Principe conditionnel de Gibbs pour des contraintes fines approchées et Inégalités de transport

Nathaël Gozlan

UNIVERSITÉ PARIS X – NANTERRE
U.F.R. SEGMI – Équipe MODAL'X

N^o attribué par la bibliothèque

|_|_|_|_|_|_|_|_|_|_|_|

THÈSE

pour l'obtention du Diplôme de

DOCTEUR DE L'UNIVERSITÉ PARIS X

Discipline : MATHÉMATIQUES

présentée par

Nathaël GOZLAN

Principe conditionnel de Gibbs pour des contraintes fines approchées
et
Inégalités de Transport

Soutenue publiquement **le 28 juin 2005**, devant le jury composé de

M.	**Patrick CATTIAUX**,	Université Paris 10,	Directeur de thèse
M.	**Francis COMETS**,	Université Paris 7,	Examinateur
M.	**Fabrice GAMBOA**,	Université Toulouse 3,	Rapporteur
M.	**Arnaud GUILLIN**,	Université Paris 9,	Examinateur
M.	**Christian LÉONARD**,	Université Paris 10,	Examinateur
M.	**Cédric VILLANI**,	E.N.S. Lyon,	Examinateur

au vu des rapports de **M. Fabrice GAMBOA** et **M. Liming WU** (Université Clermont 2).

Remerciements

Je tiens tout d'abord à exprimer ma reconnaissance à mon directeur de thèse, Patrick Cattiaux, non seulement pour ses conseils avisés sur le plan mathématique, mais aussi pour ses qualités humaines, l'enthousiasme et la curiosité qui l'animent, son humour et sa patience. J'ai passé grâce à lui quatre années de recherche stimulantes dans un climat détendu et sympathique.

Je voudrais remercier également Sylvie Méléard, qui après avoir dirigé mon mémoire de DEA, m'a encouragé à faire une thèse et m'a mis en contact avec Patrick Cattiaux.

J'adresse mes remerciements à Fabrice Gamboa et Li-Ming Wu qui ont accepté d'évaluer ce travail de recherche, ainsi qu'à Francis Comets, Arnaud Guillin, Christian Léonard et Cédric Villani qui me font l'honneur de faire partie de mon jury de thèse.

Il m'a été très agréable de travailler avec Christian Léonard. Je le remercie pour les nombreuses discussions que nous avons pu avoir tout au long de ces quatre années ainsi que pour ces passionnantes journées passées ensemble, lors de la préparation de notre article.

Je tiens à saluer tous les doctorants et ATER que j'ai croisé pendant ces années passées à Nanterre. Ils ont été de bien sympathiques compagnons de route.

Mes remerciements vont pour finir à ma famille et mes amis qui m'ont soutenu sans faillir tout au long de ma thèse, et à Laurence qui partage ma vie.

A la mémoire de ma mère

Table des matières

CHAPITRE I

Introduction Générale

Cette thèse est consacrée à deux sujets distincts : l'étude des principes conditionnels de type Gibbs et les inégalités de transport. Le matériel constituant ce travail est issu de trois articles :

- *Deviations bounds and Gibbs conditional principle for thin sets*, article écrit en collaboration avec Patrick Cattiaux.

- *Conditional principles for random weighted measures*, à paraître dans la revue *ESAIM P&S*.

- *A large deviation approach to some transportation cost inequalities*, article écrit en collaboration avec Christian Léonard.

Première partie : principes conditionnels

La théorie des Grandes Déviations étudie le taux de décroissance exponentielle des probabilités de certains systèmes aléatoires. D'une manière informelle, une suite de variables aléatoires $(N_i)_{i\in\mathbb{N}^*}$ à valeurs dans un espace Σ suit un Principe de Grandes Déviations (P.G.D) s'il existe une fonction $\mathrm{I} : \Sigma \to \mathbb{R}^+ \cup \{+\infty\}$ telle que pour tout ensemble C mesurable, on ait

$$\mathbb{P}\left(N_n \in C\right) \approx e^{-n\,\mathrm{I}(C)}, \qquad \text{lorsque } n \to +\infty,$$

en notant $\mathrm{I}(C) = \inf\{\mathrm{I}(x), x \in C\}$. La fonction I est appelée la *fonction de taux* du P.G.D.

La définition rigoureuse d'un P.G.D est énoncée ci-dessous :

Définition. Soit $(\Sigma, \mathcal{B})$ un espace mesurable muni d'une topologie séparée. On dit qu'une suite de variables aléatoires $(N_n)_n$ à valeurs dans Σ suit un Principe de Grandes Déviations de bonne fonction de taux I, si

1. La fonction $\mathrm{I} : \Sigma \to \mathbb{R}^+$ est une fonction inf-compact, *ie* $\forall r \geq 0, \{\mathrm{I} \leq r\}$ est compact.
2. Pour tout $C \in \mathcal{B}$, on a

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in C) \geq -\inf \left\{ \mathrm{I}(\sigma) : \sigma \in \mathring{C} \right\}.$$

et

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in C) \leq -\inf \left\{ \mathrm{I}(\sigma) : \sigma \in \overline{C} \right\}.$$

Dans certaines situations, on veut non seulement estimer les probabilités d'événements rares, mais aussi être capable de décrire l'évolution la plus probable du système lorsqu'un tel événement se produit. On s'intéresse alors au comportement asymptotique d'objets de la forme :

$$\mathcal{L}\left(N_n \,|N_n \in C\right). \tag{I.1}$$

Un théorème qui précise le comportement de ce type d'objet est appelé dans la littérature *Principe conditionnel.*

Le conditionnement $N_n \in C$ peut se comprendre de deux manières différentes :

- Ce conditionnent peut représenter une évolution particulièrement indésirable du système ; connaître sa réalisation la plus probable peut permettre de reparamétrer le système pour éviter des dégâts.
- Ce conditionnement peut également faire partie intégrante de la modélisation en représentant une contrainte matérielle effective. Prenons l'exemple de N utilisateurs partageant k ressources : si les ressources étaient infinies, les ressources utilisées par les N utilisateurs seraient modélisées par N vecteurs aléatoires indépendants et identiquement distribués à valeurs dans $\mathbb{N}^k$: $X_1, \ldots, X_N$; ces ressources étant finies la loi réelle d'un utilisateur typique est

$$\mathcal{L}\left(X_1 \left| \sum_{i=1}^{N} X_i \in C \right.\right),$$

avec $C = \prod_{i=1}^{k} [0, Nr_i]$. Le nombre d'utilisateurs étant supposé très grand, on cherchera à calculer

$$\lim_{N\to+\infty} \mathcal{L}\left(X_1 \left| \frac{1}{N} \sum_{i=1}^{N} X_i \in \prod_{i=1}^{k} [0, r_i] \right.\right).$$

Le calcul de cette limite relève du principe conditionnel de Gibbs que nous allons voir plus loin.

D'une manière générale, la suite de probabilités (I.1) s'accumule exponentiellement rapidement sur l'ensemble des minimisants de la fonction de taux I sur C, comme le montre la proposition suivante que l'on doit à D.W. Stroock et O. Zeitouni (voir [64]).

Notation : Pour tout ensemble A de Σ, nous noterons $\mathrm{I}(A) = \inf\{\mathrm{I}(\sigma) : \sigma \in A\}$.

Proposition. *Soit Σ un espace polonais muni de sa tribu borélienne et $(N_n)_n$ une suite de variables aléatoires à valeurs dans Σ qui satisfait un P.G.D. de bonne fonction de taux* I. *Si C un ensemble mesurable tel que $I_C = \mathrm{I}(\overset{\circ}{C}) = \mathrm{I}(\overline{C})$, alors $\mathbb{P}(N_n \in C) > 0$ pour tout n assez grand, et en posant $\mathcal{I} := \{\sigma \in \overline{C} : \mathrm{I}(\sigma) = I_C\}$, on a pour tout ouvert Γ tel que $\mathcal{I} \subset \Gamma$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(N_n \in \Gamma^c \,|N_n \in C\right) < 0.$$

En particulier, si $\mathcal{I} = \{\sigma^*\}$, alors

$$\mathcal{L}\left(N_n \,|N_n \in C\right) \xrightarrow[n\to+\infty]{} \delta_{\sigma^*}, \tag{I.2}$$

au sens de la convergence étroite sur $\mathcal{P}\left(\Sigma\right)$.

Démonstration. Si Γ est un ouvert tel que $\mathcal{I} \subset \Gamma$, alors

$$\frac{1}{n} \log \mathbb{P}\left(N_n \in \Gamma^c \,|N_n \in C\right) = \frac{1}{n} \log \mathbb{P}\left(N_n \in \Gamma^c \cap C\right) - \frac{1}{n} \log \mathbb{P}\left(N_n \in C\right).$$

Grâce au principe de grandes déviations, on en déduit que

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(N_n \in \Gamma^c \,|N_n \in C\right) \leq -\,\mathrm{I}(\Gamma^c \cap \overline{C}) + \mathrm{I}(\overset{\circ}{C}).$$

On voit facilement que $\mathrm{I}(\Gamma^c \cap \overline{C}) > I_C$ et, par conséquent,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(N_n \in \Gamma^c \,|N_n \in C\right) < 0.$$

En particulier, si $\mathcal{I} = \{\sigma^*\}$, alors pour tout ensemble F fermé, on a

$$\limsup_{n\to+\infty} \mathbb{P}\left(N_n \in F \,|N_n \in C\right) \leq \delta_{\sigma^*}(F),$$

ce qui signifie que $\mathcal{L}\left(N_n \,|N_n \in C\right) \xrightarrow[n\to+\infty]{} \delta_{\sigma^*}$, étroitement dans $\mathcal{P}\left(\Sigma\right)$. $\square$

Le cas où la fonction de taux I est strictement convexe sur son domaine et l'ensemble C est convexe est particulièrement favorable, puisque dans ce cas $\mathcal{I}$ contient au plus un point.

Quelques principes conditionnels classiques

Voyons les principes conditionnels associés aux principes de grandes déviations classiques.

Principe conditionnel pour la moyenne empirique

Soit μ une mesure de probabilité sur un espace de Banach B. Sur le dual topologique B', on définit la Log-Laplace de μ par :

$$\forall \lambda \in B', \quad \Lambda_\mu(\lambda) = \log \int_B e^{\langle \lambda, x \rangle} \, d\mu.$$

La transformée de Cramér Λ_μ^* de μ est par définition la transformée de Fenchel-Legendre de Λ_μ, c'est-à-dire

$$\forall x \in B, \quad \Lambda_\mu^*(x) = \sup_{\lambda \in B'} \{ \langle \lambda, x \rangle - \Lambda_\mu(\lambda) \}.$$

Le théorème de Cramér affirme que si $(X_i)_i$ est une suite de variables aléatoires i.i.d de loi μ, et si $0 \in \overset{\circ}{\mathrm{dom}} \Lambda_\mu$, alors la moyenne empirique $M_n = \dfrac{1}{n} \displaystyle\sum_{i=1}^n X_i$ suit un principe de grandes déviations sur B de bonne fonction de taux Λ_μ^*.

Sous l'hypothèse supplémentaire

$$\forall t > 0, \quad \int_B e^{t\|x\|} \, d\mu < +\infty,$$

on peut montrer que Λ_μ^* est strictement convexe sur son domaine. Le principe conditionnel associé à ce P.G.D, appelé le plus souvent *loi faible des grands nombres conditionnelle*[1], affirme alors que, pour tout ouvert convexe C tel que $C \cap \mathrm{dom}\, \Lambda_\mu^* \neq \emptyset$,

$$\mathcal{L}(M_n | M_n \in C) \xrightarrow[n \to +\infty]{} \delta_{x^*}, \quad \text{étroitement sur } \mathcal{P}(\mathcal{X}), \tag{I.3}$$

où x^* est l'unique minimisant de Λ_μ^* sur $\overline{C}$. Ce point x^* est appelé *point dominant de* C. Cette notion a été introduite et étudiée en dimension finie par P. Ney dans [52, 53], puis généralisée par U. Einmalhl et J. Kuelbs dans [31] et [40]. Elle permet d'obtenir un raffinement des bornes de grandes déviations de la forme :

$$\alpha_1 n^{-1/2} e^{-n\Lambda_\mu^*(x^*)} \leq \mathbb{P}(M_n \in C) \leq \alpha_2 n^{-1/2} e^{-n\Lambda_\mu^*(x^*)},$$

les constantes α_1 et α_2 dépendant, entre autre, de manière subtile de la géométrie de C au voisinage de x^*. Dans [41], J. Kuelbs et A. Meda ont utilisé cette technologie pour

[1] en anglais, *Conditional weak law of large numbers*.

démontrer des versions plus précises de (I.3) : ils obtiennent, sous diverses hypothèses, des vitesses ε_n explicites telles que

$$\mathbb{P}\left(\|M_n - x^*\| \leq \varepsilon_n \,|M_n \in C\right) \xrightarrow[n\to+\infty]{} 1.$$

Le principe conditionnel de Gibbs

Le principe conditionnel de Gibbs a pour objet le comportement limite de la mesure empirique d'une suite de variables aléatoires $(X_i)_i$ indépendantes et identiquement distribuées :

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

sous la contrainte $L_n \in C$. Le célèbre théorème de Sanov affirme que si les X_i sont i.i.d de loi μ et à valeurs dans un espace polonais $\mathcal{X}$, alors la suite $(L_n)_n$ satisfait un P.G.D de bonne fonction de taux $\mathrm{H}\left(\,.\,|\,\mu\right)$ définie par

$$\mathrm{H}\left(\nu|\,\mu\right) = \begin{cases} \int_{\mathcal{X}} \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{si } \nu \ll \mu \\ +\infty & \text{sinon.} \end{cases},$$

ceci pour la topologie de la convergence étroite et la tribu borélienne associée (voir le théorème II.21 pour des extensions). La fonction $\mathrm{H}\left(\,.\,|\,\mu\right)$ s'appelle *distance de Kullback* ou *entropie relative*. Là encore, si C est un ensemble convexe tel que $\mathrm{H}(\overset{\circ}{C}|\mu) = \mathrm{H}\left(\overline{C}\,|\,\mu\right)$, alors

$$\mathcal{L}\left(L_n \,|L_n \in C\right) \xrightarrow[n\to+\infty]{} \delta_{\mu^*}, \quad \text{étroitement sur } \mathcal{P}\left(\mathcal{P}(\mathcal{X})\right), \tag{I.4}$$

où μ^* est l'unique minimisant de $\mathrm{H}\left(\,.\,|\,\mu\right)$ sur $\overline{C}$. La probabilité μ^* est appelée *I-projection de μ sur C*. Le chapitre II de cette thèse sera consacré à cette notion introduite et étudiée par I. Csiszár dans [18, 19]. C'est également à I. Csiszár que l'on doit la première démonstration de (I.4) pour des ensembles C convexes (voir [19]).
C'est une question de Mécanique Statistique qui a motivé l'étude de $\mathcal{L}(L_n|L_n \in C)$: on suppose que les $(X_i)_i$ représentent des particules, chaque particule ayant une énergie $F(X_i)$ et on s'intéresse à la loi conditionnelle de $(X_1, \ldots, X_k)$ (k fixé) sachant que le nuage de particules a une énergie moyenne donnée :

$$\langle L_n, F\rangle = \frac{1}{n}\sum_{i=1}^{n} F(X_i) \in [a,b].$$

Le nombre de particules étant très grand, le problème mathématique se résume à calculer la limite suivante :

$$\lim_{n\to+\infty} \mathcal{L}(X_1, \ldots, X_k|L_n \in C), \tag{I.5}$$

avec $C = \left\{\nu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} F\,d\nu \in [a,b]\right\}$. Comme le montre le lemme suivant, déterminer la limite de (I.5) pour tout $k \in \mathbb{N}^*$, revient à déterminer la limite de $\mathcal{L}(L_n|L_n \in C)$, lorsque $n \to +\infty$.

Lemme (Propagation du chaos). *Si $\mathcal{X}$ est un espace polonais et si, pour tout $n \in \mathbb{N}^*$, μ^n est une probabilité symétrique sur $\mathcal{X}^n$ (ie μ^n est invariante par permutations des coordonnées), alors les deux propositions suivantes sont équivalentes :*

1. *La loi de $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ sous μ^n converge étroitement vers δ_{μ^*}.*
2. *Pour tout $k \in \mathbb{N}^*$ et pour toutes fonctions $f_1, \ldots, f_k$ continues bornées sur $\mathcal{X}$, on a*

$$\int_{\mathcal{X}^k} f_1(x_1)\cdots f_k(x_k)\, d\mu^n \xrightarrow[n\to+\infty]{} \int_{\mathcal{X}^k} f_1(x_1)\cdots f_k(x_k)\, d\mu^{*\otimes k}.$$

Démonstration. Voir l'annexe A ou la preuve du lemme 3.1 de [65]. □

En appliquant ce résultat avec $\mu^n = \dfrac{\mathbb{1}_C(L_n)}{\mu^{\otimes n}(L_n \in C)}\mu^{\otimes n}$, on voit que (I.4) équivaut à

$$\forall k \in \mathbb{N}^*, \quad \mathcal{L}(X_1, \ldots, X_k | L_n \in C) \xrightarrow[n\to+\infty]{} \mu^{*\otimes k}. \tag{I.6}$$

De plus, pour un ensemble C de la forme

$$C = \left\{\nu \in \mathcal{P}(\mathcal{X}), \int_{\mathcal{X}} F\, d\nu \in [a, b]\right\},$$

nous verrons dans le chapitre II, que la I-projection μ^* est en général une mesure de Gibbs

$$d\mu^* = Z^{-1}\exp(-\beta F)\, d\mu.$$

Ainsi, pour tout k, les variables $(X_1, \ldots, X_k)$ sont conditionnellement asymptotiquement indépendantes et identiquement distribuées, avec pour loi limite une mesure de Gibbs.

Principe conditionnel pour des mesures à poids aléatoires

Donnons nous une mesure de référence R sur un espace polonais $\mathcal{X}$, ainsi qu'une famille de points $(x_i^n)_{i=1\ldots n}$ choisis de telle sorte que

$$\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n} \xrightarrow[n\to+\infty]{} R,$$

(on peut prendre par exemple les réalisations d'une suite i.i.d de loi R) et posons

$$L_n = \frac{1}{n}\sum_{i=1}^n Z_i \delta_{x_i^n}, \tag{I.7}$$

avec $(Z_i)_i$ une suite de variables aléatoires à valeurs réelles i.i.d de loi μ. Ces mesures à poids aléatoires ont été introduites en mécanique statistique par Ellis *et al.* dans [32] et en théorie de l'estimation par Gamboa *et al.* dans [22, 35, 36, 21].

Si $\operatorname{dom}\Lambda_\mu = \mathbb{R}$, la suite $(L_n)_n$ suit un P.G.D sur $\mathcal{M}(\mathcal{X})$ (ensemble des mesures finies sur $\mathcal{X}$) équipé de la topologie de la convergence étroite de bonne fonction de taux

$$\mathrm{I}_\mu\left(P\,|R\right) = \int_{\mathcal{X}} \Lambda_\mu^* \left(\frac{dP}{dR}\right) dR.$$

On peut trouver une preuve de ce résultat dans [26] (thm 7.2.3). Si l'hypothèse $\operatorname{dom}\Lambda_\mu = \mathbb{R}$ n'est plus vérifiée, la fonction de taux fait apparaître des termes singuliers (voir [32] et [50]).

Sans surprise, si C est un convexe de $\mathcal{M}(\mathcal{X})$ tel que $\mathrm{I}_\mu(\overset{\circ}{C}|R) = \mathrm{I}_\mu\left(\overline{C}\,|R\right)$, on a

$$\mathcal{L}(L_n|L_n \in C) \xrightarrow[n\to+\infty]{} \delta_{R^*} \tag{I.8}$$

la mesure R^* étant l'unique minimisant de $\mathrm{I}_\mu\left(\,.\,|R\right)$ sur C.

L'intérêt théorique de ce résultat est qu'il donne une interprétation probabiliste de certaines procédures de selection utilisées en statistique. Une question fréquente en modélisation est la suivante : comment retrouver la loi d'un phénomène aléatoire à partir de certaines observations moyennes de celui-ci ? Ce problème est le plus souvent mal posé et il s'agit de sélectionner un élément dans l'ensemble C, généralement très grand, de toutes les mesures (de probabilité ou non) conformes aux observations empiriques. Dans certains cas, on dispose d'un modèle *a priori* R. L'objectif est de modifier R de telle sorte qu'il s'ajuste aux observations. Dans [20], I. Csiszár a posé les axiomes de ce qu'on est en droit d'attendre d'une *procédure de sélection avec a priori*. Il ressort de ce travail qu'une telle procédure est le fruit de la minimisation sous contraintes de deux types de fonctionnelles. Ces deux classes de fonctionnelles sont les distances de Bregman sur lesquelles nous ne reviendrons pas et les γ-divergences, c'est-à-dire les fonctionnelles de la forme

$$\mathrm{I}_\gamma\left(P\,|R\right) = \int_{\mathcal{X}} \gamma \left(\frac{dP}{dR}\right) dR,$$

la fonction γ étant convexe et positive. Cette classe de fonctionnelle contient notamment l'entropie relative, obtenue pour la fonction $\gamma(x) = x\log x + 1 - x$. Les fonctions de taux des P.G.D associés aux mesures aléatoires L_n (définies par (I.7)) sont des γ-divergences. On remarquera, en particulier, que l'entropie relative est obtenue en prenant des poids Z_i poissonniens de moyenne 1. Le principe conditionnel (I.8) permet ainsi de comprendre de manière plus probabiliste le minimisant de $\mathrm{I}_\mu\left(\,.\,|R\right)$ sur C. Celui-ci est théoriquement simulable grâce à une méthode d'acceptation-rejet basée sur les observations de L_n. Une telle méthode est, bien entendu, irréalisable en pratique puisque l'événement $L_n \in C$ se produit avec une probabilité tendant exponentiellement rapidement vers 0...

Présentation des chapitres

Le problème auquel s'attache cette thèse est celui des contraintes fines. Comment donner un sens à

$$\mathcal{L}\left(N_n \left| N_n \in C\right.\right)$$

lorsque $\mathbb{P}(N_n \in C) = 0$ pour une infinité d'entiers n ?

L'idée la plus satisfaisante d'un point de vue théorique serait de définir cette probabilité en utilisant une désintégration exacte de la mesure. Ce point de vue a été développé dans [69, 74, 11] dans le cas particulier de l'étude de

$$\mathcal{L}(X_1 | X_1 + \cdots + X_n = c_n), \tag{I.9}$$

où X_i est une suite i.i.d de variables aléatoires, et c_n une suite de nombres réels. Dans [69], T. Tjur a montré que si $c_n = n\mathbb{E}[X_1]$, alors (I.9) converge vers $\mathcal{L}(X_1)$. Dans [74], S. Zabell a étudié la convergence de (I.9) lorsque $c_n = n\mathbb{E}[X] + d_n$, d_n étant une suite de limite nulle. Il a obtenu des vitesses explicites pour d_n garantissant la convergence de (I.9) vers $\mathcal{L}(X_1)$. Enfin, dans [11], J. Van Campenhout et T. Cover ont étendu les résultats précédents à des suites c_n de la forme $c_n = nx + d_n$, x pouvant être différent de $\mathbb{E}[X_1]$. Cette approche, fondée sur une désintégration exacte, semble difficile à mener en toute généralité.

Un point de vue plus raisonnable est celui adopté par Stroock et Zeitouni dans [64]. Il consiste à grossir la contrainte fine C, en considérant une famille croissante $(C_\varepsilon)_\varepsilon$ d'ensembles mesurables et à étudier

$$\lim_{\varepsilon \to 0} \lim_{n \to +\infty} \mathbb{P}\left(N_n \in \,.\, \left| N_n \in C_\varepsilon\right.\right).$$

Quand la famille $(C_\varepsilon)_\varepsilon$ est bien choisie, cette limite est celle qu'on attend, à savoir le minimisant de la fonction de taux sur l'ensemble C. Ce point de vue n'est pas toujours satisfaisant. Prenons l'exemple du principe conditionnel de Gibbs, *ie* $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, supposons que C soit fermé pour la topologie de la convergence étroite et tel que $\mathrm{H}\left(C\right|\mu) < +\infty$ et posons $C_\varepsilon = \{\nu \in \mathcal{P}(\mathcal{X}), \bar{d}(\nu, C) < \varepsilon\}$, où $\bar{d}(\,.\,,\,.\,)$ est une distance métrisant la convergence étroite. A ε fixé, $\mathcal{L}(L_n | L_n \in C_\varepsilon)$ converge étroitement vers $\delta_{\mu_\varepsilon^*}$, μ_ε^* étant la I-projection de μ sur $\overline{C_\varepsilon}$ (cela résulte des premiers résultats de Csiszár sur le principe conditionnel de Gibbs). Par ailleurs, on voit facilement, en utilisant certains résultats de Csiszár sur la géométrie des I-projections, que $\delta_{\mu_\varepsilon^*} \underset{\varepsilon \to 0}{\longrightarrow} \delta_{\mu^*}$. Dans ce cas précis, on voit que la formulation en double limite n'apporte rien de nouveau.

L'objectif de cette première partie est d'obtenir une formulation en limite simple de certains principes conditionnels. Partant d'une contrainte fine convexe C, on cherchera à construire explicitement une suite décroissante C_n de convexes dont l'intersection est C

et telle que $\mathcal{L}(N_n|N_n \in C_n)$ converge vers le minimisant de la fonction de taux sur C. Sous cette forme, nous adoptons un point de vue intermédiaire entre celui hypothétique de la désintégration et celui de la double limite. Dans l'exemple précédent, nous serons en mesure, sous certaines hypothèses, de construire explicitement des suites ε_n de limite nulle telles que $\mathcal{L}(L_n|L_n \in C_{\varepsilon_n})$ converge quand $n \to +\infty$ vers δ_{μ^*}. Si, dans le cas d'une contrainte convexe C épaisse, la convergence de $\mathcal{L}\left(N_n\,|N_n \in C\right)$ vers le minimisant de la fonction de taux sur C relevait de manière directe du principe de grandes déviations satisfait par N_n, ce n'est plus le cas avec notre approche. Celle-ci requiert des bornes exactes, c'est-à-dire non-asymptotiques, pour le contrôle des petites probabilités.

Cette première partie comporte quatre chapitres. Le chapitre II est un chapitre préliminaire sur l'entropie relative. Les chapitres III et IV sont consacrés au principe conditionnel de Gibbs et le chapitre V au principe conditionnel pour des mesures à poids aléatoires. Voyons, maintenant plus en détail le contenu de chacun d'eux.

Résumé du chapitre III

Dans ce chapitre, $L_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \delta_{X_i}$ est la mesure empirique d'une suite i.i.d de loi μ sur un certain espace polonais $\mathcal{X}$. L'objectif de chapitre est de donner des conditions suffisantes pour que

$$\mathcal{L}\left(L_n\,|L_n \in C_n\right) \xrightarrow[n\to+\infty]{} \delta_{\mu^*}$$

avec C_n une suite décroissante d'ensembles convexes de $\mathcal{P}(\mathcal{X})$ d'intersection C et μ^* la I-projection de μ sur C (c'est-à-dire l'unique minimisant de $\mathrm{H}\left(\,.\,|\,\mu\right)$ sur C).

En fait, nous étudierons ce problème sous une autre forme (qui est équivalente à la précédente, tant qu'on ne s'intéresse qu'à la convergence étroite) : nous chercherons à démontrer que

$$\forall k \in \mathbb{N}^*, \quad \mu^n_{C_n,k} := \mathcal{L}(X_1, \ldots, X_k|L_n \in C_n) \xrightarrow[n\to+\infty]{} \mu^{*\otimes k}. \tag{I.10}$$

Ce qui rend cette forme plus agréable est que l'on dispose de l'inégalité suivante

$$\mathrm{H}\left(\mu^n_{C_n,k}\middle|\,\mu_n^{*\otimes k}\right) \leq -\frac{1}{\left[\frac{n}{k}\right]} \log\left(\mathbb{P}\left(L_n \in C_n\right) e^{n\,\mathrm{H}(C_n|\mu)}\right),$$

la probabilité μ_n^* étant la I-projection de μ sur C_n. Cette inégalité qui est due à I. Csiszár, s'applique dès que les C_n sont fermés en un certain sens. Grâce à ce contrôle, nous verrons au théorème III.36 que pour des topologies raisonnables, la condition

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mathbb{P}\left(L_n \in C_n\right) \geq -\,\mathrm{H}\left(C|\,\mu\right). \tag{I.11}$$

est suffisante pour avoir (I.10). Cette condition assez naturelle ne relève pas du théorème de Sanov. Cependant, en reprenant sous une forme un peu modifiée, la technique classique du recentrage exponentiel, on montre à la proposition III.46 qu'une condition suffisante pour (I.11) est

$$\lim_{n\to+\infty} \mu^{*\otimes n}(L_n \in C_n) = 1. \tag{I.12}$$

Comme μ^* appartient à C_n pour tout n, il s'agit donc de préciser la loi faible des grands nombres pour L_n sous $\mu^{*\otimes n}$.

Lorsque C est défini par une contrainte de type moment, c'est-à-dire lorsque C est de la forme

$$C = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} F \, d\nu \in K \right\},$$

avec F une fonction mesurable à valeurs dans un Banach séparable et K un convexe, une manière naturelle de grossir C est de poser, pour tout $\varepsilon > 0$

$$C_\varepsilon = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} F \, d\nu \in K^\varepsilon \right\}$$

où K^ε est un ε-voisinage de K. Il s'agit ensuite de trouver des suites ε_n telles que $C_n := C_{\varepsilon_n}$ vérifie (I.12). Pour cela, nous ferons appel à des inégalités de type Bernstein (en dimension finie) ou Yurinskii (en dimension infinie) qui garantissent que si Y_i est une suite i.i.d de loi μ^*,

$$\mu^{*\otimes n}(L_n \in C_n) \geq \mathbb{P}\left(\left\| \frac{1}{n} \sum_{i=1}^n F(Y_i) - \int_{\mathcal{X}} F \, d\mu^* \right\| \leq \varepsilon_n \right) \approx 1 - e^{-n\varepsilon_n^2}.$$

Typiquement, nous pourrons donc autoriser des vitesses de rétrécissement ε_n en $\frac{1}{n^a}$, avec $a < \frac{1}{2}$. Pour ce type de contraintes, le résultat le plus intéressant de ce chapitre est le théorème III.61 qui traite de la dimension finie. Sous des hypothèses très peu restrictives, nous obtenons la convergence en entropie de $\mu^n_{C_n,k}$ vers $\mu^{*\otimes k}$ et pour $k = 1$ la convergence a lieu en un sens encore plus fort.

Pour aborder le cas d'une contrainte convexe fine C générale, nous allons tirer partie de la métrisabilité de la topologie de la convergence étroite et poser pour tout $\varepsilon > 0$

$$C^\varepsilon = \{ \nu \in \mathcal{P}(\mathcal{X}) : \bar{d}(\nu, C) \leq \varepsilon \},$$

$\bar{d}$ étant une distance métrisant cette topologie (on considérera les métriques de Prokhorov et de Fortet-Mourier). En utilisant des résultats de S.J. Kulkarni et O. Zeitouni, nous verrons que si $\mathcal{X}$ est compact, on dispose de la borne suivante :

$$\mu^{*\otimes n}(L_n \in C^\varepsilon) \geq 1 - N_{\mathcal{P}(\mathcal{X})}\left(\bar{d}, \frac{\varepsilon}{4}\right) e^{-n\frac{\varepsilon^2}{8}}, \tag{I.13}$$

où $N_{\mathcal{P}(\mathcal{X})}\left(\bar{d},\varepsilon\right)$ est le nombre minimal de boules de rayon ε (pour la distance $\bar{d}$) nécessaire pour recouvrir l'espace compact $\mathcal{P}(\mathcal{X})$. En un mot, pour obtenir (I.13), l'idée est de recouvrir le complémentaire de C^{ε} par des boules B_i de rayon $\varepsilon/4$, d'utiliser la majoration classique $\mu^{*\otimes n}(L_n \in B_i) \le e^{-n\,\mathrm{H}(B_i|\mu^*)}$ suivie de l'inégalité de Pinsker $\bar{d}(\nu,\mu^*) \le \sqrt{2\,\mathrm{H}\,(\nu|\,\mu^*)}$. Clairement, pour que $C_n := C^{\varepsilon_n}$ vérifie (I.12), il faut que la suite ε_n tende vers 0 suffisamment lentement pour permettre au terme de grandes déviations $e^{-n\varepsilon_n^2/8}$ de compenser la croissance du nombre de boules. Des estimations "tractables" de $N_{\mathcal{P}(\mathcal{X})}\left(\bar{d},\varepsilon\right)$ en fonction de $N_{\mathcal{X}}(\varepsilon)$ existent (voir le lemme III.93). Elles permettent, à chaque fois que l'on sait estimer $N_{\mathcal{X}}(\varepsilon)$, de calculer des vitesses de rétrécissement ε_n explicites (voir le corollaire III.101 et la proposition III.105).

Si l'espace $\mathcal{X}$ n'est plus compact, on peut mettre en place une procédure d'approximation de μ^* par des probabilités à supports compacts et déduire des résultats précédents des conditions suffisantes sur ε_n pour que C_n vérifie (I.12). C'est l'objet des propositions III.106 et III.109. Cette fois, un autre facteur entre un jeu : il faut que μ^* soit rapidement approchée par des probabilités portées par des compacts dont l'entropie métrique n'explose pas trop rapidement. Ceci requiert une bonne connaissance de μ^* (typiquement de sa queue de distribution).

Nous terminons ce chapitre par une application de ces méthodes dans un cadre physique plus concret : une interprétation statistique des ponts de Schrödinger et des processus de Nelson. On s'intéresse aux comportements étranges de grands nuages de particules browniennes. Si $X_1,\ldots,X_N$ sont N particules browniennes indépendantes, le problème est de déterminer l'évolution la plus probable du nuage sachant que celui-ci a été trouvé avec une distribution approximativement égale à ν_t aux instants $t \in I$ (I étant un sous ensemble de $[0,1]$). Posant

$$C(\nu_t) = \left\{\mathcal{V} \in \mathcal{P}\left(C([0,1],\mathbb{R}^q)\right) : \forall t \in I, \mathcal{V}_t = \nu_t\right\},$$

il s'agit d'estimer

$$\lim_{N\to+\infty} \mathcal{L}(L_N|L_N \in C(\nu_t)).$$

Ceci reste bien sûr formel, puisque la contrainte $C(\nu_t)$ est une contrainte (convexe) fine. Pour de bons flots de marginales $(\nu_t)_{t\in I}$, le problème de l'existence de la I-projection $\mathcal{W}^*$ de $\mathcal{W}$ (mesure de Wiener sur $C([0,1],\mathbb{R}^q)$) sur $C(\nu_t)$ a été étudié par différents auteurs. Dans le cas où $I = \{0,1\}$, on parle de ponts de Schrödinger et pour $I = [0,1]$, de processus de Nelson. Dans les deux cas, nous montrons comment construire des suites ε_N explicites telles que

$$\lim_{N\to+\infty} \mathcal{L}\left(L_N\,|L_N \in C(\nu_t)^{\varepsilon_N}\right) = \delta_{\mathcal{W}^*}.$$

Résumé du chapitre IV

Le chapitre IV donne une interprétation en terme de principe conditionnel de Gibbs d'une méthode de calibration destinée à la finance et proposée par M. Avellaneda, C. Friedman, R. Holmes et D. Samperi dans [2]. Le problème est de modéliser un actif financier par un processus de diffusion de loi notée $\mathbb{Q}_\sigma$ solution d'une équation différentielle stochastique :

$$dX_t = \sigma(t, X_t)\, dB_t + b_0\, dt \tag{I.14}$$

et vérifiant $\mathbb{E}\left[F(X_T)\right] = 1$ pour une fonction F donnée et une date T fixée. Ici, le drift b_0 est fixé par l'absence d'arbitrage.

Le drift b_0 étant fixé, on ne peut jouer que sur le coefficient de diffusion, ce qui, d'après le théorème de Girsanov, ferme la porte à une méthode de calibration fondée sur la minimisation de l'entropie relative par rapport à une diffusion à priori $\mathbb{Q}_{\sigma_0}$. L'idée développée par Avellaneda *et al.* dans l'introduction de [2] est de minimiser l'entropie relative sur des versions discrétisées des processus. Supposons donnée, pour tout σ, une suite $\mathbb{Q}^n_\sigma$ de chaînes de Markov convergeant vers $\mathbb{Q}_\sigma$. Certains schémas d'approximation classiques, comme le schéma d'Euler ou les arbres trinomiaux, vérifient

$$\frac{1}{n} \mathrm{H}\left(\mathbb{Q}^n_\sigma \middle| \mathbb{Q}^n_{\sigma_0}\right) \xrightarrow[n\to+\infty]{} \mathrm{I}(\sigma|\sigma_0) = \mathbb{E}_\sigma\left[\int_0^1 q(\sigma^2(X_t, t), \sigma_0^2(t, X_t))\, dt\right], \tag{I.15}$$

où la fonction q dépend du schéma d'approximation choisi. Se fondant sur cette propriété, Avellaneda et ses coauteurs proposent de minimiser les fonctionnelles de la forme $\mathrm{I}(\,.\,|\sigma_0)$ sous la contrainte $\mathbb{E}_\sigma\left[F(X_T)\right] = 1$, où $\mathbb{E}_\sigma[\,.\,]$ désigne l'espérance par rapport à la loi $\mathbb{Q}_\sigma$.

Les problèmes de minimisation sous contraintes de l'entropie relative étant naturellement liés au principe conditionnel de Gibbs, nous chercherons à interpréter le minimisant $\mathbb{Q}^*$ de $\mathrm{I}(\,.\,|\sigma_0)$ sous la contrainte $\mathbb{E}_\sigma\left[F(X_T)\right] = 1$ comme une limite de la forme

$$\mathbb{Q}^* = \lim_{n\to+\infty} \mathbb{E}_{\left(\mathbb{Q}^n_{\sigma_0}\right)^{\otimes m_n}}\left[L_{m_n} \middle| L_{m_n} \in \mathcal{Q}^n\right], \tag{I.16}$$

où

- $L_m : C([0,1],\mathbb{R})^m \to \mathcal{P}(C([0,1],\mathbb{R})) : (\omega_1, \dots \omega_m) \mapsto \frac{1}{m}\sum_{i=1}^m \delta_{\omega_i}$,
- $\mathcal{Q}^n$ est l'ensemble des $\mathbb{Q}^n_\sigma$ vérifiant la contrainte $\mathbb{E}^n_\sigma\left[F(X_T)\right] \simeq 1$,
- m_n est une suite d'entiers à préciser.

Ce résultat paraît raisonnable, puisqu'à n fixé,

$$\lim_{m\to+\infty} \mathbb{E}_{\left(\mathbb{Q}^n_{\sigma_0}\right)^{\otimes m}}\left[L_m \middle| L_m \in \mathcal{Q}^n\right] \in \mathrm{Argmin}\left\{\mathrm{H}\left(\mathbb{Q} \middle| \mathbb{Q}^n_{\sigma_0}\right), \mathbb{Q} \in \mathcal{Q}^n\right\}$$

et qu'au vu de (I.15), on peut espérer que ce dernier ensemble soit proche de $\mathbb{Q}^*$.

Nous ne serons en mesure de démontrer une convergence du type (I.16) que pour un schéma d'approximation donné : les arbres trinomiaux (voir le théorème IV.29). En particulier, pour diverses raisons, notre preuve ne permet pas de traiter le schéma d'Euler. Néanmoins, grâce à ce résultat, la minimisation sous contrainte des fonctionnelles de la forme $\mathrm{I}(\,.\,|\sigma_0)$ trouve une justification plus rigoureuse.

Résumé du chapitre V

Dans le chapitre V, nous nous plaçons dans le cadre des mesures à poids aléatoires, *ie*

$$L_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \delta_{x_i^n},$$

où l'on rappelle que les Z_i sont i.i.d de loi μ et les x_i^n tels que $\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i^n}$ converge vers une certaine probabilité de référence R sur l'espace $\mathcal{X}$ considéré.

Ici, nous chercherons à démontrer des convergences de la forme

$$\mathbb{E}[L_n | L_n \in C_{\varepsilon_n}] \xrightarrow[n\to+\infty]{} R^*, \tag{I.17}$$

où C est une contrainte convexe fine et R^* est le minimisant de $\mathrm{I}_\mu\,(\,.\,|R)$ sur C. En fait, nous ne pourrons considérer que des ensembles C définis par des contraintes de type moment, c'est-à-dire de la forme

$$\mathcal{S}(F,K) := \left\{ P \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} F\, dP \in K \right\}, \text{ avec } F : \mathcal{X} \to \mathbb{R}^k \text{ et } K \text{ convexe de } \mathbb{R}^k,$$

ensemble que nous grossirons en $\mathcal{S}(F, K^\varepsilon) := \left\{ P \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} F\, dP \in K^\varepsilon \right\}$. La raison de cette restriction est qu'ici, contrairement au principe conditionnel de Gibbs, la forme algébrique particulière de R^* est utilisée dans la preuve et cette forme n'est connue que dans ce cas précis.

Pour démontrer (I.17), nous chercherons à coller au plus près à ce qui a été fait dans le cadre du principe conditionnel de Gibbs. L'outil clef du chapitre III était l'inégalité de Csiszár

$$\mathrm{H}\left(\mu_C^n |\, \mu^*\right) \le -\frac{1}{n}\log\left(\mathbb{P}\left(L_n \in C\right) e^{n\,\mathrm{H}(C|\mu)}\right), \tag{I.18}$$

où $\mu_C^n = \mathcal{L}(X_1 | L_n \in C) = \mathbb{E}[L_n | L_n \in C]$ et μ^* est la I-projection de μ sur C. Grâce à l'inégalité de Pinsker, on déduisait de (I.18) que

$$\|\mu_C^n - \mu^*\|_{VT} \le \sqrt{-\frac{2}{n}\log\left(\mathbb{P}\left(L_n \in C\right) e^{n\,\mathrm{H}(C|\mu)}\right)}. \tag{I.19}$$

Dans les raisonnements, c'est cette dernière inégalité que nous utilisions effectivement, et c'est donc une inégalité du même style que nous voulons obtenir dans le cadre des mesures à poids aléatoires. Si $R_{n,\varepsilon} := \mathbb{E}[L_n | L_n \in \mathcal{S}(F, K^\varepsilon)]$ jouera le rôle de μ_C^n, celui de μ^* sera joué non pas par R^*, mais par une certaine mesure $R^*_{n,\varepsilon}$ appelée *minimisant de l'entropie sur la moyenne*. Ces mesures ont été introduites et étudiées par Gamboa *et al.* dans [22, 35, 36, 21]. Lorsque $\operatorname{dom} \Lambda_\mu = \mathbb{R}$, l'une des manières de les définir est la suivante : en notant $\overline{R}_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$, la mesure $R^*_{n,\varepsilon}$ est le minimisant de la γ-divergence discrétisée

$$\mathrm{I}_\mu(P|\overline{R}_n) := \int_{\mathcal{X}} \Lambda_\mu^* \left(\frac{dP}{d\overline{R}_n}\right) d\overline{R}_n$$

sur l'ensemble $\mathcal{S}(F, K^\varepsilon)$. La suite de fonctions $\mathrm{I}_\mu(\,.\,|\overline{R}_n)$ converge en un sens suffisamment fort vers $\mathrm{I}_\mu(\,.\,|R)$ pour que la suite de ces minimisants sous contrainte converge également vers le minimisant sous contrainte de $\mathrm{I}_\mu(\,.\,|R)$. Autrement dit, les $R^*_{n,\varepsilon}$ convergent vers R^*_ε (voir le théorème V.8). L'inégalité qui généralise (I.19), et qui est le résultat principal de ce chapitre, est de la forme suivante :

$$\left\| R_{n,\varepsilon} - R^*_{n,\varepsilon} \right\|_{VT} \leq Q\left(\frac{-1}{n}\log\left[\mathbb{P}(L_n \in \mathcal{S}(F, K^\varepsilon)) e^{n\,\mathrm{I}_\mu(R^*_{n,\varepsilon}|\overline{R}_n)}\right]\right), \tag{I.20}$$

avec Q une fonction concave dépendant de μ (voir la proposition V.26). Si ε_n est une suite de limite nulle, la suite R^*_{n,ε_n} converge vers R^* (voir le théorème V.8). Ainsi, pour montrer (I.17), il suffit de contrôler le membre de droite de (I.20). Cette dernière étape fait intervenir des outils déjà utilisés dans le chapitre III : recentrage exacte et bornes à la Bernstein.

La démonstration de (I.20) est assez proche de celle de (I.18). L'ingrédient nouveau est donné par la proposition V.17 qui dit essentiellement que pour toute mesure de probabilité μ sur $\mathbb{R}$, on peut construire une fonction Q concave, positive croissante et nulle en 0 telle que

$$\forall \nu \in \mathcal{P}(\mathbb{R}), \qquad \left| \int_{\mathbb{R}} x\, d\nu - \int_{\mathbb{R}} x\, d\mu \right| \leq Q\left(\mathrm{H}\left(\nu | \, \mu\right)\right).$$

Ce résultat, qui est largement inspiré des travaux de S.G. Bobkov et F. Götze sur l'inégalité de transport $\mathbb{T}_1$ (voir [4]), est aussi ce qui a orienté cette thèse vers une étude des inégalités de transport et de leurs liens avec les grandes déviations.

Seconde partie : Inégalités de transport

Si ν et μ sont deux probabilités sur un espace mesurable $\mathcal{X}$ et si $c : \mathcal{X} \to \mathbb{R}^+$ est une fonction mesurable, on définit le coût de transport optimal $\mathcal{T}_c(\mu, \nu)$ de μ sur ν de la manière suivante :

$$\mathcal{T}_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \iint_{\mathcal{X}^2} c(x, y)\, d\pi(x, y), \tag{I.21}$$

où l'ensemble $\Pi(\mu, \nu)$ est l'ensemble des mesures de probabilité sur $\mathcal{X}^2$ ayant μ pour première marginale et ν pour seconde. Pour faciliter les écritures, nous supposerons toujours que c est symétrique, c'est-à-dire qu'elle vérifie $c(x, y) = c(y, x)$. De la sorte, $\mathcal{T}_c(\mu, \nu) = \mathcal{T}_c(\nu, \mu)$. L'appellation *coût de transport optimal* vient de ce qu'en interprétant $d\pi(x, y)$ comme une masse prise en x et déposée en y et en considérant qu'un tel transport élémentaire coûte le prix $c(x, y)$, on peut voir $\iint_{\mathcal{X}^2} c(x, y)\, d\pi(x, y)$ comme le coût total engendré par l'opération et $\mathcal{T}_c(\mu, \nu)$ comme le meilleur coût possible. Si le centre d'intérêt principal en théorie du transport est l'étude des plans de transport optimaux, c'est-à-dire des couplages π réalisant l'infimum dans (I.21), un autre sujet a pris ces dernières années un essor certain, c'est celui des *inégalités de transport*. On dit que μ vérifie une inégalité de transport s'il existe une fonction F telle que

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_c(\nu, \mu) \leq F\left(\mathrm{H}\left(\nu | \, \mu\right)\right). \tag{I.22}$$

Ces inégalités ont été introduites par K. Marton et M. Talagrand dans [47] et [68]. La raison de l'étude de ce genre d'inégalités est leurs liens avec les inégalités de concentration. Le chapitre VI comportant une introduction assez détaillée sur le sujet, nous nous permettrons de ne pas alourdir celle-ci et de passer à la présentation succincte de nos résultats.

Résumé du chapitre VI

Ce chapitre introduit la notion d'inégalités de transport convexes (I.T.C). Une probabilité μ sur un espace $\mathcal{X}$ satisfait l'I.T.C $T_c(\theta^*, a)$, où θ est une fonction convexe appartenant à une certaine classe $\mathcal{C}$, si

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^*\left(\frac{\mathcal{T}_c(\nu, \mu)}{a}\right) \leq \mathrm{H}\left(\nu | \, \mu\right), \tag{I.23}$$

la fonction θ^* étant la conjuguée convexe de θ. Les diverses inégalités de transport démontrées ces dernières années peuvent toutes se mettre sous cette forme. Le premier objectif de ce chapitre est d'étendre au cas général un certain nombre de résultats démontrés uniquement dans des cas particuliers. Nous obtiendrons, notamment une formulation duale à la Bobkov-Götze ainsi qu'une formule générale de tensorisation à la Marton-Talagrand. Le second objectif est d'établir des liens entre ces I.T.C et la théorie des Grandes Déviations : nous montrerons comment certaines techniques de Grandes Déviations permettent d'étudier les inégalités de transport et inversement, comment ces inégalités de transport permettent d'obtenir des inégalités de déviations.

Résumé du chapitre VII

Dans ce chapitre nous démontrons des conditions nécessaires et suffisantes pour qu'une probabilité μ vérifie (I.23). Notre résultat principal (le théorème VII.50) dit essentiellement que si θ se comporte comme x^2 au voisinage de 0, alors pour toute fonction de coût $c(x,y) = q(d(x,y))$ avec q une fonction convexe positive sur $\mathbb{R}^+$ n'explosant pas trop rapidement, l'I.T.C (I.23) est équivalente à une propriété d'intégrabilité de la forme :

$$\exists \delta > 0, \qquad \iint_{\mathcal{X}^2} e^{\theta^*(\delta c(x,y))}\, d\mu(x) d\mu(y) < +\infty.$$

Ce résultat généralise complètement les résultats de Djellout, Guillin et Wu sur l'inégalité de transport $\mathbb{T}_1$, ainsi que ceux, plus généraux, de Bolley et Villani (voir [27] et [5]).

Première partie

Principes Conditionnels

CHAPITRE II

Entropie relative, théorème de Sanov et projections entropiques

Sommaire

II.1 Introduction

Ce chapitre a pour but de regrouper les différents résultats concernant l'*entropie relative* dont nous aurons besoin dans cette thèse. Également appelée *distance de Kullback*, l'entropie relative entre deux mesures de probabilité ν et μ est définie par

$$\mathrm{H}\left(\nu|\,\mu\right) = \begin{cases} \int_{\mathcal{X}} \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{si } \nu \ll \mu \\ +\infty & \text{sinon.} \end{cases} .$$

Cette fonction joue un rôle fondamentale dans différents domaines des mathématiques : théorie de l'information, théorie des grandes déviations, inégalités fonctionnelles (Inégalités Sobolev-Logarithmiques, Inégalités de transport), concentration de la mesure, calibration de modèles...

Après avoir passé en revue dans la section II.2 quelques propriétés de bases de l'entropie relative et notamment l'importante formule de décomposition (II.4), nous aborderons l'aspect métrique de la distance de Kullback, avec l'inégalité de Pinsker (II.13) et son extension récente (II.16) qui comparent la convergence au sens de la norme en variation à la convergence en entropie.

La section II.3 est consacrée au théorème de Sanov, qui affirme que pour diverses topologie, $\mathrm{H}\left(\,.\,|\,\mu\right)$ contrôle les grandes déviations de la mesure empirique

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

d'une suite de variables $(X_i)_i$ indépendantes et identiquement distribuées de loi μ. Grâce à ce théorème, pour un ensemble A donné, les points $\mu^* \in A$ tels que

$$\mathrm{H}\left(\mu^*|\,\mu\right) = \inf\{\mathrm{H}\left(\nu|\,\mu\right),\ \nu \in A\},$$

apparaissent comme les scénarios les plus probables de la grande déviation $L_n \in A$. Lorsque A est convexe, il existe au plus un tel μ^* qui s'appelle *projection entropique* de μ sur A.

La section II.4 présente différents résultats, que l'on doit principalement à I. Csiszár, concernant les projections entropiques, également appelées I-projections ou projections de Csiszár. La projection en entropie jouit notamment d'une propriété rappelant l'inégalité de Pythagore de la projection euclidienne (voir (II.26)). Dans le théorème II.41, nous verrons que, sous certaines hypothèses, on dispose d'une formule explicite pour la projection entropique sur un convexe défini par une contrainte de type moment. Comme nous utiliserons ce théorème à de multiples reprises, nous en donnerons une preuve complète reposant sur des résultats élémentaires d'analyse convexe.

II.2 Entropie relative

II.2.1 Définition et premières propriétés

Dans ce chapitre, $(\mathcal{X}, \mathcal{B})$ est un espace mesurable, $\mathcal{M}(\mathcal{X})$ désigne l'ensemble des mesures finies sur $(\mathcal{X}, \mathcal{B})$, et $\mathcal{P}(\mathcal{X})$ celui des mesures de probabilité sur $(\mathcal{X}, \mathcal{B})$.

Définition II.1. Soient $\nu, \mu \in \mathcal{P}(\mathcal{X})$. L'entropie relative de ν par rapport à μ, notée $\mathrm{H}\left(\nu|\,\mu\right)$ est définie par

$$\mathrm{H}\left(\nu|\,\mu\right) = \begin{cases} \int_{\mathcal{X}} \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{si } \nu \ll \mu \\ +\infty & \text{sinon.} \end{cases}$$

Proposition II.2. *Pour toute $\mu \in \mathcal{P}(\mathcal{X})$, $\mathrm{H}\left(\,.\,|\,\mu\right)$ est une fonction convexe positive, ne s'annulant qu'en μ et strictement convexe sur $\{\mathrm{H}\left(\,.\,|\,\mu\right) < +\infty\}$.*

Nous conviendrons d'appeler la formule (II.4) de la proposition suivante *Formule de décomposition de l'entropie relative* :

Proposition II.3. *Soient $\mu \in \mathcal{P}(\mathcal{X})$ et $\nu \in \mathcal{P}(\mathcal{X}^n)$, $n \in \mathbb{N}^*$. On a, en désignant par ν_i la $i^{\text{ème}}$ marginale de ν,*

$$\mathrm{H}\left(\nu\,\middle|\,\mu^{\otimes n}\right) = \mathrm{H}\left(\nu\,\middle|\,\nu_1 \otimes \cdots \otimes \nu_n\right) + \sum_{i=1}^{n} \mathrm{H}\left(\nu_i\,\middle|\,\mu\right) \tag{II.4}$$

Démonstration. Voir, par exemple, la preuve du lemme 7.3.25 de [26]. □

II.2.2 Entropie relative et norme en variation

Norme en variation

On désignera par $B(\mathcal{X})$, l'ensemble des fonctions mesurables bornées sur $(\mathcal{X}, \mathcal{B})$. $B(\mathcal{X})$ sera muni de la norme $\|\,.\,\|_\infty$,

$$\|f\|_\infty = \sup_{x \in \mathcal{X}} |f(x)|$$

Définition II.5. Pour toute $\nu \in \mathcal{M}(\mathcal{X})$, la norme en variation de ν, notée $\|\nu\|_{VT}$ est définie par :

$$\|\nu\|_{VT} = \sup\left\{\int_{\mathcal{X}} f\, d\nu : f \in B(\mathcal{X}), \|f\|_\infty \leq 1\right\}. \tag{II.6}$$

Remarque II.7.

Clairement $\mathcal{M}(\mathcal{X})$ est inclus dans le dual topologique de $B(\mathcal{X})$; d'après la formule (II.6), la norme en variation de ν n'est autre que sa norme en tant que forme linéaire continue sur $B(\mathcal{X})$.

On dispose d'autres formules pour la norme en variation :

Proposition II.8.

1. *Si α est une mesure positive finie, et $\nu \in \mathcal{M}(\mathcal{X})$ est absolument continue par rapport à α, alors*
$$\|\nu\|_{VT} = \int_{\mathcal{X}} \left|\frac{d\nu}{d\alpha}\right| d\alpha \tag{II.9}$$

2. *Si $\mu, \nu \in \mathcal{P}(\mathcal{X})$,*
$$\|\nu - \mu\|_{VT} = \frac{1}{2} \sup_{A \in \mathcal{B}} \{\nu(A) - \mu(A)\} \tag{II.10}$$

Inégalité de Pinsker

L'application $(\nu, \mu) \mapsto \mathrm{H}\left(\nu|\, \mu\right)$ n'est pas une distance, néanmoins on peut lui associer une notion de convergence :

Définition II.11. On dit qu'une suite $(\nu_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{P}(X)$ converge en entropie vers $\mu \in \mathcal{P}(\mathcal{X})$ si, et seulement si, $\lim\limits_{n\to+\infty} \mathrm{H}\left(\nu_n|\, \mu\right) = 0$.

La convergence en entropie est une convergence en un sens assez fort, comme le montrent les propositions suivantes.
Commençons par la célèbre inégalité de Pinsker :

Proposition II.12 (Pinsker, [55])**.** *Pour toutes $\nu, \mu \in \mathcal{P}(\mathcal{X})$,*

$$\|\nu - \mu\|_{VT} \leq \sqrt{2\,\mathrm{H}\left(\nu|\, \mu\right)} \tag{II.13}$$

En particulier, si ν_n converge en entropie vers μ, alors $\|\nu_n - \mu\|_{VT} \xrightarrow[n\to+\infty]{} 0$.

On peut aller plus loin grâce à la proposition

Proposition II.14. *Si $(\nu_n)_{n\in\mathbb{N}}$ converge en entropie vers μ, alors pour toute fonction mesurable $f : \mathcal{X} \to \mathbb{R}$ telle que $\int_{\mathcal{X}} e^{t|f|}\, d\mu < +\infty$ pour un certain $t > 0$, on a : $\lim\limits_{n\to+\infty} \int_{\mathcal{X}} f\, d\nu_n = \int_{\mathcal{X}} f\, d\mu$.*

Démonstration. Voir, par exemple, la preuve du lemme 3.1 de [18]. □

Pour finir, citons un résultat récent de F. Bolley et C. Villani qui propose une version pondérée de l'inégalité de Pinsker :

Proposition II.15 (Bolley-Villani, [5] thm 1)**.** *Soit $\chi : \mathcal{X} \to \mathbb{R}^+$ une fonction mesurable. Il existe une constante numérique $C > 0$ indépendante de χ telle que pour toute $\mu \in \mathcal{P}(\mathcal{X})$, on ait :*

$$\begin{aligned}&\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \forall \delta > 0,\\ &\qquad \|\chi\nu - \chi\mu\|_{VT} \leq \frac{C}{\delta}\left(1 + \log \int_{\mathcal{X}} e^{\delta f}\, d\mu\right)\left(\sqrt{\mathrm{H}\left(\nu|\,\mu\right)} + \mathrm{H}\left(\nu|\,\mu\right)\right). \end{aligned} \tag{II.16}$$

Remarque II.17.

Nous utiliserons (II.13) et (II.16) dans le chapitre suivant consacré au **Principe Conditionnel de Gibbs**, et nous reviendrons sur ces inégalités dans la seconde partie de cette thèse consacrée aux **Inégalités de Transport**. Nous y verrons en particulier une autre preuve de (II.16). A titre documentaire, nous incluons ci-dessous une preuve classique de (II.13).

Démonstration de la proposition II.12.
Si $\mathrm{H}(\nu|\,\mu) = +\infty$, l'inégalité est vraie.
Supposons donc que $\mathrm{H}(\nu|\,\mu) < +\infty$ et notons $h = \frac{d\nu}{d\mu}$. D'après (II.9),

$$\|\nu - \mu\|_{VT} = \int_{\mathcal{X}} |h - 1|\, d\mu$$

Or, pour tout $x > 0$,

$$3(x-1)^2 \leq (4 + 2x)(x\log(x) - x + 1). \tag{II.18}$$

Donc

$$|h - 1| \leq \frac{1}{\sqrt{3}}\sqrt{4 + 2h}\sqrt{h\log h - h + 1}.$$

Donc, d'après l'inégalité de Cauchy-Schwarz,

$$\begin{aligned}\|\nu - \mu\|_{VT} &\leq \frac{1}{\sqrt{3}}\left[\int_{\mathcal{X}} 4 + 2h\, d\mu\right]^{\frac{1}{2}}\left[\int_{\mathcal{X}} h\log h - h + 1\, d\mu\right]^{\frac{1}{2}}\\ &= \sqrt{2\,\mathrm{H}\left(\nu|\,\mu\right)}.\end{aligned}$$

□

II.3 Le théorème de Sanov

II.3.1 La version classique

Le théorème suivant donne la version la plus classique du théorème de Sanov. Ici, $\mathcal{X}$ est un espace polonais, l'ensemble $\mathcal{P}(\mathcal{X})$ des probabilités sur $\mathcal{X}$ est muni de la topologie de la convergence étroite, *ie* la moins fine rendant continues les applications

$$\mathcal{P}(\mathcal{X}) \to \mathbb{R} : \nu \mapsto \int_{\mathcal{X}} g\, d\nu, \qquad g \in C_b(\mathcal{X}),$$

$C_b(\mathcal{X})$ étant l'ensemble des applications continues bornées sur $\mathcal{X}$. On munit $\mathcal{P}(\mathcal{X})$ de sa tribu borélienne.

Théorème II.19. *Si $(X_i)_{i\in\mathbb{N}^*}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi μ, alors la suite $L_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ suit un principe de grandes déviations sur $\mathcal{P}(\mathcal{X})$, muni de la topologie de la convergence étroite et de sa tribu borélienne, de bonne fonction de taux $\mathrm{H}\left(\,.\, | \,\mu\right)$. Autrement dit, pour tout ensemble A mesurable, on a*

$$-\inf\left\{\mathrm{H}\left(\nu|\,\mu\right), \nu \in \overset{\circ}{A}\right\} \leq \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(L_n \in A)$$

et

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(L_n \in A) \leq -\inf\left\{\mathrm{H}\left(\nu|\,\mu\right), \nu \in \overline{A}\right\}.$$

II.3.2 Extensions du théorème de Sanov

Le théorème II.19 a été généralisé par différents auteurs pour des topologies plus fortes que la topologie de la convergence étroite.

Cadre : Nous nous donnerons une classe G, d'applications mesurables sur $\mathcal{X}$ et à valeurs réelles et nous poserons

$$\mathcal{P}_G(\mathcal{X}) = \left\{\nu \in \mathcal{P}(\mathcal{X}) : \forall g \in G, \quad \int_{\mathcal{X}} |g|\, d\nu < +\infty\right\}.$$

Nous munirons $\mathcal{P}_G(\mathcal{X})$ de

- **la G-topologie**, *ie* la moins fine rendant continues les applications

$$\mathcal{P}_G(\mathcal{X}) \to \mathbb{R} : \nu \mapsto \int_{\mathcal{X}} g\, d\nu, \qquad g \in G$$

- **la G-tribu**, *ie* la tribu engendrée par ces mêmes applications.

Nous supposerons que G contient $B(\mathcal{X})$, l'ensemble des applications mesurables bornées sur $\mathcal{X}$. Sous cette hypothèse, on voit facilement que $\mathcal{P}_G(\mathcal{X})$ est séparé.

Nous dirons que $\mu \in \mathcal{P}_G(\mathcal{X})$ vérifie l'*hypothèse de Cramér forte*, si

$$\forall g \in G, \quad \forall t > 0, \quad \int_{\mathcal{X}} e^{t|g|}\, d\mu < +\infty. \tag{II.20}$$

La version suivante du théorème de Sanov est due à P. Eichelsbacher et U. Schmock.

Théorème II.21 (Eichelsbacher-Schmock, [30], thm. 1.7)**.** *Si μ vérifie l'hypothèse de Cramér forte, alors pour toute suite $(X_i)_{i\in\mathbb{N}^*}$ de variables aléatoires i.i.d de loi μ, la suite $L_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ suit un principe de grandes déviations sur $\mathcal{P}_G(\mathcal{X})$, muni de la G-topologie et de la G-tribu, de bonne fonction de taux $\mathrm{H}\left(\,.\,|\,\mu\right)$.*

Remarque II.22.

- D'après le point 1 de la proposition II.34, sous l'hypothèse (II.20), $\mathrm{H}\left(\nu|\,\mu\right) < +\infty \Rightarrow \nu \in \mathcal{P}_G(\mathcal{X})$.
- Le théorème II.21 n'est pas la dernière généralisation du théorème de Sanov : dans [46], C. Léonard et J. Najim ont montré comment on pouvait s'affranchir de l'hypothèse de Cramér forte.

II.4 Projections entropiques

II.4.1 Définition et relation de Pythagore

Notation : Pour toute partie A de $\mathcal{P}(X)$, nous noterons :

$$\mathrm{H}\left(A|\,\mu\right) := \inf\{\mathrm{H}\left(\nu|\,\mu\right) : \nu \in A\} \in \mathbb{R}^+ \cup \{+\infty\}$$

Définition II.23. Soient $\mu \in \mathcal{P}(\mathcal{X})$ et C un convexe de $\mathcal{P}(\mathcal{X})$ tel que $\mathrm{H}\left(C|\,\mu\right) < +\infty$. On appelle *I-projection* ou *projection entropique* de μ sur C tout élément $\nu \in C$ tel que :

$$\mathrm{H}\left(\nu|\,\mu\right) = \mathrm{H}\left(C|\,\mu\right)$$

Remarque II.24.

- La fonction $\mathrm{H}\left(\,.\,|\,\mu\right)$ étant strictement convexe sur $\{\mathrm{H}\left(\,.\,|\,\mu\right) < +\infty\}$, une mesure de probabilité μ admet au plus une I-projection sur C. Nous noterons, en général, μ^* cette I-projection.

- Le théorème de Sanov permet d'interpréter cette notion de I-projection : en écrivant schématiquement que pour tout A mesurable,

$$\mathbb{P}(L_n \in A) \approx e^{-n\,\mathrm{H}(A|\mu)},$$

on voit que pour un ensemble convexe C,

$$\mathbb{P}(L_n \in C) \approx \mathbb{P}(L_n \simeq \mu^*).$$

La I-projection μ^* de μ sur C apparaît donc comme le scénario le plus probable de la grande déviation $L_n \in C$. Nous verrons, au chapitre suivant, une autre interprétation des I-projections grâce au Principe Conditionnel de Gibbs.

Le théorème suivant, que l'on doit à I. Csiszár, établit une sorte de relation de Pythagore pour les I-projections :

Théorème II.25 (Csiszár, [18], thm. 2.2)**.** *Soient $\mu \in \mathcal{P}(\mathcal{X})$ et C un ensemble convexe de $\mathcal{P}(\mathcal{X})$ tel que $\mathrm{H}\left(C|\,\mu\right) < +\infty$. Si μ possède une I-projection μ^* sur C, alors*

$$\forall \nu \in C, \quad \mathrm{H}\left(\nu|\,\mu\right) \geq \mathrm{H}\left(\nu|\,\mu^*\right) + \mathrm{H}\left(\mu^*|\,\mu\right). \tag{II.26}$$

II.4.2 Projections entropiques généralisées

Théorème II.27 (Csiszár, [18], thm. 2.1)**.** *Soient $\mu \in \mathcal{P}(\mathcal{X})$ et C un ensemble convexe de $\mathcal{P}(\mathcal{X})$ tel que $\mathrm{H}\left(C|\,\mu\right) < +\infty$. Il existe une unique probabilité μ^* appartenant à l'adhérence de C pour la norme en variation vers laquelle converge en variation toute suite $(\nu_n)_{n\in\mathbb{N}}$ d'éléments de C telle que $\lim_{n\to+\infty} \mathrm{H}\left(\nu_n|\,\mu\right) = \mathrm{H}\left(C|\,\mu\right)$.*

Définition II.28. On appelle la probabilité μ^* du théorème précédent la *I-projection généralisée*, ou la *projection entropique généralisée* de μ sur C.

Remarque II.29.

- En général, si μ^* est la I-projection généralisée de μ sur C, l'inégalité

$$\mathrm{H}\left(\mu^*|\,\mu\right) \leq \mathrm{H}\left(C|\,\mu\right)$$

peut être stricte.
- Il résulte du théorème II.27 que μ possède une I-projection sur tout ensemble convexe C fermé pour la norme en variation tel que $\mathrm{H}\left(C|\,\mu\right) < +\infty$. Nous verrons, dans la section suivante, d'autres critères topologiques garantissant l'existence d'une I-projection.

La proposition suivante caractérise les I-projections généralisées par une relation de Pythagore :

Proposition II.30 (Topsoe, [70], thm. 8)**.** *Soient $\mu \in \mathcal{P}(\mathcal{X})$ et C un ensemble convexe de $\mathcal{P}(\mathcal{X})$ tel que $\mathrm{H}\left(C|\, \mu\right) < +\infty$. Une mesure de probabilité $\alpha \in C$ est la I-projection généralisée de μ sur C si, et seulement si,*

$$\forall \nu \in C, \quad \mathrm{H}\left(\nu|\, \mu\right) \geq \mathrm{H}\left(\nu|\, \alpha\right) + \mathrm{H}\left(C|\, \mu\right). \tag{II.31}$$

II.4.3 Critères d'existence d'une projection entropique

Nous avons vu au théorème II.27 précédent qu'une condition suffisante pour qu'une mesure admette une I-projection sur un ensemble convexe C était la fermeture de C pour la norme en variation. Nous allons présenter dans cette section des critères pour d'autres topologies.

Plaçons nous dans le cadre de la section II.3.2 :

Nous dirons que $\mu \in \mathcal{P}_G(\mathcal{X})$ vérifie l'*hypothèse de Cramér faible*, si

$$\forall g \in G, \quad \exists t > 0, \quad \int_{\mathcal{X}} e^{t|g|}\, d\mu < +\infty. \tag{II.32}$$

Rappelons que $\mu \in \mathcal{P}_G(\mathcal{X})$ vérifie l'*hypothèse de Cramér forte*, si

$$\forall g \in G, \quad \forall t > 0, \quad \int_{\mathcal{X}} e^{t|g|}\, d\mu < +\infty, \tag{II.33}$$

La proposition suivante est due à P. Eichelsbacher et U. Schmock :

Proposition II.34 (Eichelsbacher-Schmock, [30], thm. 1.7)**.**

1. *Si $\mu \in \mathcal{P}_G(\mathcal{X})$ vérifie l'hypothèse de Cramér faible, alors pour tout $a \geq 0$,*

   $$\{\nu \in \mathcal{P}(\mathcal{X}) :\ \mathrm{H}\left(\nu|\, \mu\right) \leq a\}$$

   est inclus dans $\mathcal{P}_G(\mathcal{X})$

2. *Si $\mu \in \mathcal{P}_G(\mathcal{X})$ vérifie l'hypothèse de Cramér forte, alors pour tout $a \geq 0$,*

   $$\{\nu \in \mathcal{P}(\mathcal{X}) :\ \mathrm{H}\left(\nu|\, \mu\right) \leq a\}$$

   est de plus compact et séquentiellement compact pour la G-topologie.

On en déduit les corollaires

Corollaire II.35. *Si μ vérifie l'hypothèse de Cramér faible (II.32) et si C est un convexe de $\mathcal{P}(\mathcal{X})$ tel que* $\mathrm{H}(C|\,\mu) < +\infty$, *alors C et $C_G := C \cap \mathcal{P}_G(\mathcal{X})$ ont la même projection généralisée.*

Démonstration.
Tout d'abord, grâce au point (1) de la proposition II.34, $\mathrm{H}(C|\,\mu) = \mathrm{H}(C_G|\,\mu)$. Ensuite, si ν_n est une suite d'éléments de C telle que $\mathrm{H}(\nu_n|\,\mu) \xrightarrow[n\to+\infty]{} \mathrm{H}(C|\,\mu)$, alors c'est également une suite d'éléments de C_G telle que $\mathrm{H}(\nu_n|\,\mu) \xrightarrow[n\to+\infty]{} \mathrm{H}(C_G|\,\mu)$. On en déduit, grâce à la proposition II.27, que C et C_G ont la même projection généralisée. □

Corollaire II.36.
Si μ vérifie l'hypothèse Cramér forte (II.20), alors μ possède une I-projection sur tout ensemble convexe $C \subset \mathcal{P}_G(\mathcal{X})$ fermé pour la G-topologie tel que $\mathrm{H}(C|\,\mu) < +\infty$.

Démonstration.
Soit $(\nu_n)_{n\in\mathbb{N}}$ une suite d'éléments de C telle que $\mathrm{H}(\nu_n|\,\nu) \xrightarrow[n\to+\infty]{} \mathrm{H}(C|\,\mu)$. Si M est un majorant de $\mathrm{H}(\nu_n|\,\mu)$, alors pour tout $n \in \mathbb{N}$, $\nu_n \in C \cap \{\mathrm{H}(\,.\,|\,\mu) \leq M\}$, et ce dernier ensemble est séquentiellement compact pour la G-topologie. Par conséquent, on peut extraire de ν_n une sous-suite convergeant vers un certain $\nu \in C$. Comme pour tout $\varepsilon > 0$, $\nu_n \in \{\mathrm{H}(\,.\,|\,\mu) \leq \mathrm{H}(C|\,\mu) + \varepsilon\}$ pour tout n assez grand, on en déduit que $\mathrm{H}(\nu|\,\mu) \leq \mathrm{H}(C|\,\mu) + \varepsilon$; ceci étant vrai pour tout ε, on a $\mathrm{H}(\nu|\,\mu) \leq \mathrm{H}(C|\,\mu)$, et par conséquent ν est la I-projection de μ sur C. □

II.4.4 Représentation des projections entropiques

Dans cette sous-section, nous allons donner l'expression de la I-projection (généralisée) μ^* d'une probabilité μ sur un ensemble convexe C défini par une contrainte de type moment, *ie* de la forme

$$C = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} F(x)\, d\nu \in K \right\}$$

où $F : \mathcal{X} \to B$ est une application à valeurs dans un espace de Banach B et C est un convexe fermé de B.

Le théorème II.41 est dû à I. Csiszár (voir [18] thm. 3.3 et [19] thm. 2 et 3). La preuve que nous proposons de ce résultat est différente de la preuve de Csiszár et repose sur quelques notions élémentaires d'analyse convexe (théorème de Fenchel, sous-différentiabilité, etc.). On pourra consulter les articles [43, 45, 44] de C. Léonard pour des résultats très généraux concernant la représentation des I-projections (et autres minimisants de fonctionnelles d'énergie).

Cadre et notations

- $(B, \|\,.\,\|)$ sera un espace de Banach séparable, muni de sa tribu borélienne. Le dual topologique de B, B' sera muni de la topologie forte.
- $F : \mathcal{X} \to B$ sera une application mesurable.
- Nous désignerons par μ_F, l'image de μ par l'application F. La transformée de Laplace de μ_F sera notée Z_F, elle est définie par :

$$\forall \lambda \in B', \quad Z_F(\lambda) = \int_{\mathcal{X}} e^{\langle \lambda, F \rangle} d\mu,$$

On désignera par Λ_F la Log-Laplace de μ_F définie par $\Lambda_F := \log Z_F$ et par Λ_F^*, la transformée de Cramér de μ_F, qui vaut par définition :

$$\Lambda_F^*(x) = \sup_{\lambda \in B'} \{\langle \lambda, x \rangle - \Lambda_F(\lambda)\}$$

- K sera un convexe fermé de B et nous poserons

$$C = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \quad \int_{\mathcal{X}} \|F\| \, d\nu < +\infty \quad \text{et} \quad \int_{\mathcal{X}} F \, d\nu \in K \right\},$$

où $\int_{\mathcal{X}} F \, d\nu$ est l'intégrale au sens de Böchner.

Nous ferons l'hypothèse suivante :

Hypothèse II.37.

1. Il existe $t > 0$ tel que $\displaystyle\int_{\mathcal{X}} e^{t\|F\|} \, d\mu < +\infty$,
2. Le domaine de λ_F, défini par $\operatorname{dom} \Lambda_F := \{\lambda \in B', \; \Lambda_F(\lambda) < +\infty\}$, est ouvert dans B'.

Remarque II.38.

- Sous l'hypothèse (II.37), on voit facilement que Λ_F est Gâteaux-différentiable sur dom Λ_F et que

$$\forall \lambda \in B', \quad \nabla \Lambda_F(\lambda) = \frac{1}{Z_F(\lambda)} \int_{\mathcal{X}} F e^{\langle \lambda, F \rangle} d\mu$$

- Si pour tout $t > 0$, $\int_{\mathcal{X}} e^{t\|F\|} \, d\mu < +\infty$, on sait d'après le corollaire II.36 (en prenant $G = B(\mathcal{X}) \cup \{\|F\|\}$), que μ admet une I-projection sur C (qui est fermé pour la G-topologie), à condition bien sûr que $\mathrm{H}\,(C|\,\mu) < +\infty$.

Nous aurons besoin du lemme suivant :

Lemme II.39. *Sous l'hypothèse II.37, si la fonction*

$$H(\lambda) = \Lambda_F(\lambda) - \inf_{y\in K}\langle \lambda, y\rangle$$

atteint son minimum, alors $\mathrm{H}\left(C|\,\mu\right) = \sup_{\lambda\in B'}\left\{\inf_{y\in K}\langle\lambda,y\rangle - \Lambda_F(\lambda)\right\}$ *et* μ *admet une I-projection* μ^* *sur* C *qui s'écrit :*

$$\mu^* = \frac{e^{\langle\lambda^*,F\rangle}}{Z_F(\lambda^*)}\mu,$$

pour tout λ^* *minimisant* H.

Démonstration. On pourra consulter les livres [38], [58] et [59] pour une définition de la notion de sous-différentiabilité utilisée ci-dessous. Soit λ^* un minimisant de H. Posons $\sigma_K(\lambda) = -\inf_{y\in K}\langle\lambda,y\rangle$. Pour tout $\lambda\in B'$, et tout $t>0$, on a :

$$\frac{\sigma_K(\lambda^*+t\lambda)-\sigma_K(\lambda^*)}{t} \geq -\frac{\Lambda_F(\lambda^*+t\lambda)-\Lambda_F(\lambda^*)}{t}. \tag{II.40}$$

La fonction Λ_F étant Gâteaux-différentiable sur son domaine, le second membre de (II.40) a pour limite $-\langle\lambda,\nabla\Lambda_F(\lambda^*)\rangle$ quand $t\to 0^+$. On en déduit, en notant $\sigma_K'(\lambda^*;\lambda)$, la dérivée directionnelle de σ_K selon le vecteur λ, que

$$\forall\lambda\in B',\quad \sigma_K'(\lambda^*;\lambda)\geq\langle\lambda,-\nabla\Lambda_F(\lambda^*)\rangle.$$

Autrement dit, $-\nabla\Lambda_F(\lambda^*)\in\partial\sigma_K(\lambda^*)$ (le sous-différentiel de σ_K en λ^*). Or σ_K n'est autre que la fonction de support de $-K$, et d'après [59] p. 35-36,

$$\partial\sigma_K(\lambda^*) = \left\{z\in -K,\ \langle\lambda^*,z\rangle = -\inf_{y\in K}\langle\lambda^*,y\rangle\right\}$$

Par conséquent,

$$\nabla\Lambda_F(\lambda^*)\in K \qquad \text{et} \qquad \langle\lambda^*,\nabla\Lambda_F(\lambda^*)\rangle = \inf_{y\in K}\langle\lambda^*,y\rangle.$$

Posons $\mu^* = \frac{e^{\langle\lambda^*,F\rangle}}{Z_F(\lambda^*)}\mu$, alors $\int_{\mathcal{X}} F\,d\mu^* = \nabla\Lambda_F(\lambda^*)\in K$ et donc $\mu^*\in C$.
De plus, pour toute $\nu\in C$, on a :

$$\begin{aligned}
\mathrm{H}\left(\nu|\,\mu\right) &= \int_{\mathcal{X}}\log\left(\frac{d\nu}{d\mu^*}\right)d\nu + \int_{\mathcal{X}}\log\left(\frac{d\mu^*}{d\mu}\right)d\nu\\
&= \mathrm{H}\left(\nu|\,\mu^*\right) + \left\langle\lambda^*,\int_{\mathcal{X}}F\,d\nu\right\rangle - \Lambda_F(\lambda^*)\\
&= \mathrm{H}\left(\nu|\,\mu^*\right) + \mathrm{H}\left(\mu^*|\,\mu\right) + \left\langle\lambda^*,\int_{\mathcal{X}}F\,d\nu\right\rangle - \left\langle\lambda^*,\int_{\mathcal{X}}F\,d\mu^*\right\rangle
\end{aligned}$$

Or, comme $\nu \in C$, on a

$$\left\langle \lambda^*, \int_{\mathcal{X}} F\, d\nu \right\rangle \geq \inf_{y\in K} \langle \lambda^*, y\rangle = \left\langle \lambda^*, \int_{\mathcal{X}} F\, d\mu^* \right\rangle.$$

Donc $\mathrm{H}\,(\nu|\,\mu) \geq \mathrm{H}\,(\nu|\,\mu^*) + \mathrm{H}\,(\mu^*|\,\mu)$, et μ^* est la I-projection de μ sur C. □

Notations : Nous noterons $\mathrm{co}\, A$, l'enveloppe convexe d'un ensemble A. Rappelons qu'en dimension finie, l'*intérieur relatif* d'un ensemble convexe A, noté $\mathrm{ri}\, A$, est l'intérieur de A pour la topologie de l'espace affine engendré par A.

Théorème II.41. *Sous l'hypothèse (II.37), si l'une des deux hypothèses suivantes est réalisée*

1. *B est de dimension finie, et* $\mathrm{ri}\, K \cap \mathrm{ri}\,\mathrm{co}\, S_F \neq \emptyset$*, S_F étant le support de μ_F,*
2. *K est d'intérieur non vide et $\overset{\circ}{K} \cap \mathrm{co}\, S_F \neq \emptyset$,*

alors $\mathrm{H}\,(C|\,\mu) = \max_{\lambda\in B'} \left\{ \inf_{y\in K} \langle \lambda, y\rangle - \Lambda_F(\lambda) \right\}$ *et pour tout λ^* où le supremum est atteint,* $\mu^* = \dfrac{e^{\langle \lambda^*, F\rangle}}{Z_F(\lambda^*)}\mu$ *est la I-projection de μ sur C.*

Remarque II.42.

On a toujours (voir par exemple le lemme 2.4 de [23]) :

$$\overline{\mathrm{dom}\, \Lambda_F^*} = \overline{\mathrm{co}\, S_F}.$$

En dimension finie, on a donc $\mathrm{ri}\,\mathrm{dom}\, \Lambda_F^* = \mathrm{ri}\,\mathrm{co}\, S_F$ (voir [38] proposition 2.1.8 p. 36). L'hypothèse 1. précédente est donc équivalente à $\mathrm{ri}\, K \cap \mathrm{ri}\,\mathrm{dom}\, \Lambda_F^* \neq \emptyset$ et l'hypothèse 2. équivaut quant à elle à $\overset{\circ}{K} \cap \mathrm{dom}\, \Lambda_F^* \neq \emptyset$.

La démonstration du théorème II.41 repose sur le *théorème de dualité de Fenchel* dont voici une version simple (voir [38] (2.3.2) p. 228 pour le point 1, et [9] thm. I.11 pour le point 2) :

Théorème II.43. *Soient $g_1, g_2 : B \to \mathbb{R} \cup \{+\infty\}$ deux fonctions convexes s.c.i non identiquement égales à $+\infty$ définies sur un espace vectoriel normé B.*
On a

$$\inf_{x\in B}\{g_1(x) + g_2(x)\} = \max_{\lambda\in B'}\{-g_1^*(-\lambda) - g_2^*(\lambda)\},$$

si l'une des deux hypothèses suivantes est réalisée :

1. *B est de dimension finie, et* $\mathrm{ri}\,\mathrm{dom}\, g_1 \cap \mathrm{ri}\,\mathrm{dom}\, g_2 \neq \emptyset$*,*
2. *Il existe $x_0 \in B$ tel que $g_1(x_0) < +\infty$, $g_2(x_0) < +\infty$, et g_1 est continue en x_0.*

Démonstration du théorème II.41 :
Notons $\imath_K$ l'indicatrice de K, définie par $\imath_K(x) = \begin{cases} 0 & \text{si } x \in K \\ +\infty & \text{sinon} \end{cases}$.
D'une part $\imath_K^*(\lambda) = \sup_{y\in K}\langle\lambda, y\rangle$, et d'autre part $(\Lambda_F^*)^* = \Lambda_F$ (voir, par exemple, [9] thm. I.10).
D'après la remarque II.42, sous l'hypothèse 1, on a $\operatorname{ri}\operatorname{dom}\imath_K \cap \operatorname{ri}\operatorname{dom}\Lambda_F^* \neq \emptyset$, et sous l'hypothèse 2, il existe $x_0 \in \overset{\circ}{K}$ tel que $\Lambda_F^*(x_0) < +\infty$ et $\imath_K$ est continue en x_0. Donc, d'après le théorème II.43, on a

$$\inf_{x\in K}\Lambda^*(x) = \inf_{x\in B}\{\Lambda^*(x) + \imath_K(x)\} = \max_{\lambda\in B'}\left\{\inf_{y\in K}\langle\lambda, y\rangle - \Lambda_F(\lambda)\right\},$$

En particulier, la fonction $H(\lambda) = \Lambda_F(\lambda) - \inf_{y\in K}\langle\lambda, y\rangle$ atteint son minimum.
On conclut grâce au lemme II.39. $\square$

Le théorème précédent n'est plus valable si l'hypothèse $\operatorname{ri} K \cap \operatorname{ri}\operatorname{co} S_F$ (resp. $\overset{\circ}{K} \cap \operatorname{co} S_F$) n'est pas satisfaite. En effet, considérons la probabilité $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 \in \mathcal{P}(\mathbb{R})$ et le convexe

$$C = \left\{\nu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} x\,d\nu \geq 1\right\}.$$

Clairement, $\operatorname{dom}\Lambda_\mu = \mathbb{R}$ est ouvert, mais $]1, +\infty[\cap[0,1] = \emptyset$.
Calculons la I-projection de μ sur C. Tout d'abord, $\delta_1 \in C$, et $\mathrm{H}\left(\delta_1 |\, \mu\right) = \frac{\log(2)}{2}$. De plus, $\nu \ll \mu \Leftrightarrow \exists\alpha \in [0,1], \nu = (1-\alpha)\delta_0 + \alpha\delta_1$. Comme

$$\int_{\mathbb{R}} x\,d\left((1-\alpha)\delta_0 + \alpha\delta_1\right) = \alpha \geq 1 \Leftrightarrow \alpha = 1,$$

on en déduit que δ_1 est la I-projection de μ sur C. Clairement δ_1 n'est pas de la forme $\dfrac{e^{sx}}{Z_\mu(s)}d\mu(x)$.

Pour conclure ce chapitre, nous allons montrer que le théorème II.41 est également mis en défaut si le domaine de Λ_μ n'est pas ouvert.

Proposition II.44. *Soit* $\mu \in \mathcal{P}(\mathbb{R})$ *telle que* $\operatorname{supp}\mu = \mathbb{R}^+$ *et* $\operatorname{dom}\Lambda_\mu =]-\infty, 1]$. *Posons* $d\mu^*(x) = \dfrac{e^x}{\Lambda_\mu(1)}\,d\mu(x)$ *et* $\alpha = \int_{\mathbb{R}} x\,d\mu^*$. *Pour tout* $a \geq \alpha$, μ^* *est la I-projection généralisée de* μ *sur le convexe* C_a *défini par*

$$C_a = \left\{\nu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |x|\,d\nu < +\infty \quad et \quad \int_{\mathbb{R}} x\,d\nu \geq a\right\}.$$

De plus, pour tout $a \geq \alpha$, *on a* $\mathrm{H}\left(C_a |\, \mu\right) = a - \Lambda_\mu(1)$.

Avant de passer à la preuve, commençons par quelques remarques :

Remarque II.45.

- La proposition précédente s'applique par exemple pour des probabilités μ de la forme $d\mu(x) = \dfrac{C}{(1+x)^b} e^{-x} \mathbb{I}_{\mathbb{R}^+}\, dx$, avec $b > 1$.
- Si $a > \alpha$, alors bien que $]a, +\infty[$ soit d'intersection non vide avec l'intérieur de l'enveloppe convexe du support de μ, la probabilité μ n'admet pas de I-projection sur C_a ($\mu^* \notin C_a$). Ceci prouve que le théorème II.41 n'est plus valable si $\operatorname{dom} \Lambda_\mu$ n'est pas ouvert.
- On a vu que pour tout $a \geq \alpha$, $\mathrm{H}\left(C_a \middle| \mu\right) = a - \Lambda_\mu(1)$. En particulier, si $a > \alpha$ on a
$$\mathrm{H}\left(C_a \middle| \mu\right) > \mathrm{H}\left(\mu^* \middle| \mu\right).$$
- Si $\alpha \leq a_1 < a_2$, alors $C_{a_2} \subset C_{a_1}$. Les ensembles C_{a_1} et C_{a_2} ont la même projection entropique généralisée μ^*. Pourtant $\mathrm{H}\left(C_{a_1} \middle| \mu\right) < \mathrm{H}\left(C_{a_2} \middle| \mu\right)$.

Démonstration. Soit $a \geq \alpha$; pour tout $n \geq 1$, posons $d\mu_n = \dfrac{\mathbb{I}_{[0,n]}}{\mu[0,n]}\, d\mu$.

Première étape :
Nous allons montrer que pour tout $t \in \mathbb{R}^+$, la suite $\left(\Lambda'_{\mu_n}(t)\right)_{n\geq 1}$ est croissante.
En effet, pour tout $t \geq 0$ fixé, on peut écrire $\Lambda'_{\mu_n}(t) = \varphi(n)$, où la fonction $\varphi :]0, +\infty[\to \mathbb{R}$ est définie par
$$\varphi(u) = \frac{\int_0^u x e^{tx}\, d\mu(x)}{\int_0^u e^{tx}\, d\mu(x)}.$$
Or,
$$\varphi'(u) = \frac{u e^{tu} \int_0^u e^{tx}\, d\mu(x) - e^{tu} \int_0^u x e^{tx}\, d\mu(x)}{\left(\int_0^u e^{tx}\, d\mu(x)\right)^2} \geq \frac{e^{tu} \int_0^u x e^{tx}\, d\mu(x) - e^{tu} \int_0^u x e^{tx}\, d\mu(x)}{\left(\int_0^u e^{tx}\, d\mu(x)\right)^2} = 0$$
Ainsi, φ est croissante, et par conséquent, $\left(\Lambda'_{\mu_n}(t)\right)_{n\geq 1}$ est aussi croissante.

Deuxième étape : Montrons que pour tout $n \geq 1$, $\Lambda'_{\mu_n}(t) \xrightarrow[t\to+\infty]{} n$.
Pour tout $\varepsilon > 0$, on a :
$$\begin{aligned}\left|\Lambda'_{\mu_n}(t) - n\right| &\leq \frac{\int_0^n |x-n| e^{tx}\, d\mu(x)}{\int_0^n e^{tx}\, d\mu(x)} \leq \varepsilon + \frac{\int_0^{n-\varepsilon} |x-n| e^{tx}\, d\mu(x)}{\int_0^n e^{tx}\, d\mu(x)} \\ &\leq \varepsilon + \frac{(n-\varepsilon) e^{t(n-\varepsilon)} \mu[0, n-\varepsilon]}{e^{t(n-\varepsilon/2)} \mu[n-\varepsilon/2, n]} = \varepsilon + (n-\varepsilon) e^{-\varepsilon t/2} \frac{\mu[0, n-\varepsilon]}{\mu[n-\varepsilon/2, n]}.\end{aligned}$$
Ainsi, pour tout $\varepsilon > 0$,
$$\limsup_{t\to+\infty} \left|\Lambda'_{\mu_n}(t) - n\right| \leq \varepsilon.$$

Comme ε est arbitraire, on en déduit que $\Lambda'_{\mu_n}(t) \xrightarrow[t\to+\infty]{} n$.

Troisième étape : Montrons qu'il existe une suite décroissante $t_n \geq 1$ définie pour tout $n \geq [a]+1$ telle que $\Lambda'_{\mu_n}(t_n) = a$.
Procédons par récurrence sur $n \geq [a]+1$:

- Pour $n = n_0 := [a]+1$, la suite $\Lambda'_{\mu_n}(1)$ étant croissante, on a

$$\Lambda'_{\mu_{n_0}}(1) \leq \lim_{n\to+\infty} \Lambda'_{\mu_n}(1) = \int_{\mathbb{R}} x\, d\mu^* = \alpha \leq a.$$

  D'autre part, $\lim\limits_{t\to+\infty} \Lambda'_{\mu_{n_0}}(t) = n_0 > a$. Donc, d'après le théorème des valeurs intermédiaires, il existe $1 \leq t_{n_0}$ tel que $\Lambda'_{\mu_{n_0}}(t_{n_0}) = a$.
- Supposons $1 \leq t_n$ construit. Comme précédemment, $\Lambda'_{\mu_{n+1}}(1) \leq a$. De plus, la suite $\Lambda'_{\mu_p}(t)$ étant croissante pour tout t, on a $\Lambda'_{\mu_{n+1}}(t_n) \geq \Lambda'_{\mu_n}(t_n) = a$. Donc, d'après le théorème des valeurs intermédiaires, il existe $1 \leq t_{n+1} \leq t_n$ tel que $\Lambda'_{\mu_{n+1}}(t_{n+1}) = a$.

Quatrième étape : Montrons que la suite t_n converge vers 1 et que $\mathrm{H}\left(C_a \middle| \mu\right) \leq a - \Lambda_\mu(1)$.
Posons

$$d\widetilde{\mu_n}(x) = \frac{e^{t_n x}}{Z_{\mu_n}(t_n)}\, d\mu_n.$$

Alors,

$$\begin{aligned}
\mathrm{H}\left(\widetilde{\mu_n} \middle| \mu\right) &= \int_{\mathbb{R}} \log \frac{d\widetilde{\mu_n}}{d\mu_n}\, d\widetilde{\mu_n} + \int_R \log \frac{d\mu_n}{d\mu}\, d\widetilde{\mu_n} \\
&= \int_{\mathbb{R}} (t_n x - \Lambda_{\mu_n}(t_n))\, d\widetilde{\mu_n}(x) - \log \mu[0,n] \\
&\overset{(*)}{=} t_n a - \Lambda_{\mu_n}(t_n) - \log \mu[0,n],
\end{aligned} \tag{II.46}$$

où $(*)$ vient de

$$\int_{\mathbb{R}} x\, d\widetilde{\mu_n}(x) = \Lambda'_{\mu_n}(t_n) = a. \tag{II.47}$$

L'équation (II.47) entraîne que $\widetilde{\mu_n} \in C_a$. En particulier, d'après (II.46), on a pour tout n

$$\mathrm{H}\left(C_a \middle| \mu\right) \leq t_n a - \Lambda_{\mu_n}(t_n) - \log \mu[0,n] \tag{II.48}$$

La suite t_n étant décroissante et minorée par 1, elle converge vers un certain $\ell \geq 1$. On obtient en utilisant le théorème de Fatou en (II.49), et (II.48) en (II.50) :

$$\begin{aligned}
\int_{\mathbb{R}} e^{\ell x}\, d\mu &= \int_{\mathbb{R}} \liminf_{n\to+\infty} \mathbb{1}_{[0,n]} \frac{e^{t_n x}}{\mu[0,n]}\, d\mu \\
&\leq \liminf_{n\to+\infty} \Lambda_{\mu_n}(t_n) && \text{(II.49)} \\
&\leq \liminf_{n\to+\infty} \left(t_n a - \log\mu[0,n] - \mathrm{H}\left(C_a|\,\mu\right)\right) && \text{(II.50)} \\
&= \ell a - \mathrm{H}\left(C_a|\,\mu\right).
\end{aligned}$$

On en déduit que $\ell \in \operatorname{dom}\Lambda_\mu =]-\infty, 1]$ et comme $\ell \geq 1$, on a $\ell = 1$. En passant à la limite dans (II.48) grâce au théorème de convergence dominée, on obtient

$$\mathrm{H}\left(C_a|\,\mu\right) \leq a - \Lambda_\mu(1) \tag{II.51}$$

Cinquième étape : Finalement montrons que $\mathrm{H}\left(C_a|\,\mu\right) = a - \Lambda_\mu(1)$ et que μ^* est la I-projection généralisée de μ sur C_a.
Pour toute $\nu \in C_a$, on a

$$\begin{aligned}
\mathrm{H}\left(\nu|\,\mu\right) &= \mathrm{H}\left(\nu|\,\mu^*\right) + \int_{\mathbb{R}} \log\frac{d\mu^*}{d\mu}\, d\nu \\
&= \mathrm{H}\left(\nu|\,\mu^*\right) + \int_{\mathbb{R}} x - \Lambda_\mu(1)\, d\nu \\
&\geq \mathrm{H}\left(\nu|\,\mu^*\right) + a - \Lambda_\mu(1) && \text{(II.52)} \\
&\geq \mathrm{H}\left(\nu|\,\mu^*\right) + \mathrm{H}\left(C_a|\,\mu\right). && \text{(II.53)}
\end{aligned}$$

Dans ce calcul, (II.52) résulte du fait que $\nu \in C_a$, et (II.53) vient de (II.51). De (II.52), on déduit que $\mathrm{H}\left(C_a|\,\mu\right) \geq a - \Lambda_\mu(1)$, ce qui d'après (II.51) entraîne que $\mathrm{H}\left(C_a|\,\mu\right) = a - \Lambda_\mu(1)$. Enfin, d'après le théorème II.30, l'inégalité (II.53) prouve que μ^* est la I-projection généralisée de μ sur C_a. □

CHAPITRE III

Principe conditionnel de Gibbs pour des contraintes fines approchées

Sommaire

III.1 Introduction

III.1.1 Présentation du problème

Le problème que nous allons aborder dans ce chapitre est issu de la Mécanique Statistique : on considère un grand nombre de particules, modélisées par des variables $X_1, \ldots, X_n$ indépendantes et identiquement distribuées de loi μ sur $\mathcal{X}$ et on cherche à déterminer la loi d'une particule typique, sous la contrainte que le nuage de particules se trouve à un niveau d'énergie moyenne donné, c'est-à-dire

$$\mathcal{L}\left(X_1 \left| \frac{1}{n} \sum_{i=1}^{n} F(X_i) = a \right.\right),$$

où $F(X_i)$ désigne l'énergie de X_i. Le nombre de particules étant élevé, le problème est de déterminer la limite quand $n \to +\infty$ de la quantité précédente. Plus généralement, on cherche à calculer

$$\lim_{n\to+\infty} \mathcal{L}\left(X_1 \left| L_n \in C\right.\right),$$

où C désigne un ensemble de probabilités, et $L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$, la mesure empirique de l'échantillon.

Si C est convexe, on montre sous de bonnes hypothèses que

$$\lim_{n\to+\infty} \mathcal{L}\left(X_1 \left| L_n \in C\right.\right) = \mu^*,$$

où μ^* est la I-projection de $\mu = \mathcal{L}(X_i)$ sur l'ensemble C. Ce résultat, démontré pour la première fois par Imre Csiszár dans [19] avec une grande généralité, porte le nom de *Principe Conditionnel de Gibbs*. L'objet $\mu_C^n := \mathcal{L}\left(X_1 \left| L_n \in C\right.\right)$ peut, grâce à l'échangeabilité des X_i, se réécrire sous la forme

$$\mu_C^n = \mathbb{E}_{\mu^{\otimes n}}[L_n | L_n \in C].$$

Sous cette forme, on voit que le Principe Conditionnel de Gibbs décrit le comportement moyen de la mesure empirique L_n lorsque l'on fait un "zoom" sur la grande déviation $L_n \in C$.

Pour que cette loi conditionnelle soit bien définie, il faut imposer que C vérifie

$$\mu^{\otimes n}(L_n \in C) > 0, \qquad \text{pour tout } n \text{ assez grand.} \tag{III.1}$$

L'objet de ce chapitre est de mettre en place des moyens permettant de considérer ce que nous appellerons des *contraintes fines*, c'est à dire des ensembles C ne vérifiant pas l'hypothèse (III.1).

III.1.2 A propos de la littérature

Avant de présenter nos résultats concernant les contraintes fines, nous allons rappeler les résultats classiques de Csiszár, Stroock et Zeitouni sur le Principe Conditionnel de Gibbs. Sauf mention contraire, nous nous placerons dans le cadre suivant : $\mathcal{X}$ est un espace mesurable ; l'ensemble $\mathcal{P}(\mathcal{X})$ des mesures de probabilité sur $\mathcal{X}$ est muni de la τ-topologie, c'est-à-dire la topologie la moins fine rendant continues les applications $\nu \mapsto \int_{\mathcal{X}} f\, d\nu$, avec f mesurable et bornée, et de la tribu engendrée par ces mêmes applications.

Les contraintes épaisses

On doit le résultat suivant à I. Csiszár.

Théorème III.2 (Csiszár, [19] thm. 1)**.** *Soient $\mu \in \mathcal{P}(\mathcal{X})$ et C un ensemble convexe mesurable de $\mathcal{P}(\mathcal{X})$ fermé pour la τ-topologie tel que* $\mathrm{H}(\overset{\circ}{C}|\, \mu) = \mathrm{H}\,(C|\, \mu) < +\infty$ *; pour toute suite $(X_i)_i$ i.i.d de loi μ et pour tout $k \in \mathbb{N}^*$,*

$$\mu^n_{C,\,k} := \mathcal{L}(X_1, \ldots, X_k | L_n \in C) \in \mathcal{P}(\mathcal{X}^k)$$

est bien définie pour n suffisamment grand et converge en entropie vers $\mu^{*\,\otimes k}$, où μ^* est la I-projection de μ sur C.

Remarque III.3.

- D'après le théorème de Sanov, la condition $\mathrm{H}(\overset{\circ}{C}|\, \mu) = \mathrm{H}\,(C|\, \mu) < +\infty$ entraîne que
$$\frac{1}{n} \log \mathbb{P}(L_n \in C) \xrightarrow[n\to+\infty]{} -\mathrm{H}\,(C|\, \mu)\,. \tag{III.4}$$
Par conséquent, $\mathbb{P}(L_n \in C) > 0$ pour tout n assez grand et $\mu^n_{C,\,k}$ est bien définie.
- Le théorème III.2 est en fait valable pour une topologie un peu plus fine que la τ-topologie et pour des ensembles *presque complètement convexes* (voir la remarque A.6 pour une définition).
- La preuve de ce théorème est une conséquence immédiate de (III.4) et de la remarquable inégalité
$$\mathrm{H}\left(\mu^n_{C,\,k}\big|\, \mu^{*\otimes k}\right) \leq -\frac{1}{[n/k]} \log\left(\mathbb{P}(L_n \in C) e^{n\,\mathrm{H}(C|\mu)}\right), \tag{III.5}$$
que nous utiliserons également de manière intensive dans ce chapitre (voir [19] thm. 1, (2.17) ou l'annexe A pour une preuve).

Les conditionnements non convexes sont traités dans le théorème suivant de D.W. Stroock et O. Zeitouni :

Théorème III.6 (Stroock-Zeitouni, [64]). *Soient $\mu \in \mathcal{P}(\mathcal{X})$ et A un ensemble mesurable de $\mathcal{P}(\mathcal{X})$ tel que $\mathrm{H}(\overset{\circ}{A}|\,\mu) = \mathrm{H}\left(\overline{A}\middle|\,\mu\right) < +\infty$. Posons $\mathcal{H} = \left\{\nu \in \overline{A} : \mathrm{H}\left(\nu|\,\mu\right) = \mathrm{H}\left(A|\,\mu\right)\right\}$. Pour tout ensemble mesurable Γ tel que $\mathcal{H} \subset \overset{\circ}{\Gamma}$, on a*

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n} \left(L_n \notin \Gamma \,|L_n \in A\right) < 0 \tag{III.7}$$

Remarque III.8.

L'inégalité (III.7), qui est une application assez simple du théorème de Sanov, signifie essentiellement que la loi conditionnelle de L_n sachant que $L_n \in A$ s'accumule exponentiellement rapidement sur l'ensemble $\mathcal{H}$.

Grâce à un argument combinatoire, l'inégalité (III.7) permet de démontrer des résultats sur la convergence de $\mathcal{L}(X_1, \ldots, X_k|L_n \in A)$. Dans la proposition suivante, $\mathcal{X}$ est un espace polonais et $\mathcal{P}(\mathcal{X})$ est muni de la topologie de convergence étroite et de sa tribu borélienne.

Proposition III.9 (Stroock-Zeitouni, [64]). *Soient $\mu \in \mathcal{P}(\mathcal{X})$, et A un ensemble mesurable de $\mathcal{P}(\mathcal{X})$ tel que $\mathrm{H}(\overset{\circ}{A}|\,\mu) = \mathrm{H}\left(\overline{A}\middle|\,\mu\right) < +\infty$.*

1. *Si $\mathcal{H} = \{\mu^*\}$, alors, pour tout $k \in \mathbb{N}^*$, $\mathcal{L}(X_1, \ldots, X_k|L_n \in A)$ converge étroitement vers ${\mu^*}^{\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$.*
2. *La suite $\mathcal{L}(X_1|L_n \in A)$ est précompacte et l'ensemble de ses valeurs d'adhérence est inclus dans $\overline{\mathrm{co}}\,\mathcal{H}$.*

On pourra consulter le chapitre 7 de [26] pour une exposition classique de ces résultats.

L'approche classique des contraintes fines

Le cadre suivant a été conçu par D.W. Stroock et O. Zeitouni dans [64] pour aborder des conditionnements fins. On se donne

- une famille croissante $(A_\delta)_{\delta>0}$ d'ensembles mesurables, c'est-à-dire telle que

$$\delta < \delta' \Rightarrow A_\delta \subset A_{\delta'},$$

- une famille croissante $(F_\delta)_\delta$ d'ensembles fermés telle que

$$\forall \delta > 0, \quad A_\delta \subset F_\delta,$$

on pose

$$A_0 = \bigcap_{\delta>0} A_\delta \qquad \text{et} \qquad F_0 = \bigcap_{\delta>0} F_\delta,$$

et on fait l'hypothèse suivante

Hypothèse III.10. Il existe $\mu^* \in A_0$ tel que

$$\mathrm{H}\,(\mu^*|\,\mu) = \mathrm{H}\,(A_0|\,\mu) = \mathrm{H}\,(F_0|\,\mu) < +\infty,$$

tel que pour tout $\delta > 0$,

$$\mu^{*\otimes n}(L_n \in A_\delta) \xrightarrow[n\to+\infty]{} 1.$$

On a alors le théorème suivant

Théorème III.11 (Stroock-Zeitouni, [64])**.** *Sous l'hypothèse III.10, pour tout ensemble mesurable* Γ *contenant* $\mathcal{H} = \{\nu \in F_0 : \ \mathrm{H}\,(\nu|\,\mu) = \mathrm{H}\,(A_0|\,\mu)\}$*, on a*

$$\lim_{\delta\to 0} \limsup_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n}\left(L_n \notin \Gamma\,|L_n \in A_\delta\right) < 0 \tag{III.12}$$

De plus, si $\mathcal{X}$ *est polonais et si* $\mathcal{H} = \{\mu^*\}$*, alors, pour tout* $k \in \mathbb{N}^*$*,*

$$\lim_{\delta\to 0} \lim_{n\to+\infty} \mathcal{L}(X_1, \ldots, X_k|L_n \in A_\delta) = \mu^*,$$

au sens de la convergence étroite sur $\mathcal{P}(\mathcal{X}^k)$*.*

Différentes extensions du Principe Conditionnel de Gibbs

Depuis les travaux de Csiszár, Stroock et Zeitouni, le Principe Conditionnel de Gibbs a été généralisé dans trois directions différentes et complémentaires :

- En généralisant le théorème de Sanov pour des topologies plus fortes que la τ-topologie, P. Eichelsbacher et U. Schmock dans [30] suivis de C. Léonard et J. Najim dans [46], ont permis de considérer de nouveaux types de contraintes.
- Dans [6], E. Bolthausen et U. Schmock ont obtenu un Principe Conditionnel de Gibbs pour les mesures d'occupations de chaînes de Markov uniformément ergodiques.
- A. Dembo et O. Zeitouni se sont intéressés dans [25] à la convergence d'un bloc de taille croissante de marginales. Ils ont montré que pour des ensembles convexes C définis par des contraintes de type moment, *ie*

  $$C = \left\{\nu \in \mathcal{P}(\mathcal{X}) : \quad \int_{\mathcal{X}} F\, d\nu \in K\right\} \quad \text{avec} \quad F : \mathcal{X} \to \mathbb{R}^d \quad \text{et} \quad K \text{ convexe},$$

  on pouvait, sous certaines hypothèses, trouver une suite $k_n \xrightarrow[n\to+\infty]{} +\infty$ d'entiers telle que

  $$\left\|\mathcal{L}(X_1, \ldots, X_{k_n}|L_n \in C) - \mu^{*\,\otimes k_n}\right\|_{VT} \xrightarrow[n\to+\infty]{} 0.$$

  Ils ont obtenu des vitesses explicites pour k_n. Cette étude a été reprise par A. Dembo et J. Kuelbs dans [24] pour une fonction F à valeurs dans un espace de Banach.

III.1.3 Survol du chapitre

Contraintes fines approchées

Dans ce chapitre, nous allons étudier un nouveau moyen d'aborder les conditionnements convexes fins. Nous nous intéresserons au comportement limite de

$$\mathcal{L}(X_1, \ldots, X_k | L_n \in C_n),$$

où $(C_n)_n$ est une suite décroissante de convexes. Nous montrerons, sous diverses hypothèses, que

$$\mathcal{L}(X_1, \ldots, X_k | L_n \in C_n) \xrightarrow[n \to +\infty]{} \mu^{*\otimes k},$$

avec μ^* la I-projection de μ sur $C = \bigcap_{n \in \mathbb{N}} C_n$.

Ici, C doit être vu comme une contrainte fine, et la suite $(C_n)_n$ comme une suite de contraintes épaisses convergeant vers C. Concrètement, nous considérerons deux types de grossissement :

1. Si C est défini par une contrainte de type moment, *ie*
$$C = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \quad \int_{\mathcal{X}} F \, d\nu \in K \right\},$$
F étant une application de $\mathcal{X}$ dans un espace vectoriel normé $(B, \| \, . \, \|)$, nous grossirons C en relaxant la contrainte à ε près :
$$C_\varepsilon = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \quad \int_{\mathcal{X}} F \, d\nu \in K^\varepsilon \right\},$$
où $K^\varepsilon = \{x \in B, \exists x' \in K, \ \|x - x'\| \leq \varepsilon\}$.
2. Si C est un ensemble convexe quelconque de $\mathcal{P}(\mathcal{X})$, muni de la topologie de la convergence étroite, nous prendrons un ε-voisinage de C, *ie* nous considérerons
$$C^\varepsilon = \{\nu \in \mathcal{P}(\mathcal{X}) : \quad \exists \nu \in C, \quad \bar{d}(\nu, C) \leq \varepsilon\},$$
où $\bar{d}(\, . \, , \, . \,)$ est une distance métrisant la topologie de la convergence étroite sur $\mathcal{P}(\mathcal{X})$.

Dans ces deux situations, nous chercherons à déterminer explicitement des vitesses de rétrécissement ε_n telles qu'en posant $C_n = C_{\varepsilon_n}$, dans le premier cas et $C_n = C^{\varepsilon_n}$, dans le second, on ait

$$\mathcal{L}(X_1, \ldots, X_k | L_n \in C_n) \xrightarrow[n \to +\infty]{} \mu^{*\otimes k}, \quad \text{au sens de la convergence en variation.}$$

La principale difficulté technique que nous rencontrerons est qu'ici, contrairement à l'approche classique développée dans le théorème III.11, *le conditionnement dépend de* n ; les bornes asymptotiques fournies par le théorème de Sanov ne pourront donc pas être directement appliquées.

Cadre et notations

Avant de passer en revue nos résultats, précisons le cadre et les notations de notre étude. Dans tout ce chapitre, $\mathcal{X}$ sera un espace polonais. L'ensemble des mesures de probabilité sur $\mathcal{X}$ sera noté $\mathcal{P}(\mathcal{X})$. Comme à la section II.3.2, nous nous donnerons G, un *sous-espace vectoriel* d'applications mesurables sur $\mathcal{X}$ et à valeurs réelles et nous poserons

$$\mathcal{P}_G(\mathcal{X}) = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \quad \forall g \in G, \quad \int_{\mathcal{X}} |g| \, d\nu < +\infty \right\}.$$

Nous munirons $\mathcal{P}_G(\mathcal{X})$ de la G-topologie et de la G-tribu (voir section II.3.2). Nous supposerons toujours que l'ensemble $C_b(\mathcal{X})$ des fonctions continues bornées sur X est inclus dans G. Concrètement, G sera dans la suite l'un des espaces suivants :

- $C_b(\mathcal{X})$ (topologie de la convergence étroite),
- $B(\mathcal{X})$, ensemble des applications mesurables bornées (τ-topologie),
- $\mathcal{L}_\tau(\mathcal{X}, \mu) = \{f : \mathcal{X} \to \mathbb{R}, \text{ mesurable tq. } \exists t > 0, \int_{\mathcal{X}} e^{t|f|} \, d\mu < +\infty\}$,
- $\mathcal{L}_\tau^a(\mathcal{X}, \mu) = \{f : \mathcal{X} \to \mathbb{R}, \text{ mesurable tq. } \forall t > 0, \int_{\mathcal{X}} e^{t|f|} \, d\mu < +\infty\}$.

Pour tout entier $n \geq 1$ et tout $x \in \mathcal{X}^n$, nous poserons

$$L_n^x = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}.$$

Nous considérerons une probabilité $\mu \in \mathcal{P}_G(\mathcal{X})$ et pour tout ensemble $A \subset \mathcal{P}_G(\mathcal{X})$ tel que $\{x : L_n^x \in A\}$ est mesurable et tel que, pour tout n, $\mu^{\otimes n}(L_n \in A) > 0$, nous définirons la mesure de probabilité $\mu_{A,k}^n$ sur $\mathcal{X}^k$ par :

$$\forall B \in \mathcal{B}^{\otimes k}, \quad \mu_{A,k}^n(B) = \frac{\mu^{\otimes n}\big(x \in \mathcal{X}^k : (x_1, \ldots, x_k) \in B \quad \text{et} \quad L_n^x \in A\big)}{\mu^{\otimes n}(L_n \in A)},$$

$\mathcal{B}$ étant la tribu borélienne de $\mathcal{X}$.
Si $(X_i)_i$ désigne une suite de variables aléatoires i.i.d de loi μ, $\mu_{A,k}^n$ n'est autre que

$$\mathcal{L}\left((X_1, \ldots, X_k) \, \middle| L_n^X \in A\right).$$

Pour $k = 1$, nous noterons μ_A^n à la place de $\mu_{A,1}^n$.
Remarquons que $\mu_A^n \in \mathcal{P}_G(\mathcal{X})$ et que, pour toute fonction $g \in G$, on a, grâce à l'échangeabilité des X_i

$$\int_{\mathcal{X}} g(x) \, d\mu_A^n(x) = \frac{\mathbb{E}\left[\langle L_n^X, g\rangle \mathbb{1}_A(L_n^X)\right]}{\mathbb{P}(L_n^X \in A)}. \tag{III.13}$$

Principaux résultats du chapitre

• Section III.2 : Résultats Généraux

Dans la section III.2, nous nous placerons dans le cadre abstrait défini ci-dessus. Le résultat principal de la section est le théorème suivant

Théorème III.14. *Soit* $(C_n)_{n\geq 1}$ *une suite décroissante d'ensembles convexes de* $\mathcal{P}_G(\mathcal{X})$ *fermés pour la G-topologie et* $C = \bigcap_{n=1}^{+\infty} C_n$.

On suppose que :

1. $\mathrm{H}\,(C|\,\mu) < +\infty$,
2. μ *admet une* I*-projection* μ^* *sur* C,
3. $\lim_{n\to+\infty} \mathrm{H}\,(C_n|\,\mu) = \mathrm{H}\,(C|\,\mu)$,
4. $\liminf_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n}(L_n \in C_n) \geq -\,\mathrm{H}\,(C|\,\mu)$.

Alors, pour tout $k \in \mathbb{N}^*$, $\mu^n_{C_n,\,k}$ *converge en variation vers* $\mu^{*\otimes k}$ *dans* $\mathcal{P}(\mathcal{X}^k)$.

Idée de la preuve.
Ce théorème se démontre assez facilement à partir de l'inégalité de Csiszár (III.5). En effet, grâce à (III.5), on obtient, en notant μ^*_n la I-projection généralisée de μ sur C_n :

$$\begin{aligned}\mathrm{H}\left(\mu^n_{C_n,\,k}\middle|\,\mu_n^{*\otimes k}\right) &\leq -\frac{k}{n}\log\left(\mu^{\otimes n}(L_n \in C_n)e^{n\,\mathrm{H}(C_n|\mu)}\right)\\ &= -\frac{k}{n}\log\left(\mu^{\otimes n}(L_n \in C_n)e^{n\,\mathrm{H}(C|\mu)}\right) + k\left[\mathrm{H}\,(C|\,\mu) - \mathrm{H}\,(C_n|\,\mu)\right].\end{aligned}$$

Les conditions 3 et 4 du théorème III.14 entraînent que le membre de droite tend vers 0. L'inégalité de Pinsker permet de conclure que $\left\|\mu^n_{C_n,\,k} - \mu_n^{*\otimes k}\right\|_{VT} \xrightarrow[n\to+\infty]{} 0$. Enfin, les conditions 2 et 3 entraînent facilement que $\left\|\mu^{*\otimes k} - \mu_n^{*\otimes k}\right\|_{VT} \xrightarrow[n\to+\infty]{} 0$. $\square$

Si μ vérifie la condition de Cramér forte (II.20), alors, d'après le théorème II.21, les grandes déviations de L_n sont contrôlées par la bonne fonction de taux $\mathrm{H}\,(\,.\,|\,\mu)$. Par ailleurs, grâce à la régularité de $\mathrm{H}\,(\,.\,|\,\mu)$, les conditions 2 et 3 du théorème précédent sont automatiquement vérifiées. En revanche, même dans ce cadre régulier, la vérification de la condition

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n}(L_n \in C_n) \geq -\,\mathrm{H}\,(C|\,\mu) \tag{III.15}$$

ne relève pas du théorème de Sanov.

Pour obtenir (III.15), nous aurons besoin de bornes inférieures non-asymptotiques (valables pour tout n) pour les probabilités de grandes déviations de L_n.
La borne inférieure suivante (voir proposition III.44), due à D.W. Stroock et J.D. Deuschel,

$$\frac{1}{n}\log\left(\mu^{\otimes n}(L_n\in C_n)e^{n\,\mathrm{H}(\mu^*|\mu)}\right)\geq -\,\mathrm{H}\left(\mu^*|\,\mu\right)\frac{\mu^{*\otimes n}(L_n\in C_n^c)}{\mu^{*\otimes n}(L_n\in C_n)}$$
$$+\frac{1}{n}\log\mu^{*\otimes n}(L_n\in C_n)-\frac{1}{ne\mu^{*\otimes n}(L_n\in C_n)},\quad \text{(III.16)}$$

permet de remplacer la condition (III.15) du théorème III.14 par la condition plus simple

$$\lim_{n\to+\infty}\mu^{*\otimes n}(L_n\in C_n)=1. \quad \text{(III.17)}$$

Dans la mesure où $\mu^*\in C_n$ pour tout n, la condition (III.17) est une condition de type loi des grands nombres.

Toujours dans la section III.2, nous essaierons d'améliorer la convergence de $\mu^n_{C_n}$ vers μ^*. Dans le cas où G est l'espace d'Orlicz $\mathcal{L}_\tau(\mathcal{X},\mu)$ (voir la page 65 pour des rappels sur les espaces d'Orlicz), nous nous intéresserons à la convergence forte de $\mu^n_{C_n}$ vue comme une forme linéaire continue sur $\mathbb{L}_\tau(\mathcal{X},\mu)$. Nous poserons pour tout $\ell\in\mathbb{L}_\tau(\mathcal{X},\mu)'$ (le dual topologique de $\mathbb{L}_\tau(\mathcal{X},\mu)$),

$$\|\ell\|_\tau^* := \sup_{\substack{f\in\mathbb{L}_\tau(\mathcal{X},\mu)\\ \|f\|_\tau\leq 1}}\langle \ell,f\rangle,$$

où $\|\,.\,\|_\tau$ est la norme de Luxembourg sur l'espace d'Orlicz $\mathbb{L}_\tau(\mathcal{X},\mu)$.
La proposition suivante donne une condition générale pour obtenir la convergence de $\mu^n_{C_n}$ vers μ^* au sens de la norme $\|\,.\,\|_\tau^*$:

Proposition III.18. *Sous les hypothèses du théorème III.14, notons $h_n=\frac{d\mu_n^*}{d\mu}$, où μ_n^* est la I-projection généralisée de μ sur C_n et supposons que $(h_n)_n$ soit une suite bornée de $\mathbb{L}_p(\mathcal{X},\mu)$ pour un certain $p>1$, alors $\left\|\mu^n_{C_n}-\mu^*\right\|_\tau^*\xrightarrow[n\to+\infty]{}0$.*

Idée de la preuve. En utilisant la généralisation II.16 de l'inégalité de Pinsker, on montre que

$$\forall \nu_1,\nu_2\in\mathcal{P}_\tau(\mathcal{X}),$$
$$\|\nu_1-\nu_2\|_\tau^*\leq C_p\left(1+\log\int_{\mathcal{X}}\frac{d\nu_2}{d\mu}^p d\mu\right)\left(\sqrt{\mathrm{H}\left(\nu_1|\,\nu_2\right)}+\mathrm{H}\left(\nu_1|\,\nu_2\right)\right),\quad \text{(III.19)}$$

où $\mathcal{P}_\tau(\mathcal{X})=\mathcal{P}_{\mathcal{L}_\tau(\mathcal{X},\mu)}(\mathcal{X})$ et C_p est une constante ne dépendant que de p.
On obtient le résultat en reprenant pas à pas la preuve du théorème III.14 pour $k=1$ mais en en utilisant cette fois l'inégalité (III.19) (avec $\nu_2=\mu_n^*$) à la place de l'inégalité de Pinsker. $\square$

• Section III.3 : Conditionnement par des contraintes de type moment

Dans cette section, $G = \mathcal{L}_\tau(\mathcal{X}, \mu)$ et on s'intéresse au cas particulier important d'un conditionnement de la forme

$$C = \left\{ \nu \in \mathcal{P}_\tau(\mathcal{X}) : \int_{\mathcal{X}} F \, d\nu \in K \right\}$$

avec F une application mesurable à valeurs dans un espace de Banach séparable B muni de sa tribu borélienne et K un convexe fermé de B. On supposera que $\|F\| \in \mathcal{L}_\tau(\mathcal{X}, \mu)$, de sorte que C est fermé.

Comme nous l'avons expliqué plus haut, nous grossirons C de la manière suivante :

$$C_n = \left\{ \nu \in \mathcal{P}_\tau(\mathcal{X}) : \int_{\mathcal{X}} F \, d\nu \in K^{\varepsilon_n} \right\},$$

où $K^{\varepsilon_n} = \{x \in B : d(x, C) \leq \varepsilon_n\}$ et ε_n est une suite de réels positifs décroissant lentement vers 0.

Dans cette section, Z_F, Λ_F et Λ_F^* seront respectivement la transformée de Laplace, la Log-Laplace et la transformée de Cramér de μ_F, image de μ par F.

Le résultat principal de cette section est le théorème suivant, où B est un espace de dimension finie.

Théorème III.20.
On suppose que

- *B est de dimension finie,*
- *$\operatorname{dom} \Lambda_F := \{\lambda \in B' : \Lambda_F(\lambda) < +\infty\}$ est ouvert dans B',*
- *L'enveloppe convexe du support de μ_F, $\operatorname{co} S_F$ est d'intérieur non vide.*

Si K est un convexe fermé de B tel que $K \cap \operatorname{co} \overset{\circ}{S_F} \neq \emptyset$, alors

1. *μ possède une I-projection μ^* sur $C = \left\{ \nu \in \mathcal{P}_\tau(\mathcal{X}),\ \int_{\mathcal{X}} F \, d\nu \in K \right\}$,*
2. *Il existe $\bar{c} \in \mathbb{R}^+$ tel que pour toute suite $\varepsilon_n \in \mathbb{R}^+$ de limite nulle telle que $\lim\limits_{n \to +\infty} n\varepsilon_n^2 \in]\bar{c}, +\infty]$, la suite $\mu^n_{C_n,k}$ converge en variation vers $\mu^{*\otimes k}$, pour tout $k \in \mathbb{N}^*$, où $C_n = \left\{ \nu \in \mathcal{P}_\tau(\mathcal{X}),\ \int_{\mathcal{X}} F \, d\nu \in K^{\varepsilon_n} \right\}$.*
3. *De plus, $\left\| \mu^n_{C_n} - \mu^* \right\|_\tau^* \xrightarrow[n \to +\infty]{} 0$.*
4. *Enfin, pour tout k, $\mu^n_{C_n,k}$ converge en entropie vers $\mu^{*\otimes k}$.*

Idée de la preuve. Tout d'abord, on montre, en utilisant les résultats de la section II.4.4 du chapitre précédent, que μ admet sur C (resp. sur C_n) une I-projection μ^* (resp. μ_n^*) qui s'écrit

$$\mu^* = \frac{e^{\langle \lambda^*, F\rangle}}{Z_F(\lambda^*)}\mu \qquad \left(\text{resp.} \qquad \mu_n^* = \frac{e^{\langle \lambda_n^*, F\rangle}}{Z_F(\lambda_n^*)}\mu\right),$$

avec λ^*, (resp. λ_n^*) l'unique minimisant de la fonction

$$H(\lambda) = \Lambda_F(\lambda) - \inf_{y\in K} \langle \lambda, y\rangle \qquad \left(\text{resp.} \qquad H_n(\lambda) = \Lambda_F(\lambda) - \inf_{y\in K^{\varepsilon_n}} \langle \lambda, y\rangle\right).$$

De plus, en utilisant des techniques classiques d'optimisation convexe, on voit que $\lim_{n\to+\infty} \lambda_n^* = \lambda^*$. Cela entraîne facilement que

$$\lim_{n\to+\infty} \mathrm{H}\left(C_n | \, \mu\right) = \mathrm{H}\left(C | \, \mu\right).$$

D'après les résultats de la section III.2, la seule chose à vérifier pour obtenir la convergence en variation de $\mu^n_{C_n,k}$ vers $\mu^{*\otimes k}$ est que

$$\mu^{*\otimes n}(L_n \in C_n) \xrightarrow[n\to+\infty]{} 1.$$

Or,

$$\mu^{*\otimes n}(L_n \in C_n) \geq \mathbb{P}\left(\left\| \frac{1}{n}\sum_{i=1}^n F(Y_i) - \int_{\mathcal{X}} F d\mu^* \right\| \leq \varepsilon_n\right),$$

avec Y_i une suite i.i.d de loi μ^*.
On voit facilement qu'il existe $\delta > 0$ tel que $\int_{\mathcal{X}} e^{\delta\|F\|}\, d\mu^* < +\infty$. On peut donc appliquer l'inégalité de Bernstein et conclure que

$$\mathbb{P}\left(\left\| \frac{1}{n}\sum_{i=1}^n F(Y_i) - \int_{\mathcal{X}} F d\mu^* \right\| > \varepsilon_n\right) \simeq e^{-n\varepsilon_n^2}.$$

Ainsi, si $n\varepsilon_n^2 \xrightarrow[n\to+\infty]{} +\infty$, la convergence en variation est démontrée. Nous verrons qu'en travaillant plus finement, on peut même prendre des suites $\varepsilon_n \sim \frac{c}{\sqrt{n}}$.

Pour montrer que $\left\|\mu^n_{C_n} - \mu^*\right\|^*_\tau \xrightarrow[n\to+\infty]{} 0$, il suffit de s'assurer que $h_n = \frac{d\mu_n^*}{d\mu}$ est bornée dans $\mathbb{L}_p(\mathcal{X}, \mu)$, pour un certain $p > 1$. Ceci découle facilement de

$$\int_{\mathcal{X}} \left(\frac{d\mu_n^*}{d\mu}\right)^p d\mu = \frac{Z_F(p\lambda_n^*)}{Z_F(\lambda_n^*)^p},$$

de la convergence de λ_n^* vers λ^* et du fait que $\operatorname{dom} Z_F$ est ouvert. $\square$

Si $B = \mathbb{R}$, on peut améliorer la vitesse de grossissement ε_n. On montre à la proposition III.70, en utilisant l'inégalité de Berry-Esseen (voir (III.72)), que les conclusions du théorème III.20 restent valables pour $\varepsilon_n = \frac{1}{n^a}$, avec $0 < a < 1$.

Le reste de la section III.3 est consacrée à des généralisations du théorème précédent pour des fonctions F à valeurs dans un espace de Banach de dimension infinie. Sous de bonnes hypothèses, la convergence en variation de $\mu^n_{C_n,k}$ vers $\mu^{*\otimes k}$ est démontrée au théorème III.76. La preuve est sensiblement la même que celle esquissée ci-dessus, à ceci près que l'inégalité de Bernstein est remplacée par sa généralisation infini-dimensionnelle donnée par le théorème de Yurinskii (voir théorème III.77). En revanche, la convergence pour la norme $\| \,.\, \|^*_\tau$ semble pour l'instant hors de portée.

• Section III.4 : Contraintes plus générales - Contrôles par recouvrement.

Dans cette section, nous revenons au cadre classique où $G = C_b(\mathcal{X})$ et nous mettons en place une méthode permettant de traiter le cas d'une contrainte convexe fine C générale. Nous munirons $\mathcal{P}(\mathcal{X})$ d'une distance $\bar{d}$ métrisant la topologie de la convergence étroite. *Dans tout ce qui suit, $\bar{d}$ sera ou bien la distance de Fortet-Mourier (voir (III.87)) ou bien la distance de Prokhorov (voir (III.88)).*

Les grossissements de C considérés dans cette section sont de la forme

$$C_n = \{\nu \in \mathcal{P}(\mathcal{X}), \bar{d}(\nu, C) \le \varepsilon_n\};$$

l'objectif étant de construire explicitement des suites $(\varepsilon_n)_n$ de limite nulle telles que

$$\mu^n_{C_n,k} \xrightarrow[n\to+\infty]{} \mu^{*\otimes k},$$

au sens de la convergence en variation sur $\mathcal{P}(\mathcal{X}^k)$. D'après les résultat généraux de la section III.2, la seule chose à montrer est que

$$\mu^{*\otimes n}(\bar{d}(L_n, \mu^*) \le \varepsilon_n) \xrightarrow[n\to+\infty]{} 1. \tag{III.21}$$

Le cas compact. Dans un premier temps, nous supposerons que $\mathcal{X}$ est un espace métrique compact. Un résultat classique (voir le théorème III.92) entraîne que $(\mathcal{P}(\mathcal{X}), \bar{d})$ est lui aussi un espace métrique compact. Pour montrer (III.21), nous allons utiliser une technique développée par S. Kulkarni et O. Zeitouni dans l'article [42]. Cette technique permet d'obtenir des contrôles non-asymptotiques faisant intervenir des nombres de recouvrement pour les probabilités de grandes déviations de L_n (voir [42], théorème 1). Rappelons que si K est une partie compacte d'un espace métrique $(\mathcal{Y}, d)$, le nombre de recouvrement de niveau ε, noté $N_{\mathcal{Y}}(d, K, \varepsilon)$, est par définition le nombre minimal de boules ouvertes de rayon ε nécessaires pour recouvrir K. La méthode de [42] permet d'obtenir la proposition suivante

Proposition III.22. *Soit A une partie mesurable de $\mathcal{P}(\mathcal{X})$. Pour tout $\nu \in \mathcal{P}(\mathcal{X})$, on a*

$$\forall \xi > 0, \quad \nu^{\otimes n}(L_n \in A^\xi) \le N_{\mathcal{P}(\mathcal{X})}(\bar{d}, A, \xi) e^{-n \operatorname{H}\left(A^{2\xi} | \nu \right)}, \tag{III.23}$$

en notant $A^\xi := \{\nu \in \mathcal{P}(\mathcal{X}) : \ \bar{d}(\nu, A) \le \xi\}$.

En appliquant la borne (III.23) avec $A = B(\mu^*, \varepsilon)^c$ et $\xi = \frac{\varepsilon}{4}$, on obtient

$$\mu^{*\otimes n}(L_n \in B(\mu^*, \varepsilon)^c) \le N_{\mathcal{P}(\mathcal{X})} \left(B(\mu^*, \varepsilon)^c, \frac{\varepsilon}{4} \right) e^{-n \operatorname{H}\left(B\left(\mu^*, \frac{\varepsilon}{2}\right)^c | \mu^* \right)}.$$

En notant $N_{\mathcal{P}(\mathcal{X})}\left(\bar{d}, \varepsilon\right) = N_{\mathcal{P}(\mathcal{X})}\left(\bar{d}, \mathcal{P}(\mathcal{X}), \varepsilon\right)$ et en utilisant l'inégalité de Pinsker

$$\bar{d}(\nu, \mu) \le \|\nu - \mu\|_{VT} \le \sqrt{2 \operatorname{H}(\nu | \mu)},$$

on obtient

$$\mu^{*\otimes n}(L_n \in B(\mu^*, \varepsilon)^c) \le N_{\mathcal{P}(\mathcal{X})} \left(\bar{d}, \frac{\varepsilon}{4} \right) e^{-n \frac{\varepsilon^2}{8}}.$$

Ainsi, la condition (III.21) est vérifiée pour toute suite $(\varepsilon_n)_n$ de limite nulle telle que

$$N_{\mathcal{P}(\mathcal{X})} \left(\bar{d}, \frac{\varepsilon_n}{4} \right) e^{-n \frac{\varepsilon_n^2}{8}} \xrightarrow[n \to +\infty]{} 0 \tag{III.24}$$

Pour rendre la condition (III.24) plus facilement vérifiable, nous utiliserons le lemme 1 de [42] qui permet de démontrer que

$$\forall \varepsilon > 0, \quad N_{\mathcal{P}(\mathcal{X})}(\bar{d}, \varepsilon) \le \left(\frac{4e}{\varepsilon} \right)^{N_{\mathcal{X}}(d, \varepsilon/2)}. \tag{III.25}$$

Grâce à la majoration (III.25), nous obtiendrons le

Corollaire III.26. *Pour toute suite $\varepsilon_n > 0$ de limite nulle telle que*

$$\frac{n\varepsilon_n^2}{8} + \log(\varepsilon_n) N_{\mathcal{X}} \left(d, \frac{\varepsilon_n}{8} \right) \xrightarrow[n \to +\infty]{} +\infty, \tag{III.27}$$

$\mu^n_{C^{\varepsilon_n}, k}$ converge en variation vers $\mu^{*\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$.

Nous verrons à la proposition III.105 que pour tout espace métrique compact $(\mathcal{X}, d)$, il existe toujours au moins une suite $(\varepsilon_n)_n$ de limite nulle vérifiant la condition (III.27). Par ailleurs, la littérature abonde en estimations des nombres de recouvrement $N_{\mathcal{X}}(d, \varepsilon)$ qui permettent via le critère (III.27) de calculer des vitesses de rétrécissement $(\varepsilon_n)_n$ explicites. Par exemple, si $\mathcal{X}$ est une partie compacte de $\mathbb{R}^q$, et d la distance euclidienne, on a la majoration classique $N_{\mathcal{X}}(d, \varepsilon) \le \frac{c}{\varepsilon^q}$ pour tout ε assez petit (voir proposition III.85), on en déduit facilement que, dans ce cas, on peut prendre $\varepsilon_n = \frac{1}{n^a}$ avec $0 < a < \frac{1}{q+2}$ (voir proposition III.104).

Extension au cas non-compact. Pour étendre les résultats précédents au cas où $(\mathcal{X}, d)$ n'est plus compact, nous allons mettre en oeuvre une technique d'approximation. On commence par approcher μ^* par la probabilité $\mu_K^* := \frac{\mathbf{1}_K}{\mu^*(K)}\mu^*$, où K est un compact de $\mathcal{X}$; pour cette probabilité $\mu_K^* \in \mathcal{P}(K)$, on dispose de la borne

$$\forall \xi > 0, \quad \mu_K^{*\otimes n}\left(L_n \in B(\mu_K^*, \xi)\right) \geq 1 - \left(\frac{16e}{\xi}\right)^{N_K\left(d, \frac{\xi}{8}\right)} e^{-n\frac{\xi^2}{8}}.$$

Un argument technique assez simple permet d'en déduire la borne suivante :

Pour tout $\xi > 0$,

$$\mu^{*\otimes n}\left(\bar{d}(L_n, C) \leq \xi + 2\mu^*(K^c)\right) \geq \mu^*(K)^n \left(1 - \left(\frac{16e}{\xi}\right)^{N_K\left(d, \frac{\xi}{8}\right)} e^{-n\frac{\xi^2}{8}}\right). \quad \text{(III.28)}$$

La borne (III.28) permet de calculer des vitesses de rétrécissement $(\varepsilon_n)_n$, comme le montre la proposition suivante :

Proposition III.29. *Soient C un convexe fermé de $\mathcal{P}(\mathcal{X})$ tel que $\mathrm{H}(C|\,\mu) < +\infty$ et μ^* la I-projection de μ sur C. S'il existe une suite $(K_n)_n$ de compacts inclus dans $\mathcal{X}$ et une suite $\xi_n > 0$ de limite nulle telles que :*

$$\mu^*(K_n)^n \xrightarrow[n\to+\infty]{} 1 \qquad \textit{et} \qquad \frac{n\xi_n^2}{8} + \log(\xi_n) N_{K_n}\left(d, \frac{\xi_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty, \quad \text{(III.30)}$$

alors, pour toute suite ε_n de limite nulle telle que $\varepsilon_n \geq \xi_n + 2\mu^*(K_n^c)$, la suite $\mu^n_{C^{\varepsilon_n},k}$ converge en variation vers $\mu^{*\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$.

Nous verrons à la proposition (III.109) que si $\frac{d\mu^*}{d\mu}$ est continue et bornée sur $\mathcal{X}$, alors le critère (III.30) peut être remplacé par la condition plus faible

$$\mu^*(K_n) \xrightarrow[n\to+\infty]{} 1 \qquad \text{et} \qquad \frac{n\xi_n^2}{8} + \log(\xi_n) N_{K_n}\left(d, \frac{\xi_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty. \quad \text{(III.31)}$$

Les critères (III.30) et (III.31) sont nettement plus difficiles à vérifier que le critère (III.27) du cas compact, le support de la probabilité μ^* devant être bien approximé par une suite de compacts pas trop gros (au sens de l'entropie métrique). Par exemple, si l'on se place dans $\mathbb{R}^q$, on doit disposer d'informations précises sur la queue de distribution de μ^* pour être en mesure de calculer des vitesses de rétrécissement explicites.

Proposition III.32. *Soient C un convexe fermé de $\mathcal{P}(\mathbb{R}^q)$ tel que $\mathrm{H}\left(C|\,\mu\right) < +\infty$ et μ^* la I-projection de μ sur C.*

1. *S'il existe $a > q$ tel que*
$$\int_{\mathcal{X}} \|x\|^a \, d\mu^*(x) < +\infty, \tag{III.33}$$
alors pour $\varepsilon_n = \frac{2}{n^b}$, avec $b < \frac{1-\frac{q}{a}}{q+2}$, $\mu^n_{C^{\varepsilon_n},k}$ converge en variation vers $\mu^{*\otimes k}$.
En particulier, s'il existe $u > 0$ tel que $\int_{\mathcal{X}} e^{u\|x\|} \, d\mu^*(x) < +\infty$, on peut prendre $b < \frac{1}{q+2}$.

2. *S'il existe $a > 0$ tel que (III.33) soit satisfaite et si on suppose en plus que $\log \frac{d\mu^*}{d\mu}$ est continue et bornée, alors on peut prendre $b < \frac{1}{q+2}$.*

La probabilité μ^* étant en général mal connue, l'hypothèse (III.33), ou tout autre hypothèse d'intégrabilité, est difficile à vérifier. On dispose néanmoins du résultat élémentaire suivant

Proposition III.34. *S'il existe $a > 0$ et $\lambda > 0$ tels que*

$$\int_{\mathcal{X}} e^{\lambda\|x\|^a} \, d\mu < +\infty, \tag{III.35}$$

et si $\nu \in \mathcal{P}(\mathcal{X})$ vérifie $\mathrm{H}\left(\nu|\,\mu\right) < +\infty$, alors $\int_{\mathcal{X}} \|x\|^a \, d\nu < +\infty$. En particulier, les conclusions de la proposition III.32 restent inchangées si l'on remplace l'hypothèse (III.33) par l'hypothèse (III.35).

Applications à l'étude des ponts de Schrödinger et des processus de Nelson. Nous terminerons ce chapitre par une interprétation des ponts de Schrödinger et des processus de Nelson. Ces processus sont les I-projections de la mesure de Wiener sur des convexes fermés de la forme

$$C(\nu_t) = \left\{\mathcal{V} \in \mathcal{P}(C([0,1], \mathbb{R}^q)) : \forall t \in I, \quad \mathcal{V}_t = \nu_t\right\},$$

où I est un sous-ensemble de $[0,1]$ et $(\nu_t)_{t\in I}$ est une famille de probabilités sur $\mathbb{R}^q$. Pour de bons flots de marginales $(\nu_t)_{t\in I}$, nous déterminerons des suites ε_n explicites telles que

$$\mathcal{W}^n_{\varepsilon_n,k} := \mathcal{L}(X_1, \ldots, X_k | L_n \in C(\nu_t)^{\varepsilon_n}) \xrightarrow[n\to+\infty]{} \mathcal{W}^{*\otimes k},$$

où X_i est une suite i.i.d de loi $\mathcal{W}$.

III.2 Résultats généraux

Rappelons que dans cette section, G désigne un sous-espace vectoriel d'applications mesurables sur l'espace polonais $(\mathcal{X}, d)$ contenant l'ensemble $\mathcal{C}_b(\mathcal{X})$ des applications continues sur $\mathcal{X}$. L'ensemble $\mathcal{P}_G(\mathcal{X})$ de toutes les mesures de probabilités ν sur $\mathcal{X}$ telles que $\forall g \in G, \quad \int_{\mathcal{X}} |g| d\nu < +\infty$ est muni de la G-topologie et de la G-tribu introduites à la section II.3.2.

Dans la suite, nous fixerons un élément μ de $\mathcal{P}_G(\mathcal{X})$ et nous étudierons le comportement asymptotiques des suites de la forme

$$\mu^n_{C_n,\,k} := \mathcal{L}(X_1, \ldots, X_k | L_n \in C_n)$$

avec $(X_i)_i$ une suite i.i.d de loi μ et C_n une suite décroissante de convexes de $\mathcal{P}_G(\mathcal{X})$.

III.2.1 Convergence en variation

Le théorème suivant a pour but de dégager un lot de conditions suffisantes garantissant la convergence en variation de $\mu^n_{C_n,\,k}$ vers $\mu^{*\otimes k}$, la probabilité μ^* étant la I-projection de μ sur $C = \bigcap_{n=1}^{+\infty} C_n$.

Théorème III.36. *Soit $(C_n)_{n\geq 1}$ une suite décroissante d'ensembles convexes de $\mathcal{P}_G(\mathcal{X})$ fermés pour la G-topologie et $C = \bigcap_{n=1}^{+\infty} C_n$.*
On suppose que :

1. $\mathrm{H}\,(C|\,\mu) < +\infty$,
2. *μ admet une I-projection μ^* sur C,*
3. $\lim_{n\to+\infty} \mathrm{H}\,(C_n|\,\mu) = \mathrm{H}\,(C|\,\mu)$,
4. $\liminf_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n}(L_n \in C_n) \geq -\,\mathrm{H}\,(C|\,\mu)$.

Alors, pour tout $k \in \mathbb{N}^*$, $\mu^n_{C_n,\,k}$ converge en variation vers $\mu^{*\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$.

La preuve de ce résultat repose sur le théorème suivant, du à I. Csiszár.

Théorème III.37. *Soit A un ensemble convexe fermé de $\mathcal{P}_G(\mathcal{X})$. On suppose que $\mathrm{H}\,(A|\,\mu) < +\infty$ et on note μ^*, la I-projection généralisée de μ sur A.*
Si $\mu^{\otimes n}(L_n \in A) > 0$, alors pour tout $k \in \{1, \ldots, n\}$, on a

$$\mathrm{H}\left(\mu^n_{A,\,k}\middle|\,\mu^{*\otimes k}\right) \leq -\frac{1}{[n/k]} \log\left(\mu^{\otimes n}(L_n \in A) e^{n\,\mathrm{H}(\,A|\mu)}\right). \qquad \text{(III.38)}$$

Démonstration. Voir l'annexe A.

Démonstration. On a, en notant μ_n^* la I-projection généralisée de μ sur C_n,

$$\begin{aligned}
\left\|\mu^n_{C_n,k} - \mu^{*\otimes k}\right\|_{VT} &\le \left\|\mu^n_{C_n,k} - \mu_n^{*\otimes k}\right\|_{VT} + \left\|\mu^{*\otimes k} - \mu_n^{*\otimes k}\right\|_{VT} && \text{(III.39)}\\
&\le \sqrt{2\,\mathrm{H}\left(\mu^n_{C_n,k}\middle|\,\mu_n^{*\otimes k}\right)} + \sqrt{2\,\mathrm{H}\left(\mu^{*\otimes k}\middle|\,\mu_n^{*\otimes k}\right)} && \text{(III.40)}\\
&= \sqrt{2\,\mathrm{H}\left(\mu^n_{C_n,k}\middle|\,\mu_n^{*\otimes k}\right)} + \sqrt{2k\,\mathrm{H}\left(\mu^*\middle|\,\mu_n^*\right)} && \text{(III.41)}
\end{aligned}$$

où (III.39) vient de l'inégalité triangulaire, (III.40) de l'inégalité de Pinsker (II.13) et (III.41) de la formule de décomposition de l'entropie (II.4).

Comme μ^* est la I-projection de μ sur C, μ^* appartient à C et donc aussi à C_n. Par conséquent, d'après l'inégalité de Csiszár (II.26),

$$\mathrm{H}\left(C|\,\mu\right) = \mathrm{H}\left(\mu^*|\,\mu\right) \ge \mathrm{H}\left(\mu^*|\,\mu_n^*\right) + \mathrm{H}\left(C_n|\,\mu\right). \qquad \text{(III.42)}$$

Ainsi, d'après l'hypothèse (3) du théorème, $\mathrm{H}\left(\mu^*|\,\mu_n^*\right)$ tend vers 0.

Pour prouver la convergence en variation de $\mu^n_{C_n,k}$ vers $\mu^{*\otimes k}$, il suffit donc, d'après (III.41), de montrer que $\lim\limits_{n\to+\infty} \mathrm{H}\left(\mu^n_{C_n,k}\middle|\,\mu_n^{*\otimes k}\right) = 0$. Or, d'après l'hypothèse (4), pour n assez grand, on a $\mu^{\otimes n}(L_n \in C_n) > 0$. On peut donc appliquer le théorème III.37 avec $A = C_n$, ce qui entraîne

$$\begin{aligned}
\mathrm{H}\left(\mu^n_{C_n,k}\middle|\,\mu_n^{*\otimes k}\right) &\le -\frac{k}{n}\log\left(\mu^{\otimes n}(L_n \in C_n)e^{n\,\mathrm{H}(C_n|\mu)}\right)\\
&= -\frac{k}{n}\log\left(\mu^{\otimes n}(L_n \in C_n)e^{n\,\mathrm{H}(C|\mu)}\right) + k\left[\mathrm{H}\left(C|\,\mu\right) - \mathrm{H}\left(C_n|\,\mu\right)\right].
\end{aligned}$$

D'après l'hypothèse (3), le dernier terme tend vers 0 et d'après l'hypothèse (4),

$$\limsup_{n\to+\infty} -\frac{1}{n}\log\left(\mu^{\otimes n}(L_n \in C_n)e^{n\,\mathrm{H}(C|\mu)}\right) \le 0.$$

□

Remarque III.43.

Notons

$$\mathcal{L}^a_\tau(\mathcal{X},\mu) = \left\{ g \text{ mesurable } : \forall s \in \mathbb{R}, \quad \int_{\mathcal{X}} e^{s|g|}\, d\mu < +\infty \right\}.$$

Si $G \subset \mathcal{L}^a_\tau(\mathcal{X},\mu)$, alors, d'après la proposition II.36, μ admet une I-projection sur le convexe fermé C vérifiant $\mathrm{H}\left(C|\,\mu\right) < +\infty$. Par ailleurs, d'après le point 2 de la proposition II.34, $\mathrm{H}\left(\,.\,|\,\mu\right)$ est une bonne fonction de taux sur $\mathcal{P}_G(\mathcal{X})$, donc, d'après le point (a) du lemme 4.1.6 de [26], on a $\mathrm{H}\left(C_n|\,\mu\right) \xrightarrow[n\to+\infty]{} \mathrm{H}\left(C|\,\mu\right)$. Ainsi, dans ce cadre régulier, il suffit de vérifier les hypothèses 1 et 4.

Pour vérifier la condition

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n}(L_n \in C_n) \geq -\operatorname{H}(C|\,\mu),$$

il est indispensable de disposer de bornes inférieures exactes (non-asymptotiques) pour le théorème de Sanov. La proposition suivante, démontrée en exercice dans le livre de J.D. Deuschel et D.W. Stroock, fournie une telle borne :

Proposition III.44. *Soient A une partie de $\mathcal{P}_G(\mathcal{X})$ telle que $\{x : L_n^x \in A\}$ est mesurable et $\nu \in \mathcal{P}_G(\mathcal{X})$ telle que $\nu \ll \mu$ et $\nu^{\otimes n}(L_n \in A) > 0$. Alors,*

$$\frac{1}{n} \log \left(\mu^{\otimes n}(L_n \in A) e^{n \operatorname{H}(\nu|\mu)}\right) \geq -\operatorname{H}(\nu|\,\mu) \frac{\nu^{\otimes n}(L_n \in A^c)}{\nu^{\otimes n}(L_n \in A)} + \frac{1}{n} \log \nu^{\otimes n}(L_n \in A) - \frac{1}{ne\nu^{\otimes n}(L_n \in A)}. \quad \text{(III.45)}$$

Démonstration. Voir l'annexe A. □

Le corollaire suivant exploite l'inégalité (III.45) et permet de remplacer l'hypothèse 4 du théorème III.36 par une condition de type loi des grands nombres :

Corollaire III.46. *Sous les hypothèses 1,2, et 3 du théorème III.36, $\mu^n_{C_n,k}$ converge en variation vers $\mu^{*\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$, dès que*

$$\lim_{n\to+\infty} \mu^{*\otimes n}(L_n \in C_n) = 1.$$

Démonstration.
Il suffit de montrer que $I_n := -\frac{1}{n} \log \left(\mu^{\otimes n}(L_n \in C_n) e^{n \operatorname{H}(C|\mu)}\right)$ est majoré par une suite de limite nulle. Or, en appliquant la proposition III.44 avec $A = C_n$ et $\nu = \mu^*$ (qui vérifie $\operatorname{H}(\mu^*|\,\mu) = \operatorname{H}(C|\,\mu)$ et $\mu^{*\otimes n}(L_n \in C_n) > 0$ pour n assez grand), on obtient :

$$I_n \leq \operatorname{H}(C|\,\mu) \frac{\mu^{*\otimes n}(L_n \in C_n^c)}{\mu^{*\otimes n}(L_n \in C_n)} - \frac{1}{n} \log \mu^{*\otimes n}(L_n \in C_n) + \frac{1}{ne\mu^{*\otimes n}(L_n \in C_n)},$$

et comme $\lim_{n\to+\infty} \mu^{*\otimes n}(L_n \in C_n) = 1$, le membre de droite tend vers 0. □

III.2.2 Convergence forte dans $\mathbb{L}_\tau(\mathcal{X}, \mu)'$

La convergence en variation donnée par le théorème III.36 n'est pas toujours satisfaisante. En effet, si l'on prend $C = \{\nu \in \mathcal{P}_G(\mathcal{X}) : \int_\mathcal{X} f\, d\nu = a\}$, avec $f \in G$ non bornée, la convergence en variation de $\mu^n_{C_n}$ vers μ^* n'est pas assez forte pour pouvoir affirmer que

$$\lim_{n\to+\infty} \int_\mathcal{X} f\, d\mu^n_{C_n} = \int_\mathcal{X} f\, d\mu^* = a.$$

En fait, la convergence en variation d'une suite ν_n vers ν n'est autre que la convergence forte de ν_n vers ν en tant que formes linéaires continues sur $B(\mathcal{X})$. Si $(G, \|\,.\,\|_G)$ est un espace vectoriel normé, la bonne notion de convergence serait la convergence pour la norme $\|\,.\,\|_G^*$, définie pour toute forme linéaire ℓ continue sur G par :

$$\|\ell\|_G^* = \sup_{\|g\|_G \le 1} \langle \ell, g \rangle.$$

La proposition III.51 suivante donne une condition suffisante qui garantit la convergence forte de $\mu_{C_n}^n$ vers μ^* dans le cas où G est l'espace d'Orlicz $\mathcal{L}_\tau(\mathcal{X}, \mu)$.

Rappels sur les espaces d'Orlicz.
Rappelons qu'une *fonction de Young* est une fonction $\theta : \mathbb{R} \to [0, +\infty]$ convexe, paire et telle que

$$\theta(0) = 0, \qquad \theta(s) \xrightarrow[s \to +\infty]{} +\infty, \qquad \exists s_0 > 0, \quad \theta(s_0) < +\infty.$$

Si μ est une mesure de probabilité sur un espace mesurable $(\mathcal{X}, \mathcal{B})$, on définit les deux espaces

$$\mathcal{L}_\theta(\mathcal{X}, \mu) = \left\{ g : \mathcal{X} \to \mathbb{R}, \text{mesurable} : \exists s > 0, \quad \int_{\mathcal{X}} \theta\left(\frac{g}{s}\right) < +\infty \right\}$$

et

$$\mathcal{L}_\theta^a(\mathcal{X}, \mu) = \left\{ g : \mathcal{X} \to \mathbb{R}, \text{mesurable} : \forall s > 0, \quad \int_{\mathcal{X}} \theta\left(\frac{g}{s}\right) < +\infty \right\}.$$

On note $\mathbb{L}_\theta(\mathcal{X}, \mu)$ (resp. $\mathbb{L}_\theta^a(\mathcal{X}, \mu)$) l'ensemble des classes d'équivalence de fonctions de $\mathcal{L}_\theta(\mathcal{X}, \mu)$ (resp. $\mathcal{L}_\theta^a(\mathcal{X}, \mu)$) pour la relation d'égalité μ-presque sûrement.

On définit sur $\mathbb{L}_\theta(\mathcal{X}, \mu)$ une norme, appelée *norme de Luxembourg*, par

$$\forall g \in \mathbb{L}_\theta(\mathcal{X}, \mu), \quad \|g\|_\theta = \inf \left\{ s > 0 : \int_{\mathcal{X}} \theta\left(\frac{g}{s}\right) d\mu \le 1 \right\}.$$

On montre que $(\mathbb{L}_\theta(\mathcal{X}, \mu), \|\,.\,\|_\theta)$ est un espace de Banach ; c'est l'*espace d'Orlicz associé à la fonction θ*.

Si θ est une fonction de Young, sa conjuguée convexe θ^* définie par

$$\theta^*(t) = \sup_{s \in \mathbb{R}} \{ st - \theta(s) \}$$

est encore une fonction de Young.

L'inégalité de Young

$$\forall s, t \in \mathbb{R}, \quad st \leq \theta(s) + \theta^*(t)$$

permet de démontrer que si $f \in \mathbb{L}_\theta(\mathcal{X}, \mu)$ et $g \in \mathbb{L}_{\theta^*}(\mathcal{X}, \mu)$, alors

$$fg \in \mathbb{L}_1(\mathcal{X}, \mu) \qquad \text{et} \qquad \int_{\mathcal{X}} |fg| \, d\mu \leq 2\|f\|_\theta \|g\|_{\theta^*}. \tag{III.47}$$

Par suite, un élément de $\mathbb{L}_{\theta^*}(\mathcal{X}, \mu)$ peut être vu comme une application linéaire continue sur $\mathbb{L}_\theta(\mathcal{X}, \mu)$. En général, le dual topologique de $\mathbb{L}_\theta(\mathcal{X}, \mu)$ est strictement plus gros que $\mathbb{L}_{\theta^*}(\mathcal{X}, \mu)$. En revanche, on a la proposition suivante :

Proposition III.48. *Si θ est une fonction de Young partout finie, alors le dual topologique de $\mathbb{L}^a_\theta(\mathcal{X}, \mu)$ peut être identifié à $\mathbb{L}_{\theta^*}(\mathcal{X}, \mu)$, c'est-à-dire que pour toute forme linéaire continue ℓ sur $\mathbb{L}^a_\theta(\mathcal{X}, \mu)$, il existe une unique fonction $g_\ell \in \mathbb{L}_{\theta^*}(\mathcal{X}, \mu)$ telle que*

$$\forall f \in \mathbb{L}^a_\theta(\mathcal{X}, \mu), \quad \ell(g) = \int_{\mathcal{X}} f g_\ell \, d\mu.$$

Dans ce qui suit, nous considérerons les espaces d'Orlicz $\mathcal{L}_\tau(\mathcal{X}, \mu)$ et $\mathcal{L}^a_\tau(\mathcal{X}, \mu)$ associés à la fonction $\tau(x) = e^{|x|} - 1 - |x|$. Pour tout $\ell \in \mathbb{L}_\tau(\mathcal{X}, \mu)'$, nous noterons

$$\|\ell\|^*_\tau = \sup\{\langle \ell, g\rangle : \quad g \in \mathbb{L}_\tau(\mathcal{X}, \mu), \|g\|_\tau \leq 1\}.$$

Dans la suite, nous supposerons que $G = \mathcal{L}_\tau(\mathcal{X}, \mu)$ et nous noterons $\mathcal{P}_\tau(\mathcal{X})$ à la place de $\mathcal{P}_{\mathcal{L}_\tau(\mathcal{X},\mu)}(\mathcal{X})$.

Si $\nu \in \mathcal{P}_\tau(\mathcal{X})$ est absolument continue par rapport à μ, alors l'application

$$\mathbb{L}_\tau(\mathcal{X}, \mu) \to \mathbb{R} : g \mapsto \int_{\mathcal{X}} g \, d\nu$$

est bien définie et est linéaire. Le lemme suivant donne une condition nécessaire et suffisante pour que cette forme soit continue :

Lemme III.49. *Une probabilité $\nu \in \mathcal{P}_\tau(\mathcal{X})$ absolument continue par rapport à μ est une forme linéaire continue sur $\mathbb{L}_\tau(\mathcal{X}, \mu)$ si, et seulement si,* $\mathrm{H}(\nu|\,\mu) < +\infty$.

Démonstration. Si $\nu \in \mathbb{L}_\tau(\mathcal{X}, \mu)'$, alors la restriction de ν à $\mathbb{L}^a_\tau(\mathcal{X}, \mu)$ appartient à $\mathbb{L}^a_\tau(\mathcal{X}, \mu)'$ et, d'après la proposition III.48, $\mathbb{L}^a_\tau(\mathcal{X}, \mu)' \simeq \mathbb{L}_{\tau^*}(\mathcal{X}, \mu)$, où $\tau^* = x\log(x) + 1 - x$. Il existe donc $h \in \mathbb{L}_{\tau^*}(\mathcal{X}, \mu)$ telle que

$$\forall g \in \mathbb{L}^a_\tau(\mathcal{X}, \mu), \quad \int_{\mathcal{X}} g \, d\nu = \int_{\mathcal{X}} gh \, d\mu,$$

et on en déduit que $\nu = h\mu$. Comme h appartient à $\mathbb{L}_{\tau^*}(\mathcal{X}, \mu)$, il existe $t > 0$ tel que

$$\int_{\mathcal{X}} (th)\log(th) + 1 - th \, d\mu = t\,\mathrm{H}(\nu|\,\mu) + 1 - t + t\log(t) < +\infty,$$

et donc $\mathrm{H}(\nu|\,\mu) < +\infty$.
Réciproquement, si $\nu \in \mathcal{P}_\tau(\mathcal{X})$ est telle que $\mathrm{H}(\nu|\,\mu) < +\infty$, alors $h = \frac{d\nu}{d\mu} \in \mathbb{L}_{\tau^*}(\mathcal{X}, \mu)$.
D'après l'inégalité (III.47), on a donc

$$\forall g \in \mathbb{L}_\tau(\mathcal{X}, \mu), \qquad \left|\int_{\mathcal{X}} g\, d\nu\right| = \left|\int_{\mathcal{X}} gh\, d\mu\right| \leq 2\|g\|_\tau \|h\|_{\tau^*},$$

ce qui prouve que $\nu \in \mathbb{L}_\tau(\mathcal{X}, \mu)'$. □

Remarque III.50.

En particulier, $\mu^n_{C_n}$ appartient à $\mathbb{L}_\tau(\mathcal{X}, \mu)'$.

La proposition suivante donne une condition suffisante pour que $\mu^n_{C_n}$ converge vers μ^* au sens de la norme $\|\,.\,\|^*_\tau$:

Proposition III.51. *Sous les hypothèses du théorème III.36, notons $h_n = \frac{d\mu^*_n}{d\mu}$ et supposons que $(h_n)_n$ soit une suite bornée de $\mathbb{L}_p(\mathcal{X}, \mu)$ pour un certain $p > 1$, alors $\left\|\mu^n_{C_n} - \mu^*\right\|^*_\tau \xrightarrow[n\to+\infty]{} 0$.*

Démonstration. Soient ν_1 et ν_2 deux éléments de $\mathcal{P}_\tau(\mathcal{X})$ et $g \in \mathbb{L}_\tau(\mathcal{X}, \mu)$ telle que $\|g\|_\tau \leq 1$. Tout d'abord,

$$\left|\int_{\mathcal{X}} g\, d\nu_1 - \int_{\mathcal{X}} g\, d\nu_2\right| \leq \Big\|\, |g|\nu_1 - |g|\nu_2 \,\Big\|_{VT}$$

D'après (II.16), pour tout $\delta > 0$, on a

$$\Big\|\, |g|\nu_1 - |g|\nu_2 \,\Big\|_{VT} \leq \frac{C}{\delta}\left(1 + \log \int_{\mathcal{X}} e^{\delta|g|}\, d\nu_2\right)\left(\sqrt{\mathrm{H}(\nu_1|\,\nu_2)} + \mathrm{H}(\nu_1|\,\nu_2)\right), \quad \text{(III.52)}$$

où C est une constante numérique.
Prenons $\nu_2 = h\mu$, avec $\left[\int_{\mathcal{X}} h^p d\mu\right]^{1/p} \leq M$, alors, d'après l'inégalité de Hölder,

$$\int_{\mathcal{X}} e^{\delta|g|}\, d\nu_2 \leq M\left[\int_{\mathcal{X}} e^{p'\delta|g|}\, d\mu\right]^{1/p'}, \quad \text{(III.53)}$$

avec p'tel que $\frac{1}{p} + \frac{1}{p'} = 1$. Comme $\|g\|_\tau \leq 1$, on a $\int_{\mathcal{X}} e^{|g|} - 1 - |g|\, d\mu \leq 1$, donc

$$\begin{aligned}\int_{\mathcal{X}} e^{|g|}\, d\mu \leq 2 + \int_{\mathcal{X}} |g|\, d\mu &\overset{(i)}{\leq} 2 + \sqrt{2}\sqrt{\int_{\mathcal{X}} \frac{g^2}{2}\, d\mu} \\ &\overset{(ii)}{\leq} 2 + \sqrt{2}\sqrt{\int_{\mathcal{X}} e^{|g|} - 1 - |g|\, d\mu} \leq 2 + \sqrt{2} \leq 4,\end{aligned}$$

(i) venant de l'inégalité de Cauchy-Schwarz et (ii) de l'inégalité $\frac{x^2}{2} \leq e^{|x|} - 1 - |x|$. Ainsi, en prenant $\delta = \frac{1}{p'}$ dans (III.53), on a $\displaystyle\int_{\mathcal{X}} e^{|g|/p'}\, d\nu_2 \leq 4^{1/p'} M$ et (III.52) donne

$$\left\| |g|\nu_1 - |g|\nu_2 \right\|_{VT} \leq p'C \left(1 + \log(4^{1/p'} M)\right) \left(\sqrt{H(\nu_1|\nu_2)} + H(\nu_1|\nu_2)\right).$$

Par conséquent, pour toute $\nu_2 \in \mathcal{P}_\tau(\mathcal{X})$ telle que $\displaystyle\left[\int_{\mathcal{X}} \left(\frac{d\nu_2}{d\mu}\right)^p d\mu\right]^{1/p}$, on a

$$\forall \nu_1 \in \mathcal{P}_\tau(\mathcal{X}),$$
$$\|\nu_1 - \nu_2\|_\tau^* \leq p'C \left(1 + \log(4^{1/p'} M)\right) \left(\sqrt{\mathrm{H}\,(\nu_1|\,\nu_2)} + \mathrm{H}\,(\nu_1|\,\nu_2)\right). \quad \text{(III.54)}$$

Pour démontrer la proposition, il suffit de reprendre mot à mot la preuve du théorème III.36, avec $k = 1$, en appliquant en (III.40) l'inégalité (III.54) (avec $\nu_1 = \mu^n_{C_n}$ et $\nu_2 = \mu^*_n$) à la place de l'inégalité de Pinsker. □

III.3 Conditionnement par des contraintes de type moment

Dans cette section, $G = \mathcal{L}_\tau(\mathcal{X}, \mu)$ et nous nous intéresserons à un conditionnement défini par une contrainte de type moment, *ie* l'ensemble C sera de la forme

$$C = \left\{\nu \in \mathcal{P}_\tau(\mathcal{X}) : \int_{\mathcal{X}} F\, d\nu \in K\right\}$$

avec F une application mesurable à valeurs dans un espace de Banach séparable $(B, \|\,.\,\|)$ telle que $\|F\| \in \mathcal{L}_\tau(\mathcal{X}, \mu)$ et K un convexe fermé de B.

Nous grossirons C de la manière suivante :

$$C_n = \left\{\nu \in \mathcal{P}_\tau(\mathcal{X}) : \int_{\mathcal{X}} F\, d\nu \in K^{\varepsilon_n}\right\},$$

où $K^{\varepsilon_n} = \{x \in B : d(x, C) \leq \varepsilon_n\}$ et ε_n est une suite de réels positifs décroissant lentement vers 0. Les théorèmes III.61 et III.76 donnent des vitesses explicites pour ε_n dans un cadre fini-dimensionnel et infini-dimensionnel.

Notations. Nous désignerons par μ_F l'image de μ par l'application F. Le support de μ_F sera noté S_F. La transformée de Laplace de μ_F sera notée Z_F ; elle est définie par

$$\forall \lambda \in B', \quad Z_F(\lambda) = \int_{\mathcal{X}} e^{\langle \lambda, F \rangle} d\mu.$$

Enfin, on notera Λ_F la Log-Laplace de μ_F définie par $\Lambda_F := \log Z_F$.

Pour montrer la condition

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n}(L_n \in C_n) \geq -\operatorname{H}(C|\,\mu)\,, \tag{III.55}$$

nous utiliserons la borne inférieure exacte donnée par la proposition suivante.

Lemme III.56. *Si μ admet une I-projection μ^* sur C de la forme $\mu^* = \frac{e^{\langle \lambda^*, F\rangle}}{Z_F(\lambda^*)}\mu$, avec $\lambda^* \in B'$, alors pour tout $\varepsilon > 0$,*

$$\frac{1}{n} \log \left(\mu^{\otimes n}(L_n \in C_\varepsilon) e^{n \operatorname{H}(\mu^*|\mu)}\right) \geq \frac{1}{n} \log \mathbb{P}\left(\left\| \frac{1}{n} \sum_{i=1}^{n} F(Y_i) - \int_{\mathcal{X}} F d\mu^* \right\| \leq \varepsilon\right) - \|\lambda^*\|\varepsilon. \tag{III.57}$$

avec $(Y_i)_i$ une suite de variables i.i.d de loi μ^*.

Démonstration. Voir l'annexe A. □

Remarque III.58.
Pour obtenir (III.55), il suffit d'après l'inégalité (III.57) de montrer que

$$\frac{1}{n} \log \mathbb{P}\left(\left\| \frac{1}{n} \sum_{i=1}^{n} F(Y_i) - \int_{\mathcal{X}} F d\mu^* \right\| \leq \varepsilon_n\right) \xrightarrow[n\to+\infty]{} 0, \tag{III.59}$$

Cette dernière condition est strictement plus faible que la condition

$$\mathbb{P}\left(\left\| \frac{1}{n} \sum_{i=1}^{n} F(Y_i) - \int_{\mathcal{X}} F d\mu^* \right\| \leq \varepsilon_n\right) \xrightarrow[n\to+\infty]{} 1$$

du corollaire III.46.

III.3.1 Cas d'un espace de dimension finie

Dans cette section, nous supposerons que B est de dimension finie et nous noterons q sa dimension. Nous travaillerons sous les hypothèses suivantes.

Hypothèse III.60.

1. $\operatorname{dom} \Lambda_F := \{\lambda \in B' : \Lambda_F(\lambda) < +\infty\}$ est ouvert dans B',
2. L'enveloppe convexe du support de μ_F, $\operatorname{co} S_F$ est d'intérieur non vide,
3. K est un convexe fermé de B tel que $K \cap \operatorname{co} \overset{\circ}{S}_F \neq \emptyset$.

Ces hypothèses vont nous permettre d'utiliser les résultats de la section II.4.4 sur la représentation des projections entropiques.

Théorème III.61. *Sous les hypothèses III.60,*

1. *μ possède une I-projection μ^* sur $C = \left\{\nu \in \mathcal{P}_\tau(\mathcal{X}) : \int_\mathcal{X} F\, d\nu \in K\right\}$.*
2. *Il existe $\bar{c} \in \mathbb{R}^+$ tel que pour toute suite $\varepsilon_n \in \mathbb{R}^+$ de limite nulle telle que $\lim\limits_{n\to+\infty} n\varepsilon_n^2 \in]\bar{c}, +\infty]$, la suite $\mu^n_{C_n,k}$ converge en variation vers $\mu^{*\otimes k}$, pour tout $k \in \mathbb{N}^*$, où $C_n = \left\{\nu \in \mathcal{P}_\tau(\mathcal{X}) : \int_\mathcal{X} F\, d\nu \in K^{\varepsilon_n}\right\}$.*
3. *De plus, $\left\|\mu^n_{C_n} - \mu^*\right\|^*_\tau \xrightarrow[n\to+\infty]{} 0$.*
4. *Enfin, pour tout k, $\mu^n_{C_n,k}$ converge en entropie vers $\mu^{*\otimes k}$.*

Pour démontrer (III.59), nous ferons appel à l'inégalité de Bernstein donnée par le theorème suivant :

Théorème III.62. *Si $Y_1, \dots, Y_n$ sont des variables aléatoires réelles indépendantes centrées, telles qu'il existe $M > 0$ et $v_1, \dots, v_n > 0$ tels que*

$$\mathbb{E}\left[|Y_i|^m\right] \leq \frac{m!}{2} M^{m-2} v_i,$$

alors, pour tout $t > 0$,

$$\mathbb{P}(Y_1 + \cdots + Y_n \geq t) \leq \exp\left(-\frac{1}{2}\frac{t^2}{v + tM}\right), \quad \textit{avec} \quad v = v_1 + \cdots + v_n.$$

Démonstration. Voir par exemple [71], 2.2.11 p.103. □

Corollaire III.63. *Soit Y_i une suite de variables aléatoires réelles indépendantes, de moyenne nulle, alors*

$$\forall t \geq 0, \quad \mathbb{P}\left(\frac{Y_1 + \cdots + Y_n}{n} \geq t\right) \leq \exp\left(-\frac{nt^2}{2M(2M+t)}\right), \tag{III.64}$$

avec $M = \inf\left\{\lambda \geq 0 : \forall i = 1 \dots n, \quad \mathbb{E}\left[\tau\left(\frac{Y_i}{\lambda}\right)\right] \leq 1\right\}$, où $\tau(x) = e^{|x|} - 1 - |x|$.

Démonstration. Si $M = +\infty$, l'inégalité est vraie.
Si $M < +\infty$, alors pour tout $i = 1 \dots n$, on a pour tout $m \geq 2$:

$$\frac{\mathbb{E}\left[|Y_i|^m\right]}{M^m m!} \leq \mathbb{E}\left[\tau\left(\frac{Y_i}{M}\right)\right] \leq 1$$

et donc

$$\mathbb{E}\left[|Y_i|^m\right] \leq \frac{m!}{2} M^{m-2} v_i,$$

avec $v_i = 2M^2$. Donc, d'après le théorème III.62, on a

$$\mathbb{P}(Y_1 + \cdots + Y_n \geq nt) \leq \exp\left(-\frac{nt^2}{2M(2M+t)}\right).$$

$\square$

D'après le théorème III.36, nous aurons également besoin de certaines propriétés de continuité des I-projections par rapport au grossissement ; celles-ci sont démontrées dans le lemme ci-dessous.

Lemme III.65. *Sous les hypothèses III.60,*

1. Λ_F *est strictement convexe,*
2. μ *admet une I-projection* μ^* *sur* C*, qui s'écrit* $\mu^* = \dfrac{e^{\langle \lambda^*, F\rangle}}{Z_F(\lambda^*)}\mu$*, avec* λ^**, l'unique minimisant de la fonction* $H(\lambda) = \Lambda_F(\lambda) - \inf\limits_{y \in K} \langle \lambda, y\rangle$*,*
3. μ *admet une I-projection* μ_n^* *sur* C_n*, qui s'écrit* $\mu_n^* = \dfrac{e^{\langle \lambda_n^*, F\rangle}}{Z_F(\lambda_n^*)}\mu$*, avec* λ_n^**, l'unique minimisant de la fonction* $H_n(\lambda) = \Lambda_F(\lambda) - \inf\limits_{y \in K^{\varepsilon_n}} \langle \lambda, y\rangle$*,*
4. *De plus,* $\lim\limits_{n\to+\infty} \lambda_n^* = \lambda^*$ *et* $\lim\limits_{n\to+\infty} \mathrm{H}\left(C_n | \, \mu\right) = \mathrm{H}\left(C | \, \mu\right)$.

Démonstration.
1. Si $\lambda_1, \lambda_2 \in \operatorname{dom} \Lambda_F$, en posant $g(t) = \Lambda_F(t\lambda_1 + (1-t)\lambda_2)$, pour $t \in [0,1]$, on voit facilement que $g''(t) = \displaystyle\int \left((\lambda_2 - \lambda_1)(x) - \int (\lambda_2 - \lambda_1)(y)\, d\tilde{\mu}_F(y)\right)^2 d\tilde{\mu}_F(x)$, avec $\widetilde{\mu}_F \sim \mu_F$. Par suite, $g''(t) = 0$ si, et seulement si, $\lambda_2 - \lambda_1$ est constante sur $\operatorname{co} S_F$. Comme $\operatorname{co} S_F$ est supposé d'intérieur non vide, cela entraîne $\lambda_1 = \lambda_2$, et Λ_F est donc strictement convexe sur son domaine.

2. Par hypothèse, $K \cap \overset{\circ}{\operatorname{co} S_F} \neq \emptyset$, donc $\operatorname{ri} K \cap \overset{\circ}{\operatorname{co} S_F} \neq \emptyset$. D'après le théorème II.41, μ possède une I-projection μ^* sur

$$\widetilde{C} = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} \|F\|\, d\nu < +\infty \quad \text{et} \quad \int_{\mathcal{X}} F\, d\nu \in K \right\},$$

mais, d'après le corollaire II.35, μ^* est la I-projection généralisée de μ sur C. Comme $\mathrm{H}\left(\mu^* | \, \mu\right) < +\infty$, le point 1 de la proposition II.34 entraîne que μ^* appartient à $\mathcal{P}_\tau(\mathcal{X})$ et donc μ^* est la I-projection de μ sur C. De plus, d'après le théorème II.41, $\mu^* = \frac{e^{\langle \lambda^*, F\rangle}}{Z_F(\lambda^*)}\mu$, avec $\lambda^* \in \operatorname{Argmin} H$. Comme Λ_F est strictement convexe, il en est de même pour H qui n'admet donc qu'un seul minimisant.

3. Idem.

4. Clairement,
$$H_n(\lambda) = H(\lambda) + \varepsilon_n \|\lambda\|.$$
On en déduit que $\operatorname{dom} H_n = \operatorname{dom} H$ et que H_n converge simplement vers H sur $\operatorname{dom} H$. Admettons un instant que la suite $(\lambda_n^*)_n$ soit bornée et considérons une valeur d'adhérence $\bar{\lambda}$ de $(\lambda_n^*)_n$ ainsi qu'une sous-suite $(\lambda_{n_k}^*)_k$ convergeant vers $\bar{\lambda}$.
Pour tout k,
$$H_{n_k}(\lambda_{n_k}^*) = \inf_{\lambda \in B'} H_{n_k}(\lambda) \leq H_{n_k}(\lambda^*),$$
donc par convergence simple,
$$\limsup_{k \to +\infty} H_{n_k}(\lambda_{n_k}^*) \leq \lim_{k \to +\infty} H_{n_k}(\lambda^*) = H(\lambda^*) \tag{III.66}$$
De plus, par semi-continuité inférieure de H :
$$H(\bar{\lambda}) \leq \liminf_{k \to +\infty} H_{n_k}(\lambda_{n_k}^*). \tag{III.67}$$
De (III.66) et (III.67), on déduit que
$$H(\bar{\lambda}) \leq H(\lambda^*).$$
Comme H n'admet qu'un seul minimisant, on a nécessairement $\bar{\lambda} = \lambda^*$. La suite $(\lambda_n^*)_n$ est une suite bornée admettant λ^* pour seule valeur d'adhérence ; elle converge donc vers λ^*. En particulier, (III.66) et (III.67) sont valables pour $n_k = k$ et par conséquent, $\lim_{n \to +\infty} \inf H_n = \inf H$. Ceci entraîne, d'après le théorème II.41, que $\mathrm{H}\left(C_n \middle| \, \mu\right)$ converge vers $\mathrm{H}\left(C \middle| \, \mu\right)$.

Montrons à présent que la suite $(\lambda_n^*)_n$ est bornée. Comme $\operatorname{co} \overset{\circ}{S_F} = \operatorname{dom} \overset{\circ}{\Lambda_F^*}$ (voir la remarque II.42), il existe $x_0 \in K \cap \operatorname{dom} \overset{\circ}{\Lambda_F^*}$. Posons $\widetilde{H}(\lambda) = \Lambda_F(\lambda) - \langle \lambda, x_0 \rangle$. On a clairement $\widetilde{H} \leq H \leq H_n$. Comme $H_{n+1} \leq H_n$, la suite $\inf H_n$ est décroissante. Donc H_n atteint son minimum sur $\left\{ H_n \leq \inf H_1 + 1 \right\} \subset \left\{ \widetilde{H} \leq \inf H_1 + 1 \right\}$. Il suffit donc de montrer que pour tout $k \in \mathbb{R}$, $\left\{ \widetilde{H} \leq k \right\}$ est borné.
Or,
$$\left\{ \widetilde{H} \leq k \right\} = \left\{ \lambda \in B' : \forall x \in B, \quad \langle \lambda, x \rangle \leq k + \widetilde{H}^*(x) \right\}$$
Mais $\widetilde{H}^*(x) = \Lambda_F^*(x + x_0)$ et donc $0 \in \operatorname{dom} \overset{\circ}{\widetilde{H}^*}$. Une fonction convexe étant continue sur l'intérieur de son domaine, on en déduit que si $r > 0$ est tel que $B(0, r) \subset \operatorname{dom} \overset{\circ}{\widetilde{H}^*}$, on a pour tout $\lambda \in \left\{ \widetilde{H} \leq k \right\}$,
$$\sup_{\|x\| \leq r} \langle \lambda, x \rangle \leq k + \sup_{\|x\| \leq r} \widetilde{H}^*(x) < +\infty,$$

et donc $\left\{\widetilde{H} \leq k\right\}$ est borné. $\square$

Démonstration du théorème III.61 :
1. C'est le point 2. du lemme III.65.
2. D'après le théorème III.36, et le point 4. du lemme III.65, il suffit de montrer que

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n}(L_n \in C_n) \geq -\operatorname{H}(C|\,\mu)$$

D'après la borne inférieure exacte (III.57) du lemme III.56, si $(Y_i)_i$ est une suite i.i.d de loi μ^*,

$$\liminf_{n\to+\infty} \frac{1}{n} \log \left(\mu^{\otimes n}(L_n \in C_n) e^{n\operatorname{H}(\mu^*|\mu)}\right) \geq \liminf_{n\to+\infty} \frac{1}{n} \log \mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{n} F(Y_i) - \int_{\mathcal{X}} F d\mu^*\right\| \leq \varepsilon_n\right).$$

Soit $(e_1, \ldots, e_q)$ une base de B ; notons $f_1, \ldots, f_q$ les composantes de F sur cette base. Par équivalence des normes en dimension finie, il existe $m_1, m_2 > 0$ tels que

$$m_1 \max_{j=1\ldots q} |x_j| \leq \|x\| \leq m_2 \max_{j=1\ldots q} |x_j|.$$

On a donc

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{n} F(Y_i) - \int_{\mathcal{X}} F\, d\mu^*\right\| \leq \varepsilon_n\right) &\geq \mathbb{P}\left(\sup_{j=1\ldots q}\left|\frac{1}{n}\sum_{i=1}^{n} f_j(X_i) - \mathbb{E}\left[f_j(X_1)\right]\right| \leq \frac{1}{m_2}\varepsilon_n\right) \\
&\geq 1 - q \max_{j=1\ldots q} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} f_j(X_i) - \mathbb{E}\left[f_j(X_1)\right]\right| \geq \frac{1}{m_2}\varepsilon_n\right).
\end{aligned}$$

Comme $m_1 \max_{j=1\ldots q} |f_j| \leq \|F\|$, pour pouvoir appliquer l'inégalité de Bernstein (III.64), il suffit de montrer que $\|F\| \in \mathbb{L}_\tau(\mathcal{X}, \mu^*)$.
Or, d'après la formule de représentation du théorème II.41 et l'inégalité de Hölder, on a pour tout $p > 1$

$$\begin{aligned}
\int_{\mathcal{X}} e^{t\|F\|}\, d\mu^* &= \frac{1}{Z_F(\lambda^*)} \int_{\mathcal{X}} e^{t\|F\|} e^{\langle \lambda^*, F\rangle}\, d\mu \\
&\leq \frac{1}{Z_F(\lambda^*)} \left[\int_{\mathcal{X}} e^{tp'\|F\|}\, d\mu\right]^{\frac{1}{p'}} \left[\int_{\mathcal{X}} e^{\langle p\lambda^*, F\rangle}\, d\mu\right]^{\frac{1}{p}},
\end{aligned} \tag{III.68}$$

avec p' tel que $\frac{1}{p}+\frac{1}{p'} = 1$. Comme $\operatorname{dom}\Lambda_F$ est ouvert, il existe $p > 1$ tel que $p\lambda^* \in \operatorname{dom}\Lambda_F$. Pour un tel p, le membre de droite de (III.68) est fini pour tout t assez petit, puisque $\|F\| \in \mathbb{L}_\tau(\mathcal{X}, \mu)$.
Soit $M = \max_{j=1\ldots q} \left\|f_j - \int_{\mathcal{X}} f_j\, d\mu^*\right\|_{\mathbb{L}_\tau(\mathcal{X},\mu^*)}$, alors d'après (III.64), on a

$$\max_{j=1\ldots q} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} f_j(X_i) - \mathbb{E}[f_j(X_1)]\right| \geq \frac{1}{m_2}\varepsilon_n\right) \leq 2\exp\left(-\frac{n(\varepsilon_n/m_2)^2}{2M(2M + \varepsilon_n/m_2)}\right),$$

Donc

$$\liminf_{n\to+\infty} \frac{1}{n} \log \left(\mu^{\otimes n}(L_n \in C_n) e^{n\,\mathrm{H}(\mu^*|\mu)} \right) \geq \liminf_{n\to+\infty} \frac{1}{n} \log \left(1 - 2q \exp\left(-\frac{n(\varepsilon_n/m_2)^2}{2M(2M+\varepsilon_n/m_2)} \right) \right) \tag{III.69}$$

Posons $\bar{c} = (2m_2M)^2 \log(2q)$ et supposons que $c := \lim_{n\to+\infty} n\varepsilon_n^2 > \bar{c}$, alors

$$\frac{n(\varepsilon_n/m_2)^2}{2M(2M+\varepsilon_n/m_2)} \xrightarrow[n\to+\infty]{} \frac{c}{(2m_2M)^2}.$$

Comme $2qe^{-\frac{c}{(2m_2M)^2}}$ <1, on en déduit que le membre de droite de (III.69) est nul.

3. D'après la proposition III.51, il suffit de montrer qu'il existe $p > 1$ tel que $\left(\frac{d\mu_n^*}{d\mu}\right)_n$ soit bornée dans $\mathbb{L}_p(\mathcal{X}, \mu)$.
Or,

$$\int_{\mathcal{X}} \left(\frac{d\mu_n^*}{d\mu} \right)^p d\mu = \frac{Z_F(p\lambda_n^*)}{Z_F(\lambda_n^*)^p}.$$

Comme λ_n^* converge vers λ^* (lemme III.65), $Z_F(\lambda_n^*)$ est bornée. Par hypothèse, $\mathrm{dom}\,\Lambda_F$ est ouvert ; il existe donc $p > 1$ et $r > 0$ tels que $B(p\lambda^*, r) \subset \mathrm{dom}\,\Lambda_F$. Il existe alors n_0 tel que pour tout $n \geq n_0$, $p\lambda_n^* \in B(p\lambda^*, r)$, et donc

$$Z_F(p\lambda_n^*) \leq \sup_{x\in B(p\lambda^*,r)} \Lambda_F(x) < +\infty.$$

4. Montrons enfin la convergence en entropie :

$$\begin{aligned}
\mathrm{H}\left(\mu_{C_n,k}^n \middle| \mu_n^{*\otimes k}\right) &= \mathrm{H}\left(\mu_{C_n,k}^n \middle| \mu^{*\otimes k}\right) + \int_{\mathcal{X}} \log \frac{d\mu^{*\otimes k}}{d\mu_n^{*\otimes k}}\, d\mu_{C_n,k}^n \\
&= \mathrm{H}\left(\mu_{C_n,k}^n \middle| \mu^{*\otimes k}\right) + k \int_{\mathcal{X}} \log \frac{d\mu^*}{d\mu_n^*}\, d\mu_{C_n}^n \\
&= \mathrm{H}\left(\mu_{C_n,k}^n \middle| \mu^{*\otimes k}\right) + k\,\mathrm{H}\left(\mu^* \middle| \mu_n^*\right) + k \int_{\mathcal{X}} \log \frac{d\mu^*}{d\mu_n^*}\, d\big(\mu_{C_n}^n - \mu^*\big)\,.
\end{aligned}$$

On a vu dans la preuve du théorème III.36 que $\mathrm{H}\left(\mu_{C_n,k}^n \middle| \mu_n^{*\otimes k}\right)$ et $\mathrm{H}\left(\mu^* \middle| \mu_n^*\right)$ convergeaient vers 0. Il suffit donc de montrer que

$$J_n := \int_{\mathcal{X}} \log \frac{d\mu^*}{d\mu_n^*}\, d\big(\mu_{C_n}^n - \mu^*\big) \xrightarrow[n\to+\infty]{} 0.$$

Or,

$$J_n = \int_{\mathcal{X}} f_n\, d\big(\mu_{C_n}^n - \mu^*\big) \qquad \text{avec} \qquad f_n = \langle \lambda_n^* - \lambda^*, F \rangle.$$

Comme $\left\|\mu_{C_n}^n - \mu^*\right\|_\tau^* \xrightarrow[n\to+\infty]{} 0$, il suffit de montrer que la suite f_n est bornée dans $\mathbb{L}_\tau(\mathcal{X}, \mu)$. Comme pour, n assez grand, $\|\lambda_n^* - \lambda^*\| \leq 1$, on a $|f_n| \leq \|\lambda_n^* - \lambda^*\| \|F\| \leq \|F\|$ et donc $\|f_n\|_\tau \leq \left\| \, \|F\| \, \right\|_\tau < +\infty$. La suite $(f_n)_n$ est donc bien bornée dans $\mathbb{L}_\tau(\mathcal{X}, \mu)$. $\square$

En dimension 1, on peut améliorer la vitesse de rétrécissement ε_n :

Proposition III.70. *Si $B = \mathbb{R}$, les conclusions du théorème III.61 restent valables pour $\varepsilon_n = \frac{1}{n^a}$, avec $0 < a < 1$.*

Démonstration. Le cas $a < \frac{1}{2}$ relève du théorème III.61. On supposera donc que $a \in [1/2, 1[$. En reprenant les notations précédentes, il suffit de démontrer que

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_i\right| \leq \varepsilon_n\right) = 0, \tag{III.71}$$

avec $Z_i = F(Y_i) - \int_{\mathcal{X}} F\, d\mu^*$, Y_i i.i.d de loi μ^*, et $\varepsilon_n = \frac{1}{n^a}$, $a \in [1/2, 1[$.
On voit facilement que $\mathbb{E}\left[|Z_1|^k\right] < +\infty$ pour tout $k \in \mathbb{N}$. Notons $\sigma^2 = \mathbb{E}\left[|Z_1|^2\right]$, $\kappa = \mathbb{E}\left[|Z_1|^3\right]$ et R_n la fonction de répartition de $\dfrac{1}{\sqrt{n}\sigma}\displaystyle\sum_{i=1}^{n} Z_i$.
D'après l'inégalité de Berry-Esseen (voir par exemple le théorème 2.1.30 de [63]), on a en notant Φ la fonction de répartition de la loi gaussienne centrée réduite

$$\sup_{x\in\mathbb{R}} |R_n(x) - \Phi(x)| \leq 10\frac{\kappa}{\sqrt{n}\sigma^3} \tag{III.72}$$

Donc

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_i\right| \leq \varepsilon_n\right) &= R_n\left(\frac{\sqrt{n}\varepsilon_n}{\sigma}\right) - R_n\left(-\frac{\sqrt{n}\varepsilon_n}{\sigma}\right) \\
&\geq \Phi\left(\frac{\sqrt{n}\varepsilon_n}{\sigma}\right) - \Phi\left(-\frac{\sqrt{n}\varepsilon_n}{\sigma}\right) - 20\frac{\kappa}{\sqrt{n}\sigma^3} \\
&= \frac{2}{\sqrt{n}}\left[\frac{\sqrt{n}}{\sqrt{2\pi}}\int_0^{\frac{\sqrt{n}\varepsilon_n}{\sigma}} e^{-u^2/2}\, du - 10\frac{\kappa}{\sigma^3}\right] \\
&\geq \frac{2}{\sqrt{n}}\left[\frac{n\varepsilon_n}{\sqrt{2\pi}} e^{-n\varepsilon_n^2/2\sigma^2} - 10\frac{\kappa}{\sigma^3}\right] := \alpha_n.
\end{aligned}$$

On voit facilement que, pour $\varepsilon_n = \frac{1}{n^a}$ avec $a \in [1/2, 1[$, on a $\alpha_n \sim \frac{2}{\sqrt{2\pi}\sigma} n^{\frac{1}{2}-a}$. Par conséquent, $\displaystyle\lim_{n\to+\infty} \frac{1}{n}\log(\alpha_n) = 0$, ce qui prouve (III.71). □

III.3.2 Cas d'un espace de dimension infinie

Convergence en variation

Nous travaillerons sous les hypothèses suivantes :

Hypothèse III.73.

1. B est un espace de Banach séparable de type 2, *ie* il existe $a > 0$ tel que pour toute suite $(Z_i)_i$ de variables aléatoires indépendantes centrées et de carré intégrable, on ait
$$\mathbb{E}\left[\|Z_1 + \cdots + Z_n\|^2\right] \leq a\left[\mathbb{E}\left[\|Z_1\|^2\right] + \cdots + \mathbb{E}\left[\|Z_n\|^2\right]\right]. \qquad \text{(III.74)}$$
2. Le domaine de Λ_F est ouvert.
3. K est un convexe fermé de B tel que la fonction
$$H(\lambda) = \Lambda_F(\lambda) - \inf_{y \in K} \langle \lambda, y \rangle$$
atteigne son minimum.

Remarque III.75.

- L'hypothèse 1. nous sera utile pour utiliser le théorème de Yurinskii ; elle est bien sûr réalisée si B est un espace de Hilbert.
- L'hypothèse 3. est en particulier réalisée si $K = \{x_0\}$, avec $x_0 = \nabla \Lambda_F(\lambda_0)$.
- D'après le lemme II.39, l'hypothèse 3. précédente garantit que μ admet une I-projection μ^* sur
$$\widetilde{C} = \left\{\nu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} \|F\|\, d\nu < +\infty \quad \text{et} \quad \int_{\mathcal{X}} F\, d\nu \in K\right\}$$
qui s'écrit $\mu^* = \frac{e^{\langle \lambda^*, F\rangle}}{Z_F(\lambda^*)}\mu$, pour tout λ^* minimisant H.
- D'après le corollaire II.35, μ^* est aussi la I-projection de μ sur
$$C := \left\{\nu \in \mathcal{P}_\tau(\mathcal{X}) : \int_{\mathcal{X}} F\, d\nu \in K\right\}.$$

Théorème III.76. *Soit $\varepsilon_n = \frac{c}{\sqrt{n}}$, avec $c > \sqrt{a \operatorname{Var}_{\mu^*}(F)}$ où a est la constante de (III.74) et $C_n = \left\{\nu \in \mathcal{P}_\tau(\mathcal{X}) : \int_{\mathcal{X}} F\, d\nu \in K^{\varepsilon_n}\right\}$. Sous les hypothèses III.73, $\mu^n_{C_n,k}$ converge en variation vers $\mu^{*\otimes k}$ pour tout $k \in \mathbb{N}^*$.*

Nous utiliserons le théorème suivant dû à Yurinskii :

Théorème III.77 (Yurinskii, [73], théorème 2.1)**.** *Soit $(Z_i)_i$ une suite de variables aléatoires indépendantes à valeur dans B telle qu'il existe b et $M > 0$ tels que, pour tout $i \in \mathbb{N}^*$, on ait :*

$$\forall k \geq 2, \quad \mathbb{E}\left[\|Z_i\|^k\right] \leq \frac{k!}{2} b^2 M^{k-2} \tag{III.78}$$

Alors, en posant $S_n = \sum_{i=1}^{n} Z_i$,

$$\forall t > 0, \quad \mathbb{P}\left(\|S_n\| \geq \mathbb{E}\left[\|S_n\|\right] + nt\right) \leq \exp\left(-\frac{1}{8}\frac{nt^2}{b^2 + tM}\right). \tag{III.79}$$

Démonstration du théorème III.76 :
D'après le théorème III.36, il suffit de vérifier que $\mathrm{H}\left(C_n|\,\mu\right)$ converge vers $\mathrm{H}\left(C|\,\mu\right)$ et que $\liminf_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n}(L_n \in C_n) \geq -\mathrm{H}\left(C|\,\mu\right)$.
Montrons que $\lim_{n\to+\infty} \mathrm{H}\left(C_n|\,\mu\right) = \mathrm{H}\left(C|\,\mu\right)$:
Tout d'abord, d'après le lemme II.39, $\mathrm{H}\left(C|\,\mu\right) = -\inf H$. De plus, on voit facilement grâce au théorème de Hahn-Banach que $\int_{\mathcal{X}} F\, d\mu^* \in \overline{\mathrm{co}\, S_F}$. Par conséquent, $K \cap \overline{\mathrm{co}\, S_F} \neq \emptyset$, et *a fortiori*, $\overset{\circ}{K^{\varepsilon_n}} \cap \mathrm{co}\, S_F \neq \emptyset$. Le théorème II.41 entraîne donc en particulier que $\mathrm{H}\left(C_n|\,\mu\right) = -\inf H_n$, avec $H_n(\lambda) = \Lambda_F(\lambda) - \inf_{y \in K^{\varepsilon_n}} \langle \lambda, y\rangle$. Comme $(H_n)_n$ converge simplement en décroissant vers H sur $\mathrm{dom}\, H$, on a

$$\inf H \leq \inf H_n \leq H_n(\lambda^*) \xrightarrow[n\to+\infty]{} \inf H$$

et donc $\lim_{n\to+\infty} \inf H_n = \inf H$.
Montrons que $\liminf_{n\to+\infty} \frac{1}{n} \log \mu^{\otimes n}(L_n \in C_n) \geq -\mathrm{H}\left(C|\,\mu\right)$:
D'après le lemme III.56, il suffit de montrer que si $(Y_i)_i$ est une suite i.i.d de loi μ^*,

$$\lim_{n\to+\infty} \frac{1}{n} \log \mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{n} F(Y_i) - \int_{\mathcal{X}} F\, d\mu^*\right\| \leq \varepsilon_n\right) = 0. \tag{III.80}$$

En raisonnant comme dans la preuve du théorème III.61, on voit que $\|F\| \in \mathbb{L}_\tau(\mathcal{X}, \mu^*)$. On voit alors facilement que (III.78) est valable pour $Z_i = Y_i - \int_{\mathcal{X}} F\, d\mu^*$, avec $M = \left\|F - \int_{\mathcal{X}} F\, d\mu^*\right\|_{\mathbb{L}_\tau(\mathcal{X},\mu^*)}$ et $b = \sqrt{2}M$. Comme B est supposé être de type 2, $\mathbb{E}[\|S_n\|] \leq \sqrt{\mathbb{E}[\|S_n\|^2]} \leq \sqrt{an}\sigma$, avec $\sigma = \sqrt{\mathbb{E}[\|Z_1\|^2]}$, de sorte que, d'après (III.79),

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{n} F(Y_i) - \int_{\mathcal{X}} F\, d\mu^*\right\| \leq \frac{\sqrt{a}\sigma}{\sqrt{n}} + t\right) \geq 1 - \exp\left(-\frac{1}{8} n \frac{t^2}{2M^2 + tM}\right).$$

Ainsi, en prenant $\varepsilon_n = \frac{c}{\sqrt{n}}$, avec $c > \sqrt{a}\sigma$, (III.80) est vérifiée. □

Convergence forte dans $\mathbb{L}_\tau(\mathcal{X},\mu)'$?

On fera les hypothèses suivantes :

Hypothèse III.81.

1. B est un espace de Banach séparable de type 2,
2. $G = \mathbb{L}_\tau^a(\mathcal{X},\mu)$ et $\|F\| \in \mathbb{L}_\tau^a(\mathcal{X},\mu)$, *ie* $\forall t > 0, \quad \int_{\mathcal{X}} e^{t\|F\|}\, d\mu < +\infty,$
3. K est un convexe fermé de B tel que la fonction
$$H(\lambda) = \Lambda_F(\lambda) - \inf_{y\in K}\langle \lambda, y\rangle$$
atteigne son minimum.
4. Il existe une suite $(\lambda_n^*)_n$ <u>bornée dans B'</u>, telle que, pour tout n, λ_n^* minimise
$$H_n(\lambda) = \Lambda_F(\lambda) - \inf_{y\in K^{\varepsilon_n}}\langle \lambda, y\rangle.$$

Théorème III.82. *Sous les hypothèses précédentes, les conclusions du théorème III.61 sont valables pour toute suite $\varepsilon_n = \frac{c}{\sqrt{n}}$ avec $c > \sqrt{a\,\mathrm{Var}_{\mu^*}(F)}$ où a est la constante de (III.74).*

Démonstration.
Par rapport au théorème III.76, la seule chose nouvelle à vérifier est que $\left(\frac{d\mu_n^*}{d\mu}\right)_n$ est bornée dans $\mathbb{L}_p(\mathcal{X},\mu)$ pour un certain $p > 1$. Si $M > 0$ est tel que $\forall n \in \mathbb{N},\ \|\lambda_n^*\| \le M$, alors on a pour tout $p > 1$

$$\int_{\mathcal{X}} \left(\frac{d\mu_n^*}{d\mu}\right)^p d\mu = \frac{\int_{\mathcal{X}} e^{p\langle \lambda_n^*, F\rangle}\, d\mu}{\left(\int_{\mathcal{X}} e^{\langle \lambda^*, F\rangle}\, d\mu\right)^p} \le \frac{\int_{\mathcal{X}} e^{pM\|F\|}\, d\mu}{\left(\int_{\mathcal{X}} e^{-M\|F\|}\, d\mu\right)^p} < +\infty.$$

□

Remarque III.83.

Nous ne connaissons pas de condition suffisante raisonnable dans un espace de dimension infinie garantissant l'hypothèse 4. précédente. Lorsque B est de dimension finie, nous avons vu dans la preuve du point 4 du lemme III.65 (page 72) que la bornitude de la suite λ_n^* était vraie sous des hypothèses assez faibles. Malheureusement, les arguments que nous avons utilisés pour démontrer cette propriété ne sont plus valables en dimension infinie.

III.4 Contraintes plus générales - Contrôles par recouvrement.

Pour aborder des conditionnements définis par des contraintes plus générales que celles prises en compte dans les sections précédentes, nous allons développer une méthode basée sur les nombres de recouvrement. Dans toute la suite, $(\mathcal{X}, d)$ sera un espace polonais. L'ensemble $\mathcal{P}(X)$ des mesures de probabilité sur $\mathcal{X}$ sera muni de la topologie de la convergence étroite, *ie* $G = C_b(\mathcal{X})$ (l'ensemble des fonctions continues bornées sur $\mathcal{X}$) et de la tribu borélienne associée à cette topologie.

III.4.1 Nombres de recouvrement

Définition III.84. Soit K une partie compacte d'un espace métrique $(\mathcal{Y}, d)$. Pour tout $\varepsilon > 0$, *le nombre de recouvrement de* K *de niveau* ε, noté $N_{\mathcal{Y}}(d, K, \varepsilon)$, est le nombre minimal de boules ouvertes de rayon ε nécessaire pour recouvrir K.
Autrement dit,

$$N_{\mathcal{Y}}(d, K, \varepsilon) = \inf \left\{ p \in \mathbb{N}^* : \quad \exists B_1, \dots B_p, \text{ boules de rayon } \varepsilon \text{ tq } K \subset \bigcup_{i=1}^{p} B_i \right\}$$

Les propositions suivantes donnent des exemples classiques d'estimation des nombres de recouvrement :

Proposition III.85. *Soit* B *une boule fermée de rayon* $r > 0$ *dans* $\mathbb{R}^q$ *muni de la distance euclidienne* d*, alors*

$$\forall \varepsilon > 0, \quad N_{\mathbb{R}^q}(d, B, \varepsilon) \leq \left(1 + 2\frac{r}{\varepsilon}\right)^q.$$

En particulier,

$$\forall \varepsilon \leq r, \quad N_{\mathbb{R}^q}(d, B, \varepsilon) \leq 3^q \left(\frac{r}{\varepsilon}\right)^q.$$

Démonstration. Voir par exemple le théorème II.4 du chapitre VII de [75]. □

Dans la proposition suivante, que nous utiliserons à la fin de ce chapitre, on s'intéresse au recouvrement d'une boule hölderienne :

Proposition III.86. *Soit* $\mathcal{X}$ *l'ensemble des fonctions continues de* $[0, 1]$ *dans* $\mathbb{R}^q$ *; posons pour tout* $R, M > 0$ *et* $\alpha \in]0, 1]$

$$K(R, M, \alpha) = \left\{ x \in \mathcal{X} : |x(0)| \leq R \quad \textit{et} \quad \sup_{s \neq t} \frac{\|x(s) - x(t)\|}{|s - t|^\alpha} \leq M \right\},$$

alors

$$\forall \varepsilon > 0, \quad N_{\mathcal{X}}(\| \, . \, \|_\infty, K(R, M, \alpha), \varepsilon) \leq c_1(\alpha, q) \left(\frac{R}{\varepsilon}\right)^q \exp\left(c_2(\alpha, q)\left(\frac{M}{\varepsilon}\right)^{\frac{q}{\alpha}}\right).$$

Démonstration. Voir le théorème 2.7.1 page 155 de [71]. □

III.4.2 $\mathcal{P}(\mathcal{X})$ en tant qu'espace métrique.

Afin de calculer des nombres de recouvrement sur $\mathcal{P}(\mathcal{X})$, nous devons équiper cet ensemble d'une distance métrisant la convergence étroite. Nous considérerons deux distances classiques sur $\mathcal{P}(X)$: les distances de Prokhorov et de Fortet-Mourier.

Les distances de Prokhorov et de Fortet-Mourier.

La distance de Fortet-Mourier, que nous noterons $d_{FM}(\,.\,,\,.\,)$, est définie de la manière suivante :

$$\forall \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}), \quad d_{FM}(\nu_1, \nu_2) = \sup_{\substack{\varphi \in \mathrm{BLip}(\mathcal{X}, d) \\ \|\varphi\|_{BL} \leq 1}} \left\{ \int_{\mathcal{X}} \varphi \, d\nu_1 - \int_{\mathcal{X}} \varphi \, d\nu_2 \right\}, \tag{III.87}$$

où $\mathrm{BLip}(\mathcal{X}, d)$ est l'ensemble des fonctions Lipschitziennes bornées sur $\mathcal{X}$, et

$$\|\varphi\|_{BL} = \|\varphi\|_\infty + \|\varphi\|_{Lip},$$

avec

$$\|\varphi\|_{Lip} = \sup_{x \neq y} \left\{ \frac{|\varphi(x) - \varphi(y)|}{d(x, y)} \right\}.$$

La distance de Prokhorov, que nous noterons $d_P(\,.\,,\,.\,)$, est définie par

$$\forall \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}),$$

$$d_P(\nu_1, \nu_2) = \inf \left\{ \alpha > 0 : \sup_{A \text{ borélien}} \{\nu_1(A) - \nu_2(A^\alpha)\} \leq \alpha \right\}, \tag{III.88}$$

où $A^\alpha = \{x \in \mathcal{X} : d(x, A) \leq \alpha\}$.
La proposition suivante donne un résultat de comparaison entre d_P, d_{FM} et $\|\,.\,\|_{VT}$.

Proposition III.89. *Pour toutes $\nu_1, \nu_2 \in \mathcal{P}(\mathcal{X})$, on a en posant $\phi(t) = \frac{2t^2}{2+t}$*

$$\phi\left(d_P(\nu_1, \nu_2)\right) \leq d_{FM}(\nu_1, \nu_2) \leq 2 d_P(\nu_1, \nu_2), \tag{III.90}$$

et

$$d_{FM}(\nu_1, \nu_2) \leq \|\nu_1 - \nu_2\|_{VT} \quad \textit{et} \quad d_P(\nu_1, \nu_2) \leq \frac{1}{2} \|\nu_1 - \nu_2\|_{VT}. \tag{III.91}$$

Démonstration. Pour l'inégalité (III.90), voir le problème 5 p.312 et le corollaire II.6.5 du chapitre 11 de [29]. L'inégalité $d_{FM}(\nu_1, \nu_2) \leq \|\nu_1 - \nu_2\|_{VT}$ est immédiate.
Montrons que $d_P(\nu_1, \nu_2) \leq \frac{1}{2}\|\nu_1 - \nu_2\|_{VT}$. Pour tout $\alpha > 0$, on a

$$\sup_{A \text{ borélien}} \{\nu_1(A) - \nu_2(A^\alpha)\} \leq \sup_{A \text{ borélien}} \{\nu_1(A) - \nu_2(A)\} = \frac{1}{2}\|\nu_1 - \nu_2\|_{VT}.$$

En prenant $\alpha = \frac{1}{2}\|\nu_1 - \nu_2\|_{VT}$, et en revenant à la définition de d_P, on en déduit le résultat. □

Notation : Dans la suite, $\bar{d}$ désignera l'une ou l'autre des distances précédemment définies. Rappelons le résultat classique suivant :

Théorème III.92. *Si $(\mathcal{X}, d)$ est un espace polonais, $\bar{d}$ définie par (III.87) ou (III.88) est une distance métrisant la topologie de la convergence étroite et $(\mathcal{P}(\mathcal{X}), \bar{d})$ est un espace polonais. Si de plus $(\mathcal{X}, d)$ est un espace métrique compact, il en est de même de $(\mathcal{P}(\mathcal{X}), \bar{d})$*

Démonstration. Voir par exemple le chapitre 11 de [29]. □

Estimation des nombres de recouvrement de $\mathcal{P}(\mathcal{X})$.

Notations : Lorsque $(\mathcal{X}, d)$ est compact, nous noterons plus simplement $N_{\mathcal{X}}(\varepsilon)$ à la place de $N_{\mathcal{X}}(d, \mathcal{X}, \varepsilon)$ et $N_{\mathcal{P}(\mathcal{X})}(\bar{d}, \varepsilon)$ à la place de $N_{\mathcal{P}(\mathcal{X})}(\bar{d}, \mathcal{P}(\mathcal{X}), \varepsilon)$ (d'après le théorème III.92, $(\mathcal{P}(\mathcal{X}), \bar{d})$ est compact).

Une question naturelle est d'estimer $N_{\mathcal{P}(\mathcal{X})}(\bar{d}, \varepsilon)$ en fonction de $N_{\mathcal{X}}(\varepsilon)$, dans le cas où $(\mathcal{X}, d)$ est compact. Le lemme suivant est du à S.R. Kulkarni et O. Zeitouni.

Lemme III.93 (Kulkarni-Zeitouni, [42], lemme 1)**.** *Si $(\mathcal{X}, d)$ est un espace métrique compact, on a pour tout $\varepsilon > 0$,*

$$N_{\mathcal{P}(\mathcal{X})}(d_P, \varepsilon) \leq \left(\frac{2e}{\varepsilon}\right)^{N_{\mathcal{X}}(\varepsilon)}. \tag{III.94}$$

Grâce à l'inégalité (III.90), on voit que $B_P(\nu, \frac{\varepsilon}{2}) \subset B_{FM}(\nu, \varepsilon)$; on en déduit immédiatement le

Lemme III.95. *Si $(\mathcal{X}, d)$ est un espace métrique compact, on a pour tout $\varepsilon > 0$,*

$$N_{\mathcal{P}(\mathcal{X})}(d_{FM}, \varepsilon) \leq \left(\frac{4e}{\varepsilon}\right)^{N_{\mathcal{X}}(\varepsilon/2)}.$$

Remarque III.96.

D'après les lemmes précédents, l'inégalité

$$\forall \varepsilon > 0, \quad N_{\mathcal{P}(\mathcal{X})}(\bar{d}, \varepsilon) \leq \left(\frac{4e}{\varepsilon}\right)^{N_{\mathcal{X}}(\varepsilon/2)} \tag{III.97}$$

est valable pour $\bar{d} = d_P$ et $\bar{d} = d_{FM}$. Pour éviter un traitement séparé des deux métriques, nous utiliserons toujours la majoration (III.97) même si, dans le cas de la distance de Prokhorov, celle-ci est un peu moins fine que (III.94).

A titre indicatif, nous montrons ci-dessous comment, en s'inspirant des techniques de [42], on peut obtenir directement une version un peu moins précise du lemme III.95.
Preuve directe du lemme III.95. Soit $\varepsilon > 0$. Posons $p = N_{\mathcal{X}}(\varepsilon)$, et considérons $B_1, \ldots, B_p$, p boules de rayon ε recouvrant $\mathcal{X}$.
Pour tout $i = 1 \ldots p$, posons $A_i = B_i - (A_1 \cup \ldots \cup A_{i-1})$. Les A_i sont tous non vides (sinon cela contredirait la minimalité de p) et forment une partition de $\mathcal{X}$. On choisit dans chaque A_i un point x_i et on note δ_i, la masse de Dirac centrée en x_i.
Pour tout entier n, posons :

$$\mathcal{Y}_n = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \nu = a_1\delta_1 + \cdots + a_p\delta_p, a_i \in \left\{ 0, \frac{1}{n}, \frac{2}{n}, \ldots, 1 \right\} \right\}$$

On voit facilement que le cardinal de $\mathcal{Y}_n$ est C_{n+p-1}^{p-1}.
De l'inégalité $n! > e\left(\frac{n-1}{e}\right)^{n-1}$, on déduit pour $p \geq 2$ et $n \geq p$:

$$\begin{aligned} C_{n+p-1}^{p-1} &= \frac{(n+p-1)\cdots(n+1)}{(p-1)!} \leq \frac{(n+p-1)^{p-1}}{(p-1)!} \\ &< \frac{(n+p-1)^{p-1}}{e\left(\frac{p-1}{e}\right)^{p-1}} = e^{p-2}\left(\frac{n}{p} + \frac{p-1}{p}\right)^{p-1}\left(\frac{p}{p-1}\right)^{p-1} \\ &< e^{p-2}\left(2\frac{n}{p}\right)^{p-1} 2^{p-1} \leq \left(\frac{4en}{p}\right)^p \end{aligned}$$

Ainsi :

$$|\mathcal{Y}_n| \leq \left(\frac{4en}{p}\right)^p$$

Soit $\gamma \in \mathcal{P}(\mathcal{X})$. Pour tout $i = 1 \ldots p-1$, il existe un unique $a_i \in \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ tel que $a_i \leq \gamma(A_i) \leq a_i + \frac{1}{n}$; posons $a_p = 1 - (a_1 + \cdots + a_{p-1})$ et définissons $\nu = a_1\delta_1 + \cdots + a_p\delta_p$.

Si φ est une fonction 1-Lipschitzienne telle que $|\varphi| \leq 1$, on a

$$\begin{aligned} &\left| \int_{\mathcal{X}} \varphi\, d\gamma - \int_{\mathcal{X}} \varphi\, d\nu \right| = \left| \sum_{i=1}^p \int_{A_i} \varphi\, d\gamma - \int_{A_i} \varphi\, d\nu \right| \\ &= \left| \sum_{i=1}^p \int_{A_i} [\varphi(x) - \varphi(x_i)]\; d\gamma + \varphi(x_i)[\gamma(A_i) - a_i] \right| \\ &\leq \sum_{i=1}^p \int_{A_i} |\varphi(x) - \varphi(x_i)|\, d\gamma(x) + \sum_{i=1}^p |\varphi(x_i)|\, |\gamma(A_i) - a_i| \\ &\leq 2\varepsilon \sum_{i=1}^p \gamma(A_i) + \sum_{i=1}^{p-1} [\gamma(A_i) - a_i] + |\gamma(A_p) - a_p| = 2\varepsilon + 2\sum_{i=1}^{p-1}[\gamma(A_i) - a_i] \\ &\leq 2\varepsilon + 2\frac{p-1}{n} \end{aligned}$$

En prenant pour $\varepsilon \leq 1$, $n = E(p/\varepsilon) > 0$, on obtient :

$$d_{FM}(\gamma, \nu) \leq 4\varepsilon$$

et

$$|\mathcal{Y}_n| \leq \left(\frac{4e}{\varepsilon}\right)^p$$

on en déduit

$$N_{\mathcal{P}(\mathcal{X})}(d_{FM}, \varepsilon) \leq \left(\frac{16e}{\varepsilon}\right)^{N_{\mathcal{X}}(\varepsilon/4)}$$

□

III.4.3 Le cas compact

Dans cette sous-section, $(\mathcal{X}, d)$ est un espace métrique compact. Pour tout A ensemble mesurable de $\mathcal{P}(\mathcal{X})$, nous noterons pour tout $\varepsilon > 0$,

$$A^{\varepsilon} := \{\nu \in \mathcal{P}(\mathcal{X}) : \ \bar{d}(\nu, A) \leq \varepsilon.\}.$$

La proposition suivante est démontrée dans [42] :

Proposition III.98. *Soit A une partie mesurable de $\mathcal{P}(\mathcal{X})$. Pour tout $\nu \in \mathcal{P}(\mathcal{X})$, on a :*

$$\forall \varepsilon > 0, \quad \nu^{\otimes n}(L_n \in A) \leq N_{\mathcal{P}(\mathcal{X})}(\bar{d}, A, \varepsilon) e^{-n\,\mathrm{H}\left(\,A^{2\varepsilon}|\nu\right)}.$$

Démonstration. L'espace $(\mathcal{P}(\mathcal{X}), \bar{d})$ étant compact, A est une partie totalement bornée de $\mathcal{P}(\mathcal{X})$. Soit $\varepsilon > 0$; posons $p = N_{\mathcal{P}(\mathcal{X})}(\bar{d}, A, \varepsilon)$ et considérons $B_1, \ldots, B_p$ des boules fermées de rayon ε recouvrant A.
On a clairement

$$\nu^{\otimes n}(L_n \in A) \leq \sum_{i=1}^{p} \nu^{\otimes n}(L_n \in B_i)$$

Or, d'après le théorème III.37, pour tout ensemble convexe fermé B, on a

$$\nu^{\otimes n}(L_n \in B) \leq e^{-n\,\mathrm{H}(\,B|\nu)}.$$

Les boules B_i étant convexes et fermées, on en déduit que pour tout $i = 1 \ldots p$,

$$\nu^{\otimes n}(L_n \in B_i) \leq e^{-n\,\mathrm{H}(\,B_i|\nu)},$$

et comme $B_i \subset A^{2\varepsilon}$, $\mathrm{H}\left(\,B_i|\,\nu\right) \geq \mathrm{H}\left(\,A^{2\varepsilon}|\,\nu\right)$. □

Corollaire III.99. *Soient C un convexe fermé de $\mathcal{P}(\mathcal{X})$, tel que* $\mathrm{H}\left(\,C|\,\mu\right) < +\infty$, *et μ^* la I-projection de μ sur C. Alors, pour tout $n \in \mathbb{N}^*$,*

$$\mu^{*\otimes n}(L_n \in C^{\varepsilon}) \geq 1 - N_{\mathcal{P}(\mathcal{X})}\left(\bar{d}, \frac{\varepsilon}{4}\right) e^{-n\frac{\varepsilon^2}{8}}. \tag{III.100}$$

Démonstration. En notant $B(\mu^*, \varepsilon)$ la boule ouverte de rayon ε, on a

$$\mu^{*\otimes n}(L_n \in C^\varepsilon) \geq \mu^{*\otimes n}(L_n \in B(\mu^*, \varepsilon)) = 1 - \mu^{*\otimes n}(L_n \in B(\mu^*, \varepsilon)^c).$$

D'après la proposition III.98, pour tout $\xi > 0$, on a

$$\mu^{*\otimes n}(L_n \in B(\mu^*, \varepsilon)^c) \leq N_{\mathcal{P}(\mathcal{X})}(B(\mu^*, \varepsilon)^c, \xi) e^{-n \operatorname{H}\left(B(\mu^*, \varepsilon)^{c\,2\xi} \middle| \mu^* \right)}.$$

Prenons $\xi = \frac{\varepsilon}{4}$, alors

$$N_{\mathcal{P}(\mathcal{X})}\left(\bar{d}, B(\mu^*, \varepsilon)^c, \frac{\varepsilon}{4}\right) \leq N_{\mathcal{P}(\mathcal{X})}\left(\bar{d}, \frac{\varepsilon}{4}\right) \quad \text{et} \quad B(\mu^*, \varepsilon)^{c\,\varepsilon/2} = B\left(\mu^*, \frac{\varepsilon}{2}\right)^c.$$

Or, pour tout $\nu \in B\left(\mu^*, \frac{\varepsilon}{2}\right)^c$, d'après le point 2 de la proposition III.89 et l'inégalité de Pinsker (II.13), on a

$$\begin{aligned}
&\text{si } \bar{d} = d_{FM}, && \operatorname{H}(\nu | \mu^*) \geq \frac{1}{2}\|\nu - \mu^*\|_{VT}^2 \geq \frac{1}{2} d_{FM}(\nu, \mu^*)^2 \geq \frac{\varepsilon^2}{8}, \\
&\text{si } \bar{d} = d_P, && \operatorname{H}(\nu | \mu^*) \geq \frac{1}{2}\|\nu - \mu^*\|_{VT}^2 \geq 2 d_P(\nu, \mu^*)^2 \geq \frac{\varepsilon^2}{2},
\end{aligned}$$

donc, dans les deux cas,

$$\mu^{*\otimes n}(L_n \in B(\mu^*, \varepsilon)^c) \leq N_{\mathcal{P}(\mathcal{X})}\left(\bar{d}, \frac{\varepsilon}{4}\right) e^{-n\frac{\varepsilon^2}{8}}.$$

□

Corollaire III.101. *Soient C un convexe fermé de $\mathcal{P}(\mathcal{X})$ tel que $\operatorname{H}(C | \mu) < +\infty$, et μ^* la I-projection de μ sur C. Pour toute suite $(\varepsilon_n)_n$ de réels strictement positifs de limite nulle telle que $N_{\mathcal{P}(\mathcal{X})}\left(\bar{d}, \frac{\varepsilon_n}{4}\right) e^{-n\frac{\varepsilon_n^2}{8}} \xrightarrow[n\to+\infty]{} 0$, on a $\mu^n_{C^{\varepsilon_n}, k} \xrightarrow[n\to+\infty]{} \mu^{*\otimes k}$ en variation dans $\mathcal{P}(\mathcal{X}^k)$.*

Démonstration. D'après le corollaire III.99,

$$\mu^{*\otimes n}(L_n \in C^{\varepsilon_n}) \geq 1 - N_{\mathcal{P}(\mathcal{X})}\left(\bar{d}, \frac{\varepsilon_n}{4}\right) e^{-n\frac{\varepsilon_n^2}{8}},$$

et donc $\mu^{*\otimes n}(L_n \in C^{\varepsilon_n}) \xrightarrow[n\to+\infty]{} 1$. On conclut en utilisant le corollaire III.14. □

En utilisant la majoration (III.97) de $N_{\mathcal{P}(\mathcal{X})}(\bar{d}, \varepsilon)$, on obtient sans peine le

Corollaire III.102. *Si $\varepsilon_n > 0$ est une suite de limite nulle telle que*

$$\frac{n\varepsilon_n^2}{8} + \log(\varepsilon_n) N_{\mathcal{X}}\left(\frac{\varepsilon_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty, \tag{III.103}$$

alors $\mu^n_{C^{\varepsilon_n}, k}$ converge en variation vers $\mu^{*\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$.

La condition (III.103) est assez simple à utiliser pour déterminer des vitesses de rétrécissement explicites :

Proposition III.104. *Si pour tout ε assez petit, $N_{\mathcal{X}}(\varepsilon) \leq \frac{\alpha}{\varepsilon^q}$, alors on peut prendre $\varepsilon_n = \frac{1}{n^a}$, pour tout $0 < a < \frac{1}{q+2}$.*

Démonstration.

$$\begin{aligned}\frac{n\varepsilon_n^2}{8} + \log(\varepsilon_n) N_{\mathcal{X}}\left(\frac{\varepsilon_n}{8}\right) &\geq \frac{n\varepsilon_n^2}{8} + \alpha 8^q \log(\varepsilon_n)\frac{1}{\varepsilon_n^q} \\ &= n^{1-2a}\left(\frac{1}{8} - \alpha 8^q a \log(n) n^{a(q+2)-1}\right) \xrightarrow[n\to+\infty]{} +\infty\end{aligned}$$

□

D'après la proposition III.85, le résultat précédent s'applique en particulier si $\mathcal{X}$ est un compact de $\mathbb{R}^q$. Dès que l'on dispose d'une estimation explicite des nombres de recouvrement $N_{\mathcal{X}}(\varepsilon)$ (et la littérature sur le sujet est assez abondante), on peut calculer des vitesses de rétrécissement ε_n explicites. Le point fort du critère (III.103) est qu'il est toujours applicable, comme le montre le résultat théorique suivant :

Proposition III.105. *Pour tout espace métrique compact $(\mathcal{X}, d)$, il existe au moins une suite $(\varepsilon_n)_n$ décroissante à valeurs dans $[0, 1[$ telle que*

$$\frac{n\varepsilon_n^2}{8} + \log(\varepsilon_n) N_{\mathcal{X}}\left(\frac{\varepsilon_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty,$$

Démonstration. En posant $N(\varepsilon) = 8N_{\mathcal{X}}\left(\frac{\varepsilon}{8}\right)$, il s'agit de montrer qu'il existe une suite ε_n telle que

$$n\varepsilon_n^2 + \log(\varepsilon_n) N(\varepsilon_n) \xrightarrow[n\to+\infty]{} +\infty.$$

Considérons la fonction

$$f :]0, 1] \to \mathbb{R}^+ : \varepsilon \mapsto -\frac{\log(\varepsilon) N(\varepsilon)}{\varepsilon^2}.$$

Clairement, f est décroissante et $\lim\limits_{\varepsilon\to 0^+} f(\varepsilon) = +\infty$. Soit $(u_n)_n$ une suite décroissante à valeurs dans $]0, 1]$ telle que $nu_n^2 \xrightarrow[n\to+\infty]{} +\infty$; la suite $w_n := f(u_n)$ est croissante et tend vers $+\infty$.
Pour tout n, notons :

$$k_n = \max\left\{k \in \mathbb{N}^* \text{ tel que } w_k \leq \sqrt{n}\right\}.$$

Pour n assez grand, k_n est bien défini.
Premier cas :
Supposons que pour tout n assez grand, $k_n \leq n$, et posons :

$$\varepsilon_n = u_{k_n} \text{ pour tout } n \in [k_n, k_{n+p_n}[, \text{ avec } p_n := \inf\{p \geq 1 \text{ tq } k_{n+p} > k_n\}.$$

Alors, pour n assez grand, on a d'une part :

$$n\varepsilon_n^2 = nu_{k_n}^2 \geq k_n u_{k_n}^2 \xrightarrow[n\to+\infty]{} +\infty,$$

et d'autre part,

$$n\varepsilon_n^2 + \log(\varepsilon_n)N(\varepsilon_n) = n\varepsilon_n^2\left(1 - \frac{w_{k_n}}{n}\right) \geq n\varepsilon_n^2\left(1 - \frac{1}{\sqrt{n}}\right) \xrightarrow[n\to+\infty]{} +\infty.$$

Second Cas :
Supposons *a contrario*, qu'il existe une suite p_i strictement croissante telle que $k_{p_i} \geq p_i$. Cela revient à supposer qu'il existe une suite p_i telle que pour tout i, $w_{p_i} \leq \sqrt{p_i}$. Pour tout n, soit $\phi(n)$ l'unique entier tel que $n \in [p_{\phi(n)}, p_{\phi(n)+1}[$; posons $\varepsilon_n = u_{p_{\phi(n)}}$, on a alors

$$n\varepsilon_n^2 \geq p_{\phi(n)} u_{p_{\phi(n)}}^2 \xrightarrow[n\to+\infty]{} +\infty,$$

et

$$\begin{aligned} n\varepsilon_n^2 + \log(\varepsilon_n)N(\varepsilon_n) = n\varepsilon_n^2\left(1 - \frac{w_{p_{\phi(n)}}}{n}\right) &\geq n\varepsilon_n^2\left(1 - \frac{\sqrt{p_{\phi(n)}}}{n}\right) \\ &\geq n\varepsilon_n^2\left(1 - \frac{1}{\sqrt{p_{\phi(n)}}}\right) \xrightarrow[n\to+\infty]{} +\infty. \end{aligned}$$

□

III.4.4 Extension au cas non-compact

Résultats généraux

Dans cette section, $(\mathcal{X}, d)$ sera un espace polonais quelconque. Pour étendre les résultats de la section précédente, notre stratégie est, en un mot, de se ramener au cas compact en invoquant le caractère tendu d'une probabilité sur un espace polonais.

Proposition III.106. *Soient C un convexe fermé de $\mathcal{P}(\mathcal{X})$ tel que $\mathrm{H}\left(C|\,\mu\right) < +\infty$, et μ^* la I-projection de μ sur C. Pour tout compact K inclus dans $\mathcal{X}$, on a pour tout $\xi > 0$,*

$$\mu^{*\otimes n}\left(\bar{d}(L_n, C) \leq \xi + 2\mu^*(K^c)\right) \geq \mu^*(K)^n\left(1 - \left(\frac{16e}{\xi}\right)^{N_K\left(\frac{\xi}{8}\right)} e^{-n\frac{\xi^2}{8}}\right) \tag{III.107}$$

En particulier, s'il existe une suite $(K_n)_n$ de compacts inclus dans $\mathcal{X}$ et une suite $\xi_n > 0$ de limite nulle telles que :

$$\mu^*(K_n)^n \xrightarrow[n\to+\infty]{} 1 \qquad et \qquad \frac{n\xi_n^2}{8} + \log(\xi_n)N_{K_n}\left(\frac{\xi_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty, \tag{III.108}$$

alors, pour toute suite $(\varepsilon_n)_n$ de limite nulle telle que $\varepsilon_n \geq \xi_n + 2\mu^*(K_n^c)$, la suite $\mu^n_{C^{\varepsilon_n},k}$ converge en variation vers $\mu^{*\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$.

Démonstration. Posons $\mu_K^* := \dfrac{\mathbb{1}_K}{\mu^*(K)}\mu^*$.
On a

$$\begin{aligned}\bar{d}(\mu_K^*, \mu^*) \leq \|\mu_K^* - \mu^*\|_{VT} &= \int_{\mathcal{X}} \left| \frac{\mathbb{1}_K}{\mu^*(K)} - 1 \right| d\mu^* \\ &= \left(\frac{1}{\mu^*(K)} - 1\right)\mu^*(K) + \mu^*(K^c) = 2\mu^*(K^c),\end{aligned}$$

donc, d'après l'inégalité triangulaire,

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \bar{d}(\nu, \mu^*) \leq \bar{d}(\nu, \mu_K^*) + 2\mu^*(K^c).$$

Par conséquent,

$$B(\mu_K^*, \xi) \subset \left\{\nu \in \mathcal{P}(\mathcal{X}) : \bar{d}(\nu, C) \leq \xi + 2\mu^*(K^c)\right\},$$

et

$$\begin{aligned}\mu^{*\otimes n}\left(\bar{d}(L_n, C) \leq \xi + 2\mu^*(K^c)\right) &\geq \mu^{*\otimes n}\left(L_n \in B(\mu_K^*, \xi)\right) \\ &\geq \mu^{*\otimes n}\left(L_n \in B(\mu_K^*, \xi) \quad \text{et} \quad \forall i = 1\ldots n, \quad x_i \in K\right) \\ &= \mu^*(K)^n \mu_K^{*\otimes n}\left(L_n \in B(\mu_K^*, \xi)\right).\end{aligned}$$

D'après le corollaire III.99 et (III.97), on a :

$$\mu_K^{*\otimes n}\left(L_n \in B(\mu_K^*, \xi)\right) \geq 1 - N_{\mathcal{P}(K)}\left(\bar{d}, \xi/4\right) e^{-n\frac{\xi^2}{8}} \geq 1 - \left(\frac{16e}{\xi}\right)^{N_K\left(\frac{\xi}{8}\right)} e^{-n\frac{\xi^2}{8}},$$

ce qui démontre (III.107).
Si $(K_n)_n$ et $(\xi_n)_n$ vérifient (III.108), alors $\mu^{*\otimes n}(L_n \in C^{\varepsilon_n}) \xrightarrow[n\to+\infty]{} 1$, ce qui entraîne, d'après le corollaire III.14, que $\mu^n_{C^{\varepsilon_n},k}$ converge en variation vers $\mu^{*\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$. □

Sous des hypothèses plus contraignantes sur $\frac{d\mu^*}{d\mu}$, le critère (III.108) peut être un peu affaibli :

Proposition III.109. *Soient C un convexe fermé de $\mathcal{P}(\mathcal{X})$ tel que $\mathrm{H}\left(C|\,\mu\right) < +\infty$, et μ^* la I-projection de μ sur C. Si $\log \frac{d\mu^*}{d\mu}$ est continue et bornée sur $\mathcal{X}$, et s'il existe une suite $(K_n)_n$ de compact inclus dans $\mathcal{X}$, et une suite $\xi_n > 0$ de limite nulle telles que :*

$$\mu^*(K_n) \xrightarrow[n\to+\infty]{} 1 \qquad \text{et} \qquad \frac{n\xi_n^2}{8} + \log(\xi_n) N_{K_n}\left(\frac{\xi_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty, \qquad \text{(III.110)}$$

alors pour toute suite ε_n de limite nulle telle que $\varepsilon_n \geq \xi_n + 2\mu^*(K_n^c)$, la suite $\mu^n_{C^{\varepsilon_n},k}$ converge en variation vers $\mu^{*\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$.

Démonstration. Posons $h = \log \frac{d\mu^*}{d\mu}$; pour tout $\varepsilon > 0$, on a :

$$\begin{aligned}
\mu^{\otimes n}(L_n \in C^\varepsilon) &\geq \mu^{\otimes n}(L_n \in B(\mu^*, \varepsilon)) \\
&= \int_{\mathcal{X}^n} \mathbb{1}_{B(\mu^*,\varepsilon)}(L_n) e^{-n\langle L_n, h\rangle} \, d\mu^{*\otimes n} \\
&= e^{-n\,\mathrm{H}(C|\mu)} \int_{\mathcal{X}^n} \mathbb{1}_{B(\mu^*,\varepsilon)}(L_n) e^{-n\langle L_n - \mu^*, h\rangle} \, d\mu^{*\otimes n} \\
&\geq e^{-n\,\mathrm{H}(C|\mu)} e^{-n\Delta(\varepsilon)} \mu^{*\otimes n}\left(L_n \in B(\mu^*, \varepsilon)\right),
\end{aligned}$$

en notant

$$\Delta(\varepsilon) = \sup_{\nu \in B(\mu^*,\varepsilon)} \langle \nu - \mu^*, h\rangle.$$

Ainsi

$$\frac{1}{n} \log\left(\mu^{\otimes n}(L_n \in C^\varepsilon) e^{n\,\mathrm{H}(C|\mu)}\right) \geq -\Delta(\varepsilon) + \frac{1}{n} \log \mu^{*\otimes n}\big(L_n \in B(\mu^*, \varepsilon)\big). \qquad \text{(III.111)}$$

L'application $\nu \mapsto \int_{\mathcal{X}} h \, d\nu$ étant continue en μ^*, on voit sans peine que $\Delta(\varepsilon) \xrightarrow[\varepsilon \to 0]{} 0$. Par conséquent, si ε_n est une suite de $\mathbb{R}^+$ de limite nulle, on a

$$\liminf_{n\to+\infty} \frac{1}{n} \log\left(\mu^{\otimes n}(L_n \in C^{\varepsilon_n}) e^{n\,\mathrm{H}(C|\mu)}\right) \geq \liminf_{n\to+\infty} \frac{1}{n} \log \mu^{*\otimes n}\big(L_n \in B(\mu^*, \varepsilon_n)\big). \qquad \text{(III.112)}$$

Or, d'après l'inégalité (III.107), pour tout compact K et tout $\xi > 0$, on a :

$$\frac{1}{n} \log \mu^{*\otimes n}\left(L_n \in B\big(\mu^*, \xi + 2\mu^*(K^c)\big)\right) \geq \log \mu^*(K) + \frac{1}{n} \log\left(1 - \left(\frac{16e}{\xi}\right)^{N_K\left(\frac{\xi}{8}\right)} e^{-n\frac{\xi^2}{8}}\right) \qquad \text{(III.113)}$$

Par conséquent, si K_n et ξ_n sont deux suites vérifiant (III.110), on a, d'après (III.112) et (III.113), pour toute suite ε_n de limite nulle telle que $\varepsilon_n \geq \xi_n + 2\mu^*(K_n^c)$:

$$\liminf_{n\to+\infty} \frac{1}{n} \log\left(\mu^{\otimes n}(L_n \in C^{\varepsilon_n}) e^{n\,\mathrm{H}(C|\mu)}\right) \geq 0.$$

D'après le théorème III.36, ceci entraîne que $\mu^n_{C^{\varepsilon_n},k}$ converge en variation vers $\mu^{*\otimes k}$ dans $\mathcal{P}(\mathcal{X}^k)$. $\square$

Quelques exemples

Dans cette section, nous supposerons que $\mathcal{X} = \mathbb{R}^q$. La proposition suivante montre comment des renseignements sur la queue de distribution de μ^* permettent de trouver des vitesses de rétrécissement :

Proposition III.114. *Soient C un convexe fermé de $\mathcal{P}(\mathbb{R}^q)$ tel que $\mathrm{H}\left(C|\,\mu\right) < +\infty$ et μ^* la I-projection de μ sur C.*

1. *S'il existe $a > q$ tel que*
$$\int_{\mathcal{X}} \|x\|^a \, d\mu^*(x) < +\infty, \tag{III.115}$$
alors pour $\varepsilon_n = \frac{2}{n^b}$, avec $b < \frac{1-\frac{q}{a}}{q+2}$, la suite $\mu^n_{C^{\varepsilon_n},k}$ converge en variation vers $\mu^{*\otimes k}$. En particulier, s'il existe $u > 0$ tel que $\int_{\mathcal{X}} e^{u\|x\|} \, d\mu^*(x) < +\infty$, on peut prendre $b < \frac{1}{q+2}$.
2. *S'il existe $a > 0$ tels que (III.115) soit satisfaite et si on suppose en plus que $\log \frac{d\mu^*}{d\mu}$ est continue et bornée, alors on peut prendre $b < \frac{1}{q+2}$.*

Démonstration.
1) En posant $M = \int_{\mathbb{R}^q} \|x\|^a d\mu^*(x)$, on a pour tout $R > 0$

$$\mu^*(\|x\| > R) \leq \frac{M}{R^a} \qquad \text{et} \qquad \mu^*(B(0,R))^n \geq \left(1 - \frac{M}{R^a}\right)^n.$$

En prenant $R_n = n^c$, avec $c > \frac{1}{a}$, on a en posant $K_n = B(0, R_n)$:

$$\mu^*(K_n)^n \geq \left(1 - \frac{M}{n^{ac}}\right)^n \xrightarrow[n \to +\infty]{} 1.$$

De plus, d'après la proposition III.85,

$$N_{K_n}\left(\frac{\xi}{8}\right) \leq M' \left(\frac{R_n}{\xi}\right)^q = M' \frac{n^{cq}}{\xi^q}.$$

En choisissant $\xi_n = \frac{1}{n^b}$, avec $b > 0$, on a

$$\frac{n\xi_n^2}{8} + \log(\xi_n) N_{K_n}\left(\frac{\xi_n}{8}\right) \geq \frac{n^{1-2b}}{8}\left(1 - 8bM' \log(n) n^{cq+b(q+2)-1}\right),.$$

En particulier, si $b < \frac{1-cq}{q+2}$, alors, d'après la proposition III.106, la suite

$$\widetilde{\varepsilon}_n = \xi_n + 2\mu^*(K_n^c)$$

est telle que $\mu^n_{C^{\widetilde{\varepsilon}_n},k}$ converge en variation vers $\mu^{*\otimes k}$.

Comme $ac > 1$ et $b < 1$, $\widetilde{\varepsilon}_n \le \frac{1}{n^b} + \frac{2M}{n^{ac}} \le \frac{2}{n^b}$, pour n assez grand. Ainsi, la suite $\varepsilon_n = \frac{2}{n^b}$ convient pour tout $b < \frac{1-cq}{q+2}$ et $c > \frac{1}{a}$, autrement dit, pour tout $b < \frac{1-\frac{q}{a}}{q+2}$.

2) D'après la proposition III.109, l'hypothèse $c > \frac{1}{a}$ est inutile et peut être remplacée par $c > 0$. On en déduit que $\varepsilon_n = \frac{2}{n^b}$, avec $b < \frac{1}{q+2}$ convient. □

Remarque III.116.

- On voit dans cette proposition que l'hypothèse
$$\log \frac{d\mu^*}{d\mu} \text{ continue et bornée,}$$
permet d'améliorer les vitesses de rétrécissement.
- L'hypothèse (III.115) ou toute autre hypothèse d'intégrabilité portant sur μ^* n'est pas facile à vérifier. En particulier le fait que μ vérifie (III.115) n'entraîne pas nécessairement qu'il en soit de même pour μ^*. En toute généralité, il ne semble pas que l'on puisse aller au delà du résultat élémentaire suivant :

Proposition III.117. *S'il existe $a > 0$ et $\lambda > 0$ tels que*

$$\int_{\mathcal{X}} e^{\lambda \|x\|^a}\, d\mu < +\infty, \tag{III.118}$$

et si $\nu \in \mathcal{P}(\mathcal{X})$ vérifie $\mathrm{H}(\nu|\,\mu) < +\infty$, alors $\int_{\mathcal{X}} \|x\|^a\, d\nu < +\infty$. En particulier, les conclusions de la proposition III.114 restent inchangées si l'on remplace l'hypothèse (III.115) par l'hypothèse (III.118).

Démonstration.

$$\begin{aligned}\int_{\mathcal{X}} \|x\|^a\, d\nu = \frac{1}{\lambda}\int_{\mathcal{X}} \lambda\|x\|^a \frac{d\nu}{d\mu}\, d\mu &\overset{(*)}{\le} \frac{1}{\lambda}\left[\int_{\mathcal{X}} e^{\lambda\|x\|^a} - 1\, d\mu + \int_{\mathcal{X}} \frac{d\nu}{d\mu}\log\frac{d\nu}{d\mu} + 1 - \frac{d\nu}{d\mu}\, d\mu\right] \\ &= \frac{1}{\lambda}\left[\int_{\mathcal{X}} e^{\lambda\|x\|^a}\, d\mu - 1 + \mathrm{H}(\nu|\,\mu)\right] < +\infty.\end{aligned}$$

$(*)$ venant de l'inégalité de Young : $xy \le e^x - 1 + y\log(y) + 1 - y$. □

III.4.5 Applications à l'étude des ponts de Schrödinger et des processus de Nelson

Dans cette section, $\mathcal{Y}$ désignera ou bien $\mathbb{R}^q$ ou bien une variété riemanienne lisse de dimension q connexe et compacte qui sera équipée de sa mesure naturelle dv. Nous poserons $\mathcal{X} = C([0,1],\mathcal{Y})$, ensemble des fonctions continues à valeurs dans $\mathcal{Y}$. Un élément générique de $\mathcal{X}$ sera noté $(x(t))_{t\in[0,1]}$. L'espace $\mathcal{X}$ sera muni de la distance $d_\infty(x,y) = \sup_{s\in[0,1]} d(x(s),x(t))$. Ici, $\mathcal{W}$ sera la mesure brownienne sur $\mathcal{Y}$ (associée à l'opérateur de Laplace-Beltrami) de loi initiale μ_0. Le but de cette section est de donner une interprétation statistique des ponts de Schrödinger et des processus de Nelson grâce aux techniques développées dans les sections précédentes.

Dans [62], E. Schrödinger a posé la question suivante :

> *"Imaginez que vous observez un système de particules en diffusion, qui soit en équilibre thermodynamique. Admettons qu'à l'instant donné t_0 vous les ayez trouvées en répartition à peu près uniforme et qu'à $t_1 > t_0$ vous ayez trouvé un écart spontané et considérable par rapport à cette uniformité. On vous demande de quelle manière cet écart s'est produit. Qu'elle en est la manière la plus probable ?"*

A cette question, la théorie des grandes déviations peut donner des éléments de réponse. Si $X_1,\dots,X_N$ sont des variables aléatoires indépendantes de loi $\mathcal{W}$ modélisant les particules (en l'absence de contraintes), la loi de probabilité que l'on cherche à déterminer est formellement

$$\mathbb{P}\left(L_N \in \, . \, | L_N \in C(\nu_0,\nu_1)\right), \tag{III.119}$$

où $C(\nu_0,\nu_1)$ est l'ensemble des probabilités sur $\mathcal{X}$ ayant pour marginales ν_0 à l'instant $t_0 = 0$ et ν_1 à l'instant $t_1 = 1$. Le nombre de particules étant élevé, on est ramené au calcul de la limite de (III.119), quand $N \to +\infty$. Toujours formellement, cette limite est identifiée par le Principe Conditionnel de Gibbs comme étant la I-projection de $\mathcal{W}$ sur le convexe $C(\nu_0,\nu_1)$. La contrainte $L_n \in C(\nu_0,\nu_1)$ est trop fine pour pouvoir définir (III.119) ; il faut donc la relaxer. On trouvera dans le chapitre 1 du livre [1] de R. Aebi une formulation en double limite de ce principe conditionnel. Ici, nous allons grossir $C(\nu_0,\nu_1)$ en posant, pour tout $\varepsilon > 0$

$$C(\nu_0,\nu_1)_\varepsilon = \{\mathcal{V} \in \mathcal{P}(\mathcal{X}) : \bar{d}(\mathcal{V}_0,\nu_0) \le \varepsilon \quad \text{et} \quad \bar{d}(\mathcal{V}_1,\nu_1) \le \varepsilon\},$$

où $\mathcal{V}_0$ (resp. $\mathcal{V}_1$) désigne la marginale de $\mathcal{V}$ à l'instant $t = 0$ (resp. $t = 1$). Nous chercherons, comme précédemment, des vitesses ε_n telle que

$$\mathcal{W}^n_{\varepsilon_n,k} := \mathcal{L}\left(X_1,\dots,X_k \,| L_n \in C(\nu_0,\nu_1)_{\varepsilon_n}\right) \xrightarrow[n\to+\infty]{} \mathcal{W}^*,$$

$\mathcal{W}^*$ étant la I-projection de $\mathcal{W}$ sur $C(\nu_0,\nu_1)$.

Commençons par rappeler quelques résultats classiques caractérisant cette I-projection. Soit $\mathcal{V}$ appartenant à $C(\nu_0, \nu_1)$. Désignons par $\mathcal{V}_{u,\,v}$ (resp. $\mathcal{W}_{u,\,v}$) la distribution conditionnelle de $\mathcal{V}$ (resp. $\mathcal{W}$) sachant que $x(0) = u$ et $x(1) = v$. Remarquons que $\mathcal{W}_{u,\,v}$ n'est autre que la loi du pont brownien allant de u à v. Notons également $\nu_{0,\,1}$ (resp. $\mu_{0,\,1}$) la loi de $(x(0), x(1))$ sous $\mathcal{V}$ (resp. $\mathcal{W}$). En écrivant que

$$\mathrm{H}\,(\mathcal{V}|\,\mathcal{W}) = \mathrm{H}\,(\nu_{0,\,1}|\,\mu_{0,\,1}) + \int \mathrm{H}\,(\mathcal{V}_{u,\,v}|\,\mathcal{W}_{u,\,v})\ d\nu_{0,\,1}(u, v),$$

il est clair que, si elle existe, la I-projection $\mathcal{W}^*$ s'écrit :

$$\mathcal{W}^* = \int \mathcal{W}_{u,\,v}\, d\mu^*_{0,\,1}(u, v),$$

avec $\mu^*_{0,\,1}$ la I-projection de $\mu_{0,\,1}$ sur

$$\Pi(\nu_0, \nu_1) = \{\alpha \in \mathcal{P}(\mathcal{Y} \times \mathcal{Y}) : \alpha_0 = \nu_0, \alpha_1 = \nu_1\}.$$

Notons μ_0 et μ_1 les marginales de $\mathcal{W}$ aux instants 0 et 1. La probabilité $\mu_{0,\,1}$ est absolument continue par rapport à $\mu_0 \otimes \mu_1$; sa densité sera notée $p(u, v)$. Le lemme suivant donne à la fois une condition suffisante pour que $\mathrm{H}\,(\mu_{0,\,1}|\,\Pi(\nu_0, \nu_1)) < +\infty$ et une formule de représentation de $\mu^*_{0,\,1}$:

Théorème III.120. *Si* $\mathrm{H}\,(\nu_0|\,\mu_0) < +\infty$, $\mathrm{H}\,(\nu_1|\,\mu_1) < +\infty$ *et si* $\log p \in \mathbb{L}_1(\nu_0 \times \nu_1)$ *alors* $\mathrm{H}\,(\mu_{0,\,1}|\,\Pi(\nu_0, \nu_1)) < +\infty$.
De plus,

$$\frac{d\mu^*_{0,\,1}}{d\mu_{0,\,1}}(u, v) = f(u)g(v),$$

pour tout couple (f, g) *de fonctions mesurables vérifiant le système d'équations*

$$\left\{ \begin{array}{l} \frac{d\nu_0}{d\mu_0}(u) = f(u) \int p(u, v)g(v)d\mu_1(v) \\ \frac{d\nu_1}{d\mu_1}(v) = g(v) \int p(u, v)f(u)d\mu_0(u) \end{array} \right. . \tag{III.121}$$

Démonstration. Voir la proposition 6.3 de [13] et [33] p. 161-164. □

Au final, sous les hypothèses du théorème précédent, on a

$$\frac{d\mathcal{W}^*}{d\mathcal{W}} = f(x(0))g(x(1)),$$

pour tout couple (f, g) de fonctions vérifiant le système (III.121).

Proposition III.122. *Sous les hypothèses du théorème III.120, $\mathcal{W}^n_{\varepsilon_n,k}$ converge en variation vers $\mathcal{W}^{*\otimes k}$ pour toute suite ε_n de limite nulle telle que, pour toute suite (Y_i) i.i.d de loi ν_0 et toute suite (Z_i) i.i.d de loi ν_1, on ait*

$$\lim_{n\to+\infty} \mathbb{P}\left(\bar{d}\left(L_n^Y, \nu_0\right) \leq \varepsilon_n\right) = 1 \qquad \textit{et} \qquad \lim_{n\to+\infty} \mathbb{P}\left(\bar{d}\left(L_n^Z, \nu_1\right) \leq \varepsilon_n\right) = 1,$$

en notant : $L_n^Y = \dfrac{1}{n}\displaystyle\sum_{i=1}^n \delta_{Y_i}$ *et* $L_n^Z = \dfrac{1}{n}\displaystyle\sum_{i=1}^n \delta_{Z_i}$.

Démonstration. D'après le corollaire III.14, il suffit de montrer que

$$\mathcal{W}^{*\otimes n}\left(L_n \in C(\nu_0,\nu_1)_{\varepsilon_n}\right) \xrightarrow[n\to+\infty]{} 1, \tag{III.123}$$

Or,

$$\begin{aligned}
\mathcal{W}^{*\otimes n}\left(L_n \in C(\nu_0,\nu_1)_{\varepsilon_n}\right) &= \mathcal{W}^{*\otimes n}\left(\bar{d}\left(\frac{1}{n}\sum_{i=1}^n \delta_{X_i(0)}, \nu_0\right) \leq \varepsilon_n \text{ et } \bar{d}\left(\frac{1}{n}\sum_{i=1}^n \delta_{X_i(1)}, \nu_1\right) \leq \varepsilon_n\right) \\
&\geq 1 - \mathcal{W}^{*\otimes n}\left(\bar{d}\left(\frac{1}{n}\sum_{i=1}^n \delta_{X_i(0)}, \nu_0\right) > \varepsilon_n\right) - \mathcal{W}^{*\otimes n}\left(\bar{d}\left(\frac{1}{n}\sum_{i=1}^n \delta_{X_i(1)}, \nu_1\right) > \varepsilon_n\right) \\
&= 1 - \mathbb{P}\left(\bar{d}\left(L_n^Y, \nu_0\right) > \varepsilon_n\right) - \mathbb{P}\left(\bar{d}\left(L_n^Z, \nu_1\right) > \varepsilon_n\right).
\end{aligned}$$

Ainsi, (III.123) est vérifiée dès que

$$\lim_{n\to+\infty} \mathbb{P}\left(\bar{d}\left(L_n^Y, \nu_0\right) \leq \varepsilon_n\right) = 1 \qquad \text{et} \qquad \lim_{n\to+\infty} \mathbb{P}\left(\bar{d}\left(L_n^Z, \nu_1\right) \leq \varepsilon_n\right) = 1.$$

□

Corollaire III.124. *Sous les hypothèses de la proposition précédente, la convergence en variation de $\mathcal{W}^n_{\varepsilon_n,k}$ vers $\mathcal{W}^{*\otimes k}$ est assurée :*

1. *si $\mathcal{Y}$ est compacte, pour toute suite ε_n telle que*

$$\frac{n\varepsilon_n^2}{8} + \log(\varepsilon_n) N_{\mathcal{Y}}\left(\frac{\varepsilon_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty,$$

2. *si $\mathcal{Y} = \mathbb{R}^q$ et s'il existe $a > q$ tel que*

$$\forall i \in \{1,2\}, \qquad \int_{\mathcal{X}} \|x\|^a \, d\nu_i(x) < +\infty,$$

pour $\varepsilon_n = \frac{2}{n^b}$, avec $b < \frac{1-\frac{q}{a}}{q+2}$.

Démonstration.
1. Cela vient de (III.100) et de (III.97).
2. Immédiat, d'après la proposition III.114. □

Remarque III.125.
D'après la proposition III.105, dans le cas compact, il existe toujours une suite ε_n vérifiant $\frac{n\varepsilon_n^2}{8} + \log(\varepsilon_n)N_{\mathcal{Y}}\left(\frac{\varepsilon_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty$. Par exemple, si $\mathcal{Y}$ est un compact de $\mathbb{R}^q$, on peut prendre $\varepsilon_n = \frac{1}{n^a}$, pour tout $0 < a < \frac{1}{q+2}$ (d'après la proposition III.104) .

Une généralisation naturelle de la question de Schrödinger est la suivante : quelle est la distribution la plus probable du nuage de particules, sachant que toutes les marginales ν_t pour $t \in [0,1]$ sont fixées ? Que ce problème soit connecté avec l'existence de processus de diffusion de Nelson (voir [12] et [51]) a été remarqué pour la première fois par H. Föllmer. Ce point de vue a été approfondi par P. Cattiaux et C. Léonard dans la série d'articles [15, 16, 17].
Dans ce qui suit, nous supposerons $\mathcal{Y} = \mathbb{R}^q$ et nous nous donnerons une famille $(\nu_t)_t$ de probabilités sur $\mathbb{R}^q$. Nous poserons

$$C(\nu_t) = \{\mathcal{V} \in \mathcal{P}(\mathcal{X}) : \forall t \in [0,1], \mathcal{V}_t = \nu_t\},$$

et pour tout $\varepsilon > 0$,

$$C(\nu_t)^\varepsilon = \{\mathcal{V} \in \mathcal{P}(\mathcal{X}) : \bar{d}(\mathcal{V}, C((\nu_t))) \le \varepsilon\}.$$

Le théorème suivant est une application des techniques de la section précédente ; nous en discuterons les hypothèses un peu plus loin.

Théorème III.126. *Supposons que $\mathcal{W}$ possède une I-projection $\mathcal{W}^*$ sur le convexe fermé $C(\nu_t)$ et que celle-ci vérifie : $\log \frac{d\mathcal{W}^*}{d\mathcal{W}}$ est continue bornée. Si, de plus, la loi initiale μ_0 de $\mathcal{W}$ vérifie*

$$\forall R > 0, \quad \mu_0(\|x\| \ge R) \le \frac{C}{R^k}, \qquad \text{avec } k > 0,$$

alors, pour toute suite ε_n de la forme $\varepsilon_n = (\log n)^{-r}$, avec $r < \frac{1}{2q}$, la suite $\mathcal{W}^n_{\varepsilon_n,k}$ converge en variation vers $\mathcal{W}^{*\otimes k}$.

Démonstration. D'après la proposition III.109, il suffit de trouver une suite K_n de compacts de $\mathcal{X}$ et une suite ξ_n de réels strictement positifs et de limite nulle telles que

$$\mathcal{W}^*(K_n) \xrightarrow[n\to+\infty]{} 1 \qquad \text{et} \qquad \frac{n\xi_n^2}{2} + \log(\xi_n)N_{K_n}\left(\frac{\xi_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty.$$

Ceci étant fait, toute suite ε_n de limite nulle telle que $\varepsilon_n \ge \xi_n + 2\mathcal{W}^*(K_n^c)$ fera l'affaire. Comme $\frac{d\mathcal{W}^*}{d\mathcal{W}}$ est bornée par un certain $D > 0$, on a $\mathcal{W}^*(K_n^c) \le D\mathcal{W}(K_n^c)$; en particulier, il suffit de trouver K_n et ξ_n vérifiant

$$\mathcal{W}(K_n) \xrightarrow[n\to+\infty]{} 1 \qquad \text{et} \qquad \frac{n\xi_n^2}{2} + \log(\xi_n) N_{K_n}\left(\frac{\xi_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty,$$

et de prendre $\varepsilon_n \geq \xi_n + 2D\mathcal{W}(K_n^c)$.
La régularité Hölder d'ordre $\alpha < \frac{1}{2}$ des trajectoires browniennes rend naturelle l'introduction des compacts :

$$K(R,M,\alpha) := \left\{ x \in \mathcal{X} : |x(0)| \leq R \quad \text{et} \quad \sup_{s\neq t\in[0,1]} \frac{\|x(s)-x(t)\|}{|s-t|^\alpha} \leq M \right\},$$

où $R, M > 0$ et $\alpha < \frac{1}{2}$.
En appliquant le critère de Kolmogorov (voir, par exemple, le théorème (2.1) du chapitre 1 de [57]), on obtient :

$$\mathcal{W}(K(R,M,\alpha)^c) \leq \mu_0(\|x\| \geq R) + C(p,\alpha)M^{-p}, \tag{III.127}$$

pour tout $p \geq 1$.
De plus, d'après la proposition III.86, on a

$$N_{K(R,M,\alpha)}\left(\frac{\xi}{8}\right) \leq c_1(\alpha,q)\left(\frac{R}{\xi}\right)^q \exp\left(c_2(\alpha,q)\left(\frac{M}{\xi}\right)^{\frac{q}{\alpha}}\right).$$

En prenant, $K_n = K(R_n, M_n, \alpha_n)$, avec

$$R_n = (a\log n)^{\frac{\alpha}{qk}}, \quad M_n = (b\log n)^{\frac{\alpha}{q}}, \quad \text{et} \quad \xi_n = (c\log n)^{-\frac{\alpha}{q}},$$

on voit, après quelques calculs, que la quantité

$$\frac{n\xi_n^2}{2} + \log(\xi_n) N_{K_n}\left(\frac{\xi_n}{8}\right)$$

est majorée par

$$n\,(\log n)^{-\frac{2\alpha}{q}}\left[A_1 + A_2\log(c\log n)\,(\log n)^{q+\frac{2\alpha}{q}}\, n^{c_2(\alpha,q)bc-1}\right],$$

où A_1 et A_2 ne dépendent plus de n. Pour tout c fixé, on peut choisir b tel que $c_2(\alpha,q)bc-1 < 0$. Ceci étant fait, la quantité précédente tend vers $+\infty$ lorsque $n \to +\infty$. Enfin, grâce à (III.127), on a

$$\xi_n + 2D\mathcal{W}(K_n^c) \leq (c\log n)^{-\frac{\alpha}{q}} + 2CD(a\log n)^{-\frac{\alpha}{q}} + 2DC(p,\alpha)(b\log n)^{-\frac{\alpha p}{q}}$$

et pour tout $\alpha' < \alpha$, cette dernière quantité est majorée pour n assez grand par $\varepsilon_n = (\log n)^{-\frac{\alpha'}{q}}$.
□

Remarque III.128.

- L'hypothèse $\log \frac{d\mathcal{W}^*}{d\mathcal{W}}$ continue bornée est indispensable. Sans cette hypothèse, on pourrait penser appliquer la proposition III.106, quitte à obtenir des vitesses de grossissement moins bonnes. Mais pour être appliquée, cette proposition requiert que
$$\mathcal{W}^*(K_n)^n \xrightarrow[n\to+\infty]{} 1$$
et ceci impose
$$\mathcal{W}^*(K_n^c) = o\left(\frac{1}{n}\right).$$
En supposant, ce qui est raisonnable que $\frac{d\mathcal{W}^*}{d\mathcal{W}} \in \mathbb{L}_r(\mathcal{X}, \mathcal{W})$, le critère de Kolmogorov nous donne
$$\mathcal{W}^*(K(R_n, M_n, \alpha)^c) \leq \mu_0(\|x\| \geq R_n) + C(p, \alpha) M_n^{-p}.$$
En particulier, on doit prendre M_n en n^a, $a > 0$. On se convaincra qu'un tel choix de M_n n'est plus compatible avec l'existence d'une suite ξ_n vérifiant
$$\frac{n\xi_n^2}{2} + \log(\xi_n) N_{K_n}\left(\frac{\xi_n}{8}\right) \xrightarrow[n\to+\infty]{} +\infty.$$
- Cette condition, $\log \frac{d\mathcal{W}^*}{d\mathcal{W}}$ continue bornée, est difficile à vérifier. En effet, en général, on sait simplement que cette densité a la forme d'une densité de Girsanov :
$$\frac{d\mathcal{W}^*}{d\mathcal{W}} = \frac{d\nu_0}{d\mu_0} G,$$
avec
$$G = \exp\left(\int_0^1 B(t, x(t)) dx(t) - \int_0^1 |B(t, x(t))|^2 dt\right)$$
et G n'est pas continue en général. Pour clore cette section et ce chapitre, nous nous contenterons de donner un exemple simple de flot $(\nu_t)_t$ pour lequel la I-projection est connue et vérifie cette hypothèse de continuité.

Soit $U : \mathbb{R}^q \to \mathbb{R}$ une fonction bornée de classe $\mathcal{C}^3$ à dérivées bornées. L'équation différentielle stochastique
$$dX_t = dB_t - \nabla U(X_t) dt,$$
admet, pour toute variable aléatoire X_0 donnée, une unique solution (forte). Notons $\mathcal{V}^0$ la loi de cette solution, et pour tout t posons $\nu_t = \mathcal{L}(X_t)$. On supposera, en outre, $\log \frac{d\nu_0}{d\mu_0}$ est continue bornée.
On a alors la

Proposition III.129. *La probabilité $\mathcal{V}^0$ est la I-projection de $\mathcal{W}$ sur $C(\nu_t)$ et $\log \frac{d\mathcal{V}^0}{d\mathcal{W}}$ est continue bornée sur $\mathcal{X}$. En particulier, les conclusions du théorème III.126 sont valables.*

Démonstration. Le premier point est assez classique et est démontré par exemple dans [15]. Le second point résulte de la forme explicite de la densité (voir, par exemple, le lemme 2.2.21 de [61]) :

$$\frac{d\mathcal{V}^0}{d\mathcal{W}} = \frac{d\nu_0}{d\mu_0} \exp\left(U(x(0)) - U(x(1)) - \frac{1}{2}\int_0^1 \left[|\nabla U|^2 - \Delta U \right] (x(s))\, ds \right).$$

□

CHAPITRE IV

A propos d'une méthode de calibration en finance

Sommaire

IV.1 Introduction

IV.1.1 Une méthode de calibration

Un problème important en mathématiques financières est celui de la calibration :
On cherche à modéliser un actif financier par un processus de diffusion solution d'une équation différentielle stochastique :

$$dS_t = \sigma(t, S_t)\, dB_t + b(t, S_t)\, dt. \tag{IV.1}$$

Pour des raisons d'ordre économique, le drift b est fixé : $b(t,x) = b_0 \in \mathbb{R}$. Il s'agit de trouver un coefficient de diffusion σ tel que

$$\mathbb{E}[F(S_T)] = 1, \tag{IV.2}$$

où $T \in]0,1[$ est un instant fixé et F une fonction positive.

Dans [2], M. Avellaneda, C. Friedman, R. Holmes et D. Samperi ont proposé la méthode suivante :

- on se donne un modèle *a priori* σ_0 et une fonction $q : \mathbb{R}^2 \to \mathbb{R}^+$ continue nulle sur la diagonale,
- on prend comme solution du problème de calibration, la fonction σ^* qui minimise la fonctionnelle :

$$\sigma \mapsto \mathrm{I}(\sigma|\sigma_0) = \mathbb{E}_\sigma \left[\int_0^1 q(\sigma^2(X_t, t), \sigma_0^2(t, X_t))\, dt \right],$$

  sous la contrainte (IV.2), où $(X_t)_{t\in[0,1]}$ désigne le processus canonique, et $\mathbb{E}_\sigma[\,.\,]$ l'espérance par rapport à la loi de la solution de (IV.1).

Le choix de ces fonctionnelles $\mathrm{I}(\,.\,|\sigma_0)$ repose sur un raisonnement heuristique, mené dans l'introduction de [2], que nous allons retranscrire ci dessous.

IV.1.2 Justification heuristique de cette méthode

Posons Σ, l'ensemble des fonctions $\sigma : \mathbb{R} \times [0,1] \to \mathbb{R}^{+*}$ continues telles que $\inf \sigma > 0$ et $\sup \sigma < +\infty$. Pour tout $\sigma \in \Sigma$, il y a existence faible et unicité en loi pour l'équation différentielle stochastique :

$$dS_t = \sigma(t, S_t)\, dB_t + b_0\, dt. \tag{IV.3}$$

Nous noterons $\mathbb{Q}_\sigma$ la mesure de probabilité sur $\mathcal{P}(C[0,1])$ ainsi définie. Pour tout $t \in [0,1]$, on posera : $\forall \omega \in C[0,1], X_t(\omega) := \omega(t)$.

Pour déterminer une solution au problème de calibration exposé plus haut, une première idée consisterait à utiliser la méthode de minimisation de l'entropie relative, à savoir, fixer un modèle *a priori* $\mathbb{Q}_{\sigma_0}$, avec $\sigma_0 \in \Sigma$, et prendre comme solution la probabilité $\mathbb{Q}^*$ minimisant $\mathrm{H}\left(\,.\, | \, \mathbb{Q}_{\sigma_0}\right)$ sous la contrainte

$$\int F(X_T)\, d\mathbb{Q} = 1.$$

Cette approche est totalement inadaptée. En effet, d'après le théorème de Girsanov, $\mathbb{Q}^*$ sera solution de

$$dS_t = \sigma_0(t, S_t)\, dB_t + b\, dt. \tag{IV.4}$$

avec $b \neq b_0$. Ainsi la méthode de minimisation de l'entropie relative fournie une réponse au problème "orthogonal" qui est de maintenir fixe le coefficient de diffusion et de changer le drift afin de garantir (IV.2).

L'idée proposée par Avellaneda et ses coauteurs est de minimiser l'entropie relative sur des versions discrétisées des processus. Pour tout $\sigma \in \Sigma_0 \subset \Sigma$, Σ_0 étant un certain sous-ensemble de Σ, ils supposent donnée une suite $(\mathbb{Q}^n_\sigma)_n$ de mesures de probabilité sur $C[0,1]$ telles que :

1. $\mathbb{Q}^n_\sigma \xrightarrow[n\to+\infty]{} \mathbb{Q}_\sigma$, au sens de la convergence étroite,
2. $\mathbb{Q}^n_\sigma \left(X_{\frac{k}{n}} \middle| X_{\frac{k-1}{n}}, \ldots, X_{\frac{1}{n}}, X_0 \right) = \mathbb{Q}^n_\sigma \left(X_{\frac{k}{n}} \middle| X_{\frac{k-1}{n}} \right)$
3. $\mathbb{Q}^n_\sigma \left(X_t = X_{\frac{k}{n}} + (nt-k) \left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}} \right], \frac{k}{n} \le t \le \frac{k+1}{n} \right) = 1.$

Ils remarquent que certains schémas d'approximation classiques (schéma d'Euler, arbres trinomiaux...) vérifient en outre :

4. $\forall \sigma_1, \sigma_2 \in \Sigma_0,\ \mathbb{Q}^n_{\sigma_1} \sim \mathbb{Q}^n_{\sigma_2}$,
5. Pour tout $(\sigma_0, \sigma) \in \Sigma_0^2$,

$$\frac{1}{n}\,\mathrm{H}\left(\mathbb{Q}^n_\sigma \middle| \mathbb{Q}^n_{\sigma_0} \right) \xrightarrow[n\to+\infty]{} \mathbb{E}_\sigma \left[\int_0^1 q(\sigma^2(X_t,t), \sigma_0^2(t,X_t))\, dt \right] := \mathrm{I}(\sigma|\sigma_0), \tag{IV.5}$$

où $q : \mathbb{R}^2 \to \mathbb{R}^+$ est une fonction convexe nulle sur la diagonale, dépendant du schéma d'approximation choisi.

Se fondant sur (IV.5), ils proposent alors de minimiser sous contraintes $\mathrm{I}(\,.\,|\sigma_0)$ pour résoudre le problème de calibration, car il paraît naturel de penser que la solution du problème de minimisation sous contraintes de $\mathrm{H}\left(\,.\, \middle| \mathbb{Q}^n_{\sigma_0} \right)$ va converger quand n tend vers l'infini vers la solution du problème de minimisation sous contraintes de $\mathrm{I}(\,.\,|\sigma_0)$.

Le but de cette section est d'éclaircir un certain nombre de points délicats de ce raisonnement heuristique et de connecter cette approche à un principe conditionnel de Gibbs multi-échelles.

Une interprétation en terme de Principe Conditionnel de Gibbs est naturelle. En effet, si l'on pose

$$L_m : \mathcal{C}[0,1]^m \to \mathcal{P}(\mathcal{C}[0,1]) : (\omega_1, \dots, \omega_m) \mapsto \frac{1}{m}\sum_{i=1}^{m} \delta_{\omega_i},$$

alors, *pour n fixé*, on s'attend à ce que

$$\mathbb{E}_{\mathbb{Q}_{\sigma_0}^{n\otimes m}}\left[L_m \,|\langle L_m, F(X_T)\rangle = 1 \quad \text{et} \quad L_m \text{ proche de } \{\mathbb{Q}_\sigma^n, \sigma \in \Sigma_0\}\right]$$

converge, *lorsque m tend vers $+\infty$*, vers

$$\mathrm{Argmin}\left\{\mathrm{H}\left(\mathbb{Q}_\sigma^n \middle| \mathbb{Q}_{\sigma_0}^n\right), \sigma \in \Sigma_0 \quad \text{t.q.} \quad \langle \mathbb{Q}_\sigma^n, F(X_T)\rangle = 1\right\}.$$

En admettant que

$$\mathrm{Argmin}\left\{\mathrm{H}\left(\mathbb{Q}_\sigma^n \middle| \mathbb{Q}_{\sigma_0}^n\right), \sigma \in \Sigma_0 \quad \text{t.q.} \quad \langle \mathbb{Q}_\sigma^n, F(X_T)\rangle = 1\right\} \xrightarrow[n\to+\infty]{} \mathbb{Q}_{\sigma^*},$$

avec $\sigma^* = \mathrm{Argmin}\,\{\mathrm{I}(\sigma|\sigma_0), \sigma \in \Sigma_0 \quad \text{t.q.} \quad \langle \mathbb{Q}_\sigma, F(X_T)\rangle = 1\}$, on peut espérer trouver une suite m_n telle que

$$\mathbb{E}_{\mathbb{Q}_{\sigma_0}^{n\otimes m_n}}\left[L_{m_n} \,|\langle L_{m_n}, F(X_T)\rangle = 1 \quad \text{et} \quad L_{m_n} \text{ proche de } \{\mathbb{Q}_\sigma^n, \sigma \in \Sigma_0\}\right] \xrightarrow[n\to+\infty]{} \mathbb{Q}_{\sigma^*}.$$

Dans la suite, nous choisirons *les arbres trinomiaux* comme modèle d'approximation (voir la section suivante pour leur définition) et nous verrons qu'il est, malheureusement, difficile de mener à bien notre programme en dehors de ce cadre.

IV.2 Approximation d'une diffusion par un arbre trinomial

IV.2.1 Approximation d'une diffusion par une chaîne de Markov

Introduisons quelques notations. On désignera par Ω l'ensemble $C([0,1],\mathbb{R})$ des applications continues de $[0,1]$ dans $\mathbb{R}$; les applications coordonnées sur Ω seront notées X_t, $t \in [0,1]$. En notant Σ l'ensemble des fonctions $\sigma : \mathbb{R} \times [0,1] \to \mathbb{R}^{+*}$ continues telles que $\inf \sigma > 0$ et $\sup \sigma < +\infty$, on a le résultat classique suivant :

Théorème IV.6. *Soient $\sigma \in \Sigma$ et $b : [0,1] \times \mathbb{R} \to \mathbb{R}$ continue, alors l'équation différentielle stochastique*

$$dS_t = \sigma(t, S_t)\, dB_t + b(t, S_t)\, dt, \quad S_0 = x_0 \tag{IV.7}$$

admet au moins une solution faible et il y a, de plus, unicité en loi.
Pour tout $x_0 \in \mathbb{R}$, nous noterons $\mathbb{Q}_{\sigma, b, x_0} \in \mathcal{P}(\Omega)$ la loi commune de toutes les solutions de (IV.7) issues de x_0.

Le théorème suivant, dû à D.W. Stroock et S.R.S Varadhan, donne un moyen pour approximer les $\mathbb{Q}_{\sigma,b,x_0}$ par des chaînes de Markov :

Théorème IV.8. (Stroock et Varadhan)
Pour tout $n \in \mathbb{N}^*$ et $t \in [0,1]$, soit $\left(\Pi^n(t,x,\,.\,)\right)_x$ un noyau de transition de $\mathbb{R}$ dans $\mathbb{R}$. Si $(\mathbb{Q}^n)_n$ est une suite de mesures de probabilité sur Ω, vérifiant

$$\begin{cases} (1) \quad \mathbb{Q}^n(X_0 = x_0) = 1, \\ (2) \quad \mathbb{Q}^n\left(X_t = X_{\frac{k}{n}} + (nt-k)\left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}}\right], \frac{k}{n} \le t \le \frac{k+1}{n}\right) = 1, \\ (3) \quad \mathbb{Q}^n\left(X_{\frac{k+1}{n}} \in \,.\, \middle| X_{\frac{k}{n}}, \dots, X_0\right) = \Pi^n\left(\frac{k}{n}, X_{\frac{k}{n}}, \,.\,\right) \end{cases},$$

et s'il existe $\sigma \in \Sigma$ et $b : [0,1]\times\mathbb{R} \to \mathbb{R}$ continue bornée telles que

a. $\displaystyle \sup_{\substack{n\in\mathbb{N},\, x\in\mathbb{R}\\ t\in[0,1]}} n \int_{|y-x|\le 1} (y-x)^2\, \Pi^n(t,x,dy) < +\infty$
et
$\displaystyle \sup_{\substack{n\in\mathbb{N},\, x\in\mathbb{R}\\ t\in[0,1]}} n \left| \int_{|y-x|\le 1} (y-x)\, \Pi^n(t,x,dy)\right| < +\infty,$

b. $\displaystyle \forall R > 0, \sup_{\substack{x\in[-R,R]\\ t\in[0,1]}} \left| n \int_{|y-x|\le 1} (y-x)^2\, \Pi^n(t,x,dy) - \sigma^2(t,x)\right| \xrightarrow[n\to+\infty]{} 0,$

c. $\displaystyle \forall R > 0, \sup_{\substack{x\in[-R,R]\\ t\in[0,1]}} \left| n \int_{|y-x|\le 1} (y-x)\, \Pi^n(t,x,dy) - b(t,x)\right| \xrightarrow[n\to+\infty]{} 0,$

d. $\displaystyle \forall \varepsilon > 0, \sup_{x\in\mathbb{R},\, t\in[0,1]} n\Pi^n(t,x,\mathbb{R} - [x-\varepsilon, x+\varepsilon]) \xrightarrow[n\to+\infty]{} 0,$

alors $\mathbb{Q}^n$ converge étroitement vers $\mathbb{Q}_{\sigma,b,x_0}$.

Dans tout ce qui suit, $x_0 = 0$ et nous noterons $\mathbb{Q}_{\sigma,b,}$ à la place de $\mathbb{Q}_{\sigma,b,0}$.

IV.2.2 Définition des arbres trinomiaux

Donnons nous deux nombres $0 < \sigma_{\min} < \sigma_{\max}$ et $b_0 \in \mathbb{R}^+$. Posons

$$\Sigma_0 = \{\sigma : [0,1]\times\mathbb{R} \to]\sigma_{\min}, \sigma_{\max}[, \text{ continues}\}$$

et pour $\varepsilon < b_0$,

$$\mathcal{B}_\varepsilon = \{b : [0,1]\times\mathbb{R} \to]b_0 - \varepsilon, b_0 + \varepsilon[, \text{ continues}\}.$$

Précisons que l'ensemble $C([0,1]\times\mathbb{R},\mathbb{R})$ des applications continues de $[0,1]\times\mathbb{R}$ dans $\mathbb{R}$ sera toujours muni de la topologie de la convergence uniforme sur tout compact.

Nous allons maintenant définir une classe de processus appelés *arbres trinomiaux* permettant d'approximer les diffusions $\mathbb{Q}_{\sigma,b}$, avec $\sigma \in \overline{\Sigma_0}$ et $b \in \overline{\mathcal{B}_\varepsilon}$. Pour cela, nous fixerons deux nombres α et s vérifiant

$$\alpha > 0, \quad b_0 > s > 0, \quad 0 < \sigma_{\text{min}} < \sigma_{\text{max}} < \alpha.$$

et nous poserons, pour tout $n \in \mathbb{N}^*$ et tout $(y,z) \in \mathbb{R}^2$

$$\begin{cases} m^n(y,z) = \frac{y^2}{2\alpha^2} + \frac{z}{2\alpha\sqrt{n}} \\ d^n(y,z) = \frac{y^2}{2\alpha^2} - \frac{z}{2\alpha\sqrt{n}} \\ r^n(y,z) = 1 - \frac{y^2}{\alpha^2} \end{cases}.$$

Il est clair qu'il existe $n_0 \in \mathbb{N}^*$ ne dépendant que de σ_{min}, σ_{max}, b_0 et s tel que, pour tout $(y,z) \in [\sigma_{\text{min}}, \sigma_{\text{max}}] \times [b_0 - s, b_0 + s]$, le vecteur $[m^n(y,z), r^n(y,z), d^n(y,z)]$ soit un vecteur de probabilité à coefficients tous strictement positifs.
Définissons pour tout $(\sigma, b) \in \overline{\Sigma_0} \times \overline{\mathcal{B}_s}$, $n \geq n_0$ et $(t,x) \in [0,1] \times \mathbb{R}$,

$$\Pi^n_{\sigma,b}(t,x,\,.\,) = m^n(\sigma,b)(t,x)\delta_{x+\frac{\alpha}{\sqrt{n}}} + r^n(\sigma,b)(t,x)\delta_x + d^n(\sigma,b)(t,x)\delta_{x-\frac{\alpha}{\sqrt{n}}}.$$

Pour tout t, $(\Pi^n_{\sigma,b}(t,x,\,.\,))_x$ est un noyau de transition de $\mathbb{R}$ dans $\mathbb{R}$.
Pour tout $(\sigma,b) \in \overline{\Sigma_0} \times \overline{\mathcal{B}_s}$, on considère la probabilité $\mathbb{Q}^n_{\sigma,b}$ sur $(\Omega, \mathcal{G})$ définie par :

$$\begin{cases} (1) \quad \mathbb{Q}^n_{\sigma,b}(X_0 = 0) = 1, \\ (2) \quad \mathbb{Q}^n_{\sigma,b}\left(X_t = X_{\frac{k}{n}} + (nt-k)\left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}}\right], \frac{k}{n} \leq t \leq \frac{k+1}{n}\right) = 1, \\ (3) \quad \mathbb{Q}^n_{\sigma,b}\left(X_{\frac{k+1}{n}} \in \,.\, \middle| X_{\frac{k}{n}}, \ldots, X_0\right) = \Pi^n_{\sigma,b}\left(\frac{k}{n}, X_{\frac{k}{n}},\,.\,\right) \end{cases} \tag{IV.9}$$

Les processus $\mathbb{Q}^n_{\sigma,b}$ sont appelés **arbres trinomiaux** (issus de 0). Nous noterons $\mathbb{E}^n_{\sigma,b}[\,.\,]$, l'espérance par rapport à $\mathbb{Q}^n_{\sigma,b}$. Le support de $\mathbb{Q}^n_{\sigma,b}$ est clairement l'ensemble $\Omega_n \subset \Omega$ défini par

$$\Omega_n = \left\{ \omega \in \Omega : \left[\begin{array}{ll} -\omega(0) = 0 & \\ -\omega\left(\frac{i+1}{n}\right) - \omega\left(\frac{i}{n}\right) \in \left\{-\frac{\alpha}{\sqrt{n}}, 0, \frac{\alpha}{\sqrt{n}}\right\}, & \text{pour } i = 0, \ldots, n-1 \\ -\omega \text{ affine sur } \left[\frac{i}{n}, \frac{i+1}{n}\right], & \text{pour } i = 0, \ldots, n-1 \end{array} \right. \right\}$$

Ω_n est un ensemble fini (de cardinal 3^n).

IV.2.3 Convergence des arbres trinomiaux

Proposition IV.10. *Soit (ε_n) une suite de réels strictement positifs convergeant vers zéro, avec $\varepsilon_n \leq s$. Pour toute suite $(\sigma_n)_n$ d'éléments de $\overline{\Sigma_0}$ convergeant vers $\sigma \in \overline{\Sigma_0}$ uniformément sur tout compact et toute suite $b_n \in \overline{\mathcal{B}_{\varepsilon_n}}$, la suite $\left(\mathbb{Q}^n_{\sigma_n,b_n}\right)_{n \geq n_0}$ converge étroitement vers $\mathbb{Q}_{\sigma,b_0}$.*

Démonstration. On voit facilement que pour n assez grand

$$n\int_{|y-x|\leq 1}(y-x)^2\,\Pi^n_{\sigma,\,b}(t,x,dy) = n\frac{\alpha^2}{n}\left[m^n(\sigma(t,x),b(t,x)) + d^n(\sigma(t,x),b(t,x))\right]$$
$$= \sigma^2(t,x),$$

$$n\int_{|y-x|\leq 1}(y-x)\,\Pi^n_{\sigma,\,b}(t,x,dy) = n\frac{\alpha}{\sqrt{n}}\left[m^n(\sigma(t,x),b(t,x)) - d^n(\sigma(t,x),b(t,x))\right]$$
$$= b(t,x),$$

et

$$n\Pi^n_{\sigma,\,b}(t,x,\mathbb{R}-[x-\varepsilon,x+\varepsilon]) = 0.$$

Le résultat découle alors du théorème IV.8. $\square$

IV.3 Principe conditionnel de Gibbs

IV.3.1 Introduction

Introduisons quelques notations supplémentaires. Pour tout $\varepsilon > 0$, $\mathcal{E}^n_\varepsilon$ désignera le sous-ensemble de $\mathcal{P}(\Omega)$ défini par

$$\mathcal{E}^n_\varepsilon = \left\{\mathbb{Q}\in\mathcal{P}(\Omega) : \left|\int F\left(X_{\frac{[nT]}{n}}\right)d\mathbb{Q} - 1\right| < \varepsilon\right\}, \tag{IV.11}$$

et $\mathcal{D}^n_\varepsilon$, l'ensemble des probabilités $\mathbb{Q}$ sur Ω vérifiant les propriétés suivantes

$$\left\{\begin{array}{ll} (1) & \mathbb{Q}(X_0 = 0) = 1, \\ (2) & \mathbb{Q}\left(X_t = X_{\frac{k}{n}} + (nt-k)\left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}}\right], \frac{k}{n}\leq t\leq \frac{k+1}{n}\right) = 1, \\ (3') & \exists(\sigma,b)\in\Sigma_0\times\mathcal{B}_\varepsilon \text{ tels que } \mathbb{Q}\left(X_{\frac{p+1}{n}}\in\ .\ \middle| X_{\frac{p}{n}}\right) = \Pi^n_{\sigma,\,b}\left(\frac{p}{n}, X_{\frac{p}{n}},\ .\ \right) \end{array}\right. \tag{IV.12}$$

Nous poserons

$$\mathcal{F}^n_\varepsilon = \mathcal{E}^n_\varepsilon\cap\mathcal{D}^n_\varepsilon. \tag{IV.13}$$

Enfin, pour $\varepsilon > 0$ et $m\in\mathbb{N}^*$, la probabilité $\mathbb{R}^n_{\varepsilon,\,m}\in\mathcal{P}(\Omega)$ est définie (quand cela est possible) par

$$\mathbb{R}^n_{\varepsilon,\,m}(\omega) = \mathbb{E}_{(\mathbb{Q}^n_{\sigma_0,\,b_0})^{\otimes m}}\left[L_m\,|L_m\in\mathcal{F}^n_\varepsilon\right]$$
$$= \frac{(\mathbb{Q}^n_{\sigma_0,\,b_0})^{\otimes m}\{(\omega_1,\ldots,\omega_m)\in\Omega_n : \omega_1=\omega, L_m(\omega_1,\ldots,\omega_m)\in\mathcal{F}^n_{\varepsilon_n}\}}{(\mathbb{Q}^n_{\sigma_0,\,b_0})^{\otimes m}\{(\omega_1,\ldots,\omega_m)\in\Omega_n : L_m(\omega_1,\ldots,\omega_m)\in\mathcal{F}^n_\varepsilon\}},$$

où

$$L_m : (\Omega_n)^m \to \mathcal{P}(\Omega_n) : (\omega_1, \ldots, \omega_m) \mapsto L_m(\omega_1, \ldots, \omega_m) = \frac{1}{m} \sum_{i=1}^{m} \delta_{\omega_i}$$

Nous ferons plus loin des commentaires sur les raisons du choix de l'ensemble $\mathcal{D}_\varepsilon^n$ (voir section IV.3.4). Avant cela, détaillons le contenu de ce qui va suivre dans les prochaines sections.

Dans la section IV.3.2, nous nous intéresserons au comportement asymptotique de $\mathbb{R}_{\varepsilon,m}^n$ lorsque ε et n sont fixés et m tend vers $+\infty$. Pour cela, nous montrerons que $\mathcal{D}_\varepsilon^n$ est un ouvert convexe de $\mathcal{P}(\Omega_n)$ (voir proposition IV.15), ce qui nous permettra de conclure dans la proposition IV.18, grâce à une version du Principe Conditionnel de Gibbs (théorème IV.19), qu'à $\varepsilon > 0$ et n fixés,

$$\mathbb{R}_{\varepsilon,m}^n \xrightarrow[m \to +\infty]{} \mathbb{Q}_\varepsilon^{n*},$$

où $\mathbb{Q}_\varepsilon^{n*}$ est la I-projection de $\mathbb{Q}_{\sigma_0, b_0}^n$ sur $\overline{\mathcal{F}_\varepsilon^n}$, *ie* l'unique probabilité $\mathbb{Q} \in \overline{\mathcal{F}_\varepsilon^n}$ telle que $\mathrm{H}\left(\mathbb{Q} \middle| \mathbb{Q}_{\sigma_0, b_0}^n\right) = \mathrm{H}\left(\overline{\mathcal{F}_\varepsilon^n} \middle| \mathbb{Q}_{\sigma_0, b_0}^n\right)$.

Dans la section IV.3.3, nous étudierons les probabilités $\mathbb{Q}_\varepsilon^{n*}$. Nous montrerons dans la proposition IV.20 que $\mathbb{Q}_\varepsilon^{n*}$ est un arbre trinomial. Ensuite nous nous intéresserons au comportement asymptotique des $\mathbb{Q}_{\varepsilon_n}^{n*}$ lorsque n tend vers $+\infty$. Dans la proposition IV.21, nous montrerons que $\sigma \mapsto \dfrac{1}{n} \mathrm{H}\left(\mathbb{Q}_{\sigma,b}^n \middle| \mathbb{Q}_{\sigma_0, b_0}^n\right)$ converge, en un sens proche de la Γ-convergence, vers

$$\sigma \mapsto \mathrm{I}(\sigma|\sigma_0) = \mathbb{E}_\sigma \left[\int_0^1 q(\sigma^2(X_t, t), \sigma_0^2(t, X_t))\, dt \right],$$

avec

$$q(x, y) = \log\left(\frac{x}{y}\right) \frac{x}{\alpha^2} + \log\left(\frac{\alpha^2 - x}{\alpha^2 - y}\right) \left[1 - \frac{x}{\alpha^2}\right].$$

Grâce à cela, nous montrerons que si, pour une suite $(\varepsilon_n)_n$ bien choisie, la suite $\mathbb{Q}_\varepsilon^{n*}$ s'exprime sous la forme

$$\mathbb{Q}_\varepsilon^{n*} = \mathbb{Q}_{\sigma_n^*, b_n^*}^n \qquad \text{avec} \qquad \sigma_n^* \ \underline{\text{précompacte}}, \tag{IV.14}$$

alors ses valeurs d'adhérence sont de la forme $\mathbb{Q}_{\sigma^*, b_0}$ avec σ^* un minimisant de $\mathrm{I}(\,.\,|\sigma_0)$ sous la contrainte (IV.2) (voir proposition IV.23).

A partir des résultats des sections précédentes, nous serons en mesure de montrer dans la section IV.3.4, sous l'hypothèse (IV.14), que toutes les valeurs d'adhérences de $\mathbb{R}_{\varepsilon_n, m_n}^n$, m_n étant une suite d'entiers tendant vers $+\infty$, sont également de la forme $\mathbb{Q}_{\sigma^*, b_0}$ (voir proposition IV.24). En particulier, si le problème de minimisation de $\mathrm{I}(\,.\,|\sigma_0)$ possède une unique solution σ^*, nous aurons

$$\mathbb{R}^n_{\varepsilon_n,\,m_n} \xrightarrow[n\to+\infty]{} \mathbb{Q}_{\sigma^*,\,b_0},$$

ce qui apportera une interprétation partielle de la méthode d'Avellaneda : la probabilité $\mathbb{Q}_{\sigma^*,\,b_0}$ fournie par cette méthode de calibration est la limite d'une suite de probabilités conditionnelles définies à partir d'une suite de discrétisations de la diffusion de référence $\mathbb{Q}_{\sigma_0,\,b_0}$. Dans cette section nous essaierons également de lever l'hypothèse (IV.14) qui est difficilement vérifiable. Pour cela, nous remplacerons Σ_0 par un sous-ensemble compact Σ_1 bien choisi. Cela aura un prix : la perte de la convexité de $\mathcal{D}^n_\varepsilon$. En faisant l'hypothèse que $\mathrm{I}\,(\,.\,|\sigma_0)$ admet un *unique* minimisant, nous établirons un résultat de convergence satisfaisant pour $\mathbb{R}^n_{\varepsilon_n,\,m_n}$. Nous terminerons cette section par un résultat de convergence valable pour des schémas d'approximations plus généraux, mais le cadre dans lequel nous nous placerons sera encore trop restrictif pour accueillir les schémas de type Euler.

IV.3.2 Convexification des arbres trinomiaux et Principe Conditionnel de Gibbs à n fixé

Considérons l'ensemble $\mathcal{T}^n_\varepsilon$ défini par

$$\mathcal{T}^n_\varepsilon = \left\{Q^n_{\sigma,\,b}, \sigma\in\Sigma_0, b\in\mathcal{B}_\varepsilon\right\},$$

qui est l'ensemble des arbres trinomiaux sur Ω_n associés à des diffusions ayant un drift dans la bande $]b_0-\varepsilon, b_0+\varepsilon[$. Cet ensemble n'est pas convexe, car une combinaison convexe de processus de Markov n'est plus un processus de Markov. Nous allons chercher à inclure $\mathcal{T}^n_\varepsilon$ dans un ensemble convexe qui ne soit pas trop gros :

Proposition IV.15. *L'ensemble $\mathcal{D}^n_\varepsilon$ défini par (IV.12) est un ouvert convexe de $\mathcal{P}(\Omega_n)$ qui contient $\mathcal{T}^n_\varepsilon$.*

Démonstration. Il est clair que $\mathcal{D}^n_\varepsilon$ contient $\mathcal{T}^n_\varepsilon$.
Montrons que $\mathcal{D}^n_\varepsilon$ est convexe. Soient $\mathbb{Q}_1$ et $\mathbb{Q}_2$ dans $\mathcal{D}^n_\varepsilon$ vérifiant la propriété (3)' de (IV.12) avec (σ_1, b_1) et (σ_2, b_2). Pour tout $u\in[0,1]$, posons

$$\mathbb{Q}_{1+u} = (1-u)\mathbb{Q}_1 + u\mathbb{Q}_2$$

Les propriétés (1) et (2) de (IV.12) sont trivialement vérifiées par $\mathbb{Q}_{1+u}$. Montrons que

$\mathbb{Q}_{1+u}$ vérifie aussi (3)' :

$$\mathbb{Q}_{1+u}\left(X_{\frac{i+1}{n}} = \frac{\alpha j}{\sqrt{n}} \middle| X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right) =$$

$$\frac{(1-u)\Pi^n_{\sigma_1, b_1}\left(\frac{i}{n}, \frac{\alpha k}{\sqrt{n}}, \frac{\alpha j}{\sqrt{n}}\right)\mathbb{Q}_1\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right) + u\Pi^n_{\sigma_2, b_2}\left(\frac{i}{n}, \frac{\alpha k}{\sqrt{n}}, \frac{\alpha j}{\sqrt{n}}\right)\mathbb{Q}_2\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}{(1-u)\mathbb{Q}_1\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right) + u\mathbb{Q}_2\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}$$

$$= \Pi_{\sigma_{1+u}, b_{1+u}}\left(\frac{i}{n}, \frac{\alpha k}{\sqrt{n}}, \frac{\alpha j}{\sqrt{n}}\right)$$

avec

$$\begin{aligned}\sigma^2_{1+u}\left(\frac{i}{n}, \frac{\alpha k}{\sqrt{n}}\right) &= \frac{(1-u)\mathbb{Q}_1\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}{(1-u)\mathbb{Q}_1\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right) + u\mathbb{Q}_2\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}\sigma^2_1\left(\frac{i}{n}, \frac{\alpha k}{\sqrt{n}}\right)\\ &+ \frac{u\mathbb{Q}_2\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}{(1-u)\mathbb{Q}_1\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right) + u\mathbb{Q}_2\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}\sigma^2_2\left(\frac{i}{n}, \frac{\alpha k}{\sqrt{n}}\right)\end{aligned}$$

et

$$\begin{aligned}b_{1+u}\left(\frac{i}{n}, \frac{\alpha k}{\sqrt{n}}\right) &= \frac{(1-u)\mathbb{Q}_1\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}{(1-u)\mathbb{Q}_1\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right) + u\mathbb{Q}_2\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}b_1\left(\frac{i}{n}, \frac{\alpha k}{\sqrt{n}}\right)\\ &+ \frac{u\mathbb{Q}_2\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}{(1-u)\mathbb{Q}_1\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right) + u\mathbb{Q}_2\left(X_{\frac{i}{n}} = \frac{\alpha k}{\sqrt{n}}\right)}b_2\left(\frac{i}{n}, \frac{\alpha k}{\sqrt{n}}\right)\end{aligned}$$

On voit facilement sur ces formules que $(\sigma_{1+u}, b_{1+u}) \in \Sigma_0 \times \mathcal{B}_\varepsilon$.

Montrons que $\mathcal{D}^n_\varepsilon$ est un ouvert de $\mathcal{P}(\Omega_n)$.
Tout d'abord, on voit sans peine qu'il existe $c > 0$ ne dépendant que de σ_{min}, σ_{max}, b_0, s et α tel que, pour tout $\mathbb{Q} \in \mathcal{D}^n_\varepsilon$ et tout $|j| \le k \le n$,

$$\mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) > c.$$

Posons, quand cela est possible, pour $|j| \le k \le n$ et $\mathbb{Q} \in \mathcal{P}(\Omega_n)$:

$$F_{k,j}(\mathbb{Q}) = \alpha\sqrt{n}\frac{\mathbb{Q}\left(X_{\frac{k+1}{n}} = \frac{(j+1)\alpha}{\sqrt{n}}, X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) - \mathbb{Q}\left(X_{\frac{k+1}{n}} = \frac{(j-1)\alpha}{\sqrt{n}}, X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right)}{\mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right)} \tag{IV.16}$$

et

$$G_{k,j}(\mathbb{Q}) = \alpha^2 \frac{\mathbb{Q}\left(X_{\frac{k+1}{n}} = \frac{(j+1)\alpha}{\sqrt{n}}, X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) + \mathbb{Q}\left(X_{\frac{k+1}{n}} = \frac{(j-1)\alpha}{\sqrt{n}}, X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right)}{\mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right)} \tag{IV.17}$$

Ces applications sont continues sur l'ensemble ouvert

$$\left\{\mathbb{Q} \in \mathcal{P}(\Omega_n) : \forall |j| \leq k \leq n, \quad \mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) > c\right\}$$

et on voit facilement que

$$\mathbb{Q} \in \mathcal{D}_\varepsilon^n \Leftrightarrow \forall |j| \leq k \leq n, \quad \left\{\begin{array}{l} \mathbb{Q}\left(X_{\frac{k}{n}} = \frac{j\alpha}{\sqrt{n}}\right) > c, \\ F_{k,j}(Q) \in]b_0 - \varepsilon, b_0 + \varepsilon[, \\ G_{k,j}(Q) \in]\sigma^2_{\text{min}}, \sigma^2_{\text{max}}[. \end{array}\right.$$

On en déduit facilement que $\mathcal{D}_\varepsilon^n$ est ouvert dans $\mathcal{P}(\Omega_n)$. $\square$

Proposition IV.18.
Soit $\varepsilon > 0$; si $\mathcal{F}_\varepsilon^n$ est non vide, alors $\mathbb{R}^n_{\varepsilon,m}$ est bien définie pour m assez grand et converge quand m tend vers $+\infty$ vers la I-projection $\mathbb{Q}_\varepsilon^{n*}$ de $\mathbb{Q}^n_{\sigma_0, b_0}$ sur $\overline{\mathcal{F}_\varepsilon^n}$.

Cette proposition repose sur la version suivante du Principe Conditionnel de Gibbs :

Théorème IV.19.
Soient $\mathcal{X}$ un ensemble fini et μ une probabilité sur $\mathcal{X}$ chargeant tous les points de $\mathcal{X}$. Si C est un ensemble ouvert convexe non vide de $\mathcal{P}(\mathcal{X})$, alors $\mu_C^m = \mathbb{E}_{\mu^{\otimes m}}[L_m | L_m \in C]$ est bien définie pour m assez grand et converge lorsque m tend vers $+\infty$ vers la I-projection μ^* de μ sur $\overline{C}$.

Démonstration.
Comme μ charge tous les points de $\mathcal{X}$, on voit facilement que $\mathrm{H}(\nu|\,\mu) < +\infty$ pour toute $\nu \in \mathcal{P}(\mathcal{X})$. Par conséquent, $\mathrm{H}(C|\,\mu) < +\infty$. De plus, l'application $\nu \mapsto \mathrm{H}(\nu|\,\mu)$ est continue sur $\mathcal{P}(\mathcal{X})$, donc $\mathrm{H}(C|\,\mu) = \mathrm{H}\left(\overline{C}\middle|\,\mu\right)$. D'après la proposition III.9, on en déduit que $\mu_C^m \xrightarrow[m\to+\infty]{} \mu^*$. $\square$

Démonstration de la proposition IV.18.
L'ensemble $\mathcal{F}_\varepsilon^n$ est un ouvert convexe. De plus, on voit facilement que $\mathbb{Q}^n_{\sigma_0, b_0}$ charge tous les points de Ω_n. Le résultat découle donc directement du théorème IV.19. $\square$

IV.3.3 Etude des I-projections de $\mathbb{Q}^n_{\sigma_0,\, b_0}$ sur $\overline{\mathcal{F}^n_\varepsilon}$

Etude à n fixé

Comme on vient de le voir, si $\mathcal{F}^n_\varepsilon$ est non vide, la I-projection de $\mathbb{Q}^n_{\sigma_0,\, b_0}$ sur $\overline{\mathcal{F}^n_\varepsilon}$. Nous la noterons $\mathbb{Q}^{n*}_\varepsilon$. La proposition suivante établit que $\mathbb{Q}^{n*}_\varepsilon$ est un arbre trinomial issu de 0.

Proposition IV.20. *Posons pour tout $\sigma \in \overline{\Sigma_0}$, $b \in \overline{\mathcal{B}_\varepsilon}$, $\varepsilon \leq s$:*

$$q^n_{\sigma,\, b\,;\, \sigma_0,\, b_0}(t,x,y) = \frac{d\Pi^n_{\sigma,\, b}(t,x,\,.\,)}{d\Pi^n_{\sigma_0,\, b_0}(t,x,\,.\,)}(y)$$

et

$$h^n_{\sigma,\, b\,;\, \sigma_0,\, b_0}(t,x) = \mathrm{H}\left(\Pi^n_{\sigma,\, b}(t,x,\,.\,)\middle|\, \Pi^n_{\sigma_0,\, b_0}(t,x,\,.\,)\right).$$

Alors

1.
$$\frac{d\mathbb{Q}^n_{\sigma,\, b}}{d\mathbb{Q}^n_{\sigma_0,\, b_0}} = \prod_{i=0}^{n-1} q^n_{\sigma,\, b\,;\, \sigma_0,\, b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}, X_{\frac{i+1}{n}}\right)$$

2.
$$\mathrm{H}\left(\mathbb{Q}^n_{\sigma,\, b}\middle|\, \mathbb{Q}^n_{\sigma_0,\, b_0}\right) = \sum_{i=0}^{n-1} \mathbb{E}^n_{\sigma,\, b}\left[h^n_{\sigma,\, b\,;\, \sigma_0,\, b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right]$$

3. *Si $\mathbb{Q}$ vérifie la propriété (IV.12) pour des fonctions $\sigma \in \overline{\Sigma_0}$ et $b \in \overline{\mathcal{B}_\varepsilon}$, alors*
$$\mathcal{L}_{\mathbb{Q}}\left(X_{\frac{i}{n}}\right) = \mathcal{L}_{\mathbb{Q}^n_{\sigma,\, b}}\left(X_{\frac{i}{n}}\right),$$
pour tout $i = 0, \ldots, n-1$.
En particulier,
$$\int F\left(X_{\frac{[nT]}{n}}\right) d\mathbb{Q} = \int F\left(X_{\frac{[nT]}{n}}\right) d\mathbb{Q}^n_{\sigma,\, b}.$$

4. *De plus, on a la formule*
$$\mathrm{H}\left(\mathbb{Q}\middle|\, \mathbb{Q}^n_{\sigma_0,\, b_0}\right) = \mathrm{H}\left(\mathbb{Q}\middle|\, \mathbb{Q}^n_{\sigma,\, b}\right) + \mathrm{H}\left(\mathbb{Q}^n_{\sigma,\, b}\middle|\, \mathbb{Q}^n_{\sigma_0,\, b_0}\right).$$

5. *La I-projection de $\mathbb{Q}^n_{\sigma_0,\, b_0}$ sur $\overline{\mathcal{F}^n_\varepsilon} \neq \emptyset$, notée $\mathbb{Q}^{n*}_\varepsilon$ s'écrit sous la forme $\mathbb{Q}^n_{\sigma^*_n,\, b^*_n}$ avec $\sigma^*_n \in \overline{\Sigma_0}$ et $b \in \overline{\mathcal{B}_\varepsilon}$.*

Démonstration. (1) et (2) se vérifient simplement.

(3) Procédons par récurrence sur i :
- c'est vrai pour $i = 0$: $\mathcal{L}_{\mathbb{Q}}(X_0) = \mathcal{L}_{\mathbb{Q}^n_{\sigma,b}}(X_0) = \delta_0$.
- supposons que pour un certain $i \in \{0, \ldots, n-1\}$, on ait :

$$\mathcal{L}_{\mathbb{Q}}\left(X_{\frac{i}{n}}\right) = \mathcal{L}_{\mathbb{Q}^n_{\sigma,b}}\left(X_{\frac{i}{n}}\right),$$

Alors, pour toute fonction f continue,

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\left[f\left(X_{\frac{i+1}{n}}\right)\right] &= \mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{Q}}\left[f\left(X_{\frac{i+1}{n}}\right)\middle|X_{\frac{i}{n}}\right]\right] = \mathbb{E}_{\mathbb{Q}}\left[\Pi^n_{\sigma,b}\left(\frac{i}{n}, X_{\frac{i}{n}}, f\right)\right] \\
&= \mathbb{E}_{\mathbb{Q}^n_{\sigma,b}}\left[\Pi^n_{\sigma,b}\left(\frac{i}{n}, X_{\frac{i}{n}}, f\right)\right] = \mathbb{E}_{\mathbb{Q}^n_{\sigma,b}}\left[\mathbb{E}_{\mathbb{Q}^n_{\sigma,b}}\left[f\left(X_{\frac{i+1}{n}}\right)\middle|X_{\frac{i}{n}}\right]\right] \\
&= \mathbb{E}_{\mathbb{Q}^n_{\sigma,b}}\left[f\left(X_{\frac{i+1}{n}}\right)\right].
\end{aligned}$$

En particulier,

$$\int F\left(X_{\frac{[nT]}{n}}\right) d\mathbb{Q} = \int F\left(X_{\frac{[nT]}{n}}\right) d\mathbb{Q}^n_{\sigma,b}$$

et

$$\mathbb{Q} \in \mathcal{F}^n_\varepsilon \Leftrightarrow \mathbb{Q}^n_{\sigma,b} \in \mathcal{F}^n_\varepsilon.$$

(4)

$$\begin{aligned}
\mathrm{H}\left(\mathbb{Q}\middle|\,\mathbb{Q}^n_{\sigma_0,b_0}\right) &= \int \log\left(\frac{d\mathbb{Q}}{d\mathbb{Q}^n_{\sigma_0,b_0}}\right) d\mathbb{Q} \\
&= \int \log\left(\frac{d\mathbb{Q}}{d\mathbb{Q}^n_{\sigma,b}}\right) d\mathbb{Q} + \int \log\left(\frac{d\mathbb{Q}^n_{\sigma,b}}{d\mathbb{Q}^n_{\sigma_0,b_0}}\right) d\mathbb{Q} \\
&= \mathrm{H}\left(\mathbb{Q}\middle|\,\mathbb{Q}^n_{\sigma,b}\right) + \int \log\left(\frac{d\mathbb{Q}^n_{\sigma,b}}{d\mathbb{Q}^n_{\sigma_0,b_0}}\right) d\mathbb{Q}
\end{aligned}$$

Mais

$$\begin{aligned}
&\int \log\left(\frac{d\mathbb{Q}^n_{\sigma,b}}{d\mathbb{Q}^n_{\sigma_0,b_0}}\right) d\mathbb{Q} = \int \sum_{i=0}^{n-1} \log\left[q^n_{\sigma,b\,;\,\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}, X_{\frac{i+1}{n}}\right)\right] d\mathbb{Q} \\
&= \mathbb{E}_{\mathbb{Q}}\left[\sum_{i=0}^{n-1} \int \log\left[q^n_{\sigma,b\,;\,\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}, y\right)\right] \Pi^n_{\sigma,b}\left(\frac{i}{n}, X_{\frac{i}{n}}, dy\right)\right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}_{\mathbb{Q}}\left[h^n_{\sigma,b\,;\,\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right] = \sum_{i=0}^{n-1} \mathbb{E}^n_{\sigma,b}\left[h^n_{\sigma,b\,;\,\sigma_0,b_0}\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right] \\
&= \mathrm{H}\left(\mathbb{Q}^n_{\sigma,b}\middle|\,\mathbb{Q}^n_{\sigma_0,b_0}\right)
\end{aligned}$$

D'où la formule :

$$\mathrm{H}\left(\mathbb{Q}\middle|\, \mathbb{Q}^n_{\sigma_0,\, b_0}\right) = \mathrm{H}\left(\mathbb{Q}\middle|\, \mathbb{Q}^n_{\sigma,\, b}\right) + \mathrm{H}\left(\mathbb{Q}^n_{\sigma,\, b}\middle|\, \mathbb{Q}^n_{\sigma_0,\, b_0}\right).$$

(5) Comme $\mathbb{Q}^{n*}_\varepsilon$ appartient à $\overline{\mathcal{F}^n_\varepsilon}$, il existe un couple $(\sigma^*_n, b^*_n) \in \overline{\Sigma_0} \times \overline{\mathcal{B}_\varepsilon}$ tel que $\mathbb{Q}^{n*}_\varepsilon$ vérifie (IV.12). D'après le point (4),

$$\mathrm{H}\left(\mathbb{Q}^{n*}_\varepsilon\middle|\, \mathbb{Q}^n_{\sigma_0,\, b_0}\right) = \mathrm{H}\left(\mathbb{Q}^{n*}_\varepsilon\middle|\, \mathbb{Q}^n_{\sigma^*_n,\, b^*_n}\right) + \mathrm{H}\left(\mathbb{Q}^n_{\sigma^*_n,\, b^*_n}\middle|\, \mathbb{Q}^n_{\sigma_0,\, b_0}\right)$$

Pour conclure, il suffit donc de montrer que $\mathbb{Q}^n_{\sigma^*_n,\, b^*_n} \in \overline{\mathcal{F}^n_\varepsilon}$.
Soit $\left(\mathbb{Q}_p\right)_p$ une suite d'éléments de $\mathcal{F}^n_\varepsilon$ convergeant vers $\mathbb{Q}^{n*}_\varepsilon$. Chaque $\mathbb{Q}_p$ est associée à un couple $(\sigma_p, b_p) \in \Sigma_0 \times \mathcal{B}_\varepsilon$. Or, pour tout $|j| \leq k \leq n$,

$$b_p\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right) = F_{k,\, j}(\mathbb{Q}_p)$$

et

$$\sigma_p^2\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right) = G_{k,\, j}(\mathbb{Q}_p),$$

où les fonctions $F_{k,\, j}$ et $G_{k,\, j}$ sont définies par (IV.16) et (IV.17). Ces fonctions étant continues, on a, pour tout $|j| \leq k \leq n$,

$$b_p\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right) \xrightarrow[p\to+\infty]{} b^*_n\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right)$$

et

$$\sigma_p^2\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right) \xrightarrow[p\to+\infty]{} (\sigma^*_n)^2\left(\frac{k}{n}, \frac{\alpha j}{\sqrt{n}}\right).$$

On en déduit aisément que

$$\mathbb{Q}^n_{\sigma_p,\, b_p} \xrightarrow[p\to+\infty]{} \mathbb{Q}^n_{\sigma^*_n,\, b^*_n}.$$

D'après le point (3),

$$\mathbb{Q}_p \in \mathcal{F}^n_\varepsilon \Rightarrow \mathbb{Q}^n_{\sigma_p,\, b_p} \in \mathcal{F}^n_\varepsilon,$$

ce qui prouve que $\mathbb{Q}^n_{\sigma^*_n,\, b^*_n}$ est adhérent à $\mathcal{F}^n_\varepsilon$. $\square$

Etude asymptotique

Dans cette section on étudie, pour un bon choix de $(\varepsilon_n)_n$ les valeurs d'adhérence de $\left(\mathbb{Q}^{n*}_{\varepsilon_n}\right)_n$.

Convergence de $\frac{1}{n}\mathrm{H}\left(\,.\,\middle|\,\mathbb{Q}^n_{\sigma_0,\,b_0}\right)$.
Pour $\sigma \in \overline{\Sigma_0}$, on pose :

$$\mathrm{I}(\sigma|\sigma_0) = \mathbb{E}_{\sigma,\,b}\left[\int_0^1 q(\sigma^2(t,X_t),\sigma_0^2(t,X_t))\,dt\right],$$

avec

$$q(x,y) = \log\left(\frac{x}{y}\right)\frac{x}{\alpha^2} + \log\left(\frac{\alpha^2-x}{\alpha^2-y}\right)\left[1-\frac{x}{\alpha^2}\right].$$

Proposition IV.21.

1. *Si $(\varepsilon_n)_n$ est une suite de réels positifs convergeant vers zéro, alors pour toute suite $b_n \in \mathcal{B}_{\varepsilon_n}$, et tout $\sigma \in \overline{\Sigma_0}$, on a :*

$$\frac{\mathrm{H}\left(\mathbb{Q}^n_{\sigma,\,b_n}\middle|\,\mathbb{Q}^n_{\sigma_0,\,b_0}\right)}{n} \xrightarrow[n\to+\infty]{} \mathrm{I}(\sigma|\sigma_0).$$

2. *Si $(\sigma_n)_n$ est une suite d'éléments de $\overline{\Sigma_0}$ convergeant vers $\sigma \in \overline{\Sigma_0}$ uniformément sur tout compact, alors, sous les mêmes hypothèses*

$$\liminf_{n\to+\infty}\frac{\mathrm{H}\left(\mathbb{Q}^n_{\sigma_n,\,b_n}\middle|\,\mathbb{Q}^n_{\sigma_0,\,b_0}\right)}{n} \geq \mathrm{I}(\sigma|\sigma_0).$$

Démonstration.
1. Montrons qu'il existe une constante $K > 0$, ne dépendant que de α, $\sigma_{\min}$, $\sigma_{\max}$, b_0 et s, telle que :

$$\left|h^n_{\sigma,\,b\,;\,\sigma_0,\,b_0} - q(\sigma^2,\sigma_0^2)\right|\left(\frac{k}{n},x\right) \leq \frac{K}{n} \tag{IV.22}$$

pour tout $(k,x) \in \{0,\ldots,n-1\} \times \frac{\alpha}{\sqrt{n}}\mathbb{Z}$ et $(\sigma,b) \in \overline{\Sigma_0} \times \overline{\mathcal{B}_s}$.
En effet, pour tout $(\sigma,b) \in \overline{\Sigma_0} \times \overline{\mathcal{B}_s}$:

$$\begin{aligned}\log\left[\frac{m^n(\sigma,b)}{m^n(\sigma_0,b_0)}\right] m^n(\sigma,b) = &\left[\log\left(\frac{\sigma^2}{\sigma_0^2}\right) + \log\left(1+\frac{b\alpha}{\sqrt{n}\sigma^2}\right) - \log\left(1+\frac{b_0\alpha}{\sqrt{n}\sigma_0^2}\right)\right]\\ &\times\left[\frac{\sigma^2}{2\alpha^2}+\frac{b}{2\alpha\sqrt{n}}\right]\end{aligned}$$

$$\log\left[\frac{d^n(\sigma,b)}{d^n(\sigma_0,b_0)}\right] d^n(\sigma,b) = \left[\log\left(\frac{\sigma^2}{\sigma_0^2}\right) + \log\left(1-\frac{b\alpha}{\sqrt{n}\sigma^2}\right) - \log\left(1-\frac{b_0\alpha}{\sqrt{n}\sigma_0^2}\right)\right] \times \left[\frac{\sigma^2}{2\alpha^2} - \frac{b}{2\alpha\sqrt{n}}\right]$$

$$\log\left[\frac{r^n(\sigma,b)}{r^n(\sigma_0,b_0)}\right] r^n(\sigma,b) = \log\left(\frac{\alpha^2-\sigma^2}{\alpha^2-\sigma_0^2}\right)\left[1-\frac{\sigma^2}{\alpha^2}\right]$$

Or, on voit sans peine, en écrivant la formule de Taylor avec reste intégral à l'ordre 2, que pour $\varepsilon \in \{-1,1\}$,

$$\sup_{\substack{x\in[\sigma^2_{\min},\sigma^2_{\max}]\\ y\in[b_0-s,b_0+s]}} \left|\log\left(1+\frac{\varepsilon y\alpha}{\sqrt{n}x}\right) - \frac{\varepsilon y\alpha}{\sqrt{n}x} + \frac{1}{2}\left(\frac{\varepsilon y\alpha}{\sqrt{n}x}\right)^2\right| \le \frac{K}{n\sqrt{n}},$$

avec K qui ne dépend que de α, $\sigma_{\max}$, $\sigma_{\min}$, b_0 et s.
On en déduit (IV.22), après quelques calculs.

Posons

$$\Phi^n = \frac{1}{n}\sum_{i=0}^{n-1} q\left(\sigma^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right), \sigma_0^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right)$$

et

$$\Phi = \int_0^1 q(\sigma^2(t,X_t), \sigma_0^2(t,X_t))\,dt.$$

La fonction q est continue bornée sur $[\sigma^2_{\min}, \sigma^2_{\max}]^2$.
La suite $(\Phi^n)_n$ est une suite de fonctions continues sur Ω, uniformément bornées, convergeant simplement vers Φ, qui est aussi continue bornée sur Ω.
Montrons que la convergence de Φ^n vers Φ est uniforme sur tout compact. La fonction q est Lipschitzienne sur $[\sigma^2_{\min}, \sigma^2_{\max}]^2$; nous noterons M une constante telle que

$$|q(x,y) - q(x',y')| \le M(|x-x'| + |y-y'|).$$

Nous noterons Δ le module de continuité de σ^2, *ie*

$$\Delta(u) = \sup_{|t-s|+|y-x|\le u} |\sigma^2(s,x) - \sigma^2(t,y)|,$$

et Δ_0 celui de σ_0^2.

Avec ces notations, on a

$$\begin{aligned}
|\Phi^n - \Phi| &= \left| \frac{1}{n} \sum_{i=0}^{n-1} q\left(\sigma^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right), \sigma_0^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right) - \int_0^1 q(\sigma^2(t, X_t), \sigma_0^2(t, X_t))\, dt \right| \\
&\leq \sum_{i=0}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n}} \left| q\left(\sigma^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right), \sigma_0^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right) - q(\sigma^2(t, X_t), \sigma_0^2(t, X_t)) \right| dt \\
&\leq M \sum_{i=0}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n}} \left| \sigma^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right) - \sigma^2(t, X_t) \right| + \left| \sigma_0^2\left(\frac{i}{n}, X_{\frac{i}{n}}\right) - \sigma_0^2(t, X_t) \right| dt \\
&\leq M \left[\sup_{|s-t| \leq \frac{1}{n}} \left| \sigma^2(s, X_s) - \sigma^2(t, X_t) \right| + \sup_{|s-t| \leq \frac{1}{n}} \left| \sigma_0^2(s, X_s) - \sigma_0^2(t, X_t) \right| \right] \\
&\leq M \left[\sup_{|s-t| \leq \frac{1}{n}} \Delta\left(|s-t| + |X_s - X_t|\right) + \sup_{|s-t| \leq \frac{1}{n}} \Delta_0\left(|s-t| + |X_s - X_t|\right) \right] \\
&\leq M \left[\Delta\left(\frac{1}{n} + \sup_{|s-t| \leq \frac{1}{n}} |X_s - X_t|\right) + \Delta_0\left(\frac{1}{n} + \sup_{|s-t| \leq \frac{1}{n}} |X_s - X_t|\right) \right]
\end{aligned}$$

D'après le théorème d'Ascoli, si $\mathcal{A}$ est un compact de Ω, alors

$$\sup_{\omega \in \mathcal{A}} \sup_{|t-s| \leq \frac{1}{n}} |X_s - X_t| \xrightarrow[n \to +\infty]{} 0.$$

On en déduit que

$$\sup_{\omega \in \mathcal{A}} |\Phi^n(\omega) - \Phi(\omega)| \xrightarrow[n \to +\infty]{} 0.$$

On a, d'après (IV.22) :

$$\left| \frac{1}{n} \mathrm{H}\left(\mathbb{Q}^n_{\sigma, b_n} \middle| \mathbb{Q}^n_{\sigma_0, b_0}\right) - \mathbb{E}^n_{\sigma, b_n}\left[\Phi^n\right] \right| \leq \frac{K}{n}$$

où K ne dépend que de α, $\sigma_{\max}$, $\sigma_{\min}$, b_0 et s. On en déduit facilement, en utilisant la convergence uniforme sur tout compact de la suite $(\Phi^n)_n$ et la tension de la suite $\mathbb{Q}^n_{\sigma, b_n}$ (Ω est polonais) que :

$$\lim_{n \to \infty} \frac{1}{n} \mathrm{H}\left(\mathbb{Q}^n_{\sigma, b_n} \middle| \mathbb{Q}^n_{\sigma_0, b_0}\right) = \mathrm{I}(\sigma | \sigma_0).$$

2.

$$
\begin{aligned}
\frac{1}{n}\,\mathrm{H}\left(\mathbb{Q}^n_{\sigma_n,\,b_n}\middle|\,\mathbb{Q}^n_{\sigma_0,\,b_0}\right) &= \frac{1}{n}\int \log\left(\frac{d\mathbb{Q}^n_{\sigma_n,\,b_n}}{d\mathbb{Q}^n_{\sigma_0,\,b_0}}\right) d\mathbb{Q}^n_{\sigma_n,\,b_n} \\
&= \frac{1}{n}\int \log\left(\frac{d\mathbb{Q}^n_{\sigma_n,\,b_n}}{d\mathbb{Q}^n_{\sigma,\,b_n}}\right) d\mathbb{Q}^n_{\sigma_n,\,b_n} + \frac{1}{n}\int \log\left(\frac{d\mathbb{Q}^n_{\sigma,\,b_n}}{d\mathbb{Q}^n_{\sigma_0,\,b_0}}\right) d\mathbb{Q}^n_{\sigma_n,\,b_n} \\
&= \frac{1}{n}\,\mathrm{H}\left(Q^n_{\sigma_n,\,b_n}\middle|\,Q^n_{\sigma,\,b_n}\right) + \frac{1}{n}\int \log\left(\frac{d\mathbb{Q}^n_{\sigma,\,b_n}}{d\mathbb{Q}^n_{\sigma_0,\,b_0}}\right) d\mathbb{Q}^n_{\sigma_n,\,b_n} \\
&\geq \frac{1}{n}\int \log\left(\frac{d\mathbb{Q}^n_{\sigma,\,b_n}}{d\mathbb{Q}^n_{\sigma_0,\,b_0}}\right) d\mathbb{Q}^n_{\sigma_n,\,b_n}
\end{aligned}
$$

D'après la proposition IV.20,

$$
\frac{1}{n}\int \log\left(\frac{d\mathbb{Q}^n_{\sigma,\,b_n}}{d\mathbb{Q}^n_{\sigma_0,\,b_0}}\right) d\mathbb{Q}^n_{\sigma_n,\,b_n} = \mathbb{E}^n_{\sigma_n,\,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1} k^n\left(\frac{i}{n}, X_{\frac{i}{n}}\right)\right],
$$

en posant

$$
\begin{aligned}
k^n = \log\left(\frac{m^n(\sigma,b_n)}{m^n(\sigma_0,b_0)}\right) m^n(\sigma_n,b_n) + \log\left(\frac{r^n(\sigma,b_n)}{r^n(\sigma_0,b_0)}\right) r^n(\sigma_n,b_n) \\
+ \log\left(\frac{d^n(\sigma,b_n)}{d^n(\sigma_0,b_0)}\right) d^n(\sigma_n,b_n)
\end{aligned}
$$

On voit facilement qu'il existe une constante K ne dépendant que de α, σ_{min}, σ_{max}, b_0 et s telle que pour tout $R > 0$,

$$
\sup_{|x|\leq R,\, t\in[0,1]} |k^n - h^n_{\sigma,\,b_n\,;\,\sigma_0,\,b_0}|(t,x) \leq K \sup_{|x|\leq R,\, t\in[0,1]} |\sigma_n - \sigma|(t,x).
$$

Comme $\mathbb{Q}^n_{\sigma_n,\,b_n}$ converge étroitement vers $\mathbb{Q}_{\sigma,\,b}$, c'est une suite tendue. On en déduit, en particulier, que pour tout $\beta > 0$, il existe $R > 0$ tel que

$$
\mathbb{Q}^n_{\sigma_n,\,b_n}\left(\sup_{t\in[0,1]} |X_t| \leq R\right) \geq 1 - \beta.
$$

Par suite, comme $|k^n|$ et $\left|h^n_{\sigma,\,b_n\,;\,\sigma_0,\,b_0}\right|$ sont bornées par M ne dépendant que de α, σ_{min},

$\sigma_{\max}$, b_0 et s, on a

$$\begin{aligned}
&\left|\mathbb{E}^n_{\sigma_n,\,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1}k^n\left(\frac{i}{n},X_{\frac{i}{n}}\right)\right]-\mathbb{E}^n_{\sigma_n,\,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1}h^n_{\sigma,\,b_n\,;\,\sigma_0,\,b_0}\left(\frac{i}{n},X_{\frac{i}{n}}\right)\right]\right|\\
&\leq \mathbb{E}^n_{\sigma_n,\,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1}\left|h^n_{\sigma,\,b_n\,;\,\sigma_0,\,b_0}-k^n\right|\mathbb{1}_{[0,R]}(\sup_{t\in[0,1]}|X_t|)\right]+2M(1-\beta)\\
&\leq K\sup_{|x|\leq R,\,t\in[0,1]}|\sigma_n-\sigma|(t,x)+2M(1-\beta).
\end{aligned}$$

On en déduit que

$$\mathbb{E}^n_{\sigma_n,\,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1}k^n\left(\frac{i}{n},X_{\frac{i}{n}}\right)\right]-\mathbb{E}^n_{\sigma_n,\,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1}h^n_{\sigma,\,b_n\,;\,\sigma_0,\,b_0}\left(\frac{i}{n},X_{\frac{i}{n}}\right)\right]\xrightarrow[n\to+\infty]{}0$$

et le même raisonnement qu'au point 1. montre que

$$\mathbb{E}^n_{\sigma_n,\,b_n}\left[\frac{1}{n}\sum_{i=0}^{n-1}h^n_{\sigma,\,b_n\,;\,\sigma_0,\,b_0}\left(\frac{i}{n},X_{\frac{i}{n}}\right)\right]\xrightarrow[n\to+\infty]{}\mathrm{I}(\sigma|\sigma_0).$$

$\square$

Convergence des I-projections.
Notons

$$\mathcal{M}_F=\mathrm{Argmin}\left\{\mathrm{I}(\sigma|\sigma_0),\quad \sigma\in\overline{\Sigma_0},\quad \int F(X_T)\,d\mathbb{Q}_{\sigma,\,b_0}=1\right\}$$

et supposons que

$$\mathcal{M}_F\cap\Sigma_0\neq\emptyset.$$

Soit $\bar\sigma\in\mathcal{M}_F\cap\Sigma_0$, on pose

$$\varepsilon_n=\min\left(\left|\int F\left(X_{\frac{[nT]}{n}}\right)d\mathbb{Q}^n_{\bar\sigma,\,b_0}-1\right|+1/n,s\right).$$

La suite $(\varepsilon_n)_n$ est une suite de réels strictement positifs majorés par s et convergeant vers zéro.

Proposition IV.23. *Supposons qu'il existe une suite $(\sigma^*_n)_n$ d'éléments de $\overline{\Sigma_0}$, précompacte dans $\overline{\Sigma_0}$ (pour la topologie de la convergence uniforme sur tout compact) et une suite $(b^*_n)_n$ d'éléments de $\overline{\mathcal{B}_{\varepsilon_n}}$ telles que $\mathbb{Q}^{n*}_{\varepsilon_n}=\mathbb{Q}^n_{\sigma^*_n,\,b^*_n}$. Alors les valeurs d'adhérence de $(\mathbb{Q}^{n*}_{\varepsilon_n})_n$ sont de la forme $\mathbb{Q}_{\sigma^*,\,b_0}$, avec $\sigma^*\in\mathcal{M}_F$.*

Démonstration.

Grâce à la précompacité de la suite σ_n^*, on voit facilement, d'après la proposition IV.10, que la suite $\left(\mathbb{Q}_{\varepsilon_n}^{n*}\right)_n$ est précompacte et que ses valeurs d'adhérence sont de la forme $\mathbb{Q}_{\sigma^*,\, b_0}$, avec $\sigma^* \in \overline{\Sigma_0}$ tel que $\int F(X_T)\, d\mathbb{Q}_{\sigma^*,\, b_0} = 1$.

Prenons $\mathbb{Q}_{\sigma^*,\, b_0}$ une valeur d'adhérence et $\varphi : \mathbb{N} \to \mathbb{N}$ strictement croissante telle que $\sigma^*_{\varphi(n)} \xrightarrow[n\to+\infty]{} \sigma^*$.
Comme $\mathbb{Q}^n_{\bar\sigma,\, b_0} \in \mathcal{F}^n_{\varepsilon_n}$, on a

$$\frac{\mathrm{H}\left(\mathbb{Q}^{\varphi(n)*}_{\varepsilon_{\varphi(n)}} \middle| \mathbb{Q}^{\varphi(n)}_{\sigma_0,\, b_0}\right)}{\varphi(n)} \le \frac{\mathrm{H}\left(\mathbb{Q}^{\varphi(n)}_{\bar\sigma,\, b_0} \middle| \mathbb{Q}^{\varphi(n)}_{\sigma_0,\, b_0}\right)}{\varphi(n)}.$$

Le membre de droite converge vers $\mathrm{I}(\bar\sigma|\sigma_0)$ et, d'après la proposition IV.21,

$$\liminf_{n\to+\infty} \frac{\mathrm{H}\left(\mathbb{Q}^{\varphi(n)*}_{\varepsilon_{\varphi(n)}} \middle| \mathbb{Q}^{\varphi(n)}_{\sigma_0,\, b_0}\right)}{\varphi(n)} \ge \mathrm{I}(\sigma^*|\sigma_0).$$

Donc

$$\mathrm{I}(\sigma^*|\sigma_0) \le \mathrm{I}(\bar\sigma|\sigma_0)$$

et par conséquent $\sigma^* \in \mathcal{M}_F$. □

IV.3.4 Principe conditionnel de Gibbs (suite et fin)

Un premier résultat de convergence pour les arbres trinomiaux

Nous pouvons à présent démontrer la

Proposition IV.24. *Supposons que l'ensemble $\mathcal{M}_F \cap \Sigma_0 \neq \emptyset$ et posons*

$$\varepsilon_n = \min\left(\left|\int F\left(X_{\frac{[nT]}{n}}\right) d\mathbb{Q}^n_{\bar\sigma, b_0} - 1\right| + 1/n, s\right),$$

où $\bar\sigma$ est un élément de $\mathcal{M}_F \cap \Sigma_0$. Supposons de plus qu'il existe une suite $(\sigma_n^*)_n$ d'éléments de $\overline{\Sigma_0}$, précompacte dans $\overline{\Sigma_0}$ (pour la topologie de la convergence uniforme sur tout compact) et une suite $(b_n^*)_n$ d'éléments de $\overline{\mathcal{B}_{\varepsilon_n}}$ telles que la I-projection $\mathbb{Q}^{n*}_{\varepsilon_n}$ de $\mathbb{Q}^n_{\sigma_0,\, b_0}$ sur $\overline{\mathcal{F}^n_{\varepsilon_n}}$ s'écrive $\mathbb{Q}^{n*}_{\varepsilon_n} = \mathbb{Q}^n_{\sigma_n^*,\, b_n^*}$. Sous ces hypothèses, il existe au moins une suite $(m_n)_n$ d'entiers, $m_n \xrightarrow[n\to+\infty]{} +\infty$ telle que les valeurs d'adhérence de la suite $\left(\mathbb{R}^n_{\varepsilon_n,\, m_n}\right)_n$ soient de la forme $\mathbb{Q}_{\sigma^*,\, b_0}$, avec $\sigma^* \in \mathcal{M}_F$.

Démonstration.
Tout d'abord, $\mathbb{Q}^n_{\bar{\sigma},\,b_0} \in \mathcal{F}^n_{\varepsilon_n}$. L'ensemble $\mathcal{F}^n_{\varepsilon_n}$ étant non vide, $\mathbb{Q}^{n*}_{\varepsilon_n}$ est bien définie. D'après la proposition IV.18,

$$\mathbb{R}^n_{\varepsilon_n,\,m} \xrightarrow[m \to +\infty]{} \mathbb{Q}^{n*}_{\varepsilon_n},$$

dans $\mathcal{P}(\Omega_n)$. On voit facilement, en utilisant un théorème de prolongement des fonctions continues, que la convergence a lieu également dans $\mathcal{P}(\Omega)$. Si $d_{FM}(\,.\,,\,.\,)$ désigne la distance de Fortet-Mourier sur $\mathcal{P}(\Omega)$, il existe donc m_n tel que

$$d_{FM}\left(\mathbb{R}^n_{\varepsilon_n,\,m_n}, \mathbb{Q}^{n*}_{\varepsilon_n}\right) \xrightarrow[n \to +\infty]{} 0.$$

Par conséquent $\left(\mathbb{R}^n_{\varepsilon_n,\,m_n}\right)_n$ et $\left(\mathbb{Q}^{n*}_{\varepsilon_n}\right)_n$ ont les mêmes valeurs d'adhérence dans $\mathcal{P}(\Omega)$. D'après la proposition IV.23, celles-ci sont de la forme $\mathbb{Q}_{\sigma^*,\,b_0}$, avec $\sigma^* \in \mathcal{M}_F$. □

Remarque IV.25.

L'hypothèse selon laquelle les I-projections $\mathbb{Q}^{n*}_{\varepsilon_n}$ s'écrivent sous la forme $\mathbb{Q}^n_{\sigma^*_n,\,b^*_n}$, avec σ^*_n une suite précompacte de Σ_0 est difficilement vérifiable. Une idée naturelle pour éviter cette hypothèse serait de remplacer dans la définition de $\mathcal{D}^n_\varepsilon$ l'ensemble Σ_0 par un sous-ensemble compact (pour la topologie de la convergence uniforme). Cela conduit à une autre difficulté : $\mathcal{D}^n_\varepsilon$ n'est plus convexe. En effet, en se reportant à la preuve de la proposition IV.15, on voit que la propriété assurant la convexité de $\mathcal{D}^n_\varepsilon$ est la suivante :

Si $\sigma_1, \sigma_2 \in \Sigma_0$ et si pour tout $t \in \left\{0, \frac{1}{n}, \dots, 1\right\}$ et tout $x \in \left\{\frac{\alpha k}{\sqrt{n}},\ k \in [-n, n]\right\}$, $\varepsilon_{t,\,x} \in [0,1]$, alors il existe $\sigma \in \Sigma_0$ telle que

$$\sigma^2(t,x) = \varepsilon_{t,\,x}\sigma_1^2(t,x) + (1 - \varepsilon_{t,\,x})\sigma_2^2(t,x), \tag{IV.26}$$

pour tout $(t,x) \in \left\{0, \frac{1}{n}, \dots, 1\right\} \times \left\{\frac{\alpha k}{\sqrt{n}},\ k = -n \dots n\right\}$.

Clairement, (IV.26) ne peut pas être satisfaite par un sous-ensemble compact de Σ_0 non réduit à un point.
Avant de voir dans quelle mesure on peut se passer de la convexité de $\mathcal{D}^n_\varepsilon$, remarquons que celle-ci découle de la forme très particulière des noyaux de transitions utilisés pour définir les arbres trinomiaux (plus précisément leur linéarité par rapport à (σ^2, b)). Si par exemple, $\mathbb{Q}^n_{\sigma,\,b}$ est un schéma d'Euler, $\Pi^n_{\sigma,\,b}\left(\frac{i}{n}, x, .\right)$ est une loi gaussienne. Une combinaison convexe de lois gaussiennes n'étant plus gaussienne, on voit, en se reportant à la preuve de la proposition IV.15, que $\mathcal{D}^n_\varepsilon$ n'est plus convexe.

Un second résultat de convergence pour les arbres trinomiaux

Nous ferons l'hypothèse suivante :

$$\mathcal{M}_F = \{\sigma^*\}, \qquad \text{avec} \qquad \sigma^* \in \Sigma_0.$$

Pour tout $\sigma \in \overline{\Sigma_0}$, désignons par $\Delta_{n,\sigma}$ le module de continuité de σ sur le compact $[0,1] \times [-\alpha\sqrt{n}, \alpha\sqrt{n}]$, *ie*

$$\Delta_{n,\sigma}(\varepsilon) = \sup\Big\{|\sigma(t,x) - \sigma(s,y)| : s,t \in [0,1], x,y \in \left[-\alpha\sqrt{n}, \alpha\sqrt{n}\right], \\ |t-s| + |x-y| \leq \varepsilon\Big\}$$

et posons

$$\Sigma_1 = \{\sigma \in \Sigma_0 : \forall n \in \mathbb{N}^*, \quad \Delta_{n,\sigma} < 2\Delta_{n,\sigma^*}\}.$$

D'après le théorème d'Ascoli, on voit facilement que Σ_1 est précompact pour la topologie de la convergence uniforme sur tout compact.

L'ensemble $\mathcal{D}^n_{\varepsilon,\Sigma_1}$ est l'ensemble des probabilités $\mathbb{Q}$ sur Ω vérifiant

$$\left\{\begin{array}{ll} (1) & \mathbb{Q}(X_0 = 0) = 1, \\ (2) & \mathbb{Q}\left(X_t = X_{\frac{k}{n}} + (nt-k)\left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}}\right], \frac{k}{n} \leq t \leq \frac{k+1}{n}\right) = 1, \\ (3') & \exists(\sigma,b) \in \Sigma_1 \times \mathcal{B}_\varepsilon \text{ tels que } \mathbb{Q}\left(X_{\frac{p+1}{n}} \in \,.\, \middle| X_{\frac{p}{n}}\right) = \Pi^n_{\sigma,b}\left(\frac{p}{n}, X_{\frac{p}{n}}, \,.\,\right) \end{array}\right. \tag{IV.27}$$

Nous poserons

$$\mathcal{F}^n_{\varepsilon,\Sigma_1} = \mathcal{E}^n_\varepsilon \cap \mathcal{D}^n_{\varepsilon,\Sigma_1} \tag{IV.28}$$

et

$$\mathbb{R}^n_{\varepsilon,m} = \mathbb{E}_{(\mathbb{Q}^n_{\sigma_0,b_0})^{\otimes m}}\left[L_m \,\middle|\, L_m \in \mathcal{F}^n_{\varepsilon,\Sigma_1}\right].$$

On a alors le théorème suivant

Théorème IV.29. *Si* $\varepsilon_n = \min\left(\left|\int F\left(X_{\frac{[nT]}{n}}\right) d\mathbb{Q}^n_{\sigma^*,b_0} - 1\right| + 1/n, s\right)$*, alors il existe au moins une suite* $(m_n)_n$ *d'entiers,* $m_n \xrightarrow[n\to+\infty]{} +\infty$ *telle que la suite* $\left(\mathbb{R}^n_{\varepsilon_n,m_n}\right)_n$ *converge vers* $\mathbb{Q}_{\sigma^*,b_0}$.

Démonstration. L'ensemble $\mathcal{D}^n_{\varepsilon,\Sigma_1}$ est ouvert ; en effet, on voit facilement que $\mathcal{D}^n_{\varepsilon,\Sigma_1}$ est l'intersection de l'ouvert $\mathcal{D}^n_\varepsilon$ et de l'ensemble des probabilités $\mathbb{Q} \in \mathcal{P}(\Omega_n)$ vérifiant pour tout $|j| \leq k \leq n$,

$$\left|\sqrt{G_{k,j}(\mathbb{Q})} - \sqrt{G_{p,q}(\mathbb{Q})}\right| < 2\Delta_{n,\sigma^*}\left(\left|\frac{k}{n} - \frac{p}{n}\right| + \left|\frac{\alpha j}{\sqrt{n}} - \frac{\alpha q}{\sqrt{n}}\right|\right),$$

où les fonctions $G_{k,j}$ sont définies par (IV.17). On en déduit facilement que $\mathcal{D}^n_{\varepsilon,\Sigma_1}$ est ouvert. L'ensemble $\mathcal{F}^n_{\varepsilon_n,\Sigma_1}$ est donc lui aussi ouvert dans $\mathcal{P}(\Omega_n)$ et contient $\mathbb{Q}^n_{\sigma^*,b_0}$. La fonction $\mathcal{P}(\Omega_n) \to \mathbb{R}^+ \cup \{+\infty\} : \mathbb{Q} \mapsto \mathrm{H}\left(\mathbb{Q} \middle| \mathbb{Q}^n_{\sigma_0,b_0}\right)$ étant convexe et partout finie ($\mathbb{Q}^n_{\sigma_0,b_0}$ charge tous les points de Ω_n), elle est continue ($\mathcal{P}(\Omega_n)$ est un simplexe de $\mathbb{R}^{3^n}$), et on a

$$\mathrm{H}\left(\mathcal{F}^n_{\varepsilon_n,\Sigma_1} \middle| \mathbb{Q}^n_{\sigma_0,b_0}\right) = \mathrm{H}\left(\overline{\mathcal{F}^n_{\varepsilon_n,\Sigma_1}} \middle| \mathbb{Q}^n_{\sigma_0,b_0}\right).$$

D'après le point 2 de la proposition III.9, la suite $\left(\mathbb{R}^n_{\varepsilon_n,m}\right)_m$ est bien définie pour m assez grand et on a

$$d_{FM}\left(\mathbb{R}^n_{\varepsilon_n,m}, \overline{\mathrm{co}}\,\mathcal{M}^n_F\right) \xrightarrow[m\to+\infty]{} 0,$$

où $\mathcal{M}^n_F$ désigne

$$\left\{\mathbb{Q} \in \overline{\mathcal{F}^n_{\varepsilon_n,\Sigma_1}} : \mathrm{H}\left(\mathbb{Q} \middle| \mathbb{Q}^n_{\sigma_0,b_0}\right) = \mathrm{H}\left(\mathcal{F}^n_{\varepsilon_n,\Sigma_1} \middle| \mathbb{Q}^n_{\sigma_0,b_0}\right)\right\}.$$

Comme,

$$d_{FM}\left(\mathbb{R}^n_{\varepsilon_n,m}, \mathbb{Q}_{\sigma^*,b_0}\right) \leq d_{FM}\left(\mathbb{R}^n_{\varepsilon_n,m}, \overline{\mathrm{co}}\,\mathcal{M}^n_F\right) + \sup_{\mathbb{Q}\in\overline{\mathrm{co}}\,\mathcal{M}^n_F} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right),$$

il suffit de montrer que

$$\sup_{\mathbb{Q}\in\overline{\mathrm{co}}\,\mathcal{M}^n_F} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right) \xrightarrow[n\to+\infty]{} 0.$$

L'application $\mathbb{Q} \mapsto d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right)$ étant convexe et continue, on a

$$\sup_{\mathbb{Q}\in\overline{\mathrm{co}}\,\mathcal{M}^n_F} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right) = \sup_{\mathbb{Q}\in\mathcal{M}^n_F} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right).$$

L'ensemble $\mathcal{M}^n_F$ étant compact, il existe $\mathbb{Q}^{n*} \in \mathcal{M}^n_F$, tel que

$$\sup_{\mathbb{Q}\in\mathcal{M}^n_F} d_{FM}\left(\mathbb{Q}, \mathbb{Q}_{\sigma^*,b_0}\right) = d_{FM}\left(\mathbb{Q}^{n*}, \mathbb{Q}_{\sigma^*,b_0}\right).$$

En raisonnant de la même manière qu'au point (5) de la proposition IV.20, on voit qu'il existe $(\sigma^*_n, b^*_n) \in \overline{\Sigma_1} \times \mathcal{B}_{\varepsilon_n}$ tel que

$$\mathbb{Q}^{n*} = \mathbb{Q}^n_{\sigma^*_n,b^*_n}.$$

En raisonnant comme dans la proposition IV.23, on voit que

$$\mathbb{Q}^{n*} \xrightarrow[n\to+\infty]{} \mathbb{Q}_{\sigma^*,b_0}.$$

□

Un résultat général de convergence

Plaçons nous dans un cadre plus général et supposons donnés un ensemble compact $\mathcal{K}$ de Σ (pour la topologie de la convergence uniforme sur tout compact) avec $\sigma_0 \in \mathcal{K}$ et pour tout $\sigma \in \mathcal{K}$ et $b \in \overline{\mathcal{B}_\varepsilon}$, une suite $\left(\mathbb{Q}^n_{\sigma,\,b}\right)_n$ de $\mathcal{P}(\Omega)$ vérifiant les hypothèses suivantes :

Hypothèse IV.30.

1. Pour tout $n \in \mathbb{N}, \sigma \in \mathcal{K}, b \in \overline{\mathcal{B}_\varepsilon}, t \in [0,1]$, il existe un noyau de transition $\left(\Pi^n_{\sigma,\,b}(t,x,\,.\,)\right)_x$ de $\mathbb{R}$ dans $\mathbb{R}$ tel que

$$\begin{cases} (1) & \mathbb{Q}^n_{\sigma,\,b}(X_0 = 0) = 1, \\ (2) & \mathbb{Q}^n_{\sigma,\,b}\left(X_t = X_{\frac{k}{n}} + (nt-k)\left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}}\right], \frac{k}{n} \le t \le \frac{k+1}{n}\right) = 1, \\ (3) & \mathbb{Q}^n_{\sigma,\,b}\left(X_{\frac{k+1}{n}} \in \,.\, \Big| X_{\frac{k}{n}}, \dots, X_0\right) = \Pi^n_{\sigma,\,b}\left(\frac{k}{n}, X_{\frac{k}{n}},\,.\,\right) \end{cases} \tag{IV.31}$$

2. Si $(\varepsilon_n)_n$ est une suite de réels strictement positifs de limite nulle, alors pour toute suite $(\sigma_n)_n$ d'éléments de $\mathcal{K}$ convergeant vers $\sigma \in \mathcal{K}$ uniformément sur tout compact et toute suite $b_n \in \overline{\mathcal{B}_{\varepsilon_n}}$, la suite $\left(\mathbb{Q}^n_{\sigma_n,\,b_n}\right)_n$ converge étroitement vers $\mathbb{Q}_{\sigma,\,b}$.

3. Pour tout $(\sigma, b) \in \mathcal{K} \times \overline{\mathcal{B}_\varepsilon}$,

$$\mathrm{H}\left(\mathbb{Q}^n_{\sigma,\,b}\middle|\, \mathbb{Q}^n_{\sigma_0,\,b_0}\right) < +\infty.$$

   De plus, il existe une fonction $q : \mathbb{R}^2 \to \mathbb{R}^+$ continue et nulle sur la diagonale, telle qu'en posant

$$\mathrm{I}(\sigma|\sigma_0) = \mathbb{E}_\sigma\left[\int_0^1 q(\sigma^2(X_t,t), \sigma_0^2(t,X_t))dt\right],$$

   on ait, pour toute suite $(\varepsilon_n)_n$ de limite nulle et toute suite $b_n \in \overline{\mathcal{B}_{\varepsilon_n}}$,

$$\forall \sigma \in \mathcal{K}, \quad \frac{\mathrm{H}\left(\mathbb{Q}^n_{\sigma,\,b_n}\middle|\, \mathbb{Q}^n_{\sigma_0,\,b_0}\right)}{n} \xrightarrow[n\to+\infty]{} \mathrm{I}(\sigma|\sigma_0), \tag{IV.32}$$

   et pour toute suite $(\sigma_n)_n$ d'éléments de $\mathcal{K}$ convergeant vers $\sigma \in \mathcal{K}$ uniformément sur tout compact,

$$\liminf_{n\to+\infty} \frac{\mathrm{H}\left(\mathbb{Q}^n_{\sigma_n,\,b_n}\middle|\, \mathbb{Q}^n_{\sigma_0,\,b_0}\right)}{n} \ge \mathrm{I}(\sigma|\sigma_0). \tag{IV.33}$$

4. L'ensemble $\mathcal{D}^n_{\varepsilon,\mathcal{K}}$ est l'ensemble des probabilités $\mathbb{Q}$ sur Ω vérifiant

$$\begin{cases} (1) & \mathbb{Q}(X_0 = 0) = 1, \\ (2) & \mathbb{Q}\left(X_t = X_{\frac{k}{n}} + (nt-k)\left[X_{\frac{k+1}{n}} - X_{\frac{k}{n}}\right], \frac{k}{n} \le t \le \frac{k+1}{n}\right) = 1, \\ (3') & \exists (\sigma, b) \in \mathcal{K} \times \mathcal{B}_\varepsilon \text{ tels que } \mathbb{Q}\left(X_{\frac{p+1}{n}} \in \,.\, \Big| X_{\frac{p}{n}}\right) = \Pi^n_{\sigma,\,b}\left(\frac{p}{n}, X_{\frac{p}{n}},\,.\,\right) \end{cases} \tag{IV.34}$$

Nous poserons

$$\mathcal{F}^n_{\varepsilon,\mathcal{K}} = \mathcal{E}^n_\varepsilon \cap \mathcal{D}^n_{\varepsilon,\mathcal{K}}, \tag{IV.35}$$

avec $\mathcal{E}^n_\varepsilon$ défini, comme précédemment, par (IV.11).
Nous supposerons que pour tout n, il existe un compact Ω_n de Ω tel que pour tout ε, on ait pour toute $\mathbb{Q} \in \mathcal{D}^n_{\varepsilon,\mathcal{K}}$, $\mathbb{Q}(\Omega_n) = 1$. Nous supposerons, de plus, que $\mathcal{D}^n_{\varepsilon,\mathcal{K}}$ est un fermé d'intérieur non vide de $\mathcal{P}(\Omega_n)$.

5. Nous supposerons que la fonction $\mathrm{I}(\sigma|\sigma_0)$ atteint son minimum en un unique point σ^* de $\{\sigma \in \mathcal{K} : \int F(X_T)\, d\mathbb{Q}_{\sigma,b_0} = 1\}$.
6. Enfin, nous poserons $\varepsilon_n = \left| \int F\left(X_{\frac{[nT]}{n}}\right) d\mathbb{Q}^n_{\sigma^*,b_0} - 1\right| + \frac{1}{n}$ et nous supposerons que

$$\mathrm{H}\left(\mathcal{F}^n_{\varepsilon_n,\mathcal{K}} \middle| Q^n_{\sigma_0,b_0}\right) = \mathrm{H}\left(\overset{\circ}{\mathcal{F}}{}^n_{\varepsilon_n,\mathcal{K}} \middle| Q^n_{\sigma_0,b_0}\right), \tag{IV.36}$$

où $\overset{\circ}{\mathcal{F}}{}^n_{\varepsilon_n,\mathcal{K}}$ désigne l'intérieur de $\mathcal{F}^n_{\varepsilon_n,\mathcal{K}}$ dans $\mathcal{P}(\Omega_n)$.

Sous ces hypothèses, nous avons le résultat suivant

Théorème IV.37. *Il existe au moins une suite $(m_n)_n$ d'entiers, $m_n \xrightarrow[n\to+\infty]{} +\infty$ telle que la suite $\left(\mathbb{R}^n_{\varepsilon_n,m_n}\right)_n$ converge vers $\mathbb{Q}_{\sigma^*,b_0}$.*

Démonstration.
Notons $\mathcal{M}^n_F$ l'ensemble des minimisants de $\mathrm{H}\left(\,.\,\middle|\,\mathbb{Q}^n_{\sigma_0,b_0}\right)$ sur $\mathcal{F}^n_{\varepsilon_n,\mathcal{K}}$.
Grâce à l'hypothèse (IV.36), on a d'après le point 2. du théorème III.9,

$$d_{FM}\left(\mathbb{R}^n_{\varepsilon_n,m}, \overline{\mathrm{co}}\,\mathcal{M}^n_F\right) \xrightarrow[m\to+\infty]{} 0.$$

On voit, en raisonnant comme dans la preuve du théorème (IV.29), qu'il suffit de montrer que

$$\sup_{\mathbb{Q}\in\overline{\mathcal{M}^n_F}} d_{FM}\left(\mathbb{Q},\mathbb{Q}_{\sigma^*,b_0}\right) \xrightarrow[n\to+\infty]{} 0.$$

Soit $\mathbb{Q}^{n*} \in \mathcal{M}^n_F$ tel que $\sup\limits_{\mathbb{Q}\in\mathcal{M}^n_F} d_{FM}\left(\mathbb{Q},\mathbb{Q}_{\sigma^*,b_0}\right) = d_{FM}\left(\mathbb{Q}^{n*},\mathbb{Q}_{\sigma^*,b_0}\right)$; montrons que $\mathbb{Q}^{n*} \in \mathcal{F}^n_{\varepsilon_n,\mathcal{K}}$. On voit de la même manière qu'au point (4) de la proposition IV.20 que pour toute $\mathbb{Q} \in \mathcal{F}^n_{\varepsilon_n,\mathcal{K}}$, il existe $(\sigma, b) \in \mathcal{K} \times \overline{\mathcal{B}_{\varepsilon_n}}$ tel que

$$\begin{cases} \mathrm{H}\left(\mathbb{Q}\middle|\mathbb{Q}^n_{\sigma_0,b_0}\right) = \mathrm{H}\left(\mathbb{Q}\middle|\mathbb{Q}^n_{\sigma,b}\right) + \mathrm{H}\left(\mathbb{Q}^n_{\sigma,b}\middle|\mathbb{Q}^n_{\sigma_0,b_0}\right) \\ \text{et} \\ \mathbb{Q}^n_{\sigma,b} \in \mathcal{F}^n_{\varepsilon_n,\mathcal{K}} \end{cases}$$

et on en déduit, en particulier, qu'il existe $(\sigma^*_n, b^*_n) \in \mathcal{K} \times \overline{\mathcal{B}_{\varepsilon_n}}$ tel que $\mathbb{Q}^{n*} = \mathbb{Q}^n_{\sigma^*_n,b^*_n}$. En raisonnant comme dans la proposition IV.23, on voit que

$$\mathbb{Q}^{n*} \xrightarrow[n\to+\infty]{} \mathbb{Q}_{\sigma^*,b_0}.$$

$\square$

CHAPITRE V

Principes conditionnels de type Gibbs pour des mesures à poids aléatoires

Sommaire

V.1 Introduction

V.1.1 Méthodes d'analyse convexe pour des problèmes inverses mal posés

Le problème d'identifier un modèle régissant un certain phénomène sur la base d'observations partielles se pose dans de très nombreux domaines, comme la tomographie, l'astronomie, ou encore la finance. Nous nous concentrerons dans la suite sur le problème inverse suivant appelé *Problème des moments* :
Retrouver une mesure finie P sur un espace mesurable $(\mathcal{X}, \mathcal{B})$ satisfaisant

$$\int_{\mathcal{X}} F(x)\, dP(x) \in K \tag{V.1}$$

avec $F = (f_1, \ldots, f_p)$ une application mesurable à valeurs dans $\mathbb{R}^k$ et K un convexe de $\mathbb{R}^k$.

Dans de nombreuses situations, on dispose d'un modèle de référence R sur $(\mathcal{X}, \mathcal{B})$ qu'il s'agit de modifier pour qu'il satisfasse (V.1).
Afin de sélectionner un élément de

$$\mathcal{S}(F, K) := \left\{ P \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} F(x)\, dP(x) \in K \right\},$$

une méthode classique consiste à minimiser une fonction de coût $\mathrm{I}\,(\,.\,|R)$ convexe positive et nulle en R. L'une des méthodes les plus populaires est de minimiser l'entropie relative par rapport à R, *ie* de prendre $\mathrm{I}(P|R) = \mathrm{H}\,(P|\,R)$ (à condition que P et R soient des probabilités.). Dans les célèbres articles [18, 19], I. Csiszár a donné des résultats précis sur la forme algébrique du minimisant (la I-projection de R sur $\mathcal{S}(F, K)$) et dans [20], le même auteur a fourni une justification axiomatique de cette méthode.

Plus récemment, J.M. Borwein et A.S. Lewis ont étudié dans [7, 8], la minimisation sous contraintes de fonctionnelles $\mathrm{I}\,(\,.\,|R)$ ayant la forme suivante :

$$\mathrm{I}_\gamma\,(P\,|R) = \int_{\mathcal{X}} \gamma\left(\frac{dP_a}{dR}\right) dR + b_\psi P_s^+(\mathcal{X}) - a_\psi P_s^-(\mathcal{X})$$

où R est une probabilité sur $\mathcal{X}$, $\gamma : \mathbb{R} \to [0, +\infty]$ est une fonction convexe, P_a est la partie absolument continue de P par rapport à R, P_s sa partie singulière et $P_s = P_s^+ - P_s^-$ est la décomposition de Jordan de P_s (voir section V.2 pour la définition de a_ψ et b_ψ). Borwein et Lewis ont obtenu la représentation des minimisants de $\mathrm{I}_\gamma\,(\,.\,|R)$ sur des ensembles de la forme $\mathcal{S}(F, K)$ (voir [7, 8], [21] thm 2.2 et 2.4, et [43, 44] pour des extensions de ces résultats). L'intérêt des γ-divergences tient dans la possibilité d'imposer, par un bon choix de γ, des contraintes non-linéaires supplémentaires à la densité de la solution (voir [21] pour plus d'informations sur le sujet).

V.1.2 Une interprétation probabiliste de ces méthodes

La théorie des grandes déviations fournit une belle interprétation de la méthode de minimisation de l'entropie relative, via le théorème de Sanov et le principe conditionnel de Gibbs : si X_i est une suite i.i.d de loi R, alors pour de bons ensembles convexes C de $\mathcal{P}(\mathcal{X})$, la loi conditionnelle X_1 sachant que $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ appartient à C converge étroitement vers la I-projection R^* de R sur C. Autrement dit : *Si l'on force la mesure empirique de $(X_1, X_2, \dots, X_n)$ à appartenir à C, la loi de X_1 est modifiée de telle manière qu'elle converge vers la I-projection R^* de R sur C.*

Dans [36], F. Gamboa et E. Gassiat ont établi qu'une grande classe de γ-divergences vérifient des propriétés analogues : elles gouvernent les grandes déviations d'une suite de mesures aléatoires, et pour ce P.G.D, un principe conditionnel de type Gibbs est valable.

Avant d'exposer leurs résultats, introduisons quelques notations :

Pour toute mesure de probabilité ν sur $\mathbb{R}^q$, nous noterons Z_ν, Λ_ν et Λ_ν^* la transformée de Laplace, la Log-Laplace et la transformée de Cramér de ν, définies respectivement par :

$$\begin{aligned}
\forall s \in \mathbb{R}^q, &\quad Z_\nu(s) = \int \exp\langle s, x\rangle d\nu(x) \in \mathbb{R}^+ \cup \{+\infty\} \\
\forall s \in \mathbb{R}^q, &\quad \Lambda_\nu(s) = \log(Z_\nu)(s) \in \mathbb{R} \cup \{+\infty\} \\
\forall t \in \mathbb{R}^q, &\quad \Lambda_\nu^*(t) = \sup_{s\in\mathbb{R}^n}\{\langle s, t\rangle - \Lambda_\nu(s)\} \in \mathbb{R}^+ \cup \{+\infty\}
\end{aligned}$$

Rappelons que le domaine d'une fonction convexe $f : V \to \mathbb{R} \cup \{+\infty\}$, noté $\operatorname{dom} f$ est l'ensemble défini par :

$$\operatorname{dom} f = \{x \in V : f(x) < +\infty\}.$$

Théorème V.2. (Gamboa, Gassiat, [36] thm 3.4)
Soient $\mathcal{X}$ un espace métrique compact, R une probabilité sur $\mathcal{X}$ dont le support est l'espace X tout entier et $(x_i^n)_{\substack{1\le i\le n\\ n\in\mathbb{N}^*}}$ une famille de points de $\mathcal{X}$ telle que $\displaystyle\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$ converge étroitement vers R. Soit μ une mesure de probabilité sur $\mathbb{R}$ telle que $\operatorname{dom} Z_\mu =]-\alpha, \beta[$, *avec $\alpha, \beta > 0$.*
Si $(Z_i)_i$ est une suite i.i.d de loi μ, alors la suite $(L_n)_n$ de mesures à poids aléatoires définie par

$$L_n = \frac{1}{n}\sum_{i=1}^n Z_i \delta_{x_i^n}$$

satisfait un principe de grandes déviations sur $\mathcal{M}(\mathcal{X})$, muni de la topologie de la convergence étroite, de bonne fonction de taux :

$$\mathrm{I}_\mu(P\,|R) = \int_{\mathcal{X}} \Lambda_\mu^*\left(\frac{dP_a}{dR}\right) dR + \alpha P_s^-(\mathcal{X}) + \beta P_s^+(\mathcal{X}).$$

(voir également [26] thm 7.2.3, [32] et [50] pour un résultat plus général.)

De plus, en supposant que $\mathbb{P}(L_n \in \mathcal{S}(F,K)) > 0$ pour tout n assez grand et en posant

$$R_n = \mathbb{E}[L_n | L_n \in \mathcal{S}(F,K)] := \frac{\mathbb{E}[L_n \mathbb{1}_{\mathcal{S}(F,K)}(L_n)]}{\mathbb{P}(L_n \in \mathcal{S}(F,K))},$$

ils ont montré, sous certaines hypothèses sur lesquelles nous reviendrons plus tard, que R_n convergeait vers R^*, l'unique minimisant de $\mathrm{I}_\mu(\,.\,|R)$ sur $\mathcal{S}(F,K)$ (voir [36] et le théorème V.15 pour une formulation plus précise).

Remarque V.3.

Ce principe conditionnel de type Gibbs donne un sens bayesien à la minimisation de γ-divergences :
R est un modèle a priori, ne satisfaisant pas la contrainte $\int_{\mathcal{X}} F\,dR \in K$. On va modifier R de la manière suivante : on commence par discrétiser R en se donnant une famille $(x_i^n)_{\substack{1\leq i\leq n\\ n\in\mathbb{N}^*}}$ de points de $\mathcal{X}$ telle que $\widetilde{L}_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$ converge étroitement vers R (x_i^n est par exemple une suite de réalisations indépendantes de R), puis on repondère $\widetilde{L}_n$ de manière aléatoire : $L_n = \frac{1}{n}\sum_{i=1}^n Z_i\delta_{x_i^n}$.
La suite

$$R_n = \mathbb{E}[L_n | L_n \in \mathcal{S}(F,K)],$$

qui est la moyenne de toutes les réalisations de L_n satisfaisant $\int_{\mathcal{X}} F\,dL_n \in K$, converge alors vers le minimisant de la γ-divergence $\mathrm{I}_\mu(\,.\,|R)$ sur $\mathcal{S}(F,K)$.

V.1.3 Le problème des contraintes fines

Comme pour le principe conditionnel de Gibbs, se pose le problème de donner un sens à

$$R_n = \mathbb{E}[L_n | L_n \in \mathcal{S}(F,K)],$$

lorsque $\mathbb{P}(L_n \in \mathcal{S}(F,K)) = 0$ et quand on ne dispose pas d'une désintégration explicite. Pour autoriser ce genre de conditionnement, nous allons reprendre la même idée que celle développée dans le chapitre III, à savoir : relaxer la contrainte en prenant un ε_n-voisinage de K, avec une suite $(\varepsilon_n)_n$ convergeant suffisamment lentement vers 0 pour garantir que $\mathbb{P}(L_n \in \mathcal{S}(F,K^{\varepsilon_n})) > 0$, pour tout $n \in \mathbb{N}^*$. Nous prouverons dans le théorème V.16 que, sous certaines hypothèses,

$$\lim_{n\to+\infty} \mathbb{E}[L_n | L_n \in \mathcal{S}(F,K^{\varepsilon_n})] = R^*, \tag{V.4}$$

avec $\varepsilon_n \gg \frac{1}{\sqrt{n}}$.

La preuve de ce résultat est, dans ses grandes lignes, analogue à celle du théorème III.61 du chapitre précédent. La principale nouveauté est la proposition V.26 qui va jouer le rôle du théorème III.37 de Csiszár. La preuve de cette proposition s'inspire des travaux de Bobkov et Götze sur l'inégalité de transport $\mathbb{T}_1$. Nous reviendrons en détails sur ce sujet dans la seconde partie de cette thèse consacrée aux inégalités de transport.

V.2 Minimisation sous contraintes des γ-divergences et procédé M.E.M

Cette section est consacrée à la minimisation sous contraintes des γ-divergences. Nous présenterons des résultats de Borwein et Lewis (théorème V.6) et l'approche de la Minimisation de l'Entropie en Moyenne (M.E.M.) (théorème V.8) de Gamboa et Gassiat.

Nous ferons les hypothèses suivantes :

Hypothèse V.5.

1. $\mathcal{X}$ est un espace métrique compact ; l'ensemble $\mathcal{M}(\mathcal{X})$ des mesures de Borel finies sur $\mathcal{X}$ est muni de la topologie de la convergence étroite, *ie* la moins fine rendant continues les applications $P \mapsto \langle P, f\rangle$, f continue sur $\mathcal{X}$,
2. R est une mesure de probabilité sur $\mathcal{X}$ dont le support est l'espace $\mathcal{X}$ tout entier,
3. $F = (f_1, \ldots, f_k) : \mathcal{X} \to \mathbb{R}^k$ est une application continue sur $\mathcal{X}$ ayant des composantes linéairement indépendantes,
4. K est un convexe compact de $\mathbb{R}^k$.

Rappelons que

$$\mathcal{S}(F, K) = \left\{ P \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} F\, dP \in K \right\}$$

Théorème V.6. (Borwein-Lewis, [8])
Soit $\gamma : \mathbb{R} \to [0, +\infty]$ *une fonction convexe s.c.i et notons* $a_\gamma < b_\gamma$ *les extrémités de* $\operatorname{dom}\gamma$*. On suppose que* γ *est derivable, strictement convexe sur* $\overset{\circ}{\operatorname{dom}}\gamma$ *et s'annule en un point de* $\overset{\circ}{\operatorname{dom}}\gamma$*. Soit* ψ *la conjuguée convexe de* γ*, ie*

$$\psi(s) = \gamma^*(s) = \sup_{t \in \mathbb{R}} \{st - \gamma(t)\}.$$

Notons $a_\psi < 0 < b_\psi$ *les extrémités de* $\operatorname{dom}\psi$*.*

Supposons qu'il existe $S \in \mathcal{S}(F,K)$ telle que $S \ll R$ et $\frac{dS}{dR} \in]a_\gamma, b_\gamma[\quad R\,ps.$

Sous ces hypothèses, la fonctionnelle $\mathrm{I}_\gamma\,(\,.\,|R)$, définie sur $\mathcal{M}(\mathcal{X})$ par

$$\mathrm{I}_\gamma\,(P\,|R) = \int_{\mathcal{X}} \gamma\left(\frac{dP_a}{dR}\right) dR + b_\psi P_s^+(\mathcal{X}) - a_\psi P_s^-(\mathcal{X})$$

atteint son minimum sur $\mathcal{S}(F,K)$.
De plus, tout minimisant R^* de $\mathrm{I}_\gamma\,(\,.\,|R)$ sur $\mathcal{S}(F,K)$ est de la forme :

$$R^* = g^* R + \sigma,$$

où

- $g^*(x) = \psi'\,\langle v^*, F(x)\rangle$,
- *v^* est l'unique minimisant de la fonction*

$$H(v) = \int_{\mathcal{X}} \psi\,\langle v, F(x)\rangle\, dR(x) - \inf_{y\in K}\langle v, y\rangle,$$

- *σ est singulière par rapport à R.*

De plus, si v^* appartient à l'intérieur de $\left\{v : \int_{\mathcal{X}} \psi\,\langle v, F(x)\rangle\, dR(x) < +\infty\right\}$, alors l'unique minimisant de $\mathrm{I}_\gamma\,(\,.\,|R)$ sur $\mathcal{S}(F,K)$ est $R^* = g^* R$. C'est en particulier le cas lorsque $\mathrm{dom}\,\psi = \mathbb{R}$.

(Pour une preuve, voir [8], ou l'appendice A de [21] ; voir CL pour une extension).

Le théorème suivant présente le procédé de Minimisation de l'Entropie sur la Moyenne (M.E.M) développé dans [22, 34, 35, 36] par D. Dacunha-Castelle, F. Gamboa et E. Gassiat, qui donne un autre point de vue sur la minimisation des γ-divergences. Nous ferons les hypothèses suivantes :

Hypothèse V.7.

1. μ est probabilité sur $\mathbb{R}$ telle que $\mathrm{dom}\,\Lambda_\mu =]-\alpha, \beta[$, avec $\alpha, \beta \in \mathbb{R}_+^* \cup \{+\infty\}$,
2. $(x_i^n)_{\substack{1\le i\le n\\ n\in\mathbb{N}^*}} \subset \mathcal{X}$ est une famille de points de $\mathcal{X}$ telle que $\dfrac{1}{n}\displaystyle\sum_{i=1}^n \delta_{x_i^n}$ converge étroitement vers R,
3. Il existe $g_0 : \mathcal{X} \to]a_\mu, b_\mu[$ continue, telle que $g_0 R \in \mathcal{S}(F,K)$, où $a_\mu < b_\mu$ sont les extrémités de l'enveloppe convexe fermée du support de μ,
4. La fonction H, définie sur $\mathbb{R}^k$ par

$$H(v) = \int_{\mathcal{X}} \Lambda_\mu\,\langle v, F(x)\rangle\, dR(x) - \inf_{y\in K}\langle v, y\rangle,$$

atteint son minimum en un unique point v^* appartenant à l'intérieur de son domaine.

Nous regroupons dans le théorème suivant différents résultats prouvés dans [35] et [36], avec un petit raffinement aux points 4 et 5 :

Théorème V.8. (Gamboa-Gassiat [36], thm. 2.1)
Pour tout $n \in \mathbb{N}^*$, soit $L_n : \mathbb{R}^n \to \mathcal{M}(\mathcal{X})$ définie par $L_n(z) = \dfrac{1}{n}\sum_{i=1}^{n} z_i \delta_{x_i^n}$.
Pour tout $\varepsilon \geq 0$, soit $K^\varepsilon = \{x \in \mathbb{R}^k : \exists y \in K, d_\infty(x,y) \leq \varepsilon\}$, où $d_\infty(x,y) = \max(|x_i - y_i|, i = 1 \dots k)$. Pour tout $n \geq 1$ et $\varepsilon \geq 0$, soit

$$\Pi_n(K^\varepsilon) = \{\nu \in \mathcal{P}(\mathbb{R}^n) : \mathbb{E}_\nu[\langle L_n, F\rangle] \in K^\varepsilon\}$$

Alors, sous les hypothèses (V.5) et (V.7), on a :

1. *Il existe $n_0 \geq 1$ tel que pour tout $\varepsilon \geq 0$, $\mu^{\otimes n}$ admet une I-projection $\mu^*_{n,\varepsilon}$ sur $\Pi_n(K^\varepsilon)$.*

2. *Pour $n \geq n_0$, $\mu^*_{n,\varepsilon}$ a l'expression suivante :*

$$\mu^*_{n,\varepsilon} = \frac{\exp\left\langle w^*_{n,\varepsilon}, \,.\,\right\rangle}{Z_{\mu^{\otimes n}}(w^*_{n,\varepsilon})}\mu^{\otimes n} \qquad \textit{avec} \qquad w^*_{n,\varepsilon} = \begin{bmatrix} \left\langle v^*_{n,\varepsilon}, F(x_1^n)\right\rangle \\ \vdots \\ \left\langle v^*_{n,\varepsilon}, F(x_n^n)\right\rangle \end{bmatrix}$$

   et $v^*_{n,\varepsilon}$ est un minimisant de la fonction $H_{n,\varepsilon}$ définie sur $\mathbb{R}^k$ par

$$H_{n,\varepsilon}(v) = \frac{1}{n}\sum_{i=1}^{n} \Lambda_\mu \langle v, F(x_i^n)\rangle - \inf_{y \in K^\varepsilon} \langle v, y\rangle.$$

3. *Pour tout $n \geq n_0$, on a :*

$$R^*_{n,\varepsilon} := \mathbb{E}_{\mu^*_{n,\varepsilon}}[L_n] = \frac{1}{n}\sum_{i=1}^{n} \Lambda'_\mu \left\langle v^*_{n,\varepsilon}, F(x_i^n)\right\rangle \delta_{x_i^n}.$$

4. *Pour toute suite $\varepsilon_n \in \mathbb{R}^+$ convergeant vers 0, v^*_{n,ε_n} converge vers v^* (l'unique minimisant de H)*

5. *Pour toute suite $\varepsilon_n \in \mathbb{R}^+$ convergeant vers 0, la suite R^*_{n,ε_n} converge étroitement vers R^*, l'unique minimisant de $\mathrm{I}_\mu\,(\,.\,|R)$ sur $\mathcal{S}(F,K)$, qui s'écrit*

$$R^* = \Lambda'_\mu \langle v^*, F(\,.\,)\rangle R.$$

(On trouvera une preuve de ce théorème dans l'annexe B.)

Remarque V.9.

- On notera plus simplement μ_n^*, R_n^*, v_n^*, *etc* à la place de $\mu_{n,0}^*$, $R_{n,0}^*$, $v_{n,0}^*$, *etc*
- Les $R_{n,\varepsilon}^*$ seront appelés les *estimateurs M.E.M.*.
- Si $\operatorname{dom}\Lambda_\mu = \mathbb{R}$, l'hypothèse (4) de (V.7) est automatiquement satisfaite .
- Si l'hypothèse (4) de (V.7) n'est pas satisfaite, les estimateurs M.E.M. ne convergent pas en général, (voir [36] thm 2.1 pour des résultats sur les points d'accumulation).

La proposition suivante permet de mieux comprendre les estimateurs M.E.M :

Proposition V.10. *On suppose que* $\operatorname{dom}\Lambda_\mu = \mathbb{R}$*, et on pose* $\overline{R}_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$*. Soit* $\mathcal{S}$ *un ensemble convexe de* $\mathcal{P}(\mathcal{X})$*. Les deux propositions suivantes sont équivalentes :*

1. *La fonction* $\mathrm{I}_\mu(\,.\,|\overline{R}_n)$*, définie sur* $\mathcal{M}(\mathcal{X})$ *par*
$$\mathrm{I}_\mu(P|\overline{R}_n) := \int_{\mathcal{X}} \Lambda_\mu^*\left(\frac{dP}{d\overline{R}_n}\right) d\overline{R}_n,$$
atteint son minimum sur $\mathcal{S}$ *en un point* R_n^**.*
2. *La mesure de probabilité* $\mu^{\otimes n}$ *admet une I-projection* μ_n^* *sur le convexe*
$$\Pi_n = \{\nu \in \mathcal{P}(\mathbb{R}^n) : \mathbb{E}_\nu[L_n] \in \mathcal{S}\}\,.$$

Dans ces conditions, R_n^* *est unique et on a la relation :*
$$R_n^* = \mathbb{E}_{\mu_n^*}[L_n].$$

Remarque V.11.

En revenant aux notations et aux hypothèses du théorème V.8 et en supposant en plus que $\operatorname{dom}\Lambda_\mu = \mathbb{R}$, on en déduit en particulier que pour tout $\varepsilon > 0$, la mesure $R_{n,\varepsilon}^*$ est l'unique minimisant de la fonction
$$\mathrm{I}_\mu(P|\overline{R}_n) := \int_{\mathcal{X}} \Lambda_\mu^*\left(\frac{dP}{d\overline{R}_n}\right) d\overline{R}_n,$$
sous la contrainte $P \in \mathcal{S}(F, K^\varepsilon)$.

Démonstration. Remarquons, tout d'abord, que
$$\mathrm{I}_\mu(P|\overline{R}_n) < +\infty \Rightarrow \left(\exists z \in \mathbb{R}^n, \quad P = \frac{1}{n}\sum_{i=1}^n z_i \delta_{x_i^n} := L_n(z)\right).$$

De plus, pour tout $z \in \mathbb{R}^n$, on a
$$I_\mu\left(L_n(z)\,\middle|\overline{R}_n\right) = \frac{1}{n}\sum_{i=1}^n \Lambda_\mu^*(z_i) = \frac{1}{n}\Lambda_{\mu^{\otimes n}}^*(z). \tag{V.12}$$

Comme $\operatorname{dom}\Lambda_\mu = \mathbb{R}$, on a l'identité classique suivante

$$\Lambda^*_{\mu^{\otimes n}}(z) = \inf\left\{\mathrm{H}\left(\nu \middle| \mu^{\otimes n}\right) : \nu \in \mathcal{P}(\mathbb{R}^n) \quad \text{telle que} \quad \int x\, d\nu = z\right\}, \tag{V.13}$$

et pour tout $z \in \operatorname{dom}\Lambda^*_{\mu^{\otimes n}}$, l'inf est atteint.
(Voir, par exemple, le théorème 5.2 de [28] ; on peut aussi appliquer la version II.21 du théorème de Sanov pour une suite i.i.d de loi $\mu^{\otimes n}$, avec G contenant la fonction identité de $\mathbb{R}^n$, et conclure grâce au principe de contraction et au corollaire II.36.)
Ainsi, pour tout $z \in \operatorname{dom}\Lambda^*_{\mu^{\otimes n}}$, il existe un unique $\nu_z \in \mathcal{P}(\mathbb{R}^n)$ tel que

$$\begin{cases} \mathrm{I}_\mu\left(L_n(z) \,\middle|\overline{R}_n\right) = \frac{1}{n}\mathrm{H}\left(\nu_z \middle| \mu^{\otimes n}\right) \\ \text{et} \\ \int_{\mathbb{R}^n} y\, d\nu_z(y) = z \end{cases}$$

Clairement, si $z \in \mathcal{S}$, alors $\nu_z \in \Pi_n$. On en déduit, en particulier, que

$$\inf\{\mathrm{I}_\mu(P|\overline{R}_n) : P \in \mathcal{S}\} \geq \frac{1}{n}\mathrm{H}\left(\Pi_n \middle| \mu^{\otimes n}\right). \tag{V.14}$$

Montrons que 2. implique 1. :
Soit μ_n^* la I-projection de $\mu^{\otimes n}$ sur Π_n ; d'après (V.12) et (V.13), on a

$$\mathrm{I}_\mu\left(\mathbb{E}_{\mu_n^*}[L_n] \,\middle|\overline{R}_n\right) = \frac{1}{n}\Lambda^*_{\mu^{\otimes n}}\left(\int_{\mathbb{R}^n} y\, d\mu_n^*(y)\right) \leq \frac{1}{n}\mathrm{H}\left(\mu_n^* \middle| \mu^{\otimes n}\right) = \frac{1}{n}\mathrm{H}\left(\Pi_n \middle| \mu^{\otimes n}\right).$$

D'après (V.14), on en déduit que $\mathrm{I}_\mu(\,.\,|\overline{R}_n)$ atteint son minimum sur $\mathcal{S}$ au point $R_n^* = \mathbb{E}_{\mu_n^*}[L_n]$.

Montrons que 1. implique 2. :
Soit $z^* \in \mathbb{R}^n$ tel que

$$\inf\{\mathrm{I}_\mu(P|\overline{R}_n) : P \in \mathcal{S}\} = \mathrm{I}_\mu(L_n(z^*)|\overline{R}_n) = \frac{1}{n}\mathrm{H}\left(\nu_{z^*} \middle| \mu^{\otimes n}\right).$$

Si $\nu \in \Pi_n$, alors

$$\frac{1}{n}\mathrm{H}\left(\nu_{z^*} \middle| \mu^{\otimes n}\right) \leq \mathrm{I}_\mu(\mathbb{E}_\nu[L_n]|\overline{R}_n) \leq \frac{1}{n}\mathrm{H}\left(\nu \middle| \mu^{\otimes n}\right).$$

La probabilité ν_{z^*} est donc la I-projection de $\mu^{\otimes n}$ sur Π_n et on a $L_n(z^*) = \mathbb{E}_{\mu_n^*}[L_n]$. $\square$

V.3 Résultats principaux

Le résultat que nous voulons étendre est le suivant :

Théorème V.15. (Gamboa-Gassiat, [36] thm 2.3)
Sous les hypothèses (V.5) et (V.7), si K est d'intérieur non vide alors l'estimateur bayesien

$$R_n := \frac{\mathbb{E}_{\mu^{\otimes n}}[L_n \mathbb{1}_{\mathcal{S}(F,K)}(L_n)]}{\mu^{\otimes n}(L_n \in \mathcal{S}(F,K))}$$

est bien défini pour tout n suffisamment grand et converge étroitement vers R^*, l'unique minimisant de $\mathrm{I}_\mu\,(\,.\,|R)$ sur $\mathcal{S}(F,K)$.

Notre résultat principal est le suivant :

Théorème V.16. *Sous les hypothèses (V.5) et (V.7), si ε_n est suite de réels strictement positifs convergeant vers 0 et telle que $\lim\limits_{n\to+\infty} n\varepsilon_n^2 = +\infty$, alors l'estimateur bayesien*

$$R_{n,\,\varepsilon_n} := \frac{\mathbb{E}_{\mu^{\otimes n}}[L_n \mathbb{1}_{\mathcal{S}(F,K^{\varepsilon_n})}(L_n)]}{\mu^{\otimes n}(L_n \in \mathcal{S}(F,K^{\varepsilon_n}))}$$

est bien défini pour tout n assez grand et converge étroitement vers R^*, l'unique minimisant de $\mathrm{I}_\mu\,(\,.\,|R)$ sur $\mathcal{S}(F,K)$.

Introduisons des notations supplémentaires :

- Pour tout $u \in \operatorname{dom} Z_\mu$, μ_u est la mesure de probabilité sur $\mathbb{R}$ définie par :

$$\frac{d\mu_u}{d\mu}(x) = \frac{\exp ux}{Z_\mu(u)},$$

  et pour tout $n \geq 2$ et tout $u \in \operatorname{dom} Z_\mu^n$,

$$\mu_u^{\otimes n} = \mu_{u_1} \otimes \cdots \otimes \mu_{u_n}$$

- $\mathcal{Q}$ désigne l'ensemble des fonctions continues, concaves, croissantes, nulles en 0 et non bornées définies sur $\mathbb{R}^+$.

La preuve du théorème V.16 repose sur la proposition suivante dont la démonstration est très proche de celle du théorème de Bobkov et Götze sur l'inégalité de transport $\mathbb{T}_1$ (voir [4] thm 3.1) :

Proposition V.17. *Pour tout segment $J \subset]-\alpha,\beta[$, il existe une fonction $Q_J \in \mathcal{Q}$ telle que, pour tout $u \in J$ et $\nu \in \mathcal{P}(\mathbb{R})$:*

$$\left| \int_{\mathbb{R}} x\, d\nu(x) - \int_{\mathbb{R}} x\, d\mu_u(x) \right| \leq Q_J(\mathrm{H}\,(\nu|\,\mu_u)).$$

Remarque V.18.

Si μ est telle que, pour tout $s \in \mathbb{R}$, $\Lambda''_\mu(s) \leq M$ (par exemple, si μ a un support compact ou si μ est une mesure gaussienne), on peut prendre $Q_J(x) = \sqrt{2Mx}$. Dans ce cas, l'inégalité précédente n'est qu'un cas particulier de l'inégalité de transport $\mathbb{T}_1$ (voir [4], théorème 3.1). D'autres exemples explicites seront donnés dans la section V.4.2. Nous reviendrons plus en détail sur ce type d'inégalités dans la seconde partie de cette thèse.

En utilisant notamment les inégalités de Csiszár (II.4) et (II.26), nous déduirons de ce résultat une majoration de la norme en variation entre R_{n,ε_n} et R^*_{n,ε_n} de la forme suivante :

$$\left\|R_{n,\varepsilon_n} - R^*_{n,\varepsilon_n}\right\|_{VT} \leq Q\left(\frac{-1}{n}\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon_n})e^{\mathrm{H}\left(\mu^*_{n,\varepsilon_n}|\mu^{\otimes n}\right)}\right]\right) \tag{V.19}$$

où $Q \in \mathcal{Q}$ ne dépend pas de n (voir proposition V.26). Cette inégalité est l'analogue du théorème III.37 de Csiszár. Comme, d'après le théorème V.8, R^*_{n,ε_n} converge vers R^*, il suffira de montrer que le membre de droite de (V.19) tend vers 0, pour montrer que R_{n,ε_n} converge également vers R^*. Le contrôle du membre de droite de (V.19) se fera par des moyens analogues à ceux mis en oeuvre dans la preuve du théorème III.61 du chapitre III : une borne inférieure exacte de déviation (lemme V.27) et une inégalité de type Bernstein (lemme V.25).

V.4 Inégalités de type transport

V.4.1 Résultats généraux

Nous aurons besoin du lemme suivant :

Lemme V.20. *Si* $k : [0, r[\to \mathbb{R}_+$, $r \in \mathbb{R}_+^* \cup \{+\infty\}$ *est telle que* $\lim_{t\to 0} k(t) = 0$ *et* $\lim_{t\to r} k(t) = +\infty$*, alors la fonction Q définie par*

$$\forall a \in \mathbb{R}_+, \quad Q(a) = \inf_{t\in]0,r[}\left\{\frac{a}{t} + k(t)\right\}$$

appartient à $\mathcal{Q}$*.*

Démonstration.

- Pour tout $a \geq 0$, $t \mapsto \dfrac{a}{t} + k(t)$ est une fonction positive donc $Q(a) = \inf_{0<t<r}\left\{\dfrac{a}{t} + k(t)\right\} \in \mathbb{R}_+$ et Q est bien définie sur $\mathbb{R}_+$. De plus, $Q(0) = \inf_{0<t<r}\{k(t)\}$; or $\lim_{t\to 0} k(t) = 0$, donc $Q(0) = 0$.

- Q étant un infimum de fonctions affines, elle est concave.

- Si $0 \leq a \leq a' < r$, alors, pour tout $0 < t < r$, on a

$$\frac{a}{t} + k(t) \leq \frac{a'}{t} + k(t).$$

En passant à l'infimum, on obtient $Q(a) \leq Q(a')$ et on en déduit que Q est croissante.

- Soit $(a_n)_n$ telle que $a_n \xrightarrow[n\to+\infty]{} 0$; pour tout $0 < t < r$, on a $Q(a_n) \leq \frac{a_n}{t} + k(t)$ et donc $\limsup\limits_{n\to+\infty} Q(a_n) \leq k(t)$. Comme $\inf\limits_{0<t<r} k(t) = 0$, il s'ensuit que $\limsup\limits_{n\to+\infty} Q(a_n) = 0$ et Q est donc continue en 0.

- Enfin soit $(a_n)_n$ telle que $a_n \xrightarrow[n\to+\infty]{} +\infty$; montrons que $Q(a_n) \xrightarrow[n\to+\infty]{} +\infty$. Q étant croissante, il suffit de prouver que $Q(a_n)$ n'est pas bornée. Pour tout n, $t \mapsto \frac{a_n}{t} + k(t)$ est une fonction admettant $+\infty$ comme limite en 0 et en r, il existe donc t_n tel que

$$Q(a_n) = \frac{a_n}{t_n} + k(t_n).$$

Par conséquent,

$$\limsup_{n\to+\infty} Q(a_n) \geq \limsup_{n\to+\infty} \frac{a_n}{t_n} \vee \limsup_{n\to+\infty} k(t_n).$$

Si $(t_n)_n$ est bornée,

$$\limsup_{n\to+\infty} \frac{a_n}{t_n} = +\infty$$

et si $(t_n)_n$ ne l'est pas ($r = +\infty$),

$$\limsup_{n\to+\infty} k(t_n) = +\infty.$$

Dans tous les cas, $Q(a_n)$ n'est pas bornée.
Ainsi Q est un élément de $\mathcal{Q}$. $\square$

Démonstration de la proposition V.17 : Pour tout $u \in]-\alpha, \beta[$,

$$Z_{\mu_u}(t) = \int_{\mathbb{R}} \exp(tx) \frac{\exp(ux)}{Z_\mu(u)}\, d\mu(x) = \frac{Z_\mu(u+t)}{Z_\mu(u)}$$

donc $\operatorname{dom} Z_{\mu_u} =]-\alpha - u, \beta - u[$.

Soit $t \in]-\alpha - u, \beta - u[$,

$$t\left(\int_{\mathbb{R}} x\, d\nu(x) - \int_{\mathbb{R}} x\, d\mu_u(x)\right) = \int_{\mathbb{R}} g_t(x)\, d\nu(x) + \log\left[\int_{\mathbb{R}} e^{t\left(x - \int_{\mathbb{R}} y\, d\mu_u(y)\right)}\, d\mu_u(x)\right],$$

en posant

$$g_t(x) = t\left(x - \int_{\mathbb{R}} y\, d\mu_u(y)\right) - \log\left[\int_{\mathbb{R}} e^{t\left(x - \int_{\mathbb{R}} y\, d\mu_u(y)\right)}\, d\mu_u(x)\right].$$

Clairement,

$$\int_{\mathbb{R}} \exp\, g_t\, d\mu_u = 1.$$

Or, d'après la formulation variationnelle de l'entropie relative, on a

$$\mathrm{H}\left(\nu |\, \mu_u\right) = \sup\left\{\int_{\mathbb{R}} g\, d\nu : \int_{\mathbb{R}} \exp\, g\, d\mu_u \leq 1\right\}.$$

Par conséquent,

$$\int_{\mathbb{R}} g_t\, d\nu \leq \mathrm{H}\left(\nu |\, \mu_u\right).$$

De plus, en remarquant que $\Lambda'_\mu(u) = \int_{\mathbb{R}} y\, d\mu_u(y)$, on voit facilement que

$$\log\left[\int_{\mathbb{R}} e^{t\left(x - \int_{\mathbb{R}} y\, d\mu_u(y)\right)}\, d\mu_u(x)\right] = \Lambda_\mu(t+u) - \Lambda_\mu(u) - t\Lambda'_\mu(u) := q(t,u).$$

Ainsi, pour tout $t \in]0, \beta - u[$,

$$\int_{\mathbb{R}} x\, d\nu(x) - \int_{\mathbb{R}} x\, d\mu_u(x) \leq \frac{\mathrm{H}\left(\nu |\, \mu_u\right)}{t} + \frac{q(t,u)}{t}$$

et, pour $t \in]0, \alpha + u[$,

$$\int_{\mathbb{R}} x\, d\mu_u(x) - \int_{\mathbb{R}} x\, d\nu(x) \leq \frac{\mathrm{H}\left(\nu |\, \mu_u\right)}{t} + \frac{q(-t,u)}{t}.$$

La fonction Λ_μ étant convexe, q est positive.
Si $J = [a,b]$, posons $r = \min(\alpha + a, \beta - b) \in \mathbb{R}_+^* \cup \{+\infty\}$; alors, pour tout $0 < t < r$, on peut écrire

$$\left|\int_{\mathbb{R}} x\, d\nu(x) - \int_{\mathbb{R}} x\, d\mu_u(x)\right| \leq \frac{\mathrm{H}\left(\nu |\, \mu_u\right)}{t} + \frac{q(t,u) + q(-t,u) + t^2}{t}.$$

Posons

$$k(t) = \frac{\max_{u \in J}(q(t,u) + q(-t,u)) + t^2}{t}.$$

Alors, pour tout $u \in J$,

$$\left|\int_{\mathbb{R}} x\, d\nu(x) - \int_{\mathbb{R}} x\, d\mu_u(x)\right| \leq \frac{\mathrm{H}\left(\nu |\, \mu_u\right)}{t} + k(t).$$

En passant à l'infimum pour $0 < t < r$, on obtient

$$\left| \int_{\mathbb{R}} x \, d\nu(x) - \int_{\mathbb{R}} x \, d\mu_u(x) \right| \leq Q_J(\mathrm{H}\,(\nu | \, \mu_u)),$$

avec

$$Q_J(a) = \inf_{0<t<r} \left\{ \frac{a}{t} + k(t) \right\}.$$

Montrons que k vérifie les hypothèses du lemme 4.1 :

- Si $r = +\infty$, $k(t) \geq t$ donc $\lim\limits_{t \to +\infty} k(t) = +\infty$.
- Si $r = \alpha + a < +\infty$, alors

  $$k(t) \geq \frac{q(-t, a)}{t} = \frac{\Lambda_\mu(a - t) - \Lambda_\mu(a)}{t} + \Lambda'_\mu(a).$$

  Comme $\lim\limits_{t \to \alpha + a} \Lambda_\mu(a - t) = +\infty$, on a $\lim\limits_{t \to \alpha + a} k(t) = +\infty$.
- Si $r = \beta - b < +\infty$, on voit de même que $\lim\limits_{t \to \beta - b} k(t) = +\infty$.

Donc, dans tous les cas, $\lim\limits_{t \to r} k(t) = +\infty$.

Montrons que $\lim\limits_{t \to 0} k(t) = 0$.
Soit $0 < t_n < r$ telle que $t_n \xrightarrow[n \to +\infty]{} 0$; pour tout n, il existe $u_n \in J$ tel que

$$k(t_n) = \frac{q(t_n, u_n) + q(-t_n, u_n)}{t_n} + t_n$$

Supposons que pour tout n, $k(t_n) \geq \varepsilon > 0$. La suite $(u_n)_n$ étant bornée, il existe ϕ tel que $u_{\phi(n)} \xrightarrow[n \to +\infty]{} u_0 \in K$. Or, d'après la formule de Taylor-Lagrange, Λ''_μ étant positive, on a

$$q(t_{\phi(n)}, u_{\phi(n)}) + q(-t_{\phi(n)}, u_{\phi(n)}) \leq t^2_{\phi(n)} \sup \left\{ \Lambda''_\mu(u), u \in [u_{\phi(n)} - t_{\phi(n)}, u_{\phi(n)} + t_{\phi(n)}] \right\}$$

Donc $k(t_{\phi(n)}) \xrightarrow[n \to +\infty]{} 0$. Contradiction, donc $\lim\limits_{t \to 0} k(t) = 0$ et $Q_J \in \mathcal{Q}$. $\square$

Corollaire V.21. *Pour tout segment J inclus dans $]-\alpha, \beta[$ et tout $u \in J^n$, on a pour toute $\nu \in \mathcal{P}(\mathbb{R}^n)$*

$$\frac{1}{n} \left\| \int_{\mathbb{R}^n} x \, d\nu(x) - \int_{\mathbb{R}^n} x \, d\mu_u^{\otimes n}(x) \right\|_1 \leq Q_J \left(\frac{\mathrm{H}\,(\nu | \, \mu_u^{\otimes n})}{n} \right),$$

en posant

$$\mu_u^{\otimes n} = \mu_{u_1} \otimes \cdots \otimes \mu_{u_n}$$

et

$$\|x\|_1 = \sum_{i=1}^n |x_i|.$$

Démonstration. Nous noterons ν_i, la $i^{\text{ème}}$ marginale de ν.
On a

$$\begin{aligned}\frac{1}{n}\left\|\int_{\mathbb{R}^n} x\,d\nu(x)-\int_{\mathbb{R}^n} x\,d\mu_u^{\otimes n}(x)\right\|_1 &= \frac{1}{n}\sum_{i=1}^n\left|\int_{\mathbb{R}^n} x_i\,d\nu(x)-\int_{\mathbb{R}^n} x_i\,d\mu_u^{\otimes n}(x)\right| \\ &= \frac{1}{n}\sum_{i=1}^n\left|\int_{\mathbb{R}} x\,d\nu_i(x)-\int_{\mathbb{R}} x\,d\mu_{u_i}(x)\right|.\end{aligned}$$

Comme pour tout $i\in\{1,\ldots,n\}$, $u_i\in J$, on a, d'après la proposition V.17,

$$\left|\int_{\mathbb{R}} x\,d\nu_i(x)-\int_{\mathbb{R}} x\,d\mu_{u_i}(x)\right|\leq Q_J(\mathrm{H}\left(\nu_i|\,\mu_{u_i}\right)).$$

Donc

$$\frac{1}{n}\left\|\int_{\mathbb{R}^n} x\,d\nu(x)-\int_{\mathbb{R}^n} x\,d\mu_u^{\otimes n}(x)\right\|_1\leq\frac{1}{n}\sum_{i=1}^n Q_J(\mathrm{H}\left(\nu_i|\,\mu_{u_i}\right)).$$

La fonction Q_J étant concave, on a, d'après l'inégalité de Jensen,

$$\frac{1}{n}\left\|\int_{\mathbb{R}^n} x\,d\nu(x)-\int_{\mathbb{R}^n} x\,d\mu_u^{\otimes n}(x)\right\|_1\leq Q_J\left(\frac{\sum_{i=1}^n \mathrm{H}\left(\nu_i|\,\mu_{u_i}\right)}{n}\right).$$

D'après la formule de décomposition entropique (II.4),

$$\mathrm{H}\left(\nu|\,\mu_u^{\otimes n}\right)=\mathrm{H}\left(\nu|\,\nu_1\otimes\cdots\otimes\nu_n\right)+\sum_{i=1}^n\mathrm{H}\left(\nu_i|\,\mu_{u_i}\right).$$

En particulier,

$$\sum_{i=1}^n\mathrm{H}\left(\nu_i|\,\mu_{u_i}\right)\leq\mathrm{H}\left(\nu|\,\mu_u^{\otimes n}\right).$$

La fonction Q_J étant croissante, on en déduit que

$$\frac{1}{n}\left\|\int_{\mathbb{R}^n} x\,d\nu(x)-\int_{\mathbb{R}^n} x\,d\mu_u^{\otimes n}(x)\right\|_1\leq Q_J\left(\frac{\mathrm{H}\left(\nu|\,\mu_u^{\otimes n}\right)}{n}\right).$$

$\square$

V.4.2 Quelques majorations explicites

Nous donnons dans cette section quelques majorations élémentaires de la fonction Q intervenant dans la proposition V.17.

Proposition V.22. *Si μ est telle que, pour tout $u \in \mathbb{R}$, $\Lambda''_\mu(u) \leq M$, alors on a pour toute $\nu \in \mathcal{P}(\mathbb{R})$ et tout $u \in \mathbb{R}$:*

$$\left| \int_{\mathbb{R}} x \, d\nu(x) - \int_{\mathbb{R}} x \, d\mu_u(x) \right| \leq \sqrt{2M \, \mathrm{H}\,(\nu |\, \mu_u)}$$

Démonstration. D'après la formule de Taylor-Lagrange, pour tout $u, t \in \mathbb{R}$, il existe a tel que

$$q(t,u) = \Lambda_\mu(u+t) - \Lambda_\mu(u) - t\Lambda'_\mu(u) = \tfrac{t^2}{2}\Lambda''_\mu(a).$$

Donc $q(t,u) \leq \frac{t^2 M}{2}$; en reprenant la preuve de la proposition V.17, on voit que l'on peut prendre $k(t) = \frac{tM}{2}$. Un simple calcul donne alors : $Q(x) = \sqrt{2Mx}$. □

Exemples :

- μ est à support inclus dans $[a, b]$

Le support de μ_u est également inclus dans $[a, b]$ et $\Lambda''_\mu(u) = \mathrm{Var}(\mu_u) \leq (b-a)^2$.
Dans ce cas, on peut donc prendre $Q(x) = (b-a)\sqrt{2x}$.

- $\mu = Z^{-1}e^{-U}dx$, avec $U'' \geq c > 0$:

La probabilité μ satisfait alors une inégalité de Poincaré de constante $\frac{1}{c}$ (pas nécessairement optimale), *ie*

$$\mathrm{Var}_\mu(f) \leq \frac{1}{c}\int_{\mathbb{R}} (f')^2(x) \, d\mu(x).$$

Or $\mu_u = \frac{e^{-U+ux}}{ZZ_\mu(u)}dx$ et $V = U(x) + ux$ vérifie également $V'' \geq c > 0$, donc μ_u vérifie également une inégalité de Poincaré avec la même constante.
En particulier, en prenant $f(x) = x$, on obtient

$$\Lambda''_\mu(u) = \mathrm{Var}(\mu_u) = \mathrm{Var}_{\mu_u}(x) \leq \frac{1}{c}.$$

Dans ce cas, on peut donc prendre $Q(x) = \sqrt{\frac{2x}{c}}$.

Le lemme suivant va nous permettre, dans certains cas, de majorer la fonction Q par une fonction continue, croissante, positive, nulle en 0, mais non concave en général.

Lemme V.23. *Soit $k : [0, +\infty[\to \mathbb{R}^+$ une fonction de classe C^2 telle que $k(0) = k'(0) = 0$ et $k'' \geq c > 0$.*
Posons $\Psi(t) = \int_0^t uk''(u)du = tk'(t) - k(t)$.
Alors

1. *Pour tout* $a \in \mathbb{R}^+$,

$$Q(a) = \inf\left\{\frac{a}{t} + \frac{k(t)}{t}\right\} = k'(\Psi^{-1}(a))$$

2. *De plus, pour tout* $a \in \mathbb{R}^+$, $Q(a) \leq k'\left(\sqrt{\frac{2a}{c}}\right)$

Démonstration. 1) Pour tout $a > 0$, $g_a : t \mapsto \dfrac{a}{t} + \dfrac{k(t)}{t}$ admet $+\infty$ comme limite en 0 et $+\infty$; g_a atteint donc son minimum en un point t_a tel que $g_a'(t_a) = 0$, c'est-à-dire tel que $\Psi(t_a) = a$. La fonction Ψ étant strictement croissante, on a $t_a = \Psi^{-1}(a)$ et ceci reste vrai pour $a = 0$. De plus,

$$Q(a) = \frac{a}{t_a} + \frac{k(t_a)}{t_a} = \frac{k'(t_a)t_a - k(t_a)}{t_a} + \frac{k(t_a)}{t_a} = k'(t_a) = k'(\Psi^{-1}(a)).$$

2) $a = \int_0^{t_a} uk''(u)du \geq \int_0^{t_a} cudu = c\frac{t_a^2}{2}$. Donc $t_a \leq \sqrt{\frac{2a}{c}}$ et k' étant croissante, on a

$$Q(a) = k'(t_a) \leq k'\left(\sqrt{\frac{2a}{c}}\right).$$

□

Exemples :
- μ est la loi de Poisson de paramètre $\lambda > 0$

On a $\Lambda_\mu(u) = \lambda(e^u - 1)$ et $\Lambda_\mu(u+t) + \Lambda_\mu(u-t) - 2\Lambda_\mu(u) = 2\lambda e^u[\cosh(t) - 1]$.
Soit $M > 0$; en posant $k(t) = 2\lambda e^M[\cosh(t) - 1]$, on a en reprenant la preuve de la proposition V.17, pour tout $u \in [-M, M]$ et toute $\nu \in \mathcal{P}(\mathbb{R})$

$$\left|\int_{\mathbb{R}} x\, d\nu(x) - \int_{\mathbb{R}} x\, d\mu_u(x)\right| \leq Q_M(\mathrm{H}\,(\nu|\,\mu_u)),$$

avec $Q_M(a) = \inf\left\{\frac{a}{t} + \frac{k(t)}{t}\right\}$.
De plus $k'(t) = 2\lambda e^M \sinh(t)$ et $k''(t) = 2\lambda e^M \cosh(t) \geq 2\lambda e^M$, donc, d'après le lemme précédent,

$$Q_M(a) \leq 2\lambda e^M \sinh\left(\sqrt{\frac{e^{-M}a}{\lambda}}\right).$$

Ainsi, pour tout $u \in [-M, M]$ et toute $\nu \in \mathcal{P}(\mathbb{R})$,

$$\left|\int_{\mathbb{R}} x\, d\nu(x) - \int_{\mathbb{R}} x\, d\mu_u(x)\right| \leq 2\lambda e^M \sinh\left(\sqrt{\frac{e^{-M}\,\mathrm{H}\,(\nu|\,\mu_u)}{\lambda}}\right).$$

- μ est la loi exponentielle de paramètre λ

En adaptant légèrement la preuve du lemme précédent, on obtient :
Pour tout $u \leq b < \lambda$ et toute $\nu \in \mathcal{P}(\mathbb{R})$ telle que $\mathrm{H}(\nu|\, \mu_u) < 1$,

$$\left| \int_{\mathbb{R}} x \, d\nu(x) - \int_{\mathbb{R}} x \, d\mu_u(x) \right| \leq \frac{2}{\lambda - b} \frac{\sqrt{\mathrm{H}(\nu|\, \mu_u)}}{1 - \mathrm{H}(\nu|\, \mu_u)}$$

V.5 Principe conditionnel

V.5.1 Majoration de la distance en variation entre l'estimateur bayesien et l'estimateur M.E.M.

D'après le théorème V.8, il existe n_0 tel que, pour tout $n \geq n_0$ et tout $\varepsilon \geq 0$, la probabilité $\mu^*_{n,\varepsilon}$ est bien définie et s'écrit

$$\mu^*_{n,\varepsilon} = \mu^{\otimes n}_{w^*_{n,\varepsilon}}.$$

Lemme V.24. *Pour toute suite ε_n de réels positifs convergeant vers 0, il existe $m \geq n_0$ et un segment $J \subset]-\alpha, \beta[$ tel que*

$$\forall n \geq m, \quad w^*_{n,\varepsilon_n} \in J^n \qquad \textit{et} \qquad \forall x \in \mathcal{X}, \quad \langle v^*, F(x) \rangle \in J$$

Démonstration. D'après le point (2) du théorème V.8,

$$w^*_{n,\varepsilon_n} = \begin{bmatrix} \left\langle F(x^n_1), v^*_{n,\varepsilon_n} \right\rangle \\ \vdots \\ \left\langle F(x^n_n), v^*_{n,\varepsilon_n} \right\rangle \end{bmatrix}.$$

La fonction F étant continue sur le compact $\mathcal{X}$, il existe $N > 0$ tel que $\|F(x)\| \leq N$ pour tout $x \in \mathcal{X}$. Pour tout $i \in \{1, \ldots, n\}$, on a, d'après l'inégalité de Cauchy-Schwarz,

$$\left| \left\langle F(x^n_i), v^*_{n,\varepsilon_n} \right\rangle - \left\langle F(x^n_i), v^* \right\rangle \right| \leq N \left\| v^*_{n,\varepsilon_n} - v^* \right\|$$

et donc

$$\inf_{x \in \mathcal{X}} \langle v^*, F(x) \rangle - N \left\| v^*_{n,\varepsilon_n} - v^* \right\| \leq \left(w^*_{n,\varepsilon_n} \right)_i \leq \sup_{x \in \mathcal{X}} \langle v^*, F(x) \rangle + N \left\| v^*_{n,\varepsilon_n} - v^* \right\|.$$

D'après l'hypothèse (4) de (V.7), $v^* \in \overset{\circ}{\mathrm{dom}}\, H$. On voit facilement que

$$\overset{\circ}{\mathrm{dom}}\, H = \left\{ v \in \mathbb{R}^k : \forall x \in \mathcal{X}, quad \langle v, F(x) \rangle \in]-\alpha, \beta[\right\}.$$

Grâce à la compacité de $\mathcal{X}$, on a

$$-\alpha < \inf_{x \in \mathcal{X}} \langle v^*, F(x) \rangle \leq \sup_{x \in \mathcal{X}} \langle v^*, F(x) \rangle < \beta.$$

D'après le point (5) du théorème V.8, v^*_{n,ε_n} converge vers v^* ; le résultat en découle facilement. $\square$

Lemme V.25. *Il existe $M > 0$ et $n_1 \geq n_0$ tels que pour tout $\varepsilon > 0$ et $n \geq n_1$,*

$$\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon) \geq 1 - 2k \exp\left(-\frac{n\varepsilon^2}{2M(2M+\varepsilon)}\right),$$

(où k est la dimension de fonction $F = (f_1, \ldots, f_k)$.)

Démonstration.
Première étape :
Montrons que pour tout segment $J \subset]-\alpha, \beta[$, il existe $M > 0$ tel que, pour tout $u \in J$ et $j \geq 2$,

$$\int_{\mathbb{R}} \left| z - \int_{\mathbb{R}} x \, d\mu_u(x) \right|^j d\mu_u(z) \leq j! M^j.$$

En notant $\tau(x) = e^{|x|} - 1 - |x|$ et $I(u, M) = \int_{\mathbb{R}} \tau\left(\frac{z - \int_{\mathbb{R}} x d\mu_u(x)}{M}\right) d\mu_u(z)$, on voit facilement que

$$\sup_{u \in J}(I(u, M)) \xrightarrow[M \to +\infty]{} 0.$$

Par conséquent, il existe $M > 0$ tel que $\sup\limits_{u \in J}(I(u, M)) \leq 1$.

Or,

$$I(u, M) = \sum_{j=2}^{+\infty} \frac{\int_{\mathbb{R}} \left| z - \int_{\mathbb{R}} x \, d\mu_u(x) \right|^j d\mu_u(z)}{M^j j!},$$

donc, pour tout $u \in J$ et $j \geq 2$, on a

$$\frac{\int_{\mathbb{R}} \left| z - \int_{\mathbb{R}} x \, d\mu_u(x) \right|^j d\mu_u(z)}{M^j j!} \leq I(u, M) \leq 1.$$

Deuxième étape :
Montrons que pour tout segment $J \subset]-\alpha, \beta[$ et tout $N > 0$, il existe $M > 0$ tel que, pour toute suite $Z_1, \ldots, Z_n$ de variables aléatoires indépendantes avec $\mathcal{L}(Z_i) = \mu_{u_i}$, $u_i \in J$ et toute suite $\alpha_1, \ldots \alpha_n \in \mathbb{R}$ telle que $|\alpha_i| \leq N$, on ait :

$$\mathbb{P}\left(\left|\overline{Z} - m\right| > \varepsilon\right) \leq 2 \exp\left(-\frac{n\varepsilon^2}{2M(2M+\varepsilon)}\right)$$

où $\overline{Z} = \frac{1}{n} \sum_{i=1}^{n} \alpha_i Z_i$ et $m = \mathbb{E}\left[\overline{Z}\right]$.

D'après la première étape, il existe $M_0 > 0$ ne dépendant que de J tel que pour tout i, on ait

$$\forall j \geq 2, \quad \mathbb{E}\left[|Z_i - E[Z_i]|^j\right] \leq j! M_0^j.$$

On en déduit que pour tout i,

$$\forall j \geq 2, \quad \mathbb{E}\left[|\alpha_i(Z_i - E[Z_i])|^j\right] \leq j!(M_0 N)^j.$$

En prenant $M = M_0 N$, le résultat découle de l'inégalité (III.64) du corollaire III.63.

Troisième étape :
A présent, montrons le lemme.
Soit $c_n = (c_{n,1}, \ldots, c_{n,k}) := \mathbb{E}_{\mu_n^*}[\langle L_n, F\rangle] \in K$.
Alors,

$$\begin{aligned}\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon) &\geq \mu_n^*(\|\langle L_n, F\rangle - c_n\|_\infty \leq \varepsilon)\\ &= 1 - \mu_n^*(\|\langle L_n, F\rangle - c_n\|_\infty > \varepsilon)\\ &\geq 1 - \sum_{p=1}^{k} \mu_n^*\left(z : \left|\frac{1}{n}\sum_{i=1}^{n} z_i f_p(x_i^n) - c_{n,p}\right| > \varepsilon\right)\end{aligned}$$

Les fonctions f_p étant continues sur le compact $\mathcal{X}$, il existe $N > 0$ tel que $|f_p(x)| \leq N$ pour tout p et $x \in \mathcal{X}$. De plus, d'après le lemme V.24 appliqué à la suite $\varepsilon_n = 0$, il existe $n_1 \geq n_0$ et un segment $J \subset]-\alpha, \beta[$ tel que, pour tout $n \geq n_1$, $w_n^* \in J^n$. Ainsi, d'après la deuxième étape, on peut conclure qu'il existe $M > 0$ tel que, pour tout $\varepsilon > 0$ et tout $n \geq n_1$, on ait

$$\forall \varepsilon \geq 0, \quad \mu_n^*(\langle L_n, F\rangle \in K^\varepsilon) \geq 1 - 2k \exp\left[-\frac{n\varepsilon^2}{2M(2M+\varepsilon)}\right].$$

□

Nous pouvons maintenant prouver la

Proposition V.26. *Si ε_n est une suite de réels strictement positifs de limite nulle telle que $\lim_{n\to+\infty} n\varepsilon_n^2 = +\infty$, alors*

1. *Il existe $n_2 \geq n_0$ tel que, pour tout $n \geq n_2$, R_{n,ε_n} et R^*_{n,ε_n} sont bien définies.*

2. *Il existe $Q \in \mathcal{Q}$ telle que, pour tout $n \geq n_2$,*

$$\left\|R_{n,\varepsilon_n} - R^*_{n,\varepsilon_n}\right\|_{VT} \leq Q\left(\frac{-1}{n}\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon_n}) e^{\mathrm{H}\left(\mu^*_{n,\varepsilon_n} | \mu^{\otimes n}\right)}\right]\right)$$

Démonstration.
(1) Pour $n \geq n_0$, μ_n^* et μ^*_{n,ε_n} sont bien définies. De plus, d'après le lemme V.25, il existe $n_1 \geq n_0$ et $M > 0$ tels que, pour tout $n \geq n_1$,

$$\mu_n^*(\langle L_n, F\rangle \in K^{\varepsilon_n}) \geq 1 - 2k \exp\left[-\frac{n\varepsilon_n^2}{2M(2M+\varepsilon_n)}\right].$$

Comme $n\varepsilon_n^2 \xrightarrow[n\to+\infty]{} +\infty$, il est clair que $\mu_n^*(\langle L_n, F\rangle \in K^{\varepsilon_n}) \xrightarrow[n\to+\infty]{} 1$. En particulier, il existe $m_1 \geq n_1$ tel que, pour tout $n \geq m_1$, $\mu_n^*(\langle L_n, F\rangle \in K^{\varepsilon_n}) > 0$. Comme $\mu^{\otimes n}$ est équivalente à μ_n^*, on en déduit que pour tout $n \geq m_1$, $\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon_n}) > 0$ et en particulier, $R_{n,\,\varepsilon_n}$ est bien définie.

(2) D'après le lemme V.24, il existe un segment $J \subset]-\alpha, \beta[$ et $m_2 \geq n_0$ tels que, pour tout $n \geq m_2$, $w_{n,\,\varepsilon_n}^* \in J^n$. Soit $\nu_{n,\,\varepsilon_n} \in \mathcal{P}(\mathbb{R}^n)$ définie par

$$\nu_{n,\,\varepsilon_n} = \frac{\mathbb{1}_{\mathcal{S}(F,K^{\varepsilon_n})}(L_n)}{\mu^{\otimes n}(L_n \in \mathcal{S}(F, K^{\varepsilon_n}))}.$$

D'après le corollaire V.21, on a pour tout $n \geq n_2 = \max(m_1, m_2)$, en posant $Q = Q_J$

$$\frac{1}{n}\left\|\int_{\mathbb{R}^n} x\, d\nu_{n,\,\varepsilon_n}(x) - \int_{\mathbb{R}^n} x\, d\mu_{n,\,\varepsilon_n}^*(x)\right\|_1 \leq Q\left(\frac{\mathrm{H}\left(\nu_{n,\,\varepsilon_n}\middle|\, \mu_{n,\,\varepsilon_n}^*\right)}{n}\right)$$

Mais

$$\begin{aligned}\left\|R_{n,\,\varepsilon_n} - R_{n,\,\varepsilon_n}^*\right\|_{VT} &= \left\|\frac{1}{n}\sum_{i=1}^n \left(\int_{\mathbb{R}^n} z_i\, d\nu_{n,\,\varepsilon_n}(dz) - \int_{\mathbb{R}^n} z_i\, d\mu_{n,\,\varepsilon_n}^*(dz)\right)\delta_{x_i^n}\right\|_{VT} \\ &\leq \frac{1}{n}\sum_{i=1}^n \left|\int_{\mathbb{R}^n} z_i\, d\nu_{n,\,\varepsilon_n}(dz) - \int_{\mathbb{R}^n} z_i\, d\mu_{n,\,\varepsilon_n}^*(dz)\right| \\ &= \frac{1}{n}\left\|\int_{\mathbb{R}^n} x\, d\nu_{n,\,\varepsilon_n}(x) - \int_{\mathbb{R}^n} x\, d\mu_{n,\,\varepsilon_n}^*(x)\right\|_1.\end{aligned}$$

Donc, pour tout $n \geq n_2$,

$$\left\|R_{n,\,\varepsilon_n} - R_{n,\,\varepsilon_n}^*\right\|_{VT} \leq Q\left(\frac{\mathrm{H}\left(\nu_{n,\,\varepsilon_n}\middle|\, \mu_{n,\,\varepsilon_n}^*\right)}{n}\right).$$

Or, on voit facilement que

$$\nu_{n,\,\varepsilon_n} \in \Pi_n(K^{\varepsilon_n}).$$

En appliquant l'inégalité (II.26) de Csiszár, on a

$$\mathrm{H}\left(\nu_{n,\,\varepsilon_n}\middle|\, \mu^{\otimes n}\right) \geq \mathrm{H}\left(\nu_{n,\,\varepsilon_n}\middle|\, \mu_{n,\,\varepsilon_n}^*\right) + \mathrm{H}\left(\mu_{n,\,\varepsilon_n}^*\middle|\, \mu^{\otimes n}\right).$$

De plus, un simple calcul montre que

$$\mathrm{H}\left(\nu_{n,\,\varepsilon_n}\middle|\, \mu^{\otimes n}\right) = -\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon_n})\right]$$

et donc

$$\mathrm{H}\left(\nu_{n,\,\varepsilon_n}\middle|\, \mu_{n,\,\varepsilon_n}^*\right) \leq -\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon_n})e^{\mathrm{H}\left(\mu_{n,\,\varepsilon_n}^*\middle|\, \mu^{\otimes n}\right)}\right].$$

La fonction Q étant croissante, on obtient, pour tout $n \geq n_2$,

$$\left\|R_{n,\,\varepsilon_n} - R_{n,\,\varepsilon_n}^*\right\|_{VT} \leq Q\left(\frac{-1}{n}\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon_n})e^{\mathrm{H}\left(\mu_{n,\,\varepsilon_n}^*\middle|\, \mu^{\otimes n}\right)}\right]\right).$$

□

V.5.2 Convergence des estimateurs bayesiens

Nous aurons besoin du lemme suivant, très similaire à la proposition III.44 :

Lemme V.27. *Dès que* $\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon) > 0$*, on a*

$$\frac{1}{n}\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^\varepsilon)e^{\mathrm{H}\left(\mu_n^*|\,\mu^{\otimes n}\right)}\right] \geq \frac{\mathrm{H}\left(\mu_n^*|\,\mu^{\otimes n}\right)}{n}\left(1 - \frac{1}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}\right) + \frac{1}{n}\log\left[\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)\right] - \frac{1}{ne}\frac{1}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}$$

Démonstration. La probabilité $\mu^{\otimes n}$ étant équivalente à μ_n^*, on a

$$\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon) > 0 \Rightarrow \mu^{\otimes n}(\langle L_n, F\rangle \in K^\varepsilon) > 0.$$

On a

$$\begin{aligned}
&\frac{1}{n}\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^\varepsilon)\right] = \frac{1}{n}\log\left[\int_{\mathbb{R}^n} \mathbb{1}_{K^\varepsilon}(\langle L_n, F\rangle)\, d\mu^{\otimes n}\right] \\
&= \frac{1}{n}\log\left[\int_{\mathbb{R}^n} \mathbb{1}_{K^\varepsilon}(\langle L_n, F\rangle)\frac{d\mu^{\otimes n}}{d\mu_n^*}\, d\mu_n^*\right] \\
&= \frac{1}{n}\log\left[\int_{\mathbb{R}^n} \frac{d\mu^{\otimes n}}{d\mu_n^*}\frac{\mathbb{1}_{K^\varepsilon}(\langle L_n, F\rangle)}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}\, d\mu_n^*\right] + \frac{1}{n}\log\left[\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)\right].
\end{aligned}$$

En appliquant l'inégalité de Jensen avec la probabilité $\dfrac{\mathbb{1}_{K^\varepsilon}(\langle L_n, F\rangle)}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}\mu_n^*$, on obtient

$$\frac{1}{n}\log\left[\int_{\mathbb{R}^n}\frac{d\mu^{\otimes n}}{d\mu_n^*}\frac{\mathbb{1}_{K^\varepsilon}(\langle L_n, F\rangle)}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}d\mu_n^*\right] \geq \frac{1}{n}\int_{\mathbb{R}^n}\log\left(\frac{d\mu^{\otimes n}}{d\mu_n^*}\right)\frac{\mathbb{1}_{K^\varepsilon}(\langle L_n, F\rangle)}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}d\mu_n^*.$$

De plus, en posant $I_n = \dfrac{1}{n}\displaystyle\int_{\mathbb{R}^n}\log\left(\frac{d\mu^{\otimes n}}{d\mu_n^*}\right)\frac{\mathbb{1}_{K^\varepsilon}(\langle L_n, F\rangle)}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}\, d\mu_n^*$, on a

$$\begin{aligned}
I_n &= \frac{1}{n\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}\int_{\mathbb{R}^n}\log\left(\frac{d\mu^{\otimes n}}{d\mu_n^*}\right)d\mu_n^* - \frac{1}{n}\int_{\mathbb{R}^n}\log\left(\frac{d\mu^{\otimes n}}{d\mu_n^*}\right)\frac{\mathbb{1}_{(K^\varepsilon)^c}(\langle L_n, F\rangle)}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}d\mu_n^* \\
&= \frac{-\mathrm{H}\left(\mu_n^*|\mu^{\otimes n}\right)}{n\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)} + \frac{1}{n}\int_{\mathbb{R}^n}\log\left(\frac{d\mu_n^*}{d\mu^{\otimes n}}\right)\frac{d\mu_n^*}{d\mu^{\otimes n}}\frac{\mathbb{1}_{(K^\varepsilon)^c}(\langle L_n, F\rangle)}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}d\mu^{\otimes n}.
\end{aligned}$$

Mais la fonction $x \mapsto x\log(x)$ étant minorée par $-\frac{1}{e}$, on a

$$\begin{aligned}
\frac{1}{n}\int_{\mathbb{R}^n}\log\left(\frac{d\mu_n^*}{d\mu^{\otimes n}}\right)\frac{d\mu_n^*}{d\mu^{\otimes n}}\frac{\mathbb{1}_{(K^\varepsilon)^c}(\langle L_n, F\rangle)}{\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}\, d\mu^{\otimes n} &\geq -\frac{\mu^{\otimes n}(\langle L_n, F\rangle \notin K^\varepsilon)}{ne\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)} \\
&\geq -\frac{1}{ne\mu_n^*(\langle L_n, F\rangle \in K^\varepsilon)}.
\end{aligned}$$

Ainsi,

$$\frac{1}{n}\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon})\right] \geq \frac{-\,\mathrm{H}\left(\mu_n^*\middle|\,\mu^{\otimes n}\right)}{n\mu_n^*(\langle L_n, F\rangle \in K^{\varepsilon})} + \frac{1}{n}\log\left[\mu_n^*(\langle L_n, F\rangle \in K^{\varepsilon})\right] - \frac{1}{ne}\frac{1}{\mu_n^*(\langle L_n, F\rangle \in K^{\varepsilon})}$$

et on obtient le résultat en ajoutant $\frac{\mathrm{H}\left(\mu_n^*\middle|\mu^{\otimes n}\right)}{n}$ aux deux membres. □

Démonstration du théorème V.16.
Il suffit de montrer que $\lim\limits_{n\to+\infty} \left\|R_{n,\,\varepsilon_n} - R^*_{n,\,\varepsilon_n}\right\|_{VT} = 0$.
D'après le point (2) de la proposition V.26, il existe $Q \in \mathcal{Q}$ et n_2 tel que, pour tout $n \geq n_2$,

$$\left\|R_{n,\,\varepsilon_n} - R^*_{n,\,\varepsilon_n}\right\|_{VT} \leq Q\left(\frac{-1}{n}\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon_n})e^{\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\middle|\,\mu^{\otimes n}\right)}\right]\right).$$

La fonction Q étant continue, croissante et nulle en 0, il suffit de majorer

$$B_n := \frac{-1}{n}\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon_n})e^{\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\middle|\mu^{\otimes n}\right)}\right]$$

par une quantité convergeant vers 0.
Écrivons

$$B_n = B_n^1 + B_n^2,$$

avec

$$B_n^1 = \frac{-1}{n}\log\left[\mu^{\otimes n}(\langle L_n, F\rangle \in K^{\varepsilon_n})e^{\mathrm{H}\left(\mu^*_n\middle|\mu^{\otimes n}\right)}\right]$$

et

$$B_n^2 = \frac{1}{n}\left[\mathrm{H}\left(\mu_n^*\middle|\,\mu^{\otimes n}\right) - \mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\middle|\,\mu^{\otimes n}\right)\right].$$

Par un simple calcul,

$$\frac{\mathrm{H}\left(\mu_n^*\middle|\,\mu^{\otimes n}\right)}{n} = \frac{1}{n}\sum_{i=1}^{n}\Big(\langle F(x_i^n), v_n^*\rangle\Lambda'_\mu\left\langle F(x_i^n), v_n^*\right\rangle - \Lambda_\mu\left\langle F(x_i^n), v_n^*\right\rangle\Big),$$

$$\frac{\mathrm{H}\left(\mu^*_{n,\,\varepsilon_n}\middle|\,\mu^{\otimes n}\right)}{n} = \frac{1}{n}\sum_{i=1}^{n}\Big(\langle F(x_i^n), v^*_{n,\,\varepsilon_n}\rangle\Lambda'_\mu\left\langle F(x_i^n), v^*_{n,\,\varepsilon_n}\right\rangle - \Lambda_\mu\left\langle F(x_i^n), v^*_{n,\,\varepsilon_n}\right\rangle\Big).$$

Grâce à l'hypothèse (2) de (V.7), au point (4) du théorème V.8 et au lemme V.24, on voit facilement que $\frac{\mathrm{H}\left(\mu_n^*\middle|\mu^{\otimes n}\right)}{n}$ et $\frac{\mathrm{H}\left(\mu_{n,\varepsilon_n}^*\middle|\mu^{\otimes n}\right)}{n}$ convergent vers la même limite I, lorsque n tend vers $+\infty$: [1]

$$I = \int_{\mathcal{X}} \langle F(x), v^* \rangle \Lambda'_\mu \langle F(x), v^* \rangle \, dR(x) - \int_{\mathcal{X}} \Lambda_\mu \langle F(x), v^* \rangle \, dR(x).$$

En particulier,

$$B_n^2 \xrightarrow[n\to+\infty]{} 0.$$

Finalement, grâce aux lemmes V.25 et V.27, on voit facilement que B_n^1 est majoré par une quantité convergeant vers 0. $\square$

[1] Remarque : $I = \mathrm{I}_\mu(R^*|R)$

Deuxième partie

Inégalités de transport

CHAPITRE VI

Inégalités de transport convexes - Résultats préliminaires

Sommaire

VI.1 Transport de masse

VI.1.1 Le problème de Monge-Kantorovich

Le problème de trouver le moyen le plus économique de boucher un trou avec un tas de sable a été proposé vers 1780 par l'ingénieur Gaspard de Monge. Si sa formulation initiale peut sembler un peu désuète, cette question a posé et pose encore des problèmes mathématiques d'une grande difficulté et est à l'origine de théorèmes puissants ayant des répercussions dans des domaines tels que la théorie des probabilités, les équations aux dérivées partielles, l'analyse fonctionnelle ou l'isopérimétrie.

Dans la formulation qu'en a donné Kantorovich, le tas de sable est représenté par un espace de probabilité $(\mathcal{X}, \mu)$ et le trou, par un espace de probabilité $(\mathcal{Y}, \nu)$.

- Le coût nécessaire pour acheminer de la masse de $\mathcal{X}$ sur $\mathcal{Y}$ est représenté par une fonction $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$, appelée *fonction de coût*.

- Un *plan de transfert de μ sur ν* est une probabilité $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ ayant pour première marginale μ et pour seconde ν.

- Le *coût de transport* associé à ce plan de transfert est

$$I_c[\pi] := \iint_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, d\pi(x, y).$$

Dans cette dernière intégrale, $d\pi(x, y)$ représente la quantité de masse prise en x et déposée en y et $c(x, y)\, d\pi(x, y)$, le coût engendré par cette opération. La quantité $d\mu(x)$ représente la masse totale au point x ; dire que π admet μ pour première marginale, s'écrit formellement $d\mu(x) = \int_{\mathcal{Y}} d\pi(x, y)$, ce qui s'interprète en disant que la totalité de la masse en x a été distribuée. De la même manière, $d\nu(y)$ représente la quantité de masse que peut recevoir y et $d\nu(y) = \int_{\mathcal{X}} d\pi(x, y)$ signifie que y reçoit exactement cette masse.

- Le *coût de transport optimal* est

$$\mathcal{T}_c(\mu, \nu) = \inf \{I_c[\pi] : \pi \in \Pi(\mu, \nu)\},$$

  où $\Pi(\mu, \nu)$ est l'ensemble des plans de transfert de μ sur ν.

Le problème de Monge, dans la formulation de Kantorovich, est donc de trouver des plans de transfert π optimaux, *ie* tels que

$$I_c[\pi] = \mathcal{T}_c(\mu, \nu).$$

On pourra consulter les deux ouvrages ([56] et [72]) de référence sur le sujet pour des résultats caractérisant les plans de transferts optimaux pour certaines fonctions de coût.

Pour la suite, nous n'aurons besoin que du résultat basique suivant (voir [72], thm. 2.18 p. 74) :

Théorème VI.1. *Soit c une fonction de coût sur $\mathbb{R}$ de la forme $c(x,y) = q(x-y)$ avec q une fonction convexe positive paire. Si $\mu, \nu \in \mathcal{P}(\mathbb{R})$ ont pour fonctions de répartition F et G, la probabilité $\pi^* \in \mathcal{P}(\mathbb{R}^2)$ de fonction de répartition $H(x,y) = \min(F(x), G(y))$ appartient à $\Pi(\mu,\nu)$ et*

$$\mathcal{T}_c(\mu,\nu) = \iint_{\mathbb{R}^2} c(x,y)\, d\pi^*.$$

VI.1.2 La dualité de Kantorovich-Rübinstein

Le théorème suivant, appelé théorème de Kantorovich-Rübinstein, donne une formulation duale du coût de transport optimal :

Théorème VI.2.
Soient $\mathcal{X}$ et $\mathcal{Y}$ des espaces polonais, $\mu \in \mathcal{P}(\mathcal{X})$ et $\nu \in \mathcal{P}(\mathcal{Y})$, et soit $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}^+ \cup \{+\infty\}$, une fonction de coût continue.

Posons :

- $\Pi(\mu,\nu)$, *l'ensemble des mesures de probabilité π sur $\mathcal{X} \times \mathcal{Y}$, telles que π a pour première marginale μ et pour seconde ν,*

- Φ_c, *l'ensemble des couples de fonctions (φ, ψ), φ (resp. ψ) continue bornée sur $\mathcal{X}$ (resp. $\mathcal{Y}$), vérifiant :*

$$\forall x, y \in \mathcal{X}, \quad \varphi(x) + \psi(y) \leq c(x,y),$$

alors

$$\inf_{\pi\in\Pi(\mu,\nu)} \left\{ \iint_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, d\pi(x,y) \right\} = \sup_{(\varphi,\psi)\in\Phi_c} \left\{ \int_{\mathcal{X}} \varphi\, d\mu + \int_{\mathcal{Y}} \psi\, d\nu \right\}, \tag{VI.3}$$

et l'infimum dans le membre de gauche de (VI.3) est atteint.

De plus, si $(\mathcal{X}, d)$ est un espace polonais, alors

$$\inf_{\pi\in\Pi(\mu,\nu)} \left\{ \iint_{\mathcal{X}\times\mathcal{X}} d(x,y)\, d\pi(x,y) \right\} = \sup \left\{ \int_{\mathcal{X}} \varphi\, d\mu - \int_{\mathcal{X}} \varphi\, d\nu : \varphi \in \mathit{BLip}_1(\mathcal{X}, d) \right\}, \tag{VI.4}$$

où $\mathit{BLip}_1(\mathcal{X}, d)$ est l'ensemble des fonctions 1-Lipschitziennes, bornées sur $\mathcal{X}$.

Remarque VI.5.

En désignant par Φ_c^s l'ensemble des couples (φ, ψ) de fonctions semi-continues supérieurement sur $\mathcal{X}$ et $\mathcal{Y}$ vérifiant $\varphi(x) + \psi(y) \leq c(x, y)$, pour tout $(x, y) \in \mathcal{X} \times \mathcal{Y}$, on a aussi $\mathcal{T}_c(\nu, \mu) = \sup_{(\varphi,\psi)\in\Phi_c^s} \left\{ \int_{\mathcal{X}} \varphi\, d\mu + \int_Y \psi\, d\nu \right\}$.

Exemple : Dans cet exemple, nous allons nous placer dans une situation qui ne relève pas du théorème précédent. Soient $\mathcal{X}$ un espace mesurable et $\chi : \mathcal{X} \to \mathbb{R}^+$ une application mesurable. Définissons une semi-métrique d_χ sur $\mathcal{X}$ par

$$\forall x, y \in \mathcal{X}, \quad d_\chi(x, y) = (\chi(x) + \chi(y))\, \mathbb{1}_{x \neq y}. \tag{VI.6}$$

On voit facilement que si χ s'annule en au plus un point, d_χ est une vraie distance sur $\mathcal{X}$. La proposition suivante exprime le coût de transport optimal associé à d_χ.

Proposition VI.7. *Si $B_\chi(\mathcal{X})$ désigne l'ensemble des fonctions φ mesurables bornées sur $\mathcal{X}$ telles que $\forall x \in \mathcal{X},\ |\varphi(x)| \leq \chi(x)$, alors*

$$\forall \mu, \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_{d_\chi}(\mu, \nu) = \sup_{\varphi \in B_\chi(\mathcal{X})} \left\{ \int_{\mathcal{X}} \varphi\, d\mu - \int_{\mathcal{X}} \varphi\, d\nu \right\}.$$

En particulier, si $\int_{\mathcal{X}} \chi\, d\nu < +\infty$ et $\int_{\mathcal{X}} \chi\, d\mu < +\infty$, alors

$$\mathcal{T}_{d_\chi}(\mu, \nu) = \|\chi\mu - \chi\nu\|_{VT}$$

Démonstration. (Voir aussi la preuve de la proposition 7.10 de [72] et le lemme 7 page 23 de [49])

Tout d'abord, si $\varphi \in B_\chi(\mathcal{X})$, on a clairement $\varphi(x) - \varphi(y) \leq d_\chi(x, y)$; donc, pour tout $\pi \in \Pi(\mu, \nu)$, on a :

$$\int_{\mathcal{X}} \varphi\, d\mu - \int_{\mathcal{X}} \varphi\, d\nu = \iint_{\mathcal{X}^2} \varphi(x) - \varphi(y)\, d\pi \leq \iint_{\mathcal{X}^2} d_\chi(x, y)\, d\pi(x, y).$$

On en déduit que $\mathcal{T}_{d_\chi}(\mu, \nu) \geq \sup_{\varphi \in B_\chi(\mathcal{X})} \left\{ \int_{\mathcal{X}} \varphi\, d\mu - \int_{\mathcal{X}} \varphi\, d\nu \right\}$.

Pour montrer l'inégalité opposée, considérons la probabilité $\pi^* \in \mathcal{P}(\mathcal{X}^2)$, définie pour toute f mesurable bornée sur $\mathcal{X}^2$ par

$$\begin{aligned} \iint_{\mathcal{X}^2} f(x, y)\, d\pi^*(x, y) = &\int_{\mathcal{X}} f(x, x)\, d(\mu \wedge \nu)(x) \\ &+ \frac{1}{\alpha} \iint_{\mathcal{X}^2} f(x, y)\, d(\mu - \nu)_+(x)\, d(\mu - \nu)_-(y), \end{aligned} \tag{VI.8}$$

où $\mu \wedge \nu = \mu - (\mu - \nu)_+$ et $\alpha = (\mu - \nu)_+(\mathcal{X}) = (\mu - \nu)_-(\mathcal{X})$.
On vérifie facilement que $\pi^* \in \Pi(\mu, \nu)$.
De plus,

$$\begin{aligned}\iint_{\mathcal{X}^2} d_\chi(x,y)d\pi^* &= \frac{1}{\alpha}\iint_{\mathcal{X}^2} (\chi(x)+\chi(y))\mathbb{1}_{x\neq y}\, d(\mu-\nu)_+(x)\, d(\mu-\nu)_-(y)\\ &\leq \frac{1}{\alpha}\iint_{\mathcal{X}^2} (\chi(x)+\chi(y))\, d(\mu-\nu)_+(x)\, d(\mu-\nu)_-(y)\\ &= \int_{\mathcal{X}} \chi\, d(\mu-\nu)_+ + \int_{\mathcal{X}} \chi\, d(\mu-\nu)_-\\ &= \int_{\mathcal{X}} \chi\, d|\mu-\nu|\end{aligned}$$

et on voit facilement que $\displaystyle\int_{\mathcal{X}} \chi\, d|\mu-\nu| = \sup_{\varphi\in B_\chi(\mathcal{X})}\left\{\int_{\mathcal{X}} \varphi\, d\mu - \int_{\mathcal{X}} \varphi\, d\nu\right\}$. □

Remarque VI.9.

- Si $\chi = 1$,

$$\mathcal{T}_{d_1}(\mu,\nu) = \|\mu-\nu\|_{VT} = 2\inf\{\mathbb{P}(X\neq Y) : \mathcal{L}(X)=\mu, \mathcal{L}(Y)=\nu\}.$$

- Si on se place dans un cadre discret $\mathcal{X} = \{x_1, \ldots, x_n, \ldots\}$, π^* est défini par :

$$\pi^*(x,y) = \min(\mu(x),\nu(x)),\ \text{si } x=y; \quad \frac{1}{\alpha}(\mu-\nu)_+(x)(\mu-\nu)_-(y),\ \text{sinon}$$

  et correspond à la stratégie qui consiste à laisser en place la masse commune ($\min(\mu(x),\nu(x))$) et à distribuer l'excédent de μ par rapport à ν ($(\mu-\nu)_+(x)$) aux endroits y où $\mu(y)\leq\nu(y)$ proportionnellement au déficit de μ par rapport à ν ($\frac{1}{\alpha}(\mu-\nu)_-(y)$).

Lemme VI.10. *Une fonction ψ est 1-Lipschitzienne pour d_χ si, et seulement si, elle s'écrit $\psi = a+\varphi$, avec $a\in\mathbb{R}$ et $|\varphi|\leq\chi$.*

Démonstration. Il est clair que toute fonction $\psi = a+\varphi$, avec $a\in\mathbb{R}$ et $|\varphi|\leq\chi$ est 1-Lipschitzienne pour d_χ. Réciproquement, si ψ est une fonction 1-Lipschitzienne pour d_χ, alors pour tout $(x,y)\in\mathcal{X}^2$, on a

$$\psi(x)-\chi(x)\leq\psi(y)+\chi(y),$$

donc

$$a = \sup\{\psi(x)-\chi(x),\ \ x\in\mathcal{X}\} < +\infty.$$

Posons $\varphi = \psi - a$, alors

$$\varphi(x)-\chi(x) = \psi(x)-\chi(x)-\sup\{\psi(x)-\chi(x),\ \ x\in\mathcal{X}\}\leq 0.$$

Ainsi $\varphi \leq \chi$. De plus, pour tout $x, y \in \mathcal{X}$,

$$\psi(x) - \psi(y) \geq -\chi(x) - \chi(y),$$

donc

$$\varphi(x) + \chi(x) \geq \varphi(y) - \chi(y),$$

et par suite

$$\varphi(x) + \chi(x) \geq \sup\{\varphi(y) - \chi(y),\ \ y \in \mathcal{X}\} = \sup\{\psi(y) - \chi(y),\ \ y \in \mathcal{X}\} - a = 0.$$

Donc $\varphi \geq -\chi$. □

Remarque VI.11.

En notant, $\mathrm{BLip}_1(\mathcal{X}, d_\chi)$ l'ensemble des fonctions mesurables bornées et 1-Lipschitziennes pour d_χ, la proposition VI.7 s'énonce :

$$\mathcal{T}_{d_\chi}(\mu, \nu) = \sup_{\varphi \in \mathrm{BLip}_1(\mathcal{X}, d_\chi)} \left\{ \int_{\mathcal{X}} \varphi\, d\mu - \int_{\mathcal{X}} \varphi\, d\nu \right\}.$$

VI.1.3 Inégalités de Transport

Le sujet que nous allons aborder dans ce chapitre et le suivant est celui des *Inégalités de Transport*. Fixons $\mathcal{X}$ un espace mesurable (en général, $\mathcal{X}$ sera un espace polonais) et une fonction de coût $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$ sur $\mathcal{X}$ *symétrique*, *ie* telle que

$$\forall x, y \in \mathcal{X}, \quad c(x, y) = c(y, x).$$

Sous cette hypothèse de symétrie, nous aurons

$$\forall \nu, \mu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_c(\nu, \mu) = \mathcal{T}_c(\mu, \nu).$$

Nous dirons, provisoirement, qu'une probabilité μ vérifie une inégalité de transport, s'il existe une fonction F telle que

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_c(\nu, \mu) \leq F(\mathrm{H}(\nu | \mu)). \tag{VI.12}$$

L'étude des inégalités de transport est un sujet assez récent, initié par les travaux de K. Marton et M. Talagrand.

Bref historique sur les inégalités de transport.

L'inégalité de Pinsker (1964). La première inégalité de transport est l'inégalité de Pinsker : si $\mathcal{X}$ est un espace mesurable, on a

$$\forall \nu, \mu \in \mathcal{P}(\mathcal{X}), \quad \|\nu - \mu\|_{VT} \leq \sqrt{2\,\mathrm{H}\,(\nu|\,\mu)}.$$

C'est une inégalité de transport dans la mesure où, comme on l'a vu à la proposition VI.7, la norme en variation est le coût de transport optimal associé à la fonction de coût $c(x,y) = 2\mathbb{1}_{\{x\neq y\}}$.

Les premiers travaux de K. Marton (1986). Dans l'article [47], K. Marton obtient la généralisation suivante de l'inégalité de Pinsker :

Théorème VI.13. *Soit $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \cdots \times \mathcal{X}_n$ un produit d'espaces mesurables ; on définit sur $\mathcal{X}$, la distance de Hamming, notée $d_H(\,.\,,\,.\,)$, par la formule*

$$d_H(x,y) = \sum_{i=1}^{n} \mathbb{1}_{\{x_i \neq y_i\}}.$$

Si pour tout $i = 1\ldots n$, $\mu_i \in \mathcal{P}(\mathcal{X}_i)$, alors en posant $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$, on a

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_{d_H}(\nu,\mu) \leq \sqrt{\frac{n}{2}\,\mathrm{H}\,(\nu|\,\mu)}.$$

Remarque VI.14.

Remarquons que, pour $n = 1$, on retrouve bien l'inégalité de Pinsker.

Pour démontrer ce théorème, K. Marton utilise un argument de couplage astucieux sur lequel nous reviendrons dans la section VI.2.4. Le résultat précédent répond, dans un cas particulier, à la question suivante :

Si pour tout $i = 1\ldots n$, μ_i vérifie (VI.12) avec une fonction F_i, quelle inégalité de transport vérifie $\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$?

Nous aborderons en détail ce problème de la *tensorisation* des inégalités de transport dans la section VI.2.4.

Une conséquence intéressante du théorème VI.13, est l'obtention de résultats de concentration assez fins pour les mesures produit. Grâce à un argument d'une grande simplicité, appelé depuis *argument de Marton* (voir la proposition VI.81), K. Marton déduit du théorème VI.13 le résultat suivant :

Proposition VI.15. *Si X est un vecteur aléatoire à composantes indépendantes à valeurs dans $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \cdots \times \mathcal{X}_n$, alors pour tout ensemble mesurable A, on a*

$$\forall t \geq 0, \quad \mathbb{P}\left(d_H(X, A) \geq t + \sqrt{\frac{n}{2} \log\left(\frac{1}{\mathbb{P}(X \in A)}\right)}\right) \leq e^{-\frac{2t^2}{n}}.$$

Ce résultat est très proche des résultats de concentration de M. Talagrand (voir les articles [66] et [67]). Dans [48], K. Marton étend les résultats précédents au cas Markovien (μ (resp. X) est une probabilité Markovienne (resp. une chaîne de Markov)).

Travaux autour de l'inégalité $\mathbb{T}_2$. Soit $(\mathcal{X}, d)$ un espace polonais ; nous dirons que $\mu \in \mathcal{P}(\mathcal{X})$ satisfait l'inégalité de transport $\mathbb{T}_2(c)$, si

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_{d^2}(\nu, \mu) \leq c \,\mathrm{H}\left(\nu |\, \mu\right). \tag{VI.16}$$

L'inégalité (VI.16) est plus couramment écrite sous la forme équivalente suivante :

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathrm{W}_2(\nu, \mu) \leq \sqrt{c \,\mathrm{H}\left(\nu |\, \mu\right)}, \tag{VI.17}$$

où $\mathrm{W}_2(\nu, \mu) = \sqrt{\mathcal{T}_{d^2}(\nu, \mu)}$.

M. Talagrand est le premier à avoir démontré (VI.16) pour les mesures gaussiennes sur $\mathbb{R}^n$ muni de la distance euclidienne standard.

Théorème VI.18 (Talagrand, [68])**.** *La loi gaussienne standard sur $\mathbb{R}^n$ vérifie l'inégalité $\mathbb{T}_2(2)$ sur $\mathbb{R}^n$ muni de sa distance euclidienne.*

Pour démontrer le théorème précédent, Talagrand commence par démontrer, par des moyens assez élémentaires, que la loi gaussienne standard sur $\mathbb{R}$ vérifie $\mathbb{T}_2(2)$ pour $d(x, y) = |x - y|$. Il constate ensuite que l'inégalité $\mathbb{T}_2$ jouit d'une remarquable propriété de *tensorisation avec invariance de la constante*. En reprenant les techniques de couplage de Marton, il obtient la

Proposition VI.19. *Si pour tout $i = 1 \ldots n$, μ_i est une probabilité sur $\mathbb{R}$ vérifiant $\mathbb{T}_2(c)$, alors la probabilité $\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$ vérifie aussi l'inégalité $\mathbb{T}_2(c)$ sur $\mathbb{R}^n$ muni de sa distance euclidienne.*

Le théorème VI.18 découle alors immédiatement du cas $n = 1$ et de cette propriété de tensorisation. Par ailleurs, grâce à l'argument de Marton, le théorème VI.18 lui permet de montrer que pour tout Borélien B,

$$\forall \varepsilon \geq \sqrt{2 \log \frac{1}{\gamma(B)}}, \quad \gamma(B^\varepsilon) \geq 1 - \exp\left(-\frac{1}{2}\left(\varepsilon - \sqrt{2 \log \frac{1}{\gamma(B)}}\right)^2\right),$$

où γ est la loi gaussienne standard sur $\mathbb{R}^n$, et $B^\varepsilon = \{x \in \mathbb{R}^n, d(x,B) \leq \varepsilon\}$. Ce résultat de concentration est quasi optimal.

Dans [54], F. Otto et C. Villani ont étudié les liens existant entre l'inégalité $\mathbb{T}_2$ et les inégalités de Sobolev-logarithmiques et de Poincaré. Ils ont obtenu le résultat suivant

Théorème VI.20 (Otto-Villani (2000), [54])**.** *Soient Φ une application de $\mathbb{R}^n$ dans $\mathbb{R}$ telle que $e^{-\Phi}$ soit intégrable et μ la mesure de probabilité sur $\mathbb{R}^n$ définie par*

$$\frac{d\mu}{dx} = Z^{-1}e^{-\Phi},$$

avec $Z = \int e^{-\Phi}\, dx$.

1. *Si μ vérifie une inégalité de Sobolev logarithmique de constante c, c'est-à-dire que pour toute fonction f de classe C^1,*
$$\mathrm{Ent}_\mu(f^2) \leq c \int |\nabla f|^2 \, d\mu,$$
alors μ vérifie l'inégalité $\mathbb{T}_2(c)$ sur $\mathbb{R}^n$ muni de la distance euclidienne.
2. *Si μ vérifie l'inégalité $\mathbb{T}_2(c)$, alors μ vérifie l'inégalité de Poincaré de constante $\frac{c}{2}$, c'est-à-dire que pour toute fonction f de classe C^1,*
$$\mathrm{Var}_\mu(f) \leq \frac{c}{2} \int |\nabla f|^2 \, d\mu.$$

Ces résultats ont été redémontrés de manière plus simple par S.G. Bobkov, I. Gentil et M. Ledoux dans [3]. Le problème de savoir si l'inégalité $\mathbb{T}_2$ est équivalente à l'inégalité de Sobolev-Logarithmique ou non, n'a pas encore été résolu. On pourra consulter [14] pour des éléments de réponse.

Travaux autour de l'inégalité $\mathbb{T}_1$. Soit $(\mathcal{X}, d)$ un espace polonais ; on dit que $\mu \in \mathcal{P}(\mathcal{X})$ vérifie l'inégalité de transport $\mathbb{T}_1(c)$, si

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_d(\nu, \mu) \leq \sqrt{c\,\mathrm{H}\,(\nu|\,\mu)}. \tag{VI.21}$$

Cette inégalité de transport est strictement plus faible que l'inégalité $\mathbb{T}_2$. En effet, grâce à l'inégalité de Jensen, il est clair que

$$\forall \nu, \mu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_d(\nu, \mu) \leq \sqrt{\mathcal{T}_{d^2}(\nu, \mu)},$$

et par conséquent,

$$\mu \text{ satisfait } \mathbb{T}_2(c) \Rightarrow \mu \text{ satisfait } \mathbb{T}_1(c).$$

D'après l'argument de Marton, $\mathbb{T}_1$ est associée à un phénomène de concentration gaussienne : grossièrement, si μ satisfait une inégalité $\mathbb{T}_1$, alors pour tout ensemble mesurable A tel que $\mu(A) \geq \frac{1}{2}$, on a

$$\mu(A^\varepsilon) \geq 1 - e^{-C\varepsilon^2}, \quad \text{pour tout } \varepsilon \text{ assez grand},$$

où $A^\varepsilon = \{x \in \mathcal{X} : d(x, A) \geq \varepsilon\}$ (voir la proposition VI.81 pour un énoncé précis).

Dans [4], S. G. Bobkov et F. Götze ont obtenu un critère dual pour (VI.21). Ils ont montré le résultat suivant :

Théorème VI.22 (Bobkov-Götze (1999), [4], thm. 3.1)**.** *Une probabilité μ sur $\mathcal{X}$ vérifie $\mathbb{T}_1(c)$ si, et seulement si, pour toute fonction φ 1-Lipschitzienne, on a*

$$\forall s \in \mathbb{R}, \qquad \int_{\mathcal{X}} e^{s\varphi}\, d\mu \leq \exp\left(s \int_{\mathcal{X}} \varphi\, d\mu + c\frac{s^2}{4}\right). \tag{VI.23}$$

A la différence de l'inégalité $\mathbb{T}_2$, qui est en relation avec d'autres inégalités fonctionnelles non triviales, l'inégalité de transport $\mathbb{T}_1$ se résume à une propriété d'intégrabilité, comme le montre le théorème suivant, dû à H. Djellout, A. Guillin et L. Wu.

Théorème VI.24 (Djellout-Guillin-Wu,[27], thm. 3.1)**.** *Soit μ une probabilité sur $\mathcal{X}$; il y a équivalence entre les deux propositions suivantes :*

1. *Il existe $c > 0$ tel que μ vérifie $\mathbb{T}_1(c)$.*
2. *Il existe $\varepsilon > 0$ tel que $\int_{\mathcal{X}} e^{\varepsilon d(x,y)^2} d\mu(x)d\mu(y) < +\infty$.*

Nous préciserons plus loin le lien qui existe entre c et ε.

Dans [5], F. Boley et C. Villani, ont obtenu des versions pondérées de l'inégalité de Pinsker :

Théorème VI.25 (Bolley-Villani, [5], thm. 1)**.**
Soit $\chi : \mathcal{X} \to \mathbb{R}^+$, une fonction mesurable. Alors pour toute $\nu \in \mathcal{P}(\mathcal{X})$,

(i) $\|\chi\nu - \chi\mu\|_{VT} \leq \left(\frac{3}{2} + \log \int_{\mathcal{X}} e^{2\chi(x)}\, d\mu(x)\right) \left(\sqrt{\mathrm{H}\left(\nu|\,\mu\right)} + \frac{1}{2}\mathrm{H}\left(\nu|\,\mu\right)\right)$;

(ii) $\|\chi\nu - \chi\mu\|_{VT} \leq \sqrt{1 + \log \int_{\mathcal{X}} e^{\chi(x)^2} d\mu(x)}\sqrt{2\,\mathrm{H}\left(\nu|\,\mu\right)}$.

Remarquons que, d'après la proposition VI.7,

$$\|\chi\nu - \chi\mu\|_{VT} = \mathcal{T}_{d_\chi}(\nu, \mu),$$

avec d_χ définie par (VI.6). Si l'inégalité (ii) est une inégalité $\mathbb{T}_1$ au sens classique, l'inégalité (i) est une inégalité de transport faisant intervenir la fonction $F(x) = x + \sqrt{x}$ et non plus la fonction $F(x) = \sqrt{x}$.

Grâce à ces deux généralisations de l'inégalité de Pinsker, Bolley et Villani ont pu affiner le lien entre les constantes c et ε du théorème VI.24. Elles leur ont, par ailleurs, permis d'obtenir toute une famille d'inégalités de transport pour des coûts de la forme $c(x, y) = d^p(x, y), p > 1$.

Survol du chapitre

Ce chapitre a pour but d'introduire la notion d'*inégalités de transport convexes*, notion qui englobe tous les cas particuliers introduits plus haut, d'étudier certaines de leurs propriétés (on établira, notamment, une formule générale de tensorisation) et de les mettre en relations avec des inégalités de type Grandes Déviations.

Si θ est une fonction convexe appartenant à une certaine classe $\mathcal{C}$ que nous définirons plus loin, et si c est une fonction de coût symétrique sur un espace mesurable $\mathcal{X}$, on dira que $\mu \in \mathcal{P}(\mathcal{X})$ satisfait l'inégalité de transport convexe $T_c(\theta^*, a)$, si

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^*\left(\frac{\mathcal{T}_c(\nu, \mu)}{a}\right) \leq \mathrm{H}\left(\nu \,|\, \mu\right), \tag{VI.26}$$

la fonction θ^* étant la conjuguée convexe de la fonction convexe θ.

Par ailleurs, si Φ désigne une classe de fonctions mesurables bornées sur un espace mesurable $\mathcal{X}$ telle que $\varphi \in \Phi \Rightarrow -\varphi \in \Phi$, nous poserons

$$\forall \nu_1, \nu_2 \in \mathcal{P}(\mathcal{X}), \quad \|\nu_1 - \nu_2\|_\Phi^* = \sup_{\varphi \in \Phi} \left\{ \int_{\mathcal{X}} \varphi \, d\nu_1 - \int_{\mathcal{X}} \varphi \, d\nu_2 \right\}$$

et nous dirons que $\mu \in \mathcal{P}(\mathcal{X})$ vérifie l'inégalité $T_\Phi(\theta^*, a)$ si

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^*\left(\frac{\|\nu - \mu\|_\Phi^*}{a}\right) \leq \mathrm{H}\left(\nu \,|\, \mu\right). \tag{VI.27}$$

Les inégalités de la forme (VI.27) ne sont plus, à proprement parler, des inégalités de transport. Les semi-normes $\| \,.\, \|_\Phi^*$ sont des généralisations naturelles des coûts de transport optimaux associés à des fonctions de coûts métriques.

• Section VI.2 : Inégalités de transport convexes.

Dans la section VI.2.2, nous démontrerons une généralisation du critère (VI.23) de Bobkov et Götze. Si c est continue sur un espace polonais $(\mathcal{X}, d)$, nous verrons au théorème VI.38 que μ satisfait $T_c(\theta^*, a)$ si, et seulement si, pour tout couple $(\varphi, \psi) \in \Phi_c$, on a

$$\forall s \geq 0, \quad \int_{\mathcal{X}} \exp s(\varphi(x) + \langle \mu, \psi \rangle) \, d\mu(x) \leq \exp \theta(as). \tag{VI.28}$$

En particulier, si $c = d$, μ satisfait $T_d(\theta^*, a)$ si, et seulement si, pour toute fonction $\varphi \in BLip_1(\mathcal{X}, d)$, on a

$$\forall s \geq 0, \quad \int_{\mathcal{X}} \exp s(\varphi(x) - \langle \mu, \varphi \rangle) \, d\mu(x) \leq \exp \theta(as). \tag{VI.29}$$

De même, μ vérifie l'inégalité $T_\Phi(\theta^*, a)$ si, et seulement si, pour toute fonction $\varphi \in \Phi$, on a

$$\forall s \geq 0, \quad \int_{\mathcal{X}} \exp s(\varphi(x) - \langle \mu, \varphi \rangle)\, d\mu(x) \leq \exp \theta(as). \tag{VI.30}$$

Les critères précédents n'ont pas un caractère pratique, mais se révéleront d'une grande utilité théorique, notamment pour les questions de tensorisation. La preuve que nous donnons de (VI.28) est très différente de celle de Bobkov et Götze ; elle utilise des outils classiques en Théorie des Grandes Déviations : théorèmes de Cramér et Sanov, principe de contraction, *etc*. On pourra consulter [37] pour plus de détails sur les liens entre les Inégalités de Transport et les Grandes Déviations.

La proposition VI.48, de la section VI.2.3, donne une interprétation probabiliste des inégalités de la forme $T_\Phi(\theta^*, a)$. Nous montrons qu'il y a équivalence entre

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^* \left(\|\nu - \mu\|_\Phi^* \right) \leq \mathrm{H}\left(\nu \middle| \mu\right).$$

et

$$\forall t > 0, \forall n \in \mathbb{N}^*, \quad \sup_{\varphi \in \Phi} \mathbb{P} \left(\frac{\varphi(X_1) + \cdots + \varphi(X_n)}{n} \geq \langle \mu, \varphi \rangle + t \right) \leq e^{-n\theta^*(t)},$$

avec $(X_k)_{k \geq 1}$ une suite i.i.d de loi μ.
Cette correspondance entre inégalités de type transport et bornes de déviation non asymptotiques permet, par exemple, de retrouver l'inégalité de Pinsker à partir de l'inégalité de Hoeffding, et l'inégalité (i) du théorème VI.25 à partir d'une version de l'inégalité de Bernstein.

Dans la section VI.2.4, nous démontrons une propriété générale de tensorisation des inégalités de transport convexes. Si c_1 est une fonction de coût sur $\mathcal{X}_1$ et c_2 une fonction de coût sur $\mathcal{X}_2$, nous noterons $c_1 \oplus c_2$ la fonction de coût définie sur $\mathcal{X}_1 \times \mathcal{X}_2$ par

$$\forall (x, y) \in (\mathcal{X}_1 \times \mathcal{X}_2)^2, \quad c_1 \oplus c_2(x, y) = c_1(x_1, y_1) + c_2(x_2, y_2).$$

D'une façon assez générale, nous montrerons que si pour $i \in \{1, 2\}$, μ_i est une probabilité sur $\mathcal{X}_i$ vérifiant l'inégalité de transport convexe

$$\forall \nu \in \mathcal{P}(\mathcal{X}_i), \quad \theta_i^*(\mathcal{T}_{c_i}(\nu, \mu_i)) \leq \mathrm{H}\left(\nu \middle| \mu_i\right),$$

alors, la probabilité $\mu_1 \otimes \mu_2$ vérifie

$$\forall \nu \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2), \quad (\theta_1 + \theta_2)^*(\mathcal{T}_{c_1 \oplus c_2}(\nu, \mu_1 \otimes \mu_2)) \leq \mathrm{H}\left(\nu \middle| \mu_1 \otimes \mu_2\right).$$

En particulier, si μ vérifie $T_c(\theta^*, a)$ sur $\mathcal{X}$, alors $\mu^{\otimes n}$ vérifie :

$$\forall \nu \in \mathcal{P}(\mathcal{X}^n), \quad n\theta^* \left(\frac{\mathcal{T}_{\oplus^n c}(\nu, \mu^{\otimes n})}{na} \right) \leq \mathrm{H}\left(\nu \middle| \mu^{\otimes n}\right), \tag{VI.31}$$

en notant

$$\forall x, y \in \mathcal{X}^n, \quad \oplus^n c(x,y) = \sum_{i=1}^{n} c(x_i, y_i).$$

D'après (VI.31), une condition suffisante, pour qu'il y ait *tensorisation avec invariance de la constante* est donc que θ^* soit linéaire (c'est bien sûr le cas pour $\mathbb{T}_2$).

Nous proposerons deux manières de démontrer cette propriété de tensorisation :

- soit de manière directe, en construisant un couplage astucieux de ν sur $\mu_1 \otimes \mu_2$ (le couplage de Marton),
- soit de manière indirecte, en utilisant le critère dual (VI.28).

La première méthode, due à K. Marton, a de loin le plus fort contenu intuitif et théorique. En revanche, elle pose des problèmes de mesurabilité assez délicats. La seconde, due à M. Ledoux, est nettement moins intuitive. Elle est, par contre, beaucoup plus rapide à mettre en oeuvre et permet d'éviter ce problème de mesurabilité.

• Section VI.3 : Applications des I.T.C.
Cette section est consacrée aux liens entre les inégalités de transport convexes associées à un coût métrique ($c = d$) les inégalités de concentration et les inégalités de déviations.

La proposition VI.81 est une version générale de l'argument de Marton. On montre que si μ est une probabilité sur un espace polonais $(\mathcal{X}, d)$ qui vérifie l'inégalité $T_d(\theta^*, a)$, alors pour tout ensemble mesurable $A \subset \mathcal{X}$, tel que $\mu(A) \geq \frac{1}{2}$, on a

$$\mu(A^\varepsilon) \geq 1 - \exp\left(-\frac{1}{a}\theta^*(\varepsilon - r)\right), \tag{VI.32}$$

avec $r = \theta^{*-1}(a\log(2))$, et $A^\varepsilon = \{x \in \mathcal{X} : d(x, A) \leq \varepsilon\}$.

La suite de cette section montre comment la propriété de tensorisation des inégalités de transport associées à un coût métrique permet d'obtenir des inégalités de déviations pour des fonctions de variables aléatoires indépendantes.
Le point de départ est le résultat élémentaire suivant :
Si μ vérifie l'inégalité $T_d(\theta^*, a)$, alors pour toute fonction φ 1-Lipschitzienne, on a

$$\forall t > 0, \quad \mu\left(\varphi \geq \langle \mu, \varphi \rangle + t\right) \leq e^{-\theta^*\left(\frac{t}{a}\right)}. \tag{VI.33}$$

(Voir la proposition VI.83.)
Par tensorisation, on en déduit que si $F : \mathcal{X}^n \to \mathbb{R}$ est une fonction 1-Lipschitzienne pour la distance $\oplus^n d$, alors

$$\forall t > 0, \quad \mathbb{P}\left(F(X_1, \ldots, X_n) \geq \mathbb{E}[F(X_1, \ldots, X_n)] + t\right) \leq e^{-n\theta^*\left(\frac{t}{an}\right)}. \tag{VI.34}$$

En particulier, en appliquant VI.34 à

$$F(x_1, \ldots, x_n) = \sup_{\varphi \in \Phi} \left\{ \varphi(x_1) + \ldots + \varphi(x_n) - n \int_{\mathcal{X}} \varphi \, d\mu \right\},$$

où Φ est un ensemble dénombrable de fonctions 1-Lipschitziennes, on obtient

$$\forall t > 0, \forall n \in \mathbb{N}^*, \quad \mathbb{P}\left(\sup_{\varphi \in \Phi} \langle L_n - \mu, \varphi \rangle \geq \mathbb{E}\left[\sup_{\varphi \in \Phi} \langle L_n - \mu, \varphi \rangle \right] + t \right) \leq e^{-n\theta^*\left(\frac{t}{a}\right)},$$

en notant $L_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$.
Par cette approche, on peut obtenir des versions (un peu moins précises) de résultats comme le théorème de Yurinskii ou des bornes à la Talagrand-Ledoux-Massart pour les processus empiriques.

Remarque VI.35.
Il va sans dire que les résultats de ce chapitre n'ont d'intérêt que si l'on dispose de critères effectifs permettant de démontrer qu'une probabilité μ satisfait une inégalité de transport donnée. Le chapitre suivant est consacré à ce problème. On y démontre notamment des conditions nécessaires et suffisantes pour les inégalités de transport convexes associées à un coût métrique.

VI.2 Inégalités de transport convexes

VI.2.1 Définitions

- Nous noterons $\mathcal{C}$, la classe des fonctions $\theta : \mathbb{R} \to \mathbb{R}^+ \cup \{+\infty\}$, convexes, semi-continues inférieurement, $\theta(0) = 0$, $\operatorname{dom} \theta = [0, a_\theta[$, avec $a_\theta \in]0, +\infty]$. Remarquons que si $\theta \in \mathcal{C}$, alors θ est non bornée sur son domaine.
- Pour $\theta \in \mathcal{C}$, la fonction convexe conjuguée de θ sera notée θ^*, elle est définie par :

$$\forall t \in \mathbb{R}, \quad \theta^*(t) = \sup\{st - \theta(s)\},$$

  θ^* est convexe, positive, s.c.i, et on voit facilement que θ^* est identiquement nulle sur $\mathbb{R}^-$.
- Dans tout ce qui suit, les fonctions de coût sur $\mathcal{X}$ seront toujours supposées *symétriques*, *ie*

$$\forall x, y \in \mathcal{X}, \quad c(x, y) = c(y, x).$$

  Sous cette hypothèse,

$$\forall (\mu, \nu) \in \mathcal{P}(\mathcal{X})^2, \quad \mathcal{T}_c(\mu, \nu) = \mathcal{T}_c(\nu, \mu).$$

Définition VI.36. Soit $\theta \in \mathcal{C}$. Nous dirons que $\mu \in \mathcal{P}(\mathcal{X})$ satisfait l'inégalité de transport convexe (I.T.C) $T_c(\theta^*, a)$, si

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^*\left(\frac{\mathcal{T}_c(\nu,\mu)}{a}\right) \leq \mathrm{H}\left(\nu|\,\mu\right). \tag{VI.37}$$

VI.2.2 Formulation duale des I.T.C

Le théorème suivant généralise le théorème VI.22 de Bobkov et Götze. Il permet d'obtenir, grâce au théorème VI.2, une traduction de (VI.37).

Théorème VI.38. *Soient $(\mathcal{X}, d)$ un espace polonais, $\theta \in \mathcal{C}$, $\mu \in \mathcal{P}(\mathcal{X})$ et c une fonction de coût continue sur $\mathcal{X}$.*
Il y a équivalence entre :

1. *μ satisfait $T_c(\theta^*, a)$,*
2. *Pour tout $(\varphi, \psi) \in \Phi_c$ et tout $s \geq 0$,*

$$\int_{\mathcal{X}} \exp s(\varphi(x) + \langle \mu, \psi \rangle)\, d\mu(x) \leq \exp \theta(as).$$

En particulier, si $c(x, y) = d(x, y)$, il y a équivalence entre :

1. *μ satisfait $T_d(\theta^*, a)$,*
2. *Pour tout $\varphi \in BLip_1(\mathcal{X}, d)$ et tout $s \geq 0$,*

$$\int_{\mathcal{X}} \exp s(\varphi(x) - \langle \mu, \varphi \rangle)\, d\mu(x) \leq \exp \theta(as).$$

Démonstration. D'après la formule de dualité, μ satisfait $T_c(\theta^*, a)$ si, et seulement si,

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^*\left(\frac{1}{a} \sup_{(\varphi,\psi)\in \Phi_c} \left\{\int_{\mathcal{X}} \varphi\, d\nu + \int_{\mathcal{X}} \psi\, d\mu\right\}\right) \leq \mathrm{H}\left(\nu|\,\mu\right).$$

La fonction θ^* étant continue et croissante, ceci équivaut à

$$\forall (\varphi, \psi) \in \Phi_c,\ \forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^*\left(\frac{\int_{\mathcal{X}} \varphi\, d\nu + \int_{\mathcal{X}} \psi\, d\mu}{a}\right) \leq \mathrm{H}\left(\nu|\,\mu\right),$$

soit, pour tout $(\varphi, \psi) \in \Phi_c$,

$$\forall t \in \mathbb{R}, \quad \theta^*(t) \leq \inf\left\{\mathrm{H}\left(\nu|\,\mu\right) : \nu \in \mathcal{P}(\mathcal{X}), \int_{\mathcal{X}} \varphi\, d\nu + \int_{\mathcal{X}} \psi\, d\mu = at\right\}$$

Soit $(X_i)_i$ une suite i.i.d de loi μ ; posons $L_n = \dfrac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$. D'après le théorème de Sanov, $(L_n)_n$ suit un P.G.D sur $\mathcal{P}(\mathcal{X})$ muni de la τ-topologie de bonne fonction de taux

$\mathrm{H}\,(\,.\,|\,\mu)$. La fonction φ étant bornée, l'application : $\mathcal{P}(\mathcal{X}) \to \mathbb{R} : \nu \mapsto \int_{\mathcal{X}} \varphi\, d\nu$ est continue. D'après le principe de contraction, $\displaystyle\int_{\mathcal{X}} \varphi\, dL_n = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i)$ suit un P.G.D de bonne fonction de taux

$$\mathrm{I}(t) = \inf\left\{\mathrm{H}\,(\nu|\,\mu) : \nu \in \mathcal{P}(\mathcal{X}), \int_{\mathcal{X}} \varphi\, d\nu = t\right\}.$$

Or, d'après le théorème de Cramér, $\displaystyle\frac{1}{n}\sum_{i=1}^{n} \varphi(X_i)$ suit un P.G.D de bonne fonction de taux Λ^*_φ définie par

$$\Lambda^*_\varphi(t) = \sup_{s\in\mathbb{R}} \{ts - \Lambda_\varphi(s)\},$$

avec

$$\Lambda_\varphi(s) = \log\left(\int_{\mathcal{X}} e^{s\varphi(x)}\, d\mu(x)\right).$$

Par conséquent, par unicité de la fonction de taux, $\mathrm{I}(t) = \Lambda^*_\varphi(t)$. En particulier,

$$\inf\left\{\mathrm{H}\,(\nu|\,\mu) : \nu \in \mathcal{P}(\mathcal{X}), \int_{\mathcal{X}} \varphi\, d\nu + \int_{\mathcal{X}} \psi\, d\mu = at\right\} = \Lambda^*_\varphi\left(at - \int_{\mathcal{X}} \psi\, d\mu\right).$$

Ainsi μ satisfait $T_c(\theta^*, a)$ si, et seulement si, pour tout $(\varphi, \psi) \in \Phi_c$,

$$\forall t \in \mathbb{R}, \quad \theta^*(t) \le \Lambda^*_\varphi\left(at - \int_{\mathcal{X}} \psi\, d\mu\right), \tag{VI.39}$$

ce qui équivaut à

$$\forall s \in \mathbb{R}, \quad \theta(as) \ge \Lambda_\varphi(s) + s\int_{\mathcal{X}} \psi\, d\mu$$

soit

$$\forall s \in \mathbb{R}, \quad \int_{\mathcal{X}} \exp s(\varphi(x) + \langle \mu, \psi\rangle)\, d\mu(x) \le \exp\theta(as)$$

et comme $\theta(s) = +\infty$ pour $s < 0$, on obtient le résultat. $\square$

Remarque VI.40.

Pour démontrer le théorème VI.38, il est également possible de reprendre la preuve originale du théorème 3.1 de [4].

Nous étudierons plus particulièrement le cas d'un coût métrique sur un espace polonais, cas pour lequel on dispose de la formule :

$$\forall \nu, \mu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_d(\nu, \mu) = \sup_{\varphi\in \mathrm{BLip}_1(\mathcal{X},d)} \left\{\int_{\mathcal{X}} \varphi\, d\mu - \int_{\mathcal{X}} \varphi\, d\nu\right\}.$$

Si maintenant Φ désigne une classe quelconque de fonctions mesurables bornées sur un espace mesurable $\mathcal{X}$ quelconque, telle que

$$\varphi \in \Phi \Rightarrow -\varphi \in \Phi, \tag{VI.41}$$

alors, en posant

$$\|\mu - \nu\|_\Phi^* = \sup_{\varphi \in \Phi} \left\{ \int_{\mathcal{X}} \varphi \, d\mu - \int_{\mathcal{X}} \varphi \, d\nu \right\},$$

on obtient une classe plus générale de fonctionnelles sur $\mathcal{P}(\mathcal{X})^2$ englobant en particulier les $\mathcal{T}_d(\, . \, , \, . \,)$. Une classe de fonction Φ vérifiant (VI.41) sera dite *symétrique*.

Pour les fonctionnelles $\| \, . \, \|_\Phi^*$, on a la

Proposition VI.42. *Soit $\theta \in \mathcal{C}$, $\mu \in \mathcal{P}(\mathcal{X})$. Il y a équivalence entre :*

1. *μ satisfait $T_\Phi(\theta^*, a)$, ie*

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^* \left(\frac{\|\nu - \mu\|_\Phi^*}{a} \right) \leq \mathrm{H}\left(\nu | \, \mu\right)$$

2. *Pour toute $\varphi \in \Phi$ et tout $s \geq 0$,*

$$\int_{\mathcal{X}} \exp s(\varphi(x) - \langle \mu, \varphi \rangle) \, d\mu(x) \leq \exp \theta(as).$$

Démonstration. Idem. □

VI.2.3 Quelques exemples

Dans cette sous-section, nous allons voir comment utiliser le critère dual pour retrouver certaines I.T.C bien connues.

Inégalité de Pinsker

La preuve de l'inégalité de Pinsker que nous allons donner est issue de [49] . Le lemme suivant porte le nom de lemme d'Hoeffding :

Lemme VI.43. *Si X est une variable aléatoire à valeurs dans $[a, b]$, alors*

$$\forall s \geq 0, \quad \mathbb{E}\left[e^{sX}\right] \leq e^{s\mathbb{E}[X] + \frac{s^2(b-a)^2}{8}}. \tag{VI.44}$$

Démonstration. Posons $\Lambda(s) = \log \mathbb{E}\left[e^{sX}\right]$. Il est clair que, $\Lambda(0) = 0$ et $\Lambda'(0) = \mathbb{E}[X]$. De plus, si μ désigne la loi de X, on voit facilement que $\Lambda''(s)$ est la variance de la probabilité μ_s définie par :

$$\frac{d\mu_s}{d\mu}(x) = \frac{\exp(sx)}{\Lambda(s)}.$$

Or, si Y est une variable aléatoire à valeurs dans $[a, b]$, on a $\left|Y - \frac{a+b}{2}\right| \leq \frac{(b-a)}{2}$, donc

$$\text{Var}(Y) = \inf_a \mathbb{E}[(Y-a)^2] \leq \mathbb{E}\left[\left(Y - \frac{a+b}{2}\right)^2\right] \leq \frac{(b-a)^2}{4}.$$

Comme μ_s a son support dans $[a, b]$, on en déduit que $\Lambda''(s) \leq \frac{(b-a)^2}{4}$. □

Corollaire VI.45. *Si μ et ν sont des probabilités sur un espace mesurable $\mathcal{X}$, on a*

$$\frac{1}{2}\|\mu - \nu\|_{VT}^2 \leq \mathrm{H}\left(\nu \mid \mu\right) \tag{VI.46}$$

Démonstration. Remarquons que $\|\mu - \nu\|_{VT} = \|\mu - \nu\|_{B_1(\mathcal{X})}^*$, avec $B_1(\mathcal{X})$ l'ensemble des fonctions mesurables φ telles que $|\varphi| \leq 1$. Or, d'après le lemme VI.43, pour toute $\varphi \in B_1(\mathcal{X})$, on a pour tout $s \geq 0$,

$$\int_{\mathcal{X}} \exp s(\varphi - \langle \mu, \varphi \rangle)\, d\mu \leq \exp \frac{s^2}{2},$$

ce qui entraîne (VI.46), d'après la proposition VI.42. □

Remarque VI.47.

On voit dans cette preuve que l'inégalité de Pinsker (VI.46), et l'inégalité de Hoeffding :

$$\mathbb{P}\left(\frac{Y_1 + \cdots + Y_n}{n} \geq t\right) \leq e^{-nt^2/2},$$

valable pour toute suite Y_i de variables aléatoires indépendantes centrées et à valeurs dans un segment de longueur 2, reposent toutes deux sur le lemme VI.43. Il y a en fait un lien général entre les I.T.C et les bornes de déviations exactes, comme le montre la proposition suivante.

Un lien général entre I.T.C et inégalités de déviations

Proposition VI.48. *Soit Φ une classe symétrique de fonctions mesurables bornées sur un espace mesurable $\mathcal{X}$. Les trois propositions suivantes sont équivalentes :*

1. $\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^* \left(\|\mu - \nu\|_\Phi^*\right) \leq \mathrm{H}\left(\nu | \, \mu\right),$
2. $\forall \varphi \in \Phi, \ \forall s \geq 0, \qquad \displaystyle\int_{\mathcal{X}} \exp s(\varphi - \langle \mu, \varphi \rangle) \, d\mu \leq \exp \theta(s),$
3. $\forall \varphi \in \Phi, \ \forall n \geq 1, \ \forall t \geq 0, \quad \mathbb{P}\left(\dfrac{\varphi(X_1) + \cdots + \varphi(X_n)}{n} \geq \langle \mu, \varphi \rangle + t \right) \leq e^{-n\theta^*(t)},$
   avec $(X_k)_{k\geq 1}$ une suite i.i.d de loi μ.

Démonstration. On a déjà vu dans la proposition VI.42 que les propositions (1) et (2) étaient équivalentes.
Montrons l'équivalence de (2) et (3). Tout d'abord, d'après l'inégalité de Chernoff classique, on a, pour tout n et tout $t \geq 0$,

$$\mathbb{P}\left(\frac{\varphi(X_1) + \cdots + \varphi(X_n)}{n} \geq \langle \mu, \varphi \rangle + t \right) \leq e^{-n\gamma^*(t + \langle \mu, \varphi \rangle)},$$

où γ^* est la transformée de Cramér de $\varphi(X)$, X de loi μ. Or (2) entraîne immédiatement que

$$\forall t \geq 0, \quad \theta^*(t) \leq \gamma^*(t + \langle \mu, \varphi \rangle).$$

Par conséquent, (2) implique (3).
Réciproquement, d'après la borne inférieure du théorème de Cramér, (3) entraîne que :

$$\forall t \geq 0, \quad -\inf\{\gamma^*(u), u \in]\langle \mu, \varphi \rangle + t, +\infty[\} \leq -\theta^*(t),$$

donc, si $\langle \mu, \varphi \rangle + t \in \overset{\circ}{\mathrm{dom}}\, \gamma^*$, γ^* étant croissante sur $]\langle \mu, \varphi \rangle, +\infty[$, on a

$$\gamma^*(\langle \mu, \varphi \rangle + t) \geq \theta^*(t),$$

inégalité qui reste vraie pour tout $t \geq 0$, à cause du caractère s.c.i des deux fonctions. Enfin la propriété

$$\forall t \geq 0, \quad \theta^*(t) \leq \gamma^*(\langle \mu, \varphi \rangle + t)$$

entraîne facilement (2) par conjugaison convexe. □

Remarque VI.49.

Cette proposition établit un pont entre les I.T.C et certaines bornes exactes de déviations. La propriété de tensorisation des I.T.C développée dans la section VI.2.4 va nous permettre d'établir des bornes exactes de déviations pour une plus grande classe d'objets. Avant cela, nous allons voir comment la généralisation de l'inégalité de Pinsker (VII.10) proposée par F. Bolley et C. Villani peut se retrouver à partir d'une version de l'inégalité de Bernstein.

Inégalité de Pinsker pondérée et inégalité de Bernstein

Dans [5], F. Bolley et C. Villani, ont obtenu, par des moyens purement analytiques, une version pondérée de l'inégalité de Pinsker :

Proposition VI.50. *Soit χ une fonction mesurable positive sur un espace de mesurable $\mathcal{X}$. Si $\mu, \nu \in \mathcal{P}(\mathcal{X})$ sont telles que $\int_{\mathcal{X}} \chi\, d\mu < +\infty$ et $\int_{\mathcal{X}} \chi\, d\nu < +\infty$, alors*

$$\|\chi\nu - \chi\mu\|_{VT} \leq \left(\frac{3}{2} + \log \int_{\mathcal{X}} e^{2\chi}\, d\mu\right) \left(\sqrt{\mathrm{H}\left(\nu|\,\mu\right)} + \frac{1}{2}\, \mathrm{H}\left(\nu|\,\mu\right)\right) \tag{VI.51}$$

A l'instar de l'inégalité de Pinsker qui était une traduction de l'inégalité de Hoeffding, nous allons voir que (VI.51) est une traduction (à un facteur numérique près) de la version suivante de l'inégalité de Bernstein.

Proposition VI.52.

1. *Si X une variable aléatoire réelle centrée et $M = \inf\left\{\lambda > 0 : \mathbb{E}\left[e^{\frac{|X|}{\lambda}}\right] \leq 2\right\}$, alors*

$$\forall s \in \mathbb{R}, \quad \mathbb{E}\left[e^{sX}\right] \leq e^{\theta_1(Ms)}, \qquad \textit{avec} \qquad \theta_1(s) = \begin{cases} \frac{s^2}{1-s} & \textit{si } s \in [0,1[\\ +\infty & \textit{sinon} \end{cases} \tag{VI.53}$$

2. *En particulier, si $X_1, \ldots, X_n$ sont des variables réelles indépendantes centrées, en posant $M = \inf\left\{\lambda > 0 : \forall i = 1 \ldots n,\ \mathbb{E}\left[e^{\frac{|X_i|}{\lambda}}\right] \leq 2\right\}$, on a*

$$\forall t \geq 0, \quad \mathbb{P}(X_1 + \cdots + X_n \geq nt) \leq e^{-n\left(\sqrt{1+t/M}-1\right)^2} \tag{VI.54}$$

Démonstration.
(1) Par définition de M, on a

$$1 \geq \mathbb{E}\left[e^{\frac{|X|}{M}} - 1\right] = \sum_{k=1}^{+\infty} \frac{\mathbb{E}\left[|X|^k\right]}{k! M^k}.$$

Donc, pour tout $k \geq 2$, $\dfrac{\mathbb{E}[|X|^k]}{k!} \leq M^k$. Par conséquent, pour tout $s \in \left[0, \frac{1}{M}\right[$,

$$\begin{aligned} \mathbb{E}\left[e^{sX}\right] &= 1 + \sum_{k=2}^{+\infty} s^k \frac{\mathbb{E}\left[X^k\right]}{k!} \leq 1 + \sum_{k=2}^{+\infty} s^k \frac{\mathbb{E}\left[|X|^k\right]}{k!} \\ &\leq 1 + \sum_{k=2}^{+\infty} s^k M^k = 1 + \frac{(sM)^2}{1 - sM} \leq e^{\theta_1(sM)}. \end{aligned}$$

(2) On déduit du premier point que $\mathbb{E}\left[e^{s(X_1+\cdots+X_n)}\right] \leq e^{n\theta_1(sM)}$. Le résultat en découle facilement en calculant : $\theta_1^*(t) = \begin{cases} \left(\sqrt{1+t} - 1\right)^2 & \text{si } t \in \mathbb{R}^+ \\ 0 & \text{sinon} \end{cases}$ $\square$

Remarque VI.55.

L'inégalité (VI.54) n'est pas la véritable inégalité de Bernstein. La forme habituelle de cette inégalité est donnée dans le théorème suivant

Théorème VI.56. *Si $X_1, \ldots, X_n$ sont des variables aléatoires réelles indépendantes centrées, telles qu'il existe $M > 0$ et $v_1, \ldots, v_n > 0$ tels que*

$$\mathbb{E}\left[|X_i|^m\right] \leq \frac{m!}{2} M^{m-2} v_i, \tag{VI.57}$$

alors, pour tout $t > 0$,

$$\mathbb{P}(X_1 + \cdots + X_n \geq t) \leq e^{-\frac{1}{2}\frac{t^2}{v+tM}}, \quad \textit{avec} \quad v = v_1 + \cdots + v_n. \tag{VI.58}$$

Si les variables X_i ne sont pas bornées, l'hypothèse (VI.57) n'est pas évidente à vérifier. Une condition suffisante plus tractable est la condition de type Orlicz suivante

$$\mathbb{E}\left[e^{|Y_i|/M} - 1 - \frac{|Y_i|}{M}\right] M^2 \leq \frac{1}{2} v_i. \tag{VI.59}$$

En affaiblissant encore (VI.59), on obtient l'inégalité de la proposition VI.52, ou aucun terme de variance n'apparaît. D'une manière générale, nous ne serons pas en mesure d'inclure des termes de variance dans nos inégalités.

Introduisons l'espace d'Orlicz $\mathbb{L}_\rho(\mathcal{X}, \mu)$ associé à la fonction de Young $\rho(t) = e^{|t|} - 1$ et munissons le de sa norme de jauge $\|\,.\,\|_\rho$ (voir p. 65).

Proposition VI.60. *Soit Φ une classe symétrique de fonctions mesurables bornées sur un espace de probabilité $(\mathcal{X}, \mu)$. Si $\tilde{\Phi} = \{\varphi - \langle \mu : \varphi \rangle, \varphi \in \Phi\}$ est une partie bornée de $\mathbb{L}_\rho(\mathcal{X}, \mu)$, alors μ vérifie $T_\Phi(\theta_1^*, M)$, avec $M = \sup_{\varphi \in \Phi} \|\varphi - \langle \mu, \varphi \rangle\|_\rho$.*

Autrement dit,

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \|\nu - \mu\|_\Phi^* \leq 2M \left(\sqrt{\mathrm{H}(\nu|\,\mu)} + \frac{1}{2} \mathrm{H}(\nu|\,\mu) \right).$$

Démonstration. D'après l'inégalité (VI.53), pour toute $\varphi \in \Phi$ on a :

$$\forall s \in \mathbb{R}, \quad \int_{\mathcal{X}} \exp s(\varphi - \langle \mu, \varphi \rangle)\, d\mu \leq \exp \theta_1(Ms),$$

donc, d'après la proposition VI.42,

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta_1^* \left(\frac{\|\nu - \mu\|_\Phi^*}{M} \right) \leq \mathrm{H}(\nu|\,\mu)$$

Comme $\theta_1^{*-1}(x) = 2\sqrt{x} + x$, on a de manière équivalente

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \|\nu - \mu\|_\Phi^* \leq 2M \left(\sqrt{\mathrm{H}\left(\nu|\, \mu\right)} + \frac{1}{2}\,\mathrm{H}\left(\nu|\, \mu\right) \right)$$

□

Corollaire VI.61. *Si $d(\,.\,,\,.\,)$ est une distance mesurable sur un espace mesurable $\mathcal{X}$ et $\mu \in \mathcal{P}(\mathcal{X})$ telle que*

$$\exists \delta > 0, \quad \iint_{\mathcal{X}^2} e^{\delta d(x,y)}\, d\mu(x)\, d\mu(y) < +\infty,$$

alors, en posant $M = \|d(\,.\,,\,.\,)\|_{\mathbb{L}_\rho(\mathcal{X}^2,\,\mu^{\otimes 2})}$, on a en notant $BLip_1(\mathcal{X}, d)$ l'ensemble fonctions mesurables bornées 1-Lipschitziennes pour d

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \|\nu - \mu\|_{BLip_1(\mathcal{X},d)}^* \leq 2M \left(\sqrt{\mathrm{H}\left(\nu|\, \mu\right)} + \frac{1}{2}\,\mathrm{H}\left(\nu|\, \mu\right) \right) \qquad \text{(VI.62)}$$

Démonstration. Remarquons que pour tout $\lambda > 0$, on a pour toute $\varphi \in \mathrm{BLip}_1(\mathcal{X}, d)$

$$\begin{aligned} \int_{\mathcal{X}} \rho\left(\frac{|\varphi - \langle \mu, \varphi\rangle|}{\lambda}\right) d\mu &\overset{(*)}{\leq} \iint_{\mathcal{X}^2} \rho\left(\frac{|\varphi(x) - \varphi(y)|}{\lambda}\right) d\mu(x)\, d\mu(y) \\ &\leq \iint_{\mathcal{X}^2} \rho\left(\frac{d(x,y)}{\lambda}\right) d\mu(x)\, d\mu(y), \end{aligned}$$

$(*)$ venant de l'inégalité de Jensen. Ainsi $\sup\limits_{\varphi \in \mathrm{BLip}_1(\mathcal{X},d)} \|\varphi - \langle \mu, \varphi\rangle\|_\rho \leq M$ et le résultat découle de la proposition VI.60. □

Remarque VI.63.

Nous verrons à la section VII.4.1 du prochain chapitre que

$$\|d(\,.\,,\,.\,)\|_{\mathbb{L}_\rho(\mathcal{X}^2,\,\mu^{\otimes 2})} \leq 1 + \frac{\log \iint_{\mathcal{X}^2} e^{d(x,y)}\, d\mu(x)\, d\mu(y)}{\log(2)}.$$

En particulier, pour $d = d_\chi$, on obtient, sous les hypothèses de la proposition VI.50

$$\begin{aligned} &\forall \nu \in \mathcal{P}(\mathcal{X}), \\ &\qquad \|\chi\nu - \chi\mu\|_{VT} \leq 2\left(1 + \frac{2\log \int_{\mathcal{X}} e^{\chi}\, d\mu}{\log(2)}\right) \left(\sqrt{\mathrm{H}\left(\nu|\, \mu\right)} + \frac{1}{2}\,\mathrm{H}\left(\nu|\, \mu\right) \right), \end{aligned}$$

inégalité qui ne diffère de (VI.51) que par des facteurs numériques.

VI.2.4 Tensorisation des I.T.C

Dans cette sous-section, nous chercherons à répondre à la question suivante : si μ_1 et μ_2 sont deux probabilités satisfaisant chacune une I.T.C, quelle I.T.C vérifie la mesure produit $\mu_1 \otimes \mu_2$?

Introduisons quelques notations :

- Si $c_1, \ldots, c_n$ sont des fonctions de coût définies sur respectivement sur des espaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$, nous noterons $\oplus_{i=1}^n c_i$ ou plus rapidement $\oplus_i c_i$, la fonction de coût définie sur $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ par

$$\forall (x,y) \in (\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)^2, \quad \oplus_{i=1}^n c_i(x,y) = \sum_{i=1}^n c_i(x_i, y_i).$$

- Si $f_1, \ldots, f_n$ sont des fonctions convexes s.c.i définies sur $\mathbb{R}$, leur inf-convolution est la fonction notée $f_1 \square f_2 \cdots \square f_n$ ou encore $\square_i f_i$, et définie pour tout $x \in \mathbb{R}$ par

$$f_1 \square f_2 \cdots \square f_n(x) = \inf\{f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) : x = x_1 + x_2 + \cdots + x_n\}$$

Théorème VI.64. *Si pour tout $i = 1 \ldots n$, μ_i est une probabilité sur un espace polonais $\mathcal{X}_i$ satisfaisant l'I.T.C*

$$\forall \nu \in \mathcal{P}(\mathcal{X}_i), \quad \theta_i^* \left(\mathcal{T}_{c_i}(\nu, \mu_i)\right) \leq \mathrm{H}\left(\nu | \, \mu_i\right),$$

avec pour tout i, c_i une fonction de coût continue symétrique sur $\mathcal{X}_i$ telle que

$$\forall x_i \in \mathcal{X}_i, \quad c_i(x_i, x_i) = 0$$

et $\theta_i \in \mathcal{C}$, alors $\mu_1 \otimes \cdots \otimes \mu_n \in \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$ satisfait l'I.T.C

$$\forall \nu \in \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n), \quad \theta_1^* \square \theta_2^* \cdots \square \theta_n^* \left[\mathcal{T}_{\oplus_i c_i}(\nu, \otimes_i \mu_i)\right] \leq \mathrm{H}\left(\nu | \otimes_i \mu_i\right). \tag{VI.65}$$

Nous donnerons deux preuves de ce résultat. La première, qui utilise un argument de couplage dû à K. Marton, est la plus satisfaisante d'un point de vue théorique, mais elle pose un problème de mesurabilité peu évident sur lequel nous reviendrons. La seconde, qui utilise la version duale des I.T.C donnée par le théorème VI.38 généralise un argument de M. Ledoux.

Preuve par couplage :
Nous nous restreindrons au cas $\mathcal{X}_1 = \cdots = \mathcal{X}_n = \mathbb{R}$.

Si $\nu \in \mathcal{P}(\mathbb{R}^n)$, nous noterons ν_1 sa marginale sur $\mathbb{R}^{n-1}$ et $y \mapsto \nu_2(\,.\,|y)$ désignera un noyau de transition de $\mathbb{R}^{n-1}$ dans $\mathbb{R}$ tel que

$$\nu(dx) = \nu_2(dx_n|x_1,\ldots,x_{n-1})\nu_1(dx_1,\ldots,dx_{n-1}).$$

Autrement dit, si $X = (X_1,\ldots,X_n)$ est de loi ν, alors ν_1 est la loi de $(X_1,\ldots,X_{n-1})$ et $\nu_2(\,.\,|y)$ est une version régulière de la loi conditionnelle de X_n sachant $(X_1,\ldots,X_{n-1})$.

On a alors les propositions suivantes :

Proposition VI.66. *Si $\nu \in \mathcal{P}(\mathbb{R}^n)$, $\alpha_1 \in \mathcal{P}(\mathbb{R}^{n-1}) et \alpha_2 \in \mathcal{P}(\mathbb{R})$, alors*

$$\mathrm{H}\left(\nu|\,\alpha_1\otimes\alpha_2\right) = \mathrm{H}\left(\nu_1|\,\alpha_1\right) + \int_{\mathbb{R}} \mathrm{H}\left(\nu_2(\,.\,|y)|\,\alpha_2\right)\, d\nu_1(y). \tag{VI.67}$$

Démonstration. Voir par exemple la preuve du théorème D.13 de [26]. □

Proposition VI.68. *Si c_1 est une fonction de coût sur $\mathbb{R}^{n-1}$ et c_2 une fonction de coût sur $\mathbb{R}$ de la forme $c_2(x,y) = q(x-y)$, avec $q : \mathbb{R} \to \mathbb{R}^+$ une fonction convexe paire, alors, pour toute $\nu \in \mathcal{P}(\mathbb{R}^n)$, $\alpha_1 \in \mathcal{P}(\mathbb{R}^{n-1})$, $\alpha_2 \in \mathcal{P}(\mathbb{R})$, on a*

$$\mathcal{T}_{c_1\oplus c_2}(\nu,\alpha_1\otimes\alpha_2) \leq \mathcal{T}_{c_1}(\nu_1,\alpha_1) + \int_{\mathbb{R}} \mathcal{T}_{c_2}(\nu_2(\,.\,|y),\alpha_2)\, d\nu_1(y) \tag{VI.69}$$

Démonstration. Pour tout $y \in \mathbb{R}^{n-1}$, soit π_2^y la probabilité sur $\mathbb{R}^2$ ayant pour fonction de répartition

$$H^y(s,t) = \min\left\{\alpha_2(]-\infty,s]),\nu_2(]-\infty,t]|y)\right\}.$$

D'après le théorème VI.1,

$$\pi_2^y \in \Pi(\alpha_2,\nu_2(\,.\,|y)) \qquad \text{et} \qquad \mathcal{T}_{c_2}(\alpha_2,\nu_2(\,.\,|y)) = \int_{\mathbb{R}^2} c_2\, d\pi_2^y.$$

Comme pour tout $t \in \mathbb{R}$, $y \mapsto \nu_2(]-\infty,t]|y)$ est mesurable, on en déduit que pour tout $(s,t) \in \mathbb{R}^2$, la fonction

$$y \mapsto \pi_2^y(]-\infty,s]\times]-\infty,t]) \quad (= H^y(s,t))$$

est mesurable. Par un argument de classe monotone, on en déduit que pour tout A Borélien de $\mathbb{R}^2$, la fonction

$$y \mapsto \pi_2^y(A)$$

est mesurable. Pour tout $\pi_1 \in \Pi(\alpha_1, \nu_1)$, on peut donc définir une probabilité π sur $\mathbb{R}^n \times \mathbb{R}^n = (\mathbb{R}^{n-1} \times \mathbb{R})^2$ par

$$\int f\, d\pi = \int f(x_1, x_2, x_3, x_4)\, d\pi_2^{x_3}(x_2, x_4)\, d\pi_1(x_1, x_3).$$

Clairement, $\pi \in \Pi(\alpha_1 \otimes \alpha_2, \nu)$. De plus,

$$\begin{aligned}\int c_1 \oplus c_2\, d\pi &= \int c_1(x_1, x_3)\, d\pi_2^{x_3}(x_2, x_4)\, d\pi_1(x_1, x_3)\\ &\quad + \int c_2(x_2, x_4)\, d\pi_2^{x_3}(x_2, x_4)\, d\pi_1(x_1, x_3)\\ &= \int c_1\, d\pi_1 + \int \mathcal{T}_{c_2}(\nu_2(\,.\,|x_3), \alpha_2)\, d\pi_1(x_1, x_3)\\ &= \int c_1\, d\pi_1 + \int \mathcal{T}_{c_2}(\nu_2(\,.\,|x_3), \alpha_2)\, d\nu_1(x_3)\end{aligned}$$

On en déduit que pour tout $\pi_1 \in \Pi(\alpha_1, \nu_1)$,

$$\mathcal{T}_{c_1 \oplus c_2}(\nu, \alpha_1 \otimes \alpha_2) \leq \int c_1\, d\pi_1 + \int \mathcal{T}_{c_2}(\nu_2(\,.\,|x), \alpha_2)\, d\nu_1(x),$$

d'où le résultat en optimisant en π_1. $\square$

Remarque VI.70.

La même preuve fonctionne sur des espaces plus généraux s'il existe un noyau de transition $y \mapsto \pi_2^y$ de $\mathcal{X}_1 \times \cdots \times \mathcal{X}_{n-1}$ dans $\mathcal{X}_n$ tel que pour tout y, $\pi_2^y \in \Pi(\alpha_2, \nu_2(\,.\,|y))$ et $\mathcal{T}_{c_2}(\alpha_2, \nu_2(\,.\,|y)) = \int_{\mathcal{X}_n} c_2\, d\pi_2^y$. C'est le cas en particulier, si $c_i = d_{\mathcal{X}_i}$, comme nous le verrons à la proposition VI.73.

Proposition VI.71. *Si pour tout $i = 1 \ldots n$, μ_i est une probabilité sur $\mathbb{R}$ satisfaisant l'I.T.C*

$$\forall \nu \in \mathcal{P}(\mathbb{R}), \quad \theta_i^* \left(\mathcal{T}_{c_i}(\nu, \mu_i)\right) \leq \mathrm{H}\left(\nu |\, \mu_i\right),$$

avec pour tout i, c_i une fonction de coût de la forme $c_i(x, y) = q_i(x - y)$ avec q_i une fonction convexe positive paire, alors $\mu_1 \otimes \cdots \otimes \mu_n \in \mathcal{P}(\mathbb{R}^n)$ satisfait l'I.T.C

$$\forall \nu \in \mathcal{P}(\mathbb{R}^n), \quad \theta_1^* \Box \theta_2^* \cdots \Box \theta_n^* \left[\mathcal{T}_{\oplus_i c_i}(\nu, \otimes_i \mu_i)\right] \leq \mathrm{H}\left(\nu |\, \otimes_i \mu_i\right). \tag{VI.72}$$

Démonstration. Par récurrence sur n.
Posons $c_0 = \oplus_{i=1}^{n-1} c_i$ qui est une fonction de coût sur $\mathbb{R}^{n-1}$, $\alpha_1 = \mu_1 \otimes \cdots \otimes \mu_{n-1} \in \mathcal{P}(\mathbb{R}^{n-1})$ et $\theta_0^* = \theta_1^* \Box \theta_2^* \cdots \Box \theta_{n-1}^*$. Supposons que

$$\forall \nu \in \mathcal{P}(\mathbb{R}^{n-1}), \quad \theta_0^* \left[\mathcal{T}_{c_0}(\nu, \alpha_1)\right] \leq \mathrm{H}\left(\nu |\, \alpha_1\right).$$

Soit $\nu \in \mathcal{P}(\mathbb{R}^n)$; définissons ν_1 et $\nu_2(\,.\,|y)$ comme précédemment. D'après l'inégalité (VI.69), on a

$$\mathcal{T}_{c0\oplus c_n}(\nu, \alpha_1 \otimes \mu_n) \leq \mathcal{T}_{c_0}(\nu_1, \alpha_1) + \int \mathcal{T}_{c_n}(\nu_2(\,.\,|x), \mu_n)\, d\nu_1(x).$$

Donc

$$\begin{aligned}
\theta_0^* \square \theta_n^* \left(\mathcal{T}_{c0\oplus c_n}(\nu, \alpha_1 \otimes \mu_n)\right) &\overset{(i)}{\leq} \theta_0^* \square \theta_n^* \left(\mathcal{T}_{c_0}(\nu_1, \alpha_1) + \int \mathcal{T}_{c_n}(\nu_2(\,.\,|x), \mu_n)\, d\nu_1(x)\right) \\
&\overset{(ii)}{\leq} \theta_0^* \left(\mathcal{T}_{c_0}(\nu_1, \alpha_1)\right) + \theta_n^* \left(\int \mathcal{T}_{c_n}(\nu_2(\,.\,|x), \mu_n)\, d\nu_1(x)\right) \\
&\overset{(iii)}{\leq} \theta_0^* \left(\mathcal{T}_{c_0}(\nu_1, \alpha_1)\right) + \int \theta_n^* \left(\mathcal{T}_{c_n}(\nu_2(\,.\,|x), \mu_n)\right)\, d\nu_1(x) \\
&\overset{(iv)}{\leq} \mathrm{H}\left(\nu_1 |\, \alpha_1\right) + \int \mathrm{H}\left(\nu_2(\,.\,|x) |\, \mu_n\right)\, d\nu_1(x) \\
&\overset{(v)}{=} \mathrm{H}\left(\nu |\, \mu_1 \otimes \cdots \otimes \mu_n\right),
\end{aligned}$$

où (i) vient de la croissance de $\theta_0^* \square \theta_n^*$, (ii) de la définition de l'inf-convolution, (iii) de l'inégalité de Jensen, (iv) de l'hypothèse de récurrence et de l'I.T.C satisfaite par μ_n et (v) de la formule (VI.67). $\square$

Comme nous l'avons annoncé plus haut, la preuve précédente reste valable pour la tensorisation des coûts $\mathcal{T}_{d_\chi}$:

Proposition VI.73. *Si pour tout $i = 1 \ldots n$, μ_i est une probabilité sur un espace mesurable $\mathcal{X}_i$ satisfaisant l'I.T.C*

$$\forall \nu \in \mathcal{P}(\mathbb{R}), \quad \theta_i^* \left(\mathcal{T}_{d_{\chi_i}}(\nu, \mu_i)\right) \leq \mathrm{H}\left(\nu |\, \mu_i\right),$$

avec pour tout i, χ_i une fonction mesurable positive et $\theta_i \in \mathcal{C}$, alors la probabilité $\mu_1 \otimes \cdots \otimes \mu_n \in \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$ satisfait l'I.T.C

$$\forall \nu \in \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n), \quad \theta_1^* \square \theta_2^* \cdots \square \theta_n^* \left[\mathcal{T}_{\oplus_i d_{\chi_i}}(\nu, \otimes_i \mu_i)\right] \leq \mathrm{H}\left(\nu |\, \otimes_i \mu_i\right). \quad \text{(VI.74)}$$

Démonstration. Clairement, il suffit de montrer que si $(\mathcal{X}_1, \alpha_1)$, $(\mathcal{X}_2, \alpha_2)$ sont des espaces de probabilité, $c_1(.,.)$ est une fonction de coût mesurable sur $\mathcal{X}_1 \times \mathcal{X}_1$ et $\chi : \mathcal{X}_2 \to \mathbb{R}^+$ est une fonction mesurable, alors pour toute $\nu \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$, avec $\nu \ll \alpha_1 \otimes \alpha_2$,

$$\mathcal{T}_{c_1 \oplus d_\chi}(\nu, \alpha_1 \otimes \alpha_2) \leq \mathcal{T}_{c_1}(\nu_1, \alpha_1) + \int_{\mathcal{X}_1} \mathcal{T}_{d_\chi}(\nu_2(\,.\,|x_1), \alpha_2)\, d\nu_1(x_1), \quad \text{(VI.75)}$$

avec $\nu(dx_1, dx_2) = h_1(x_1)h_2(x_2|x_1)\alpha_1(dx_1)\alpha_2(dx_2)$ et $\nu_1 = h_1.\alpha_1$, $\nu_2(\,.\,|x_1) = h_2(\,.\,|x_1).\alpha_2$. Or, en se reportant à la preuve de la proposition VI.7, on sait que

$$\mathcal{T}_{d_\chi}(\nu_2(\,.\,|x_1), \alpha_2) = \iint_{\mathcal{X}_2^2} d_\chi(s,t)\, d\pi_2^{x_1}(s,t),$$

avec $\pi_2^{x_1}$ défini par

$$\iint f(s,t)\,d\pi_2^{x_1}(s,t) = \int_{\mathcal{X}} f(s,s)\,d(\alpha_2 \wedge \nu_2(\,.\,|x_1))(s) + \frac{1}{m(x_1)} \iint_{\mathcal{X}_2^2} f(s,t)\,d(\alpha_2 - \nu_2(\,.\,|x_1))_+(s)\,d(\alpha_2 - \nu_2(\,.\,|x_1))_-(t),$$

avec $m(x_1) = (\alpha_2 - \nu_2(\,.\,|x_1))_+(\mathcal{X}_2)$. On voit alors facilement que $x_1 \mapsto \pi_2^{x_1}$ est un noyau de transition, ce qui, d'après la remarque VI.70, assure la validité de (VI.75). □

Exemple : En prenant $\chi_1 = \cdots = \chi_n = 1$ et en utilisant l'inégalité de Pinsker (VI.46) $\frac{1}{2}\|\mu - \nu\|_{VT}^2 \leq \mathrm{H}\left(\nu|\,\mu\right)$, on obtient immédiatement la généralisation suivante de l'inégalité de Pinsker due à K. Marton :

Proposition VI.76. (Marton [47]) *Soient $(\mathcal{X}_1, \mu_1), \ldots, (\mathcal{X}_n, \mu_n)$ des espaces de probabilité. Considérons la distance de Hamming sur $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ définie par*

$$d_H^n(x,y) = \sum_{i=1}^n \mathbb{1}_{x_i \neq y_i}.$$

Alors $\mu_1 \otimes \cdots \otimes \mu_n$ satisfait $T_{d_H^n}\left(\frac{x^2}{2}, \sqrt{n}\right)$, ie

$$\forall \nu \in \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n), \quad \mathcal{T}_{d_H^n}(\nu, \mu_1 \otimes \cdots \otimes \mu_n) \leq \sqrt{\frac{n}{2}\,\mathrm{H}\left(\nu|\,\mu_1 \otimes \cdots \otimes \mu_n\right)}.$$

Tensorisation via le critère dual : Soit $c(\,.\,,\,.\,)$ une fonction de coût symétrique, continue sur un espace polonais $\mathcal{X}$ telle que $c(x,x) = 0$, pour tout $x \in \mathcal{X}$. Remarquons qu'en posant pour toute fonction φ semi-continue supérieurement bornée (s.c.s.b) sur $\mathcal{X}$,

$$Q_c\varphi(x) = \inf_{y \in \mathcal{X}} \{\varphi(y) + c(x,y)\},$$

$Q_c\varphi$ est s.c.s.b ($\forall x \in \mathcal{X},\ \inf \varphi \leq Q_c\varphi(x) \leq \varphi(x)$) et on voit facilement à partir du théorème VI.2, que

$$\mathcal{T}_c(\nu, \mu) = \sup_{\varphi \text{ s.c.s.b}} \left\{ \int_{\mathcal{X}} Q_c\varphi\,d\nu - \int_{\mathcal{X}} \varphi\,d\mu \right\}.$$

Le critère du théorème VI.38 peut se reformuler sous la forme :

$$\left(\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^*\left(\mathcal{T}_c(\nu,\mu)\right) \leq \mathrm{H}\left(\nu|\,\mu\right)\right)$$
$$\Leftrightarrow$$
$$\left(\forall \varphi \text{ s.c.s.b sur } \mathcal{X}, \quad \forall s \geq 0, \quad \int_{\mathcal{X}} e^{sQ_c\varphi}\,d\mu \leq e^{\theta(s)+s\langle\mu,\varphi\rangle}\right)$$
$$\Leftrightarrow$$
$$\left(\forall \varphi \in C_b(\mathcal{X}), \quad \forall s \geq 0, \quad \int_{\mathcal{X}} e^{sQ_c\varphi}\,d\mu \leq e^{\theta(s)+s\langle\mu,\varphi\rangle}\right).$$

Démonstration du théorème VI.64.
Il suffit de traiter le cas $n = 2$. D'après la remarque précédente, on a pour $i = 1, 2$:

$$\forall \varphi \text{ s.c.s.b sur } \mathcal{X}_i, \quad \forall s \geq 0, \quad \int_{\mathcal{X}_i} e^{sQ_{c_i}\varphi}\,d\mu_i \leq e^{\theta_i(s)+s\langle\mu_i,\varphi\rangle} \tag{VI.77}$$

De plus, comme $\theta_1^*\Box\theta_2^* = (\theta_1+\theta_2)^*$ (voir par exemple la théorème 2.3.1 p. 227), il suffit de montrer que

$$\forall \varphi \in C_b(\mathcal{X}_1\times\mathcal{X}_2), \quad \forall s \geq 0, \quad \iint_{\mathcal{X}_1\times\mathcal{X}_2} e^{sQ_{c_1\oplus c_2}\varphi}\,d\mu_1\otimes\mu_2 \leq e^{\theta_1(s)+\theta_2(s)+s\langle\mu_1\otimes\mu_2,\varphi\rangle}. \tag{VI.78}$$

Or,

$$\begin{aligned}
Q_{c_1\oplus c_2}\varphi(x_1,x_2) &= \inf_{(y,z)\in\mathcal{X}_1\times\mathcal{X}_2}\{\varphi(y,z)+c_1(x_1,y)+c_2(x_2,z)\}\\
&= \inf_{y\in\mathcal{X}_1}\left\{\inf_{z\in\mathcal{X}_2}\{\varphi(y,z)+c_2(x_2,z)\}+c_1(x_1,y)\right\}\\
&= Q_{c_1}\varphi_{x_2}(x_1),
\end{aligned}$$

en posant $\varphi_{x_2}(y) = \inf_{z\in\mathcal{X}_2}\{\varphi(y,z)+c_2(x_2,z)\}$ qui est s.c.s.b sur $\mathcal{X}_1$.
Donc, d'après (VI.77),

$$\begin{aligned}
\iint_{\mathcal{X}_1\times\mathcal{X}_2} e^{sQ_{c_1\oplus c_2}\varphi}\,d\mu_1\otimes\mu_2 &= \int_{\mathcal{X}_2}\left(\int_{\mathcal{X}_1} e^{sQ_{c_1}\varphi_{x_2}(x_1)}\,d\mu_1(x_1)\right)\,d\mu_2(x_2)\\
&\leq \int_{\mathcal{X}_2} e^{s\theta_1(s)+s\left\langle\mu_1,\varphi_{x_2}(.)\right\rangle}\,d\mu_2(x_2).
\end{aligned}$$

Or,

$$\begin{aligned}
\langle\mu_1,\varphi_{x_2}(.)\rangle &= \int_{\mathcal{X}_1}\inf_{z\in\mathcal{X}_2}\{\varphi(x_1,z)+c_2(x_2,z)\}\,d\mu_1(x_1)\\
&\leq \inf_{z\in\mathcal{X}_2}\left\{\int_{\mathcal{X}_1}\varphi(x_1,z)\,d\mu_1(x_1)+c_2(x_2,z)\right\}\\
&= Q_{c_2}\tilde{\varphi}(x_2),
\end{aligned}$$

avec $\widetilde{\varphi}(z) = \int_{\mathcal{X}_1} \varphi(x_1, z)\, d\mu_1(x_1)$ qui est continue sur $\mathcal{X}_2$. En appliquant une nouvelle fois (VI.77), on obtient :

$$\begin{aligned}\iint_{\mathcal{X}_1\times\mathcal{X}_2} e^{sQ_{c_1\oplus c_2}\varphi}\, d\mu_1\otimes\mu_2 &\leq e^{\theta_1(s)} \int_{\mathcal{X}_2} e^{sQ_{c_2}\widetilde{\varphi}(x_2)}\, d\mu_2(x_2)\\ &\leq e^{\theta_1(s)+\theta_2(s)} e^{s\langle\mu_2,\widetilde{\varphi}\rangle}\\ &= e^{\theta_1(s)+\theta_2(s)+s\langle\mu_1\otimes\mu_2,\varphi\rangle}.\end{aligned}$$

□

Remarque VI.79.

Il n'y a pas de propriété de tensorisation générale des inégalités de la forme $T_\Phi(\theta^*, a)$. Néanmoins, on dispose de la proposition suivante :

Proposition VI.80. *Soient* $(X_i, \mathcal{B}_i), i = 1\ldots n$ *des espaces mesurables. Pour tout* $i = 1\ldots n$, d_i *est une métrique sur* $\mathcal{X}_i$ *et* $BLip_1(\mathcal{X}_i, d_i)$ *est l'ensemble des applications 1-Lipschitziennes pour* d_i *et* $\mathcal{B}_i$ *mesurable. Si pour tout* i, μ_i *est une probabilité sur* $(\mathcal{X}_i, \mathcal{B}_i)$ *vérifiant l'inégalité :*

$$\forall\nu\in\mathcal{P}(\mathcal{X}_i), \quad \theta_i^*\left(\|\nu-\mu\|^*_{BLip_1(\mathcal{X}_i,d_i)}\right) \leq \mathrm{H}\left(\nu|\,\mu_i\right),$$

avec $\theta_i \in \mathcal{C}$*, alors* $\mu_1\otimes\cdots\otimes\mu_n$ *vérifie*

$$\begin{aligned}&\forall\nu\in\mathcal{P}(\mathcal{X}_1\times\cdots\times\mathcal{X}_n),\\ &\qquad \theta_1^*\Box\cdots\Box\theta_n^*\left(\|\nu-\mu_1\otimes\cdots\otimes\mu_n\|^*_{BLip_1(\Pi\mathcal{X}_i,\oplus_i d_i)}\right) \leq \mathrm{H}\left(\nu|\,\mu_1\otimes\cdots\otimes\mu_n\right).\end{aligned}$$

Démonstration. Il suffit de montrer la proposition pour $n = 2$. D'après la proposition VI.42, il suffit de montrer que pour toute $\varphi \in BLip_1(\mathcal{X}_1\times\mathcal{X}_2, d_1\oplus d_2)$, on a

$$\int_{\mathcal{X}_1\times\mathcal{X}_2} e^{s\varphi - s\langle\varphi,\mu_1\otimes\mu_2\rangle}\, d\mu_1\otimes\mu_2 \leq e^{\theta_1(s)+\theta_2(s)}.$$

Or, pour tout $s > 0$,

$$\begin{aligned}\int_{\mathcal{X}_1}\int_{\mathcal{X}_2} e^{s\varphi(x_1,x_2)}\, d\mu_2(x_2) d\mu_1(x_1) &\overset{(i)}{\leq} \int_{\mathcal{X}_1} \exp\left(s\int_{\mathcal{X}_2}\varphi(x_1,x_2)\, d\mu_2(x_2) + \theta_2(s)\right) d\mu_1(x_1)\\ &\overset{(ii)}{\leq} \exp\left(\theta_1(s)+\theta_2(s) + s\int_{\mathcal{X}_1\times\mathcal{X}_2}\varphi(x_1,x_2)\, d\mu_1(x_1) d\mu_2(x_2)\right)\end{aligned}$$

où (i) vient du fait que pour tout $x_1 \in \mathcal{X}_1$, la fonction $x_2 \mapsto \varphi(x_1, x_2)$ appartient à $BLip_1(\mathcal{X}_2, d_2)$, et (ii) du fait que $x_1 \mapsto \int_{X_2} \varphi(x_1, x_2)\, d\mu_2(x_2)$ appartient à $BLip_1(\mathcal{X}_1, d_1)$.

□

VI.3 Applications des I.T.C

Dans cette section, nous allons rappeler un certain nombre d'applications bien connues des inégalités de transport pour un coût métrique.

VI.3.1 Inégalités de concentration

Le procédé utilisé dans la preuve de la proposition suivante est connu sous le nom d'*argument de Marton* :

Proposition VI.81 (Marton, [47])**.**
Soit $(\mathcal{X}, d)$ un espace polonais et $\mu \in \mathcal{P}(\mathcal{X})$. Si μ satisfait l'inégalité $T_d(\theta^*, a)$, alors pour tout ensemble mesurable $A \subset \mathcal{X}$ tel que $\mu(A) \geq \frac{1}{2}$, on a :

$$\mu(A^\varepsilon) \geq 1 - \exp\left(-\theta^*\left(\frac{\varepsilon - r}{a}\right)\right), \tag{VI.82}$$

avec $r = a\theta^{*-1}(\log(2))$ et $A^\varepsilon = \{x \in \mathcal{X} : d(x, A) \leq \varepsilon\}$.

Démonstration. Pour tout A, B mesurables tels que $\mu(A) > \frac{1}{2}, \mu(B) > 0$, notons

$$\mu_A(\,.\,) = \frac{\mu(\,.\cap A)}{\mu(A)} \qquad \text{et} \qquad \mu_B(\,.\,) = \frac{\mu(\,.\cap B)}{\mu(B)}.$$

Alors, d'après l'inégalité triangulaire (voir, par exemple, la preuve du théorème 7.3 de [72]) et l'inégalité de transport satisfaite par μ, on a :

$$\begin{aligned}
\mathcal{T}_d(\mu_A, \mu_B) &\leq \mathcal{T}_d(\mu_A, \mu) + \mathcal{T}_d(\mu_B, \mu) \leq a\theta^{*-1}\big(\mathrm{H}(\mu_A|\,\mu)\big) + a\theta^{*-1}\big(\mathrm{H}(\mu_B|\,\mu)\big) \\
&= a\theta^{*-1}\big(-\log\mu(A)\big) + a\theta^{*-1}\big(-\log\mu(B)\big) \\
&\leq a\theta^{*-1}\big(\log(2)\big) + a\theta^{*-1}\big(-\log\mu(B)\big)
\end{aligned}$$

Or, si $\pi \in \Pi(\mu_A, \mu_B)$, alors $\pi(A \times B) = 1$, car

$$\pi\left((A \times B)^c\right) \leq \pi(A^c \times \mathcal{X}) + \pi(\mathcal{X} \times B^c) = \mu_A(A^c) + \mu_B(B^c) = 0$$

En particulier, si $B = A^{\varepsilon c}$, on a pour tout $\pi \in \Pi(\mu_A, \mu_{A^{\varepsilon c}})$:

$$\iint_{\mathcal{X}^2} d(x,y)\, d\pi = \iint_{A \times A^{\varepsilon c}} d(x,y)\, d\pi \geq \varepsilon,$$

et par conséquent $\mathcal{T}_d(\mu_A, \mu_{A^{\varepsilon c}}) \geq \varepsilon$.
Ainsi,

$$\varepsilon \leq a\theta^{*-1}\big(\log(2)\big) + a\theta^{*-1}\big(-\log\mu(B)\big),$$

et l'inégalité (VI.82) s'en déduit immédiatement. $\square$

Proposition VI.83. *Soient $\mathcal{X}$ un espace mesurable, d une distance mesurable sur $\mathcal{X}$ et $\mu \in \mathcal{P}(\mathcal{X})$ telle que $\int_{\mathcal{X}} d(x_0, x)\, d\mu(x) < +\infty$ pour au moins un $x_0 \in \mathcal{X}$. Si μ satisfait $T_{BLip_1(\mathcal{X},d)}(\theta^*, a)$, alors pour toute fonction mesurable φ 1-Lipschitzienne pour d, on a*

$$\forall t \geq 0, \quad \mu\big(\varphi \geq \langle \mu, \varphi \rangle + t\big) \leq e^{-\theta^*\left(\frac{t}{a}\right)} \tag{VI.84}$$

Démonstration.
D'après la proposition VI.42, pour toute $\varphi \in \mathrm{BLip}_1(\mathcal{X}, d)$, on a

$$\forall s \geq 0, \quad \int_{\mathcal{X}} e^{s\varphi}\, d\mu \leq e^{\theta(as)+s\langle \mu, \varphi \rangle}. \tag{VI.85}$$

Si maintenant $\varphi \in \mathrm{Lip}_1(\mathcal{X}, d)$, en posant $\varphi_n = \varphi \wedge n \vee -n$, on voit, par convergence dominée, que (VI.85) reste vraie pour φ. On obtient alors (VI.84) grâce à la majoration de Chebychev :

$$\mu\left(\varphi \geq \langle \mu, \varphi \rangle + t\right) \leq \inf_{s \geq 0} \int_{\mathcal{X}} e^{s(\varphi - \langle \mu, \varphi \rangle - t)}\, d\mu \leq \inf_{s \geq 0} e^{\theta(as) - st} = e^{-\theta^*\left(\frac{t}{a}\right)}.$$

□

VI.3.2 I.T.C et inégalités de déviations

La propriété de tensorisation des I.T.C associées à des coûts métriques permet de déduire des inégalités de déviations pour une classe enrichie d'objets :

Proposition VI.86. *Soient $(\mathcal{X}, \mathcal{B})$ un espace mesurable, d une distance mesurable sur $\mathcal{X}$ et $\mu \in \mathcal{P}(\mathcal{X})$ vérifiant l'inégalité $T_{BLip_1(\mathcal{X},d)}(\theta^*, a)$ et telle que, pour tout $x \in \mathcal{X}$, $\int_{\mathcal{X}} d(x, y)\, d\mu(y) < +\infty$. Si X_i est une suite de variables aléatoires i.i.d de loi μ, alors pour toute fonction $F : \mathcal{X}^n \to \mathbb{R}$ mesurable et 1-Lipschitzienne pour la distance $\oplus^n d$ définie par*

$$\oplus^n d(x, y) = d(x_1, y_1) + \cdots + d(x_n, y_n),$$

on a

$$\forall n \in \mathbb{N}^*, \quad \forall t \geq 0, \quad \mathbb{P}(F(X_1, \ldots, X_n) \geq \mathbb{E}[F] + t) \leq e^{-n\theta^*(t/an)},$$

ou de manière équivalente,

$$\forall n \in \mathbb{N}^*, \quad \forall u \geq 0, \quad \mathbb{P}\left(F(X_1, \ldots, X_n) \geq \mathbb{E}[F] + an\theta^{*-1}\left(\frac{u}{n}\right)\right) \leq e^{-u}.$$

En particulier,

1. *si $\mathcal{F}$ est une classe dénombrable d'applications mesurables 1-Lipschitziennes pour d, alors en notant $Z_n^{\mathcal{F}} = \sup_{\varphi \in \mathcal{F}} \left| \langle L_n, \varphi \rangle - \int_{\mathcal{X}} \varphi\, d\mu \right|$, on a*

$$\forall n \in \mathbb{N}^*, \quad \forall t \geq 0, \quad \mathbb{P}\left(Z_n^{\mathcal{F}} \geq \mathbb{E}\left[Z_n^{\mathcal{F}}\right] + t\right) \leq e^{-n\theta^*(t/a)}. \tag{VI.87}$$

2. *si* $(X, \| \, . \, \|)$ *est un espace de Banach et* $d(x, y) = \|x - y\|$*, alors en notant* $Z_n = \dfrac{X_1 + \cdots + X_n}{n} - \displaystyle\int_{\mathcal{X}} x \, d\mu$*, on a*

$$\forall n \in \mathbb{N}^*, \quad \forall t \geq 0, \quad \mathbb{P}\left(\|Z_n\| \geq \mathbb{E}\left[\|Z_n\|\right] + t\right) \leq e^{-n\theta^*(t/a)}.$$

Démonstration. On voit facilement, d'après le théorème VI.64, que $\mu_1 \otimes \cdots \otimes \mu_n$ satisfait l'inégalité de transport $T_{\oplus^n d}(n\theta^*, an)$. On conclut grâce à la proposition VI.83. Pour le reste, on rappelle qu'un sup d'applications 1-Lipschitziennes est 1-Lipschitzienne. □

Exemples : S'il existe $\delta > 0$ tel que $\iint_{\mathcal{X}^2} e^{\delta d(x,y)} d\mu(x) d\mu(y) < +\infty$, alors, d'après le corollaire VI.61, μ vérifie l'inégalité $T_{BLip_1(\mathcal{X},d)}(\theta_1^*, M)$, avec $\theta_1^*(t) = \left(\sqrt{1+t} - 1\right)^2$ et $M = \inf\left\{\lambda > 0 : \iint_{\mathcal{X}^2} e^{\frac{d(x,y)}{\lambda}} d\mu(x) d\mu(y) \leq 2\right\}$. La proposition VI.86 entraîne que, pour toute classe $\mathcal{F}$ d'applications mesurables 1-Lipschitzienne pour d,

$$\forall n \in \mathbb{N}^*, \quad \forall t \geq 0, \quad \mathbb{P}\left(Z_n^{\mathcal{F}} \geq \mathbb{E}\left[Z_n^{\mathcal{F}}\right] + t\right) \leq e^{-n\left(\sqrt{1+\frac{t}{M}} - 1\right)^2}.$$

Si $\mathcal{X}$ est un espace de Banach et $d = \| \, . \, \|$, alors, sous les mêmes hypothèses :

$$\forall n \in \mathbb{N}^*, \quad \forall t \geq 0, \quad \mathbb{P}\left(\|Z_n\| \geq \mathbb{E}\left[\|Z_n\|\right] + t\right) \leq e^{-n\left(\sqrt{1+\frac{t}{M}} - 1\right)^2}.$$

Pour que les bornes de la proposition VI.86 soient utilisables, il faut être capable de montrer que le terme d'espérance $\mathbb{E}\left[Z_n^{\mathcal{F}}\right]$ tend vers 0 et d'estimer la vitesse de cette convergence.
Le résultat suivant permet de conclure lorsque d est la distance euclidienne sur $\mathbb{R}^q$:

Théorème VI.88.
Soit μ *une mesure de probabilité sur* $\mathbb{R}^q$ *telle que*

$$c := \int \|x\|^{q+5} \, d\mu < +\infty. \tag{VI.89}$$

Alors, il existe une constante D *ne dependant que de* c *et de* q*, telle que*

$$\mathbb{E}\left[\mathcal{T}_2(L_n, \mu)\right] \leq D n^{-\frac{2}{q+4}}, \tag{VI.90}$$

où $\mathcal{T}_2(\nu, \mu) = \inf\{\int \|x - y\|^2 \, d\pi(x, y) : \pi \in \Pi(\mu, \nu)\}$.

Démonstration. Voir le théorème 10.2.1 de [56] (volume II). □

En notant $\mathcal{T}_1(\nu, \mu) = \inf\{\int \|x - y\| \, d\pi(x, y) : \pi \in \Pi(\mu, \nu)\}$, on a d'après l'inégalité de Jensen :

$$\mathcal{T}_1(\nu, \mu) \leq \sqrt{\mathcal{T}_2(\nu, \mu)}.$$

Corollaire VI.91. *Soit μ une probabilité sur $\mathbb{R}^q$, vérifiant (VI.89) et l'inégalité de transport*

$$\forall \nu \in \mathcal{P}(\mathbb{R}^q), \quad \theta^*\left(\frac{\mathcal{T}_1(\nu,\mu)}{a}\right) \leq \mathrm{H}\left(\nu|\,\mu\right),$$

alors, pour toute classe $\mathcal{F}$ de fonctions 1-Lipschitziennes, on a pour tout $u > 0$,

$$\forall n \geq \left(\frac{\sqrt{D}}{u}\right)^{q+4}, \quad \mathbb{P}\left(Z_n^{\mathcal{F}} \geq u\right) \leq \exp\left(-n\theta^*\left(\frac{u}{a} - \frac{\sqrt{D}}{an^{\frac{1}{q+4}}}\right)\right),$$

où D est la constante de (VI.90).

Démonstration. Il suffit de remarquer que, d'après le théorème VI.88, on a

$$\mathbb{E}\left[Z_n^{\mathcal{F}}\right] \leq \mathbb{E}\left[\mathcal{T}_1(L_n,\mu)\right] \leq \sqrt{\mathbb{E}\left[\mathcal{T}_2(L_n,\mu)\right]} \leq \sqrt{D}n^{-\frac{1}{q+4}},$$

puis d'appliquer (VI.87). $\square$

CHAPITRE VII

Méthodes d'Orlicz pour certaines inégalités de transport convexes

Sommaire

VII.1 Introduction

VII.1.1 Cadre

Dans ce chapitre, nous nous placerons dans le cadre suivant :

- $\mathcal{X}$ sera un espace mesurable,
- Φ sera une classe de fonctions mesurables bornées sur $\mathcal{X}$ qui sera supposée symétrique *ie*, $\varphi \in \Phi \Rightarrow -\varphi \in \Phi$.
- μ sera une probabilité de référence sur $\mathcal{X}$,
- Pour toute $\nu \in \mathcal{P}(\mathcal{X})$, nous poserons

$$\|\nu - \mu\|_\Phi^* = \sup_{\varphi \in \Phi} \left\{ \int_{\mathcal{X}} \varphi \, d\nu - \int_{\mathcal{X}} \varphi \, d\mu \right\},$$

- Enfin, $\mathcal{C}$ désignera la classe des fonctions $\theta : \mathbb{R} \to \mathbb{R}^+ \cup \{+\infty\}$, convexes, semi-continues inférieurement, $\theta(0) = 0$, $\operatorname{dom} \theta = [0, a_\theta[$, avec $a_\theta \in]0, +\infty]$,

Pour $\theta \in \mathcal{C}$, nous dirons que μ satisfait l'inégalité de transport convexe $T_\Phi(\theta^*, a)$, si

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^* \left(\frac{\|\nu - \mu\|_\Phi}{a} \right) \leq \mathrm{H}\left(\nu \middle| \mu\right). \tag{VII.1}$$

L'objectif de ce chapitre est d'obtenir une condition nécessaire et suffisante pour que μ vérifie (VII.1). Nous allons voir que (VII.1) est en lien avec des propriétés d'intégrabilité exponentielle des éléments de Φ.

Introduisons l'espace d'Orlicz de type exponentiel suivant :

$$\mathbb{L}_{E\theta^*}(\mathcal{X}, \mu) = \left\{ \varphi \text{ mesurable } : \exists \lambda > 0, \int_{\mathcal{X}} \exp \theta^* \left(\frac{|\varphi|}{\lambda} \right) d\mu < +\infty \right\}$$

qui sera muni de la norme de Luxembourg :

$$\|\varphi\|_{E\theta^*}^{(r)} = \inf \left\{ \lambda > 0 : \int_{\mathcal{X}} \exp \theta^* \left(\frac{|\varphi|}{\lambda} \right) d\mu \leq r \right\}$$

où r est un nombre réel strictement supérieur à 1.

Dans ce qui suit, $\mathcal{C}_{quad}$ désignera l'ensemble des fonctions $\theta \in \mathcal{C}$ *quadratiques à l'origine* au sens suivant :

$$\exists s_\theta > 0, c_\theta > 0, \quad \forall s \in [0, s_\theta], \quad \theta(s) \geq c_\theta s^2. \tag{VII.2}$$

Le résultat principal de ce chapitre est le théorème suivant :

Théorème VII.3. *Soit $\theta \in \mathcal{C}_{quad}$, il y a équivalence entre*

1. *Il existe $a > 0$ tel que μ satisfait $T_\Phi(\theta^*, a)$,*
2. *$\widetilde{\Phi} = \{\varphi - \langle \varphi, \mu \rangle, \varphi \in \Phi\}$ est une partie bornée de $\mathbb{L}_{E\theta^*}(\mathcal{X}, \mu)$.*

Plus précisément,

$$(\mu \text{ satisfait } T_\Phi(\theta^*, a)) \Rightarrow \left(\forall \phi \in \Phi, \quad \|\varphi - \langle \varphi, \mu \rangle\|_{E\theta^*}^{(r)} \leq \frac{r+1}{r-1} a. \right)$$

et

$$\left(\forall \phi \in \Phi, \quad \|\varphi - \langle \varphi, \mu \rangle\|_{E\theta^*}^{(r)} \leq M. \right) \Rightarrow \left(\mu \text{ satisfait } T_\Phi(\theta^*, \sqrt{r} m_\theta M) \right) \tag{VII.4}$$

où

$$m_\theta = e \max \left(\frac{1}{\theta^{-1}(2)\sqrt{c_\theta(1-u)}}, \frac{1}{u} \right) \text{ où } u \in [0, 1[\text{ est tel que :}$$
$$\frac{u}{\sqrt{1-u}} \leq s_\theta \sqrt{c_\theta} \quad \text{et} \quad \frac{u^3}{1-u} \leq 2$$

La preuve de ce théorème repose sur un résultat assez ancien de Kozachenko et Ostrowski (théorème VII.25) qui fournit une majoration de la transformée de Laplace d'une variable aléatoire vérifiant une condition d'Orlicz. En prenant pour Φ la boule des fonctions 1-Lipschitziennes d'un espace polonais $(\mathcal{X}, d)$, on déduit immédiatement du théorème VII.3 un résultat concernant l'inégalité $T_d(\theta^*, a)$ (voir théorème VII.38). En utilisant une idée de F. Bolley et C. Villani, on obtiendra le théorème suivant qui concerne des I.T.C associées à des coûts non-métriques :

Théorème VII.5. *Soient $(\mathcal{X}, d)$ un espace polonais et $c(\,.\,,\,.\,)$ une fonction de coût sur $\mathcal{X}$ s'écrivant sous la forme $c(x, y) = q(d(x, y))$, avec $q : \mathbb{R}^+ \to \mathbb{R}^+$ une fonction convexe strictement croissante, satisfaisant la condition Δ_2, ie*

$$\exists K > 0, \quad \forall x \in \mathbb{R}^+, \quad q(2x) \leq K q(x),$$

Pour tout $\theta \in \mathcal{C}_{quad}$, les deux propositions suivantes sont équivalentes :

1. $\exists a > 0, \quad \forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^* \left(\frac{\mathcal{T}_c(\nu, \mu)}{a} \right) \leq \mathrm{H}\left(\nu | \, \mu\right),$
2. $\exists b > 0, \quad \iint_{\mathcal{X}^2} \exp \theta^* \left(\frac{c(x, y)}{b} \right) d\mu(x) d\mu(y) < +\infty.$

VII.1.2 A propos de la littérature.

Les liens entre intégrabilité exponentielle et inégalités de transport ont été étudiés dans deux articles récents (voir [27] et [5]). Dans [27], H. Djellout, A. Guillin et L. Wu ont établi la première condition nécessaire et suffisante pour une inégalité de transport de la forme :

$$\mathcal{T}_d(\nu,\mu) \leq a\sqrt{2\,\mathrm{H}\left(\nu|\,\mu\right)} \tag{VII.6}$$

Ils ont obtenu le

Théorème VII.7. (Djellout, Guillin, Wu, [27],thm 3.1)
Si μ vérifie (VII.6), alors pour tout $\delta \in]0, \frac{1}{4a}[$, $\displaystyle\iint_{\mathcal{X}^2} e^{\delta d(x,y)^2}\,d\mu(x)d\mu(y) < +\infty$.
Si $\displaystyle\iint_{\mathcal{X}^2} e^{\delta d(x,y)^2}\,d\mu(x)d\mu(y) < +\infty$ pour un certain $\delta > 0$, alors μ satisfait (VII.6) avec

$$a = \sup_{k\geq 1}\left[\frac{2^k k! \iint_{\mathcal{X}^2} d(x,y)^{2k}\,d\mu(x)d\mu(y)}{(2k)!}\right]^{1/2k} \tag{VII.8}$$

et on a la majoration :

$$a \leq \frac{\sqrt{2}}{\delta}\sup_{k\geq 1}\left(\frac{(k!)^2}{(2k!)}\right)^{1/2k}\left[\iint_{\mathcal{X}^2} e^{\delta^2 d(x,y)^2}\,d\mu(x)d\mu(y)\right]^{1/2k} < +\infty \tag{VII.9}$$

Dans [5], F. Bolley et C. Villani ont démontré une version pondérée de l'inégalité de Csiszar-Pinsker-Kullback :

Théorème VII.10. (Bolley, Villani, [5], thm 1)
Soit $\chi : \mathcal{X} \to \mathbb{R}^+$, une fonction mesurable. Alors pour toute $\nu \in \mathcal{P}(\mathcal{X})$,

(i) $\|\chi\nu - \chi\mu\|_{VT} \leq \left(\dfrac{3}{2} + \log\displaystyle\int_{\mathcal{X}} e^{2\chi(x)}\,d\mu(x)\right)\left(\sqrt{\mathrm{H}\left(\nu|\,\mu\right)} + \dfrac{1}{2}\mathrm{H}\left(\nu|\,\mu\right)\right)$;

(ii) $\|\chi\nu - \chi\mu\|_{VT} \leq \sqrt{1 + \log\int_{\mathcal{X}} e^{\chi(x)^2} d\mu(x)}\sqrt{2\,\mathrm{H}\left(\nu|\,\mu\right)}$.

En utilisant la majoration (voir [72], prop. 7.10)

$$\mathcal{T}_{d^p}(\nu,\mu) \leq 2^{p-1}\|d(x_0,\,.\,)^p\mu - d(x_0,\,.\,)^p\nu\|_{VT}, \tag{VII.11}$$

ils déduisent du théorème VII.10, les résultats suivants :

Corollaire VII.12. (Bolley,Villani, [5] cor. 3 et 4)
Pour toute $\nu \in \mathcal{P}(\mathcal{X})$, on a pour tout $p \geq 1$:

(i)
$$\mathcal{T}_{d^p}(\nu,\mu)^{1/p} \leq C_1 \left[\mathrm{H}\,(\nu|\,\mu)^{1/p} + \left(\frac{\mathrm{H}\,(\nu|\,\mu)}{2}\right)^{1/2p}\right],$$

avec
$$C_1 = 2 \inf_{x_0\in\mathcal{X},\,\delta>0} \left[\frac{1}{\delta}\left(\frac{3}{2} + \log \int_{\mathcal{X}} e^{\delta d(x_0,x)^p}\, d\mu(x)\right)\right]^{1/p}$$

(ii)
$$\mathcal{T}_{d^p}(\nu,\mu) \leq C_2\, \mathrm{H}\,(\nu|\,\mu)^{1/2p},$$

avec
$$C_2 = 2 \inf_{x_0\in\mathcal{X},\,\delta>0} \left[\frac{1}{2\delta}\left(1 + \log \int_{\mathcal{X}} e^{\delta d(x_0,x)^{2p}}\, d\mu(x)\right)\right]^{1/2p}$$

En particulier, pour $p = 1$, la constante C_2 figurant au point (ii) du théorème précédent est nettement meilleure que l'estimée fournie par (VII.9). Néanmoins, dans la section VII.4, nous montrerons qu'une majoration plus fine de (VII.8) permet d'obtenir, à un facteur numérique près, la constante de Bolley et Villani.

VII.2 Conditions nécessaires pour une I.T.C.

Commençons par une remarque élémentaire réduisant la classe des fonctions θ admissibles. Si Φ n'est constituée que de fonctions μ-ps constantes, $\|\nu - \mu\|_\Phi^* = 0$ pour toute probabilité $\nu \ll \mu$; nous exclurons donc ce cas d'étude triviale dans ce qui suit. On a la

Proposition VII.13. *Si μ satisfait $T_\Phi(\theta^*, a)$, alors*

$$\exists s_\theta > 0, c_\theta > 0, \quad \forall s \in [0, s_\theta], \quad \theta(s) \geq c_\theta s^2. \tag{VII.14}$$

Démonstration. On peut supposer que $a = 1$. Soit $\varphi \in \Phi$ une fonction non constante ; notons $\Lambda_\varphi(s) = \log \int_{\mathcal{X}} e^{s\varphi}\, d\mu$. Alors,

$$\lim_{s\to 0^+} \frac{\Lambda_\varphi(s) - s\langle\varphi,\mu\rangle}{s^2} = \frac{1}{2}\mathrm{Var}_\mu(\varphi) > 0.$$

Comme, d'après la proposition VI.42, $\Lambda_\varphi(s) - s\langle\varphi,\mu\rangle \leq \theta(s)$, on en déduit que $\liminf_{s\to 0^+} \dfrac{\theta(s)}{s^2} > 0$, ce qui entraîne facilement (VII.14). □

Rappelons quelques notations :

- On désignera par $\mathcal{C}_{quad}$, la classe des fonctions convexes s.c.i. $\theta : \mathbb{R} \to \mathbb{R}^+ \cup \{+\infty\}$ telles que $\theta \equiv +\infty$ sur $]-\infty, 0[$, $\theta(0) = 0$ et θ vérifie (VII.14).

- Pour $\varphi \in \Phi$, nous noterons $\widetilde{\varphi} = \varphi - \langle \varphi, \mu \rangle$, et $\widetilde{\Phi} = \{\widetilde{\varphi}, \varphi \in \Phi\}$.

Les deux propositions suivantes donnent des conditions nécessaires pour $T_\Phi(\theta^*, a)$ et $T_d(\theta^*, a)$:

Proposition VII.15. *Si μ satisfait $T_\Phi(\theta^*, a)$, alors $\widetilde{\Phi}$ est une partie bornée de $\mathbb{L}_{E\theta^*}(\mathcal{X}, \mu)$. Plus précisément, pour tout $r > 1$,*

$$\forall \varphi \in \Phi, \quad \|\varphi - \langle \varphi, \mu \rangle\|_{E\theta^*}^{(r)} \leq \frac{r+1}{r-1} a$$

Proposition VII.16. *Si $(\mathcal{X}, d)$ est un espace polonais et si μ vérifie $T_d(\theta^*, a)$, alors*

$$\int_{\mathcal{X}^2} \exp \theta^* \left(\frac{d(x,y)}{3a} \right) d\mu(x) d\mu(y) < +\infty$$

Pour prouver les propositions VII.15 et VII.16, nous aurons besoin des lemmes suivants :

Lemme VII.17. *Soit X une variable aléatoire réelle telle que $\mathbb{E}\left[e^{\delta|X|}\right] < +\infty$, pour au moins un $\delta > 0$. En notant γ^* la transformée de Cramer de X, on a*

$$\forall \varepsilon \in [0, 1[, \quad \mathbb{E}\left[e^{\varepsilon \gamma^*(X)}\right] \leq \frac{1+\varepsilon}{1-\varepsilon}.$$

Lemme VII.18. *Si φ est une fonction mesurable telle que $\langle \varphi, \mu \rangle = 0$ et si*

$$\exists a > 0, \quad \forall s \in \mathbb{R}, \quad \int_{\mathcal{X}} e^{s\varphi} \, d\mu \leq e^{\theta(a|s|)}, \tag{VII.19}$$

alors $\varphi \in \mathbb{L}_{E\theta^*}(\mathcal{X}, \mu)$ et on a, pour tout $r > 1$,

$$\|\varphi\|_{E\theta^*}^{(r)} \leq \frac{r+1}{r-1} a.$$

Démonstration du lemme VII.17. Le domaine de γ^*, $\operatorname{dom} \gamma^*$, est un intervalle d'extrémités $a < b$, $a \in \mathbb{R} \cup \{-\infty\}, b \in \mathbb{R} \cup \{+\infty\}$. Pour tout $t \geq 0$, γ^* étant convexe s.c.i, $\{\gamma^* \leq t\}$ est un intervalle fermé d'extrémités $a \leq a(t) \leq b(t) \leq b$.
Donc, pour tout $t \geq 0$

$$\mathbb{P}(\gamma^*(X) > t) = \mathbb{P}(X < a(t)) + \mathbb{P}(X > b(t))$$

Soit $m = \mathbb{E}[X]$. Comme $\gamma^*(m) = 0$, on a $a(t) \leq m$. Or pour tout $u \leq m$, il est bien connu que :

$$\mathbb{P}(X \leq u) \leq \exp(-\gamma^*(u)) \tag{VII.20}$$

Si $a(t) > a$, on voit facilement grâce à la continuité de γ^* sur $]a, b[$ que $\gamma^*(a(t)) = t$; donc, d'après (VII.20),

$$\mathbb{P}(X < a(t)) \leq e^{-t}.$$

Si $a(t) = a$, on a :

$$\mathbb{P}(X < a) = \lim_{n\to+\infty} \mathbb{P}(X < a - 1/n) \overset{(i)}{\leq} \lim_{n\to+\infty} \exp(-\gamma^*(a - 1/n)) \overset{(ii)}{=} \lim_{n\to+\infty} 0 = 0$$

(i) venant de (VII.20), et (ii) de $a - 1/n \notin \mathrm{dom}\ \gamma^*$.
Ainsi, dans tous les cas, $\mathbb{P}(X < a(t)) \leq e^{-t}$, et de même, $\mathbb{P}(X > b(t)) \leq e^{-t}$.
D'où

$$\forall t \geq 0, \quad \mathbb{P}(\gamma^*(X) > t) \leq 2e^{-t}. \tag{VII.21}$$

Enfin, une intégration par partie donne, en utilisant (VII.21) en $(*)$:

$$\begin{aligned} \mathbb{E}\left[e^{\varepsilon\gamma^*(X)}\right] &= \int_{-\infty}^{+\infty} e^t \mathbb{P}(\gamma^*(X) > t/\varepsilon)\, dt = \int_{-\infty}^{0} e^t\, dt + \int_0^{+\infty} e^t \mathbb{P}(\gamma^*(X) > t/\varepsilon)\, dt \\ &\overset{*}{\leq} 1 + 2\int_0^{+\infty} e^{(1-1/\varepsilon)t}\, dt = \frac{1+\varepsilon}{1-\varepsilon}. \end{aligned}$$

□

Démonstration du lemme VII.18. Soit X une variable aléatoire de loi μ. Notons Λ_φ la Log-Laplace de $\varphi(X)$, alors (VII.19) exprime que $\Lambda_\varphi(s) \leq \theta(a|s|)$, ce qui entraîne, en prenant les conjuguées convexes que $\forall t \in \mathbb{R},\ \theta^*\left(\frac{|t|}{a}\right) \leq \Lambda_\varphi^*(t)$. Par conséquent, d'après le lemme VII.17, on a pour tout $\varepsilon \in [0, 1[$:

$$\mathbb{E}\left[e^{\varepsilon\theta^*\left(\frac{|\varphi(X)|}{a}\right)}\right] \leq \mathbb{E}\left[e^{\varepsilon\Lambda_\varphi^*(\varphi(X))}\right] \leq \frac{1+\varepsilon}{1-\varepsilon}.$$

Or, θ^* étant convexe, on a pour tout $t \in \mathbb{R}$: $\theta^*(\varepsilon|t|) \leq \varepsilon\theta^*(t)$, et donc

$$\mathbb{E}\left[e^{\theta^*\left(\varepsilon\frac{|\varphi(X)|}{a}\right)}\right] \leq \frac{1+\varepsilon}{1-\varepsilon}.$$

Enfin

$$\frac{1+\varepsilon}{1-\varepsilon} \leq r \Leftrightarrow \varepsilon \leq \frac{r-1}{r+1},$$

donc

$$\mathbb{E}\left[e^{\theta^*\left(\frac{(r-1)|\varphi(X)|}{(r+1)a}\right)}\right] \leq r,$$

d'où

$$\|\varphi\|_{E\theta^*}^{(r)} \leq \frac{r+1}{r-1}a.$$

□

Démonstration de la proposition VII.15.
Soit $\varphi \in \Phi$; d'après la proposition VI.42, (VII.1) équivaut à

$$\forall s \geq 0, \quad \log \int_{\mathcal{X}} e^{s(\varphi - \langle \varphi, \mu \rangle)}\, d\mu \leq \theta(as).$$

Comme $-\varphi \in \Phi$, on a aussi

$$\forall s \leq 0, \quad \log \int_{\mathcal{X}} e^{s(\varphi - \langle \varphi, \mu \rangle)}\, d\mu \leq \theta(a|s|).$$

Ainsi, $\widetilde{\varphi}$ satisfait (VII.19) et donc, d'après le lemme VII.18, $\widetilde{\varphi} \in \mathbb{L}_{E\theta^*}(\mathcal{X}, \mu)$ et pour tout $r > 1$, $\|\widetilde{\varphi}\|^{(r)}_{E\theta^*} \leq \frac{r+1}{r-1} a$. □

Démonstration de la proposition VII.16.
D'après ce qui précède, pour toute fonction φ 1-Lipschitzienne bornée, on a pour tout $\varepsilon \in [0, 1[$

$$\int_{\mathcal{X}} \exp \varepsilon \theta^*(|\varphi(x) - \langle \varphi, \mu \rangle|/a)\, d\mu(x) \leq \frac{1+\varepsilon}{1-\varepsilon}.$$

En utilisant la continuité à gauche de θ^*, un argument d'approximation et le théorème de Fatou, on déduit que cette inégalité reste vraie pour toute fonction φ 1-Lipschitzienne non bornée. En particulier, pour tout $x_0 \in \mathcal{X}$ et pour tout $\varepsilon \in [0, 1[$ on a :

$$\int_{\mathcal{X}} \exp \varepsilon \theta^*(|d(x, x_0) - \langle d(\,.\,, x_0), \mu \rangle|/a)\, d\mu(x) \leq +\infty$$

Or, en notant $m = \langle d(\,.\,, x_0), \mu \rangle$, on a

$$\begin{aligned} \iint_{\mathcal{X}^2} e^{3\varepsilon\,\theta^*(d(\,.\,,\,.\,)/3a)} d\mu^{\otimes 2} &\overset{(i)}{\leq} \iint_{\mathcal{X}^2} \exp 3\varepsilon\,\theta^* \left(\frac{d(x, x_0) - m}{3a} + \frac{d(y, x_0) - m}{3a} + \frac{2m}{3a} \right) d\mu^2(x, y) \\ &\overset{(ii)}{\leq} \left(\int_{\mathcal{X}} \exp \varepsilon \theta^*(|d(x, x_0) - m|/a)\, d\mu(x) \right)^2 e^{\varepsilon\theta^*(2m/a)} < +\infty \end{aligned}$$

où (i) vient de l'inégalité triangulaire et de la croissance de θ^* et (ii) de la convexité de θ^*. Il suffit de prendre $\varepsilon = 1/3$, pour obtenir le résultat. □

VII.3 Conditions suffisantes pour une I.T.C. convexe. Critères intégraux.

Dans cette section, nous allons voir que les propositions VII.15 et VII.16 admettent des réciproques partielles dans le cas où $\underline{\theta \in \mathcal{C}_{quad}}$, hypothèse que nous ferons dans toute cette section.

VII.3.1 Majoration de la transformée de Laplace d'une variable aléatoire de $\mathbb{L}_{E\theta^*}(\mathcal{X}, \mu)$.

Les résultats que nous allons exposer maintenant sont issus du travail de Kozachenko et Ostrovski (voir [39] et [10] p. 63-68). Commençons par une

Définition VII.22. Nous dirons que φ vérifie la propriété $\mathrm{Sub}_\theta(\mathcal{X}, \mu)$ si, et seulement si, $\langle \varphi, \mu \rangle = 0$ et

$$\exists a \geq 0, \quad \forall s \in \mathbb{R}, \quad \log \int_{\mathcal{X}} e^{s\varphi}\, d\mu \leq \theta(a|s|) \tag{VII.23}$$

Clairement, une fonction mesurable φ telle que $\langle \varphi, \mu \rangle = 0$ vérifie $\mathrm{Sub}_\theta(\mathcal{X}, \mu)$ si, et seulement si,

$$\beta_\theta^1(\varphi) = \sup_{s \neq 0} \frac{\theta^{-1}\left(\log \int_{\mathcal{X}} e^{s\varphi}\, d\mu\right)}{|s|} < +\infty,$$

et dans ce cas, on voit facilement que $\beta_\theta^1(\varphi)$ est le plus petit a pour lequel (VII.23) est vérifiée.
La proposition suivante est immédiate :

Proposition VII.24. *μ satisfait $T_\Phi(\theta^*, a)$ si et seulement si pour toute $\varphi \in \Phi$, $\beta_\theta^1(\widetilde{\varphi}) \leq a$.*

Avec ces nouvelles notations, le lemme VII.18 s'énonce :

$$\left(\beta_\theta^1(\varphi) < +\infty \qquad \text{et} \qquad \langle \varphi, \mu \rangle = 0\right) \Rightarrow \frac{r-1}{r+1}\|\varphi\|_{E\theta^*}^{(r)} \leq \beta_\theta^1(\widetilde{\varphi}).$$

L'outil principal de cette section est le théorème suivant dû à Kozachenko et Ostrovski :

Théorème VII.25. *Il existe une constante m_θ ne dépendant que de la fonction θ, telle que*

$$\forall \varphi \in \mathbb{L}_{E\theta^*}(\mathcal{X}, \mu) \text{ telle que } \langle \varphi, \mu \rangle = 0, \quad \beta_\theta^1(\varphi) \leq \sqrt{r} m_\theta \|\varphi\|_{E\theta^*}^{(r)}.$$

On peut prendre :

$$m_\theta = e \max\left(\frac{1}{\theta^{-1}(2)\sqrt{c_\theta(1-u)}}, \frac{1}{u}\right),$$

où $u \in [0, 1[$ est tel que

$$\frac{u}{\sqrt{1-u}} \leq s_\theta \sqrt{c_\theta} \qquad \text{et} \qquad \frac{u^3}{1-u} \leq 2$$

Remarque VII.26.

On peut montrer (voir [10] thm 4.1) que β^1_θ est une norme sur $\mathbb{L}^0_{E\theta^*}(\mathcal{X},\mu) = \{\varphi \in \mathbb{L}_{E\theta^*}(\mathcal{X},\mu), \quad \langle \varphi, \mu\rangle = 0\}$, qui est donc, d'après le théorème VII.25, équivalente à la norme de Luxembourg $\|\,.\,\|^{(r)}_{E\theta^*}$.

Pour démontrer le théorème VII.25, nous allons introduire la quantité intermédiaire suivante :

$$\beta^2_\theta(\varphi) = \sup_{k\geq 2}\left\{\frac{\theta^{-1}(k)}{k}\|\varphi\|_k\right\} \qquad \text{avec} \qquad \|\varphi\|_k = \left[\int_{\mathcal{X}} |\varphi|^k \, d\mu\right]^{\frac{1}{k}}$$

Proposition VII.27. *Si $\varphi \in \mathbb{L}_{E\theta^*}(\mathcal{X},\mu)$, alors $\beta^2_\theta(\varphi) \leq \sqrt{r}\|\varphi\|^{(r)}_{E\theta^*}$.*

Cette proposition est immédiate au vu du lemme suivant :

Lemme VII.28. *Pour toute $\varphi \in \mathbb{L}_{E\theta^*}(\mathcal{X},\mu)$, on a pour tout $k \geq 1$:*

$$\|\varphi\|_k \leq \frac{r^{1/k}k}{\theta^{-1}(k)}\|\varphi\|^{(r)}_{E\theta^*} \tag{VII.29}$$

Démonstration du lemme VII.28. Si $k \geq 1$, alors, pour tout $x \geq 0$, on a

$$\begin{aligned} x^k e^{-\theta^*(x)} &= x^k e^{-\sup_{s\geq 0}\{sx-\theta(s)\}} = \inf_{s\geq 0}\left\{x^k e^{\theta(s)-sx}\right\} \\ &\leq \sup_{x\geq 0}\inf_{s\geq 0}\left\{x^k e^{\theta(s)-sx}\right\} \leq \inf_{s\geq 0} e^{\theta(s)} \sup_{x\geq 0}\left\{x^k e^{-sx}\right\}. \end{aligned}$$

Or, on voit facilement que pour $s > 0$, $\sup_{x\geq 0}\left\{x^k e^{-sx}\right\} = \left(\dfrac{k}{es}\right)^k$. En particulier, en prenant $s = \theta^{-1}(k)$, on a

$$x^k e^{-\theta^*(x)} \leq e^k\left(\frac{k}{e\theta^{-1}(k)}\right)^k = \left(\frac{k}{\theta^{-1}(k)}\right)^k.$$

Ainsi,

$$\forall x \geq 0, \quad x^k \leq \left(\frac{k}{\theta^{-1}(k)}\right)^k e^{\theta^*(x)}.$$

On en déduit, en prenant $x = \frac{|\varphi|}{\lambda}$, avec $\lambda > 0$ puis en intégrant par rapport à μ que

$$\|\varphi\|_k \leq \lambda\frac{k}{\theta^{-1}(k)}\left[\int_{\mathcal{X}} e^{\theta^*\left(\frac{|\varphi|}{\lambda}\right)}\, d\mu\right]^{1/k}.$$

Donc en prenant $\lambda = \|\varphi\|^{(r)}_{E\theta^*}$, on obtient (VII.29). □

Démonstration du théorème VII.25.
Grâce à la proposition VII.27, il suffit de démontrer l'inégalité $\beta_\theta^1(\varphi) \leq m_\theta \beta_\theta^2(\varphi)$.
Une majoration préliminaire :

$$\begin{aligned}
\int_{\mathcal{X}} e^{s\varphi}\,d\mu &= 1 + \sum_{k=2}^{+\infty} \frac{s^k}{k!} \int_{\mathcal{X}} \varphi^k\,d\mu \leq 1 + \sum_{k=2}^{+\infty} \frac{|s|^k}{k!} \|\varphi\|_k^k \\
&= 1 + \sum_{k=2}^{+\infty} \frac{1}{k!} \left(\frac{|s|k}{\theta^{-1}(k)}\right)^k \left(\|\varphi\|_k \frac{\theta^{-1}(k)}{k}\right)^k \\
&\leq 1 + \sum_{k=2}^{+\infty} \frac{1}{k!} \left(\frac{|s|k}{\theta^{-1}(k)}\right)^k \beta_\theta^2(\varphi)^k
\end{aligned}$$

Comme $\dfrac{k^k}{k!} \leq e^k$, on a, en posant $m = e\beta_\theta^2(\varphi)$

$$\int_{\mathcal{X}} e^{s\varphi}\,d\mu \leq 1 + \sum_{k=2}^{+\infty} \left(\frac{m|s|k}{\theta^{-1}(k)}\right)^k. \tag{VII.30}$$

Majoration pour les petites valeurs de s *:*
Dans toute la suite de la démonstration, u désignera un nombre réel appartenant à $[0,1[$ tel que :

$$\frac{u}{\sqrt{1-u}} \leq s_\theta \sqrt{c_\theta} \qquad \text{et} \qquad \frac{u^3}{1-u} \leq 2. \tag{VII.31}$$

Posons $s_1 = \dfrac{u\theta^{-1}(2)}{m}$. Pour $|s| \leq s_1$, on a, d'après (VII.30)

$$\begin{aligned}
\int_{\mathcal{X}} e^{s\varphi}\,d\mu &\leq 1 + \sum_{k=2}^{+\infty} \left(\frac{m|s|}{\theta^{-1}(2)}\right)^k = 1 + \frac{\left(\frac{m|s|}{\theta^{-1}(2)}\right)^2}{1 - \frac{m|s|}{\theta^{-1}(2)}} \leq 1 + \frac{\left(\frac{m|s|}{\theta^{-1}(2)}\right)^2}{1-u} \\
&= 1 + c_\theta \left[\frac{m|s|}{\theta^{-1}(2)\sqrt{c_\theta(1-u)}}\right]^2.
\end{aligned}$$

Or

$$\frac{m|s|}{\theta^{-1}(2)\sqrt{c_\theta(1-u)}} \leq \frac{ms_1}{\theta^{-1}(2)\sqrt{c_\theta(1-u)}} = \frac{u}{\sqrt{c_\theta(1-u)}} \overset{(*)}{\leq} s_\theta,$$

$(*)$ venant de (VII.31).
Donc, en posant $c_1 = \dfrac{m}{\theta^{-1}(2)\sqrt{c_\theta(1-u)}}$, on a pour $|s| \leq s_1$

$$\int_{\mathcal{X}} e^{s\varphi}\,d\mu \leq 1 + \theta(c_1|s|) \leq \exp\theta(c_1|s|). \tag{VII.32}$$

Majoration pour $|s| \geq s_1$:
Pour tout $|s| \geq s_1$, soit k_s l'unique entier ≥ 2, tel que :

$$\frac{m|s|}{\theta^{-1}(k_s)} \geq u \qquad \text{et} \qquad \frac{m|s|}{\theta^{-1}(k_s+1)} < u \tag{VII.33}$$

Posons

$$A_1(s) = \sum_{k=2}^{k_s} \left(\frac{m|s|}{\theta^{-1}(k)}\right)^k \qquad \text{et} \qquad A_2(s) = \sum_{k=k_s+1}^{+\infty} \left(\frac{m|s|}{\theta^{-1}(k)}\right)^k.$$

Tout d'abord, d'après (VII.33), $k_s \leq \theta\left(\frac{m|s|}{u}\right)$, donc pour tout $2 \leq k \leq k_s, \theta\left(\frac{m|s|}{u}\right)\frac{1}{k} \geq 1$.
Par conséquent, θ^{-1} étant concave et croissante, on a pour tout $2 \leq k \leq k_s$

$$\theta^{-1}(k) \geq \theta^{-1}\left(\frac{k}{\theta\left(\frac{m|s|}{u}\right)}\theta(m|s|)\right) \geq \frac{k}{\theta\left(\frac{m|s|}{u}\right)}\theta^{-1}(\theta(m|s|)) = \frac{km|s|}{\theta\left(\frac{m|s|}{u}\right)}.$$

On en déduit que pour tout $2 \leq k \leq k_s$,

$$\frac{m|s|}{\theta^{-1}(k)} \leq \frac{\theta\left(\frac{m|s|}{u}\right)}{k}.$$

D'où

$$A_1(s) \leq \sum_{k=2}^{k_s} \frac{\theta\left(\frac{m|s|}{u}\right)^k}{k^k} \leq \sum_{k=2}^{k_s} \frac{\theta\left(\frac{m|s|}{u}\right)^k}{k!}. \tag{VII.34}$$

Par ailleurs,

$$A_2(s) \overset{(i)}{\leq} \sum_{k=k_s+1}^{+\infty} u^k = \frac{u^{k+1}}{1-u} \leq \frac{u^3}{1-u} \overset{(ii)}{\leq} 2 \leq k_s \overset{(iii)}{\leq} \theta\left(\frac{m|s|}{u}\right). \tag{VII.35}$$

où (i) et (iii) découlent de (VII.33) et (ii) de (VII.31).
Finalement, d'après (VII.30), (VII.34) et (VII.35), on a

$$\begin{aligned}\int_{\mathcal{X}} e^{s\varphi}\, d\mu &\leq 1 + A_1(s) + A_2(s) \leq 1 + \sum_{k=2}^{k_s} \frac{\theta\left(\frac{m|s|}{u}\right)^k}{k!} + \theta\left(\frac{m|s|}{u}\right) \\ &= \sum_{k=0}^{k_s} \frac{\theta\left(\frac{m|s|}{u}\right)^k}{k!} \leq \exp\theta\left(\frac{m|s|}{u}\right)\end{aligned}$$

Ainsi, d'après (VII.32), pour tout $s \in \mathbb{R}$,

$$\int_{\mathcal{X}} e^{s\varphi}\, d\mu \leq \exp\theta(c_2|s|),$$

avec

$$c_2 = e\beta_\theta^2(\varphi) \max\left(\frac{1}{\theta^{-1}(2)\sqrt{c_\theta(1-u)}}, \frac{1}{u}\right),$$

ce qui entraîne $\beta_\theta^1(\varphi) \leq m_\theta \beta_\theta^2(\varphi)$. $\square$

VII.3.2 Applications aux I.T.C.

Grâce au théorème VII.25 et à la proposition VII.24, on déduit sans peine le

Théorème VII.36. *Soit $\theta \in \mathcal{C}_{quad}$, il y a équivalence entre*

1. *Il existe $a > 0$ tel que μ satisfait $T_\Phi(\theta^*, a)$*
2. *$\widetilde{\Phi} = \{\varphi - \langle \varphi, \mu\rangle, \varphi \in \Phi\}$ est une partie bornée de $\mathbb{L}_{E\theta^*}(\mathcal{X}, \mu)$.*

Plus précisément,

$$\mu \textit{ satisfait } T_\Phi(\theta^*, a) \Rightarrow \left(\forall \varphi \in \Phi, \quad \|\varphi - \langle \varphi, \mu\rangle\|_{E\theta^*}^{(r)} \leq \frac{r+1}{r-1} a.\right)$$

et

$$\left(\forall \varphi \in \Phi, \quad \|\varphi - \langle \varphi, \mu\rangle\|_{E\theta^*}^{(r)} \leq M.\right) \Rightarrow \mu \textit{ satisfait } T_\Phi(\theta^*, \sqrt{r} m_\theta M) \qquad \text{(VII.37)}$$

où m_θ est la constante définie à la proposition VII.25.

De même, dans le cas d'un coût métrique, on a le

Théorème VII.38. *Soient $(\mathcal{X}, d)$ un espace polonais et $\theta \in \mathcal{C}_{quad}$. Il y a équivalence entre*

1. *Il existe $a > 0$ tel que μ satisfait $T_d(\theta^*, a)$.*
2. *Il existe $b > 0$ tel que $\iint_{\mathcal{X}^2} \exp\theta^*\left(\frac{d(x,y)}{b}\right) d\mu(x)d\mu(y) < +\infty$.*

Plus précisément,

$$\mu \textit{ satisfait } T_d(\theta^*, a) \Rightarrow \iint_{\mathcal{X}^2} \exp\theta^*\left(\frac{d(x,y)}{3a}\right) d\mu(x)d\mu(y) < +\infty$$

et

$$2. \Rightarrow \mu \textit{ satisfait } T_d(\theta^*, \sqrt{r} m_\theta M) \qquad \text{(VII.39)}$$

avec $M := \|d(.,.)\|_{\mathbb{L}_{E\theta^*(\mathcal{X}^2, \mu^2)}}^{(r)}$

Démonstration. Il suffit de montrer que pour toute φ 1-Lipschitzienne pour $d(.,.)$, on a

$$\|\varphi - \langle \varphi, \mu \rangle\|_{E\theta^*}^{(r)} \leq \|d(.,.)\|_{\mathbb{L}_{E\theta^*(\mathcal{X}^2,\mu^2)}}^{(r)}.$$

Or, pour tout $\lambda > 0$, on a

$$\begin{aligned}\int e^{\theta^*(|\varphi(x)-\langle\varphi,\mu\rangle|/\lambda)}\, d\mu(x) &\overset{(i)}{\leq} \iint_{\mathcal{X}^2} e^{\theta^*(|\varphi(x)-\varphi(y)|/\lambda)} d\mu(x)d\mu(y) \\ &\overset{(ii)}{\leq} \iint_{\mathcal{X}^2} e^{\theta^*(d(x,y)/\lambda)} d\mu(x)d\mu(y)\end{aligned}$$

On obtient (i) grâce à l'inégalité de Jensen appliquée à la fonction convexe $U(x) = \exp(\theta^*(|x|))$ et (ii) vient du caractère 1-Lipschitzien de φ. □

VII.4 Exemples et estimation des constantes.

VII.4.1 Estimations des normes de jauge.

Le lemme suivant donne une majoration élémentaire des normes de Luxembourg intervenant dans les résultats précédents.

Lemme VII.40. *Soit $\theta \in \mathcal{C}_{quad}$, et $r > 1$.*

1. *Si* $\mathrm{dom}\,\theta^* = \mathbb{R}$*, alors pour toute $\varphi \in \mathbb{L}_{E\theta^*}(\mathcal{X}, \mu)$,*

$$\forall \delta > 0, \quad \|\varphi\|_{E\theta^*}^{(r)} \leq \max\left(\frac{1}{\delta}, \frac{\log \int_{\mathcal{X}} \exp(\theta^*(\delta|\varphi|))d\mu}{\delta \log(r)}\right)$$

2. *Si* $\mathrm{dom}\,\theta^*$ *est majoré, alors $\mathbb{L}_{E\theta^*}(\mathcal{X}, \mu) = \mathbb{L}_\infty(\mathcal{X}, \mu)$ et*

$$a^{-1}\|\varphi\|_\infty \leq \|\varphi\|_{E\theta^*}^{(r)} \leq r_{\theta^*}^{-1}\|\varphi\|_\infty$$

avec a la borne supérieure de $\mathrm{dom}\,\theta^*$ *et* $r_{\theta^*} = \sup\{x : \theta^*(x) \leq \log(r)\}$.

Démonstration.
(1) Posons $\lambda = \|\varphi\|_{\theta^*}^{(r)}$. Si $\lambda \leq \dfrac{1}{\delta}$ ou si $\displaystyle\int_{\mathcal{X}} \exp\theta^*(\delta|\varphi|)\, d\mu = +\infty$, il n'y a rien à montrer.

Supposons donc que $\lambda \geq \dfrac{1}{\delta}$ et que $\displaystyle\int_{\mathcal{X}} \exp\theta^*(\delta|\varphi|)\, d\mu < +\infty$.
On a alors

$$r^{\lambda\delta} \overset{(i)}{=} \left[\int_{\mathcal{X}} \exp\theta^*\left(\frac{|\varphi|}{\lambda}\right) d\mu\right]^{\lambda\delta} \overset{(ii)}{\leq} \int_{\mathcal{X}} \exp\lambda\delta\theta^*\left(\frac{|\varphi|}{\lambda}\right) d\mu \overset{(iii)}{\leq} \int_{\mathcal{X}} \exp\theta^*\left(\delta|\varphi|\right) d\mu$$

où (i) vient de la définition de la norme de jauge, (ii) de l'inégalité de Jensen, et (iii) de l'inégalité $\theta^*(|x|/M) \leq \theta^*(|x|)/M$, pour tout $M \geq 1$.

(2) Tout d'abord,

$$\left(\int_{\mathcal{X}} e^{\theta^*(|\varphi|/\lambda)}\, d\mu < +\infty\right) \Rightarrow (|\varphi| \leq a\lambda \quad \mu \text{ p.s.})$$

Ainsi, $\mathbb{L}_{E\theta^*}(\mathcal{X},\mu) \subset \mathbb{L}_\infty(\mathcal{X},\mu)$, et en prenant $\lambda = \|\varphi\|^{(r)}_{E\theta^*}$, on a $\|\varphi\|_\infty \leq a\|\varphi\|^{(r)}_{E\theta^*}$. Par ailleurs,

$$\int_{\mathcal{X}} e^{\theta^*(|\varphi|/\lambda)}\, d\mu \leq e^{\theta^*(\|\varphi\|_\infty/\lambda)}$$

Donc en prenant $\lambda = \dfrac{\|\varphi\|_\infty}{r_{\theta^*}}$, le membre de droite est majoré par r et on en déduit que

$$\mathbb{L}_\infty(\mathcal{X},\mu) \subset \mathbb{L}_{E\theta^*}(\mathcal{X},\mu) \qquad \text{et} \qquad \|\varphi\|^{(r)}_{E\theta^*} \leq \frac{\|\varphi\|_\infty}{r_{\theta^*}}.$$

□

Remarque VII.41. Il est facile de voir que si $\theta \in \mathcal{C}_{quad}$, $\mathrm{dom}\ \theta^*$ est borné si, et seulement si, $\mathrm{dom}\ \theta = \mathbb{R}$ et $\lim\limits_{s\to+\infty} \dfrac{\theta(s)}{s} = a < +\infty$.

VII.4.2 Exemples.

Nous allons étudier les I.T.C. associées à la fonction : $\theta_2(s) = \begin{cases} \frac{s^2}{2} & \text{si } s \in \mathbb{R}^+ \\ +\infty & \text{sinon} \end{cases}$ et donner, dans ce cas particulier, un contrôle plus approprié des constantes.
Un calcul immédiat donne $\theta_2^*(t) = \begin{cases} \frac{t^2}{2} & \text{si } t \in \mathbb{R}^+ \\ 0 & \text{sinon} \end{cases}$. Dans un premier temps nous allons voir comment raffiner l'approche de Djellout, Guillin et Wu pour obtenir les bornes de Bolley et Villani, à un facteur numérique près. Nous aurons besoin de la proposition suivante :

Proposition VII.42. *Soit X une variable aléatoire centrée telle que $\mathbb{E}[e^{\delta X^2}] < +\infty$ pour un certain $\delta > 0$. Alors, pour tout $s \geq 0$,*

$$\mathbb{E}\left[e^{sX}\right] \leq \exp\theta_2\left(\sqrt{2}scM\right),$$

avec $c = \begin{cases} 1 & \textit{si } X \textit{ est symétrique,} \\ \sqrt[4]{3.1} & \textit{sinon.} \end{cases}$, et $M = \inf\left\{\lambda \geq 0 : \mathbb{E}\left[\exp\left(\frac{X^2}{2\lambda^2}\right)\right] \leq e\right\}$.
De plus, on a la majoration

$$\forall\delta > 0, \quad M \leq \frac{1}{\delta}\sqrt{1 + \log\mathbb{E}\left[e^{\frac{\delta^2 X^2}{2}}\right]}. \qquad \text{(VII.43)}$$

Démonstration. Tout d'abord, il est démontré dans [10], page 7, que pour tout $s \geq 0$, on a

$$\mathbb{E}\left[e^{sX}\right] \leq \sum_{k=0}^{+\infty} \frac{(sc)^{2k}\mathbb{E}\left[X^{2k}\right]}{(2k)!},$$

c valant 1 dans le cas où la variable X est symétrique, et $\sqrt[4]{3.1}$ dans le cas contraire.
En posant $\beta(X) = \sup_{k\geq 1}\left(\sqrt[2k]{\frac{2^k.k!\mathbb{E}[X^{2k}]}{(2k)!}}\right)$, on a clairement

$$\mathbb{E}\left[e^{sX}\right] \leq \exp\theta_2\left(sc\beta(X)\right).$$

Montrons que $\beta(X) \leq \sqrt{2}M$:
En utilisant l'inégalité $x^{2k} \leq \left(\frac{2k}{e}\right)^k e^{\frac{x^2}{2}}$, on en déduit

$$\mathbb{E}\left[X^{2k}\right] \leq e\left(\frac{2k}{e}\right)^k M^{2k}$$

(en particulier, $\sqrt{\mathbb{E}[X^2]} \leq \sqrt{2}M$). Par conséquent, pour tout $k \geq 1$,

$$\frac{2^k.k!\mathbb{E}\left[X^{2k}\right]}{(2k)!} \leq e.\frac{2^{2k}.k!.k^k}{(2k)!e^k}M^{2k}.$$

En utilisant la formule de Stirling, *ie*

$$\forall p \geq 1, \quad \exists|\theta_p| \leq \frac{1}{12p}, \quad p! = \sqrt{2\pi p}.p^p.e^{-p+\theta_p},$$

on trouve facilement

$$\frac{2^k.k!\mathbb{E}\left[X^{2k}\right]}{(2k)!} \leq \frac{e^{1+\frac{1}{8k}}}{\sqrt{2}}M^{2k},$$

puis pour $k \geq 2$,

$$\sqrt[2k]{\frac{2^k.k!\mathbb{E}\left[X^{2k}\right]}{(2k)!}} \leq \frac{e^{\frac{17}{64}}}{2^{\frac{1}{8}}}M \leq \sqrt{2}M.$$

Montrons l'inégalité

$$M \leq \sqrt{1+\log\mathbb{E}\left[e^{\frac{X^2}{2}}\right]}. \tag{VII.44}$$

Si $M \leq 1$ est vraie. Supposons $M > 1$; on a

$$e^{M^2} = \mathbb{E}\left[e^{\frac{X^2}{2M^2}}\right]^{M^2} \leq \mathbb{E}\left[e^{\frac{X^2}{2}}\right].$$

Donc $M^2 \leq \log\mathbb{E}\left[e^{\frac{X^2}{2}}\right] \leq 1+\log\mathbb{E}\left[e^{\frac{X^2}{2}}\right]$. On obtient, ensuite (VII.43) en appliquant (VII.44) à la variable aléatoire δX. □

On en déduit le

Corollaire VII.45.

1. *Si* $\widetilde{\Phi}$ *est une partie de* $\mathbb{L}_{E\theta_2^*}(\mathcal{X},\mu)$*,*

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \|\nu-\mu\|_\Phi^* \le \sqrt{2}\sqrt[4]{3.1}M\sqrt{2\,\mathrm{H}\,(\nu|\,\mu)} \tag{VII.46}$$

où $M=\sup\Big\{\|\varphi-\langle\varphi,\mu\rangle\|_{E\theta_2^*}^{(e)}, \quad \varphi\in\Phi\Big\}$.

2. *Si* $(\mathcal{X},d)$ *est un espace polonais et s'il existe* $\delta>0$ *tel que*

$$\iint_{\mathcal{X}^2} e^{\delta d^2(x,y)}\,d\mu(x)d\mu(y) < +\infty,$$

alors,

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_d(\nu,\mu) \le \sqrt{2}\|d(\,.\,,\,.\,)\|_{\mathbb{L}_{E\theta_2^*}(\mathcal{X}^2,\mu^2)}^{(e)}\sqrt{2\,\mathrm{H}\,(\nu|\,\mu)} \tag{VII.47}$$

De plus,

$$\|d(\,.\,,\,.\,)\|_{\mathbb{L}_{E\theta_2^*}(\mathcal{X}^2,\mu^2)}^{(e)} \le \frac{1}{\delta}\sqrt{1+\log\int_{\mathcal{X}^2} e^{\delta^2 d(x,y)^2}\,d\mu(x)d\mu(y)}.$$

Pour terminer cette section, nous allons voir comment obtenir directement les bornes de Bolley et Villani sans passer par l'estimation des normes de jauge. Nous aurons besoin du lemme suivant :

Proposition VII.48. *Si* X *une variable aléatoire symétrique et centrée telle que* $\mathbb{E}[e^{X^2}]<+\infty$*, alors*

$$\forall s\ge 0, \quad \mathbb{E}\left[e^{sX}\right] \le \exp\left(\frac{(sM)^2}{2}\right),$$

avec $M=\sqrt{1+2\log\mathbb{E}\left[e^{X^2/2}\right]}$.

Démonstration.
Pour $s\le 1$, on a

$$\mathbb{E}\left[e^{sX}\right] \le 1+\sum_{k=1}^{+\infty}\frac{s^{2k}\mathbb{E}\left[X^{2k}\right]}{(2k)!} \overset{(i)}{\le} 1+\sum_{k=1}^{+\infty}\frac{s^{2k}\mathbb{E}\left[X^{2k}\right]}{2^k.k!} = \mathbb{E}\left[e^{s^2X^2/2}\right] \overset{(ii)}{\le} \mathbb{E}\left[e^{X^2/2}\right]^{s^2},$$

en utilisant l'inégalité $(2k)!\ge 2^k.k!$ en (i), et l'inégalité de Jensen en (ii).
Pour $s\ge 1$,

$$\mathbb{E}\left[e^{sX}\right] \le \mathbb{E}\left[e^{s^2/2+X^2/2}\right] \le e^{s^2/2}\mathbb{E}\left[e^{X^2/2}\right]^{s^2}.$$

Ainsi, pour tout $s \geq 0$, on a

$$\mathbb{E}\left[e^{sX}\right] \leq e^{s^2/2}\mathbb{E}\left[e^{X^2/2}\right]^{s^2} = \exp\left(\frac{(sM)^2}{2}\right),$$

avec $M = \sqrt{1 + 2\log \mathbb{E}\left[e^{X^2/2}\right]}$. □

On en déduit facilement le corollaire suivant.

Corollaire VII.49. *Soit $\chi : \mathcal{X} \to \mathbb{R}^+$, une fonction mesurable. Alors, pour toute $\nu \in \mathcal{P}(\mathcal{X})$,*

$$\|\chi\nu - \chi\mu\|_{VT} \leq \sqrt{1 + 4\log \int_{\mathcal{X}} e^{\chi(x)^2} d\mu(x)}\sqrt{2\,\mathrm{H}\left(\nu|\,\mu\right)}.$$

VII.5 I.T.C. convexes pour des fonctions de coût non métriques.

Dans cette section, nous allons utiliser les résultats des sections VII.2 et VII.3 pour étudier les I.T.C. associées à des coûts de transport de la forme $c(x, y) = q(d(x, y))$.

Dans toute la suite, $q : \mathbb{R}^+ \to \mathbb{R}^+$ sera une fonction convexe strictement croissante, et $(\mathcal{X}, d)$ un espace polonais. Nous poserons $c(x, y) = q(d(x, y))$ et nous noterons $\mathcal{T}_c$ le coût de transport optimal associé à c.

Le résultat principal de cette section est le théorème suivant :

Théorème VII.50. *Si $\theta \in \mathcal{C}_{quad}$ et si q satisfait la condition Δ_2, ie*

$$\exists K > 0, \quad \forall x \in \mathbb{R}^+, \quad q(2x) \leq Kq(x),$$

alors les deux propositions suivantes sont équivalentes :

1. $\exists a > 0, \quad \forall \nu \in \mathcal{P}(\mathcal{X}), \quad \theta^*\left(\frac{\mathcal{T}_c(\nu, \mu)}{a}\right) \leq \mathrm{H}\left(\nu|\,\mu\right),$
2. $\exists b > 0, \quad \iint_{\mathcal{X}^2} \exp\theta^*\left(\frac{c(x, y)}{b}\right) d\mu(x)d\mu(y) < +\infty.$

Pour démontrer le théorème VII.50, nous allons généraliser l'approche développée dans [5], en commençant par étendre l'inégalité (VII.11) à d'autres transformations convexes q que les fonctions puissances :

Proposition VII.51. *Soit $x_0 \in \mathcal{X}$, et posons pour tout $x \in \mathcal{X}$, $\chi(x) = \frac{1}{2}q(2d(x, x_0))$, alors*

$$\forall \nu \in \mathcal{P}(X), \quad q(\mathcal{T}_d(\nu, \mu)) \leq \mathcal{T}_c(\nu, \mu) \leq \|\chi\nu - \chi\mu\|_{VT}. \tag{VII.52}$$

Démonstration. Tout d'abord, pour tout $\pi \in \Pi(\nu,\mu)$, on a, d'après l'inégalité de Jensen, $q\left(\int_{\mathcal{X}^2} d(x,y)\,d\pi(x,y)\right) \leq \iint_{\mathcal{X}^2} q(d(x,y))\,d\pi(x,y)$; on en déduit immédiatement la première inégalité.
Pour tout $x, y \in \mathcal{X}$, on a en utilisant l'inégalité triangulaire et la convexité de q

$$\begin{aligned} c(x,y) &= q(d(x,y)) \leq q(d(x,x_0)+d(y,y_0)) \\ &\leq \frac{1}{2}\left[q(2d(x,x_0)) + q(2d(y,x_0))\right] \\ &= \chi(x) + \chi(y). \end{aligned}$$

Donc, $c(x,y) \leq d_\chi(x,y)$, et par conséquent, $\mathcal{T}_c(\nu,\mu) \leq \mathcal{T}_{d_\chi}(\nu,\mu) = \|\chi\nu - \chi\mu\|_{VT}$, (d'après la proposition VI.7). □

Démonstration du théorème VII.50.
Montrons que (1) entraîne (2). D'après l'inégalité VII.52, (1) implique que pour toute $\nu \in \mathcal{P}(\mathcal{X})$, on a

$$\theta^*\left(\frac{q\left(\mathcal{T}_d(\nu,\mu)\right)}{a}\right) \leq \mathrm{H}\left(\nu|\,\mu\right).$$

Comme $x \mapsto \theta^*\left(\frac{q(x)}{a}\right)$ est convexe s.c.i, le théorème VII.16 entraîne qu'il existe $\tilde{a} > 0$ tel que $\iint_{\mathcal{X}^2} \exp\theta^*\left(\frac{q(d(x,y)/\tilde{a})}{a}\right) d\mu^2(x,y) < +\infty$. Soit n un entier naturel tel que $2^n \geq \tilde{a}$; on a alors, en utilisant la condition Δ_2

$$\forall x \in \mathbb{R}^+,\ \ q(x) = q\left(\tilde{a}\frac{x}{\tilde{a}}\right) \leq q\left(2^n\frac{x}{\tilde{a}}\right) \leq K^n q\left(\frac{x}{\tilde{a}}\right).$$

Par conséquent,

$$\iint_{\mathcal{X}^2} \exp\theta^*\left(\frac{c(x,y)}{K^n a}\right) d\mu(x)d\mu(y) \leq \iint_{\mathcal{X}^2} \exp\theta^*\left(\frac{q(d(x,y)/\tilde{a})}{a}\right) d\mu(x)d\mu(y) < +\infty.$$

Montrons que (2) implique (1). D'après le théorème VII.38 appliqué à d_χ, il suffit de montrer qu'il existe $x_0 \in \mathcal{X}$ et $u > 0$ tels que

$$\iint_{\mathcal{X}^2} \exp\theta^*\left(\frac{d_\chi(\,.\,,\,.\,)}{u}\right) d\mu^2 \leq \iint_{\mathcal{X}^2} \exp\theta^*\left(\frac{q(2d(\,.\,,x_0)) + q(2d(\,.\,,x_0))}{2u}\right) d\mu^2 < +\infty.$$

Or, en utilisant une nouvelle fois la condition Δ_2 et la convexité de q, on voit sans peine que la dernière intégrale est majorée par $\left[\int_{\mathcal{X}} \exp\theta^*\left(\frac{K}{u}c(\,.\,,x_0)\right) d\mu\right]^2$. Mais, par hypothèse, $\iint_{\mathcal{X}^2} \exp\theta^*\left(\frac{c(x,y)}{b}\right) d\mu(x)d\mu(y) < +\infty$, donc en particulier, pour μ presque tout $x_0 \in \mathcal{X}$, $\int_{\mathcal{X}} \exp\theta^*\left(\frac{1}{b}c(\,.\,,x_0)\right) d\mu < +\infty$, d'où le résultat, en prenant $u = Kb$. □

ANNEXE A

Annexe du chapitre III

A.1 Preuve du lemme Propagation du chaos

Montrons le lemme suivant que nous avons utilisé dans l'introduction :

Lemme (Propagation du chaos). *Soit $\mathcal{X}$ un espace polonais, et pour tout $n \in \mathbb{N}^*$, soit μ^n une probabilité sur $\mathcal{X}^n$. On suppose que chaque μ^n est symétrique, ie pour toute permutation σ de $\{1, \ldots, n\}$, $\mu^n \circ f_\sigma^{-1} = \mu^n$, en notant $f_\sigma : (x_1, \ldots, x_n) \mapsto (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. Il y a équivalence entre les propositions suivantes :*

1. *La loi de $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ sous μ^n converge étroitement vers δ_{μ^*}.*
2. *Pour tout $k \in \mathbb{N}^*$ et pour toutes fonctions $f_1, \ldots, f_k$ continues bornées sur $\mathcal{X}$, on a*

$$\int_{\mathcal{X}^k} f_1(x_1)\cdots f_k(x_k)\, d\mu^n \xrightarrow[n\to+\infty]{} \int_{\mathcal{X}^k} f_1(x_1)\cdots f_k(x_k)\, d\mu^{*\otimes k}.$$

Démonstration. Montrons que 1 implique 2 :
Soit $f_1, \ldots, f_k \in C_b(\mathcal{X})$,

$$\begin{aligned}\left|\int_{\mathcal{X}^k} f_1(x_1)\cdots f_k(x_k)\, d\mu^n - \int_{\mathcal{X}^k} f_1(x_1)\cdots f_k(x_k)\, d\mu^{*\otimes k}\right| \\ &\leq \left|\int_{\mathcal{X}^k} f_1(x_1)\cdots f_k(x_k)\, d\mu^n - \int_{\mathcal{X}^n} \prod_{i=1}^k \langle L_n, f_i\rangle\, d\mu^n\right| \\ &+ \left|\int_{\mathcal{X}^n} \prod_{i=1}^k \langle L_n, f_i\rangle\, d\mu^n - \int_{\mathcal{X}^k} f_1(x_1)\cdots f_k(x_k)\, d\mu^{*\otimes k}\right|\end{aligned}$$

Le deuxième terme tend vers 0 par hypothèse ; reste à voir qu'il en est de même du premier. Or, celui ci peut s'écrire :

$$I = \left| \int_{X^k} \left[\frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} f_1(x_{\sigma(1)}) \cdots f_k(x_{\sigma(k)}) - \prod_{i=1}^{k} \left\langle \frac{1}{n} \sum_{j=1}^{n} \delta_{x_j}, f_i \right\rangle \right] d\mu^n \right|,$$

où $\mathfrak{S}_n$ désigne l'ensemble des permutations de $\{1, \ldots, n\}$.

Soit M un majorant des f_i, on a en notant $\mathfrak{F}(k,n)$ l'ensemble des applications de $\{1, \ldots, k\}$ dans $\{1, \ldots, n\}$:

$$\begin{aligned} I &= \left| \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} f_1(x_{\sigma(1)}) \cdots f_k(x_{\sigma(k)}) - \frac{1}{n^k} \sum_{\alpha \in \mathfrak{F}(k,n)} f_1(x_{\alpha(1)}) \cdots f_k(x_{\alpha(k)}) \right| \\ &\leq \left| \sum_{\substack{\alpha \in \mathfrak{F}(k,n) \\ \text{injectives}}} \left(\frac{(n-k)!}{n!} - \frac{1}{n^k} \right) f_1(x_{\alpha(1)}) \cdots f_k(x_{\alpha(k)}) \right| + \left| \frac{1}{n^k} \sum_{\substack{\alpha \in \mathfrak{F}(k,n) \\ \text{non injectives}}} f_1(x_{\alpha(1)}) \cdots f_k(x_{\alpha(k)}) \right| \\ &\leq M^k \left[\left(\frac{(n-k)!}{n!} - \frac{1}{n^k} \right) \frac{n!}{(n-k)!} + \frac{1}{n^k} \left(n^k - \frac{n!}{(n-k)!} \right) \right] = 2M^k \left(1 - \frac{n!}{n^k (n-k)!} \right), \end{aligned}$$

qui tend vers 0 quand $n \to \infty$.

Montrons que 2 implique 1 :
Notons $Q_n = \mathcal{L}_{\mu^n}(L_n)$. Pour montrer que Q_n converge étroitement vers δ_{μ^*}, il faut montrer que pour tout ouvert $\mathcal{O}$ de $\mathcal{P}(\mathcal{X})$, on a

$$\liminf_{n \to +\infty} Q_n(\mathcal{O}) \geq \delta_{\mu^*}(\mathcal{O}).$$

Cela revient à démontrer que pour tout ouvert $\mathcal{O}$ contenant μ^*, on a

$$Q_n(\mathcal{O}) \xrightarrow[n \to +\infty]{} 1. \tag{A.1}$$

Par définition de la topologie de la convergence étroite, il suffit de montrer que (A.1) est vraie pour $\mathcal{O}$ de la forme

$$\bigcap_{i=1}^{p} \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \left| \int_{\mathcal{X}} f_i \, d\nu - \int_{\mathcal{X}} f_i \, d\mu^* \right| < \alpha_i \right\},$$

avec $\alpha_i \in \mathbb{R}^+$ et $f_i \in C_b(\mathcal{X})$. Comme (A.1) est stable par intersection finie, il suffit de traiter le cas $p = 1$.
Or, si $f \in C_b(\mathcal{X})$, alors

$$\int_{\mathcal{X}^n} \langle L_n - \mu^*, f \rangle^2 \, d\mu^n = \frac{1}{n^2} \sum_{i,j} \int_{\mathcal{X}^n} f(x_i) f(x_j) \, d\mu^n - \frac{2}{n} \langle \mu^*, f \rangle \sum_i \int_{X^n} f(x_i) \, d\mu^n + \langle \mu^*, f \rangle^2$$

$$\frac{1}{n} \int_{\mathcal{X}^n} f(x_1)^2 \, d\mu^n + \frac{n-1}{n} \int_{\mathcal{X}^n} f(x_1) f(x_2) \, d\mu^n - 2 \langle \mu^*, f \rangle \int_{X^n} f(x_i) \, d\mu^n + \langle \mu^*, f \rangle^2$$

qui tend vers 0, d'après 2.
Grâce à l'inégalité de Markov, on en déduit que

$$Q_n \left\{ \nu : \left| \int_{\mathcal{X}} f \, d\nu - \int_{\mathcal{X}} f \, d\mu^* \right| < \alpha \right\} = \mu_n \left(\left| \int_{\mathcal{X}} f \, dL_n - \int_{\mathcal{X}} f \, d\mu^* \right| < \alpha \right) \xrightarrow[n \to +\infty]{} 1.$$

□

A.2 Contrôles non-asymptotiques pour le théorème de Sanov

A.2.1 Bornes supérieures exactes :

Le premier théorème de cette annexe est dû à I. Csiszár.

Théorème A.2 (Csiszár, [19] thm. 1)**.** *Soit A un ensemble convexe fermé de $\mathcal{P}_G(\mathcal{X})$. On suppose que $\mathrm{H}\left(A \middle| \mu\right) < +\infty$ et on note μ^* la I-projection généralisée de μ sur A.*
Si $\mu^{\otimes n}(L_n \in A) > 0$, alors pour tout $k \in \{1, \ldots, n\}$, on a

$$\mathrm{H}\left(\mu^n_{A,k} \middle| \mu^{*\otimes k}\right) \leq -\frac{1}{[n/k]} \log \left(\mu^{\otimes n}(L_n \in A) e^{n \mathrm{H}(A|\mu)} \right). \tag{A.3}$$

On en déduit immédiatement le corollaire suivant dont nous nous servirons dans la section III.4 :

Corollaire A.4. *Si A est un convexe fermé, tel que $\mathrm{H}\left(A \middle| \mu\right) < +\infty$ alors pour tout $n \geq 1$,*

$$\mu^{\otimes n}(L_n \in A) \leq e^{-n \mathrm{H}(A|\mu)}. \tag{A.5}$$

Démonstration. Tout d'abord,

$$\frac{d\mu^n_{A,n}}{d\mu^{\otimes n}} = \frac{\mathbb{1}_A(L_n)}{\mu^{\otimes n}(L_n \in A)}$$

et on calcule facilement

$$\mathrm{H}\left(\mu^n_{A,n} \middle| \mu^{\otimes n}\right) = -\log \mu^{\otimes n}(L_n \in A).$$

De plus, les marginales unidimensionnelles de $\mu^n_{A,n}$ étant toutes égales à μ^n_A, on a, d'après la proposition II.4

$$\mathrm{H}\left(\mu^n_{A,n} \middle| \mu^{\otimes n}\right) = \mathrm{H}\left(\mu^n_{A,n} \middle| (\mu^n_A)^{\otimes n}\right) + n \, \mathrm{H}\left(\mu^n_A \middle| \mu\right),$$

et d'autre part,

$$\mathrm{H}\left(\mu^n_{A,n} \middle| \mu^{*\otimes n}\right) = \mathrm{H}\left(\mu^n_{A,n} \middle| (\mu^n_A)^{\otimes n}\right) + n \, \mathrm{H}\left(\mu^n_A \middle| \mu^*\right).$$

On en déduit que

$$\mathrm{H}\left(\mu_{A,n}^n \middle| \mu^{\otimes n}\right) = \mathrm{H}\left(\mu_{A,n}^n \middle| \mu^{*\otimes n}\right) + n\left[\mathrm{H}\left(\mu_A^n \middle| \mu\right) - \mathrm{H}\left(\mu_A^n \middle| \mu^*\right)\right].$$

Admettons un instant que $\mu_A^n \in A$; alors, d'après la proposition II.26 , on a

$$\mathrm{H}\left(\mu_A^n \middle| \mu\right) - \mathrm{H}\left(\mu_A^n \middle| \mu^*\right) \geq \mathrm{H}\left(A \middle| \mu\right)$$

et donc

$$\mathrm{H}\left(\mu_{A,n}^n \middle| \mu^{\otimes n}\right) \geq \mathrm{H}\left(\mu_{A,n}^n \middle| \mu^{*\otimes n}\right) + n\,\mathrm{H}\left(A \middle| \mu\right).$$

Soit

$$-\log\left(\mu^{\otimes n}(L_n \in A)e^{n\,\mathrm{H}(A|\mu)}\right) \geq \mathrm{H}\left(\mu_{A,n}^n \middle| \mu^{*\otimes n}\right).$$

En appliquant encore une fois la proposition II.4 , on voit facilement que

$$\mathrm{H}\left(\mu_{A,n}^n \middle| \mu^{*\otimes n}\right) \geq [n/k]\,\mathrm{H}\left(\mu_{A,k}^n \middle| \mu^{*\otimes k}\right).$$

D'où le résultat. Pour finir, montrons que μ_A^n appartient à A. Pour cela, posons

$$\mathcal{M}_G(\mathcal{X}) = \left\{\nu \in \mathcal{M}(\mathcal{X}) : \forall g \in G,\ \int_{\mathcal{X}} |g|\, d|\nu| < +\infty\right\}.$$

L'ensemble $\mathcal{M}_G(\mathcal{X})$ sera muni de la G-topologie, *ie* la moins fine rendant continues les applications $\nu \mapsto \int_{\mathcal{X}} g\, d\nu$, avec $g \in G$. Pour cette topologie, $\mathcal{M}_G(\mathcal{X})$ est un espace vectoriel topologique localement convexe qui a pour dual topologique :

$$\mathcal{M}_G(\mathcal{X})' = \{\nu \mapsto \langle \nu, g\rangle : g \in G\}.$$

Par hypothèse, G contient l'ensemble $C_b(\mathcal{X})$ des applications continues bornées ; on en déduit facilement que $\mathcal{P}_G(\mathcal{X})$ est fermé, et que $\mathcal{M}_G(\mathcal{X})$ est séparé.
Si μ_A^n n'était pas dans A (qui est fermé dans $\mathcal{M}_G(\mathcal{X})$), il existerait, d'après le théorème de Hahn-Banach, une fonction $g \in G$ telle que :

$$\langle \mu_A^n, g\rangle < \inf\{\langle \nu, g\rangle : \nu \in A\} := \alpha.$$

Or,

$$\langle \mu_A^n, g\rangle = \frac{\mathbb{E}[\langle L_n^X, g\rangle \mathbb{1}_A(L_n^X)]}{\mathbb{P}(L_n^X \in A)} \geq \frac{\mathbb{E}[\alpha \mathbb{1}_A(L_n^X)]}{\mathbb{P}(L_n^X \in A)} = \alpha$$

- contradiction. □

Remarque A.6.
Dans [19], I. Csiszár, a établi l'inégalité (III.38), sans hypothèse topologique sur A, mais pour des ensembles A *presque complètement convexes* :

- Un ensemble A est dit *complètement convexe* si pour tout espace de probabilité, $(\Omega, \mathcal{A}, P)$ et tout noyau de transition $N : \Omega \to A$, la mesure de probabilité $N.P \in \mathcal{P}(\mathcal{X})$ définie par $N.P(B) = \int_\Omega N(\omega, B)\, dP(\omega)$, appartient à A.
- Un ensemble A est dit *presque complètement convexe* s'il existe une suite croissante A_n de sous-ensemble complètement convexes de A telle que

$$A \cap \mathcal{P}_f(\mathcal{X}) \subset \bigcup_n A_n,$$

où $\mathcal{P}_f(\mathcal{X})$ désigne les mesures de probabilité ne chargeant qu'un nombre fini de points.

A.2.2 Bornes inférieures exactes :

La proposition suivante, démontrée en exercice dans le livre de J.D. Deuschel et D.W. Stroock, donne une borne inférieure non-asymptotique pour le théorème de Sanov.

Proposition A.7. *Soient A une partie de $\mathcal{P}_G(\mathcal{X})$ telle que $\{x : L_n^x \in A\}$ est mesurable, $\nu \in \mathcal{P}_G(\mathcal{X})$, avec $\nu \ll \mu$ et $\nu^{\otimes n}(L_n \in A) > 0$. Alors,*

$$\frac{1}{n} \log\left(\mu^{\otimes n}(L_n \in A) e^{n\,\mathrm{H}(\nu|\mu)}\right) \geq -\,\mathrm{H}\left(\nu|\,\mu\right) \frac{\nu^{\otimes n}(L_n \in A^c)}{\nu^{\otimes n}(L_n \in A)} + \frac{1}{n} \log \nu^{\otimes n}(L_n \in A) - \frac{1}{ne\nu^{\otimes n}(L_n \in A)} \tag{A.8}$$

Démonstration.
Posons $h = \frac{d\nu^{\otimes n}}{d\mu^{\otimes n}}$, et $\widetilde{A} = \{x \in \mathcal{X}^n : L_n^x \in A \quad \text{et} \quad h(x) > 0\}$.
Alors,

$$\mu^{\otimes n}(L_n \in A) \geq \mu^{\otimes n}(\widetilde{A}) = \int_{\widetilde{A}} h(x)\, d\nu^{\otimes n}(x) = \nu^{\otimes n}(\tilde{A}) \frac{\int_{\widetilde{A}} e^{-\log h(x)}\, d\nu^{\otimes n}(x)}{\nu^{\otimes n}(\widetilde{A})}.$$

Donc, d'après l'inégalité de Jensen,

$$\log \mu^{\otimes n}(L_n \in A) \geq \log \nu^{\otimes n}(\widetilde{A}) - \frac{\int_{\widetilde{A}} \log h(x)\, d\nu^{\otimes n}}{\nu^{\otimes n}(\widetilde{A})}.$$

Comme $\mathrm{H}\left(\nu^{\otimes n}|\,\mu^{\otimes n}\right) = \int \log h(x)\, d\nu^{\otimes n}$, on en déduit que

$$\log \mu^{\otimes n}(L_n \in A) \geq \log \nu^{\otimes n}(\widetilde{A}) - \frac{\mathrm{H}\left(\nu^{\otimes n}|\,\mu^{\otimes n}\right)}{\nu^{\otimes n}(\widetilde{A})} + \frac{\int_{\widetilde{A}^c} \log h(x) h(x)\, d\mu^{\otimes n}}{\nu^{\otimes n}(\widetilde{A})} \tag{A.9}$$

Or, pour tout $x > 0,\ x\log x \geq -\frac{1}{e}$, donc

$$\frac{\int_{\widetilde{A}^c} \log h(x) h(x)\, d\mu^{\otimes n}}{\nu^{\otimes n}(\widetilde{A})} \geq -\frac{\mu^{\otimes n}(\widetilde{A})}{e\nu^{\otimes n}(\widetilde{A})} \geq -\frac{1}{e\nu^{\otimes n}(\widetilde{A})}. \tag{A.10}$$

Enfin, en reportant (A.10) dans (A.9) et en utilisant les relations suivantes :

$$\mathrm{H}\left(\nu^{\otimes n}\middle|\, \mu^{\otimes n}\right) = n\,\mathrm{H}\left(\nu\middle|\, \mu\right) \qquad \text{et} \qquad \nu^{\otimes n}(\widetilde{A}) = \nu^{\otimes n}(L_n \in A),$$

on obtient facilement (A.8). □

Considérons à présent le cas particulier d'un convexe C défini par des contraintes de type moment *ie*, C est de la forme

$$C = \left\{\nu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} \|F\|\, d\nu < +\infty \quad \text{et} \quad \int_{\mathcal{X}} F\, d\nu \in K\right\},$$

avec $F : \mathcal{X} \to B$ une application mesurable à valeurs dans un espace de Banach séparable muni de sa tribu borélienne et K un convexe fermé de B.
Pour tout $\varepsilon > 0$, nous poserons

$$C_\varepsilon = \left\{\nu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} \|F\|\, d\nu < +\infty \quad \text{et} \quad \int_{\mathcal{X}} F\, d\nu \in K^\varepsilon\right\},$$

où $K^\varepsilon = \{x \in B : d(x, K) \leq \varepsilon\}$.
Nous noterons Z_F la transformée de Laplace de μ_F, image de μ par F, et Λ_F, sa Log-Laplace.

Lemme A.11. *Si μ admet une I-projection μ^* sur C s'écrivant $\mu^* = \frac{e^{\langle\lambda^*, F\rangle}}{Z_F(\lambda^*)}\mu$, avec $\lambda^* \in B'$, alors pour tout $\varepsilon > 0$,*

$$-\frac{1}{n}\log\left(\mu^{\otimes n}(L_n \in C_\varepsilon)e^{n\,\mathrm{H}(\mu^*|\mu)}\right) \leq -\frac{1}{n}\log\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^n F(Y_i) - \int_{\mathcal{X}} F d\mu^*\right\| \leq \varepsilon\right) + \|\lambda^*\|\varepsilon. \tag{A.12}$$

avec $(Y_i)_i$ une suite de variables i.i.d de loi μ^*.

Démonstration.

$$\begin{aligned}
\mu^{\otimes n}(L_n \in C_\varepsilon) &= \int \mathbb{1}_{C_\varepsilon}(L_n^x)\frac{d\mu}{d\mu^*}(x_1)\cdots\frac{d\mu}{d\mu^*}(x_n)\, d\mu^{*\otimes n}(x) \\
&= \int \mathbb{1}_{C_\varepsilon}(L_n^x)\exp\left(-n\left\langle L_n, \log\frac{d\mu^*}{d\mu}\right\rangle\right) d\mu^{*\otimes n}(x) \\
&= e^{-n\,\mathrm{H}(\mu^*|\mu)}\int \mathbb{1}_{C_\varepsilon}(L_n^x)\exp\left(-n\left\langle L_n - \mu^*, \log\frac{d\mu^*}{d\mu}\right\rangle\right) d\mu^{*\otimes n}(x)
\end{aligned}$$

Or, $\log \frac{d\mu^*}{d\mu} = \langle \lambda^*, F \rangle - \Lambda_F(\lambda^*)$, et donc

$$\left\langle L_n - \mu^*, \log \frac{d\mu^*}{d\mu} \right\rangle = \left\langle \lambda^*, \frac{1}{n} \sum_{i=1}^n F(x_i) - \int_{\mathcal{X}} F \, d\mu^* \right\rangle.$$

Posons

$$\widetilde{C}_\varepsilon = \left\{ \nu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} \|F\| d\nu < +\infty \quad \text{et} \quad \left\| \int_{\mathcal{X}} F \, d\nu - \int_{\mathcal{X}} F \, d\mu^* \right\| \leq \varepsilon \right\} \subset C_\varepsilon,$$

on voit que

$$\begin{aligned} \mu^{\otimes n}(L_n \in C_\varepsilon) e^{n \,\mathrm{H}(\mu^* | \mu)} &\geq \int \mathbb{1}_{\widetilde{C}_\varepsilon}(L_n^x) e^{-n \left\langle \lambda^*, \frac{1}{n} \sum_{i=1}^n F(x_i) - \int_{\mathcal{X}} F \, d\mu^* \right\rangle} d\mu^{*\otimes n}(x) \\ &\geq e^{-n \|\lambda^*\| \varepsilon} \int \mathbb{1}_{\widetilde{C}_\varepsilon}(L_n^x) \, d\mu^{*\otimes n}(x) \\ &= e^{-n \|\lambda^*\| \varepsilon} \mathbb{P}\left(\left\| \frac{1}{n} \sum_{i=1}^n F(Y_i) - \int_{\mathcal{X}} F \, d\mu^* \right\| \leq \varepsilon \right) \end{aligned}$$

□

ANNEXE B

Preuve du théorème V.8

La preuve du théorème V.8 est contenue en plusieurs morceaux dans les articles de F. Gamboa et E. Gassiat ([34, 22, 35, 36]). Par soucis de clarté, nous donnons ci-dessous une preuve complète de ce théorème.

Nous aurons besoin du lemme suivant qui donne la convergence des solutions d'une suite de problèmes de minimisation de fonctions convexes (voir [60] pour des résultats plus généraux).

Lemme B.1. *Soit* $(H_n)_n$ *une suite de fonctions convexes définies sur* $\mathbb{R}^k$ *à valeurs dans* $\mathbb{R} \cup \{+\infty\}$ *et* H *une fonction convexe sur* $\mathbb{R}^k$ *à valeurs dans* $\mathbb{R} \cup \{+\infty\}$.
Supposons que

- *pour tout* n, $\emptyset \neq \overset{\circ}{\mathrm{dom}}\, H \subset \mathrm{dom}\, H_n$,
- *pour tout* n *suffisamment grand, l'ensemble* $\mathrm{Argmin}\, H_n$ *de tous les minimisants de* H_n *soit non vide,*
- H *admet un unique minimisant* v^* *appartenant à* $\overset{\circ}{\mathrm{dom}}\, H$,
- *la suite* $(H_n)_n$ *converge simplement vers* H *sur* $\overset{\circ}{\mathrm{dom}}\, H$,

alors, pour tout $\varepsilon > 0$, *il existe* $N \in \mathbb{N}$ *tel que pour tout* $n \geq N$,

$$\mathrm{Argmin}\, H_n \subset B(v^*, \varepsilon)$$

Démonstration. Raisonnons par l'absurde, et supposons qu'il existe $r > 0$ tel que $B(v^*, r) \subset \overset{\circ}{\mathrm{dom}}\, H$ et une suite $(v_n^*)_n$ telle que, pour tout n, $v_n^* \in \operatorname{Argmin} H_n$ et $|v_n^* - v^*| > r$.

Première étape :
Soit $\bar{v}_n \in B\left(v^*, \frac{r}{3}\right)$ telle que

$$H_n(\bar{v}_n) = \min\left\{H_n(v) : v \in B\left(v^*, \frac{r}{3}\right)\right\}$$

La suite $(\bar{v}_n)_n$ est bornée ; soit $\bar{v}$ une valeur d'adhérence de cette suite, et ϕ telle que $\lim\limits_{n\to+\infty} \bar{v}_{\phi(n)} = \bar{v}$. $(H_n)_n$ est une suite de fonction convexes convergeant simplement vers H sur $\overset{\circ}{\mathrm{dom}}\, H$, la convergence est donc uniforme sur tout compact inclus dans $\overset{\circ}{\mathrm{dom}}\, H$ (voir par exemple [38], Thm 3.1.4 p.105). En particulier,

$$\left|H_{\phi(n)}(\bar{v}_{\phi(n)}) - H(\bar{v}_{\phi(n)})\right| \leq \left\|H_{\phi(n)} - H\right\|_{\infty, B\left(v^*, \frac{r}{3}\right)} \xrightarrow[n\to+\infty]{} 0$$

De plus, par continuité de H, $H(\bar{v}_{\phi(n)}) \xrightarrow[n\to+\infty]{} H(\bar{v})$, donc $H_{\phi(n)}(\bar{v}_{\phi(n)}) \xrightarrow[n\to+\infty]{} H(\bar{v})$. Or, $H_{\phi(n)}(\bar{v}_{\phi(n)}) \leq H_{\phi(n)}(v^*)$, donc en passant à la limite, $H(\bar{v}) \leq H(v^*)$. La fonction H n'atteignant son minimum qu'au point v^*, on en déduit $\bar{v} = v^*$. Par conséquent $(\bar{v}_n)_n$ converge vers v^*.

Deuxième étape :
Pour tout $n \in \mathbb{N}$, la fonction $h_n : [0, 1] \to \mathbb{R} : t \mapsto H_n(v_n^* + t(\bar{v}_n - v_n^*))$, est croissante.
Soit $t_n \in [0, 1]$ tel que $\frac{2r}{3} \leq |v_n^* + t_n(\bar{v}_n - v_n^*) - v^*| \leq r$.
Posons $z_n = v_n^* + t_n(\bar{v}_n - v_n^*)$, alors pour tout n,

$$H_n(z_n) \leq H_n(\bar{v}_n) \qquad \text{et} \qquad \frac{2r}{3} \leq |z_n - v^*| \leq r \tag{B.2}$$

Quitte à extraire une sous-suite, on peut supposer que $(z_n)_n$ converge vers z vérifiant $\dfrac{2r}{3} \leq |z - v^*| \leq r$. La suite $(H_n)_n$ convergeant uniformément vers H sur $B(v^*, r)$, on conclut facilement que $\lim\limits_{n\to+\infty} H_n(z_n) = H(z)$ et, en passant à la limite dans l'inégalité (B.2), que $H(z) \leq H(v^*)$, ce qui entraîne que $z = v^*$ - absurde. □

Démonstration du théorème V.8.
Preuve des points 1. et 2.
Pour toute $\nu \in \mathcal{P}(\mathbb{R}^n)$,

$$\begin{aligned}
\mathbb{E}_\nu[\langle L_n, F\rangle] &= \mathbb{E}_\nu\left[\frac{1}{n}\sum_{i=1}^n z_i f_1(x_i^n), \ldots, \frac{1}{n}\sum_{i=1}^n z_i f_k(x_i^n)\right] \\
&= \frac{1}{n}\begin{pmatrix} f_1(x_1^n) & \ldots & f_1(x_n^n) \\ \vdots & \ldots & \vdots \\ f_k(x_1^n) & \ldots & f_k(x_n^n)\end{pmatrix}\mathbb{E}_\nu[Z] \\
&= A_n\mathbb{E}_\nu[Z],
\end{aligned}$$

donc

$$\Pi_n(K^\varepsilon) = \{\nu \in \mathcal{P}(\mathbb{R}^n) : A_n\mathbb{E}_\nu[Z] \in K^\varepsilon\}.$$

Notons $S_{\mu^{\otimes n}}$, le support de $\mu^{\otimes n}$, et admettons un instant que

$$\exists n_0, \quad \forall n \geq n_0, \quad A_n^{-1}(K) \cap \mathrm{co}\,\overset{\circ}{S}_{\mu^{\otimes n}} \neq \emptyset. \tag{B.3}$$

Nous prouverons (B.3) plus loin. Remarquons que l'on a aussi, pour tout $\varepsilon \geq 0$,

$$\forall n \geq n_0, \quad A_n^{-1}(K^\varepsilon) \cap \mathrm{co}\,\overset{\circ}{S}_{\mu^{\otimes n}} \neq \emptyset \tag{B.4}$$

$\mathrm{dom}\, Z_{\mu^{\otimes n}} =]-\alpha, \beta[^n$ étant ouvert, on peut appliquer le théorème II.41 et conclure que

- $\mu^{\otimes n}$ admet une I-projection $\mu^*_{n,\varepsilon}$ sur $\Pi_n(K^\varepsilon)$, ce qui prouve le point 1.,
- $\mu^*_{n,\varepsilon}$ vérifie
$$\frac{d\mu^*_{n,\varepsilon}}{d\mu^{\otimes n}} = \frac{\exp\left\langle A_n^t u^*_{n,\varepsilon}, \,.\,\right\rangle}{Z_{\mu^{\otimes n}}(A_n^t u^*_{n,\varepsilon})},$$
  où $u^*_{n,\varepsilon} \in \mathbb{R}^k$ est un minimisant de
$$G_{n,\varepsilon}(u) = \Lambda_{\mu^{\otimes n}}(A_n^t u) - \inf_{y \in K^\varepsilon}\langle u, c\rangle.$$

Mais, pour tout $x \in]-\alpha, \beta[^n$

$$\Lambda_{\mu^{\otimes n}}(x) = \Lambda_\mu(x_1) + \cdots + \Lambda_\mu(x_n)$$

et pour tout $u \in \mathbb{R}^k$,

$$A_n^t u = \begin{bmatrix} \left\langle F(x_1^n), \frac{u}{n}\right\rangle \\ \vdots \\ \left\langle F(x_n^n), \frac{u}{n}\right\rangle \end{bmatrix}$$

Par conséquent,

$$\begin{aligned} G_{n,\varepsilon}(u) &= n\left[\frac{1}{n}\sum_{i=1}^{n}\Lambda_\mu\left\langle F(x_i^n),\frac{u}{n}\right\rangle - \inf_{y\in K^\varepsilon}\left\langle\frac{u}{n},y\right\rangle\right] \\ &= nH_{n,\varepsilon}\left(\frac{u}{n}\right) \end{aligned}$$

donc $u^*_{n,\varepsilon}$ minimise $G_{n,\varepsilon}$ si, et seulement si, $\frac{u^*_{n,\varepsilon}}{n}$ minimise $H_{n,\varepsilon}$.

En posant $v^*_{n,\varepsilon} = \frac{u^*_{n,\varepsilon}}{n}$ et $w^*_{n,\varepsilon} = \begin{bmatrix} \langle F(x_1^n), v^*_{n,\varepsilon}\rangle \\ \vdots \\ \langle F(x_n^n), v^*_{n,\varepsilon}\rangle \end{bmatrix}$, on obtient le point 2.

Preuve du point 3.

$$R^*_{n,\varepsilon} = \mathbb{E}_{\mu^*_{n,\varepsilon}}[L_n] = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}} z_i\, d\mu^*_{n,\varepsilon}(dz)\delta_{x_i^n},$$

mais, pour tout $w \in]-\alpha,\beta[$,

$$\int_{\mathbb{R}} x\, d\mu_w = \Lambda'_\mu(w),$$

donc, pour tout i

$$\int_{\mathbb{R}} z_i\, d\mu^*_{n,\varepsilon}(z) = \int_{\mathbb{R}} z\, d\mu_{(w^*_{n,\varepsilon})_i} = \Lambda'_\mu((w^*_{n,\varepsilon})_i) = \Lambda'_\mu\left\langle v^*_{n,\varepsilon}, F(x_i^n)\right\rangle$$

et

$$R^*_{n,\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\Lambda'_\mu\left\langle v^*_{n,\varepsilon}, F(x_i^n)\right\rangle\delta_{x_i^n}$$

Preuve de (B.3).
Montrons qu'il existe n_0 tel que pour tout $n \geq n_0$,

$$A_n^{-1}(K)\cap \operatorname{co}\overset{\circ}{S}_{\mu^{\otimes n}} \neq \emptyset$$

Soit J_μ l'enveloppe convexe (fermée) du support de μ. On voit facilement que $\operatorname{co} S_{\mu^{\otimes n}} = J_\mu^n$.
Montrons donc, que pour tout n assez grand, il existe $z^n \in (\overset{\circ}{J}_\mu)^n$ tel que $A_n z^n \in K$.

Notons $C_\mu(\mathcal{X})$ l'ensemble des fonctions continues sur $\mathcal{X}$ et à valeurs dans $\overset{\circ}{J}_\mu$.
Pour toute $g \in C_\mu(\mathcal{X})$, nous poserons :

$$z^n(g) = (g(x_1^n),\ldots,g(x_n^n)) \in (\overset{\circ}{J}_\mu)^n$$

Remarquons que pour toute $g \in C_\mu(\mathcal{X})$,

$$A_n z^n(g) = \left[\frac{1}{n}\sum_{i=1}^n g(x_i^n) f_1(x_i^n), \ldots, \frac{1}{n}\sum_{i=1}^n g(x_i^n) f_k(x_i^n)\right]$$

On en déduit, d'après l'hypothèse (2) de (V.7), que pour toute $g \in C_\mu(\mathcal{X})$,

$$A_n z^n(g) \xrightarrow[n\to+\infty]{} \int_{\mathcal{X}} g(x)F(x)\,dR(x).$$

Or, d'après l'hypothèse (5) de (V.7), il existe $g_0 \in C_\mu(\mathcal{X})$ telle que

$$c_0 := \int_{\mathcal{X}} g_0(x)F(x)\,dR(x) \in K.$$

Raisonnons par l'absurde et supposons qu'il existe une suite strictement croissante d'entiers $(n_p)_p$ telle que pour tout p et toute $g \in C_\mu(\mathcal{X})$,

$$A_{n_p} z_{n_p}(g) \neq c_0.$$

Pour tout p, $\{A_{n_p} z_{n_p}(g) : g \in C_\mu(\mathcal{X})\} \subset \mathbb{R}^k$ est convexe et ne contient pas c_0.
D'après le théorème de Hahn-Banach, il existe $u_{n_p} \in \mathbb{R}^k$ tel que $\|u_{n_p}\| = 1$ et

$$\langle u_{n_p}, c_0\rangle \geq \sup_{g\in C_\mu(\mathcal{X})} \left\langle u_{n_p}, A_{n_p} z_{n_p}(g)\right\rangle.$$

Par compacité, on peut supposer que u_{n_p} converge vers u.
Pour tout $g \in C_\mu(\mathcal{X})$, $\langle u_{n_p}, c_0\rangle \geq \left\langle u_{n_p}, A_{n_p} z_{n_p}(g)\right\rangle$ donc, en passant à la limite dans cette inégalité, on obtient

$$\langle u, c_0\rangle \geq \left\langle u, \int_{\mathcal{X}} g(x)F(x)\,dR(x)\right\rangle.$$

Par suite pour toute $g \in C_\mu(\mathcal{X})$,

$$\left\langle u, \int_{\mathcal{X}} (g-g_0)(x)F(x)\,dR(x)\right\rangle \leq 0.$$

Soit B la boule unité de $C(\mathcal{X})$ (ensemble des fonctions continues sur $\mathcal{X}$).
Alors pour $r > 0$ assez petit, $g_0 + rB \subset C_\mu(\mathcal{X})$. On en déduit que pour toute $g \in rB$, $\left\langle u, \int_{\mathcal{X}} g(x)F(x)\,dR(x)\right\rangle \leq 0$ ce qui entraîne par symétrie et homogénéité que, pour toute $g \in C(\mathcal{X})$,

$$\int_{\mathcal{X}} g(x)\,\langle u, F(x)\rangle\, dR(x) = 0.$$

On en déduit que

$$R\left(\langle u, F(x)\rangle = 0\right) = 1$$

et ceci entraîne, d'après l'hypothèse (1) de (V.7), que

$$\langle u, F(x)\rangle = 0 \text{ pour tout } x \in \mathcal{X},$$

ce qui contredit l'hypothèse (3) de (V.7).

Preuve du point (4).
La fonction

$$H(\,.\,) = \int_{\mathcal{X}} \Lambda_\mu \langle\, .\,, F(x)\rangle \, dR(x) - \inf_{y\in K} \langle\, .\,, y\rangle$$

vérifie

$$\overset{\circ}{\operatorname{dom}} H = \left\{ v \in \mathbb{R}^k : \forall x \in U, \quad \langle v, F(x)\rangle \in]-\alpha, \beta[\right\}$$

et on a clairement

$$\overset{\circ}{\operatorname{dom}} H \subset \operatorname{dom} H_{n,\varepsilon_n},$$

où H_{n,ε_n} est la fonction convexe donnée par

$$H_{n,\varepsilon_n}(v) = \frac{1}{n}\sum_{i=1}^{n} \Lambda_\mu \langle v, F(x_i^n)\rangle - \inf_{y\in K^{\varepsilon_n}} \langle v, y\rangle.$$

Pour tout $v \in \overset{\circ}{\operatorname{dom}} H$, la fonction $\Lambda_\mu \langle v, F(\,.\,)\rangle$ est bornée, donc d'après l'hypothèse (2) de (V.7), $(H_{n,\varepsilon_n})_n$ converge simplement vers H sur $\overset{\circ}{\operatorname{dom}} H$. De plus, d'après l'hypothèse (6), la fonction H atteint son minimum en un unique point $v^* \in \overset{\circ}{\operatorname{dom}} H$.
On peut donc conclure, en utilisant le lemme B.1, que v^*_{n,ε_n} converge vers v^*.

Preuve du point (5).
Pour toute $g \in C(\mathcal{X})$, on a

$$\left\langle R^*_{n,\varepsilon_n}, g\right\rangle = \frac{1}{n}\sum_{i=1}^{n} \Lambda'_\mu \left\langle v^*_{n,\varepsilon_n}, F(x_i^n)\right\rangle g(x_i^n).$$

Le lemme V.24 entraîne qu'il existe un segment J inclus dans $]-\alpha, \beta[$ et m tel que pour tout $n \geq m$,

$$\forall n \geq m, \quad \left\langle v^*_{n,\varepsilon_n}, F(x_i^n)\right\rangle \in J \qquad \text{et} \qquad \forall x \in \mathcal{X}, \quad \langle v^*, F(x)\rangle \in J.$$

Si $M = \sup_{x\in J} \Lambda''_\mu(x)$, on a donc, d'après l'inégalité des accroissements finis,

$$\left| \left\langle R^*_{n,\varepsilon_n}, g\right\rangle - \frac{1}{n}\sum_{i=1}^{n} \Lambda'_\mu \langle v^*, F(x_i^n)\rangle \, g(x_i^n) \right| \leq M \sup |g| . \sup \|F\| . \|v^* - v^*_{n,\varepsilon_n}\| \xrightarrow[n\to+\infty]{} 0.$$

Enfin,

$$\frac{1}{n}\sum_{i=1}^{n}\Lambda'_\mu\left\langle v^*, F(x_i^n)\right\rangle g(x_i^n) = \left\langle \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i^n}, \Lambda'_\mu\left\langle v^*, F(\,.\,)\right\rangle g(\,.\,)\right\rangle$$

et comme $\Lambda'_\mu\left\langle v^*, F(\,.\,)\right\rangle g(\,.\,) \in C(\mathcal{X})$, on a d'après l'hypothèse (2) de (V.7)

$$\left\langle R^*_{n,\varepsilon_n}, g\right\rangle \xrightarrow[n\to+\infty]{} \int_{\mathcal{X}} \Lambda'_\mu\left\langle v^*, F(x)\right\rangle g(x)\, dR(x),$$

ceci pour toute $g \in C(\mathcal{X})$. $\square$

Bibliographie

[1] R. Aebi. *Schrödinger diffusion processes*. Birkhäuser, Basel-Berlin-Boston, 1996.

[2] M. Avellaneda, C. Friedman, R. Holmes, and D. Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance*, 4(1) :37–64, 1997.

[3] S. G. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *Journal de Mathématiques Pures et Aplliquées*, 80(7) :669–696, 2001.

[4] S.G. Bobkov and F. Gotze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *Journal of Functional Analysis.*, 163 :1–28, 1999.

[5] F. Bolley and C. Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *à paraître aux Annales de la Faculté des Sciences de Toulouse*, 2005.

[6] E. Bolthausen and U. Schmock. On the maximum entropy principle for uniformly ergodic Markov chains. *Stochastic Processes and their applications*, 33 :1–27, 1989.

[7] J.M. Borwein and A.S. Lewis. Duality relationships for entropy-like minimization problems. *SIAM Journal of Control and Optimization*, 29 :325–338, 1991.

[8] J.M. Borwein and A.S. Lewis. Partially-finite programming in L_1 and the exitence of maximum entropy estimates. *SIAM Journal of Optimization*, 3 :248–267, May 1993.

[9] H. Brezis. *Analyse Fonctionnelle*. Masson, 1983.

[10] V. V. Buldygin and Yu.V. Kozachenko. *Metric characterization of random variables and random processes*. American Mathematical Society, 2000.

[11] J. Van Campenhout and T. Cover. Maximum entropy and conditional probability. *IEEE Transactions on Information Theory*, 27(4) :483–489, 1981.

[12] E. Carlen. Conservative diffusions. *Communications in Mathematical Physic*, 94 :293–316, 1984.

[13] P. Cattiaux and F. Gamboa. Large deviations and variational theorems for marginal problems. *Bernoulli*, 5 :81–108, 1999.

[14] P. Cattiaux and A. Guillin. Talagrand's like quadratic transportation cost inequalities. preprint, 2004.

[15] P. Cattiaux and C. Léonard. Minimization of the Kullback information of diffusion processes. *Annales de l'Institut Henri Poincaré*, 30(1) :83–132, 1994. and correction in *Ann. Inst. Henri Poincaré* vol.31, p.705-707, 1995.

[16] P. Cattiaux and C. Léonard. Large deviations and Nelson processes. *Formum Mathematicum*, 7 :95–115, 1995.

[17] P. Cattiaux and C. Léonard. Minimization of the Kullback information for general Markov processes. Séminaire de Probas XXX. *Lectures Notes in Maths*, 1626 :283–311, 1996.

[18] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3 :146–158, 1975.

[19] I. Csiszár. Sanov property, generalized I-projection and a conditional limit theorem. *Annals of Probability*, 12 :768–793, 1984.

[20] I. Csiszár. Why least squares and maximum entropy ? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19 :2032–2066, 1991.

[21] I. Csiszár, F. Gamboa, and E. Gassiat. MEM pixel correlated solutions for generalized moment and interpolation problems. *IEEE Transactions on Information Theory*, 45(7) :2253–2270, 1999.

[22] D. Dacunha-Castelle and F. Gamboa. Maximum d'entropie et problèmes des moments. *Annales de l'Institut Henri Poincaré*, 26 :567–596, 1990.

[23] A. de Acosta. On large deviations of sums of independent random variables. In Lecture Notes in Math. 1153, 1985. Springer-Verlag.

[24] A. Dembo and J. Kuelbs. Refined Gibbs conditioning principle for certain infinite dimensional statistics. *Studia Scientiarum Mathematicarum Hungarica*, 34 :107–126, 1998.

[25] A. Dembo and O. Zeitouni. Refinements of the Gibbs conditioning principle. *Probability Theory and Related Fields*, 104 :1–14, 1996.

[26] A. Dembo and O. Zeitouni. *Large deviations techniques and applications. Second edition*. Springer Verlag, 1998.

[27] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities for random dynamical systems and diffusions. *Annals of Probability*, 32(3B) :2702–2732, 2004.

[28] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, III. *Comm. Pure Appl. Math.*, 36 :389–461, 1976.

[29] R.M. Dudley. *Real analysis and probability*. Wadsworth & Brooks/Cole, 1989.

[30] P. Eichelsbacher and U. Schmock. Large deviations of U-empirical measures in strong topologies and applications. *Annales de l'Institut Henri Poincaré*, 38(5) :779–797, 2002.

[31] U. Einmahl and J. Kuelbs. Dominating points and large deviations for random vectors. *Probability Theory and Related Fields*, 105 :529–543, 1996.

[32] R.S. Ellis, J. Gough, and J.V. Pulé. The large deviation principle for measures with random weights. *Reviews in Mathematical Physics*, 5 :659–692, 1993.

[33] H. Föllmer. Random fields and diffusion processes, Ecole d'été de Saint-Flour. *Lectures Notes in Maths*, 1362 :101–204, 1988.

[34] F. Gamboa. *Méthode du maximum d'entropie sur la moyenne et applications*. Thèse Orsay, 1989.

[35] F. Gamboa and E. Gassiat. Maximum d'entropie et problèmes des moments : Cas multidimensionnel. *Probability and Mathematical Statistics*, 12 :67–83, 1991.

[36] F. Gamboa and E. Gassiat. Bayesian methods and maximum entropy for ill-posed inverse problems. *The Annals of Statistics*, 25 :328–350, 1997.

[37] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. preprint, 2005.

[38] J.B. Hirriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Springer Verlag, 2001.

[39] Yu.V. Kozachenko and E.I. Ostrovskii. Banach spaces of random variables of subgaussian type. *Theor. Probability and Math. Statist.*, 3. :45–56, 1986.

[40] J. Kuelbs. Large deviation probabilities and dominating points in open convex sets : non-logarithmic behavior. *The Annals of Probability*, 28(3) :1259–1279, 2000.

[41] J. Kuelbs and A. Meda. Rates of convergence for the Nummelin conditional weak law of large numbers. *Stochastic Processes and their Applications*, 98(2) :229–252, 2002.

[42] S. Kulkarni and O. Zeitouni. A general classification rule for probability measures. *The Annals of Statistics*, 23(4) :1393–1407, 1995.

[43] C. Léonard. Minimizer of energy functionals. *Acta Mathematica Hungarica*, 93(4) :281–325, 2001.

[44] C. Léonard. A convex optimization problem arising from probabilistic questions. Prépublications de l'Université Paris 10 - Nanterre, 2004.

[45] C. Léonard. Dominating points and entropic projections. Prépublications de l'Université Paris 10 - Nanterre, 2004.

[46] C. Léonard and J. Najim. An extension of Sanov's theorem : application to the Gibbs conditioning principle. *Bernoulli*, 8(6) :721–743, 2002.

[47] K. Marton. A simple proof of the blowing-up lemma. *IEEE Transactions on Information Theory*, 32 :445–446, 1986.

[48] K. Marton. Bounding $\bar{d}$-distance by informational divergence : a way to prove measure concentration. *Annals of Probability*, 24 :857–866, 1996.

[49] P. Massart. Saint-Flour Lecture Notes. 2003.

[50] J. Najim. A Cramer type theorem for weighted random variables. *Electronic Journal of Probability*, 7, 2002.

[51] E. Nelson. Stochastic mechanics and random fields, Ecole d'été de Saint-Flour. *Lectures Notes in Maths*, 1362 :429–450, 1988.

[52] P. Ney. Dominating points and the asymptotics of large deviations for random walks on $\mathbb{R}^d$. *The Annals of Probability*, 11 :158–167, 1983.

[53] P. Ney. Convexity and large deviations. *The Annals of Probability*, 12 :903–906, 1984.

[54] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173 :361–400, 2000.

[55] M. S. Pinsker. *Information and information stability of random variables and processes*. Holden-Day, San Francisco, 1964.

[56] S. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol I : Theory, Vol. II : Applications*. Probability and its applications. Springer Verlag, New York, 1998.

[57] D. Revuz and M. Yor. *Continuous martingales and brownian motion*. Springer, third edition, 1998.

[58] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

[59] R.T. Rockafellar. *Conjugate Duality and Optimization*. Society for Industrial and Applied Mathematics, 1974.

[60] R.T. Rockafellar and R. Wets. *Variational Analysis*. Springer Verlag, 1997.

[61] G. Royer. *Une initiation aux inégalités de Sobolev logarithmiques*. SMF, 1999.

[62] E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Annales de l'Institut Henri Poincaré*, 2 :269–310, 1932.

[63] D.W. Stroock. *Probability theory : an analytic view*. Cambridge University Press, 1993. revised version.

[64] D.W. Stroock and O. Zeitouni. Microcanonical distributions, Gibbs states and the equivalence of ensembles. In R. Durret and H. Kesten editors, *Festschrift in honour of F. Spitzer*. p.399-424, 1991. Birkhäuser.

[65] A.S. Sznitman. Equations de type de Boltzmann spatialement homogènes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 66 :559–592, 1984.

[66] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'I.H.E.S.*, 81 :73–203, 1995.

[67] M. Talagrand. New concentration inequalities for product spaces. *Inventionnes Mathematicae*, 126 :505–563, 1996.

[68] M. Talagrand. Transportation cost for gaussian and other product measures. *Geometric and Functional Analysis*, 6 :587–600, 1996.

[69] T. Tjur. *Conditional Probability Distributions*. PhD thesis, Univ. Copenhagen, 1974.

[70] F. Topsoe. Information theoretical optimization techniques. *Kybernetika*, 15 :8–27, 1979.

[71] A. Van Der Vaart and J. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer, 1995.

[72] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.

[73] V.V. Yurinskii. Exponential inequalities for sums of random vectors. *Journal of multivariate analysis*, 6 :473–499, 1976.

[74] S.L. Zabell. Rates of convergence for conditional expectations. *Annals of Probability*, 8 :928–941, 1980.

[75] C. Zuily and H. Queffélec. *Agrégation de Mathématiques - Eléments d'analyse*. Dunod.