of 2^X

$\mathcal{C}(X)$ = the set of closed subsets of X

$\mathcal{C}_c(X)$ = the set of closed convex subsets of X

2.1 Addition and scalar multiplication of sets

$$C,\, C_1\,,\, C_2 \in 2^X \qquad\qquad \lambda \in \mathbf{R}$$

The *(Minkowski) sum* of C_1 and C_2

$$C_1 + C_2 = \{x_1 + x_2 : x_1 \in C_1\ ,\ x_2 \in C_2\}$$

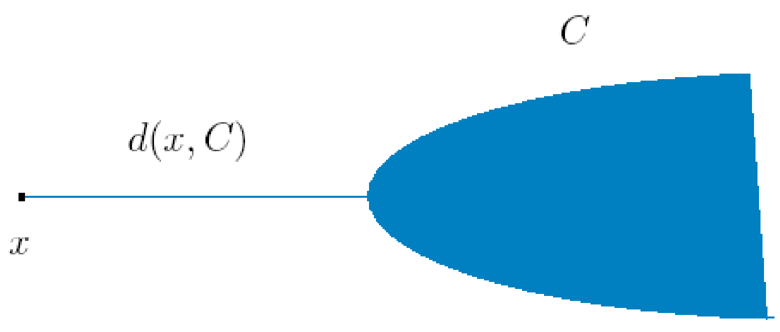
The *scalar product* of λ by C

$$\lambda\, C = \{\lambda\, x : x \in C\}$$

2.2 The distance function and the support function

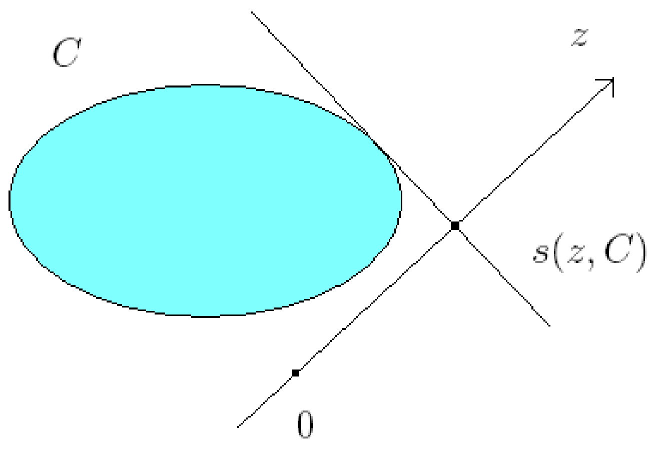
The distance function

$$d(x, C) = \inf_{y \in C} \|x - y\| \qquad x \in X \qquad C \in 2^X$$



The support function

$$s(z, C) = \sup_{x \in C} < x, z > \qquad z \in X^* \ \text{(the dual space of } X\text{)}$$



2.3 Convergence of sets

The Painlevé-Kuratowski convergence

$$(C_n)_{n\geq 1} \text{ a sequence of subsets of } X \qquad\qquad C \in 2^X$$

The *lower (inferior) limit*

$$Li\, C_n = \{x \in X : x = \lim x_n \ , \ x_n \in C_n \qquad \forall n \geq 1\}$$

The *upper (superior) limit*

$$Ls\, C_n = \{x \in X : x = \lim x_k \ , \ x_k \in C_{n_k} \qquad \forall k \geq 1\}$$

where C_{n_k} is a subsequence of (C_n)

Properties

- The sets $Li\, C_n$ and $Ls\, C_n$ are closed.

- The following inclusion holds

$$Li\, C_n \subseteq Ls\, C_n$$

- If the C_n are convex, so is $Li\, C_n$

Definition 2.1 *The sequence* (C_n) *is said to PK-converge to* C *if*

$$C = Li\, C_n = Ls\, C_n$$

Notation

$$C_n \xrightarrow{PK} C \qquad\qquad \text{or} \qquad\qquad C = PK - \lim_{n\to+\infty} C_n$$

This is equivalent to

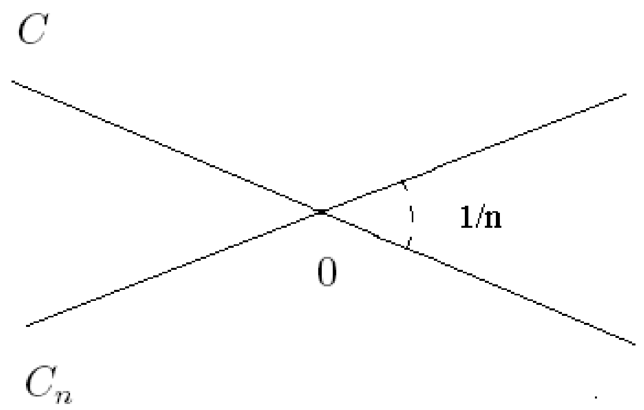
$$Ls\, C_n \subseteq C \subseteq Li\, C_n$$

Characterization of PK-convergence in term of distance functions

$$C_n \xrightarrow{PK} C \quad \Leftrightarrow \quad d(x, C) = \lim_{n \to +\infty} d(x, C_n) \qquad \forall x \in X$$

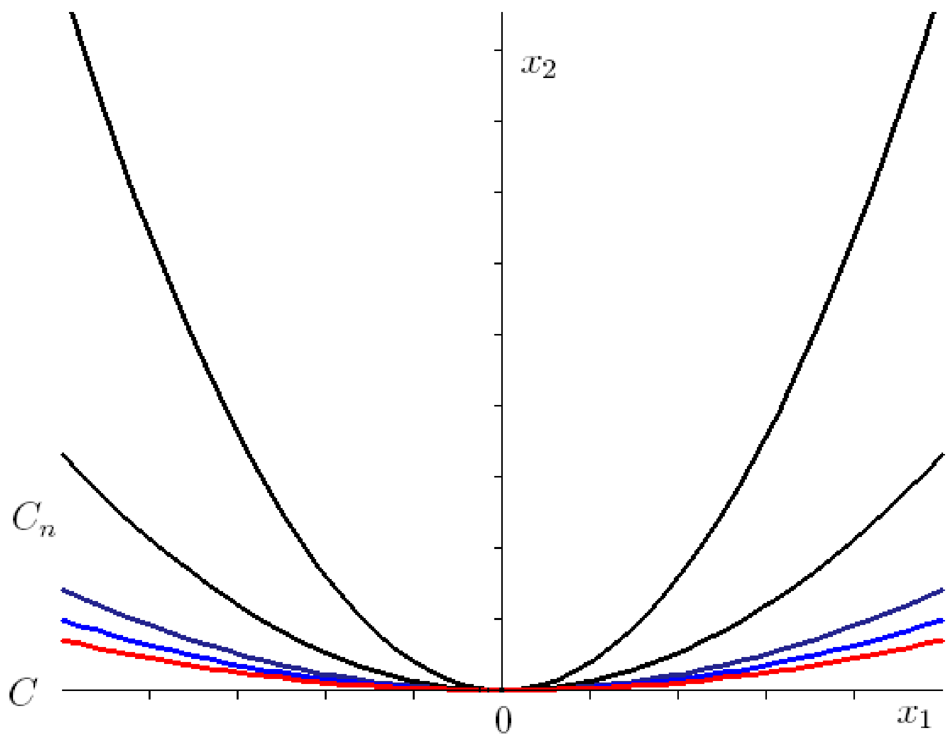
PK-convergence is the pointwise convergence of distance functions on X

Example 1



Example 2

$$C_n = \{x = (x_1\,,\,x_2) : x_2 = \frac{1}{n}\,x_1^2\,\} \qquad\qquad C = \{x = (x_1\,,\,x_2) : x_2 = 0\}$$



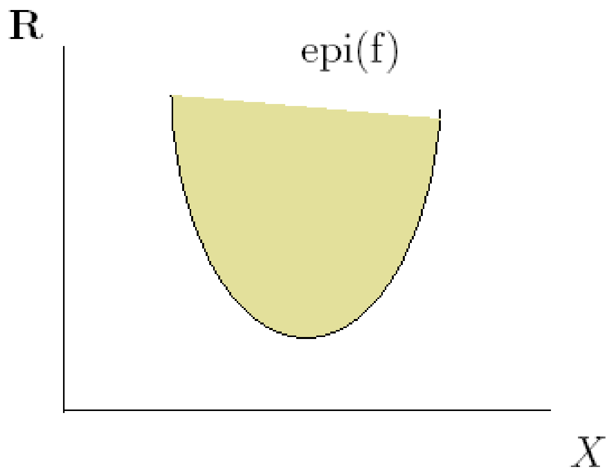
2.4 Extended real-valued functions

f an extended-real valued function

$$f : X \to \overline{\mathbf{R}}$$

Epigraph

$$\text{epi}(f) = \{(x, \lambda) \in X \times \mathbf{R} : f(x) \leq \lambda\}$$



The strict epigraph

$$\text{epi'}(f) = \{(x, \lambda) \in X \times \mathbf{R} : f(x) < \lambda\}$$

The epigraph and the strict epigraph have the same closure

$$\text{cl}\big(\text{epi'}(f)\big) = \text{cl}\big(\text{epi}(f)\big)$$

Lower semicontinuity

f is said to be *lower semi-continuous (lsc)* at $x_0 \in X$ if for each $\varepsilon > 0$ one can find $\alpha > 0$ such that

$$\|x - x_0\| < \alpha \Rightarrow f(x) \geq f(x_0) - \varepsilon$$

Properties

TFAE

(i) f is lsc at each point of X

(ii) The epigraph of f is closed

The lower semi-continuous (convex) regularization

Given $f : X \to \overline{\mathbf{R}}$, the *lsc regularization* of f, denoted by $\mathrm{cl}\, f$, is the largest lsc function smaller than or equal to f

$$(\mathrm{cl}\, f)(x) = \liminf_{y \to x} f(y) \qquad\qquad x \in X$$

The *convex lsc regularization* of f is denoted by $\mathrm{cl}\,\mathrm{co}\, f$

It is the largest convex lsc function smaller than or equal to f

Translation in terms of epigraphs

$$\mathrm{epi}\,(\mathrm{cl} f) = \mathrm{cl}\,(\mathrm{epi} f) \qquad\qquad \text{and} \qquad\qquad \mathrm{epi}\,(\mathrm{cl}\,\mathrm{co} f) = \mathrm{cl}\,\mathrm{co}\,(\mathrm{epi} f)$$

The Young-Fenchel transform (or conjugate) of f is denoted by f^*

$$X^* = \text{the dual space of X}$$

$$f^*(z) = \sup\{< x, z > -f(x) : x \in X\} \qquad z \in X^*$$

Remark

f^* is convex and lsc on X^* (as a supremum of continuous affine functions).

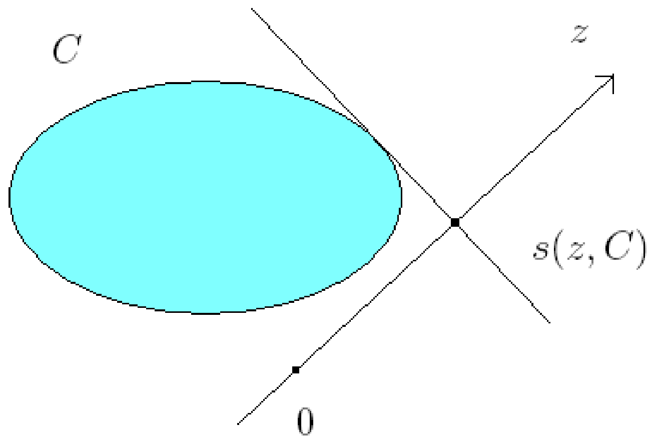
Example

The *indicator function* of C, denoted by χ_C

$$\chi_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases}$$

The conjugate of χ_C is the support function of C

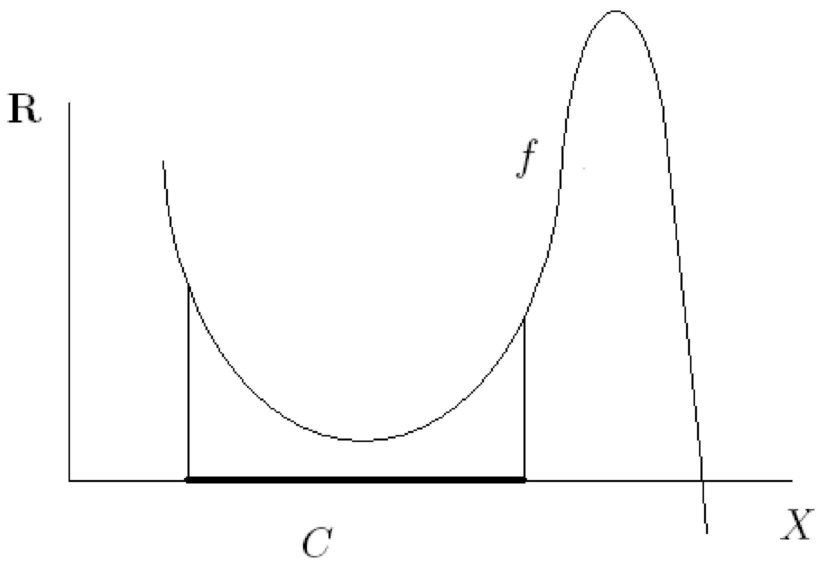
$$(\chi_C)^*(z) = s(z, C) \qquad z \in X^*$$

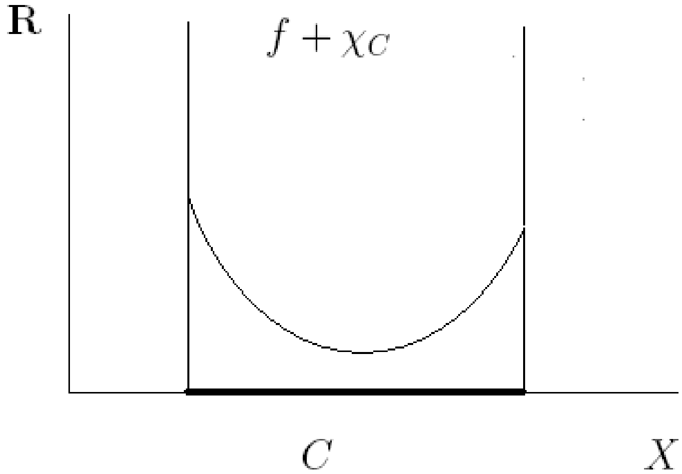


Remark: *the indicator function allows for modeling the constraints*

Minimizing f on C, a subset of X, is equivalent to minimizing $f + \chi_C$ on X

$$\min_{x \in C} f(x) = \min_{x \in X} \big(f(x) + \chi_C(x) \big)$$





The biconjugate

$$f^{**}(x) = \sup\{< x, z > - f^*(z) : z \in X^*\} \qquad\qquad x \in X$$

If f is not identically $+\infty$ and is minorized by an affine function

$$f^{**} = \mathrm{cl}\,\mathrm{co}\, f$$

Sums and scalar multiplication of epigraphs

The epi-sum (alias inf-convolution)

$$f,\, g : X \to \overline{\mathbf{R}}$$

Definition

$$(f \overset{e}{+} g)(x) = \inf_{y \in X} \{f(y) + g(x - y)\}$$

$f \overset{e}{+} g$ is also denoted $f \bigtriangledown g$

Properties

$$\boxed{\text{epi'}(f \overset{e}{+} g) = \text{epi'}(f) + \text{epi'}(g)}$$

$$\boxed{(f \overset{e}{+} g)^* = f^* + g^*}$$

Remark

The inf-convolution is involved in applications to various domains: Geology, Financial Mathematics, Environmental Sciences ...

The *right scalar multiplication* of f by $\lambda \neq 0$ is denoted $f\lambda$ and defined by

$$(f\lambda)(x) = \lambda f(x/\lambda)$$

Translation in terms of (strict) epigraphs

$$\boxed{\text{epi'}(f\lambda) = \lambda\, \text{epi'(f)}}$$

Optimization problem

$$(\mathcal{P}) \begin{cases} \text{minimize} \quad f(x) \\ \\ \quad x \in X \end{cases}$$

A solution of $(\mathcal{P})$ is a *minimizer* of f, i.e. a point $x_0 \in X$ such that

$$f(x_0) = \inf\{f(x) : x \in X\}$$

Approximate minimizers

Given $\varepsilon > 0$, an $\varepsilon-$*minimizer* of f is a point x_ε of X such that

$$f(x_\varepsilon) \leq \max\{ \inf_{x\in X} f(x) + \varepsilon\,,\ -1/\varepsilon\}$$

2.5 Epigraphical convergence

$$f_\infty\,,\, f_n : X \to \overline{\mathbf{R}} \qquad\qquad n \geq 1$$

Definition

(f_n) is said to *epi-converge* to f_∞ when n goes to $+\infty$ if

$$\text{epi}(f_n) \stackrel{PK}{\to} epi(f_\infty)$$

This is denoted

$$f_n \stackrel{epi}{\to} f_\infty \qquad\qquad \text{or} \qquad\qquad f_\infty = \text{epi - }\lim_{n\to+\infty} f_n$$

Remarks

(i) The epi-limit f_∞ is lsc because epi(f_∞) is closed.

(ii) For any sequence $(C_n)_{n\geq 1}$ of subsets of X, the following equivalence holds

$$C_n \stackrel{PK}{\to} C_\infty \qquad\qquad \Longleftrightarrow \qquad\qquad \chi_{C_n} \stackrel{epi}{\to} \chi_{C_\infty}$$

(iii) Synonym: $\Gamma-$convergence (De Giorgi, Dal Maso ...)

Variational properties of epi-convergence

For each $n \geq 1$ consider the optimization problem

$$(\mathcal{P}_n) \begin{cases} \text{minimize} \quad f_n(x) \\ \\ x \in X \end{cases}$$

Theorem 1

Hypotheses:

(a) $f_n \overset{epi}{\to} f_\infty$

(b) For each $n \geq 1$, x_n is a minimizer of f_n

Conclusion:

Under the above hypotheses

(i) If x_∞ is a cluster point of (x_n) then x_∞ is a minimizer of f_∞ and

$$f_\infty(x_\infty) = \limsup_{n \to +\infty} f_n(x_n)$$

(ii) If (x_n) converges to x_∞ then x_∞ is a minimizer of f_∞ and

$$f_\infty(x_\infty) = \lim_{n \to +\infty} f_n(x_n)$$

Remark

The same conclusions hold if hypothesis (b) is replaced by

(b') For each $n \geq 1$, x_n is an ε_n-minimizer of f_n, where (ε_n) is a sequence of positive numbers converging to 0.

Theorem 2

Epi-convergence is preserved by the Young-Fenchel transform.

$$f_n \ , \ f_\infty : X \to \overline{\mathbf{R}} \qquad\qquad f_n^* \ , \ f_\infty^* : X^* \to \overline{\mathbf{R}}$$

$$f_n \stackrel{epi}{\to} f_\infty \implies f_n^* \stackrel{epi}{\to} f_\infty^*$$

Remarks

- Epi-convergence is weaker than uniform convergence, namely uniform convergence implies epi-convergence.

- Epi-convergence cannot be compared with pointwise convergence: Epi-convergence neither implies, nor is implied by, pointwise convergence (see e.g. [[9]]).

3 Random sets and random functions

3.1 Definitions

Consider a probability space $(\Omega, \mathcal{A}, P)$

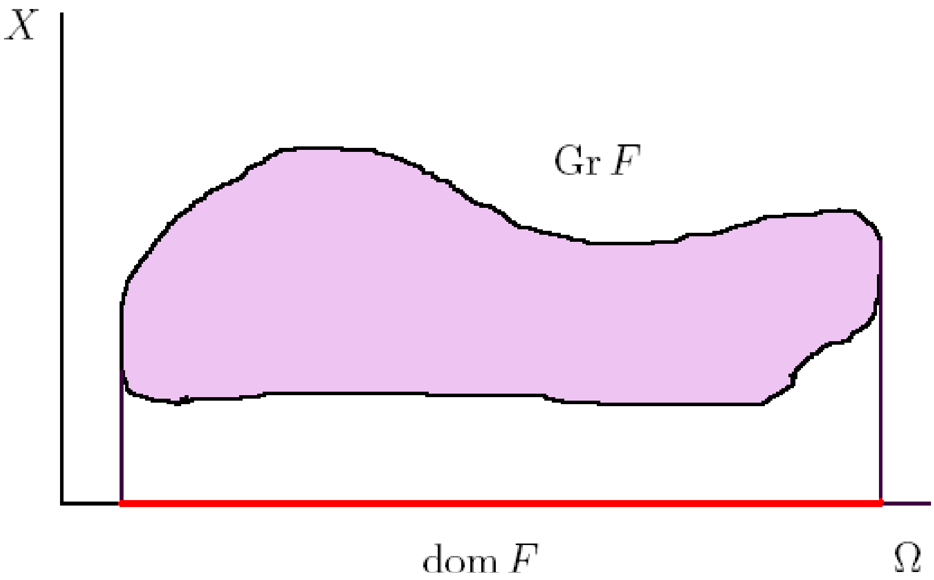
A map $F : \Omega \to 2^X$ is called a set-valued map, alias multifunction, multiapplication, correspondence ...

The domain of a multifunction

$$\mathrm{dom}\, F = \{\omega \in \Omega : F(\omega) \neq \emptyset\}$$

The graph

$$\mathrm{Gr}\, F = \{(\omega, x) \in \Omega \times X : x \in F(\omega)\}$$



Basically, a random set is a random variable whose values are subsets of X

Example: the set of constraints of a stochastic optimization problem or the epigraph of a random function (in $X \times \mathbf{R}$).

Definition

Assume that a $\sigma-$field $\mathcal{T}$ is given on 2^X.

A *random set* (r.s.) is a measurable map F from $(\Omega, \mathcal{A})$ into $(2^X, \mathcal{T})$

$$F^{-1}(S) = \{\omega \in \Omega : F(\omega) \in S\} \in \mathcal{A} \qquad \forall\, S \in \mathcal{T}$$

A random set is also called a measurable multifunction, a set-valued random variable ...

Definition

The *(probability) distribution* of F, denoted by P_F is defined by

$$P_F(S) = P\big(F^{-1}(S)\big) \qquad S \in \mathcal{T}$$

Two random sets F and G are said to be *independent* if for all $S, T \in \mathcal{T}$

$$P\big(F^{-1}(S) \cap G^{-1}(T)\big) = P\big(F^{-1}(S)\big)\, P\big(G^{-1}(T)\big)$$

In the sequel, we consider the $\sigma-$field $\mathcal{E}$ generated on 2^X by the distance functions

$\mathcal{E}$ is called the Effrös $\sigma-$field. It is also generated by the family of subsets

$$S = U^- = \{C \in 2^X : C \cap U \neq \emptyset\} \qquad U \text{ open subset of } X$$

A multifunction $F : \Omega \to 2^X$ is (Effrös)- measurable if

$$F^{-1}(U^-) = \{\omega \in \Omega : F(\omega) \cap U \neq \emptyset\} \in \mathcal{A} \qquad \forall \text{ open set } U$$

If $(\Omega, \mathcal{A}, P)$ is complete and if the values of F are closed, then the following equivalence holds:

$$F \text{ is measurable} \Longleftrightarrow \mathrm{Gr}(F) \in \mathcal{A} \otimes \mathcal{B}(X)$$

where $\mathcal{B}(X)$ is the Borel$-\sigma-$field of X.

3.2 Measurable selections

Given $F : \Omega \to 2^X$, a map $f : \Omega \to X$ is a *selection* of F if

$$f(\omega) \in F(\omega) \qquad\qquad \forall\, \omega \in \operatorname{dom} F$$

The set of $\mathcal{A}$-measurable selections of F is denoted by $\mathcal{S}(F, \mathcal{A})$

Existence of measurable selections:

The Kuratowski-Ryll-Nardewski Theorem

If $F : \Omega \to \mathcal{C}(X)$ is a random closed set, then $\mathcal{S}(F, \mathcal{A})$ is non empty.

3.3 Set-valued integration

The Aumann integral

Consider the space of (classes of) $P-$integrable functions $f : \Omega \to X$

$$L^1(X) = L^1(\Omega, \mathcal{A}, P; X)$$

$$L^1(X) = \{f : \Omega \to X, \text{ measurable } : \int_\Omega \|f\|\, dP < +\infty\}$$

The set of all $\mathcal{A}-$measurable and integrable selections of F is denoted by

$$S^1(F, \mathcal{A}) = S(F, \mathcal{A}) \cap L^1(X)$$

A random set F is said to be *integrable* if $S^1(F, \mathcal{A})$ is nonempty

An integrability criterion: the random set F is integrable if and only if the distance function is integrable, i.e.

$$d(0, F(.)) \in L^1$$

This is a immediate consequence of the equality

$$\int_\Omega \inf\{\, \|x\| : x \in F(\omega)\}\, dP = \inf\Big\{ \int_\Omega \|f\|\, dP : f \in S^1(F, \mathcal{A})\Big\}$$

The *set-valued (Aumann) integral* of F is defined by

$$\int_\Omega F dP = \{\int_\Omega f dP : f \in S^1(F, \mathcal{A})\}$$

3.4 Random function

Given a function $\varphi : \Omega \times X \to \overline{\mathbf{R}}$ the multifunction defined on Ω by

$$\omega \to \operatorname{epi} \varphi(\omega, .)$$

with values in $X \times \mathbf{R}$, is called the *epigraphical multifunction* of φ.

φ is called a *random function* if its epigraphical multifunction is measurable.

Of course, φ may be defined up to a $P-$null set, but the $P-$null set is assumed not to depend on the second variable.

A convenient characterization

If $(\Omega, \mathcal{A}, P)$ is complete and $\varphi(\omega, .)$ is lower semicontinuous for all $\omega \in \Omega$, the following equivalence holds:

$\varphi : \Omega \times X \to \overline{\mathbf{R}}$ is a random function $\Longleftrightarrow$ φ is $\mathcal{A} \otimes \mathcal{B}(X)-$measurable

where $\mathcal{B}(X)$ is the Borel$-\sigma-$field of X.

4 A multivalued version of the strong law of large numbers

Theorem 3

Hypotheses :

- a random set F with values in $\mathcal{C}(X) \setminus \{\emptyset\}$

- a sequence (F_n) of independent random sets having the same distribution as F

Consider the following two statements :

(a) F is integrable, i.e. $S^1(F, \mathcal{A}) \neq \emptyset$

(b) there exists $C \in \mathcal{C}_c(X) \setminus \{\emptyset\}$ such that

$$C = PK - \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} F_i(\omega) \qquad \text{with probability 1}$$

Then, implication (a) $\Rightarrow$ (b) always holds with C given by

$$C = \operatorname{cl} \operatorname{co} \int_\Omega F \, dP$$

Conversely, if F satisfies the condition:

(NL) *there exists a fixed subset* $L \in \mathcal{C}_c(X)$ *such that*

(i) L contains no lines (it may contains half-lines only)

(ii) $F(\omega) \subseteq L \qquad \forall \omega \in \Omega$.

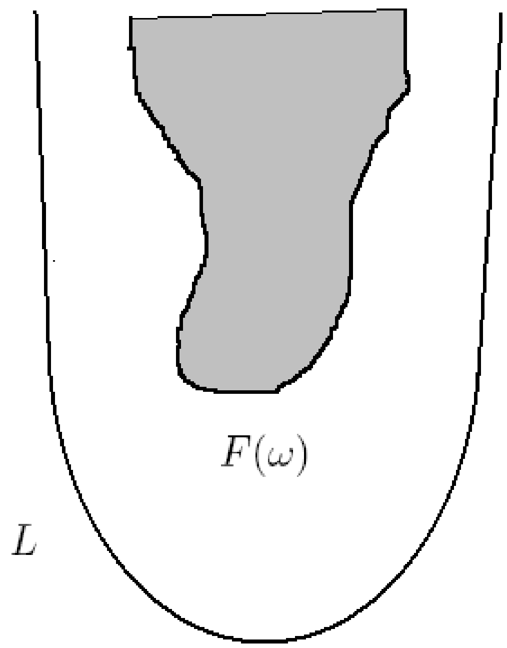
then (b) $\Rightarrow$ (a).

Remarks

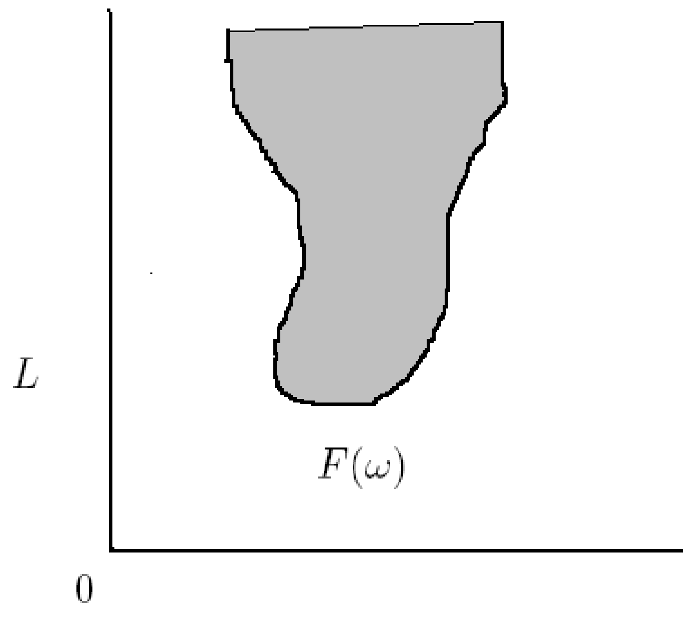
- For each $n \geq 1$, $(F_1 , \ldots , F_n)$ is a random sample of F

- *Convexification effect:* the limit is convex even if the values of F are not convex

- Condition (NL)



For example, if $X = \mathbf{R}^m$ one can take $L = \mathbf{R}^m_+$



5 Approximation of continuous epi-sum by discrete epi-sum

5.1 Integrable random functions

A random function $\varphi : \Omega \times X \to \overline{\mathbf{R}}$ is said to be *integrable* if the multifunction

$$\omega \to \operatorname{epi} \varphi(\omega, .))$$

is integrable, i.e. admits at least one integrable selection.

Proposition

TFAE

(a) the random function φ is integrable

(b) there exists $f \in \text{Ł}^1(X)$ such that $\varphi(.\,,\, f(.))^+ \in L^1$

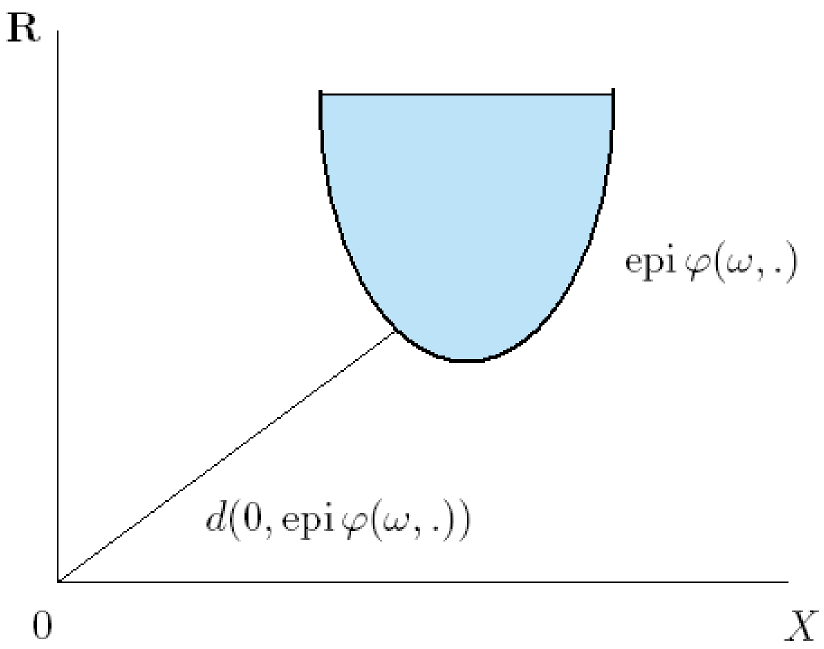
(c) the distance function (in $X \times \mathbf{R}$)

$$\omega \to d(0, \operatorname{epi}\ \varphi(\omega, .)) = \inf_{x \in X} \left(\|x\| + \varphi(\omega, x)^+ \right)$$

is a member of L^1

In statement (c), it is assumed that the norm on $X \times \mathbf{R}$ is defined by

$$\|(x, \lambda)\| = \|x\| + |\lambda|$$



5.2 The approximation result

Consider

- an integrable random lsc function $\varphi : \Omega \times X \to \overline{\mathbf{R}}$

- the multifunction F defined by

$$F(\omega) = \operatorname{epi} \varphi(\omega, .)$$

with closed values in $Y = X \times \mathbf{R}$

- a sequence $(\varphi_n)_{n \geq 1}$ of random lsc functions that are independent and have the same distribution as φ

Equivalently, the multifunctions F_n defined by

$$F_n(\omega) = \operatorname{epi} \varphi_n(\omega, .) \qquad \omega \in \Omega \qquad n \geq 1$$

are independent and have the same distribution as F

By Theorem 3 (set-valued strong law of large number)

$$(5.1) \qquad \text{cl co} \int_\Omega F\, dP = PK - \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} F_i(\omega) \qquad \text{with probability 1}$$

The values of F and F_n are contained in $X \times \mathbf{R}$.

Translation in terms of random functions

The right-hand side of (5.1) satisfies

$$\text{cl} \left\{ \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{F}_{\mathrm{n}}(\omega) \right\} = \text{epi cl } \psi_{\mathrm{n}}(\omega, .) = \text{cl epi } \psi_{\mathrm{n}}(\omega, .)$$

where

$$\psi_n(\omega, .) = \big(\varphi_1(\omega, .) \overset{e}{+} \cdots \overset{e}{+} \varphi_n(\omega, .)\big).\big(\frac{1}{n}\big) = \big\{\text{epi} \sum\nolimits_{i=1}^{n} \varphi_i(\omega, .)\big\}.\big(\frac{1}{n}\big)$$

or for all $x \in X$

$$\psi_n(\omega,x) = \inf\left\{\frac{1}{n}\ \sum_{i=1}^{n} \varphi_i(\omega,x_i) : (x_1\,,\,\ldots,x_n) \in X^n\,,\ \frac{1}{n}\ \sum_{i=1}^{n} x_i = x\right\}$$

ψ_n is a random discrete averaged epi-sum (inf-convolution).

The left-hand side of (5.1) satisfies

$$\text{cl co} \int_\Omega F\,dP = \text{cl}\,\text{co epi}\ \psi = \text{epi}\ (\text{cl}\,\text{co}\ \psi)$$

where

$$\psi(x) = \inf\left\{\int_\Omega \varphi(\omega,f(\omega))\,dP : f \in \mathrm{L}^1(X),\ \int_\Omega f(\omega)\,dP = x\right\}$$

ψ is the *continuous averaged epi-sum (inf-convolution) of* φ

In conclusion, the continuous epi-sum can express as the almost sure limit of stochastic discrete epi-sum:

Theorem 4

If $\varphi : \Omega \times X \rightarrow \overline{\mathbf{R}}$ is an integrable random lsc function and $(\varphi_n)_{n\geq 1}$ is a sequence independent identically distributed random lsc functions (with the same distribution as φ), then

(5.2) $$\text{cl co}\ \psi = \text{epi} - \lim_{n\to+\infty} \psi_n(\omega,.) \qquad \text{with probability 1}$$

Consequence

Under suitable compactness asssumptions, (approximate) minimizers for problem

$$\min\{\psi_n(\omega, x) : x \in X\}$$

will provide approximation of minimizers for the problem

$$\min\{\mathrm{cl\ co}\ \psi(x) : x \in X\}$$

or even (relaxation) for the problem

$$\min\{\psi(x) : x \in X\}$$

6 Approximation of the expectation functional by finite sums

Consider

- a random convex lsc $\varphi : \Omega \times X \to \overline{\mathbf{R}}$ satisfying a suitable minorization condition e. g. nonnegative (see below)

- a sequence $(\varphi_n)_{n\geq 1}$ of random lsc functions that are independent and have the same distribution as φ

- the multifunctions (with closed convex values in $X^* \times \mathbf{R}$) defined by

$$G(\omega) = \text{ epi } \varphi^*(\omega, .) \qquad \text{and} \qquad G_n(\omega) = \text{ epi } \varphi_n^*(\omega, .) \qquad n \geq 1$$

where the random convex lsc function φ^* is defined by

$$\varphi^*(\omega, z) = \sup\{< x, z > -\varphi(\omega, x) : x \in X\} \qquad \omega \in \Omega \qquad z \in X^*$$

- the *expectation (mean) functional* of φ which is denoted by $E(\varphi)$ and defined by

$$E(\varphi)(x) = \int_\Omega \varphi(\omega, x)\, dP \qquad x \in X$$

It is assumed that φ satisfies the following hypothesis:

(H) There exists $g_0 \in L^1(X)$ and $\alpha_0 \in L^1$ such that with probability 1

$$\varphi(\omega, x) \;\geq < x, g_0(\omega) > \,-\, \alpha_0(\omega) \qquad \forall\, x \in X$$

Translated in terms of multifunction, hypothesis **(H)** means that $(g_0\,, \alpha_0)$ is an integrable selection of the random set G

$$G(\omega) = \text{ epi } \varphi^*(\omega, .)$$

It is thus possible to apply Theorem 3 (SLLN) to the sequence (G_n), as in the proof of Theorem 4

$$\text{(6.1)} \qquad \text{cl} \int_\Omega G\, dP = PK - \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} G_i(\omega) \qquad \text{with probability 1}$$

The right-hand side of (6.1) corresponds to the random discrete averaged epi-sum

$$\theta_n(\omega, .) = \left\{\text{epi} \sum\nolimits_{i=1}^{n} \varphi_i^*(\omega, .)\right\}.\left(\frac{1}{n}\right) \qquad \omega \in \Omega$$

namely

$$\theta_n(\omega, z) = \inf \left\{\frac{1}{n} \sum_{i=1}^{n} \varphi_i^*(\omega, z_i) : (z_1\,,\, \ldots, z_n) \in X^n\,,\ \frac{1}{n} \sum_{i=1}^{n} z_i = z\right\} \qquad z \in X^*$$

The left-hand side of (6.1) corresponds to the continuous averaged epi-sum

$$\theta(z) = \inf \left\{ \int_\Omega \varphi^*(\omega, g(\omega))\, dP : g \in \text{Ł}^1(X^*),\ \int_\Omega g(\omega)\, dP = z\right\} \qquad z \in X^*$$

By Theorem 4, we have

$$\text{(6.2)} \qquad \theta = \text{epi-} \lim_{n\to+\infty} \theta_n(\omega, .) \qquad \text{with probability 1}$$

Applying the Young-Fenchel transform in both sides of (6.2) yields

$$\theta^* = \text{epi} - \lim_{n\to+\infty} \theta_n^*(\omega, .) \qquad \text{with probability 1}$$

By conjugacy, the discrete epi-sum is transformed into an ordinary sum and the continuous epi-sum is transformed into an integral. Consequently

$$E(\varphi^{**})(.) = \text{epi} - \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} \varphi_i^{**}(\omega, .)$$

Since φ is a random convex lsc function, we have proved

Theorem 5

Under the above hypotheses, especially Hypothesis **(H)**, one has

$$E(\varphi)(.) = \mathrm{epi} - \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} \varphi_i(\omega, .) \tag{6.3}$$

Remarks

(i) Alternate proof of equality (6.3).

This equality has been proved directly, without using the Young-Fenchel transform.

The main argument of the proof is based on the approximation of a random lsc function by a non decreasing sequence of random Lipschitz functions.

It remains valid for lsc **non convex** functions.

(ii) Consequence: under suitable compactness asssumptions, (approximate) minimizers for problem

$$\min\{ \frac{1}{n} \sum_{i=1}^{n} \varphi_i(\omega, x) : x \in X\}$$

will provide approximations of minimizers for the problem

$$\min\{ E(\varphi)(x) : x \in X\}$$

Applications

- Stochastic optimization and control (see [23])
- Stochastic homogeneization for heterogeneous materials (see [20])
- Statistics: almost sure convergence of estimators (see e.g. [14]).

7 Set-valued conditional expectation

As we have seen, the Aumann integral is the set-valued extension of the usual real or vector valued integral or expectation. We have shown that the strong law of large numbers is valid for this integral.

Now, this suggests to look for the extension of the set-valued conditional expectation. This will allow for the formulation of other probabilistic limit theorems such as martingale convergence and ergodic theorems.

Given a sub$-\sigma-$field $\mathcal{B}$ of $\mathcal{A}$, and an $\mathcal{A}-$measurable integrable random set F with values in $\mathcal{C}(X)$, Hiaï and Umegaki [19] showed the existence of a $\mathcal{B}-$measurable integrable random set G such that

$$S^1(G, \mathcal{B}) = \mathrm{cl}\{E(f/B) : f \in S^1(F, \mathcal{A})\}$$

where the closure is taken in $L^1(X)$.

G is the (set-valued) conditional expectation of F relative to $\mathcal{B}$ and is denoted by $E(F|\mathcal{B})$. As in the single-valued case, it is defined up to a $P-$null set.

Special case: if $\mathcal{B} = \{\Omega, \emptyset\}$, then $E(F|\mathcal{B}) = E(F) = \mathrm{cl} \int_\Omega F.dP$

Some basic properties

Proposition 7.1 *If F and G are two integrable random sets with values in $\mathcal{C}(X)$, and $\mathcal{B}$ a sub$-\sigma-$field of $\mathcal{A}$, then the following properties hold :*

(a) $E(F + G/\mathcal{B}) = \mathrm{cl}\{E(F/\mathcal{B}) + E(G/\mathcal{B})\}$ *a. s.*

(b) if r is a real $\mathcal{B}-$measurable function such that $r\,F$ is integrable, then

$$E(r\,F/\mathcal{B}) = rE(F/\mathcal{B}) \qquad a.\ s.$$

(c) $E(\mathrm{cl\,co}\,F/\mathcal{B}) = \mathrm{cl\,co}\,E(F/\mathcal{B})$ *a. s.*

(d) if $h : \Omega \to X^*$ is a bounded $\mathcal{B}-$measurable function then

$$s\big(h, E(F/\mathcal{B})\big) = E\big(s(h,F)/\mathcal{B}\big) \qquad a.\ s.$$

(e) Let F be $\mathcal{B}-$measurable, with values in $\mathcal{C}_c(X)$, and r an $\mathcal{A}-$measurable non-negative function such that $r\,F$ is integrable, then

$$E(r\,F/\mathcal{B}) = E(r/\mathcal{B})F \qquad a.\ s.$$

In particular,

$$E(F/\mathcal{B}) = F$$

Remarks

(i) Property (e) is no longer true if the values of F are not convex.

Counter-example

Assume that $(\Omega, \mathcal{A}, P)$ is nonatomic, $F(\omega) = \{0,1\} \subseteq \mathbf{R}$ and $\mathcal{B} = \{\Omega, \emptyset\}$.

By the Lyapunov Convexity Theorem: $E(F/\mathcal{B}) = \int_\Omega F.dP = [0,1] \neq F$

(ii) In Property (e), the assumption of r being nonnegative cannot be removed (see [19]).

8 Set-valued martingales

Let $(\mathcal{B}_n)_{n\geq 1}$ be a non decreasing sequence of sub$-\sigma-$fields of $\mathcal{A}$

A sequence $(F_n)_{n\geq 1}$ of measurable multifunctions with values in $\mathcal{C}_c(X)$ is said to be *adapted* to $(\mathcal{B}_n)$ if, for any $n \geq 1$, F_n is $\mathcal{B}_n-$measurable,
i.e. $F_n^{-1}(U^-) \in \mathcal{B}_n$ for every open set U of X.

Such an adapted sequence is said to be a *set-valued martingale* if the following two conditions hold :

(a) for all $n \geq 1$, $S^1(F_n, \mathcal{B}_n)$ is non empty

(b) for all $n \geq 1$, $F_n = E(F_{n+1}/\mathcal{B}_n)$

Set-valued submartingales and supermartingales

- If condition (b) is replaced with

$$F_n \subseteq E(F_{n+1}/\mathcal{B}_n)$$

(F_n) is said to be a *set-valued submartingale*

- If condition (b) is replaced with

$$F_n \supseteq E(F_{n+1}|\mathcal{B}_n))$$

(F_n) is said to be a *set-valued supermartingale*

Remark

If a set-valued sub- (or super-) martingale is single-valued in X, then it is a martingale (in the classical sense).

Theorem 6

Let $(F_n)_{n\geq 1}$ be a multivalued supermartingale, with values in $\mathcal{C}_c(X)$, satisfying

$$\text{(8.1)} \qquad \sup_{n\geq 1} E\, d(0, F_n) < +\infty$$

Then, one can find an integrable measurable multifunction F, with values in $\mathcal{C}_c(X)$, such that

$$F(\omega) = PK - \lim_{n\geq 1} F_n(\omega) \qquad \text{a. s.}$$

Remarks

- Condition (8.1) is the analog of Doob's condition that appears in the usual a.s. convergence theorems for real-valued martingales.

- Adaptations of the previous result to martingales of random lsc functions exist. It can be also useful in stochastic optimization.

Let us mention the following definition. A sequence $(\varphi_n)_{n\geq 1}$ of random convex lsc functions is said to be a *martingale* or an *epi-martingale* if the sequence $(\text{epi}\,\varphi_n)_{n\geq 1}$ is a set-valued martingale (see e.g. [12]).

9 Ergodic theorems for random sets

Consider a probability space $(\Omega, \mathcal{A}, P)$. An $\mathcal{A}-$measurable map $T : \Omega \to \Omega$ is said to be *measure-preserving* if

$$P(T^{-1}(A)) = P(A) \qquad\qquad \forall A \in \mathcal{A}$$

A set $A \in \mathcal{A}$ is *invariant* (with respect to T) if $T^{-1}(A) = A$

The class of invariant sets is a sub$-\sigma-$field of $\mathcal{A}$, denoted by $\mathcal{I}$.

T is called *ergodic* if for every $A \in \mathcal{I}$,

$$P(A) = 0 \ \text{ or } \ P(A) = 1$$

Thus, one has $\mathcal{I} = \{\Omega, \emptyset\}$ (up to $P-$null sets).

The following result is a multivalued analog of Birkhoff's ergodic theorem for random sets with values in $\mathcal{C}_c(X)$.

Theorem 7

If T is a measure-preserving transformation on $(\Omega, \mathcal{A}, P)$ and

$$F : \Omega \to \mathcal{C}_c(X)$$

an integrable random set, the one has

$$E(F/\mathcal{I})(\omega) = PK - \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} F(T^i\omega) \qquad\qquad \text{with probability 1}$$

In particular, if T is ergodic

$$\mathrm{cl} \int_\Omega F\, dP = PK - \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} F(T^i\omega) \qquad\qquad \text{with probability 1}$$

Now, we present a version of Birkhoff's Ergodic Theorem for random lsc function in the ergodic case

Theorem 8

Let $\beta \in L^1$ and assume that T is ergodic. If φ is a random lsc function such that

$$(9.1) \qquad \varphi(\omega, x) \geq \beta(\omega) \qquad x \in X \qquad \omega \in \Omega$$

then, with probability 1

$$E(\varphi)(.) = \text{epi} - \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} \varphi(T^i \omega, .)$$

Remarks

- Due to the local character of epi-convergence, condition (9.1) can be weakened. Indeed, it can be localized on a suitable neighborhood of each point of X.

- Extension to the non ergodic case requires the appropriate definition of the conditional expectation for random variables depending on a parameter (see e.g. [6]).

10 Other topics and concluding remarks

(i) There are several characterizations of equi-distribution and independence of random sets and random functions. For example:

Proposition 10.1 *If F and G are two random sets with values in $\mathcal{C}(X)$, the following statements are equivalent:*

(a) F and G have the same distribution on $(\mathcal{C}(X), \mathcal{E})$

(b) For each open subset U of X

$$P\big(F^{-1}(U^-)\big) = P\big(G^{-1}(U^-)\big)$$

where

$$U^- = \{C \in \mathcal{C}(X) : C \cap U \neq \emptyset\}$$

(c) For any finite subset $Y = \{x_1\,,\,\ldots,\,x_k\}$ of X, the $\mathbf{R}^k-$valued random variables

$$\big(d(x_1\,,\,F)\ldots,d(x_k\,,\,F)\big) \qquad \text{and} \qquad \big(d(x_1\,,\,G)\ldots,d(x_k\,,\,G)\big)$$

have the same distribution.

Characterizations in terms of measurable selections also exist (see e.g. [15]).

(ii) When X is an infinite dimensional Banach space (e.g. a function space), it is necessary to make more precise the definition of the Painlevé-Kuratowski convergence in order to obtain useful results.

A major example of infinite dimensional extension of PK-convergence is the *Mosco convergence* (see [22], [9])

Denote by s the strong (norm) topology of X and by w the weak-topology. Further, denote by $PK(s)$ (resp. $PK(w)$) the Painlevé-Kuratowski convergence with respect to topology s (resp. w).

A sequence (C_n) of subsets of X is said to *Mosco-converge* to C if

$$C = PK(s) - \lim_{n\to+\infty} C_n \qquad \text{and} \qquad C = PK(w) - \lim_{n\to+\infty} C_n$$

Even if X is finite dimensional, the consideration of an infinite dimensional setting may be necessary, e.g. for functional integrals.

Given a random lsc function $\varphi : \Omega \times X \rightarrow \overline{\mathbf{R}}$, the *functional integral* associated with φ is denoted I_φ and defined by

$$I_\varphi(f) = \int_\Omega \varphi(\omega, f(\omega))\, dP$$

where $f : \Omega \rightarrow X$ belongs to some function space Y, so that the above integral makes sense.

Example: $Y = L^p(X)$ $(p \geq 1)$

Here, the weak topology of Y is $w = \sigma(Y, Y^*)$, where $Y^* = L^q(X^*)$ and

$$1/p + 1/q = 1 \qquad\qquad (q = \infty \text{ if } p = 1).$$

Consider the random lsc functions φ_∞ and φ_n $(n \geq 1)$ and the following two statements.

(a) For all (or $P-$almost all) $\omega \in \Omega$, $\varphi_n(\omega, .) \stackrel{epi}{\rightarrow} \varphi_\infty(\omega, .)$ in X

(b) $I_{\varphi_n} \stackrel{Mosco}{\rightarrow} I_{\varphi_\infty}$ in Y

There are results asserting that under appropriate conditions,
implication $(a) \Rightarrow (b)$ holds (see e.g. [7]).

(iii) If the constraints appear explicitely in optimization problems $(\mathcal{P}_n)$ and $(\mathcal{P}_\infty)$

$$(\mathcal{P}_n)\ \begin{cases} \text{minimize} \quad f_n(x) \\ \quad x \in C_n \end{cases}$$

$$(\mathcal{P}_\infty)\ \begin{cases} \text{minimize} \quad f_\infty(x) \\ \quad x \in C_\infty \end{cases}$$

it is necessary to study epi-convergence of the sequence

$$(f_n + \chi_{C_n}) \qquad \text{to} \qquad f_\infty + \chi_{C_\infty}$$

In general, it is not true that if $f_n \stackrel{epi}{\to} f_\infty$ and $g_n \stackrel{epi}{\to} g_\infty$ imply

$$f_n + g_n \stackrel{epi}{\to} f_\infty + g_\infty$$

but it is true in some particular situations (see [9]).

For example, assume that f_n $(n \geq 1)$ and f_∞ are finite everywhere on X. If (f_n) converges uniformly to f_∞ and $g_n = \chi_{C_n} \stackrel{epi}{\to} g = \chi_{C_\infty}$, we have

$$f_n + g_n \stackrel{epi}{\to} f_\infty + g_\infty$$

(iv) In this presentation, we have only considered sets whose values are unbounded, because our objective was the application to epigraphs. Let us mention that there exist an abundant literature and numerous results concerning random sets whose values are bounded. In that framework, set-convergence is often defined by the Hausdorff distance, which is stronger than the Painlevé-Kuratowski convergence.

(v) Set-valued extensions of Fatou's Lemma have been also proved for sequences of random sets. They consist of specific set-inclusions that are useful in several fields, especially for the study of equilibria in Mathematical Economics (see e.g. [3]).

11 References

References

[1] H. Attouch, Variational convergence for functions and operators, Pitman Advanced Pub. Program (1984)

[2] Z. Artstein and S. Hart, Law of Large Numbers for Random Sets and Allocation Processes, Mathematics of Operations Research, Vol. 6, No. 4, pp. 485-492 (1981)

[3] E. Balder and C. Hess, On the unbounded multivalued version of Fatou's Lemma, Mathematics of Operations Research, Vol. 20, No. 1 (1995)

[4] C. Castaing, C. Hess and M. Saadoune, Tightness conditions and integrability of the sequential weak upper limit of a sequence of multifunctions, Advances in Mathematical Economics 11, p. 1-34 (2008)

[5] C. Castaing and M. Valadier, Convex Analysis and Measurable Multifunctions, Lecture Notes in Mathematics, Springer, Berlin, Germany, Vol. 580 (1977)

[6] C. Choirat, C. Hess and R. Seri, A version of the Birkhoff Ergodic Theorem for a normal integrand: a variational approach, Annals of Probability, Vol. 31, No. 1, p. 63-92 (2003)

[7] J. Couvreux and C. Hess, Mosco approximation of integrands and integral functionals, Journal of Optimization Theory and Applications, Vol. 90, No. 2, p. 335-356 (1996)

[8] J. Couvreux and C. Hess, A Lévy type martingale convergence theorem for random sets with unbounded values, Journal of Theoretical Probability, Vol. 12, No. 4, p. 933-969 (1999)

[9] G. Dal Maso, An Introduction to $\Gamma-$convergence, Birkhäuser (1993)

[10] C. Hess, Measurability and Integrability of the weak upper limit of a sequence of multifunctions, Journal of Mathematical Analysis and Applications, Vol. 153, No. 1, p. 226-249 (1990)

[11] C. Hess, Convergence of conditional expectations for unbounded random sets, integrands and integral functionals, Mathematics of Operations Research, vol. 16 No. 3, p. 627-649 (1991)

[12] C. Hess, On multivalued martingales whose values may be unbounded : martingale selectors and Mosco convergence, Journal of Multivariate Analysis, Vol. 39, No. 1, (1991)

[13] C. Hess, Multivalued strong laws of large numbers in the slice topology. Application to integrands., Set Valued Analysis, 2, 183-205 (1994)

[14] C. Hess, Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator, Annals of Statistics Vol. 24, No. 3, 1298-1315 (1996)

[15] C. Hess, From the Distribution of Unbounded Random Sets to the Multivalued Strong Law of Large Numbers in Nonreflexive Banach Spaces, Journal of Convex Analysis, Vol. 6, No. 1, p. 163-182 (1999)

[16] C. Hess, Conditional expectation and martingales of random sets, Journal of Pattern Recognition, 32, p. 1543-1567 (1999)

[17] C. Hess, Set-Valued Integration and Set-Valued Probability Theory : An Overview, in “Handbook of Measure Theory”, North Holland, Elsevier Science B.V. p. 617-673, edited by Prof. Endre Pap (2002)

[18] C. Hess, On the parametrized integral of a multifunction : the unbounded case, Set-Valued Analysis, Vol. 15, p. 1-20 (2007)

[19] F. Hiaï and H. Umegaki, Integrals, Conditional Expectations and Martingales of Multivalued Functions, Journal of Multivariate Analysis, Vol. 7, pp. 149-182 (1977)

[20] L. Modica, Stochastic homogenization and ergodic theory, Lecture Notes in Mathematics, Volume 1190 (1986)

[21] I. Molchanov, Theory of Random Sets, Springer, London (2005)

[22] U. Mosco., On the Continuity of the Young-Fenchel Transform, Journal of Mathematical Analysis and Applications, Vol. 35, pp. 518-535, 1971.

[23] R.T. Rockafellar and R. J.-B. Wets, Variational Analysis, Springer (1997)