published in Idemptocency (J. Gunawardena, ed.) Cambridge Univ. press, 1998

Duality between Probability and Optimization

Marianne Akian, Jean-Pierre Quadrat and Michel Viot

1 Introduction

Following the theory of idempotent measures of Maslov, a formalism analogous to probability calculus is obtained for optimization by replacing the classical structure of real numbers $(\mathbb{R}, +, \times)$ by the idempotent semi-field obtained by endowing the set $\mathbb{R} \cup \{+\infty\}$ with the "min" and "+" operations. To the probability of an event corresponds the cost of a set of decisions. To random variables correspond decision variables.

Weak convergence, tightness and limit theorems of probability have an optimization counterpart which is useful to approximate Hamilton Jacobi Bellman (HJB) equation and to obtain asymptotics for this equation. The introduction of tightness for cost measures and its consequences is the main contribution of this paper. The link between the weak convergence and the epigraph convergence used in convex analysis is done.

The Cramer transform used in the large deviation literature is defined as the composition of the Laplace transform by the logarithm by the Fenchel transform. It transforms convolution into inf-convolution. Probabilistic results about processes with independent increments are then transformed into similar results on dynamic programming equations. Cramer transform gives new insight on the Hopf method used to compute explicit solutions of some HJB equations. It also explains the limit theorems obtained directly as the image of the classic limit theorems of probability.

Bibliographic notes are given at the end of the paper.

2 Cost Measures and Decision Variables

Let us denote by $\mathbb{R}_{\min}$ the idempotent semifield $(\mathbb{R} \cup \{+\infty\}, \min, +)$ and by extension the metric space $\mathbb{R} \cup \{+\infty\}$ endowed with the exponential distance $d(x, y) = |\exp(-x) - \exp(-y)|$. We start by defining cost measures which can be seen as normalized idempotent measures of Maslov in $\mathbb{R}_{\min}$ [24].

Definition 2.1. We call a *decision space* the triplet $(U, \mathcal{U}, \mathbb{K})$ where U is a topological space, $\mathcal{U}$ the set of open sets of U and $\mathbb{K}$ a mapping from $\mathcal{U}$ to $\mathbb{R}_{\min}$ such that

1. $\mathbb{K}(U) = 0$,
2. $\mathbb{K}(\emptyset) = +\infty$,
3. $\mathbb{K}\left(\bigcup_n A_n\right) = \inf_n \mathbb{K}(A_n)$ for any $A_n \in \mathcal{U}$.

The mapping $\mathbb{K}$ is called a *cost measure.*

A function $c : U \to \mathbb{R}_{\min}$ such that $\mathbb{K}(A) = \inf_{u\in A} c(u)\ \forall A \in \mathcal{U}$ is called a *cost density* of the cost measure $\mathbb{K}$.

The set $D_c \stackrel{\text{def}}{=} \{u \in U \mid c(u) \neq +\infty\}$ is called the *domain* of c.

Theorem 2.2. *Given a l.s.c. c with values in $\mathbb{R}_{\min}$ such that $\inf_u c(u) = 0$, the mapping $A \in \mathcal{U} \mapsto \mathbb{K}(A) = \inf_{u\in A} c(u)$ defines a cost measure on $(U, \mathcal{U})$. Conversely any cost measure defined on open sets of a second countable topological space*[1] *admits a unique minimal extension $\mathbb{K}_*$ to $\mathcal{P}(U)$ (the set of subsets of U) having a density c which is a l.s.c. function on U satisfying $\inf_u c(u) = 0$.*

Proof. This precise result is proved in Akian [1]. See also Maslov [24] and Del Moral [15] for the first part and Maslov and Kolokoltsov [23, 25] for the second part. □

Remark 2.3. This theorem shows that on second countable spaces there is a bijection between l.s.c. functions and cost measures. In this paper, we will consider cost measures on $\mathbb{R}^n$, $\mathbb{R}^{\mathbb{N}}$, separable Banach spaces and separable reflexive Banach spaces with the weak topology which are all second countable topological spaces.

Example 2.4. We will use very often the two following cost densities defined on $\mathbb{R}^n$ with $\|.\|$ the euclidian norm.

1. $\chi_m(x) \stackrel{\text{def}}{=} \begin{cases} +\infty & \text{for } x \neq m. \\ 0 & \text{for } x = m, \end{cases}$
2. $\mathcal{M}^p_{m,\sigma}(x) \stackrel{\text{def}}{=} \frac{1}{p}\|\sigma^{-1}(x-m)\|^p$ for $p \geq 1$ with $\mathcal{M}^p_{m,0} \stackrel{\text{def}}{=} \chi_m$.

By analogy with conditional probability we define the conditional cost excess.

Definition 2.5. The *conditional cost excess* to take the best decision in A knowing that it must be taken in B is

$$\mathbb{K}(A|B) \stackrel{\text{def}}{=} \mathbb{K}(A \cap B) - \mathbb{K}(B)\ .$$

[1] i.e. a topological space with a countable basis of open sets.

By analogy with random variables we define decision variables and related notions.

Definition 2.6. 1. A *decision variable* X on $(U, \mathcal{U}, \mathbb{K})$ is a mapping from U to E (a second countable topological space). It induces a cost measure $\mathbb{K}_X$ on $(E, \mathcal{B})$ ($\mathcal{B}$ denotes the set of open sets of E) defined by $\mathbb{K}_X(A) = \mathbb{K}_*(X^{-1}(A))$ for all $A \in \mathcal{B}$. The cost measure $\mathbb{K}_X$ has a l.s.c. density denoted c_X. When $E = \mathbb{R}$, we call X a real decision variable; when $E = \mathbb{R}_{\min}$, we call it a *cost variable.*

2. Two decision variables X and Y are said *independent* when:
$$c_{X,Y}(x, y) = c_X(x) + c_Y(y) \ .$$

3. The *conditional cost excess* of X knowing Y is defined by:
$$c_{X|Y}(x, y) \stackrel{\text{def}}{=} \mathbb{K}_*(X = x \mid Y = y) = c_{X,Y}(x, y) - c_Y(y) \ .$$

4. The *optimum* of a decision variable is defined by
$$\mathbb{O}(X) \stackrel{\text{def}}{=} \arg\min_{x \in E} \operatorname{conv}(c_X)(x)$$
when the minimum exists. Here conv denotes the l.s.c. convex hull and arg min the point where the minimum is reached. When a decision variable X with values in a linear space satisfies $\mathbb{O}(X) = 0$ we say that it is *centered.*

5. When the optimum of a decision variable X with values in $\mathbb{R}^n$ is unique and when near the optimum, we have
$$\operatorname{conv}(c_X)(x) = \frac{1}{p} \|\sigma^{-1}(x - \mathbb{O}(X))\|^p + o(\|x - \mathbb{O}(X)\|^p) \ ,$$
we say that X is of *order* p and we define its *sensitivity of order* p by $\mathbb{S}^p(X) \stackrel{\text{def}}{=} \sigma$. When $\mathbb{S}^p(X) = I$ (the identity matrix) we say that X is of *order* p *and normalized.*

6. The *value* of a cost variable X is $\mathbb{V}(X) \stackrel{\text{def}}{=} \inf_x (x + c_X(x))$, *the conditional value* is $\mathbb{V}(X \mid Y = y) \stackrel{\text{def}}{=} \inf_x (x + c_{X|Y}(x, y))$.

Example 2.7. For a real decision variable X of cost $\mathcal{M}^p_{m,\sigma}$ with $p > 1$ and $1/p + 1/p' = 1$, we have
$$\mathbb{O}(X) = m, \ \mathbb{S}^p(X) = \sigma, \ \mathbb{V}(X) = m - \frac{1}{p'} \sigma^{p'} \ .$$

3 Vector Spaces of Decision Variables

Theorem 3.1. *For $p > 0$, the numbers*

$$|X|_p \stackrel{\text{def}}{=} \inf \left\{ \sigma \mid c_X(x) \geq \frac{1}{p} |(x - \mathbb{O}(X))/\sigma|^p \right\} \textit{and} \|X\|_p \stackrel{\text{def}}{=} |X|_p + |\mathbb{O}(X)|$$

define respectively a seminorm and a norm on the vector space $\mathbb{L}^p$ of classes[2] *of real decision variables having a unique optimum and such that $\|X\|_p$ is finite.*

Proof. Let us denote $X' = X - \mathbb{O}(X)$ and $Y' = Y - \mathbb{O}(Y)$. We first remark that $\sigma > |X|_p$ implies

$$c_X(x) \geq \frac{1}{p}(|x - \mathbb{O}(X)|/\sigma)^p \quad \forall x \in \mathbb{R} \Leftrightarrow \mathbb{V}(-\frac{1}{p}|X'/\sigma|^p) \geq 0 \;. \tag{3.1}$$

If there exists $\sigma > 0$ and $\mathbb{O}(X)$ such that (3.1) holds, then $c_X(x) < 0$ for any $x \neq \mathbb{O}(X)$ and $c_X(x)$ tends to 0 implies x tends to $\mathbb{O}(X)$ therefore $\mathbb{O}(X)$ is the unique optimum of X. Moreover $|X|_p$ is the smallest σ such that (3.1) holds.

If $X \in \mathbb{L}^p$, $\lambda \in \mathbb{R}$ and $\sigma > |X|_p$ we have

$$\mathbb{V}(-\frac{1}{p}|\lambda X'/\lambda\sigma|^p) = \mathbb{V}(-\frac{1}{p}|X'/\sigma|^p) \geq 0 \;,$$

then $\lambda X \in \mathbb{L}^p$, $\mathbb{O}(\lambda X) = \lambda \mathbb{O}(X)$ and $|\lambda X|_p = |\lambda||X|_p$.

If X and $Y \in \mathbb{L}^p$, $\sigma > |X|_p$ and $\sigma' > |Y|_p$,

$$\mathbb{V}(-\frac{1}{p}(\max(|X'/\sigma|^p, |Y'/\sigma'|^p))) = \min(\mathbb{V}(-\frac{1}{p}|X'/\sigma|^p), \mathbb{V}(-\frac{1}{p}|Y'/\sigma'|^p)) \geq 0$$

and

$$\frac{|X' + Y'|}{\sigma + \sigma'} \leq \frac{\sigma}{\sigma + \sigma'}\frac{|X'|}{\sigma} + \frac{\sigma'}{\sigma + \sigma'}\frac{|Y'|}{\sigma'} \leq \max(\frac{|X'|}{\sigma}, \frac{|Y'|}{\sigma'}) \;,$$

then

$$\mathbb{V}(-\frac{1}{p}(|X' + Y'|/(\sigma + \sigma'))^p) \geq 0 \;.$$

Therefore we have proved that $X + Y \in \mathbb{L}^p$ with $\mathbb{O}(X + Y) = \mathbb{O}(X) + \mathbb{O}(Y)$ and $|X + Y|_p \leq |X|_p + |Y|_p$.

Then $\mathbb{L}^p$ is a vector space, $|.|_p$ and $\|.\|_p$ are seminorms and $\mathbb{O}$ is a linear continuous operator from $\mathbb{L}^p$ to $\mathbb{R}$. Moreover, $\|X\|_p = 0$ implies $c_X = \chi$ thus $X = 0$ up to a set of infinite cost. □

[2]for the almost sure equivalence relation: $X \stackrel{\text{a.s.}}{=} Y \;\Leftrightarrow\; \mathbb{K}_*(X \neq Y) = +\infty$.

Theorem 3.2. *For two independent real decision variables X and Y and $k \in \mathbb{R}$ we have (as soon as the right and left hand sides exist)*

$$\mathbb{O}(X+Y) = \mathbb{O}(X) + \mathbb{O}(Y), \quad \mathbb{O}(kX) = k\mathbb{O}(X), \quad \mathbb{S}^p(kX) = |k|\mathbb{S}^p(X) \ ,$$

$$[\mathbb{S}^p(X+Y)]^{p'} = [\mathbb{S}^p(X)]^{p'} + [\mathbb{S}^p(Y)]^{p'}, \quad (|X+Y|_p)^{p'} \le (|X|_p)^{p'} + (|Y|_p)^{p'} \ ,$$

where $1/p + 1/p' = 1$.

Proof. Let us prove only the last inequality. Consider X and Y in $\mathbb{L}^p$ and $\sigma > |X|_p$ and $\sigma' > |Y|_p$. Let us denote $\sigma'' = (\sigma^{p'} + \sigma'^{p'})^{1/p'}$, $X' = X - \mathbb{O}(X)$ and $Y' = Y - \mathbb{O}(Y)$. The Hölder inequality $a\alpha + b\beta \le (a^p + b^p)^{1/p}(\alpha^{p'} + \beta^{p'})^{1/p'}$ implies

$$(|X' + Y'|/\sigma'')^p \le |X'/\sigma|^p + |Y'/\sigma'|^p \ ,$$

then by the independency of X and Y we get

$$\mathbb{V}(-\frac{1}{p}(|X' + Y'|/\sigma'')^p) \ge 0 \ ,$$

and the inequality is proved. □

Theorem 3.3 (Chebyshev)**.** *For a decision variable belonging to* $\mathbb{L}^p$ *we have*

$$\mathbb{K}(|X - \mathbb{O}(X)| \ge a) \ge \frac{1}{p}(a/|X|_p)^p \ ,$$

$$\mathbb{K}(|X| \ge a) \ge \frac{1}{p}((a - \|X\|_p)^+/\|X\|_p)^p \ .$$

Proof. The first inequality is a straightforward consequence of the inequality $c_Y(y) \ge (|y|/|Y|_p)^p/p$ applied to the centered decision variable $Y = X - \mathbb{O}(X)$.

The second inequality comes from the nonincreasing property of the function $x \in \mathbb{R}^+ \mapsto (a-x)^+/x$. □

4 Convergence of Decision Variables and Law of Large Numbers

Definition 4.1. A *sequence of independent and identically costed (i.i.c.) real decision variables of cost c* on $(U, \mathcal{U}, \mathbb{K})$ is an application X from U to $\mathbb{R}^{\mathbb{N}}$ which induces the density cost

$$c_X(x) = \sum_{i=0}^{\infty} c(x_i), \quad \forall x = (x_0, x_1, \ldots) \in \mathbb{R}^{\mathbb{N}} \ .$$

Remark 4.2. The cost density is finite only on minimizing sequences of c, elsewhere it is equal to $+\infty$.

Remark 4.3. We have defined a decision sequence by its density and not by its value on the open sets of $\mathbb{R}^{\mathbb{N}}$ because the density always exists and can be defined easily.

In order to state limit theorems, we define several type of convergence of sequences of decision variables.

Definition 4.4. For the sequence of real decision variables $\{X_n, n \in \mathbb{N}\}$ we say that

1. $X_n \in \mathbb{L}^p$ *converges in p-norm* towards $X \in \mathbb{L}^p$ denoted $X_n \xrightarrow{\mathbb{L}^p} X$, if $\lim_n \|X_n - X\|_p = 0$;

2. X_n *converges in cost* towards X, denoted $X_n \xrightarrow{\mathbb{K}} X$, if for all $\epsilon > 0$ we have $\lim_n \mathbb{K}\{u \mid |X_n(u) - X(u)| \geq \epsilon\} = +\infty$;

3. X_n *converges almost surely* towards X, denoted $X_n \xrightarrow{\text{a.s.}} X$, if we have $\mathbb{K}\{u \mid \lim_n X_n(u) \neq X(u)\} = +\infty$.

Some relations between these different kinds of convergence are given in the following theorem.

Theorem 4.5. *1. Convergence in p-norm implies convergence in cost but the converse is false.*

2. Convergence in cost implies almost sure convergence and the converse is false.

Proof. See Akian [2] for points 1 and 2 and Del Moral [15] for point 2. □

We have the analogue of the law of large numbers.

Theorem 4.6. *Given a sequence $\{X_n,\ n \in \mathbb{N}\}$ of i.i.c. decision variables belonging to $\mathbb{L}^p$, $p \geq 1$, we have*

$$\lim_{N\to\infty} Y_N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=0}^{N-1} X_n = \mathbb{O}(X_0) ,$$

where the limit can be taken in the sense of almost sure, cost and p-norm convergence.

Proof. We have only to estimate the convergence in p-norm. The result follows from simple computation of the p-seminorm of Y_N. Thanks to Theorem 3.2 we have $(|Y_N|_p)^{p'} \leq N(|X_0|_p)^{p'}/N^{p'}$ which tends to 0 as N tends to infinity. □

5 Weak Convergence and Tightness of Decision Variables

In this section we introduce the notions of weak convergence and tightness of cost measures and show the relations between the weak convergence and epigraph convergence of functions introduced in convex analysis [5, 4, 22]. Weak convergence and tightness of decision variables will mean weak convergence of their cost measures.

Definition 5.1. 1. Let $\mathbb{K}_n$ and $\mathbb{K}$ be cost measures on $(U,\mathcal{U})$. We say that $\mathbb{K}_n$ *converges weakly* towards $\mathbb{K}$, denoted $\mathbb{K}_n \xrightarrow{\text{w}} \mathbb{K}$, if for all f in $\mathcal{C}_b(U)$[3] we have $\lim_n \mathbb{K}_n(f) = \mathbb{K}(f)$[4].

2. Let c_n and c be functions from U (a first countable topological space[5]) to $\mathbb{R}_{\min}$, we say that c_n converges in the epigraph sense (*epi-converges*) towards c, denoted $c_n \xrightarrow{\text{epi}} c$ if

$$\forall u, \quad \forall u_n \to u, \quad \liminf_n \ c_n(u_n) \geq c(u) \ , \tag{5.1}$$

$$\forall u, \quad \exists u_n \to u : \limsup_n \ c_n(u_n) \leq c(u) \ . \tag{5.2}$$

3. If U is a reflexive Banach space, we say that c_n *Mosco-epi-converges* towards c, denoted $c_n \xrightarrow{\text{M-epi}} c$, if the convergence of u_n holds for the weak topology in (5.1) and for the strong topology in (5.2).

Theorem 5.2. *Let $\mathbb{K}_n$, $\mathbb{K}$ be cost measures on a metric space U. Then the three following conditions are equivalent*

1. $\mathbb{K}_n \xrightarrow{\text{w}} \mathbb{K}$ *;*

2.

$$\liminf_n \mathbb{K}_n(F) \geq \mathbb{K}(F) \quad \forall F \, closed \, , \tag{5.3}$$

$$\limsup_n \mathbb{K}_n(G) \leq \mathbb{K}(G) \quad \forall G \, open \, ; \tag{5.4}$$

3. $\lim_n \mathbb{K}_n(A) = \mathbb{K}(A)$ *for any set A such that* $\mathbb{K}(\overset{\circ}{A}) = \mathbb{K}(\bar{A})$.

Proof. The proof is similar to those of classical probability theory. The mains ingredients in both theories are : 1) U is normal, 2) a probability on the Borel sets of U or a cost measure on the open sets of U is "regular", 3) any bounded continuous function from U to $\mathbb{R}$ or $\mathbb{R}_{\min}$ may be approximated

[3] $\mathcal{C}_b(U)$ denotes the set of continuous and lower bounded functions from U to $\mathbb{R}_{\min}$.

[4] $\mathbb{K}(f) \stackrel{\text{def}}{=} \inf_u(f(u) + c(u))$ where c is the density of $\mathbb{K}$.

[5] Each point admits a countable basis of neighborhoods.

above and below by a $\mathbb{R}$ or $\mathbb{R}_{\min}$-linear combination of characteristic functions of "measurable" sets. Properties 1) and 2) are used in showing 1. $\Rightarrow$ 2. and the equivalence 2. $\Leftrightarrow$ 3. and property 3) in showing 2. $\Rightarrow$ 1.. The main difficulty in optimization theory compared to classical probability is that cost measures are not continuous for the nonincreasing convergence of sets.

Let us precise properties 1–3) in the case of optimization theory. Firstly, since U is a metric space, U is normal in the classical sense. Equivalently, by using the bi-continuous application $t \mapsto -\log(t)$ from $[0,1]$ to the subset $[0,+\infty]$ of $\mathbb{R}_{\min}$, U is normal with respect to $\mathbb{R}_{\min}$ (see Maslov [24] for this notion), that is for any open set G and closed set F such that $F \subset G$, there exists a continuous function f from U to $\mathbb{R}_{\min}$ such that $f \geq 0$, $f = 0$ on F and $f = +\infty$ on G^c and then $\chi_G \leq f \leq \chi_F$. A typical function f is:

$$f(u) = -\log\left(\frac{d(u,G^c)}{d(u,F)+d(u,G^c)}\right) .$$

Secondely, the regularity property of classical probabilities may be translated here in the two following conditions :

$$\mathbb{K}(F) = \sup_{G \supset F,\, G \in \mathcal{U}} \mathbb{K}(G) \quad \forall F \text{ closed} \tag{5.5}$$

and

$$\mathbb{K}(G) = \inf_{F \subset G,\, F \text{ closed}} \mathbb{K}(F) \quad \forall G \in \mathcal{U} . \tag{5.6}$$

The first one is a consequence of the definition of the minimal extension, the second of the fact that in a metric space, any open set is a countable union of closed sets and of the continuity of cost measures for the nondecreasing convergence of sets. Let us note that in classical probability conditions (5.5) and (5.6) are equivalent which is not the case here.

Finally, any lower bounded continuous function may be approximated by simple functions : above by a $\mathbb{R}_{\min}$-linear combination of characteristic functions of open sets, below by an $\mathbb{R}_{\min}$-linear combination of characteristic functions of closed sets. The first approximation follows easily from upper semi-continuity of continuous functions. The second one uses the relative compactness of lower bounded sets in $\mathbb{R}_{\min}$. □

Definition 5.3. A set of cost measures $\mathcal{K}$ is said *tight* if

$$\sup_{C \text{ compact } \subset U} \inf_{\mathbb{K} \in \mathcal{K}} \mathbb{K}(C^c) = +\infty .$$

A sequence $\mathbb{K}_n$ of cost measures is said *asymptotically tight* if

$$\sup_{C \text{ compact } \subset U} \liminf_n \mathbb{K}_n(C^c) = +\infty .$$

Theorem 5.4. *On asymptotically tight sequences $\mathbb{K}_n$ over a metric space U, the weak convergence of $\mathbb{K}_n$ towards $\mathbb{K}$ is equivalent to (5.4) and*

$$\liminf_n \mathbb{K}_n(C) \geq \mathbb{K}(C) \quad \forall C\, compact\,. \tag{5.7}$$

Remark 5.5. In a locally compact space conditions (5.7) (5.4) are equivalent to the condition $\lim_n \mathbb{K}_n(f) = \mathbb{K}(f)$ for any continuous function with compact support. This is the definition of weak convergence used by Maslov and Samborski in [27]. These conditions are also equivalent to the epigraph convergence of densities (see Theorem 5.7 below). This type of convergence does not insure that a weak-limit of cost measures is a cost measure (the infimum of the limit is not necessarily equal to zero).

Theorem 5.6. *Let us denote by $\mathcal{K}(U)$ the set of cost measures on U (a metric space) endowed with the topology of the weak convergence. Any tight set K of $\mathcal{K}(U)$ is relatively sequentially compact*[6].

Proof. It is sufficient to prove that from any asymptotically tight sequence $\{\mathbb{K}_n\}$, we can extract a weakly convergent subsequence.

Let C_k be a compact set such that: $\liminf_n \mathbb{K}_n(C_k^c) \geq k$ and $V = \cup_k C_k$, then $\liminf_n \mathbb{K}_n(V^c) = +\infty$. The convergence of $\mathbb{K}_n$ is then equivalent to the convergence of $\mathbb{K}_n$ on V, which is a separable metric space. Since $\mathbb{K}_n$ is still asymptotically tight on V, we suppose now $U = V$.

Let $\mathcal{B}$ be a countable basis of open sets of U. Since $\mathbb{K}_n$ takes its values in $[0, +\infty]$ (which is a compact set of $\mathbb{R}_{\min}$) and $\mathcal{B}$ is countable, we may extract a subsequence of $\mathbb{K}_n$, denoted also $\mathbb{K}_n$, such that $\lim_n \mathbb{K}_n(B) = \widetilde{\mathbb{K}}(B)\ \forall B \in \mathcal{B}$.

Since any open set is a countable union of elements of $\mathcal{B}$, we define $\mathbb{K}$ on $\mathcal{U}$ by:

$$\mathbb{K}(A) = \sup_{\mathcal{A}} \inf_{B \in \mathcal{A}} \widetilde{\mathbb{K}}(B)\,,$$

where the supremum is taken over subsets $\mathcal{A}$ of $\mathcal{B}$ such that $\cup_{B \in \mathcal{A}} B = A$. $\mathbb{K}$ is the minimal cost measure on $\mathcal{U}$ greater than $\widetilde{\mathbb{K}}$ on $\mathcal{B}$.

Its minimal extension to $\mathcal{P}(U)$ is

$$\mathbb{K}(A) = \sup_{\mathcal{A}} \inf_{B \in \mathcal{A}} \widetilde{\mathbb{K}}(B)\,.$$

where this times $\mathcal{A}$ satisfies $\cup_{B \in \mathcal{A}} B \supset A$.

Let us show that $\mathbb{K}_n \xrightarrow{\text{w}} \mathbb{K}$. By Theorem 5.4 it is enough to prove (5.4) and (5.7). If G is an open set, then for any $B \in \mathcal{B}$ such that $B \subset G$, we have

$$\limsup_n \mathbb{K}_n(G) \leq \limsup_n \mathbb{K}_n(B) = \widetilde{\mathbb{K}}(B)\,.$$

[6]that is any sequence of K contains a weakly convergent subsequence

Therefore if $G = \cup_{B\in\mathcal{A}} B$ with $\mathcal{A} \subset \mathcal{B}$, we have

$$\limsup_n \mathbb{K}_n(G) \leq \inf_{B\in\mathcal{A}} \widetilde{\mathbb{K}}(B) \leq \mathbb{K}(G) \ .$$

If F is a compact set, and if $\cup_{B\in\mathcal{A}} B \supset F$, we may restrict $\mathcal{A}$ to be finite, then we have

$$\liminf_n \mathbb{K}_n(F) \geq \liminf_n \inf_{B\in\mathcal{A}} \mathbb{K}_n(B) = \inf_{B\in\mathcal{A}} \liminf_n \mathbb{K}_n(B) = \inf_{B\in\mathcal{A}} \widetilde{\mathbb{K}}(B) \ .$$

By taking the supremum over all sets $\mathcal{A}$, we obtain condition (5.7). □

Theorem 5.7. *On a first countable topological space, the epi-convergence of l.s.c. densities c_n of $\mathbb{K}_n$ towards the density c of $\mathbb{K}$ is equivalent to conditions (5.4) and (5.7).*

Proof. We prove (5.1)⇔ (5.7) and (5.2)⇔ (5.4).

1. (5.1)⟹ (5.7).

   Let u_n be a point where c_n reaches its optimum in compact set C. For any converging subsequence $\{u_{n_k}\}$, the limit u belongs to C and we have $\liminf_k c_{n_k}(u_{n_k}) \geq c(u) \geq \mathbb{K}(C)$, therefore $\liminf_n \mathbb{K}_n(C) \geq \mathbb{K}(C)$.

2. (5.7)⟹ (5.1).

   The sets $C_N = \{u_n, n \geq N\} \cup \{u\}$ are compact and we have

   $$\liminf_n c_n(u_n) \geq \liminf_n \mathbb{K}_n(C_N) \geq \mathbb{K}(C_N) \ .$$

   As the l.s.c of c implies $\sup_N \mathbb{K}(C_N) \geq c(u)$, the result follows.

3. (5.2)⟹ (5.4).

   Let us prove this assertion by contraposition. Let us suppose there exists an open set G and $\epsilon > 0$ such that $\limsup_n \mathbb{K}_n(G) > \mathbb{K}(G)+\epsilon$. By definition of the infimum there exists $u \in G$ such that $c(u) \leq \mathbb{K}(G) + \epsilon$. Therefore for any sequence u_n converging towards u, $u_n \in G$ for n big enough and we have $\limsup_n c_n(u_n) \geq \limsup_n \mathbb{K}_n(G) > \mathbb{K}(G) + \epsilon \geq c(u)$ which contradicts the hypothesis.

4. (5.4)⟹ (5.2).

   For all u there exists a decreasing family of open sets $\{G_k, k \in \mathbb{N}\}$ such that $\cap_k G_k = \{u\}$. By definition of the infimum, there exists u_n^k such that $\mathbb{K}_n(G_k) \geq c_n(u_n^k) - 1/n$. Then we have $\limsup_n c_n(u_n^k) \leq \limsup_n \mathbb{K}_n(G_k) \leq \mathbb{K}(G_k) \leq c(u)$. By diagonal extraction we obtain a sequence $u_n^{k(n)}$ which satisfies (5.2). □

Remark 5.8. In Attouch [4] another definition of epigraph convergence is given in a general topological space and is mostly related to conditions (5.4) and (5.7).

Proposition 5.9. *If $\mathbb{K}_n$ and $\mathbb{K}'_n$ are (asymptotically) tight sequences of cost measures on U and U' and $\mathbb{K}_n \xrightarrow{\mathrm{w}} \mathbb{K}$ and $\mathbb{K}'_n \xrightarrow{\mathrm{w}} \mathbb{K}'$ then $\mathbb{K}_n \times \mathbb{K}'_n$ is (asymptotically) tight and $\mathbb{K}_n \times \mathbb{K}'_n \xrightarrow{\mathrm{w}} \mathbb{K} \times \mathbb{K}'$.*

Proof. The product of two measures $\mathbb{K}$ and $\mathbb{K}'$ is defined as in probability, then if $\mathbb{K}$ and $\mathbb{K}'$ have densities $c(u)$ and $c'(u')$, $\mathbb{K}\times\mathbb{K}'$ has density $c(u)+c'(u')$. In probability theory, tightness is not necessary, but the technique of proof does not work here. We need to impose the tightness condition, but in this case weak convergence is equivalent to epigraph convergence, for which the result is clear. □

Theorem 5.10. *If $X_n \xrightarrow{\mathbb{K}} X$ and X is tight then $X_n \xrightarrow{\mathrm{w}} X$. More generally if $X_n \xrightarrow{\mathrm{w}} X$, $X_n - Y_n \xrightarrow{\mathbb{K}} 0$ and X is tight, then $Y_n \xrightarrow{\mathrm{w}} X$.*

Proof. see [2]. □

6 Characteristic Functions

The role of the Laplace or Fourier transforms in probability calculus is played by the Fenchel transform in decision calculus.

Definition 6.1.
1. Let $c \in \mathcal{C}_{\mathrm{x}}$, where $\mathcal{C}_{\mathrm{x}}$ denotes the set of l.s.c. and proper[7] convex functions from E (a reflexive Banach space with dual E') to $\mathbb{R}_{\min}$. Its *Fenchel transform* is the function from E' to $\mathbb{R}_{\min}$ defined by $\hat{c}(\theta) \stackrel{\mathrm{def}}{=} [\mathcal{F}(c)](\theta) \stackrel{\mathrm{def}}{=} \sup_x[\langle\theta, x\rangle - c(x)]$.
2. The *characteristic function* of a decision variable is $\mathbb{F}(X) \stackrel{\mathrm{def}}{=} \mathcal{F}(c_X)$.
3. Given two functions f and g from E to $\mathbb{R}_{\min}$, the *inf-convolution of f and g*, denoted $f \,\square\, g$, is the function $z \in E \mapsto \inf_{x,y}\,[f(x)+g(y) \mid x+y = z]$.

Theorem 6.2. *1. For $f, g \in \mathcal{C}_{\mathrm{x}}$ we have*

(a) $\mathcal{F}(f) \in \mathcal{C}_{\mathrm{x}}$,

(b) *$\mathcal{F}$ is an involution that is* $\mathcal{F}(\mathcal{F}(f)) = f$,

(c) $\mathcal{F}(f \,\square\, g) = \mathcal{F}(f) + \mathcal{F}(g)$,

(d) $\mathcal{F}(f + g) = \mathcal{F}(f) \,\square\, \mathcal{F}(g)$.

[7] not always equal to $+\infty$

2. *For two independent decision variables X and Y and $k \in \mathbb{R}$, we have*

$$c_{X+Y} = c_X \,\square\, c_Y, \quad \mathbb{F}(X+Y) = \mathbb{F}(X) + \mathbb{F}(Y), \quad [\mathbb{F}(kX)](\theta) = [\mathbb{F}(X)](k\theta) ,$$

3. *A decision variable with values in $\mathbb{R}^n$ is of order p if we have:*

$$\mathbb{F}(X)(\theta) = \langle \mathbb{O}(X), \theta \rangle + \frac{1}{p'} \|\mathbb{S}^p(X)\theta\|^{p'} + o(\|\theta\|^{p'}) ,$$

with $1/p + 1/p' = 1$.

Remark 6.3. The Fenchel transform (for l.s.c proper convex functions) is bi-continuous for the Mosco-epi-convergence [22].

Theorem 6.4. *Let $\mathbb{K}_n$ and $\mathbb{K}$ be cost measures on a separable reflexive Banach space with (proper) l.s.c convex densities c_n and c, then $c_n \overset{\text{M-epi}}{\longrightarrow} c$ iff the two conditions (5.4) and*

$$\liminf_n \mathbb{K}_n(C) \geq \mathbb{K}(C) \quad \forall C \textit{bounded closed and convex} \tag{6.1}$$

hold.

Proof. In the proof of Theorem 5.7 we see easily that we can replace compact sets by bounded closed convex sets which are weakly compact on a reflexive Banach space and let open sets be those of the strong topology. □

Corollary 6.5. *For an asymptotically tight sequence X_n of decision variables with l.s.c. convex cost densities on a separable reflexive Banach space, X_n converges weakly towards X iff $\mathbb{F}(X_n)$ Mosco-epi-converges towards $\mathbb{F}(X)$.*

Proof. By the tightness property and previous result, the weak convergence of X_n towards X is equivalent the Mosco-epi-convergence of X_n towards X and then to the Mosco-epi-convergence of $\mathbb{F}(X_n)$ towards $\mathbb{F}(X)$. □

This may be used for proving the central limit theorem in a Banach space. For simplicity let us state it in finite dimensional situation where epigraph and Mosco-epigraph convergences are equivalent.

Theorem 6.6 (Central limit theorem)**.** *Let $\{X_n,\ n \in \mathbb{N}\}$ be an i.i.c. sequence centered of order p with l.s.c. convex cost density and $1/p + 1/p' = 1$, we have*

$$Z_N \overset{\text{def}}{=} \frac{1}{N^{1/p'}} \sum_{n=0}^{N-1} X_n \overset{\text{w}}{\longrightarrow} \mathcal{M}^p_{0,\mathbb{S}^p(X_0)} .$$

Proof. We have $\lim_N [\mathbb{F}(Z_N)](\theta) = \frac{1}{p'} \|\mathbb{S}^p(X_0)\theta\|^{p'}$, where the convergence can be taken in the pointwise, uniform on any bounded set or epigraph sense. In order to obtain the weak convergence we have to prove the tightness of Z_N. But as the convergence is uniform on $B = \{\|\theta\| \leq 1\}$ we have for $N \geq N_0$, $\mathbb{F}(Z_N) \leq C$ on B where C is a constant. Therefore $c_{Z_N}(x) \geq \|x\| - C$ for $N \geq N_0$ and Z_N is asymptotically tight. □

The central limit theorem may be generalized to the case of non convex cost densities. This generalization essentially uses the strict convexity of the limiting cost density and was suggested by the Gärtner-Ellis theorem on large deviations of dependent random variables [18, 21]. Indeed, the large deviation principle for probabilities $\mathbb{P}_n$ with entropy I may be considered as the weak convergence of "measures" $\mathbb{K}_n = -h(n)\log \mathbb{P}_n$ (with $\lim_n h(n) = 0$) towards the cost measure with density I. We first need the following result which is proved in [2].

Proposition 6.7. *If $X_n \xrightarrow{\mathrm{w}} X$ in $\mathbb{R}^p$ and $(\mathbb{F}(X_n)(\theta))_n$ is upper bounded for any $\theta \in \mathcal{O}$ where $\mathcal{O}$ is an open convex neighborhood of 0 in $\mathbb{R}^p$, then*

$$\mathbb{F}(X_n)(\theta) \underset{n\to+\infty}{\longrightarrow} \mathbb{F}(X)(\theta) \quad \forall \theta \in \mathcal{O} .$$

In general, a l.s.c. function c on $\mathbb{R}^p$ is not characterized by its Fenchel transform, but when the Fenchel transform is essentially smooth, the convex hull of c is essentially strictly convex, thus c is necessarily convex (see Rockafellar [30] for definitions). A generalization of this remark leads to a result equivalent to Gärtner-Ellis theorem.

Proposition 6.8. *If X_n is a sequence of decision variables with values in $\mathbb{R}^p$ such that*

$$\mathbb{F}(X_n)(\theta) \underset{n\to+\infty}{\longrightarrow} \varphi(\theta) \quad \forall \theta \in \mathbb{R}^p ,$$

where φ is an essentially smooth proper l.s.c. convex function such that $0 \in \overset{\circ}{D}_\varphi$. Then, $X_n \xrightarrow{\mathrm{w}} \mathcal{F}(\varphi)$.

The particular case, where $\varphi(\theta) = \frac{1}{p'}\|\sigma\theta\|^{p'}$ which has a strictly convex Fenchel transform leads to the general central limit theorem.

7 Bellman Chains and Processes

We can generalize i.i.c. sequences to the analogue of Markov chains that we call Bellman chains.

Definition 7.1. A finite valued Bellman chain (E, C, ϕ) with

1. E a finite set of $|E|$ elements called the state space,
2. $C : E \times E \mapsto \mathbb{R}_{\min}$ satisfying $\inf_y C_{xy} = 0$ called the transition cost,
3. ϕ a cost measure on E called the initial cost,

is a decision sequence $X = \{X_n,\ n \in \mathbb{N}\}$ taking its values in $E^{\mathbb{N}}$, such that

$$c_X(x \overset{\text{def}}{=} (x_0, x_1, \ldots)) = \phi_{x_0} + \sum_{i=0}^{\infty} C_{x_i x_{i+1}} ,\ \forall x \in E^{\mathbb{N}} .$$

Theorem 7.2. *For any function f from E to $\mathbb{R}_{\min}$, a Bellman chain satisfies the Markov property* $\mathbb{V}\{f(X_n) \mid X_0, \ldots, X_{n-1}\} = \mathbb{V}\{f(X_n) \mid X_{n-1}\}$.

The analogue of the forward Kolmogorov equation giving a way to compute recursively the marginal probability to be in a state at a given time is the following Bellman equation.

Theorem 7.3. *The marginal cost* $v^n_x = \mathbb{K}(X_n = x)$ *of a Bellman chain is given by the recursive forward equation:* $v^{n+1} = v^n \otimes C \stackrel{\text{def}}{=} \min_{x\in E}(v^n_x + C_{x.})$ *with* $v^0 = \phi$.

Remark 7.4. The cost measure of a Bellman chain is normalized which means that its infimum on all the trajectories is 0. In some applications we would like to avoid this restriction. This can be done by introducing the analogue of the multiplicative functionals of the trajectories of a stochastic process.

We can easily define continuous time decision processes which correspond to deterministic controlled processes. We discuss here only decision processes with continuous trajectories.

Definition 7.5.

1. A *continuous time Bellman process* X_t with continuous trajectories is a decision variable with values in $\mathcal{C}(\mathbb{R}^+)$[8] having the cost density
$$c_X(x(\cdot)) \stackrel{\text{def}}{=} \phi(x(0)) + \int_0^\infty c(t, x(t), x'(t))dt ,$$
with $c(t,\cdot,\cdot)$ a family of transition costs (that is a function c from $\mathbb{R}^3$ to $\mathbb{R}_{\min}$ such that $\inf_y c(t,x,y) = 0,\ \forall t, x)$ and ϕ a cost density on $\mathbb{R}$. When the integral is not defined the cost is by definition equal to $+\infty$.
2. The Bellman process is said *homogeneous* if c does not depend on time t.
3. The Bellman process is said *with independent increments* if c does not depend on state x. Moreover if this process is homogeneous, c is reduced to the cost density of a decision variable.
4. The *p-Brownian decision process*, denoted by B^p_t, is the process with independent increments and transition cost density $c(t,x,y) = \frac{1}{p}|y|^p$.

As in the discrete time case, the marginal cost to be in state x at time t can be computed recursively using a forward Bellman equation.

[8] $\mathcal{C}(\mathbb{R}^+)$ denotes the set of continuous functions from $\mathbb{R}^+$ to $\mathbb{R}$.

Theorem 7.6. *The marginal cost* $v(t,x) \stackrel{\text{def}}{=} \mathbb{K}(X_t = x)$ *is given by the Bellman equation:*

$$\partial_t v + \hat{c}(\partial_x v) = 0, \quad v(0,x) = \phi(x) \;, \tag{7.1}$$

where $\hat{c}$ *means here* $[\hat{c}(\partial_x v)](t,x) \stackrel{\text{def}}{=} \sup_y [y\partial_x v(t,x) - c(t,x,y)]$.

Let $p > 1$ *and* $1/p + 1/p' = 1$. *For the Brownian decision process* B_t^p *starting from* 0, *the marginal cost to be in state* x *at time* t *satisfies the Bellman equation*

$$\partial_t v + (1/p')|\partial_x v|^{p'} = 0, \quad v(0,\cdot) = \chi \;.$$

Its solution can be computed explicitly, it is $v(t,x) = \mathcal{M}^p_{0,t^{1/p'}}(x)$, *therefore*

$$\mathbb{V}[f(B_t^p)] = \inf_x \left[f(x) + \frac{x^p}{p t^{\frac{p}{p'}}} \right] \;. \tag{7.2}$$

8 Tightness in $\mathcal{C}([0,1])$ and Brownian Approximation

Theorem 8.1. *A sequence of decision variables* $\{X_n, n \in \mathbb{N}\}$ *with values in* $\mathcal{C}([0,1])$ *is tight if* $X_n(t) \in \mathbb{L}^p$ *for* $t \in [0,1]$, $\|X_n(0)\|_p$ *is bounded and*

$$\lim_{\delta \to 0^+} \sup_{t \in [0,1-\delta], n \in \mathbb{N}} \|X_n(t+\delta) - X_n(t)\|_p = 0 \;. \tag{8.1}$$

Proof. By Ascoli theorem, we know that relatively compact subsets of $\mathcal{C}([0,1])$ coincide with equi-continuous subsets taking bounded values in 0 . Therefore, we can deduce a necessary and sufficient condition of tightness for a sequence of decision variables $\{X_n\}$ in $\mathcal{C}([0,1])$. The sequence $\{X_n\}$ is tight iff *i)* $X_n(0)$ is tight, that is for all η there exists a such that $\mathbb{K}(|X_n(0)| \geq a) \geq \eta$ and *ii)* for all $\eta, \epsilon > 0$ there exists $\delta > 0$ such that

$$\inf_n \inf_{\substack{t,s \in [0,1] \\ |s-t| \leq \delta}} \mathbb{K}(|X_n(t) - X_n(s)| \geq \epsilon) \geq \eta \;.$$

Then, condition *i)* is a direct consequence of the fact that $\|X_n(0)\|_p$ is bounded and of the Chebyshev inequality applied to $X_n(0)$ and condition *ii)* is a direct consequence of (8.1) and of the Chebyshev inequality applied to $X_n(t) - X_n(s)$. □

The following result shows that weak convergence in $\mathcal{C}([0,1])$ may be characterized by the convergence of finite dimensional marginal costs.

Theorem 8.2. *1. There may exists different cost measures* $\mathbb{K}$ *and* $\mathbb{K}'$ *on* $\mathcal{C}([0,1])$ *such that*

$$\mathbb{K}_\pi = \mathbb{K}'_\pi \quad \forall \pi : \mathcal{C}([0,1]) \to \mathbb{R}^k, x \mapsto (x(t_1), \ldots, x(t_k)). \tag{8.2}$$

2. *If $\mathbb{K}$ is tight and $\mathbb{K}$ and $\mathbb{K}'$ satisfy (8.2), then $\mathbb{K} = \mathbb{K}'$.*

3. *If the sequence $\mathbb{K}_n$ is asymptotically tight and if $(\mathbb{K}_n)_\pi \xrightarrow{\mathrm{w}} \mathbb{K}_\pi$ for all $\pi : \mathcal{C}([0,1]) \to \mathbb{R}^k$, $x \mapsto (x(t_1), \ldots, x(t_k))$, then $\mathbb{K}_n \xrightarrow{\mathrm{w}} \mathbb{K}$.*

Proof. Condition (8.2) is equivalent to $\mathbb{K}(U) = \mathbb{K}'(U)$ for any open set U of the form $U = \{x,\ x(t_1) \in U_1, \ldots, x(t_k) \in U_k\}$ where the U_i are open subsets of $\mathbb{R}$, that is for any open set of the pointwise convergence topology. Since any ball of $\mathcal{C}([0,1])$ is a nonincreasing limit of such open sets, we may have conclude $\mathbb{K} = \mathbb{K}'$ if cost measures were continuous for the nonincreasing convergence of sets as in classical probability. This is not the case in general, but it remains true for a sequence of closed sets if $\mathbb{K}$ is tight.

1. Let us prove the second assertion. Let $\overline{B}(x, \varepsilon)$ denotes the closed ball of center x and radius ε for the uniform convergence norm. There exists open sets U_n for the pointwise convergence topology such that $\overline{B}(x, \varepsilon) = \cap_n \overline{U_n} = \cap_n U_n$. Then, if $\mathbb{K}$ is tight

$$\mathbb{K}(\overline{B}(x, \varepsilon)) = \sup_n \mathbb{K}(\overline{U_n}) \leq \sup_n \mathbb{K}(U_n)$$

$$= \sup_n \mathbb{K}'(U_n) \leq \mathbb{K}'(\cap_n U_n) = \mathbb{K}'(\overline{B}(x, \varepsilon)) .$$

As any open set of $\mathcal{C}([0,1])$ is a countable union of closed balls, we obtain $\mathbb{K}(U) \leq \mathbb{K}'(U)$. Then, the tightness of $\mathbb{K}$ implies the tightness of $\mathbb{K}'$ which implies the converse inequality.

2. For the first assertion, it is sufficient to exhibit a cost measure $\mathbb{K}$ such that $\mathbb{K}(G) \neq 0$ for some open set G and $\mathbb{K}(G) = 0$ for any open set G of the pointwise convergence topology. Since $\mathbb{K}$ has necessarily a density c, $\mathbb{K}(G) = \inf_{x \in G} c(x) = 0$ for any open set G of the pointwise convergence topology. This means that the l.s.c. envelope of c for this topology is equal to 0, whereas those for the uniform convergence topology is non equal to 0. The function $c(x) = \exp(-\|x\|_\infty)$ satisfies this property.

3. As $\mathbb{K}_n$ is asymptotically tight, there exists a weakly converging subsequence that we denote also $\mathbb{K}_n$. Let $\mathbb{K}'$ be the limit. By the tightness of $\mathbb{K}_n$, $\mathbb{K}'$ is also tight and from $\mathbb{K}_n \xrightarrow{\mathrm{w}} \mathbb{K}'$ we have $(\mathbb{K}_n)_\pi \xrightarrow{\mathrm{w}} \mathbb{K}'_\pi$ and therefore $\mathbb{K}_\pi = \mathbb{K}'_\pi$ for any finite dimensional projection π. By the previous result and the tightness of $\mathbb{K}'$ we obtain $\mathbb{K} = \mathbb{K}'$. From the unicity of the limit we obtain $\mathbb{K}_n \xrightarrow{\mathrm{w}} \mathbb{K}$. □

The next result is the analogue of Donsker's theorem about time discretization of Brownian motion.

Theorem 8.3. *Given an i.i.c. sequence X_n of real decision variables centered with sensibilities of order p equal to $\mathbb{S}^p(X_1) = \sigma$, let $S_i = X_1 + \cdots X_i$ be the* *partial sums and Z_n be the decision variable with values in $\mathcal{C}([0,1])$ defined by :*

$$Z_n(t) = \frac{1}{\sigma n^{1/p'}}(S_{[nt]} + (nt - [nt])X_{[nt+1]}) ,$$

with $1/p + 1/p' = 1$. Suppose in addition that $X_1 \in \mathbb{L}^p$. Then, Z_n weakly converges towards the p-Brownian decision process :

$$Z_n \stackrel{\mathrm{w}}{\longrightarrow} B^p .$$

Proof. From Theorem 8.2, we only have to prove tightness of Z_n on the one hand and convergence of finite dimensional distributions of Z_n towards those of B^p on the other hand. For second point, we follow the same technique as in Billingsley's proof of the probabilistic version [13], whereas tightness is proved by using the sufficient conditions of Theorem 8.1.

The tightness of $Z_n(0)$ is obvious since $Z_n(0) \equiv 0$. Using Theorem 3.2 we obtain for any $s, t \in [0,1]$

$$\|Z_n(t) - Z_n(s)\|_p \leq (t-s)^{1/p'} \|X_1\|_p / \sigma ,$$

then $\sup_{n,t} \|Z_n(t+\delta) - Z_n(t)\|_p$ tends to 0 when δ tends to 0.

Let us prove now that the finite dimensional distributions of Z_n converge towards those of B^p, that is $\pi(Z_n) \stackrel{\mathrm{w}}{\longrightarrow} \pi(B^p)$ for any function of the form $\pi : x \mapsto (x(t_1), \dots x(t_k))$.

We first prove $Z_n(t) \stackrel{\mathrm{w}}{\longrightarrow} B^p(t)$ for $t > 0$ (it is clear for $t = 0$). By Theorem 7.6, $B^p(t)$ has cost density $\mathcal{M}^p_{0,t^{1/p'}}$. However, by central limit theorem, we have $S_{[nt]}/[nt]^{1/p'} \stackrel{\mathrm{w}}{\longrightarrow} \mathcal{M}^p_{0,\sigma}$, then

$$Y_n \stackrel{\mathrm{def}}{=} \frac{t^{1/p'} S_{[nt]}}{\sigma [nt]^{1/p'}} \stackrel{\mathrm{w}}{\longrightarrow} B^p(t) .$$

Since

$$Z_n(t) = (\frac{[nt]}{nt})^{1/p'} Y_n + (\frac{nt - [nt]}{\sigma n^{1/p'}}) X_{[nt]+1} ,$$

$\lim_n \|Z_n(t) - Y_n\|_p = 0$ and by Chebyshev inequality $Z_n(t) - Y_n \stackrel{\mathbb{K}}{\longrightarrow} 0$, then the convergence of $Z_n(t)$ towards $B^p(t)$ follows from Theorem 5.10.

Let us prove now $\pi(Z_n) \stackrel{\mathrm{w}}{\longrightarrow} \pi(B^p)$ for any function π. By using a bicontinuous transformation, we may replace π by $x \mapsto (x(t_1),\ x(t_2) - x(t_1), \dots x(t_k) - x(t_{k-1}))$ that we also denote π. Now, by the same type of approximation as before, we may replace $Z_n(t_i) - Z_n(t_{i-1})$ (with $i = 1, \dots, k$ and $t_0 = 0$) by

$$Y_{n,i} = \frac{(t_i - t_{i-1})^{1/p'} (S_{[nt_i]} - S_{[nt_{i-1}]})}{\sigma([nt_i] - [nt_{i-1}])^{1/p'}} .$$

The decision variables $Y_{n,i}$ with $i = 1, \dots, k$ are independent for any n and separately tend to $B^p(t_i) - B^p(t_{i-1})$ which are also independent. Since the sequences $Y_{n,i}$ are tight for any i, the convergence of $\pi(Z_n)$ follows from Proposition 5.9. □

See Maslov [24] for a weaker result when $p = 2$. See Dudnikov and Samborski [17] for an analogous result when the state space is also discretized.

9 Inf-Convolution and Cramer Transform

Definition 9.1. The Cramer transform $\mathcal{C}$ is a function from $\mathcal{M}$, the set of positive measures on $E = \mathbb{R}^n$, to $\mathcal{C}_{\mathrm{x}}$ defined by $\mathcal{C} \overset{\mathrm{def}}{=} \mathcal{F} \circ \log \circ \mathcal{L}$, where $\mathcal{L}$ denotes the Laplace transform[9].

From the properties of the Laplace and Fenchel transforms the following result is clear.

Theorem 9.2. *For $\mu, \nu \in \mathcal{M}$ we have $\mathcal{C}(\mu * \nu) = \mathcal{C}(\mu) \,\square\, \mathcal{C}(\nu)$.*

The Cramer transform transforms convolutions into inf-convolutions and consequently independent random variables into independent decision variables. In Table 1 we summarize the main properties and examples concerning the Cramer transform when $E = \mathbb{R}$. The difficult results of this table can be found in Azencott [6]. In this table we have denoted

$$H(x) \overset{\mathrm{def}}{=} \begin{cases} 0 & \text{for } x \geq 0, \\ +\infty & \text{elsewhere.} \end{cases}$$

Let us give an example of utilization of these results in the domain of partial differential equations (PDE). Processes with independent increments are transformed into decision processes with independent increments. This implies that a generator $\hat{c}(-\partial_x)$ of a stochastic process is transformed into the generator of the corresponding decision process $v \mapsto -\hat{c}(\partial_x v)$.

Theorem 9.3. *The Cramer transform v of the solution r of the PDE on $E = \mathbb{R}$*

$$-\partial_t r + [\hat{c}(-\partial_x)](r) = 0, \ r(0,.) = \delta \ ,$$

(with $\hat{c} \in \mathcal{C}_{\mathrm{x}}$) satisfies the HJB equation

$$\partial_t v + \hat{c}(\partial_x v) = 0, \ v(0,.) = \chi \ . \tag{9.1}$$

This last equation is the forward HJB equation of the control problem of dynamic $x' = u$, instantaneous cost $c(u)$ and initial cost χ.

Remark 9.4. First let us remark that $\hat{c}$ is convex l.s.c. and not necessarily polynomial which means that fractional derivatives may appear in the PDE.

[9] $\mu \mapsto \int_E e^{\langle \theta, x \rangle} \mu(dx)$.

Table 1: Properties of the Cramer transform.

$\mathcal{M}$	$\log(\mathcal{L}(\mathcal{M})) = \mathcal{F}(\mathcal{C}(\mathcal{M}))$	$\mathcal{C}(\mathcal{M})$
μ	$\hat{c}_\mu(\theta) = \log \int e^{\theta x} d\mu(x)$	$c_\mu(x) = \sup_\theta(\theta x - \hat{c}(\theta))$
0	$-\infty$	$+\infty$
δ_a	θa	χ_a
$\lambda e^{-\lambda x - H(x)}$	$H(\lambda - \theta) + \log(\lambda/(\lambda - \theta))$	$H(x) + \lambda x - 1 - \log(\lambda x)$
$p\delta_0 + (1-p)\delta_1$	$\log(p + (1-p)e^\theta)$	$x \log(\frac{x}{1-p})$ $+(1-x)\log(\frac{1-x}{p})$ $+H(x) + H(1-x)$
stable distrib.	$m\theta + \frac{1}{p'}\lvert\sigma\theta\rvert^{p'} + H(\theta)$ $1 < p' < 2$	$c(x) = \mathcal{M}^p_{m,\sigma}$, $x \geq m$ $c(x) = 0,\ x < m,$ $1/p + 1/p' = 1$
Gauss distrib.	$m\theta + \frac{1}{2}\lvert\sigma\theta\rvert^2$	$\mathcal{M}^2_{m,\sigma}$
$\mu * \nu$	$\hat{c}_\mu + \hat{c}_\nu$	$c_\mu \,\square\, c_\nu$
$k\mu$	$\log(k) + \hat{c}$	$c - \log(k)$
$\mu \geq 0$	$\hat{c}$ convex l.s.c.	c convex l.s.c.
$m_0 \stackrel{\text{def}}{=} \int \mu$	$\hat{c}(0) = \log(m_0)$	$\inf_x c(x) = -\log(m_0)$
$m_0 = 1$	$\hat{c}(0) = 0$	$\inf_x c(x) = 0$
$S_\mu \stackrel{\text{def}}{=} \overline{\text{cvx}(\text{supp}(\mu))}$ $m_0 = 1$	$\hat{c}$strictly convex in $D_{\hat{c}}$ $\hat{c}$isC^∞in $\overset{\circ}{D}_{\hat{c}}$	$\overset{\circ}{D}_c = \overset{\circ}{S}_\mu$ $c$isC^1in $\overset{\circ}{D}_c$
$m_0 = 1,\ m \stackrel{\text{def}}{=} \int x\mu$	$\hat{c}'(0) = m$	$c(m) = 0$
$m_0 = 1,\ m_2 \stackrel{\text{def}}{=} \int x^2\mu$	$\hat{c}''(0) = \sigma^2 \stackrel{\text{def}}{=} m_2 - m^2$	$c''(m) = 1/\sigma^2$
$m_0 = 1,\ 1 < p' < 2$ $\hat{c} = \lvert\sigma\theta\rvert^{p'}/p' + o(\lvert\theta\rvert^{p'})$ $+H(\theta)$	$\hat{c}^{(p')}(0^+) = \Gamma(p')\sigma^{p'}$	$c^{(p)}(0^+) = \Gamma(p)/\sigma^p$

Proof. The Laplace transform of r denoted q satisfies:

$$-\partial_t q(t,\theta) + \hat{c}(\theta) q(t,\theta) = 0,\ q(0,.) = 1\ .$$

Therefore $w = \log(q)$ satisfies:

$$-\partial_t w(t,\theta) + \hat{c}(\theta) = 0,\ w(0,.) = 0\ , \tag{9.2}$$

which can be easily integrated. As soon as $\hat{c}$ is l.s.c and convex w is l.s.c and convex and can be considered as the Fenchel transform of a function v. The function v satisfies a PDE which can be easily computed. Indeed we have:

$$w(t,\theta) = \sup_x(\theta x - v(t,x)) \Longrightarrow \left\{ \begin{array}{c} \theta = \partial_x v\ , \\ \partial_t w = -\partial_t v\ . \end{array} \right.$$

Therefore v satisfies equation (9.1). This equation is the forward HJB equation of the control problem with dynamic $x' = u$, instantaneous cost $c(u)$ and initial cost χ because $\hat{c}$ is the Fenchel transform of c and the HJB equation of this control problem is

$$-\partial_t v + \min_u \{-u\partial_x v + c(u)\} = 0,\ v(0,.) = \chi\ . \quad \square$$

If $\hat{c}$ is independent of time the optimal trajectories are straight lines and $v(x) = tc(x/t)$. This can be obtained by using (9.2).

Solution of linear PDE with constant coefficients can be computed explicitly by Fourier transform. The previous theorem shows that that nonlinear convex first order PDE with constant coefficients are isomorphic to linear PDE with constant coefficients and therefore can be computed explicitly. Such explicit solutions of HJB equation are known as Hopf formulas [9]. Let us develop the computations on a non trivial example.

Example 9.5. Let us consider the HJB equation

$$\partial_t v + \frac{1}{2}(\partial_x v)^2 + \frac{2}{3}(|\partial_x v|)^{\frac{3}{2}} = 0,\ v(0,.) = \chi\ .$$

From (9.2) we deduce that :

$$w(t,\theta) = t(\frac{1}{2}\theta^2 + \frac{2}{3}|\theta|^{\frac{3}{2}})\ ,$$

therefore using the fact that the Fenchel transform of a sum is an inf-convolution we obtain:

$$v(t,x) = \frac{x^2}{2t} \,\square\, \frac{|x|^3}{3t^2}\ .$$

We can verify on this explicit formula a continuous time version of central limit theorem. Using the scaling $x = yt^{2/3}$, we have

$$\lim_{t\to+\infty} v(t, yt^{2/3}) = y^3/3\ ,$$

since the shape around zero of the corresponding instantaneous cost $c(u) = (u^2/2) \,\square\, (|u|^3/3)$ is $|u|^3/3$. Indeed a simple computation shows that $c(u)$ is obtained from

$$\begin{cases} c = y^4/2 + |y|^3/3\ , \\ u = |y|y + y\ , \end{cases}$$

by elimination of y. This system may be also considered as a parametrical definition of $c(u)$.

Notes and Comments. Bellman [11] was aware of the interest of the Fenchel transform (which he calls max transform) for the analytic study of the dynamic programming equations. The bicontinuity of the Fenchel transform has been well studied in convex analysis [22, 5, 4].

Maslov has started the study of idempotent integration in [24]. He has been followed in particular by [23, 25, 26, 16, 15, 10, 3, 1, 2] and independently by [28]. In [27] idempotent Sobolev spaces have been introduced as a way to study HJB equation as a linear object. In this paper the min-plus weak convergence has been also introduced but for compact support test functions. This weak convergence is used in [17] for the approximation of HJB equations. In [29] and [7] the law of large numbers and the central limit theorem for decision variables has been given in the particular case $p = 2$. In two independent works [16, 15] and [10] the study of decision variables has been started. The second work has been continued in [3]. A lot of results announced in [3] are proved in [1] and [2].

The Cramer transform is an important tool in large deviations literature [6, 21, 31, 18]. In [16, 7, 3] Cramer transform has been used in the min-plus context.

Some aspects of [32, 33, 8, 12], for instance the morphism between LQG and LEQG problems presented in [32, Section 6.1] and the separation principle developed in [12], provide other illustrations of the analogy between probability and decision calculus.

References

[1] Akian, M.: Densities of idempotent measures and large deviations. INRIA Report **2534** (1995).

[2] Akian, M.: Theory of cost measures: convergence of decision variables. INRIA Report **2611** (1995).

[3] Akian, M., Quadrat J.P., Viot M.: Bellman processes, in 11th Inter. Conf. on Analysis and Optimization of Systems. L.N. in Control and Inf. Sc. No.199, Springer Verlag (1994).

[4] Attouch, H: Variational convergence for functions and operators. Pitman (1984).

[5] Attouch, H., Wets, R.J.B.: Isometries for the Legendre-Fenchel transform. Transactions of the American Mathematical Society **296**, No. 1 (1986) 33–60.

[6] Azencott, R., Guivarc'h, Y., Gundy, R.F.: Ecole d'été de Saint Flour 8. Lect. Notes in Math., Springer-Verlag, Berlin (1978).

[7] Baccelli, F., Cohen, G., Olsder, G.J., Quadrat, J.P.: Synchronization and linearity: an algebra for discrete event systems. John Wiley and Sons, New York (1992).

[8] Basar, T., Bernhard, P.: H_∞ optimal control and relaxed minimax design problems. Birkhaüser (1991).

[9] Bardi, M., Evans, L.C. On Hopf's formulas for solutions of Hamilton-Jacobi equations. Nonlinear Analysis, Theory, Methods & Applications, Vol.8., No. 11 (1984) 1373–1381.

[10] Bellalouna, F.: Un point de vue linéaire sur la programmation dynamique. Détection de ruptures dans le cadre des problèmes de fiabilité. Thesis dissertation, University of Paris-IX Dauphine (1992).

[11] Bellman, R., Karush, W.: Mathematical programming and the maximum transform. SIAM Journal of Applied Mathematics **10** (1962).

[12] Bernhard, P.: Discrete and continuous time partial information minimax control. Submitted to Annals of ISDG (1994).

[13] Billingsley, P.: Convergence of probability measures. John Wiley & Sons, New York (1968).

[14] Cuninghame-Green, R.: Minimax algebra. Lecture Notes in Economics and Mathematical Systems No. 166, Springer Verlag (1979).

[15] Del Moral, P.: Résolution particulaire des problèmes d'estimation et d'optimisation non-linéaires. Thesis dissertation, Toulouse, France (1994).

[16] Del Moral, P., Thuillet, T., Rigal, G., Salut, G.: Optimal versus random processes: the non-linear case. LAAS Report, Toulouse, France (1990).

[17] Dudnikov, P.I., Samborski, S.: Networks methods for endomorphisms of semimodules over min-plus Algebras, in 11th Inter. Conf. on Analysis and Optimization of Systems. L.N. in Control and Inf. Sc. No.199, Springer Verlag (1994).

[18] Ellis, R. S.: Entropy, large deviations, and statistical mechanics. Springer Verlag, New York (1985).

[19] Feller, W.: An introduction to probability theory and its applications, John Wiley and Sons, New York (1966).

[20] Fenchel, W.: On the conjugate convex functions. Canadian Journal of Mathematics **1** (1949) 73–77.

[21] Freidlin, M.I., Wentzell, A.D.: Random perturbations of dynamical systems. Springer-Verlag, Berlin (1984).

[22] Joly, J.L.: Une famille de topologies sur l'ensemble des fonctions convexes pour lesquelles la polarité est bicontinue. J. Math. pures et appl. **52** (1973) 421–441.

[23] Kolokoltsov, V. N. and Maslov, V. P.: The general form of the endomorphisms in the space of continuous functions with values in a numerical commutative semiring (with the operation $\oplus$=max). Soviet Math. Dokl. Vol. 36, No. 1 (1988) 55–59.

[24] Maslov, V.: Méthodes opératorielles. Éditions MIR, Moscou (1987).

[25] Maslov, V. P. and Kolokoltsov, V. N.: Idempotent analysis and its applications to optimal control theory. Moskow, Nauka (1994) in Russian.

[26] Maslov, V., Samborski, S.N.: Idempotent analysis. Advances In Soviet Mathematics **13**, Amer. Math. Soc., Providence (1992).

[27] Maslov, V., Samborski, S.N.: Stationary Hamilton-Jacobi and Bellman equations (existence and uniqueness of solutions), in Idempotent analysis. Advances In Soviet Mathematics **13**, Amer. Math. Soc., Providence (1992).

[28] Pap, E.: Solution of nonlinear differential and difference equations. EUFIT'93, Aachen, Sept 7–10 (1993) 498–503.

[29] Quadrat, J.P.: Théorèmes asymptotiques en programmation dynamique. Note CRAS **311** Paris (1990) 745–748.

[30] Rockafellar, R.T.: Convex analysis. Princeton University Press Princeton, N.J. (1970).

[31] Varadhan, S.R.S.: Large deviations and applications. CBMS-NSF Regional Conference Series in Applied Mathematics No. 46, SIAM Philadelphia, Penn.

(1984).

[32] Whittle, P.: Risk sensitive optimal control. John Wiley and Sons, New York (1990).

[33] Whittle, P.: A risk-sensitive maximum principle : the case of imperfect state observation. IEEE Trans. Auto. Control, AC-36 (1991) 793–801.