

N° d'ordre: 9983

THÈSE

Présentée pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES
DE L'UNIVERSITÉ PARIS-SUD XI

Spécialité : Mathématiques

par

Pierre PETIT

Sur la théorie de Cramér et sa généralisation aux champs asymptotiquement découplés

Soutenue le 22 octobre 2010 devant la Commission d'examen :

M.	Alano ANCONA	
M.	Raphaël CERF	(Directeur de thèse)
M.	Francis COMETS	
M.	Jean-Dominique DEUSCHEL	
M.	Charles-Edouard PFISTER	(Rapporteur)
M.	Wendelin WERNER	

Rapporteurs :

M.	Charles-Edouard Pfister
M.	Sandy Zabell



Thèse préparée au
Département de Mathématiques d'Orsay
Laboratoire de Mathématiques (UMR 8628), Bât. 425
Université Paris-Sud 11
91 405 Orsay CEDEX

RÉSUMÉ

La présente thèse s'inscrit dans une suite de travaux sur la théorie fondamentale des grandes déviations. Cramér (1938) a montré que les moyennes empiriques d'une suite de variables aléatoires réelles indépendantes et de même loi vérifient un principe de grandes déviations (PGD). Et Chernoff (1952) a identifié l'entropie du PGD et l'opposée de la fonction convexe-conjuguée de la pression ($s = -p^*$). Donsker et Varadhan (1966) ont proposé un cadre généralisant l'obtention du PGD, d'où découle l'égalité $s = -p^*$. Leur formalisme a été approfondi dans les ouvrages classiques d'Azencott (1980), de Acosta (1985), Deuschel et Stroock (1989) et Dembo et Zeitouni (1993). Reprenant les idées de Bahadur et Zabell (1979), la présente thèse donne un cadre plus optimal pour l'égalité $s = -p^*$. Ce travail permet de mieux comprendre les outils pertinents pour la théorie de Cramér (sous-additivité, convexité, convexe-tension). Au passage, nous donnons une nouvelle preuve, plus simple, du résultat originel de Cramér sur la droite réelle. D'autre part, nous étendons la théorie de Cramér aux champs asymptotiquement découplés introduits par Pfister (2002) : nous relaxons donc l'hypothèse d'indépendance, tout en conservant une forme de sous-additivité. Le cadre finalement obtenu contient les théories de Cramér et de Sanov pour des variables indépendantes, ainsi que les principes de grandes déviations pour les chaînes de Markov (Donsker et Varadhan) et les mesures de Gibbs (Comets, Orey, Pelikan, Föllmer, Ort et Olla).

Mots-clefs : Grandes déviations, théorie de Cramér, découplage asymptotique

REMERCIEMENTS

C'est à l'heure de rédiger cette page que je prends conscience du nombre phénoménal de personnes qui habitent le « nous » des démonstrations qui suivent. Mission impossible que de citer toutes ces personnes sans qui je n'aurais pu mener à bien ce travail de trois années. Tâche ardue, mais joyeuse, que de mentionner les plus présentes, à ma mémoire et à mon cœur.

Merci avant tout à Raphaël Cerf qui a su faire grandir, dans la confiance, mon goût pour la recherche. Il a été un directeur présent. Depuis mon Master 2, il oriente mes recherches, par des remarques venant à point et un grand souci de clarté, tout en me laissant ce précieux temps de scruter les problèmes en profondeur, comme il le fait lui-même pour préparer et donner ses cours et ses exposés. Et il a toujours été prévoyant pour m'éviter toute pression inutile. Je le remercie également d'être le témoin d'une belle vie de famille, joyeuse et accueillante, notamment pendant les séjours qu'il m'a permis de faire à Göteborg et à Padoue.

Merci à Charles-Edouard Pfister et à Sandy Zabell. J'ai eu un grand plaisir à lire et à approfondir leurs travaux, dont j'apprécie la clarté. Et je leur suis maintenant très reconnaissant d'avoir accepté la lourde tâche de relire la présente thèse durant l'été et de l'évaluer. Je regrette que nous n'ayons pu nous arranger pour que M. Zabell puisse assister à la soutenance et remercie beaucoup M. Pfister de s'être libéré pour venir.

Merci à Wendelin Werner qui, depuis cinq ans maintenant, est présent à chaque moment clef de mes études. En Master 2, son cours et celui de M. Cerf ont séduit mon penchant pour la physique statistique. M. Werner m'a alors orienté vers M. Cerf pour mon mémoire et je constate avec joie la justesse de son discernement. Il est à nouveau présent, avec enthousiasme, dans mon jury de thèse et je l'en remercie de tout cœur.

Merci à Jean-Dominique Deuschel qui a accepté de se déplacer pour faire partie de mon jury. La présente thèse doit beaucoup à ses travaux, que mon directeur m'a souvent cités comme première référence. Merci aussi à Francis Comets et à Alano Ancona de témoigner un vif intérêt pour ce travail en ayant également accepté de faire partie du jury.

Merci à Marie Théret, ma grande sœur de thèse. Attentive à ce que je ne sois pas perdu pour mes premiers pas dans le monde de l'école doctorale, elle a su répondre à mes questions. J'en profite pour remercier également Valérie Lavigne de son aide précieuse pour les doctorants. Marie Théret a aussi fait preuve d'initiative pour que nous nous exposions mutuellement nos domaines de recherche, ce dont je lui suis très reconnaissant. Plus généralement, je souhaite remercier ici tous ceux avec qui, à un moment ou à un autre, j'ai eu la joie d'explorer le monde des mathématiques : mes professeurs de mathématiques, au premier rang desquels mes parents Françoise et Frédéric, les organisateurs des divers concours de mathématiques, et ceux qui ont été sur les bancs avec moi, notamment Rémi, Pascal, Vincent, Pierre, Aurélie, Aurélien, Nicolas et Sébastien.

Merci à tous ceux qui m'ont donné le goût d'observer, de manipuler, de comprendre, de chercher. A toutes les personnes déjà citées, j'aimerais ajouter mes professeurs du conservatoire à qui je dois beaucoup. L'apprentissage d'un instrument requiert patience, écoute et travail, et mes professeurs ont eu la patience de me guider sur ce chemin de la recherche du beau et du don de soi. Je veux également remercier tous ceux, professeurs, pères, frères, qui m'ont ouvert l'esprit à de larges horizons et ont contribué à me construire humainement. Je pense particulièrement à ma marraine Monique, au père Armogathe et au père Leproux. Je tiens encore à dire un merci de petit-fils à Paul Petit et Lucien Mollet, amoureux de la nature et du beau, qui m'ont donné envie d'ouvrir les yeux.

Merci à tous ceux qui m'ont un jour soutenu, souvent supporté ou fidèlement aimé, à tous ces amis, musiciens, frères et sœurs qui, sans tout comprendre à ce que je faisais assis à mon bureau, ont permis que je ne sois pas qu'un doctorant en mathématiques. Merci en particulier à ma famille et à toi, Marie, pour la joie que tu mets dans mon cœur.

Deo gratias.

Sur la théorie de Cramér
et sa généralisation
aux champs asymptotiquement découplés

TABLE DES MATIÈRES

Résumé	3
Remerciements	5
1. Introduction	11
1.1 Grandes déviations	12
1.2 L'irruption de la physique statistique	18
1.3 Contributions : un retour à la théorie de Cramér	26
1.4 Suite du programme	30
2. Une preuve élémentaire du théorème de Cramér dans $\mathbb{R}$	33
2.1 Introduction	33
2.2 Énoncé	34
2.3 Démonstration	35
3. Théorie de Cramér pour des variables indépendantes	39
3.1 Introduction	40
3.2 Cadre	42
3.3 Théorie de Cramér	46
3.4 Retour sur les hypothèses et compléments	48
3.5 Questions de mesurabilité	51
3.6 Démonstration des résultats principaux	53
3.7 Compléments	70
3.8 Suppléments techniques	78
4. Pression, entropie et transport linéaire	81
4.1 Introduction	82
4.2 Pression	82
4.3 Entropie et principes de grandes déviations	83
4.4 Espaces topologiques localement mesurables	92
4.5 Espaces vectoriels topologiques mesurables	95
4.6 Les trois fondements de la théorie de Cramér	99
4.7 Suppléments sur le PGD	103

5. Théorie de Cramér pour des champs asymptotiquement découplés	107
5.1 Introduction	108
5.2 Cadre	108
5.3 Théorie de Cramér pour des champs asymptotiquement découplés	113
5.4 Retour sur les hypothèses et compléments	115
5.5 Démonstration des résultats principaux	118
5.6 Compléments	126
6. Une preuve de $s = -p^*$ à partir du théorème de Mosco	129
6.1 Introduction	130
6.2 Lemme sous-additif	130
6.3 PGD faible	135
6.4 Pression	138
6.5 Pression et entropie	142

1

Chapitre 1

INTRODUCTION

Sommaire

1.1	Grandes déviations	12
1.1.1	« Un nouveau théorème-limite »	12
1.1.2	Principes de grandes déviations	13
1.1.3	Vers une borne supérieure plus faible	17
1.2	L'irruption de la physique statistique	18
1.2.1	Le second principe de la thermodynamique	18
1.2.2	L'argument sous-additif	19
1.2.3	De (BS_b) à (BS_h) : la convexe-régularité	21
1.2.4	Identification de s : l'équivalence d'ensembles	22
1.3	Contributions : un retour à la théorie de Cramér	26
1.3.1	Contexte, en bref	26
1.3.2	Une preuve élémentaire du théorème de Cramér dans $\mathbb{R}$	26
1.3.3	Théorie de Cramér pour des variables indépendantes	27
1.3.4	Pression, entropie et transport linéaire	29
1.3.5	Théorie de Cramér pour les champs asymptotiquement découplés	30
1.3.6	Une preuve de $s = -p^*$ à partir du théorème de Mosco	30
1.4	Suite du programme	30

1.1 Grandes déviations

1.1.1 « Un nouveau théorème-limite »

Dans son article publié en 1938 et intitulé « Sur un nouveau théorème-limite de la théorie des probabilités » [Cra38], motivé par le calcul optimal de primes d'assurance, le mathématicien suédois Harald Cramér¹ s'intéresse au comportement asymptotique de moyennes empiriques

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

où $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes, identiquement distribuées et de carré intégrable, chacune modélisant les remboursements mensuels de sinistres par la compagnie d'assurance. Plus précisément, notant m (resp. σ) l'espérance (resp. l'écart-type) de X_1 , il donne divers développements asymptotiques de

$$F_n(x_n) = \mathbb{P}\left(\overline{X}_n \leq m + \frac{x_n \sigma}{\sqrt{n}}\right)$$

pour certaines suites $(x_n)_{n \geq 1}$ de réels positifs, voulant alors généraliser ce qu'il appelle le « théorème limite classique de LAPLACE-LIAPOUNOFF (dans sa forme moderne précisée par LINDEBERG et par M. Paul LÉVY) », aujourd'hui plus connu sous le nom de *théorème-limite central*² : pour tout réel x ,

$$\lim_{n \rightarrow \infty} F_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

Le théorème-limite central fait partie, dans la terminologie actuelle, des résultats de *petites déviations*, c'est-à-dire des résultats asymptotiques pour lesquels $x_n = O(1)$, par opposition aux résultats de *grandes déviations* dans le cas où $x_n = c\sqrt{n}$. On parle aussi de *moyennes déviations*³ dans le cas où $1 \ll x_n \ll \sqrt{n}$ et de *très grandes déviations* lorsque $x_n \gg \sqrt{n}$. Cramér étudie les moyennes et grandes déviations, et obtient, dans le cas des

¹ À ne pas confondre avec le mathématicien suisse homonyme, Gabriel Cramer (1704 – 1752), à qui nous devons entre autres la fameuse méthode de résolution des systèmes linéaires. Quant à lui, Harald Cramér (1893 – 1985) s'est penché sur la théorie analytique des nombres et sur les probabilités. Mentionnons le très beau résultat, conjecturé par Lévy en 1935 et démontré par Cramér l'année suivante [Cra36] : si une variable aléatoire normale peut être décomposée en la somme de deux variables aléatoires indépendantes, alors chacun des termes de la somme suit une loi normale.

² Cette dénomination n'est peut-être pas la plus répandue dans l'école française de probabilités. Le Cam [LC86] note : « The appellation “central” is due to Pólya (1920) who used it because of the central role of the theorem in probability theory, not as the modern French do, because it describes the behavior of the center of the distribution as opposed to its tails. » L'article de George Pólya [Pól20] s'intitule en effet “Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem”.

³ Un certain nombre d'articles parus avant celui de Cramér traitent de petites et moyennes déviations, bien qu'ils emploient le terme de « grandes déviations » (« больших уклонений », « grosse Abweichungen » chez Smirnov [Smi33] par exemple).

grandes déviations, un développement très précis de la forme : pour tout réel c , il existe des coefficients $\alpha, b_0, \dots, b_{k-1}$ tels que

$$1 - F_n(c\sqrt{n}) = \frac{1}{\sqrt{n}} e^{-\alpha n} \left(b_0 + \frac{b_1}{n} + \dots + \frac{b_{k-1}}{n^{k-1}} + O\left(\frac{1}{n^k}\right) \right)$$

Sa démonstration repose essentiellement sur l'application du théorème-limite central, non directement à la suite $(X_n)_{n \geq 1}$ de variables de loi μ , mais à la suite $(\tilde{X}_n)_{n \geq 1}$ de variables dont la loi est donnée par le changement de loi

$$d\tilde{\mu}(x) = \frac{e^{tx}}{\mathbb{E}(e^{tX_1})} d\mu(x)$$

Cette preuve requiert des hypothèses raisonnables sur la loi μ de X_1 : des moments exponentiels et l'absolue continuité par rapport à la mesure de Lebesgue. Ne s'intéressant qu'à un encadrement du type

$$e^{n(s(x)-\varepsilon)} \leq \mathbb{P}(\bar{X}_n \geq x) \leq e^{ns(x)}$$

Hermann Chernoff propose en 1952 [Che52] une très belle preuve qui se passe de l'absolue continuité de μ : par approximation, il se ramène aux loi discrètes pour lesquelles le résultat est conséquence de la formule de Stirling et de la méthode des multiplicateurs de Lagrange. Les statisticiens indiens R. R. Bahadur et R. Ranga Rao vont encore affaiblir les hypothèses et finalement, en 1971, Bahadur [Bah71] publie une démonstration pour une mesure quelconque sur $\mathbb{R}$, montrant d'abord le résultat pour les lois de variables majorées.

1.1.2 Principes de grandes déviations

À la même époque, Sanov [San57] énonce un très beau résultat, l'analogue de celui de Cramér pour les mesures empiriques

$$M_n = \frac{\delta_{X_1} + \dots + \delta_{X_n}}{n}$$

Il montre que, pour des classes de probabilités A raisonnables,

$$\mathbb{P}(M_n \in A) = \exp \left(n \sup_{\nu \in A} s(\nu|\mu) + o(1) \right)$$

où

$$s(\nu|\mu) = \int \log \frac{d\mu}{d\nu} d\nu$$

est l'entropie⁴ de ν relativement à μ . Sanov pressent qu'il est possible d'unifier ses résultats et ceux de Cramér⁵ : « La valeur d'une telle méthode consistera en une approche unique d'un grand nombre de questions différentes. » C'est ce que propose le probabiliste indien S. R. Srinivasa Varadhan, qui s'est vu remettre le prix Abel en 2007 notamment pour ce travail unificateur. C'est à lui que nous devons la notion de *principe de grandes déviations* : on dit qu'une suite de mesures $(\mu_n)_{n \geq 1}$ sur un espace topologique vérifie un principe de grandes déviations (PGD) s'il existe une fonction raisonnable s (que nous appellerons *bonne fonction de taux du PGD*⁶) telle que, pour tout ensemble mesurable A , on ait une estimation du type

$$\log \mu_n(A) \sim n \sup_A s$$

Plus précisément, s doit avoir des ensembles de niveau compacts et vérifier la borne inférieure

(BI) pour tout $A \in \mathcal{F}$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \geq \sup_A s$$

et la borne supérieure

(BS) pour tout $A \in \mathcal{F}$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \leq \sup_A s$$

En particulier, on dit qu'une suite de variables aléatoires $(Z_n)_{n \geq 1}$ vérifie un PGD si la suite des lois des variables $(\mu_n)_{n \geq 1}$ vérifie un PGD. Auparavant, Chernoff [Che52] et Sanov [San57] avaient proposé de bons candidats, dans leurs cadres respectifs, pour la fonction de taux :

- si $Z_n = \overline{X}_n$ est la suite des moyennes empiriques d'une suite $(X_n)_{n \geq 1}$ de variables réelles indépendantes et identiquement distribuées,

$$s(x) = - \sup_{\lambda \in \mathbb{R}} \left(\lambda x - \log \mathbb{E}(e^{\lambda X_1}) \right)$$

⁴ Sanov [San57] mentionne le lien à la fin de son article. Notons qu'il est courant (*cf.* notamment [San57], [Föl73], [OP88], [FO88a], [FO88b], [Oll88], [DSZ91], [Cer07]) de définir l'entropie relative par $h = -s$. La terminologie n'est toujours pas unifiée, l'histoire n'ayant pas aidé. Boltzmann [Bol98] désigne par H une quantité de la forme

$$\int f \log f$$

et l'identifie à l'opposé de l'entropie S de Clausius [Cla67]. Shannon [Sha48] utilise à bon escient le terme *entropy* (sur une suggestion de von Neumann; *cf.* [TM71]), mais note cette quantité H ... Kullback et Leibler [KL51] appellent (*mean*) *information* l'opposé de l'entropie et la désignent par I . Georgii [Geo93] reprend la notation I , mais l'appelle *entropy*; Lewis, Pfister et Sullivan [LPS94a] utilisent l'expression *information gain*, mais la notent h . Peut-être faudrait-il préférer les notations S (pour l'entropie) et I (pour l'information) à la notation ambiguë H , d'autant que I est la notation utilisée pour les fonctions de taux depuis l'article fondateur de Varadhan [Var66].

⁵ Il pense aussi à ceux de Smirnov et de Petrov.

⁶ C'est habituellement $I = -s$ que l'on dénomme ainsi.

autrement dit l'opposé de la transformée de Fenchel-Legendre⁷ du logarithme de la transformée de Laplace de la loi de X_1 .

• si $Z_n = M_n$ est la suite des mesures empiriques d'une suite $(X_n)_{n \geq 1}$ de variables réelles indépendantes et de même loi μ ,

$$s(\nu) = - \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu$$

autrement dit l'entropie relativement à μ .

Varadhan a donc unifié les résultats de Cramér-Chernoff (les moments exponentiels assurent que la fonction de taux proposée est bonne) et de Sanov dans un cadre prometteur. D'une part, le concept de principe de grandes déviations jette un pont entre l'estimation d'une probabilité asymptotique et le calcul variationnel de maximisation d'une fonction s , facilement calculable dans les bons cas. D'autre part, l'approche générale de Varadhan permet d'énoncer des propriétés de transport des PGD :

Lemme de Varadhan : si $(\mu_n)_{n \geq 1}$ vérifie un PGD de bonne fonction de taux s et si f est une application continue majorée, alors la suite des mesures⁸ $(\tilde{\mu}_n)_{n \geq 1}$ définies par

$$d\tilde{\mu}_n(x) = e^{nf(x)} d\mu_n(x)$$

vérifie un PGD⁹ de bonne fonction de taux $f + s$.

Principe de contraction : si $(\mu_n)_{n \geq 1}$ vérifie un PGD de bonne fonction de taux s et si f est une application continue, alors $(\mu_n \circ f^{-1})_{n \geq 1}$ vérifie un PGD de bonne fonction de taux

$$s_f(y) = \sup_{f(x)=y} s(x)$$

⁷ Adrien-Marie Legendre (1752–1833) a d'abord étudié une transformation analogue pour des fonctions différentiables, donnée par

$$f^*(\lambda) = \lambda x(\lambda) - f(x(\lambda))$$

où $x(\lambda)$ est le paramètre conjugué de λ , solution de l'équation

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

Werner Fenchel (1905–1988) a défini la transformation

$$f^*(\lambda) = \sup_{x \in X} (\langle \lambda | x \rangle - f(x))$$

et a montré le théorème de dualité pour des fonctions convexes [Fen49]. Il s'avère que les deux transformations coïncident pour les fonctions convexes différentiables. On garde habituellement les noms des deux mathématiciens, bien que la définition soit souvent celle de Fenchel (cf. [Mor67]).

⁸ On trouve souvent la dénomination anglaise de « tilted measures ».

⁹ Le lemme de Varadhan généralise, en un sens, la méthode de Laplace qui assure que

$$\log \int e^{nf} dx \sim n \max f$$

Théorème de Dawson-Gärtner : soit $(\mu_n)_{n \geq 1}$ une suite de mesures et $(f_i)_{i \in I}$ la famille des applications canoniques d'un système projectif d'applications continues ; si, pour tout $i \in I$, $(\mu_n \circ f_i^{-1})_{n \geq 1}$ vérifie un PGD de bonne fonction de taux s_i , alors $(\mu_n)_{n \geq 1}$ vérifie un PGD de bonne fonction de taux

$$s = \inf_{i \in I} s_i \circ f_i$$

On doit les deux premiers résultats à Varadhan lui-même (cf. [Var66]) ; le dernier apparaît plus tard, dans [DG87], théorisant un outil de la démonstration de [GOR79] du théorème de Sanov. Le chapitre 4 du livre de Dembo et Zeitouni [DZ93] regroupe ces résultats, et quelques autres, avec leurs démonstrations. Pour résumer, le cadre général des principes de grandes déviations aboutit à deux conclusions :

- si une suite de mesures $(\mu_n)_{n \geq 1}$ vérifie un PGD de fonction de taux s , l'estimation asymptotique de $\mu_n(A)$ se ramène à la maximisation de s sur A ;
- les PGD se comportent bien vis à vis de transformations raisonnables.

Les propriétés de transport donnent l'espoir de ramener un grand nombre de problèmes de grandes déviations aux résultats de Cramér et de Sanov, ou à des résultats similaires. Cela ouvre la voie à deux types de généralisations : montrer des PGD dans des espaces autres que $\mathbb{R}$ ou $\mathcal{M}_1^+(\mathbb{R})$ (l'ensemble des mesures de probabilité sur $\mathbb{R}$) ; relaxer l'hypothèse d'indépendance. Donsker et Varadhan [DV75] ont commencé par généraliser le théorème de Sanov pour des processus de Markov. Plusieurs ont avancé dans cette voie, généralisant le théorème de Sanov à des champs d'interaction sur $\mathbb{Z}^d$ issus de la physique statistique, en particulier les mesures de Gibbs : c'est le cas de Comets, Orey, Pelikan, Föllmer, Ort et Olla (cf. [Com86], [OP88], [FO88a], [FO88b] et [Oll88]). Hans-Otto Georgii [Geo93] fait une synthèse de ces résultats et remplace la topologie faible par une topologie plus fine, la τ -topologie (c'est l'analogue, pour les mesures de Gibbs, de [GOR79] pour les variables i.i.d.). Le théorème-limite central qui servait à Cramér est ici remplacé par des théorèmes ergodiques. D'un autre côté, Plachky et Steinebach [PS75], puis Gärtner [Gär77], généralisent la théorie russe équivalente à celle de Donsker et Varadhan, due à Freidlin et Wentzell, en remplaçant l'indépendance par une simple condition de convergence des transformées de Laplace

$$\forall \lambda \in \mathbb{R} \quad \frac{1}{n} \log \mathbb{E}(e^{\lambda Z_n}) \rightarrow p(\lambda)$$

Gärtner met ici en évidence un nouveau lien avec la physique statistique, la fonction $p(t)$ étant, au signe près, l'analogue d'une *énergie libre spécifique*. Ellis [Ell84] utilise les idées de Donsker et Varadhan pour obtenir le théorème aujourd'hui connu sous le nom de théorème de Gärtner-Ellis¹⁰ (cf. [DZ93]), généralisé par Baldi [Bal88] à des espaces vectoriels topologiques de dimension infinie, moyennant une hypothèse de tension exponentielle : on

¹⁰ Dans [Ell85], Ellis montre le lien entre physique statistique et principes de grandes déviations dans le cadre du théorème de Gärtner-Ellis. Comme nous allons le voir, cette approche diffère de celle de la présente thèse.

dit que $(\mu_n)_{n \geq 1}$ est exponentiellement tendue si, pour tout $\gamma < 0$, il existe un compact $K(\gamma)$ tel que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(X \setminus K(\gamma)) \leq \gamma$$

La notion de tension exponentielle est un outil puissant introduit dans [DV75] permettant de passer de la borne supérieure (BS) pour les compacts à la borne supérieure pour tout ensemble. Elle est utilisée pour montrer les premières généralisations du théorème de Sanov.

1.1.3 Vers une borne supérieure plus faible

Toutefois, le cadre de Varadhan ne contient pas la théorie de Cramér dans toute sa généralité : en effet, cinq ans après le premier article de Varadhan, Bahadur [Bah71] a étendu le théorème de Cramér à toute mesure μ sur $\mathbb{R}$. Or, il existe des lois μ pour lesquelles les ensembles de niveau de la fonction de taux ne sont pas compacts : c'est le cas de la loi de Cauchy, pour laquelle $s = 0$. Dinwoodie (cf. [Din91]) a même montré l'existence, dans tout espace de dimension supérieure à 3, d'une suite $(\bar{X}_n)_{n \geq 1}$ pour laquelle il n'existe pas de fonction de taux vérifiant (BI) et (BS) (sa construction repose sur un contre-exemple de Slaby [Sla88])¹¹. On peut pourtant espérer obtenir des résultats un peu plus faibles.

L'approche de Varadhan suggère de séparer le problème des grandes déviations en deux questions successives : d'abord l'existence (abstraite) d'une fonction de taux s , puis l'expression de s . Explicitons. Pour fixer les idées, soit $(\mu_n)_{n \geq 1}$ une suite de mesures sur un espace topologique X muni de sa tribu borélienne $\mathcal{B}$. Il est assez facile de voir que la plus grande fonction qui vérifie (BI) est définie par

$$s(x) = \inf_{V \in \mathcal{V}_x} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(V)$$

où $\mathcal{V}_x$ est l'ensemble des voisinages mesurables de x . Cette définition de s est donc un troisième candidat, après $-p^*$ et $s(\cdot|\mu)$, pour une fonction de taux. Nous avons vu qu'en règle générale il n'y a pas de PGD dans la théorie de Cramér, autrement dit la borne supérieure (BS) avec la fonction s définie ci-dessus n'est pas vérifiée pour tout ensemble mesurable. Cela amène naturellement à poser le problème du PGD un peu différemment. Si $\mathcal{P}$ est un ensemble de parties mesurables de X , on définit

(BS $_{\mathcal{P}}$) pour tout $A \in \mathcal{P}$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \leq \sup_{A \in \mathcal{P}} s$$

On peut alors s'intéresser aux sous-ensembles $\mathcal{P}$ de $\mathcal{B}$ pour lesquels (BS $_{\mathcal{P}}$) est vérifiée par la fonction s définie plus haut. En particulier, la suite $(\mu_n)_{n \geq 1}$ vérifiera un PGD si et seulement si s est bonne et vérifie (BS $_{\mathcal{B}}$). Mais cette approche ne se limite pas aux cas où il y a vraiment PGD : avoir (BS $_{\mathcal{P}}$) pour des sous-ensembles stricts $\mathcal{P}$ de $\mathcal{B}$ peut suffire pour arriver à des conclusions intéressantes, justement en physique statistique.

¹¹ Notons que le théorème de Cramér assure la borne supérieure (BS) en dimension 1, mais la question reste ouverte en dimension 2.

1.2 L'irruption de la physique statistique

1.2.1 Le second principe de la thermodynamique

La fonction s , telle que nous venons de la définir, ressemble beaucoup à l'*entropie microcanonique spécifique*¹² en physique statistique. En effet, la célèbre formule que Ludwig Boltzmann propose en 1877 (*cf.* [Bol77]) relie l'entropie microcanonique d'un système, fonction de paramètres macroscopiques comme la température ou la pression, au nombre d'états microscopiques compatibles par

$$S = k \log W$$

où $k = 1,38065812 \cdot 10^{-23} J \cdot K^{-1}$ est, aujourd'hui, la constante de Boltzmann et W , pour « Wahrscheinlichkeit », *vraisemblance*¹³, une façon de mesurer le nombre d'états microscopiques ω correspondant à l'état macroscopique du système isolé observé. Interprétant W comme la probabilité d'un ensemble d'états microscopiques, divisant par le nombre de particules n du système et par la constante de Boltzmann, puis passant à la limite thermodynamique, on retrouve bien une expression analogue à la définition de s . Cette formule, gravée sur la tombe du père de la physique statistique, est-elle cohérente avec le second principe de la thermodynamique, énoncé au milieu du dix-neuvième siècle par Clausius ? Rappelons, dans un cas simple, ce à quoi conclut le second principe. Imaginons un gaz confiné dans la moitié d'une boîte isolée juste après l'ouverture de la cloison séparant la boîte en deux (la classique détente de Joule). La température t ¹⁴ du gaz évolue (en pratique, elle diminue), jusqu'à une température d'équilibre t_0 . Selon le second principe, la valeur $S(t_0)$ maximise l'entropie $S(t)$. L'équation d'état (macroscopique), du type $pv = knt$, permet de calculer S et la thermodynamique classique permet donc de trouver t_0 . L'approche de Boltzmann part de la description des interactions microscopiques du système. À chaque valeur t de la température correspond un nombre d'états microscopiques $W(t)$. Notre question se reformule donc ainsi : à l'équilibre, $\log W(t_0)$ est-il bien maximal ? Comme nous allons le voir, il s'agit d'un problème de grandes déviations. Si n est le nombre (gigantesque, de l'ordre de $N_A = 6,02 \cdot 10^{23}$) de particules dans la boîte, on note $T_n(\omega)$ la température d'un état microscopique ω ¹⁵. Soient t_0 un point où $s(t_0)$ est maximal et V un voisinage de t_0 en dehors duquel l'entropie s est majorée par $s(t_0) - \varepsilon$. Si nous montrons (BS), alors,

¹² L'adjectif *microcanonique* se rapporte à un système isolé, c'est-à-dire n'échangeant ni énergie ni particules avec l'extérieur. Quant à *spécifique*, cela signifie *par particule* : l'entropie étant une grandeur extensive, c'est l'entropie spécifique qui a du sens quand on passe à la limite thermodynamique, autrement dit quand le nombre de particules tend vers l'infini.

¹³ Le sens actuel de *probabilité* se répand petit à petit.

¹⁴ Contrairement aux usages habituels des physiciens, nous utiliserons des lettres minuscules pour les valeurs macroscopiques, réservant les lettres majuscules aux variables aléatoires, définies sur l'ensemble Ω des états microscopiques, qui apparaîtront par la suite.

¹⁵ On peut voir $T_n(\omega)$ comme l'énergie cinétique interne spécifique du système dans l'état microscopique ω .

d'une part $1 = \mathbb{P}(T_n \in \mathbb{R}) \leq e^{ns(t_0)}$, donc $s(t_0) = 0$, et d'autre part

$$\mathbb{P}(T_n \in \mathbb{R} \setminus V) \sim \exp \left(n \sup_{\mathbb{R} \setminus V} s \right) \leq e^{-n\varepsilon}$$

Donc T_n se concentre sur V , pour tout voisinage V de t_0 . Physiquement, la température T_n mesurée est t_0 . Ainsi, justifier la formule de Boltzmann revient à montrer (BS). En fait, pas tout à fait. Il suffit de montrer la borne supérieure pour des complémentaires de voisinages V particuliers pour aboutir à la même conclusion. Dans le cas de la température, il suffit de considérer un petit voisinage V de la forme $[t_0 - \delta t, t_0 + \delta t]$ (où δt est l'imprécision de la mesure). Plus, généralement, si l'on s'intéresse à plusieurs paramètres en même temps, par exemple la température et la pression, $x = (t, p)$, on se place dans $X = \mathbb{R}^2$, et il suffit de prendre

$$V(\delta) = B_\infty(x_0, \delta)$$

de sorte que $X \setminus V(\delta)$ est une réunion de quatre demi-espaces : ainsi, la borne supérieure (BS_h) pour les unions finies de demi-espaces est suffisante pour aboutir au second principe.

1.2.2 L'argument sous-additif

L'application des idées de la physique statistique à la théorie de Cramér est due au physicien Oscar E. Lanford [Lan73]. Vingt ans auparavant, van Hove [vH49] est l'un des premiers physiciens à donner une preuve rigoureuse de l'existence de la limite thermodynamique de la fonction énergie libre f caractérisant l'état d'un système : « La thermodynamique admet que pour $N \rightarrow \infty$, $f(N, D)$ tend vers une limite finie $f(v)$, indépendante de la forme de D et fonction seulement du volume spécifique v maintenu constant dans le passage à la limite. [...] Notre but est de montrer que, sous des conditions générales relatives à U_N , la fonctionnelle $f(N, D)$ tend effectivement vers une fonction $f(v)$ satisfaisant à ces exigences. » Remarquons que, au cours de sa démonstration, il mentionne qu'elle pourrait aussi se faire « en utilisant la formule de Stirling et les multiplicateurs de Lagrange », méthode utilisée par Chernoff pour démontrer le théorème de Cramér. Certains physiciens continuent dans ce souci d'obtenir des « rigorous results », notamment Yang et Lee [YL52] qui simplifient la preuve de van Hove, puis Ruelle, Robinson et Lanford, une belle synthèse étant faite par Ruelle [Rue69] en 1969. Ces derniers travaux reposent sur l'article *Correlation Functionals* [Rue65] où Ruelle met pleinement en lumière l'argument *sous-additif*¹⁶ comme outil général pour prouver l'existence de limites thermodynamiques :

Lemme 1.2.2.1 (Lemme sous-additif). *Soit $(u(n))_{n \geq 1}$ une suite réelle positive telle que, pour tous m et n supérieurs à 1,*

$$u(m+n) \leq u(m) + u(n)$$

¹⁶ La première version du lemme sous-additif est due au mathématicien israëlo-hongrois Michael (Mihály) Fekete [Fek23] et a été généralisée par Pólya et Szegő [PS54] (exercice 98 de la première partie).

Alors,

$$\lim_{n \rightarrow \infty} \frac{u(n)}{n} = \inf_{n \geq 1} \frac{u(n)}{n}$$

C'est cet argument sous-additif que Lanford [Lan73] importe dans la théorie de Cramér, dans le chapitre « A Digression : Sums of Independent Random Variables ». Si C est convexe et si les variables $(X_n)_{n \geq 1}$ sont indépendantes et de même loi, alors

$$\begin{aligned} \mathbb{P}(\overline{X}_{m+n} \in C) &\geq \mathbb{P}\left(\frac{X_1 + \cdots + X_m}{m} \in C; \frac{X_{m+1} + \cdots + X_{m+n}}{m+n} \in C\right) \\ &= \mathbb{P}(\overline{X}_m \in C) \mathbb{P}(\overline{X}_n \in C) \end{aligned}$$

Ainsi, la suite

$$-\frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C)$$

est sous-additive et converge¹⁷ vers son infimum $-\mathfrak{s}(C)$. Avec le concept de *mesure asymptotiquement découplée* et une généralisation du lemme sous-additif, Pfister [Pfi02] étend l'argument, de façon unificatrice, à de nombreux modèles de physique statistique et aux chaînes de Markov. Revenant à l'entropie définie plus haut, on en déduit que

$$s(x) = \inf_{C \in \mathcal{C}_x} \mathfrak{s}(C)$$

où $\mathcal{C}_x$ est l'ensemble des voisinages convexes ouverts de x . L'existence de cette limite est un pas décisif car la fonction $\mathfrak{s}$ vérifie¹⁸ alors ce que Lewis, Pfister et Sullivan [LPS94a] appellent le « Principle of the Largest Term¹⁹ » (PLT) : pour tous ouverts convexes C et D ,

$$\mathfrak{s}(C \cup D) = \mathfrak{s}(C) \vee \mathfrak{s}(D)$$

L'argument est le même que celui²⁰ qui montre que

$$2^n + 3^n \sim 3^n$$

En se ramenant alors à des unions finies de petits ouverts convexes, pour lesquels $\mathfrak{s}(C) \approx s(x)$ avec $x \in C$, Ruelle [Rue65] (puis Lanford [Lan73] dans le cadre du théorème de Cramér) montre la borne supérieure (BS_b) pour les compacts, ce qu'on appelle maintenant *principe de grandes déviations faible*.

¹⁷ Pour être précis, il faut supposer C ouvert pour éviter le cas où la suite prendrait occasionnellement la valeur $+\infty$.

¹⁸ Plus généralement, dans le cas d'une suite quelconque $(\mu_n)_{n \geq 1}$, c'est

$$\overline{\mathfrak{s}}(A) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A)$$

qui vérifie le PLT pour tous sous-ensembles mesurables A_1 et A_2 .

¹⁹ L'idée, qui recouvre la même réalité que la méthode de Laplace, est bien connue des physiciens. [LPS94a] cite [Hua63] où le terme « largest term » apparaît.

²⁰ On notera encore le lien avec la méthode de Laplace, qui permet de démontrer la formule de Stirling utilisée par Chernoff dans sa démonstration du théorème de Cramér. Le PLT est donc plus fondamental et permet justement de court-circuiter le recours à la formule de Stirling.

1.2.3 De (BS_b) à (BS_h) : la convexe-régularité

Nous avons vu que, pour aboutir au second principe de la thermodynamique, il serait satisfaisant de montrer la borne supérieure (BS_h) pour les demi-espaces, le PLT assurant que la borne supérieure sera vraie pour les unions finies de demi-espaces. Justement, combinant (BS_b) et la régularité de la mesure de Lebesgue, Ruelle montre que, pour tout intervalle ouvert C de $\mathbb{R}$,

$$s(C) = \sup_{x \in C} s(x)$$

autrement dit une version raffinée de la borne supérieure pour les convexes. En particulier, on a la borne supérieure pour les demi-espaces ouverts. Mission accomplie ! Lanford transpose les mêmes arguments pour le théorème de Cramér, mais il ne montre pas que l'argument de régularité fonctionne pour une mesure quelconque. C'est à Bahadur et Zabell [BZ79] que l'on doit une synthèse très complète des idées importées par Lanford, dans un cadre mathématique très général : ils étendent la théorie de Cramér aux espaces vectoriels localement convexes de dimension infinie, traitant du même coup les questions de mesurabilité, et introduisent la notion de *convexe-régularité*²¹, permettant justement de passer de (BS_b) à (BS_h) . On dit qu'une mesure μ est convexe-régulière si, pour tout ouvert convexe C et pour tout $\gamma > 0$, il existe un convexe compact K inclus dans C tel que

$$\mu(C \setminus K) \leq \gamma$$

Contrairement à la tension exponentielle, la convexe-régularité est toujours satisfaite en dimension finie et, plus généralement, dans un espace de Banach séparable ! D'autre part, elle est plus facile à vérifier que la tension exponentielle. Elle est donc plus appropriée pour les résultats énoncés dans cette partie. Toutefois, la convexe-tension a été plus ou moins délaissée par les successeurs de Bahadur et Zabell. Le cours donné par Azencott durant l'été 1978 à Saint-Flour [Aze80] reprend les idées de [BZ79] et les applique dans deux directions : il montre que le théorème de Sanov [San57] est un corollaire de la théorie de Cramér²² et s'intéresse au théorème de Cramér dans les espaces de Banach séparables, dans lesquels il montre que la condition

$$\mathbb{E}[e^{t\|X_1\|}] < \infty$$

entraîne la tension exponentielle, et donc le PGD. Autrement dit, Azencott commence à réinjecter les idées de Bahadur et Zabell dans le cadre de Donsker et Varadhan, délaissant donc la convexe-régularité pour des hypothèses plus fortes assurant (BS) . C'est aussi ce que font les suivants dans cette voie : de Acosta [DA85], puis les deux traités classiques de grandes déviations de Deuschel et Strook [DS89] et de Dembo et Zeitouni [DZ93].

²¹ C'est l'hypothèse vérifiée par la mesure de Lebesgue (et, en fait, par toute mesure sur un espace de dimension finie) qui fait fonctionner la preuve de Ruelle [Rue65] et de Lanford [Lan73]. La définition est sûrement aussi inspirée de l'idée de [Bah71] de se ramener à des variables majorées.

²² Azencott [Aze80] mentionne : « C'est un peu ce qu'ont tenté de faire Bahadur et Zabell [7], mais au prix de l'utilisation, en cours de route de résultats de Donsker-Varadhan [15], qui eux-mêmes n'étaient obtenus qu'après démonstration du théorème de Sanov ! »

1.2.4 Identification de s : l'équivalence d'ensembles

Qu'en est-il de l'expression de s ? Nous avons vu que Chernoff proposait

$$s(u) = -\sup_{\lambda \in \mathbb{R}} (\lambda u - p(\lambda)) = -p^*(u)$$

où

$$p(\lambda) = \log \mathbb{E}(e^{\lambda X_1})$$

On appelle *formule variationnelle de Gibbs* la formule duale

$$p(\lambda) = \sup_{u \in \mathbb{R}} (\lambda u + s(u)) = (-s)^*(\lambda)$$

qui est équivalente à la première dès lors que s est une fonction concave et semi-continue supérieurement. En fait, alors que s désigne l'entropie microcanonique, la fonction p s'interprète physiquement comme l'opposée de l'énergie libre canonique, ou l'opposée de la fonction grand potentiel, qui, dans le cas des fluides simples (cf. [DGLR89]), est égale (au facteur v près) à la pression grand-canonique (d'où la notation). La formule variationnelle de Gibbs exprime ce que les physiciens, après Gibbs [Gib02], appellent *équivalence d'ensembles*²³. Pour faire simple, imaginons un gaz de n particules dans une boîte de volume v . Pour fixer l'état d'équilibre du gaz, on peut par exemple fixer son énergie interne u en isolant la boîte (description microcanonique) et en laissant la température évoluer, disons jusqu'à une température t à l'équilibre thermodynamique. Une autre façon de procéder est la suivante : fixons cette fois la température du gaz à t en le mettant au contact d'un thermostat (description canonique) et laissons s'opérer des échanges d'énergie avec le thermostat. A l'équilibre, l'énergie interne du gaz se stabilise, et ce à la valeur u de la description précédente. C'est ce qu'on appelle l'équivalence entre les descriptions, ou ensembles, microcanonique et canonique, à la limite thermodynamique : les deux variables d'état u et t sont conjuguées. Plus précisément : dans la description microcanonique, la valeur de l'entropie spécifique (divisée par la constante de Boltzmann k) s à l'équilibre thermodynamique est simplement fonction de l'énergie interne u fixée au départ ; et, dans la description canonique, la valeur de l'énergie libre f , ou plutôt de la fonction $p = -f/vkt$, est seulement fonction de la température t du thermostat. Eh bien, u et t , ou plutôt u et $\lambda = 1/kt$ sont reliés, de façon biunivoque par²⁴

$$s(u) - p(\lambda) = \lambda u$$

L'équivalence d'ensembles est vérifiée de manière générale pour les modèles simples de la physique statistique²⁵. Khinchin [Khi43] l'a montrée pour les gaz parfaits, Dobrushin et

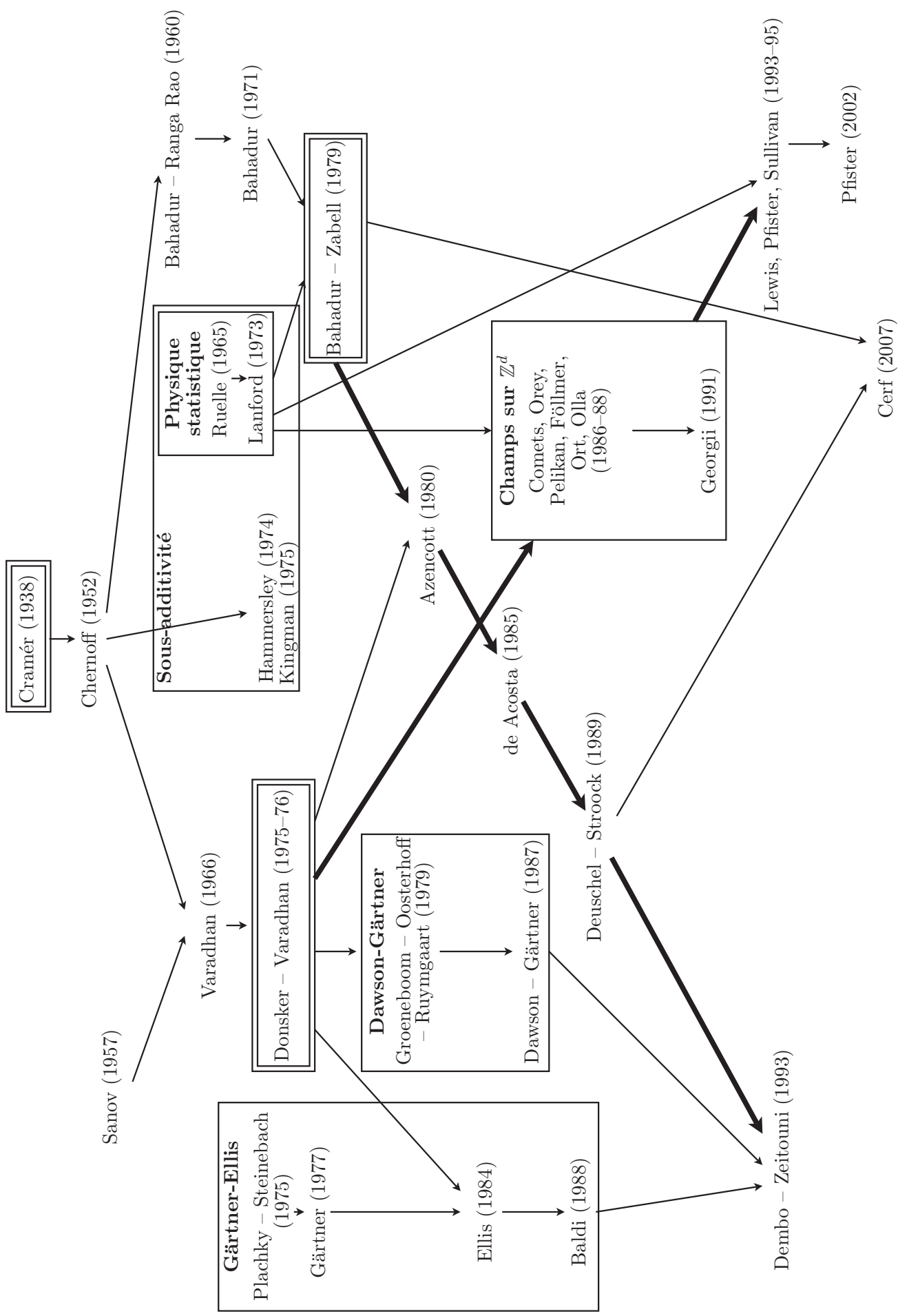
²³ On peut définir la notion d'équivalence d'ensembles à deux niveaux : au niveau des fonctions thermodynamiques, comme ici, ou au niveau des états, autrement dit des mesures sur les états microscopiques. La seconde question est beaucoup plus délicate et étudiée en détail dans [DSZ91], puis divers articles de Georgii [Geo94] [Geo95]. Pour une très belle synthèse et simplification du lien entre les deux, on renvoie aux articles de Lewis, Pfister et Sullivan [LPS94a], [LPS94b] et [Pf02].

²⁴ Pour compléter [DGLR89], une référence très claire à ce sujet est [ZRM09].

²⁵ Les interactions doivent être à portée suffisamment faible. Pour les modèles de champ moyen, par exemple, il n'y a pas équivalence d'ensembles. D'autres méthodes sont à mettre en œuvre pour identifier la fonction de taux, par exemple le théorème de Gärtner-Ellis.

Tirozzi [DT73] pour des modèles de gaz sur réseau (*lattice gas models* en anglais), tous trois utilisant une version locale du théorème-limite central, excluant notamment les cas de transition de phase du premier ordre. Il s'avère que les techniques de grandes déviations sont plus appropriées. Pour une suite $(X_n)_{n \geq 1}$ de variables indépendantes et identiquement distribuées, dans le cas où X_1 est à valeurs dans un compact, la formule variationnelle de Gibbs est une simple conséquence du lemme de Varadhan. Relaxer l'hypothèse d'indépendance est l'objectif principal des articles de mécanique statistique. [Com86], [OP88], [FO88a], [FO88b] et [Oll88] sont les premiers à utiliser des techniques de grandes déviations pour obtenir la formule variationnelle dans le cas des champs de Gibbs. Les suivants ([DSZ91], [Geo93], [LPS94a], [LPS94b] et [Pfi02]) approfondissent la question. Dans tous ces modèles, la compacité est vérifiée; seul [LP95] mentionne l'égalité plus généralement en cas de PGD. Or, grâce à la convexe-régularité, Bahadur et Zabell [BZ79] avaient étendu l'égalité $s = -p^*$ à leur cadre général. En fait, (BS) n'est pas nécessaire : (BS_h) apparaît comme une condition suffisante, naturelle et plus générale, pour l'identification entre s et $-p^*$.

Le schéma de la page suivante résume les différentes influences que nous avons mentionnées. Remarquons que nous n'avons pas du tout fait référence aux travaux de Freidlin et Wentzell (synthétisés dans [FW84]), bien que, par exemple, Gärtner travaille dans leur cadre (*cf.* [Gär77] et [DG87]). Il s'avère que leur théorie des grandes déviations, équivalente à celle de Donsker et Varadhan [DV75], est élaborée dans des espaces fonctionnels. Ayant focalisé notre étude sur la théorie de Cramér, et donc sur l'étude de suites de variables aléatoires, nous ne nous sommes pas appesantis sur leur travaux. Si nous devons placer la présente thèse sur le schéma, nous la placerions dans la droite ligne de [BZ79], via [Cer07], et de [Pfi02], avec l'influence de [DG87].



La descendance de Cramér : références principales

- [Aze80] R. AZENCOTT : *Grandes déviations et applications*. Ecole d'Été de Probabilités de Saint-Flour VIII-1978, Lecture Notes in Mathematics 774. Springer-Verlag, 1980.
- [Bah71] R. R. BAHADUR : *Some Limit Theorems in Statistics*. SIAM, 1971.
- [Bal88] P. BALDI : Large deviations and stochastic homogenization. *Ann. Math. Pura Appl.*, 151:161–177, 1988.
- [BZ79] R. R. BAHADUR et S. L. ZABELL : Large deviations of the sample mean in general vector spaces. *Ann. Prob.*, 7(4):587–621, 1979.
- [Cer07] R. CERF : *On Cramér's Theory in Infinite Dimensions*. Panoramas et Synthèses 23. Société Mathématique de France, Paris, 2007.
- [Che52] H. CHERNOFF : A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23:493–507, 1952.
- [Com86] F. COMETS : Grandes déviations pour des champs de Gibbs sur $\mathbb{Z}^d$. *C. R. Acad. Sc. Paris*, 303, sér. I(11):511–513, 1986.
- [Cra38] H. CRAMÉR : Sur un nouveau théorème limite de la théorie des probabilités. *Actual. Sci. Indust.*, 736:5–23, 1938.
- [DA85] A. DE ACOSTA : On large deviations of sums of independent random vectors. *Probability in Banach spaces V, Lecture Notes in Math.*, 1153:1–14, 1985.
- [DG87] D. A. DAWSON et J. GÄRTNER : Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20:247–308, 1987.
- [DS89] J.-D. DEUSCHEL et D. W. STROOCK : *Large Deviations*. Academic Press, 1989.
- [DV75] M. D. DONSKER et S. R. S. VARADHAN : Asymptotic evaluation of certain Markov process expectations for large time I. *Comm. Pure Applied Math.*, 28:1–47, 1975. Première partie d'une série de quatre articles sortis en 1975 et 1976.
- [DZ93] A. DEMBO et O. ZEITOUNI : *Large Deviations Techniques and Applications*. Jones and Bartlett, 1993. Deuxième édition chez Springer-Verlag en 1998.
- [Ell84] R. S. ELLIS : Large deviations for a general class of random vectors. *Ann. Prob.*, 12(1):1–12, 1984.
- [Föll73] H. FÖLLMER : On entropy and information gain in random fields. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 26:207–217, 1973.
- [FO88a] H. FÖLLMER et S. OREY : Large deviations for the empirical field of a Gibbs measure. *Ann. Prob.*, 16(3):961–977, 1988.
- [Geo93] H.-O. GEORGII : Large deviations and maximum entropy principle for interacting random fields on $\mathbb{Z}^d$. *Ann. Prob.*, 21(4):1845–1875, 1993.
- [GOR79] P. GROENEBOOM, J. OOSTERHOFF et F. H. RUYMGAART : Large deviation theorems for empirical probability measures. *Ann. Prob.*, 7(4):553–586, 1979.
- [Ham74] J. M. HAMMERSLEY : Postulates for subadditive processes. *Ann. Prob.*, 2(4):652–680, 1974.
- [Lan73] O. E. LANFORD : *Entropy and Equilibrium States in Classical Statistical Mechanics*. Lecture Notes in Physics 20. Springer, 1973.
- [LPS95] J. T. LEWIS, C.-E. PFISTER et W. G. SULLIVAN : Entropy, concentration of probability and conditional limit theorems. *Markov Processes Relat. Fields*, 1(3):319–386, 1995.
- [Oll88] S. OLLA : Large deviations for Gibbs random fields. *Probab. Theory Relat. Fields*, 77:343–357, 1988.
- [Pfi02] C.-E. PFISTER : *Thermodynamical aspects of classical lattice systems. Lectures delivered at the 4th brazilian school of probability, Mambucaba, RJ (2000), In and Out of Equilibrium, Probability with a Physics Flavor*. V. Sidoravicius, Birkhäuser, 2002.
- [Var66] S. R. S. VARADHAN : Asymptotic probabilities and differential equations. *Comm. Pure Applied Math.*, 19:261–286, 1966.

1.3 Contributions : un retour à la théorie de Cramér

1.3.1 Contexte, en bref

Pour résumer la situation, le théorème originel de Cramér [Cra38] et de Chernoff [Che52] a été principalement abordé et généralisé dans le langage unificateur proposé par Varadhan [Var66], le rapprochant du théorème de Sanov [San57]. Ce langage des *principes de grandes déviations* est le cadre des généralisations aux cas de variables non indépendantes (chaînes de Markov, mesures de Gibbs, *etc.*). Les textes classiques [Aze80] et [DS89] exposent également la théorie de Cramér dans ce cadre, reprenant de [BZ79] essentiellement l'argument sous-additif, lui-même inspiré des physiciens : le passage de (BS_b) à (BS) , ainsi que l'égalité $s = -p^*$, y sont obtenus par une hypothèse de *tension exponentielle*. Toutefois, cette hypothèse n'est pas toujours vérifiée, même dans des cas simples et non triviaux sur $\mathbb{R}$. En revanche, l'argument de *convexe-régularité* de [BZ79], beaucoup moins contraignant, assure la borne supérieure (BS_c) pour les convexes et $s = -p^*$. Il n'est repris que par [DZ93] et par [Cer07]. Dans la présente thèse, nous développons les idées de [BZ79], approfondies par [Cer07], pensant que la théorie de Cramér y trouve un cadre naturel et général. D'une part, nous substituons la notion de *convexe-tension* à celle de convexe-régularité, en simplifiant sensiblement les preuves. Et d'autre part, nous incorporons ces idées au cadre de [Pfi02], obtenant ainsi une théorie de Cramér pour des variables aléatoires *asymptotiquement découplées*. Au passage, nous établissons une série de résultats de *transport linéaire*, homologues des résultats de transport pour les principes de grandes déviations.

$$\left\{ \begin{array}{l} \textbf{Principes de grandes} \\ \textbf{déviations} \\ \text{espace topologique} \\ \text{tension exponentielle} \\ (BS) \\ s \text{ bonne fonction de taux} \end{array} \right. \quad \textit{versus} \quad \left\{ \begin{array}{l} \textbf{Théorie de Cramér} \\ \text{e.v. localement convexe} \\ \text{convexe-tension} \\ (BS_c) \\ \underline{s} = \overline{s} \end{array} \right.$$

Voici un aperçu du contenu de chacun des chapitres de la thèse.

1.3.2 Une preuve élémentaire du théorème de Cramér dans $\mathbb{R}$

Les preuves classiques du théorème de Cramér dans $\mathbb{R}$ reposent ou bien sur un autre théorème-limite (théorème-limite central dans [Cra38], loi des grands nombres dans [DS89], [DZ93]) après avoir effectué un changement de loi, ou bien sur une méthode d'approximation après avoir montré le résultat pour une famille de lois (lois discrètes via la formule de Stirling dans [Che52] et [Cer07], conditions de moment exponentiel dans [Bah71]). Les méthodes d'approximations sont faites à la main ([Che52], [Bah71], [Cer07, chapitre 15]) ou reposent sur le théorème de Mosco ([Cer07, chapitre 26] suite aux idées de [Zab92a] et [Zab92b]). Il est aussi possible d'utiliser la méthode de Laplace dans le cas où la pression p est dérivable ([Cer07, chapitre 14]). La preuve proposée ici est plus élémentaire et directe : on utilise simplement le théorème de convergence monotone, ayant conditionné les variables à être

dans un compact. L'idée s'inspire de la notion de convexe-régularité introduite dans [BZ79] et s'adapte en toute généralité aux cadres des chapitres 3 et 5 de cette thèse. Le plan de la preuve est aussi simplifié (il est assez proche de celui de [Ham74], repris par [Kin76], textes dont nous avons eu connaissance après avoir rédigé la présente preuve) et les grandes étapes apparaissent clairement. Nous montrons d'abord l'égalité duale

$$p = (-s)^*$$

où

$$p(\lambda) = \mathbb{E}(e^{\lambda X_1}) \quad \text{et} \quad s(x) = \sup_{n \geq 1} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq x)$$

Nous procédons par double inégalité : la borne inférieure ($p \geq (-s)^*$) découle de la classique inégalité de Tchebychev ; la borne supérieure ($p \leq (-s)^*$) quant à elle est montrée par conditionnement sur un compact et convergence monotone. Ensuite, l'égalité

$$s = -p^*$$

est classique en analyse convexe et repose sur le théorème de Hahn-Banach (dualité de Fenchel-Legendre ; cf. par exemple [Roc70]) dès que s est concave et semi-continue supérieurement. Nous montrons la concavité, et du même coup, l'égalité

$$s(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq x)$$

par l'argument sous-additif (on peut regrouper les deux preuves en une, mais nous estimons que la démonstration perd en clarté). La semi-continuité n'est pas montrée explicitement, mais elle est cachée dans la fin de la preuve.

Ce texte a été accepté dans la revue *American Mathematical Monthly*. Une version anglaise, un peu antérieure, est disponible à l'adresse <http://arxiv.org/abs/1002.3496>.

1.3.3 Théorie de Cramér pour des variables indépendantes

Ce chapitre n'a pas été le premier objet de recherche de cette thèse. Nous avons d'abord essayé d'exporter les résultats du cas indépendant (synthétisés dans [Cer07]) au cas asymptotiquement découplé (introduit dans [Pfi02]) : c'est l'objet du chapitre 5 de la présente thèse. Au cours de ce travail, nous avons été amené à reprendre les preuves du cas indépendant, notamment autour de la question

$$s = -p^*$$

et nous sommes arrivés à une preuve plus élémentaire que les preuves classiques, dont nous avons déjà vu le schéma général dans le chapitre 2. Motivés par la recherche d'un contre-exemple à l'égalité $s = -p^*$ (qui est vérifiée dans beaucoup de cas, notamment en dimension finie ; cf. [Cer07]), nous avons été amenés à étendre le cadre habituel de la théorie de Cramér (espace polonais muni de sa tribu borélienne) à un cadre plus général

et, somme toute, plus naturel. Nous reprenons et explicitons le cadre de [Cer07] emprunté à [LPS95]. Nous nous donnons un espace (vectoriel) X que nous munissons seulement d'une tribu²⁶ $\mathcal{F}$ de sorte que les quantités que nous avons à manipuler aient un sens : nous voulons simplement que l'addition vectorielle soit mesurable, pour que $\overline{X}_n$ soit une variable aléatoire (ce type d'hypothèse est déjà fait dans [BZ79], mais dans le cas d'une tribu borélienne). C'est pour cela que nous introduisons la notion d'*espace vectoriel localement convexe mesurable* (*e.v.l.c.m.*) : la tribu est engendrée par une famille de convexes internes. La convexité est le nerf de la sous-additivité, comme le montre [BZ79]. Nous identifions la séparabilité comme nerf de la mesurabilité²⁷. Ce n'est qu'après que l'on munit X d'une topologie, également engendrée par des convexes internes, la plus fine possible pour avoir le PGD faible et l'égalité $s = -p^*$. La construction précédente assure, par la même occasion, la mesurabilité des formes linéaires continues²⁸. C'est pour obtenir l'égalité $p = (-s)^*$ que nous introduisons l'hypothèse de *convexe-tension* : on dit qu'une probabilité μ est *convexe-tendue* si, pour tout $\gamma > 0$, il existe un convexe relativement compact $K(\gamma)$ tel que

$$\mu(X \setminus K(\gamma)) \leq \gamma$$

En utilisant l'outil *jauge* des convexes (à notre connaissance, introduit par [Cer07] dans la théorie de Cramér), nous montrons que cette hypothèse est (presque) équivalente à la convexe-régularité de [BZ79] et simplifions sensiblement les preuves et les hypothèses nécessaires (par exemple, nous nous passons de la régularité). À ce sujet, nous montrons au passage que l'énoncé de l'appendice de [BZ79] est correct (même si la preuve qu'ils proposent requiert des hypothèses supplémentaires pour appliquer le théorème de Krein), alors que [DS89] rajoute l'hypothèse $(\hat{C})$ pour corriger la démonstration, et que [DZ93] rajoute l'hypothèse 6.1.2 (b) ; de même, grâce à la jauge, [Cer07] n'avait pas besoin de la fin de l'hypothèse $(\hat{C})$ de [DS89] pour montrer l'existence de la limite dans $[-\infty, 0]$ de

$$\frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C)$$

pour tout convexe ouvert mesurable C . La convexe-tension nous donne, en prime, la borne supérieure pour les convexes (BS_c) . Enfin, ayant le souci de pouvoir déduire le théorème de Sanov (hormis l'identification de l'entropie avec $s(\cdot|\mu)$) de la théorie de Cramér, les questions de mesurabilité (et la tribu naturelle pour Sanov qui n'est plus une tribu borélienne) nous obligent à affiner les résultats quand les variables sont à valeurs dans une partie D (non nécessairement mesurable) de l'espace X . Nous obtenons ainsi une borne supérieure faible améliorée $(BS_{b,D})$ pour les parties mesurables K telles que $K \cap D$ soit relativement

²⁶ Il s'avère que la topologie est construite en même temps, mais ce n'est pas nécessaire. On renvoie à la partie « Suppléments techniques ».

²⁷ On notera que la mesurabilité de l'addition entraîne, dès que $\{0\}$ est mesurable, la mesurabilité de la diagonale $\{(x, -x); x \in X\}$. Or, si $\text{card}(X) > \aleph_1$, on sait montrer que la diagonale ne peut pas être dans $\mathcal{F} \otimes \mathcal{F}$. D'un autre côté, un espace métrique séparable est de cardinal $\leq \aleph_1$. Nous ne savons pas si l'hypothèse de séparabilité est optimale, mais elle n'en est pas loin.

²⁸ On aurait pu envisager définir p^* simplement avec les formes linéaires mesurables (continues), mais la propriété de Hahn-Banach ne serait plus automatiquement vérifiée, de même que l'égalité $s = -p^*$.

compact. Nous discutons la pertinence de ce formalisme dans la partie « Suites de Cramér et probabilités portées par une partie » et nous précisons ainsi un argument de [Cer07, chapitre 24]. Du même coup, nous obtenons l'égalité $s = -p^*$ dans tous les e.v.l.c.m. : nous nous passons donc de la complétude habituellement requise pour avoir la convexe-tension. Enfin, nous approfondissons l'étude du domaine de s , en corrigeant certaines imprécisions de [Cer07] (lemme 9.8, par exemple).

1.3.4 Pression, entropie et transport linéaire

Dans ce chapitre, aucun PGD ou PGD faible n'est prouvé. Il est plutôt question de définir un cadre naturel le plus général possible dans lequel les questions de grandes déviations peuvent être posées. Les idées générales d'une théorie des PGD faibles, déjà vues dans le chapitre 3 et utilisées dans les autres chapitres, sont exposées ici. Il manque seulement l'argument sous-additif : le lemme sous-additif est mentionné et démontré, de manière générale, sans toutefois que l'on montre ici explicitement le lien avec la théorie de Cramér. C'est l'objet des chapitres 3 et 5 où l'on voit que le lemme sous-additif sert justement à amorcer les preuves dans le cas de la théorie de Cramér, et de sa généralisation aux champs asymptotiquement découplés. Ce chapitre est donc l'analogue du chapitre 2 « Some generalities » de [DS89], du chapitre 4 « General Principles » de [DZ93], du chapitre 2 « Large Deviation Theory » de [Cer07] ou de l'article [LP95] dont le présent chapitre s'inspire et développe certains points (nous renvoyons d'ailleurs à cet article pour les nombreux autres résultats qui s'y trouvent). L'originalité de cette présentation est de définir d'abord une notion générale de pression sur un espace mesurable X : pour toute fonction mesurable $\varphi : X \rightarrow [-\infty, +\infty]$, on définit

$$\underline{p}(\varphi) := \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{v_n \varphi} d\mu_n \quad \text{et} \quad \overline{p}(\varphi) := \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{v_n \varphi} d\mu_n$$

On peut ensuite définir l'entropie à partir de la pression des indicatrice $\delta_A = -\infty 1_A$ par

$$\underline{s}(x) = \inf_{A \in \mathcal{F}} \left(\underline{p}(\delta_A) \dot{+} (-\delta_A(x)) \right) \quad \text{et} \quad \overline{s}(x) = \inf_{A \in \mathcal{F}} \left(\overline{p}(\delta_A) \dot{+} (-\delta_A(x)) \right)$$

où l'opération $\dot{+}$ étend l'addition à $[-\infty, +\infty]$ avec $(-\infty) + (+\infty) = +\infty$. Cela fait naturellement apparaître le lien entre entropie et pression et laisse entrevoir la fameuse égalité

$$s = -p^*$$

Nous énonçons dans le cas des PGD faibles (et aussi dans les cas plus agréables où $\underline{s} = \overline{s}$) des résultats de transport habituellement énoncés pour les PGD : lemme de Varadhan, principe de contraction (intéressant dans le cas i.i.d.), et surtout une version linéaire du théorème de Dawson-Gärtner, fondamentale pour l'étude de la question $s = -p^*$: l'égalité se transporte par limite projective. Ce résultat permet d'obtenir l'égalité $s = -p^*$ pour les espaces vectoriels localement convexes mesurables introduits au chapitre 3, et est utile pour la structure projective naturelle du chapitre 5.

1.3.5 Théorie de Cramér pour les champs asymptotiquement découplés

Nous nous sommes rendus compte que le cadre de [Pfi02], essentiellement voué à conserver une forme de sous-additivité (*cf.* [Pfi02, lemme 3.1]), permettait une généralisation naturelle de la théorie de Cramér (généralisation qui contient notamment les mesures de Gibbs et les chaînes de Markov). Ici, notre schéma de preuve remplace l'utilisation classique du théorème ergodique, de même qu'il remplaçait l'utilisation de la loi des grands nombres dans le cas indépendant. Essentiellement, nous étendons le cadre de Pfister au cas où l'espace X n'est pas compact. Pour cela, il nous faut ajouter deux hypothèses : une version adaptée de la convexe-tension, la *convexe-tension locale*, et une hypothèse de *contrôle local* permettant justement d'avoir l'équivalent de la limite pour les convexes.

$$\left\{ \begin{array}{l} \text{Cas i.i.d.} \\ \text{indépendance} \\ \text{même loi} \\ \text{contrôle} \\ \text{convexe-tension} \end{array} \right. \rightarrow \left\{ \begin{array}{l} \text{Cas a.d.i.c.l.} \\ \text{découplage asymptotique} \\ \text{invariance par translation} \\ \text{contrôle local} \\ \text{convexe-tension locale} \end{array} \right.$$

Le déroulement de la preuve est ensuite analogue à celui du cas i.i.d. ; la seule différence est que nous n'avons pas une sous-additivité exacte, mais une pseudo-sous-additivité, qui n'assure que (SAC_b) (notation du chapitre 4) et non la limite pour les convexes ouverts (SAC) comme dans le cas i.i.d. Le cadre inclut complètement le cas i.i.d. du chapitre 3 et [Pfi02] pour la question $s = -p^*$. Notons que le problème de mesurabilité posé au chapitre 3 vient du fait que nous voulons considérer des suites i.i.d., dont les marginales fini-dimensionnelles sont définies naturellement comme des mesures produits, sur des tribus produits $\mathcal{F}^{\otimes n}$. Nous passons ici cette question sous le tapis en considérant des tribus initiales, rendant mesurable l'addition vectorielle, car nous devons de toute façon nous donner des lois jointes. Dans le cas du chapitre 3, ces tribus initiales sont bien égales aux tribus produits.

1.3.6 Une preuve de $s = -p^*$ à partir du théorème de Mosco

La preuve donnée dans ce chapitre a été rédigée bien avant les deux chapitres précédents. Elle étend au cas asymptotiquement découplé la démonstration de [Cer07, chapitre 26] dans le cas indépendant, inspirée des idées de [Zab92a] et [Zab92b] qui jettent un pont entre le théorème de Mosco et les grandes déviations. La preuve, quoique assez lourde et chère en hypothèses, fait apparaître une certaine uniformité dans la sous-additivité et le passage à la limite qui s'ensuit. Nous ne savons pas si le théorème de Mosco pourrait être utilisé dans la théorie de Cramér pour obtenir des résultats plus fins dans certaines situations.

1.4 Suite du programme

L'égalité $s = -p^*$ est donc toujours vérifiée dans les espaces vectoriels localement convexes mesurables. Peut-on encore étendre le cadre dans l'espoir de trouver un contre-exemple à

l'égalité dans le cas i.i.d.? Nous avons simplement donné un exemple où $p \neq (-s)^*$ (*cf.* chapitre 4), mais passer à l'égalité duale n'a pas de sens. D'autres questions restent ouvertes dans le cas i.i.d. : le PGD dans $\mathbb{R}^2$, l'étude des limites pour les convexes, notamment. Nous n'avons pas abordé la question du lien avec l'entropie de Kullback-Leibler dans le cas du théorème de Sanov : dans le cadre de [Pfi02], le découplage inférieur sert à montrer la limite thermodynamique de s , le découplage supérieur la limite pour l'entropie de Kullback-Leibler (et l'un ou l'autre pour la pression). L'approche que nous proposons permettrait peut-être aussi de simplifier, ou d'adapter au cadre de la théorie de Cramér, les preuves de certains résultats classiques en grandes déviations (*cf.* [DZ93]) : le théorème de Mogulskii et le théorème de Gärtner-Ellis, entre autres.

2

Chapitre 2

UNE PREUVE ÉLÉMENTAIRE DU THÉORÈME DE CRAMÉR DANS $\mathbb{R}$

Sommaire

2.1	Introduction	33
2.2	Énoncé	34
2.3	Démonstration	35
2.3.1	Égalité duale	35
2.3.2	Propriétés de s	37
2.3.3	Conclusion	38

2.1 Introduction

Le résultat le plus fondamental de la théorie des probabilités est certainement la loi des grands nombres pour une suite $(X_n)_{n \geq 1}$ de variables aléatoires réelles indépendantes et identiquement distribuées. Définissons la *moyenne empirique* de $(X_n)_{n \geq 1}$ par

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

La loi des grands nombres affirme que la moyenne empirique $\bar{X}_n$ converge (presque sûrement) vers la moyenne théorique $\mathbb{E}(X_1)$ à condition que $\mathbb{E}(|X_1|)$ soit fini. Après la loi des grands nombres viennent le théorème-limite central et le théorème de Cramér, qui en sont

des raffinements dans deux directions différentes. Le théorème-limite central décrit les fluctuations aléatoires de $\overline{X}_n$ autour de $\mathbb{E}(X_1)$. Le théorème de Cramér évalue la probabilité que $\overline{X}_n$ s'écarte significativement de $\mathbb{E}(X_1)$:

$$\mathbb{P}(\overline{X}_n \geq \mathbb{E}(X_1) + \varepsilon) \quad \text{for } \varepsilon > 0.$$

Un tel événement est dit de « grande déviation » car il dépend du comportement de $\overline{X}_n$ loin de la moyenne. Il s'avère que sa probabilité décroît exponentiellement vite avec n . La première estimation de ce type remonte à l'article de Cramér [Cra38] qui traite le cas de variables à densité. Dans [Che52] Chernoff relaxe cette hypothèse. Enfin, Bahadur [Bah71] donne une preuve sans aucune hypothèse sur la loi de X_1 . Issu de la mécanique statistique, Lanford importe l'argument sous-additif dans la preuve [Lan73]. Le résultat de Cramér a été ensuite généralisé dans plusieurs directions. Hammersley [Ham74] s'intéresse aux suites sous-additives dans $\mathbb{R}$; utilisant la structure d'ordre de $\mathbb{R}$, il trouve une preuve assez différente, et efficace, du résultat (nous remercions Olivier Garet de nous avoir indiqué ce texte). Bahadur et Zabell [BZ79] étendent la théorie de Cramér aux espaces vectoriels de dimension infinie, introduisant la notion de *convexe-régularité*. Les preuves classiques du théorème de Cramér dans $\mathbb{R}$ (présentées dans [Aze80], [DS89], [DZ93] et [Cer07]) reposent ou bien sur la loi des grands nombres (*cf.* [DZ93]) ou bien sur le théorème de Mosco (*cf.* [Cer07, chapitre 26], reprenant une idée de [Zab92a]). Nous exposons ici une preuve directe du théorème de Cramér dans $\mathbb{R}$ qui combine les idées de [Lan73], [Ham74] et de [BZ79], avec deux simplifications : nous démontrons d'abord une version duale du théorème de Cramér (au sens des fonctions convexes) et nous nous servons d'un conditionnement sur un convexe compact. Ce dernier argument, simplifiant l'utilisation de la convexe-régularité dans [BZ79], précise la dernière inégalité de [Ham74, (3.28)]. Non seulement la preuve est plus claire, mais elle s'adapte facilement au cadre plus général des chapitres 3 et 5.

Ce texte a été accepté pour publication dans l'*American Mathematical Monthly*. Une version anglaise est disponible à l'adresse <http://arxiv.org/abs/1002.3496> (cependant, le texte a été retravaillé depuis).

2.2 Énoncé

Théorème de Cramér. *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées. Soit $\overline{X}_n$ la moyenne empirique :*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Pour tout $x \in \mathbb{R}$, la suite

$$\frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq x)$$

converge dans $[-\infty, 0]$ et

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq x) = \inf_{\lambda \geq 0} (\log \mathbb{E}(e^{\lambda X_1}) - \lambda x)$$

2.3 Démonstration

Définissons l'entropie de la suite $(X_n)_{n \geq 1}$ par

$$\forall x \in \mathbb{R} \quad s(x) = \sup_{n \geq 1} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq x)$$

et la pression de X_1 (ou log-Laplace de la loi de X_1) par

$$\forall \lambda \in \mathbb{R} \quad p(\lambda) = \log \mathbb{E}(e^{\lambda X_1})$$

L'entropie et la pression peuvent prendre des valeurs infinies.

2.3.1 Égalité duale

Nous montrons d'abord une version duale — au sens des fonctions convexes — du théorème de Cramér.

Égalité duale. Pour tout $\lambda \geq 0$,

$$p(\lambda) = \sup_{u \in \mathbb{R}} (\lambda u + s(u))$$

Démonstration : La fameuse inégalité de Tchebychev permet de montrer une inégalité. Pour montrer l'autre inégalité, nous conditionnons la variable $\overline{X}_n$ à être bornée par un réel K avant de faire tendre K vers $+\infty$. Étant donné que $X_1, \dots, X_n$ sont indépendantes et identiquement distribuées, on a, pour tout $\lambda \geq 0$,

$$\forall n \geq 1 \quad \mathbb{E}(e^{n\lambda \overline{X}_n}) = \mathbb{E}(e^{\lambda X_1})^n$$

Ainsi, pour tous $u \in \mathbb{R}$ et $n \geq 1$, l'inégalité de Tchebychev donne

$$\begin{aligned} p(\lambda) = \log \mathbb{E}(e^{\lambda X_1}) &= \frac{1}{n} \log \mathbb{E}(e^{n\lambda \overline{X}_n}) \\ &\geq \frac{1}{n} \log \left(e^{n\lambda u} \mathbb{P}(e^{n\lambda \overline{X}_n} \geq e^{n\lambda u}) \right) \geq \lambda u + \frac{1}{n} \log \mathbb{P}(\overline{X}_n \geq u) \end{aligned}$$

D'où, en prenant le supremum en $n \geq 1$, puis en $u \in \mathbb{R}$, on obtient

$$\forall \lambda \geq 0 \quad p(\lambda) \geq \sup_{u \in \mathbb{R}} (\lambda u + s(u))$$

Montrons maintenant l'égalité pour $\lambda = 0$. La proposition suivante est utile et facile à montrer.

Inégalité utile. Pour tous $x \in \mathbb{R}$ et $n \geq 1$,

$$\mathbb{P}(X_1 \geq x)^n \leq \mathbb{P}(\overline{X}_n \geq x) \leq e^{ns(x)}$$

En particulier, $s(u) \geq \log \mathbb{P}(X_1 \geq u)$. D'où, en faisant tendre u vers $-\infty$, on voit que

$$\sup_{u \in \mathbb{R}} s(u) = 0 = p(0)$$

On prouve maintenant l'autre inégalité pour $\lambda > 0$. Soient $\lambda > 0$ et $K > 0$. Pour tout $n \geq 1$, sachant que $X_1, \dots, X_n$ sont indépendantes et identiquement distribuées, on a :

$$\begin{aligned} \log \mathbb{E}(e^{\lambda X_1} 1_{|X_1| \leq K}) &= \frac{1}{n} \log \mathbb{E}(e^{\lambda(X_1 + \dots + X_n)} 1_{|X_1| \leq K} \dots 1_{|X_n| \leq K}) \\ &\leq \frac{1}{n} \log \mathbb{E}(e^{n\lambda \overline{X}_n} 1_{|\overline{X}_n| \leq K}) \\ &= \frac{1}{n} \log \mathbb{E}\left(\left(e^{-n\lambda K} + \int_{-K}^{\overline{X}_n} n\lambda e^{n\lambda u} du\right) 1_{|\overline{X}_n| \leq K}\right) \\ &\leq \frac{1}{n} \log\left(e^{-n\lambda K} + \int_{-\infty}^{+\infty} \mathbb{E}(1_{-K \leq u \leq \overline{X}_n} 1_{|\overline{X}_n| \leq K}) n\lambda e^{n\lambda u} du\right) \end{aligned}$$

la dernière ligne étant conséquence du théorème de Fubini. Etant donné que

$$\mathbb{E}(1_{-K \leq u \leq \overline{X}_n} 1_{|\overline{X}_n| \leq K}) \leq \mathbb{P}(\overline{X}_n \geq u) 1_{|u| \leq K} \leq e^{ns(u)} 1_{|u| \leq K}$$

il vient

$$\begin{aligned} \log \mathbb{E}(e^{\lambda X_1} 1_{|X_1| \leq K}) &\leq \frac{1}{n} \log\left(e^{-n\lambda K} + \int_{-K}^K n\lambda e^{n(\lambda u + s(u))} du\right) \\ &\leq \frac{1}{n} \log\left(e^{-n\lambda K} + 2Kn\lambda \exp\left(n \sup_{u \in \mathbb{R}} (\lambda u + s(u))\right)\right) \end{aligned}$$

Soit K suffisamment grand pour que (nous rappelons que le supremum de s est 0, donc il existe $u \in \mathbb{R}$ tel que $\lambda u + s(u) > -\infty$)

$$-\lambda K < \sup_{u \in \mathbb{R}} (\lambda u + s(u))$$

Faisant tendre n vers ∞ , on obtient

$$\log \mathbb{E}(e^{\lambda X_1} 1_{|X_1| \leq K}) \leq \sup_{u \in \mathbb{R}} (\lambda u + s(u))$$

Faisant enfin tendre K vers $+\infty$, il vient

$$p(\lambda) \leq \sup_{u \in \mathbb{R}} (\lambda u + s(u))$$

□

2.3.2 Propriétés de s

Pour déduire le théorème de Cramér de l'égalité duale, nous devons montrer que la fonction s est concave.

Sous-additivité. Pour tout $x \in \mathbb{R}$, la suite $-\log \mathbb{P}(\bar{X}_n \geq x)$ est sous-additive et

$$\frac{1}{n} \log \mathbb{P}(\bar{X}_n \geq x)$$

converge dans $[-\infty, 0]$ vers $s(x)$.

Démonstration : Soit $x \in \mathbb{R}$. Supposons que $\mathbb{P}(X_1 \geq x) > 0$ et fixons $m \geq 1$. Pour $n \geq m$, soit $n = mq_n + r_n$ la division euclidienne de n par m . Alors

$$\{\bar{X}_n \geq x\} \supset \bigcap_{k=0}^{q_n-1} \left\{ \frac{1}{m} \sum_{i=m(k+1)}^{m(k+1)} X_i \geq x \right\} \cap \bigcap_{i=mq_n+1}^n \{X_i \geq x\}$$

Ainsi, comme les X_i sont indépendantes et identiquement distribuées et $r < m$,

$$\mathbb{P}(\bar{X}_n \geq x) \geq \mathbb{P}(\bar{X}_m \geq x)^{q_n} \mathbb{P}(X_1 \geq x)^{r_n} \geq \mathbb{P}(\bar{X}_m \geq x)^{q_n} \mathbb{P}(X_1 \geq x)^m$$

Passant au logarithme, divisant par n et faisant tendre n vers ∞ , il vient, sachant que $q_n/n \rightarrow 1/m$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \geq x) \geq \frac{1}{m} \log \mathbb{P}(\bar{X}_m \geq x)$$

Passant au supremum en $m \geq 1$, on conclut que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \geq x) = s(x)$$

Si x vérifie $\mathbb{P}(X_1 \geq x) = 0$, alors

$$\forall n \geq 1 \quad \frac{1}{n} \log \mathbb{P}(\bar{X}_n \geq x) = -\infty$$

donc la suite converge vers $-\infty = s(x)$. □

Concavité. La fonction $s : \mathbb{R} \rightarrow [-\infty, 0]$ est concave.

Démonstration : Pour tous $x, y \in \mathbb{R}$ et $n \geq 1$,

$$\{\bar{X}_{2n} \geq \frac{1}{2}(x+y)\} \supset \left\{ \frac{1}{n} \sum_{i=1}^n X_i \geq x \right\} \cap \left\{ \frac{1}{n} \sum_{i=n+1}^{2n} X_i \geq y \right\}$$

de sorte que, prenant le logarithme de la probabilité de chaque membre, divisant par n et faisant tendre n vers ∞ , on obtient

$$s\left(\frac{1}{2}(x+y)\right) \geq \frac{1}{2}(s(x) + s(y))$$

Par récurrence, il vient, pour tout $\alpha \in [0, 1]$ rationnel dyadique,

$$s(\alpha x + (1-\alpha)y) \geq \alpha s(x) + (1-\alpha)s(y)$$

Comme s est clairement décroissante, on en déduit que s est concave. □

2.3.3 Conclusion

Concluons la démonstration du théorème de Cramér. Pour le moment, nous savons que, pour tout $x \in \mathbb{R}$,

$$\inf_{\lambda \geq 0} (p(\lambda) - \lambda x) = \inf_{\lambda \geq 0} \sup_{u \in \mathbb{R}} (\lambda(u - x) + s(u))$$

Reste à montrer que cette dernière quantité est égale à $s(x)$. C'est un résultat standard d'analyse convexe. Donnons en une preuve élémentaire dans notre cadre. Le membre de droite de l'équation précédente est clairement supérieur à $s(x)$: il suffit de prendre $u = x$. Pour montrer l'autre inégalité, on pose

$$c = \inf \{x \in \mathbb{R} : \mathbb{P}(X_1 \geq x) = 0\}$$

et on distingue les deux cas $x < c$ et $x \geq c$.

- Supposons $x < c$. Alors $s(x) > -\infty$. Comme s est concave et décroissante, soit $-\lambda = s'_g(x) \leq 0$ la dérivée à gauche de s au point x . Alors

$$\forall u \in \mathbb{R} \quad s(u) \leq s(x) - \lambda(u - x)$$

ce qui permet de conclure.

- Supposons $x \geq c$. Notons que

$$\mathbb{P}(X_1 > c) = \lim_{\varepsilon \rightarrow 0^+} \mathbb{P}(X_1 \geq c + \varepsilon) = 0$$

Alors, pour tous $\lambda \geq 0$ et $\varepsilon > 0$,

$$\begin{aligned} p(\lambda) - \lambda x &= \log \mathbb{E} \left(e^{\lambda(X_1 - x)} (1_{X_1 < x - \varepsilon} + 1_{x - \varepsilon \leq X_1 \leq c}) \right) \\ &\leq \log \left(e^{-\lambda \varepsilon} + \mathbb{P}(X_1 \geq x - \varepsilon) \right) \end{aligned}$$

Passant à l'infimum en $\lambda \geq 0$ puis faisant tendre ε vers 0, il vient

$$\inf_{\lambda \geq 0} (p(\lambda) - \lambda x) \leq \log \mathbb{P}(X_1 \geq x) \leq s(x)$$

□

3

Chapitre 3

THÉORIE DE CRAMÉR POUR DES VARIABLES INDÉPENDANTES

Sommaire

3.1	Introduction	40
3.2	Cadre	42
3.2.1	Entropie et pression	42
3.2.2	Espaces vectoriels localement convexes mesurables	44
3.2.3	Suites de Cramér et probabilités portées par une partie	45
3.3	Théorie de Cramér	46
3.3.1	Énoncé des résultats principaux	46
3.3.2	Exemples d'applications	48
3.4	Retour sur les hypothèses et compléments	48
3.4.1	E.v.l.c.m. : définition équivalente	48
3.4.2	Convexe-tension	49
3.5	Questions de mesurabilité	51
3.5.1	Version convexes internes	51
3.5.2	Version système projectif	52
3.6	Démonstration des résultats principaux	53
3.6.1	Borne inférieure	54
3.6.2	Lemme sous-additif	55
3.6.3	Propriétés de l'entropie	58
3.6.4	Borne supérieure faible	59
3.6.5	Pression et entropie	61
3.6.6	Transformation de Fenchel-Legendre	62
3.6.7	Borne supérieure pour les convexes	64
3.6.8	Domaine de s	66

3.7 Compléments	70
3.7.1 Régularité de p	70
3.7.2 Un exemple où $p \neq (-s)^*$	70
3.7.3 Autour de $\text{cosupp}(\mu)$	71
3.7.4 Pour approfondir la notion de convexe interne	72
3.7.5 Limite pour les convexes	73
3.8 Suppléments techniques	78
3.8.1 De l'utilité des hypothèses	78
3.8.2 Quelques extensions du cadre	78
3.8.3 Questions de séparation	78
3.8.4 Tension et convexe-tension	79

3.1 Introduction

Le théorème de Cramér dans $\mathbb{R}$, une fois sa forme générale démontrée par Bahadur [Bah71], a été rapidement généralisé en dimension supérieure. Le cadre des espaces vectoriels localement convexes a été introduit par [BZ79] et semble le cadre le plus naturel pour la théorie de Cramér. L'objet du présent chapitre est, au départ, une exploration de la question

$$s = -p^*$$

et une recherche de contre-exemples. Cela a conduit à définir un cadre, le plus général possible, où la question a un sens : résoudre les problèmes de mesurabilité des moyennes empiriques $\bar{X}_n$, d'une part, et des formes linéaires λ , d'autre part. Nous ne travaillons plus avec une tribu borélienne. D'ailleurs, dans le cas du théorème de Sanov, la tribu cylindrique n'est pas la tribu borélienne de la topologie faible (ce n'est vrai que si l'espace sous-jacent est polonais). Cela nous amène à introduire le concept d'*espace vectoriel localement convexe mesurable* (*e.v.l.c.m.*) où la séparabilité assure la mesurabilité voulue. La tribu considérée est la « tribu des convexes » (analogue de la tribu des boules, dans un espace métrique, et qui peut être plus petite que la tribu borélienne; cf. [Bil68, exemple 1.4.]) qui semble plus naturelle dans la théorie de Cramér : si $\mathcal{F}$ est une tribu quelconque, notre définition de l'entropie ne dépend que des valeurs de la mesure μ , loi de X_1 , sur la tribu engendrée par les convexes mesurables.

Nous travaillons avec ce que nous appelons des *suites de Cramér* (sans leur donner de nom, [BZ79] a déjà recours à ce procédé) : étant donné une mesure μ sur un espace mesurable $(X, \mathcal{F})$ quelconque, le théorème d'extension de Kolmogorov ne permet pas, en général, de construire une mesure $\mu^{\otimes \mathbb{N}}$ (la régularité de la mesure est une condition suffisante;

cf. [Bil79, section 36.]), donc il n'existe pas nécessairement de suite $(X_n)_{n \geq 1}$ de variables aléatoires i.i.d. de loi μ . On travaille donc avec, pour tout $n \geq 1$, μ_n la loi de

$$(x_1, \dots, x_n) \in X^n \longmapsto \frac{x_1 + \dots + x_n}{n}$$

sous $\mu^{\otimes n}$ (la mesure produit fini, elle, existe bien).

Nous introduisons la notion de probabilité μ portée par un ensemble D : cela revient à dire que μ est la loi d'une variable à valeurs dans D . Dire qu'une mesure est portée par D est plus fort que de dire que son support $\text{supp}(\mu)$ est inclus dans D (il existe des mesures de probabilité de support vide, et pourtant non portées par le vide). Au départ, il s'agit simplement de clarifier un point de [Cer07, chapitre 24] : le PGD faible pour une suite de mesures portées par un même ensemble relativement compact n'assure pas le PGD (des contre-exemples sont donnés au chapitre 4). Au passage, le PGD faible en tribu cylindrique est vide si l'espace est trop gros : par exemple, dans $\mathbb{R}^{\mathbb{R}}$, il n'y a pas de mesurable relativement compact car tous les mesurables non vides sont non bornés. Pourtant, dans le cas d'une suite de Cramér, on s'attend à avoir le PGD si la mesure μ est portée par une partie D relativement compacte (cas du théorème de Sanov). C'est pour ce genre d'exemples, où l'argument de sous-additivité continue à fonctionner, que nous avons rajouté ce point technique et la notion de borne supérieure $(\text{BS}_{\flat, D})$ pour les ensembles dont l'intersection avec D est relativement compacte. La notion s'avère ensuite être utile pour passer de l'égalité $s = -p^*$ dans l'espace séparable et complet $\mathcal{C}([0, 1]; \mathbb{R})$ à la même égalité dans tout sous-espace. Or, le théorème de Banach-Mazur permet de voir que tout espace vectoriel normé séparable est un sous-espace de $\mathcal{C}([0, 1]; \mathbb{R})$. On en déduit l'égalité $s = -p^*$ dans tout e.v.l.c.m., en toute généralité.

Enfin, nous empruntons à [Mor67] l'idée de travailler avec des fonctions à valeurs dans $[-\infty, +\infty]$. Dans ce cadre, une fonction convexe est automatiquement propre, ou bien constante et égale à $\pm\infty$. Cela simplifie la terminologie et évite des vérifications de finitude (sur les transformées de Fenchel-Legendre) superflues.

Le déroulement de la preuve est le suivant : le lemme sous-additif et le *principle of the largest term* permettent d'établir le PGD faible. Nous montrons même, plus généralement, un lemme de Varadhan compact 3.6.4.2. Puis, la *convexe-tension* nous permet de passer de la borne supérieure pour les compacts $(\text{BS}_{\flat})$ à la borne supérieure pour les convexes (BS_c) . En cela, nous simplifions la preuve de [BZ79] et nous nous passons de l'hypothèse additionnelle 6.1.2 (b) de [DZ93], introduite pour rectifier la preuve de [BZ79, appendice, proposition 1]. La démonstration de $p = (-s)^*$, que nous donnons avant, n'est qu'un cas particulier du lemme de Varadhan convexe 3.6.7.1. Même si cela n'apparaît pas, car nous donnons une preuve directe de l'égalité $p = (-s)^*$ similaire à celle du chapitre 2, il s'avère que la borne supérieure pour les demi-espaces (BS_h) entraîne l'égalité $p = (-s)^*$ (cela apparaît dans la preuve de [BZ79], puis de [Aze80], et la remarque est faite dans [Cer07, proposition 16.2.]). Ensuite, l'égalité $s = -p^*$ découle de la propriété d'inversion de la transformée de Fenchel-Legendre.

Nous nous intéressons enfin à d'autres questions apparaissant naturellement dans la théorie de Cramér, la première d'entre elles étant la description du domaine $\text{dom}(s)$ de l'entropie. L'hypothèse de convexe-tension nous suffit à montrer que

$$\overline{\text{dom}(s)} = \text{cosupp}(\mu)$$

Ainsi, nous nous passons de l'hypothèse de régularité de [BZ79] et approfondissons les résultats de [Cer07] (notamment lemme 9.7.¹). Nous nous intéressons également à l'existence de la limite

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(C)$$

pour les convexes mesurables C (cf. [Sla88] pour un bon résumé des résultats pour des ensembles C non nécessairement convexes). Dans les suppléments techniques, nous évoquons la question de la séparation : elle n'est pas nécessaire pour la démonstration, elle simplifie les termes employés. En fait, la borne supérieure faible est vraie pour des ensembles vérifiant une propriété de sous-recouvrement fini par des convexes ouverts : de même que la convexe-régularité s'avère plus pertinente que la régularité, de même cette propriété de sous-recouvrement fini par des convexes ouverts est plus pertinente que la propriété de Borel-Lebesgue. Ce genre de question est aussi derrière les problèmes qui nous ont amenés à introduire l'ensemble D . Il est sans doute possible de faire une théorie de Cramér sans parler de compacts, mais en identifiant des ensembles plus naturels, vérifiant une propriété convexe de Borel-Lebesgue.

3.2 Cadre

3.2.1 Entropie et pression

Soient X un espace vectoriel réel, $\mathcal{F}$ une tribu sur X et τ une topologie séparée (au sens de Hausdorff) sur X . Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d. à valeurs dans X . On définit ses moyennes empiriques par

$$\overline{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$$

L'entropie de $(X_n)_{n \geq 1}$ est la fonction $s : X \rightarrow [-\infty, 0]$ définie par

$$\forall x \in X \quad s(x) := \inf_{\substack{A \in \mathcal{F} \\ \overset{\circ}{A} \ni x}} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in A)$$

Par construction, l'entropie s est la plus grande fonction vérifiant la borne inférieure :

(BI) pour tout $A \in \mathcal{F}$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in A) \geq \sup_{\overset{\circ}{A}} s$$

¹ Remarquons que la démonstration de [Cer07, lemme 9.8.] suppose la convexe-régularité.

On dit que $(X_n)_{n \geq 1}$ vérifie un *principe de grandes déviations* (PGD) si la borne supérieure suivante est vérifiée :

(BS) pour tout $A \in \mathcal{F}$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in A) \leq \sup_{\overline{A}} s$$

De manière générale, (BS) n'est pas vérifiée pour tous les mesurables. Si $\mathcal{P}$ désigne un ensemble de parties mesurables de X , on définit la version restreinte de la borne supérieure suivante :

(BS $_{\mathcal{P}}$) pour tout $A \in \mathcal{P}$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in A) \leq \sup_{\overline{A}} s$$

En particulier, si D est une partie de X , on notera :

(BS $_{b,D}$) pour tout $K \in \mathcal{F}$ tel que $K \cap D$ soit relativement compact,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in K) \leq \sup_{\overline{K}} s$$

(BS $_{c,D}$) pour tout $C \in \mathcal{F}$ tel que $C \cap D$ soit convexe,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C) \leq \sup_{\overline{C}} s$$

Si $(X_n)_{n \geq 1}$ vérifie (BS $_{b,X}$), on dit que $(X_n)_{n \geq 1}$ vérifie un *principe de grandes déviations faible* (PGD faible). L'objectif principal de ce texte est de donner un cadre général simple pour avoir un PGD faible et une expression simple de l'entropie en fonction de la loi de X_1 . Explicitons ce dernier point. Notons X^* le dual topologique de X . La *pression* de $(X_n)_{n \geq 1}$ est l'application $p : X^* \rightarrow [-\infty, +\infty]$ définie par

$$\forall \lambda \in X^* \quad p(\lambda) := \log \mathbb{E}(e^{\langle \lambda, X_1 \rangle})$$

Dans notre cadre, l'entropie est l'opposée de la fonction convexe-conjuguée (aussi appelée transformée de Fenchel-Legendre) de la pression et a donc pour expression

$$-p^*(x) := - \sup_{\lambda \in X^*} (\langle \lambda, x \rangle - p(\lambda))$$

On notera que les définitions mêmes de s et de p posent un problème de mesurabilité. Le cadre général que nous proposons ci-après résout ce problème. Nous ne savons pas si ce cadre contient le cadre de [BZ79] et de [Cer07] (qui prennent la mesurabilité comme hypothèse), en tout cas il est explicite et plus général que leurs applications (aux espaces de Banach séparables, aux espaces polonais, ou au théorème de Sanov en τ -topologie et tribu cylindrique).

3.2.2 Espaces vectoriels localement convexes mesurables

Soit X un espace vectoriel réel. On se propose de munir X d'une topologie et d'une tribu de sorte que l'entropie et la pression soient définies pour toute suite de variables i.i.d. donnée sur X . Si C est un convexe de X contenant 0, sa jauge (ou fonctionnelle de Minkowski) est l'application $M_C : X \rightarrow [0, +\infty]$ définie par

$$\forall x \in X \quad M_C(x) = \inf\{t \geq 0; x \in tC\}$$

La proposition suivante relie les propriétés topologiques d'un ouvert convexe à ses propriétés algébriques intrinsèques.

Propriété 3.2.2.1. *Soient (X, τ) un espace vectoriel topologique réel et C un voisinage convexe de 0. Alors M_C est finie partout et*

$$\overset{\circ}{C} = \{x \in X; M_C(x) < 1\} \quad \text{et} \quad \overline{C} = \{x \in X; M_C(x) \leq 1\}$$

Pour la preuve, on renvoie par exemple à [Bou81, II.21]. Pour les deux égalités, on peut utiliser la continuité de M_C pour montrer que $M_{\overset{\circ}{C}} = M_C = M_{\overline{C}}$. Cette proposition justifie la définition suivante : un convexe C de X est un *convexe interne* si sa jauge M_C est finie partout et si

$$C = \{x \in X; M_C(x) < 1\}$$

La notion de convexe interne est centrale dans cet article : elle fait le lien entre la structure topologique, l'outil « jauge » qui permet de mener efficacement les calculs, et surtout le lemme sous-additif qui permet de définir l'entropie. Si C est un convexe interne, $\{tC; t \in \mathbb{R}^*\}$ est un système de voisinages de 0 pour une topologie localement convexe (non nécessairement séparée) τ_C sur X (cf. [Bou81, II.25]) : on dira que τ_C est *la topologie localement convexe engendrée par C* . On dit qu'un convexe C est *symétrique* si $C = -C$. On se donne maintenant une famille $\mathcal{C}_0$ de convexes de X vérifiant les trois axiomes suivants :

(EVL CM_1) pour tout $C \in \mathcal{C}_0$, C est un convexe interne symétrique² et la topologie localement convexe τ_C engendrée par C est séparable ;

(EVL CM_2) $\mathcal{C}_0$ est stable par intersection finie et par dilatation de rapport non nul ;

(EVL CM_3) L'intersection des éléments de $\mathcal{C}_0$ est réduite à $\{0\}$.

Alors $\mathcal{C}_0$ est un système fondamental de voisinages de 0 pour une topologie localement convexe séparée³ τ sur X (cf. [Bou81, II.25]). Pour ce qui est de la tribu, notons, pour tout $x \in X$,

$$\mathcal{C}_x = \{x + C; C \in \mathcal{C}_0\}$$

² L'hypothèse symétrique n'est pas fondamentale, mais on peut toujours s'y ramener, quitte à remplacer les $C \in \mathcal{C}_0$ par $C \cap (-C)$.

³ C'est l'axiome (EVL CM_3) qui assure la séparation. L'hypothèse n'est pas nécessaire (cf. Suppléments techniques) mais nous la faisons pour deux raisons. La première d'ordre culturel : pour certains auteurs, la séparation fait partie des axiomes d'une topologie d'espace vectoriel topologique. La seconde d'ordre pratique : nous jugeons plus clair de manipuler des compacts que des quasi-compacts.

un système fondamental de voisinages convexes ouverts de x . On définit

$$\mathcal{F} = \sigma \left(\bigcup_{x \in X} \mathcal{C}_x \right)$$

On dira alors que le quadruplet $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est un *espace vectoriel localement convexe mesurable* (*e.v.l.c.m.*). On notera que $\mathcal{C}_0$ est un système fondamental de voisinages de 0 convexes, symétriques, mesurables et ouverts ; et ils vérifient donc la propriété ci-dessus. On dira que $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est *l'e.v.l.c.m. associé* à $\mathcal{C}_0$. Deux remarques pour finir ce paragraphe. La première : la topologie τ n'est pas nécessairement séparable : par exemple, $X = \mathbb{R}^{2^{\mathbb{R}}}$ muni de la topologie produit n'est pas séparable et peut être obtenu avec la construction ci-dessus. La seconde : en pratique, on se donne une famille $\tilde{\mathcal{C}}_0$ vérifiant (EVLCM₁) et (EVLCM₃) et on construit $\mathcal{C}_0$ en ajoutant à $\tilde{\mathcal{C}}_0$ les dilatés de rapports non nuls des éléments de $\tilde{\mathcal{C}}_0$, puis les intersections finies des ensembles obtenus. On vérifie alors que $\mathcal{C}_0$ est bien stable par dilatation de rapport non nul et que tout élément de $\mathcal{C}_0$ est un convexe interne symétrique engendrant une topologie localement convexe séparable.

3.2.3 Suites de Cramér et probabilités portées par une partie

Soient $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m. et μ une mesure de probabilité sur $\mathcal{F}$. Soit $(X_n)_{n \geq 1}$ une suite de variables indépendantes et de même loi μ^4 . Pour tout $n \geq 1$, on note μ_n la mesure image de $\mu^{\otimes n}$ par l'application mesurable (cf. Questions de mesurabilité)

$$(x_1, \dots, x_n) \in X^n \mapsto \frac{1}{n} \sum_{k=1}^n x_k \in X$$

autrement dit la loi de la moyenne empirique

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

On dira que $(\mu_n)_{n \geq 1}$ est la *suite de Cramér associée* à μ . Si D est une partie de X , on dira que μ est *portée par* D si, pour tout $(A, B) \in \mathcal{F}^2$,

$$A \cap D = B \cap D \Rightarrow \mu(A) = \mu(B)$$

Dans ce cas, μ est la loi d'une variable aléatoire X_1 à valeurs dans D : il suffit de considérer l'inclusion

$$X_1 : (D, \mathcal{F}|_D, \mu|_D) \hookrightarrow (X, \mathcal{F})$$

⁴ L'existence d'une telle suite est assurée, en toute généralité, par un résultat énoncé par Lomnicki et Ulam [LU34] dont la première démonstration est due à von Neumann [vN35]. Il n'y a donc pas besoin du théorème d'extension de Kolmogorov dans le cas précis des mesures produits. On renvoie à la bibliographie de [SAJ48] pour plus de détails.

où $\mathcal{F}|_D = \{A \cap D; A \in \mathcal{F}\}$ et $\mu|_D(A \cap D) = \mu(A)$ pour tout $A \in \mathcal{F}$: le fait que μ soit portée par D justifie la définition de la probabilité $\mu|_D$, appelée *probabilité trace de μ sur D* . Réciproquement, si μ_D est une probabilité sur $\mathcal{F}|_D$, on peut définir une mesure μ sur $\mathcal{F}$ par

$$\forall A \in \mathcal{F} \quad \mu(A) = \mu_D(A \cap D)$$

Alors, μ est une probabilité portée par D . En outre, si μ est portée par D , alors on peut choisir la suite $(X_n)_{n \geq 1}$ à valeurs dans D : il suffit de considérer l'inclusion

$$(X_n)_{n \geq 1} : (D^{\mathbb{N}^*}, (\mathcal{F}|_D)^{\otimes \mathbb{N}^*}, (\mu|_D)^{\otimes \mathbb{N}^*}) \hookrightarrow (X^{\mathbb{N}^*}, \mathcal{F}^{\otimes \mathbb{N}^*})$$

Ainsi, si D est convexe, $\overline{X}_n$ est également à valeurs dans D . À chaque fois qu'on considérera une probabilité μ portée par une partie convexe D de X , les variables aléatoires X_n associées seront toutes supposées à valeurs dans D .

3.3 Théorie de Cramér

3.3.1 Énoncé des résultats principaux

Soient $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m., μ une mesure de probabilité sur X et $(\mu_n)_{n \geq 1}$ la suite de Cramér associée à μ . On rappelle la notation

$$\mathcal{C}_x = \{x + C; C \in \mathcal{C}_0\}$$

Le fait que $\mathcal{C}_x$ soit un système fondamental de voisinages de x montre que l'entropie (qui est bien définie ainsi ; cf. Questions de mesurabilité) a pour expression

$$\forall x \in X \quad s(x) = \inf_{C \in \mathcal{C}_x} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C)$$

La pression est également bien définie (cf. Questions de mesurabilité) et sa fonction convexe-conjuguée a pour expression

$$\forall x \in X \quad p^*(x) := \sup_{\lambda \in X^*} \left(\langle \lambda | x \rangle - \log \mathbb{E}(e^{\langle \lambda | X_1 \rangle}) \right)$$

On définit le support de μ par

$$\text{supp}(\mu) = \{x \in X; \forall C \in \mathcal{C}_x \quad \mu(C) > 0\}$$

et on note $\text{cosupp}(\mu)$ l'enveloppe fermée convexe du support de μ , *i.e.* l'intersection de tous les ensembles convexes fermés contenant $\text{supp}(\mu)$. Si D est une partie de X et ν une probabilité sur X portée par D , on note $\nu|_D$ la probabilité trace de ν sur D . On définit enfin, pour $f : X \rightarrow [-\infty, +\infty]$, le domaine de f par

$$\text{dom}(f) = \{x \in X; f(x) \in \mathbb{R}\}$$

Théorème 3.3.1.1. Soient $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m., D une partie convexe de X et μ une probabilité sur X portée par D . Soit $(\mu_n)_{n \geq 1}$ la suite de Cramér associée à μ . Alors

- la suite $(\mu_n)_{n \geq 1}$ vérifie $(BS_{b,D})$; en particulier, $(\mu_n)_{n \geq 1}$ et $(\mu_n|_D)_{n \geq 1}$ vérifient un PGD faible; et, si D est relativement compact, $(\mu_n)_{n \geq 1}$ et $(\mu_n|_D)_{n \geq 1}$ vérifient un PGD;
- le domaine de s et l'enveloppe convexe fermée du support de μ sont contenus dans l'adhérence de D , i.e.

$$\text{dom}(s) \subset \overline{D} \quad \text{et} \quad \text{cosupp}(\mu) \subset \overline{D}$$

- l'entropie est égale à l'opposée de la convexe-conjuguée de la pression, i.e.

$$s = -p^*$$

La preuve des deux premiers points repose sur les arguments généraux de grandes déviations. L'amélioration principale à ce sujet, par rapport aux textes classiques, est la résolution des problèmes de mesurabilité. Quant à l'égalité $s = -p^*$ vraie en toute généralité, elle repose sur la notion de convexe-tension relativement à des topologies bien choisies sur X . Si D est une partie de X et μ une probabilité sur X , on dit que μ est *convexe-tendue sur D* s'il existe une suite $(K_m)_{m \geq 1}$ de parties mesurables de X telle que, pour tout $m \geq 1$, $K_m \cap D$ soit convexe et relativement compact et vérifiant

$$\lim_{m \rightarrow \infty} \mu(K_m) = 1$$

Si la mesure μ est elle-même convexe-tendue, on a de plus les résultats suivants :

Théorème 3.3.1.2. Soient $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m., D une partie convexe de X et μ une probabilité sur X portée par D . Soit $(\mu_n)_{n \geq 1}$ la suite de Cramér associée à μ . Si μ est convexe-tendue sur D , alors

- la suite $(\mu_n)_{n \geq 1}$ vérifie $(BS_{c,D})$;
- l'adhérence du domaine de s est l'enveloppe convexe fermée du support de μ , i.e.

$$\overline{\text{dom}(s)} = \text{cosupp}(\mu)$$

La convexe-tension s'avère être un outil pertinent pour passer de la borne supérieure pour les convexes compacts à la borne supérieure pour les convexes. En particulier, les demi-espaces ouverts vérifient la borne supérieure, ce qui entraîne l'égalité $p = (-s)^*$. Puis, l'égalité duale $s = -p^*$ découle du théorème de Hahn-Banach dans l'espace vectoriel localement convexe (X, τ) . La notion de convexe-tension a déjà été introduite par [BZ79] et reprise par [Cer07], mais nous avons simplifié notablement son utilisation (cf. partie sur la convexe-tension).

Remarque : Voici un exemple où $s = -p^*$ alors que μ n'est pas convexe-tendue elle-même. Sur $X = \mathbb{R}^{\mathbb{R}}$ muni de la tribu cylindrique et de la topologie produit, le premier théorème assure l'égalité $s = -p^*$. Toutefois, si μ désigne la mesure image de la mesure de Lebesgue par l'application $x \in [0, 1] \mapsto 1_{[x, x+1]}$, μ n'est pas convexe-tendue.

3.3.2 Exemples d'applications

Voici les cas les plus importants de la théorie de Cramér que contient ce nouveau cadre :

- **Théorème de Cramér dans un espace de Banach séparable** : Soient (X, τ) un espace de Banach séparable, B sa boule unité ouverte, $\tilde{\mathcal{C}}_0 = \{B\}$ (cf. la section introduisant les e.v.l.c.m.) et $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ l'e.v.l.c.m. associé. Dans ce cas, $\mathcal{C}_0$ est l'ensemble des boules ouvertes de X centrées en 0 et $\mathcal{F}$ la tribu borélienne sur X . Alors, toute suite i.i.d. $(X_n)_{n \geq 1}$ sur X vérifie le PGD faible et l'égalité $s = -p^*$. En outre, toute mesure de probabilité sur X est convexe-tendue (cf. convexe-tension). Donc, on a la borne supérieure pour tous les convexes.
- **Théorème de Cramér en topologie faible** : Soient X et Y deux espaces en dualité, $\tilde{\mathcal{C}}_0$ (cf. la section introduisant les e.v.l.c.m.) l'ensemble des bandes ouvertes de la forme $\{|\langle y, \cdot \rangle| < 1\}$ pour $y \in Y$ et $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ l'e.v.l.c.m. associé. Dans ce cas, $\tau = \sigma(X; Y)$ est la topologie faible sur X , et $\mathcal{F}$ est la tribu cylindrique sur X . Alors, toute suite de Cramér $(\mu_n)_{n \geq 1}$ sur X vérifie un PGD faible et l'égalité $s = -p^*$.
- **Théorème de Sanov** : Cas particulier du point précédent, si $(E, \mathcal{E})$ est un espace mesurable, soient Y l'espace vectoriel des fonctions mesurables bornées de $(E, \mathcal{E})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, muni de la norme $\|\cdot\|_\infty$, et $X = Y^*$. Alors, X et Y sont en dualité et l'ensemble des mesures de probabilité sur $(E, \mathcal{E})$, noté $\mathcal{M}_1^+(E)$, est un sous-ensemble convexe de X . Toute suite i.i.d. $(M_n)_{n \geq 1}$ de mesures de probabilité sur $(E, \mathcal{E})$ vérifie donc un PGD faible dans X d'entropie $s = -p^*$. La suite $(M_n)_{n \geq 1}$ vérifie également un PGD faible dans $\mathcal{M}_1^+(E)$ relativement à la τ -topologie, trace de τ sur $\mathcal{M}_1^+(E)$. De plus, comme $\mathcal{M}_1^+(E)$ est relativement compact, $(M_n)_{n \geq 1}$ vérifie un PGD.

3.4 Retour sur les hypothèses et compléments

3.4.1 E.v.l.c.m. : définition équivalente

Soient X un espace vectoriel réel et

$$\overleftarrow{\mathcal{X}} = (N_i, f_i, f_{ij})_{i \leq j}$$

un système projectif d'espaces vectoriels normés séparables, autrement dit une famille telle que

(PROJ₁) les indices i et j décrivent un ensemble $(J, \leq)$ préordonné filtrant à droite ;

(PROJ₂) pour tout $i \in J$, N_i est un espace vectoriel normé séparable, f_i une application linéaire de X dans N_i et, pour tout $(i, j) \in J^2$ tel que $i \leq j$, f_{ij} est une application linéaire continue de N_j dans N_i ;

(PROJ₃) pour tout $(i, j, k) \in J^3$ tel que $i \leq j \leq k$, on a $f_{ii} = id_{N_i}$,

$$f_i = f_{ij} \circ f_j \quad \text{et} \quad f_{ik} = f_{ij} \circ f_{jk}$$

On note, pour tout $i \in J$, B_i la boule unité ouverte de N_i . On vérifie alors que

$$\mathcal{C}_0 = \{f_i^{-1}(t_i B_i); i \in J, t_i \in]0, +\infty[\}$$

satisfait aux trois axiomes (EVL CM_1), (EVL CM_2) et (EVL CM_3). De plus, on vérifie que, si $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est l'e.v.l.c.m. associé à $\mathcal{C}_0$, alors τ (resp. $\mathcal{F}$) est la topologie (resp. tribu) initiale pour la famille $(N_i, f_i)_{i \in J}$. En effet, la topologie initiale est définie ainsi. Quant à la tribu, pour tout $i \in I$, si $\mathcal{B}_i$ désigne la tribu borélienne de N_i , étant donné que N_i est normé et séparable, $\mathcal{B}_i$ est la tribu engendrée par les $tB_i + u$, pour $t > 0$ et $u \in N_i$. Avec cette approche, on pourra dire que $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est l'e.v.l.c.m. associé au système projectif $\overleftarrow{\mathcal{X}}$. Réciproquement, si $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est un e.v.l.c.m., pour tout $C \in \mathcal{C}_0$, on peut définir $N_C = X / \{x \in X; M_C(x) = 0\}$ l'espace séparé associé à (X, M_C) et f_C la surjection canonique : N_C est un espace vectoriel normé et séparable. De plus, si $B \subset C$, N_C s'injecte canoniquement dans N_B : on note f_{CB} l'injection associée. Alors, $\overleftarrow{\mathcal{X}} = (N_C, f_C, f_{CB})_{B \subset C}$ est un système projectif d'espaces vectoriels normés séparables et $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est l'e.v.l.c.m. associé à $\overleftarrow{\mathcal{X}}$.

Remarque : L'analogue d'une famille $\tilde{\mathcal{C}}_0$ serait une structure initiale d'espaces vectoriels normés séparables. A partir d'une telle famille d'espaces, il est facile de construire un système projectif définissant les mêmes topologie et tribu sur X .

Remarque : On peut également définir la notion de système projectif d'e.v.l.c.m. avec des applications linéaires mesurables. Il s'avère que la limite projective d'un tel système projectif est encore un e.v.l.c.m. (cf. chapitre 5).

3.4.2 Convexe-tension

Dans cette section, nous donnons quelques propriétés de la convexe-tension. Soient X un espace vectoriel réel, $\mathcal{F}$ une tribu sur X , τ une topologie séparée sur X , μ une probabilité sur X et D une partie convexe de X . On dit que μ est *convexe-tendue sur D* s'il existe une suite $(K_m)_{m \geq 1}$ de parties mesurables de X telle que, pour tout $m \geq 1$, $K_m \cap D$ soit un convexe relativement compact et vérifiant

$$\lim_{m \rightarrow \infty} \mu(K_m) = 1$$

Si A est une partie de X , on dit que A est convexe (resp. relativement compacte) *sur D* si $A \cap D$ est convexe (resp. relativement compacte). Ici, pour tout $m \geq 1$, K_m est convexe sur D et relativement compact sur D . Conséquence immédiate de la convexe-tension : pour tout $C \in \mathcal{F}$ convexe sur D et pour tout $m \geq 1$, $K_m \cap C$ est mesurable, convexe sur D , relativement compact sur D et inclus dans C et on a

$$\lim_{m \rightarrow \infty} \mu(K_m \cap C) = \mu(C)$$

Mentionnons une autre notion introduite dans [BZ79] : on dit que μ est *convexe-régulière sur D* si, pour tout $C \in \mathcal{F}$ convexe ouvert, il existe une suite $(K_m)_{m \geq 1}$ de parties mesurables

de X , convexes sur D et relativement compactes sur D telle que, pour tout $m \geq 1$, $\overline{K_m} \subset C$, et vérifiant

$$\lim_{m \rightarrow \infty} \mu(K_m) = \mu(C)$$

Voici le lien entre les deux notions :

Proposition 3.4.2.1. *Soient (X, τ) un espace vectoriel localement convexe séparé, $\mathcal{F}$ une tribu sur X stable par translation et dilatation de rapport non nul, μ une mesure de probabilité sur X et D une partie convexe de X . Alors μ est convexe-régulière sur D si et seulement si μ est convexe-tendue sur D .*

Remarque : Ce résultat permet de clarifier un point de l'annexe de [BZ79]. Et, par la même occasion, d'alléger l'hypothèse $(\hat{C})$ introduite par [DS89] en réponse à cette imprécision, et reprise dans les textes suivants, notamment [DZ93, 6.1.2. (b)].

Démonstration : L'implication directe est immédiate. Pour la réciproque, soit $C \in \mathcal{F}$ convexe ouvert. Comme $\mathcal{F}$ est invariante par translation, on peut supposer que $0 \in C$. Etant donné que $\mathcal{F}$ est stable par dilatation de rapport non nul, la propriété 3.2.2.1 permet d'écrire :

$$\overline{C} = \{x \in X; M_C(x) \leq 1\} = \bigcap_{\substack{r \in \mathbb{Q} \\ r > 1}} \{x \in X; M_C(x) < r\} = \bigcap_{\substack{r \in \mathbb{Q} \\ r > 1}} rC$$

Aussi $\overline{C}$ est-il mesurable, de même que ses dilatés de rapports non nuls. Définissons, pour tout $m \geq 1$,

$$C_m = \left(1 - \frac{1}{m+1}\right) \overline{C}$$

de sorte que $C_m \subset C$ et $\mu(C_m) \rightarrow \mu(C)$. Soit $(K_m)_{m \geq 1}$ une suite de parties de X mesurables, convexes sur D et relativement compactes sur D telle que $\mu(K_m) \rightarrow 1$. Alors, pour tout $m \geq 1$, $C_m \cap K_m$ est mesurable, convexe sur D , relativement compact sur D , d'adhérence incluse dans C et vérifie

$$\mu(C_m \cap K_m) \geq \mu(C_m) - (1 - \mu(K_m)) \rightarrow \mu(C) \quad \square$$

Soient maintenant $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m., D une partie convexe de X , μ une probabilité sur X portée par D et $(\mu_n)_{n \geq 1}$ la suite de Cramér associée à μ . Si μ est convexe-tendue sur D et si $(K_m)_{m \geq 1}$ est une suite de parties mesurables, convexes sur D et relativement compactes sur D telle que $\mu(K_m) \rightarrow 1$, alors, pour tout $n \geq 1$, $\mu_n(K_m) \rightarrow 1$: donc les μ_n sont toutes convexes-tendues sur D (au passage, la suite $(K_m)_{m \geq 1}$ est commune aux μ_n). En effet, si $n \geq 1$, on a :

$$\mu_n(K_m) = \mathbb{P}(\overline{X_n} \in K_m) \geq \mathbb{P}(X_1 \in K_m; \dots; X_n \in K_m) = \mu(K_m)^n \rightarrow 1$$

Cette remarque permet de simplifier les hypothèses de [Cer07, chapitre 9]. On dit qu'un e.v.l.c. (X, τ) est un espace de Fréchet s'il est complet et si 0 admet un système fondamentale de voisinages dénombrable (cette seconde condition est équivalente à la métrisabilité).

Proposition 3.4.2.2. *Si (X, τ) est un espace de Fréchet séparable et $\mathcal{F}$ sa tribu borélienne, alors toute probabilité sur X est convexe-tendue sur X .*

Les deux arguments de la preuve sont les suivants : toute probabilité sur X est tendue (cf. [Bil68, théorème 1.3.]) et l'enveloppe convexe fermée d'un compact de X est compacte, d'après le théorème de Krein (cf. [Bou81, IV.37]). En particulier, le résultat est vrai sur un espace de Banach séparable muni de sa tribu borélienne.

3.5 Questions de mesurabilité

La séparabilité (ou plutôt la seconde dénombrabilité) est le nerf de la mesurabilité. Le but de cette partie est de montrer que les fonctions s et p sont bien définies, sur X et X^* respectivement. Si $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est un e.v.l.c.m. et $(X_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans X , pour pouvoir définir l'entropie, on souhaite que les fonctions $\overline{X}_n$ soient des variables aléatoires. Etant donné que $\mathcal{F}$ est stable par dilatation de rapport non nul, la mesurabilité requise est conséquence de la mesurabilité de l'addition vectorielle. Pour ce qui est de la pression, il s'agit de vérifier que toute forme linéaire continue est mesurable. Nous proposons deux démonstrations de ces deux résultats, correspondant aux deux définitions équivalentes des e.v.l.c.m.

3.5.1 Version convexes internes

Proposition 3.5.1.1. *Soit $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m. L'addition vectorielle*

$$(x, y) \in (X^2, \mathcal{F}^{\otimes 2}) \mapsto x + y \in (X, \mathcal{F})$$

est mesurable.

Démonstration : Il suffit de vérifier que, pour tout $C \in \mathcal{C}_0$,

$$\{(x, y) \in X^2; x + y \in C\} \in \mathcal{F}^{\otimes 2}$$

Soit $C \in \mathcal{C}_0$. Soit Q une partie de X dénombrable et dense pour la topologie localement convexe τ_C engendrée par C . Montrons que

$$\{(x, y) \in X^2; x + y \in C\} = \bigcup_{\substack{u \in Q \\ r \in \mathbb{Q} \cap]0, 2[}} \left\{ (x, y) \in X^2; \begin{array}{l} M_C(x - u) < r \\ M_C(y + u) < 1 - r \end{array} \right\}$$

Pour l'inclusion $\subset$, soit $(x, y) \in X^2$ tel que $x + y \in C$. Notons $a = M_C(x + y) < 1$. Soit $r \in \mathbb{Q} \cap]0, (1 - a)/2[$. L'ouvert non vide $x + rC$ contient alors un point u de Q . Et on a $M_C(x - u) < r$ (car C est symétrique) et

$$M_C(y + u) \leq M_C(y + x) + M_C(u - x) < a + r < 1 - r$$

Pour l'inclusion $\supset$, on note que, pour tout $(x, y, u) \in X^3$,

$$M_C(x + y) \leq M_C(x - u) + M_C(y + u)$$

d'où le résultat en choisissant convenablement $u \in Q$. $\square$

Lemme 3.5.1.2. *Soit X un espace vectoriel. Soient B et C deux convexes internes de X . Si $B \subset C$ et si la topologie localement convexe τ_B engendrée par B est séparable, alors C s'écrit comme réunion dénombrable de translatés et de dilatés de rapports non nuls de B .*

Démonstration : Soit Q une partie de X dénombrable et dense relativement à τ_B . Montrons que

$$C \subset \bigcup_{\substack{u \in Q, r \in \mathbb{Q}_+ \\ u + rB \subset C}} u + rB$$

l'autre inclusion étant immédiate. Soit $x \in C$. Comme C est interne, $M_C(x) < 1$ et il existe $r \in \mathbb{Q}_+$ tel que $0 < r < 1 - M_C(x)$; ainsi $x + rB \subset C$. L'ouvert non vide $x + \frac{1}{2}(rB \cap (-rB))$ relativement à τ_B contient un point u de Q . On a alors

$$u \in x - \frac{r}{2}B \quad \text{et} \quad u + \frac{r}{2}B \subset x + rB \subset C$$

d'où

$$x \in u + \frac{r}{2}B \subset C \quad \square$$

Proposition 3.5.1.3. *Soit $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m. Toute forme linéaire continue sur X est mesurable.*

Démonstration : Soit $\lambda \in X^*$. Soit H_λ le demi-espace ouvert $\{x \in X; \langle \lambda | x \rangle < 1\}$. Comme H_λ est un voisinage de 0, il existe $C \in \mathcal{C}_0$ tel que $C \subset H_\lambda$. Le lemme précédent montre alors que H_λ est réunion dénombrable de translatés et dilatés de rapports non nuls de C , donc est mesurable. Comme $\mathcal{F}$ est stable par translation, λ est mesurable. $\square$

3.5.2 Version système projectif

Proposition 3.5.2.1. *Soit $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m. L'addition vectorielle*

$$(x, y) \in (X^2, \mathcal{F}^{\otimes 2}) \mapsto x + y \in (X, \mathcal{F})$$

est mesurable.

Démonstration : Supposons que $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ soit l'e.v.l.c.m. associé au système projectif $\overleftarrow{\mathcal{X}} = (N_i, f_i, f_{ij})_{i \leq j}$. Il suffit de montrer que, pour tout $i \in I$, l'application composée $f_i(\cdot + \cdot)$

$$(x, y) \in (X^2, \mathcal{F}^{\otimes 2}) \mapsto f_i(x) + f_i(y) \in (N_i, \mathcal{B}_i)$$

est mesurable. Cette application apparaît comme le chemin du haut du schéma commutatif

$$\begin{array}{ccc} (x, y) \in (X^2, \mathcal{F}^{\otimes 2}) & \longrightarrow & x + y \in (X, \mathcal{F}) \\ \downarrow & & \downarrow \\ (f_i(x), f_i(y)) \in (N_i^2, \mathcal{B}_i^{\otimes 2}) & \longrightarrow & f_i(x) + f_i(y) \in (N_i, \mathcal{B}_i) \end{array}$$

Or, le chemin du bas est mesurable : la flèche verticale est un produit d'applications mesurables, et la flèche horizontale, addition vectorielle dans l'espace à base dénombrable N_i , est continue, donc mesurable. $\square$

Proposition 3.5.2.2. *Soit $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m. Toute forme linéaire continue sur X est mesurable.*

Démonstration : Supposons que $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ soit l'e.v.l.c.m. associé au système projectif $\overleftarrow{\mathcal{X}} = (N_i, f_i, f_{ij})_{i \leq j}$. Il s'agit simplement de voir que

$$X^* = \{\lambda_i \circ f_i; i \in J, \lambda_i \in N_i^*\}$$

Soit $\lambda \in X^*$. Comme λ est continue, il existe $C \in \mathcal{C}_0$ tel que

$$C \subset \{x \in X; \langle \lambda | x \rangle < 1\}$$

Ecrivons C sous la forme $f_i^{-1}(t_i B_i)$. On voit alors que

$$\forall x \in X \quad f_i(x) = 0 \Rightarrow \langle \lambda | x \rangle = 0$$

En effet, si $\langle \lambda | x \rangle \neq 0$, il existe $t \in \mathbb{R}$ tel que $\langle \lambda | tx \rangle > 1$, donc $f_i(tx) \notin t_i B_i$, d'où $f_i(x) \neq 0$. On en déduit que l'on peut passer au quotient dans

$$\begin{array}{ccc} x \in X & \longrightarrow & \langle \lambda | x \rangle \in \mathbb{R} \\ \searrow & & \nearrow \\ & f_i(x) \in N_i & \end{array}$$

et l'application quotient est une forme linéaire continue sur N_i . $\square$

3.6 Démonstration des résultats principaux

Tout au long de cette partie, $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ désigne un e.v.l.c.m. et, pour tout $x \in X$,

$$\mathcal{C}_x = \{x + C; C \in \mathcal{C}_0\}$$

est un système fondamental de voisinages de x ouverts, convexes et mesurables. On se donne aussi D une partie convexe de X , μ une probabilité sur X portée par D et $(\mu_n)_{n \geq 1}$ la suite de Cramér associée à μ .

3.6.1 Borne inférieure

Nous avons vu que l'entropie s définie par

$$\forall x \in X \quad s(x) = \inf_{C \in \mathcal{C}_x} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C)$$

est la plus grande fonction vérifiant la borne inférieure (BI). Montrons un résultat un peu plus fort. On étend l'addition à $[-\infty, +\infty]$ via

$$-\infty \dagger (+\infty) = -\infty$$

De plus, si $f : X \rightarrow [-\infty, +\infty]$, on dit que f est *semi-continue inférieurement* si, pour tout $t \in [-\infty, +\infty]$, l'ensemble

$$\{x \in X ; f(x) > t\}$$

est un ouvert de X . On note $f_\bullet$ la régularisée semi-continue inférieurement de f , autrement dit la plus grande fonction semi-continue inférieurement qui soit inférieure à f .

Théorème 3.6.1.1 (Varadhan ouvert). *Pour toute fonction $f : X \rightarrow [-\infty, +\infty]$ mesurable,*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}(e^{nf(\overline{X}_n)}) \geq \sup_X (f_\bullet \dagger s)$$

Démonstration : Soit $f : X \rightarrow [-\infty, +\infty]$ mesurable. Soient $x \in X$, $\delta > 0$ et $M > 0$. Par définition de $f_\bullet$, il existe $C \in \mathcal{C}_x$ tel que, pour tout $y \in C$,

$$f(y) \geq \min(f_\bullet(x) - \delta, M)$$

On a alors, pour tout $n \geq 1$:

$$\begin{aligned} \frac{1}{n} \log \mathbb{E}(e^{nf(\overline{X}_n)}) &\geq \frac{1}{n} \log \mathbb{E}(e^{nf(\overline{X}_n)} 1_{\overline{X}_n \in C}) \\ &\geq \min(f_\bullet(x) - \delta, M) + \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C) \end{aligned}$$

Prenant la limite inférieure en n , il vient

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}(e^{nf(\overline{X}_n)}) \geq \min(f_\bullet(x) - \delta, M) + s(x)$$

On conclut en faisant tendre δ vers 0 et M vers $+\infty$, puis en prenant le supremum en $x \in X$. $\square$

En particulier, si $f : X \rightarrow \mathbb{R}$ est mesurable et $A \in \mathcal{F}$, le résultat précédent appliqué à $f - \delta_A$ (où $\delta_A = +\infty 1_{X \setminus A}$) donne

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}(e^{nf(\overline{X}_n)} 1_{\overline{X}_n \in A}) \geq \sup_A (f_\bullet \dagger s)$$

que l'on peut voir comme une borne inférieure généralisée. D'autre part, on rappelle que la pression est définie (et bien définie; cf. Questions de mesurabilité) par

$$\forall \lambda \in X^* \quad p(\lambda) := \log \mathbb{E}(e^{\langle \lambda | X_1 \rangle})$$

Corollaire 3.6.1.2. *On a l'inégalité*

$$\forall \lambda \in X^* \quad \forall x \in X \quad p(\lambda) - s(x) \geq \langle \lambda | x \rangle$$

Démonstration : Soit $\lambda \in X^*$. Il suffit de remarquer que

$$p(\lambda) = \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}(e^{\langle \lambda | \overline{X}_n \rangle})$$

et d'appliquer le théorème précédent (3.6.1.1) à la fonction mesurable et continue $f = \lambda$. $\square$

Remarque : On obtient aussi l'inégalité équivalente

$$\forall x \in X \quad s(x) \leq \inf_{\lambda \in X^*} (p(\lambda) - \langle \lambda | x \rangle) = -p^*(x)$$

Notons que notre démonstration fait bien apparaître que l'inégalité $p \geq (-s)^*$ (et $s \leq -p^*$) est une borne inférieure. La démonstration classique de ce point utilise l'inégalité de Tchebychev. Nous la reproduisons ici. Soient $\lambda \in X^*$ et $x \in X$. Définissons, pour $\varepsilon > 0$, le demi-espace ouvert (mesurable)

$$H = \{y \in X; \langle \lambda | y \rangle > \langle \lambda | x \rangle - \varepsilon\}$$

Si $n \geq 1$, l'inégalité de Tchebychev donne

$$\begin{aligned} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in H) &= \frac{1}{n} \log \mathbb{P}(n\langle \lambda | \overline{X}_n \rangle - n\langle \lambda | x \rangle + n\varepsilon > 0) \\ &\leq \frac{1}{n} \log \mathbb{E}[\exp(n\langle \lambda | \overline{X}_n \rangle - n\langle \lambda | x \rangle + n\varepsilon)] = p(\lambda) - \langle \lambda | x \rangle + \varepsilon \end{aligned}$$

car, pour tout $n \geq 1$, $\mathbb{E}(\exp\langle \lambda | n\overline{X}_n \rangle) = \mathbb{E}(\exp\langle \lambda | X_1 \rangle)^n$. Etant donné que $H \in \mathcal{C}_x$, en prenant le supremum en $n \geq 1$, on obtient $s(x) \leq p(\lambda) - \langle \lambda | x \rangle + \varepsilon$. Puis, passant à l'infimum en $\varepsilon > 0$, il vient

$$p(\lambda) - s(x) \geq \langle \lambda | x \rangle$$

3.6.2 Lemme sous-additif

Lemme 3.6.2.1. *Soit $(u(n))_{n \geq 1}$ une suite à valeurs dans $[0, +\infty]$. On suppose que*

(SA) u est sous-additive, i.e.

$$\forall m, n \geq 1 \quad u(m+n) \leq u(m) + u(n)$$

Alors,

$$\liminf_{n \rightarrow \infty} \frac{u(n)}{n} = \inf_{n \geq 1} \frac{u(n)}{n}$$

Démonstration : Par sous-additivité, pour tous $d, m \geq 1$,

$$\frac{u(dm)}{dm} \leq \frac{u(m)}{m}$$

En faisant tendre d vers ∞ , on obtient

$$\liminf_{n \rightarrow \infty} \frac{u(n)}{n} \leq \liminf_{d \rightarrow \infty} \frac{u(dm)}{dm} \leq \frac{u(m)}{m}$$

d'où le résultat en passant à l'infimum en $m \geq 1$. □

Lemme 3.6.2.2. Soit $(u(n))_{n \geq 1}$ une suite à valeurs dans $[0, +\infty]$. On suppose que

(SA) u est sous-additive ;

(C) u est contrôlée, i.e. il existe $N \geq 1$ tel que

$$\forall n \geq N \quad u(n) < +\infty$$

Alors, la suite $(u(n)/n)_{n \geq 1}$ converge vers

$$\inf_{n \geq 1} \frac{u(n)}{n}$$

Démonstration : Soient $n \geq m \geq N$. La division euclidienne de n par m s'écrit $n = mq + r$ avec $q \geq 1$ et $r \in \{0, \dots, m-1\}$; ainsi, la sous-additivité permet d'écrire

$$u(n) = u(mq + r) = u(m(q-1) + m + r) \leq (q-1)u(m) + u(m+r)$$

puis

$$\frac{u(n)}{n} \leq \frac{u(m)}{m} + \frac{1}{n} \max_{0 \leq i < m} u(m+i)$$

D'où, comme u est contrôlée, en faisant tendre n , puis m vers ∞ , on obtient

$$\limsup_{n \rightarrow \infty} \frac{u(n)}{n} \leq \liminf_{m \rightarrow \infty} \frac{u(m)}{m}$$

autrement dit la suite $(u(n)/n)_{n \geq 1}$ converge. D'après le lemme 3.6.2.1, sa limite est

$$\inf_{n \geq 1} \frac{u(n)}{n}$$

□

On étend l'addition à $[-\infty, +\infty]$ via

$$-\infty \dot{+} (+\infty) = +\infty$$

Soit $f : X \rightarrow [-\infty, +\infty]$. On dit que f est *concave sur* D si

$$\forall (x, y) \in D^2 \quad \forall \alpha \in [0, 1] \quad f(\alpha x + (1-\alpha)y) \geq \alpha f(x) \dot{+} (1-\alpha)f(y)$$

Si A est une partie de X , on dit que A est convexe (resp. relativement compacte) *sur* D si $A \cap D$ est convexe (resp. relativement compacte). La proposition suivante relie la convexité et la sous-additivité :

Proposition 3.6.2.3. *Pour toute fonction mesurable $f : X \rightarrow [-\infty, +\infty]$ concave sur D ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left(e^{nf(\overline{X}_n)} \right) = \sup_{n \geq 1} \frac{1}{n} \log \mathbb{E} \left(e^{nf(\overline{X}_n)} \right)$$

Démonstration : Soit $f : X \rightarrow [-\infty, +\infty]$ mesurable et concave sur D . On vérifie que la suite de terme général

$$u(n) = -\log \mathbb{E} \left(e^{nf(\overline{X}_n)} \right)$$

est sous-additive. Soient $m, n \geq 1$. Comme f est concave sur D et la suite $(X_k)_{1 \leq k \leq n}$ i.i.d., on a :

$$\mathbb{E} \left(e^{(m+n)f(\overline{X}_{m+n})} \right) \geq \mathbb{E} \left(e^{mf(\overline{X}_m) + nf(\overline{X}_n)} \right) = \mathbb{E} \left(e^{mf(\overline{X}_m)} \right) \mathbb{E} \left(e^{nf(\overline{X}_n)} \right)$$

On conclut à l'aide du lemme 3.6.2.1. □

Proposition 3.6.2.4. *Pour tout $C \in \mathcal{F}$ convexe sur D ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C) = \sup_{n \geq 1} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C)$$

Si, de plus, C est le translaté d'un convexe interne (en particulier si C est un convexe ouvert), alors

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C) = \sup_{n \geq 1} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C)$$

Démonstration : Soit $C \in \mathcal{F}$ convexe sur D . Pour le premier point, il suffit d'appliquer le lemme précédent à la fonction $-\delta_C$ mesurable et concave sur D . Supposons maintenant qu'il existe $x \in X$ tel que $C - x$ soit interne. Pour simplifier les notations, on se ramène au cas où $x = 0$. Ou bien $u(n) = +\infty$ pour tout $n \geq 1$ auquel cas on a bien la convergence voulue, ou bien il existe $m \geq 1$ tel que $u(m) < +\infty$. Dans ce cas, nous allons montrer que la suite $(u(n))_{n \geq 1}$ est contrôlée, i.e. vérifie (C). On a :

$$0 < \mathbb{P}(\overline{X}_m \in C) = \mathbb{P}(M_C(\overline{X}_m) < 1) = \lim_{\varepsilon \rightarrow 0} \mathbb{P}(M_C(\overline{X}_m) < 1 - \varepsilon)$$

On en déduit qu'il existe $\varepsilon > 0$ tel que

$$\mathbb{P}(M_C(\overline{X}_m) < 1 - \varepsilon) > 0$$

Soit $n \geq 1$. Ecrivons la division euclidienne de n par m sous la forme $n = mq + r$ avec $r \in \{1, \dots, m\}$. La convexité et la $\mathbb{R}_+$ -linéarité de M_C permettent de montrer que

$$M_C(\overline{X}_n) \leq \frac{m}{n} \left(\sum_{k=0}^{q-1} M_C \left(\frac{1}{m} \sum_{i=m(k+1)}^{m(k+1)} X_i \right) \right) + \sum_{i=mq+1}^n \frac{M_C(X_i)}{n}$$

Sachant que $m/n \leq 1/q$ et que $(X_k)_{1 \leq k \leq n}$ est i.i.d., on peut écrire

$$\begin{aligned} \mathbb{P}(M_C(\overline{X}_n) < 1) &\geq \mathbb{P}\left(\forall k \in \{0, \dots, q-1\} M_C\left(\frac{1}{m} \sum_{i=m(k+1)}^{m(k+1)} X_i\right) < 1 - \varepsilon ; \right. \\ &\quad \left. \forall i \in \{mq+1, \dots, n\} M_C(X_i) < \frac{n\varepsilon}{r}\right) \\ &\geq \mathbb{P}\left(M_C(\overline{X}_m) < 1 - \varepsilon\right)^q \mathbb{P}\left(M_C(X_1) < \frac{n\varepsilon}{m}\right)^m \end{aligned}$$

Comme M_C est finie partout,

$$1 = \lim_{t \rightarrow +\infty} \mathbb{P}\left(M_C(X_1) < t\right)$$

donc il existe $t \in]0, +\infty[$ tel que $\mathbb{P}(M_C(X_1) < t) > 0$. Alors,

$$\forall n \geq \frac{mt}{\varepsilon} \quad \mathbb{P}(\overline{X}_n \in C) \geq \mathbb{P}(M_C(\overline{X}_m) < 1 - \varepsilon)^q \mathbb{P}(M_C(X_1) < t)^m > 0$$

donc la suite $(u(n))_{n \geq 1}$ est contrôlée, et la conclusion découle du lemme 3.6.2.2. $\square$

Remarque : En vertu du lemme 3.6.4.1, le premier point de la proposition précédente est vrai pour C union finie de parties mesurables convexes sur D et le second pour C union finie de convexes internes mesurables.

3.6.3 Propriétés de l'entropie

Proposition 3.6.3.1. *La fonction $s : X \rightarrow [-\infty, 0]$ définie par*

$$s(x) = \inf_{C \in \mathcal{C}_x} \sup_{n \geq 1} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C)$$

est égale à l'entropie de la suite $(\mu_n)_{n \geq 1}$. De plus, elle est semi-continue supérieurement et concave.

Démonstration : En vertu de la proposition précédente,

$$s(x) = \inf_{C \in \mathcal{C}_x} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C)$$

donc s est l'entropie. Montrons la semi-continuité supérieure de s . Soient $t \in \mathbb{R}$ et $x \in X$ tels que $s(x) < t$. Par définition de $s(x)$, il existe $C \in \mathcal{C}_0$ tel que

$$\sup_{n \geq 1} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in x + C) < t$$

Alors, pour tout $y \in x + C$, il existe $\varepsilon > 0$ tel que $y + \varepsilon C \subset x + C$, et ainsi

$$s(y) \leq \sup_{n \geq 1} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in y + \varepsilon C) \leq \sup_{n \geq 1} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in x + C) < t$$

On en déduit que s est semi-continue supérieurement. Vérifions que s est concave. Soient $x, y \in X$ et $z = (x + y)/2$. Soit $C \in \mathcal{C}_z$. Par continuité de l'addition, il existe $A \in \mathcal{C}_x$ et $B \in \mathcal{C}_y$ tels que $(A + B)/2 \subset C$. Alors

$$\mathbb{P}(\overline{X}_{2n} \in C) \geq \mathbb{P}(\overline{X}_n \in A) \mathbb{P}(\overline{X}_n \in B)$$

puis

$$\lim_{n \rightarrow \infty} \frac{1}{2n} \log \mathbb{P}(\overline{X}_{2n} \in C) \geq \lim_{n \rightarrow \infty} \frac{1}{2n} \left(\log \mathbb{P}(\overline{X}_n \in A) + \log \mathbb{P}(\overline{X}_n \in B) \right) \geq \frac{s(x) + s(y)}{2}$$

On en déduit que $s((x + y)/2) \geq (s(x) + s(y))/2$. La concavité de s en découle, sachant que s est semi-continue supérieurement. $\square$

3.6.4 Borne supérieure faible

La borne supérieure pour les compacts repose sur l'interversion infimum-supremum qui fait l'objet du lemme suivant.

Lemme 3.6.4.1. *Si $(u_1(n))_{n \geq 1}, \dots, (u_r(n))_{n \geq 1}$ sont r suites à valeurs dans $[0, +\infty]$, on a l'égalité*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^r u_i(n) = \max_{1 \leq i \leq r} \limsup_{n \rightarrow \infty} \frac{1}{n} \log u_i(n)$$

Démonstration : De l'encadrement (on rappelle que les suites sont positives)

$$\max_{1 \leq i \leq r} u_i(n) \leq \sum_{i=1}^r u_i(n) \leq r \max_{1 \leq i \leq r} u_i(n)$$

on déduit que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^r u_i(n) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log \max_{1 \leq i \leq r} u_i(n)$$

Puis

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \max_{1 \leq i \leq r} u_i(n) &= \lim_{n \rightarrow \infty} \sup_{k \geq n} \frac{1}{k} \log \max_{1 \leq i \leq r} u_i(k) \\ &= \lim_{n \rightarrow \infty} \max_{1 \leq i \leq r} \left(\sup_{k \geq n} \frac{1}{k} \log u_i(k) \right) \\ &= \max_{1 \leq i \leq r} \lim_{n \rightarrow \infty} \left(\sup_{k \geq n} \frac{1}{k} \log u_i(k) \right) \\ &= \max_{1 \leq i \leq r} \limsup_{n \rightarrow \infty} \frac{1}{n} \log u_i(n) \end{aligned}$$

La troisième égalité vient du fait facile à vérifier que

$$\max : [-\infty, +\infty]^r \rightarrow [-\infty, +\infty]$$

est continue. $\square$

Pour achever la démonstration du PGD faible, on montre un résultat un peu plus fort. On étend l'addition à $[-\infty, +\infty]$ via

$$-\infty \dot{+} (+\infty) = +\infty$$

De plus, si $f : X \rightarrow [-\infty, +\infty]$, on dit que f est *semi-continue supérieurement* si, pour tout $t \in [-\infty, +\infty]$, l'ensemble

$$\{x \in X ; f(x) < t\}$$

est un ouvert de X . On note $f^\bullet$ la régularisée semi-continue supérieurement de f , autrement dit la plus petite fonction semi-continue supérieurement qui soit supérieure à f .

Théorème 3.6.4.2 (Varadhan compact). *Pour toute fonction $f : X \rightarrow [-\infty, +\infty]$ mesurable et pour tout $K \in \mathcal{F}$ relativement compact sur D ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left(e^{nf(\bar{X}_n)} 1_{\bar{X}_n \in K} \right) \leq \sup_{\bar{K} \cap \bar{D}} [f^\bullet \dot{+} s]$$

Démonstration : Soient $K \in \mathcal{F}$ relativement compact sur D , $\delta > 0$ et $M < 0$. Par définition de $f^\bullet$, pour tout $x \in X$, il existe $C(x) \in \mathcal{C}_x$ tel que, pour tout $y \in C(x)$,

$$f(y) \leq \max(f^\bullet(x) + \delta, M)$$

Par définition de $s(x)$, quitte à réduire un peu $C(x)$, on peut supposer que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \in C(x)) \leq \max(s(x) + \delta, M)$$

Du recouvrement de $\bar{K} \cap \bar{D}$ par les $C(x)$ avec $x \in \bar{K} \cap \bar{D}$, on peut extraire un sous-recouvrement fini, noté $\{C(x_i); i \in \{1, \dots, r\}\}$. Pour tout $n \geq 1$, on a :

$$\begin{aligned} \frac{1}{n} \log \mathbb{E} \left(e^{nf(\bar{X}_n)} 1_{\bar{X}_n \in K} \right) &\leq \frac{1}{n} \log \sum_{i=1}^r \mathbb{E} \left(e^{nf(\bar{X}_n)} 1_{\bar{X}_n \in C(x_i)} \right) \\ &\leq \frac{1}{n} \log \sum_{i=1}^r e^{n \max(f^\bullet(x_i) + \delta, M)} \mathbb{P}(\bar{X}_n \in C(x_i)) \end{aligned}$$

Prenant la limite supérieure en n et utilisant le lemme 3.6.4.1, on obtient :

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left(e^{nf(\bar{X}_n)} 1_{\bar{X}_n \in K} \right) &\leq \max_{1 \leq i \leq r} \left(\max(f^\bullet(x_i) + \delta, M) + \max((s(x_i) + \delta), M) \right) \\ &\leq \sup_{x \in \bar{K} \cap \bar{D}} \left(\max(f^\bullet(x) + \delta, M) + \max((s(x) + \delta), M) \right) \end{aligned}$$

ce qui donne le résultat attendu, en faisant tendre δ vers 0 et M vers $-\infty$. $\square$

Corollaire 3.6.4.3 (Borne supérieure faible). *Pour tout $K \in \mathcal{F}$ relativement compact sur D ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in K) \leq \sup_{\overline{K \cap D}} s$$

En particulier, $(\mu_n)_{n \geq 1}$ et $(\mu_n|_D)_{n \geq 1}$ vérifient un PGD faible.

Démonstration : Il suffit d'appliquer le lemme de Varadhan compact 3.6.4.2 à $f = 0$ et $K \in \mathcal{F}$ relativement compact sur D . $\square$

Notons que, si $K \in \mathcal{F}$ est convexe relativement compact sur D , la proposition 3.6.2.4 donne la borne non asymptotique

$$\sup_{n \geq 1} \frac{1}{n} \log \mathbb{P}(\overline{X}_n \in K) \leq \sup_{\overline{K \cap D}} s$$

3.6.5 Pression et entropie

Théorème 3.6.5.1. *On suppose que μ est convexe-tendue sur D . Alors la pression est la fonction convexe-conjuguée de l'opposée de l'entropie, i.e.*

$$\forall \lambda \in X^* \quad p(\lambda) = \sup_{x \in X} (\langle \lambda | x \rangle + s(x))$$

Démonstration : L'inégalité $p \geq (-s)^*$ ne nécessite pas la convexe-tension : elle repose sur l'inégalité de Tchebychev. Il suffit de passer au supremum en $x \in X$ dans 3.6.1.2 :

$$\forall \lambda \in X^* \quad p(\lambda) \geq \sup_{x \in X} (\langle \lambda | x \rangle + s(x))$$

C'est l'autre inégalité qui requiert la convexe-tension de la loi μ de X_1 . Soit $(K_m)_{m \geq 1}$ une suite de parties mesurables de X , convexes sur D et relativement compacts sur D telle que $\mu(K_m) \rightarrow 1$. L'idée est de conditionner $\overline{X}_n$ dans K_m . On remarque tout d'abord que, pour tout $n \geq 1$,

$$\begin{aligned} \log \mathbb{E} \left(e^{\langle \lambda | X_1 \rangle} 1_{X_1 \in K_m} \right) &= \frac{1}{n} \log \mathbb{E} \left(\exp \left(n \left\langle \lambda \left| \frac{1}{n} (X_1 + \dots + X_n) \right. \right\rangle \right) 1_{X_1 \in K_m} \dots 1_{X_n \in K_m} \right) \\ &\leq \frac{1}{n} \log \mathbb{E} \left(e^{n \langle \lambda | \overline{X}_n \rangle} 1_{\overline{X}_n \in K_m} \right) \end{aligned}$$

Or, le lemme de Varadhan compact 3.6.4.2 donne

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left(e^{n \langle \lambda | \overline{X}_n \rangle} 1_{\overline{X}_n \in K_m} \right) \leq \sup_{x \in \overline{K_m}} (\langle \lambda | x \rangle + s(x))$$

D'où

$$\log \mathbb{E} \left(e^{\langle \lambda | X_1 \rangle} 1_{X_1 \in K_m} \right) \leq \sup_{x \in X} (\langle \lambda | x \rangle + s(x))$$

Enfin, en faisant tendre m vers ∞ , le lemme 3.6.5.2 ci-après donne

$$p(\lambda) \leq \sup_{x \in X} (\langle \lambda | x \rangle + s(x)) \quad \square$$

Lemme 3.6.5.2. *Soit $g : X \rightarrow [0, +\infty]$ et (A_m) une suite de parties mesurables de X telle que $\mu(A_m) \rightarrow 1$. Alors*

$$\mathbb{E}(g(X_1)1_{X_1 \in A_m}) \rightarrow \mathbb{E}(g(X_1))$$

Démonstration : Soit $M > 0$. On a

$$\mathbb{E}(g(X_1)1_{X_1 \in A_m}) \geq \mathbb{E}((g(X_1) \wedge M)1_{X_1 \in A_m}) \geq \mathbb{E}(g(X_1) \wedge M) - M(1 - \mu(A_m))$$

Faisant tendre m vers ∞ , puis M vers $+\infty$, on obtient le lemme. $\square$

3.6.6 Transformation de Fenchel-Legendre

Avant de pouvoir passer à l'égalité duale, il nous faut établir un résultat classique d'analyse convexe. De bonnes références à ce sujet sont [Mor67] et [Roc70]. On se donne ici simplement (X, τ) un espace vectoriel localement convexe. Notons X^* son dual topologique. Définissons, pour $f : X \rightarrow [-\infty, +\infty]$, sa *transformée de Fenchel-Legendre* par

$$\forall \lambda \in X^* \quad f^*(\lambda) := \sup_{x \in X} (\langle \lambda | x \rangle - f(x))$$

La transformée de Fenchel-Legendre de $g : X^* \rightarrow [-\infty, +\infty]$ se définit de façon analogue. Etendons l'addition à $[-\infty, +\infty]$ via

$$\forall a \in \mathbb{R} \quad \pm \infty + a = \pm \infty \quad \text{et} \quad -\infty + (+\infty) = -\infty$$

ainsi que la multiplication par un réel via

$$\begin{array}{ll} \forall \alpha > 0 & \alpha \cdot (\pm \infty) = \pm \infty \\ \forall \alpha < 0 & \alpha \cdot (\pm \infty) = \mp \infty \end{array} \quad \text{et} \quad 0 \cdot (\pm \infty) = 0$$

On dit que $f : X \rightarrow [-\infty, +\infty]$ est *convexe* si

$$\forall (x, y) \in X^2 \quad \forall \alpha \in [0, 1] \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

On notera que la seule fonction convexe prenant la valeur $-\infty$ est la fonction constante de valeur $-\infty$. Les fonctions convexes autres que les deux constantes $\pm \infty$ sont habituellement dénommées *fonctions convexes propres*. On dit que $f : X \rightarrow [-\infty, +\infty]$ est *concave* si $-f$ est convexe. On dit que $q : X \rightarrow [-\infty, +\infty]$ est *affine* si q est convexe et concave. Les fonctions affines à valeurs dans $[-\infty, +\infty]$ sont alors les fonctions affines habituelles à valeurs dans $] -\infty, +\infty[$ et les deux fonctions constantes $\pm \infty$. Le seul résultat qui nous intéresse ici est le suivant :

Proposition 3.6.6.1. *Soit (X, τ) un espace vectoriel localement convexe et $f : X \rightarrow [-\infty, +\infty]$. Alors*

$$f^{**} = f$$

si et seulement si f est convexe et semi-continue inférieurement relativement à la topologie faible $\sigma(X, X^)$.*

Remarque : Plus précisément, la transformation de Fenchel-Legendre réalise une bijection des fonctions convexes $\sigma(X, X^*)$ -s.c.i. sur X sur les fonctions convexes $\sigma(X^*, X)$ -s.c.i. sur X^* , de sorte que, si f est convexe et $\sigma(X, X^*)$ -s.c.i., sa transformée de Fenchel-Legendre f^* prend le nom de fonction *convexe-conjuguée* de f .

Démonstration : L'implication directe découle du fait que, pour toute fonction $g : X^* \rightarrow [-\infty, +\infty]$, la fonction g^* est convexe et $\sigma(X, X^*)$ -s.c.i. Montrons la réciproque. On remarque que, si $\lambda \in X^*$, $\langle \lambda | \cdot \rangle - f^*(\lambda)$ est la plus grande fonction affine (y compris les deux fonctions affines $\pm\infty$) dirigée par λ et inférieure à f . Il s'agit donc de voir que f est la borne supérieure de l'ensemble des fonctions affines continue plus petites que f (y compris les deux fonctions affines $\pm\infty$). Si $f = \pm\infty$, le résultat est immédiat. Sinon, f est une fonction convexe propre. Définissons l'épigraphe de f par $\text{epi}(f) = \{(x, t) \in X \times \mathbb{R}; f(x) \leq t\}$. Le fait que f soit convexe et $\sigma(X, X^*)$ -s.c.i. assure que $\text{epi}(f)$ est convexe et fermé relativement à τ . Soit $(x, t) \in X \times \mathbb{R} \setminus \text{epi}(f)$. Le théorème de Hahn-Banach dans l'espace localement convexe $X \times \mathbb{R}$ donne l'existence d'un hyperplan fermé séparant strictement (x, t) et $\text{epi}(f)$. Si cet hyperplan n'est pas vertical, il correspond à une fonction affine plus petite que f . Sinon, l'hyperplan est de la forme $H \times \mathbb{R}$ où H est un hyperplan affine fermé de X et $f(x) = +\infty$. Soit alors p une fonction affine continue sur X telle que $p(x) > 0$ et $p(y) = 0$ pour tout $y \in H$. Comme f est une fonction convexe propre, il existe une fonction affine continue et finie q inférieure à f : en effet, il existe $x \in X$ tel que $f(x) \in]-\infty, +\infty[$ et un hyperplan fermé H séparant $(x, f(x) - 1)$ de $\text{epi}(f)$; cet hyperplan H n'est pas vertical et correspond à une fonction affine continue et finie q inférieure à f . Ainsi, pour tout $\alpha > 0$, $q + \alpha p$ est toujours une fonction affine continue inférieure à f . Il suffit alors de choisir α tel que $q(x) + \alpha p(x) > t$. $\square$

On en déduit enfin :

Théorème 3.6.6.2. *Soit $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m., μ une probabilité sur $\mathcal{F}$ portée par un convexe D , et $(\mu_n)_{n \geq 1}$ la suite de Cramér associée. On suppose que μ est convexe-tendue sur D . Alors l'entropie est l'opposée de la fonction convexe-conjuguée de la pression, i.e.*

$$\forall x \in X \quad s(x) = \inf_{\lambda \in X^*} (p(\lambda) - \langle \lambda | x \rangle)$$

Démonstration : Avec les notations de la proposition 3.6.6.1, le théorème 3.6.5.1 s'écrit $p = (-s)^*$. Comme $-s$ est convexe et s.c.i. (cf. la proposition 3.6.3.1), donc $\sigma(X, X^*)$ -s.c.i., la proposition 3.6.6.1 donne

$$p^* = (-s)^{**} = -s$$

$\square$

Passons enfin à l'égalité $s = -p^*$, vraie en toute généralité dans un e.v.l.c.m., sans supposer μ convexe-tendue.

Théorème 3.6.6.3. *Pour toute probabilité μ sur un e.v.l.c.m., l'entropie de la suite de Cramér associée est l'opposée de la fonction convexe-conjuguée de la pression, i.e.*

$$\forall x \in X \quad s(x) = \inf_{\lambda \in X^*} (p(\lambda) - \langle \lambda | x \rangle)$$

Démonstration : Nous nous servons ici du théorème de Dawson-Gärtner linéaire 4.5.3.3 énoncé dans le chapitre 4. Étant donné qu'un e.v.l.c.m. est la limite projective d'un système projectif d'espaces vectoriels normés séparables, il suffit de montrer le résultat pour une probabilité μ sur un espace vectoriel normé séparable N muni de sa tribu borélienne $\mathcal{B}(N)$. Pour ce faire, les théorèmes de Banach-Mazur et de Mazur-Ulam (cf. [Ban32, chapitre XI]) montrent qu'on peut voir N comme un sous-espace (non nécessairement fermé) de $C = \mathcal{C}([0, 1]; \mathbb{R})$. Si $\mathcal{B}$ désigne la tribu borélienne de C , on a :

$$\mathcal{B}(N) = \mathcal{B}|_N := \{A \cap N; A \in \mathcal{B}\}$$

En effet, l'inclusion $N \hookrightarrow C$ est continue, donc mesurable, donc $\mathcal{B}(N) \supset \mathcal{B}|_N$. L'autre inclusion est venue du fait que $\mathcal{B}(N)$ est la tribu engendrée par les ouverts de N , autrement dit les ensembles de la forme $U \cap N$ avec U ouvert de C et que $\mathcal{B}|_N$ contient ces ensembles. On définit alors une mesure $\tilde{\mu}$ sur $\mathcal{B}$ par

$$\forall A \in \mathcal{B} \quad \tilde{\mu}(A) = \mu(A \cap N)$$

qui est une probabilité (portée par N ; cf. Suites de Cramér et probabilités portées par une partie). Or, comme C est un espace de Banach séparable, $\tilde{\mu}$ est convexe-tendue sur C . Pour éviter toute ambiguïté, notons s_μ l'entropie de la suite de Cramér associée à μ et, de façon analogue, p_μ , $s_{\tilde{\mu}}$ et $p_{\tilde{\mu}}$. Le théorème précédent montre que $s_{\tilde{\mu}} = -p_{\tilde{\mu}}^*$. Reste à appliquer le théorème de Dawson-Gärtner linéaire pour le système projectif réduit à l'inclusion $N \hookrightarrow C$ pour en déduire $s_\mu = -p_\mu^*$. $\square$

3.6.7 Borne supérieure pour les convexes

Dans toute cette section, on suppose que μ est convexe-tendue sur D . On commence par énoncer une généralisation du lemme 3.6.5.1. On étend l'addition à $[-\infty, +\infty]$ via

$$-\infty \dot{+} (+\infty) = +\infty$$

On rappelle que $f : X \rightarrow [-\infty, +\infty]$ est concave sur D si

$$\forall (x, y) \in D^2 \quad \forall \alpha \in [0, 1] \quad f(\alpha x + (1 - \alpha)y) \geq \alpha f(x) \dot{+} (1 - \alpha)f(y)$$

et qu'on note

$$\text{dom}(f) = \{x \in X; f(x) \in \mathbb{R}\}$$

Si f est concave sur D , $\text{dom}(f)$ est convexe sur D (on rappelle que, si f prend la valeur $+\infty$, alors $f = +\infty$, donc $\text{dom}(f) = \emptyset$; cf. remarques précédant la proposition 3.6.6.1 en les adaptant aux fonctions concaves).

Théorème 3.6.7.1 (Varadhan convexe). *Pour toute fonction $f : X \rightarrow [-\infty, +\infty]$ mesurable et concave sur D ,*

$$\forall n \geq 1 \quad \frac{1}{n} \log \mathbb{E} \left(e^{nf(\overline{X}_n)} \right) \leq \sup_X [f^\bullet \dot{+} s]$$

Si, de plus, $\text{dom}(f)$ est un convexe ouvert et f est semi-continue supérieurement sur $\text{dom}(f) \cap \overline{D}$, alors

$$\forall n \geq 1 \quad \frac{1}{n} \log \mathbb{E} \left(e^{nf(\overline{X}_n)} \right) \leq \sup_X [f \dot{+} s]$$

Démonstration : Soient $f : X \rightarrow [-\infty, +\infty]$ mesurable et concave sur D , et $N \geq 1$. Soit $(K_m)_{m \geq 1}$ une suite de parties mesurables de X , convexes sur D et relativement compacts sur D telle que $\mu_N(K_m) \rightarrow 1$. Pour tout $m \geq 1$, le lemme 3.6.2.3 appliqué à $f - \delta_{K_m}$ donne

$$\frac{1}{N} \log \mathbb{E} \left(e^{Nf(\overline{X}_N)} 1_{\overline{X}_N \in K_m} \right) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left(e^{nf(\overline{X}_n)} 1_{\overline{X}_n \in K_m} \right)$$

Puis, le lemme de Varadhan compact 3.6.4.2 donne

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left(e^{nf(\overline{X}_n)} 1_{\overline{X}_n \in K_m} \right) \leq \sup_{\overline{K_m \cap D}} (f^\bullet \dot{+} s)$$

D'où

$$\frac{1}{N} \log \mathbb{E} \left(e^{Nf(\overline{X}_N)} 1_{\overline{X}_N \in K_m} \right) \leq \sup_X (f^\bullet \dot{+} s)$$

En faisant tendre m vers ∞ , le lemme 3.6.5.2 donne

$$\frac{1}{N} \log \mathbb{E} \left(e^{Nf(\overline{X}_N)} \right) \leq \sup_X (f^\bullet \dot{+} s)$$

Si $C = \text{dom}(f)$ est un convexe ouvert, on affine la borne supérieure en choisissant K_m de sorte que $\overline{K_m} \subset C$ et que $\mu(K_m) \rightarrow \mu(C)$ (cf. 3.4.2.1). Dans ce cas, si f est semi-continue supérieurement sur $C \cap \overline{D}$, alors $f^\bullet = f$ sur $\overline{K_m} \cap \overline{D}$. On conclut comme précédemment avec le lemme 3.6.5.2. $\square$

Corollaire 3.6.7.2. *Soit $f : X \rightarrow [-\infty, +\infty]$ une fonction mesurable et concave sur D . Si $\text{dom}(f)$ est un convexe ouvert et si f est continue sur $\text{dom}(f) \cap \overline{D}$, alors*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left(e^{nf(\overline{X}_n)} \right) = \sup_X (f \dot{+} s)$$

Remarque : Si f est concave et s.c.i. sur $\text{dom}(f)$, alors $\text{dom}(f)$ est un convexe ouvert.

Démonstration : Si $f = +\infty$, le résultat est immédiat. Sinon, comme f est concave, f ne prend pas la valeur $+\infty$. Il suffit alors de combiner les lemmes de Varadhan ouvert 3.6.1.1 et convexe 3.6.7.1. $\square$

Corollaire 3.6.7.3 (Borne supérieure pour les convexes). *Pour tout $C \in \mathcal{F}$ convexe sur D ,*

$$\forall n \geq 1 \quad \frac{1}{n} \log \mu_n(C) \leq \sup_C s$$

Si, de plus, C est un convexe ouvert,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(C) = \sup_C s$$

Remarque : La première inégalité est l'objet de [BZ79, lemma 2.6.], mentionné sans démonstration. La version affinée est appelée « inégalité de Bernstein » dans [Bár78, théorème 2].

Démonstration : Il suffit d'appliquer le lemme de Varadhan convexe 3.6.7.1 et son corollaire à la fonction mesurable $f = -\delta_C := -\infty 1_{X \setminus C}$ qui est concave sur D . $\square$

3.6.8 Domaine de s

Dans toute cette section, μ est une probabilité sur X portée par un convexe D . On définit le support de μ par

$$\text{supp}(\mu) = \{x \in X ; \forall C \in \mathcal{C}_x \quad \mu(C) > 0\}$$

Cette définition est cohérente avec la définition habituelle, car $\mathcal{C}_x$ est une base de voisinages de x . On note $\text{cosupp}(\mu)$ l'enveloppe convexe fermée de $\text{supp}(\mu)$, *i.e.* l'intersection de tous les ensembles convexes fermés contenant $\text{supp}(\mu)$. On rappelle enfin que, pour $f : X \rightarrow [-\infty, +\infty]$, le domaine de f est défini par

$$\text{dom}(f) = \{x \in X ; f(x) \in \mathbb{R}\}$$

Si f est une fonction concave, alors $\text{dom}(f)$ est convexe (on rappelle que, si f prend la valeur $+\infty$, alors $f = +\infty$, donc $\text{dom}(f) = \emptyset$; cf. remarques précédant la proposition 3.6.6.1 en les adaptant aux fonctions concaves).

Proposition 3.6.8.1. *On a les inclusions :*

$$\text{dom}(s) \subset \overline{D} \quad \text{et} \quad \text{cosupp}(\mu) \subset \overline{D}$$

Démonstration : Soit $x \in X \setminus \overline{D}$. Il existe $C \in \mathcal{C}_x$ tel que $C \cap D = \emptyset$. Pour tout $n \geq 1$, comme μ_n est portée par D , $\mu_n(C) = 0$. Donc $s(x) = 0$ et $x \notin \text{supp}(\mu)$. Comme D est convexe, on en déduit aussi que $\text{cosupp}(\mu) \subset \overline{D}$. $\square$

Remarque : On peut affiner le résultat précédent ainsi : on note $A(D)$ l'ensemble des points $x \in X$ tels que, pour tout demi-espace fermé mesurable H contenant x , $\mu(H) > 0$. Alors,

$$\text{dom}(s) \subset A(D) \subset \overline{D}$$

La preuve est analogue à celle de [Cer07, p. 88].

On suppose désormais que μ est convexe-tendue sur D .

Théorème 3.6.8.2. *On a l'égalité*

$$\overline{\text{dom}(s)} = \text{cosupp}(\mu)$$

Ce théorème complète les résultats de [BZ79] et de [Cer07]. Sa démonstration repose sur le lemme suivant :

Lemme 3.6.8.3. *Soient (X, τ) un espace vectoriel localement convexe séparé, $\mathcal{F}$ une tribu sur X stable par translation et dilatation de rapport non nul, D une partie convexe de X et μ une mesure de probabilité sur X portée par D . Si μ est convexe-tendue sur D , alors, pour tout C convexe ouvert mesurable de X ,*

$$C \cap \text{supp}(\mu) \neq \emptyset \iff \mu(C) > 0$$

Démonstration : L'implication directe découle de la définition du support de μ (la convexe-tension n'est pas requise ici). Pour ce qui est de la réciproque, supposons $C \cap \text{supp}(\mu) = \emptyset$. Comme μ est convexe-régulière sur D , il existe une suite $(K_m)_{m \geq 1}$ de parties mesurables de X , convexes sur D et relativement compactes sur D telle que $\overline{K_m} \subset C$ et $\mu(K_m) \rightarrow \mu(C)$. Comme $\overline{K_m} \cap D \subset X \setminus \text{supp}(\mu)$, pour tout $x \in \overline{K_m} \cap D$, il existe $C(x) \in \mathcal{C}_x$ tel que $\mu(C(x)) = 0$. Après extraction d'un sous-recouvrement fini $\{C(x_i) ; i \in \{1, \dots, r\}\}$ de $\overline{K_m} \cap D$, on obtient

$$\mu(K_m) = \mathbb{P}(X_1 \in K_m) \leq \mathbb{P}(X_1 \in C(x_1) \cup \dots \cup C(x_r)) \leq \sum_{i=1}^r \mu(C(x_i)) = 0$$

d'où $\mu(C) = 0$. □

Remarque : Si $C = \text{cosupp}(\mu)$ est mesurable et d'intérieur non vide, un raisonnement analogue montre que $\mu(C) = 1$. En effet, supposant que $0 \in \overset{\circ}{C}$ et notant $C_m = (1 + 1/m)\overset{\circ}{C}$, on a $\overline{K_m} \cap (X \setminus C_m) \subset X \setminus C$ et $\mu(K_m \cap (X \setminus C_m)) \rightarrow 1$.

Démonstration du théorème : Etant donné que $\text{dom}(s)$ est convexe, il suffit de montrer les deux inclusions :

$$\text{supp}(\mu) \subset \overline{\text{dom}(s)}$$

et

$$\text{dom}(s) \subset \text{cosupp}(\mu)$$

• Soit $x \in X \setminus \overline{\text{dom}(s)}$. Il existe $C \in \mathcal{C}_x$ tel que $C \cap \text{dom}(s) = \emptyset$. La borne supérieure améliorée pour les convexes ouverts mesurables donne

$$\log \mu(C) \leq \sup_C s = -\infty$$

Donc $\mu(C) = 0$ et le lemme permet d'obtenir $C \cap \text{supp}(\mu) = \emptyset$. D'où la première inclusion.

• Soient $x \in \text{dom}(s)$ et $C \in \mathcal{C}_0$. Montrons que $(x + 2C) \cap \text{cosupp}(\mu) \neq \emptyset$. Il existe $n \geq 1$ tel que $\mu_n(x + C) > 0$. Notons alors

$$\tilde{C} = \left\{ (x_1, \dots, x_n) \in X^n ; \frac{1}{n}(x_1 + \dots + x_n) \in x + C \right\}$$

L'ensemble $\tilde{C}$ est un ouvert de X^n . Soit Q une partie dénombrable de X dense pour la topologie localement convexe $\tau(C)$ engendrée par $\{tC; t \in \mathbb{R}^*\}$. Alors, $Q^n \cap \tilde{C}$ est dense dans $\tilde{C}$ pour la topologie $\tau(C)^n$ et

$$\tilde{C} \subset \bigcup_{(u_1, \dots, u_n) \in Q^n \cap \tilde{C}} \prod_{i=1}^n (u_i + C)$$

Etant donné que l'union ci-dessus est dénombrable et que $\mu^{\otimes n}(\tilde{C}) = \mu_n(x+C) > 0$, il existe $(u_1, \dots, u_n) \in \tilde{C}$ tel que, pour tout $i \in \{1, \dots, n\}$, $\mu(u_i + C) > 0$. Le lemme précédent assure l'existence, pour tout $i \in \{1, \dots, n\}$, de $y_i \in (u_i + C) \cap \text{supp}(\mu)$. D'où

$$y = \frac{1}{n}(y_1 + \dots + y_n) \in \frac{1}{n}((u_1 + C) + \dots + (u_n + C)) \subset x + 2C$$

Comme de plus $y \in \text{cosupp}(\mu)$, on obtient le résultat voulu. $\square$

Remarque : Notons μ_n la loi de $\overline{X}_n$ et introduisons l'ensemble :

$$S = \bigcup_{n \geq 1} \text{supp}(\mu_n)$$

On montre que

$$\overline{S} = \overline{\text{dom}(s)} = \text{cosupp}(\mu)$$

• Pour voir que $\overline{S} \subset \overline{\text{dom}(s)}$, il suffit de raffiner le premier point de la preuve ci-dessus : pour $x \in X \setminus \overline{\text{dom}(s)}$, il existe $C \in \mathcal{C}_x$ tel que $C \cap \text{dom}(s) = \emptyset$. Soit $n \geq 1$. La borne supérieure améliorée pour les convexes ouverts mesurables donne

$$\frac{1}{n} \log \mu_n(C) \leq \sup_C s = -\infty$$

Donc $\mu_n(C) = 0$ et le lemme permet d'obtenir $C \cap \text{supp}(\mu_n) = \emptyset$.

• Montrons que $\text{cosupp}(\mu) \subset \overline{S}$. Soit $x \in \text{cosupp}(\mu)$ et $C \in \mathcal{C}_0$. En vertu du lemme, pour montrer que $x \in \overline{S}$, il suffit de montrer qu'il existe $n \geq 1$ tel que $\mu_n(x + C) > 0$. Soient $(x_1, \dots, x_r) \in \text{supp}(\mu)^r$ et $(\alpha_1, \dots, \alpha_r) \in [0, 1]^r$ tels que

$$\sum_{i=1}^r \alpha_i = 1 \quad \text{et} \quad \sum_{i=1}^r \alpha_i x_i = x$$

Par continuité des applications d'espace vectoriel, il existe $(U_1, \dots, U_r)$ tel que, pour tout $i \in \{1, \dots, r\}$, U_i soit un voisinage de α_i pour la topologie induite sur $[0, 1]$ et tel que

$$\sum_{i=1}^r U_i x_i \subset x + C$$

Soit maintenant, pour tout $i \in \{1, \dots, r\}$, un rationnel $r_i \in U_i$. Prenant un dénominateur commun n , on écrit $r_i = k_i/n$ et on définit

$$\tilde{x} = \sum_{i=1}^r \frac{k_i}{n} x_i \in x + C$$

Il existe alors $\varepsilon > 0$ tel que $\tilde{x} + \varepsilon C \subset x + C$ et on a :

$$\mu_n(x + C) \geq \mu_n(\tilde{x} + \varepsilon C) \geq \prod_{i=1}^r \mu(x_i + \varepsilon C) > 0$$

□

3.7 Compléments

3.7.1 Régularité de p

La fonction p est convexe (conséquence de l'inégalité de Jensen) et, dans le cas où μ est convexe-tendue, p est s.c.i. (cf. [Cer07, Lemme 12.1.]). On notera que ces propriétés ne nous servent pas au cours de la démonstration.

3.7.2 Un exemple où $p \neq (-s)^*$

On se place sur $X = \ell^\infty = L^\infty(\mathbb{N}; \mathbb{R})$. On munit X de la topologie τ associée à la norme $\|\cdot\|_\infty$. Munissons X d'une tribu. On note $\mathcal{P}^*$ l'ensemble des parties de $\mathbb{N}$ dont le complémentaire est infini, et on définit

$$S = \{1_P \in X; P \in \mathcal{P}^*\}$$

Puis

$$\pi : \alpha \in [0, 1] \longmapsto (\alpha_k)_{k \in \mathbb{N}} \in S$$

où $\overline{\alpha_0 \alpha_1 \dots}$ est l'écriture binaire de α . Cette écriture est unique si l'on suppose la suite non cofinale à 1, et c'est le cas vu la définition de S . L'application π est donc une bijection. On vérifie que

$$\mathcal{F} = \left\{ A \subset X; \forall x \in X \quad \pi^{-1}((A + x) \cap S) \in \mathcal{B}(\mathbb{R}) \right\}$$

est une tribu, invariante par translation et dilatation de rapport non nul, et contenant les boules ouvertes. De plus,

$$\{(x, y) \in X^2; x + y \in B(0, 1)\} = \bigcap_{k \in \mathbb{N}} \{(x, y) \in X^2; |x_k + y_k| < 1\} \in \mathcal{F} \otimes \mathcal{F}$$

Donc, pour toute boule ouverte B de X ,

$$\{(x, y) \in X^2; x + y \in B\} \in \mathcal{F} \otimes \mathcal{F}$$

On munit X de la tribu $\mathcal{F}$. On définit maintenant, sur $\mathcal{F}$, une mesure par

$$\forall A \in \mathcal{F} \quad \mu(A) = \text{Leb} \circ \pi^{-1}(A \cap S)$$

et on note $(\mu_n)_{n \geq 1}$ la suite de Cramér associée à μ . Bien que l'addition vectorielle ne soit pas mesurable *a priori*, les considérations précédentes permettent de définir l'entropie s associée par

$$\forall x \in X \quad s(x) = \inf_{r > 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B(x, r))$$

Montrons que $s = -\infty$. Calculons $s(0)$, les autres valeurs se calculant de la même manière. Pour cela, montrons que, pour tout $n \geq 1$,

$$\mu_n(B(0, 1)) = 0$$

Pour $n = 1$, la définition de π donne :

$$\mu\{x \in X; |x_0| < 1, \dots, |x_k| < 1\} = \text{Leb}[0, 2^{-(k+1)}[= 2^{-(k+1)}$$

donc

$$\mu(B(0, 1)) = 0$$

Pour $n = 2$, on remarque que

$$\{(x, y) \in X^2; |x_0 + y_0| < 2\} \cap (S \times S) = \left\{ (1_P, 1_Q); (P, Q) \in \mathcal{P}^* \text{ et } \begin{cases} P \subset \mathbb{N} \setminus \{0\} \\ \text{ou } Q \subset \mathbb{N} \setminus \{0\} \end{cases} \right\}$$

d'où

$$\mu_2\{(x, y) \in X^2; |x_0 + y_0| < 2\} = \frac{3}{4}$$

puis

$$\mu_2\{(x, y) \in X^2; |x_0 + y_0| < 2, \dots, |x_k + y_k| < 2\} = \left(\frac{3}{4}\right)^{k+1}$$

et

$$\mu_2(B(0, 1)) = 0$$

Le même raisonnement montre que, pour tout $n \geq 1$,

$$\mu_n(B(0, 1)) = 0$$

D'où $s(0) = -\infty$, puis $s = -\infty$. Ainsi,

$$(-s)^*(\lambda) = \sup_{x \in X} (\langle \lambda | x \rangle + s(x)) = -\infty$$

Or $p(0) = 0$, donc

$$p \neq (-s)^*$$

Toutefois, on ne peut pas prendre les transformées de Fenchel-Legendre car p n'est pas définie sur tout X^* , *a priori*⁵.

3.7.3 Autour de $\text{cosupp}(\mu)$

Soit $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m., μ une mesure de probabilité sur X , et $(\mu_n)_{n \geq 1}$ la suite de Cramér associée. Notons

$$S = \bigcup_{n \geq 1} \text{supp}(\mu_n)$$

Pour toute partie A de X , on définit

$$\text{conv}_{\mathbb{Q}}(A) = \left\{ x \in X; \exists r \geq 1 \quad \exists (x_1, \dots, x_r) \in A^r \quad x = \frac{1}{r}(x_1 + \dots + x_r) \right\}$$

⁵ On remarquera que $\text{card}(X^*) = 2^{\aleph_1}$ et que la tribu engendrée par les boules ouvertes de X est de cardinal $\aleph_1$. On doit pouvoir montrer que la tribu $\mathcal{F}$ est aussi de cardinal $\aleph_1$.

On note aussi $\text{conv}(A)$ l'enveloppe convexe de A et $\text{co}(A) = \overline{\text{conv}(A)}$ l'enveloppe convexe fermée de A . On montre que $\text{conv}_{\mathbb{Q}}(A) = \text{co}(A)$. Il est facile de voir que

$$\text{conv}_{\mathbb{Q}}(\text{supp}(\mu)) \subset S$$

Il n'y a pas d'autre inclusion, en général, entre les ensembles $\text{conv}_{\mathbb{Q}}(\text{supp}(\mu))$, $\text{conv}(\text{supp}(\mu))$ et S , comme le montrent les exemples suivants :

- Sur $X = \mathbb{R}$ muni de sa topologie usuelle et de la tribu borélienne, soit $\mu = (\delta_0 + \delta_1)/2$. Alors

$$S = \mathbb{Q} \cap [0, 1] = \text{conv}_{\mathbb{Q}}(\text{supp}(\mu))$$

de sorte que $\text{conv}(\text{supp}(\mu)) = [0, 1] \not\subset S$.

- Voici un exemple de probabilité μ pour laquelle $S \not\subset \text{conv}(\text{supp}(\mu))$. On se place sur $X = \mathbb{R}^2$ muni de sa topologie usuelle τ et de la tribu borélienne $\mathcal{F}$. On définit

$$f : t \in \mathbb{R}^* \mapsto \left(t, \frac{1}{|t|} \right)$$

et on considère μ la mesure image par f de la loi $\mathcal{N}(0, 1)$. Alors, $\text{conv}(\text{supp}(\mu))$ est le demi-espace supérieur ouvert $\{(x, y) \in \mathbb{R}^2; y > 0\}$ et $\text{supp}(\mu_2)$ contient l'axe des abscisses $\{(x, y) \in \mathbb{R}^2; y = 0\}$.

- On notera que, dans $\mathbb{R}$, l'enveloppe convexe d'un fermé est fermée (ce qui n'est plus vrai dès la dimension 2 — cf. l'exemple précédent), de sorte que, dans $\mathbb{R}$,

$$\text{conv}_{\mathbb{Q}}(\text{supp}(\mu)) \subset S \subset \text{conv}(\text{supp}(\mu))$$

Voici un exemple de probabilité μ sur $\mathbb{R}$ telle que $S \neq \text{conv}_{\mathbb{Q}}(\text{supp}(\mu))$. Il suffit de prendre une mesure μ de support $\mathbb{Z}$ (par exemple, $\mu(\{0\}) = 1/2$ et, pour tout $n \geq 1$, $\mu(\{n\}) = \mu(\{-n\}) = 1/2^{-(n+2)}$). Alors $S = \mathbb{R}$ et $\text{conv}_{\mathbb{Q}}(\text{supp}(\mu)) = \mathbb{Q}$.

3.7.4 Pour approfondir la notion de convexe interne

Les outils utilisés s'inspirent principalement de [Bou81]. On trouvera aussi une autre présentation dans [RR64]. Soient C un convexe de X et $x \in C$. On dit que x est un *point interne* de C si la jauge M_{C-x} est finie partout, autrement dit si

$$X = \bigcup_{t \in [0, +\infty[} x + t(C - x)$$

Un convexe C de X contenant 0 est *interne* si et seulement si tout point de C est interne. On vérifie que, si 0 est un point interne de C , alors

$$\{x \in X; M_C(x) < 1\}$$

est l'ensemble des points internes de C . On dit qu'un convexe C de X est *absorbant* si 0 est un point interne de C . Plus généralement que celle de convexe interne, la notion de

convexe absorbant est fondamentale dès que l'on parle de topologie localement convexe : en effet, si $\mathcal{C}_0$ est une famille de convexes absorbants de X , stable par intersection finie et par dilatation de rapport non nul, alors $\mathcal{C}_0$ est un système fondamental de voisinages de 0 pour une topologie localement convexe τ sur X (cf. [Bou81, II.25]). Si de plus les convexes de $\mathcal{C}_0$ sont internes, alors $\mathcal{C}_0$ est un système fondamental de voisinages de 0 à la fois convexes et ouverts ; et c'est cette propriété supplémentaire qui nous sert dans ce chapitre (car nous avons la limite pour les convexes ouverts mesurables). Notons encore que l'ensemble des convexes internes est l'ensemble des convexes qui sont des voisinages ouverts de 0 pour une certaine topologie localement convexe sur X . Ou plutôt : un convexe C est interne si et seulement si il existe une topologie localement convexe pour laquelle C est un voisinage ouvert de 0.

Remarque : Soit C un convexe contenant 0. Si la topologie engendrée par les translatés et dilatés de rapports non nuls de C est séparable, alors C est absorbant.

Terminons cette section par les liens algébriques entre convexes internes et convexes absorbants. Si C est un convexe de X contenant 0, on peut montrer que l'ensemble des points internes de C est ou bien $\emptyset$, ou bien un convexe interne. Un convexe est donc absorbant si et seulement si il contient un convexe interne. Dans l'autre sens, si C est un convexe de X , C est interne si et seulement si C est absorbant et, pour tout $x \in C$,

$$C = \bigcup_{t \in]0,1[} x + t(C - x)$$

On a vu qu'un ouvert convexe contenant 0 est interne. Plus généralement, pour tout convexe C de X , l'intérieur pour τ de C est ou bien $\emptyset$, ou bien l'ensemble des points internes de C .

3.7.5 Limite pour les convexes

Soient $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m., μ une probabilité sur $\mathcal{F}$ et $(\mu_n)_{n \geq 1}$ la suite de Cramér associée à μ . On s'intéresse à la question : pour quels ensembles A la suite de terme général

$$\frac{1}{n} \log \mu_n(A)$$

est-elle convergente ? La proposition 3.6.2.4 montre que, de façon générale, la suite converge dès que A est le translaté d'un convexe interne. Sans cette hypothèse, la suite peut ne pas converger, même pour un convexe : il suffit de considérer, sur $\mathbb{R}$, $\mu = (\delta_{-1} + \delta_1)/2$ et $A = \{0\}$, de sorte que

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) = -\infty$$

(valeurs des termes impairs de la suite) et

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) = 0$$

De manière générale (même en dimension infinie), si C est un convexe mesurable, la démonstration du lemme sous-additif montre qu'il existe $k_C \in \mathbb{N}^* \cup \{\infty\}$ tel que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(C) = \lim_{\substack{n \rightarrow \infty \\ n \in k_C \mathbb{N}}} \frac{1}{n} \log \mu_n(C) \in] -\infty, 0]$$

et

$$\forall n \notin k_C \mathbb{N} \quad \frac{1}{n} \log \mu_n(C) = -\infty$$

Si $k_C = 1$, la suite converge dans $] -\infty, 0]$ et, si $k_C = \infty$, la suite est constamment égale à $-\infty$. Notons $\text{cosupp}(\mu)$ l'enveloppe convexe fermée du support de μ , défini par

$$\text{supp}(\mu) = \{x \in X; \forall C \in \mathcal{C}_x \quad \mu(C) > 0\}$$

On peut adapter la démonstration de 3.6.2.4 pour montrer que la suite de terme général

$$\frac{1}{n} \log \mathbb{P}(\overline{X}_n \in C)$$

converge vers son supremum plus généralement :

- si C est un convexe mesurable dont l'ensemble des points internes rencontre $\text{cosupp}(\mu)$;
- ou si $C \in \mathcal{F}$ ne rencontre pas $\text{cosupp}(\mu)$ (auquel cas le supremum vaut $-\infty$).

On en déduit le résultat suivant :

Proposition 3.7.5.1. *Soient $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m., μ une probabilité sur $\mathcal{F}$ et $(\mu_n)_{n \geq 1}$ la suite de Cramér associée à μ . Si μ ne charge pas les hyperplans affines de X , alors, pour tout convexe $C \in \mathcal{F}$ admettant un point interne, la suite de terme général*

$$\frac{1}{n} \log \mu_n(C)$$

converge dans $[-\infty, 0]$ vers son supremum.

Remarque : Le théorème de Fubini permet de voir que, si μ ne charge pas les hyperplans, il en est de même de μ_n , pour tout $n \geq 1$.

Remarque : Si le sous-espace affine $\text{aff}(C)$ engendré par C est un sous-espace affine strict de X , comme, pour tout $n \geq 1$, μ_n ne charge pas les hyperplans, la limite existe pour C , et vaut $-\infty$. Le cas intéressant est donc seulement celui où $\text{aff}(C) = X$. Or, dans ce cas, il n'est pas automatique que C admette un point interne. C'est le cas si X est de dimension finie, mais, par exemple, dans $X = \ell^2 = L^2(\mathbb{N}; \mathbb{R})$,

$$K = \left\{ x \in \ell^2; \forall n \quad u_n \geq 0 \text{ et } \sum_{n \in \mathbb{N}} n u_n \leq 1 \right\}$$

est un compact convexe (comme enveloppe convexe fermée du compact $\{0\} \cup \{e_k/k; k \geq 1\}$, où $e_k(n) = \delta_{n,k}$, dans l'espace complet séparable ℓ^2) dont le sous-espace affine engendré

est X et qui n'admet pas de point interne. En effet, si $u \in K$ était un point interne, soit $(n_k)_{k \geq 1}$ une suite d'entiers strictement croissante telle que, pour tout $k \geq 1$,

$$\sum_{n \geq n_k} nu_n \leq 3^{-k}$$

En particulier,

$$u_{n_k} \leq \frac{1}{n_k 3^k}$$

Définissons, pour tout $n \in \mathbb{N}$,

$$w_n = \begin{cases} -\frac{1}{n_k 2^k} & \text{si } n = n_k \\ 0 & \text{sinon} \end{cases}$$

Alors $w \in \ell^2$ mais il n'existe pas $v \in K$ et $t \geq 0$ tels que $w = u + t(v - u)$. En effet, il faut certainement $t \geq 1$ et on a :

$$u_{n_k} + t(v_{n_k} - u_{n_k}) \geq (1 - t)u_{n_k} \geq \frac{1 - t}{n_k 3^k} > -\frac{1}{n_k 2^k}$$

pour k assez grand.

Démonstration : Soit $C \in \mathcal{F}$ convexe. Pour simplifier les notations, supposons que 0 est un point interne de C . Dans ce cas, l'ensemble des points internes de C est $\{M_C < 1\}$. En vertu de la remarque précédente, il reste donc à traiter le cas où

$$C \cap \text{cosupp}(\mu) \neq \emptyset \quad \text{et} \quad \{M_C < 1\} \cap \text{cosupp}(\mu) = \emptyset$$

Dans ce cas, $C \cap \text{cosupp}(\mu)$ est un convexe inclus dans $\{M_C = 1\}$: il est donc inclus dans un hyperplan affine H . En effet, $0 \notin \text{aff}(\{M_C = 1\})$, sinon il existerait $x, y \in \{M_C = 1\}$ tels que

$$0 = x + t(y - x)$$

et on en déduirait que

$$M_C(y) \leq 1 - \frac{1}{t} < 1$$

Comme, pour tout $n \geq 1$, $\mu_n(H) = 0$ (notons que les hyperplans sont mesurables, car X est un e.v.l.c.m.), on en déduit que $s(C) = -\infty$ et on a la limite souhaitée. $\square$

En dimension finie, on peut affiner le résultat. On se donne désormais un entier positif d et on se place dans $X = \mathbb{R}^d$, muni de la topologie standard et de la tribu borélienne. On se donne une probabilité μ sur X et on note $(\mu_n)_{n \geq 1}$ la suite de Cramér associée⁶.

⁶ Il existe ici une suite i.i.d. de loi μ : on pourrait donc se passer de la notion de suite de Cramér.

Proposition 3.7.5.2. Soient μ une mesure de probabilité sur un espace de dimension finie et $(\mu_n)_{n \geq 1}$ la suite de Cramér associée. Si, pour tout couple d'hyperplans parallèles et distincts (F, G) ,

$$\mu(F)\mu(G) = 0$$

alors, pour tout convexe mesurable C , la suite

$$\frac{1}{n} \log \mu_n(C)$$

converge dans $[-\infty, 0]$ vers son supremum.

Remarque : Même en dimension finie, un convexe n'est pas automatiquement borélien (pour peu que sa frontière ne le soit pas). En revanche, comme sa frontière est un ensemble négligeable pour la mesure de Lebesgue, tout convexe est un lebesguien. Comme nous considérons ici la tribu borélienne, nous précisons que C est mesurable.

Démonstration : On montre le résultat plus généralement pour une mesure μ de masse inférieure à 1. La démonstration se fait par récurrence sur la dimension d de l'espace affine ambiant. Le résultat est immédiat pour $d = 0$. Supposons le résultat vrai pour tout entier strictement inférieur à d et soit μ une mesure finie sur $\mathbb{R}^d$ vérifiant la condition de l'énoncé. Soit C un convexe mesurable de $\mathbb{R}^d$. Remarquons que, pour tout $n \geq 1$, comme μ_n est régulière,

$$\mu_n(C) = \mu_n(C \cap \text{cosupp}(\mu))$$

On distingue trois cas.

- Si $C \cap \text{cosupp}(\mu) = \emptyset$, alors

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(C) = -\infty$$

- Si $\overset{\circ}{C} \cap \text{cosupp}(\mu) \neq \emptyset$, le lemme sous-additif pour les convexes internes montre que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(C) = \sup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(C) \in \mathbb{R}$$

On notera que, μ étant de masse inférieure à 1, la suite $u(n) = -\log \mu_n(C)$ est bien positive.

- Reste le cas où $C \cap \text{cosupp}(\mu) \neq \emptyset$ et $\overset{\circ}{C} \cap \text{cosupp}(\mu) = \emptyset$. Dans ce cas,

$$\overbrace{C \cap \text{cosupp}(\mu)}^{\circ} = \overset{\circ}{C} \cap \overbrace{\text{cosupp}(\mu)}^{\circ} = \emptyset$$

Ainsi, $\tilde{C} = C \cap \text{cosupp}(\mu)$, convexe non vide d'intérieur vide, engendre un sous-espace affine strict $\tilde{X}$ de X . Notons $\tilde{\mu} = \mu|_{\tilde{X}}$ et $(\tilde{\mu}_n)_{n \geq 1}$ la suite de Cramér associée. La condition sur μ assure que, pour tout $n \geq 1$,

$$\mu_n(\tilde{X}) = \tilde{\mu}_n(\tilde{X})$$

(le reste de μ_n « ne voit pas » $\tilde{X}$, qui est contenu dans un hyperplan). De plus, $\tilde{\mu}$ vérifie l'hypothèse de l'énoncé. On conclut en appliquant l'hypothèse de récurrence à $\tilde{X}$, $\tilde{\mu}$ et $\tilde{C}$. $\square$

La réciproque n'est pas vraie : dans $\mathbb{R}$, il suffit de considérer

$$\mu = \sum_{n=1}^{\infty} 2^{-n} \delta_n$$

Dans ce cas, pour tout $a \in \mathbb{Q} \cap [1, +\infty[$, $k_{\{a\}} = 1$ et, pour tout $a \in \mathbb{R} \setminus (\mathbb{Q} \cap [1, +\infty[)$, $k_{\{a\}} = \infty$. Le cas des convexes I d'intérieur non vide se traite en distinguant les cas $I \subset]-\infty, 0[$ ($k_I = \infty$) et $I \cap [0, +\infty[\neq \emptyset$ ($k_I = 1$).

3.8 Suppléments techniques

3.8.1 De l'utilité des hypothèses

La structure d'espace vectoriel localement convexe est naturelle en théorie de Cramér. Ce qui fait fonctionner la sous-additivité, c'est la structure de système fondamental de voisinages convexes (non nécessairement absorbants, *a priori*). La topologie d'espace vectoriel localement convexe (donc les ouverts convexes absorbants) est imposée d'abord car on a, *a priori*, la limite uniquement pour C translaté de convexe interne. D'autre part, la remarque de la page 73 montre que la séparabilité impose aux convexes d'être absorbants. La question de l'utilité de la séparabilité pour la mesurabilité est une question plus délicate que nous n'avons pas plus abordé que la remarque faite en introduction.

3.8.2 Quelques extensions du cadre

Soient X un espace vectoriel, $\mathcal{F}$ une tribu sur X stable par translation et dilatation de rapport non nul, et rendant continue l'addition

$$(x, y) \in (X \times X, \mathcal{F} \otimes \mathcal{F}) \longmapsto x + y \in (X, \mathcal{F})$$

Remarquons que, comme nous l'avons fait dans le premier complément, il pourrait même suffire que, pour certains ensembles $A \in \mathcal{F}$ (par exemple des convexes),

$$\{(x, y) \in X \times X ; x + y \in A\} \in \mathcal{F} \otimes \mathcal{F}$$

Désignons par $\mathcal{C}_0(\mathcal{F})$ l'ensemble de tous les convexes internes mesurables de X contenant 0 (si $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est un e.v.l.c.m., $\mathcal{C}_0 \subset \mathcal{C}_0(\mathcal{F})$, mais il n'y a pas égalité en général). Soit $\mathcal{C}_0$ un sous-ensemble de $\mathcal{C}_0(\mathcal{F})$ et τ la topologie d'espace vectoriel localement convexe engendrée par $\mathcal{C}_0$. Alors, pour toute probabilité μ sur $\mathcal{F}$, la suite de Cramér associée $(\mu_n)_{n \geq 1}$ vérifie un PGD faible relativement à τ . Autrement dit, il y a plus de topologies que celles que nous avons considérées relativement auxquelles on peut énoncer un PGD faible. Si de plus μ est convexe-tendue, alors, pour toute forme linéaire $\lambda \in X^*$ mesurable, on a

$$p(\lambda) = \sup_{x \in X} (\langle \lambda | x \rangle + s(x))$$

3.8.3 Questions de séparation

Nous avons supposé l'espace topologique (X, τ) séparé pour utiliser la notion habituelle de compacité. Toutefois nous avons seulement recours à l'axiome de Borel-Lebesgue vérifié par les compacts, et non à la séparation (cf. lemme 3.6.4.2). On dit qu'une partie Q de X est *quasi-compacte* si de tout recouvrement ouvert de Q on peut extraire un sous-recouvrement fini. On dit qu'une partie de X est *relativement quasi-compacte* si elle contenue dans une partie quasi-compacte de X . Il est alors très simple d'adapter l'ensemble du texte sans supposer l'espace (X, τ) séparé : il suffit de remplacer chaque occurrence du mot « compact »

par « quasi-compact » (dans la borne supérieure faible, dans la définition de la convexe-tension, etc.). Voici ce que devient le lemme de Varadhan convexe. On montre en réalité un résultat plus fort : pour tout $K \in \mathcal{F}$ relativement quasi-compact,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \int_K e^{nf(x)} d\mu_n(x) \leq \sup_{Q(K)} [f + s]$$

où $Q(K)$ est l'intersection de tous les quasi-compacts contenant K (cet ensemble n'est pas nécessairement quasi-compact lui-même). Sachant que l'intersection d'un quasi-compact et d'un fermé est encore quasi-compact, on a $Q(K) \subset \overline{K}$: d'où le résultat et, si X n'est pas séparé, la borne supérieure obtenue peut être bien meilleure.

3.8.4 Tension et convexe-tension

Soient X_1 (resp. X_2) un espace vectoriel réel, $\mathcal{F}_1$ (resp. $\mathcal{F}_2$) une tribu sur X_1 (resp. X_2), et τ_1 (resp. τ_2) une topologie séparée sur X_1 (resp. X_2). On munit $X_1 \times X_2$ de $\mathcal{F}_1 \otimes \mathcal{F}_2$ et de $\tau_1 \times \tau_2$ et on se donne μ une probabilité sur $X_1 \times X_2$. Si la marginale μ_1 (resp. μ_2) est convexe-tendue sur D_1 (resp. D_2), alors μ est convexe-tendue sur $D_1 \times D_2$. En effet, si $(K_{m,1})_{m \geq 1}$ (resp. $(K_{m,2})_{m \geq 1}$) est une suite de parties mesurables de X_1 (resp. X_2), convexes sur D_1 (resp. D_2) et relativement compacts sur D_1 (resp. D_2) telle que $\mu_1(K_{m,1}) \rightarrow 1$ (resp. $\mu_2(K_{m,2}) \rightarrow 1$), alors, pour tout $m \geq 1$, $K_{m,1} \times K_{m,2}$ est mesurable, convexe sur $D_1 \times D_2$ et relativement compact sur $D_1 \times D_2$, et on a :

$$\mu(X_1 \times X_2 \setminus K_{m,1} \times K_{m,2}) \leq \mu_1(X_1 \setminus K_{m,1}) + \mu_2(X_2 \setminus K_{m,2}) \rightarrow 0$$

Remarque : Soient $(X, \tilde{\tau})$ un e.v.l.c. quasi-complet et τ une topologie d'e.v.l.c. compatible avec la dualité entre X et X^* (en fait, τ est aussi quasi-complète ; cf. [Bou81, IV.5]) ; on suppose que $\mathcal{F}$ est la tribu borélienne de τ . Alors, sur $(X, \mathcal{F}, \tau)$, la convexe-tension équivaut à la *tension* : on dit qu'une probabilité μ est tendue s'il existe une suite $(K_m)_{m \geq 1}$ de compacts telle que

$$\lim_{m \rightarrow \infty} \mu(K_m) = 1$$

Le résultat est conséquence du théorème de Krein (cf. [Bou81, IV.37, théorème 3]). Le choix de $\mathcal{F}$ borélienne évite le problème de mesurabilité de l'enveloppe convexe.

4

Chapitre 4

PRESSION, ENTROPIE ET TRANSPORT LINÉAIRE

Sommaire

4.1	Introduction	82
4.2	Pression	82
4.2.1	Cadre et définitions	82
4.2.2	Convexité de la pression	82
4.3	Entropie et principes de grandes déviations	83
4.3.1	Cadre et définitions	83
4.3.2	Bornes inférieures	85
4.3.3	Borne supérieure faible	87
4.3.4	Condition suffisante de PGD faible	88
4.3.5	Condition nécessaire et suffisante de PGD faible	89
4.4	Espaces topologiques localement mesurables	92
4.4.1	Définition	92
4.4.2	Limites projectives d'espaces topologiques localement mesurables	92
4.5	Espaces vectoriels topologiques mesurables	95
4.5.1	Transformation de Fenchel-Legendre	95
4.5.2	Espaces vectoriels topologiques mesurables	96
4.5.3	Limites projectives d'espaces vectoriels topologiques mesurables	96
4.6	Les trois fondements de la théorie de Cramér	99
4.6.1	Lemme sous-additif	99
4.6.2	Intervention infimum-supremum	100
4.6.3	Transformation de Fenchel-Legendre	101
4.7	Suppléments sur le PGD	103
4.7.1	Lemme de Varadhan fermé	103

4.7.2	Conditions de PGD	103
4.7.3	Identification de $\bar{s}_\#$ et de $\bar{s}_b$	104
4.7.4	Tension exponentielle	104
4.7.5	Condition suffisante de tension exponentielle	105

4.1 Introduction

L'objectif de ce texte est de donner les définitions les plus naturelles de la pression et de l'entropie, et de faire apparaître le lien entre les deux fonctions, autrement dit les fondements de l'équivalence d'ensembles. Le but n'est pas ici de montrer l'équivalence d'ensembles, ce qui ne peut se faire que moyennant des hypothèses sur la suite $(\mu_n)_{n \geq 1}$ des mesures considérées (par exemple, loi de $\bar{X}_n$ pour des variables i.i.d. au chapitre 3, pour des variables asymptotiquement découplées au chapitre 5). Le présent chapitre ébauche un cadre dans lequel les questions des chapitres 3 et 5 se posent naturellement. Ce cadre peut être vu comme l'analogie du cadre de Varadhan pour les PGD, transposé à la théorie de Cramér, et en particulier à la question $s = -p^*$. On aboutit à une version linéaire du théorème de Dawson-Gärtner assurant le transport du PGD faible et de l'égalité $s = -p^*$ par limite projective. On notera qu'en revanche (BS_c) ne passe pas aux limites projectives, de même que l'égalité duale $p = (-s)^*$. Ce texte doit beaucoup à [LP95] pour les notations utilisées et pour plusieurs résultats montrés au passage (comme la caractérisation des PGD faibles).

4.2 Pression

4.2.1 Cadre et définitions

Soient X un ensemble et $\mathcal{F}$ une tribu sur X . Soient $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité sur $\mathcal{F}$ et $(v_n)_{n \geq 1}$ une suite de réels strictement positifs. On note $\mathcal{M}$ l'ensemble des fonctions mesurables de X dans $[-\infty, +\infty]$ et on définit, pour tout $\varphi \in \mathcal{M}$,

$$\underline{p}(\varphi) := \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{v_n \varphi} d\mu_n \quad \text{et} \quad \bar{p}(\varphi) := \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{v_n \varphi} d\mu_n$$

On appelle *pression* de la suite $(\mu_n, v_n)_{n \geq 1}$ la fonction $p = \bar{p}$.

4.2.2 Convexité de la pression

On étend l'addition à $[-\infty, +\infty]$ via

$$\forall a \in \mathbb{R} \quad a + (\pm\infty) = \pm\infty \quad \text{et} \quad (-\infty) + (+\infty) = -\infty$$

(on gardera, sauf pour $(-\infty) + (+\infty)$, la notation $+$) ainsi que la multiplication par un réel via

$$\begin{aligned} \forall \alpha > 0 \quad \alpha \cdot (\pm\infty) &= \pm\infty \\ \forall \alpha < 0 \quad \alpha \cdot (\pm\infty) &= \mp\infty \end{aligned} \quad \text{et} \quad 0 \cdot (\pm\infty) = 0$$

On munit $\mathcal{M}$ de l'addition et de la multiplication par un réel terme à terme. On dit qu'une fonction $g : \mathcal{M} \rightarrow [-\infty, +\infty]$ est *convexe* si

$$\forall (\varphi, \psi) \in X^2 \quad \forall \alpha \in [0, 1] \quad g(\alpha\varphi + (1 - \alpha)\psi) \leq \alpha g(\varphi) + (1 - \alpha)g(\psi)$$

Proposition 4.2.2.1. *La pression est une fonction convexe.*

Démonstration : Pour tout $n \geq 1$, l'inégalité de Hölder permet de montrer que

$$\varphi \in \mathcal{M} \mapsto \frac{1}{v_n} \log \int e^{v_n \varphi} d\mu_n \in [-\infty, +\infty]$$

est convexe. On conclut en remarquant qu'une limite supérieure de fonctions convexes est convexe. $\square$

4.3 Entropie et principes de grandes déviations

4.3.1 Cadre et définitions

Soient X un ensemble, $\mathcal{F}$ une tribu sur X et τ une topologie séparée¹ sur X . Soient $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité sur $\mathcal{F}$ et $(v_n)_{n \geq 1}$ une suite de réels strictement positifs. On définit, pour tout $x \in X$,

$$\underline{s}(x) := \inf_{\substack{A \in \mathcal{F} \\ A \ni x}} \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) \quad \text{et} \quad \overline{s}(x) := \inf_{\substack{A \in \mathcal{F} \\ A \ni x}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A)$$

Bien entendu, $\underline{s} \leq \overline{s}$. On appelle² *entropie* de la suite $(\mu_n, v_n)_{n \geq 1}$, relativement à τ , la fonction $s = \underline{s}$. Par construction, l'entropie est la plus grande fonction vérifiant la borne inférieure :

(BI) pour tout $A \in \mathcal{F}$,

$$\liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) \geq \sup_{\overset{\circ}{A}} s$$

On dit que $(\mu_n, v_n)_{n \geq 1}$ vérifie un *principe de grandes déviations* (PGD) si la borne supérieure suivante est vérifiée :

¹ Toutes les topologies seront supposées séparées, on ne le précisera plus. La raison est d'ordre culturel et pratique... Notons toutefois que cette hypothèse est inutile : il suffit de remplacer chaque occurrence de « compact » par « quasi-compact ».

² On dit, d'habitude, l'entropie de la suite $(\mu_n)_{n \geq 1}$, de vitesse $(v_n)_{n \geq 1}$, relativement à τ . C'est par souci de concision que nous l'écrivons ainsi. Notons aussi que [LP95] réserve le terme d'*entropie* au cas où $\underline{s} = \overline{s}$ est concave.

(BS) pour tout $A \in \mathcal{F}$,

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) \leq \sup_{\bar{A}} s$$

De manière générale, (BS) n'est pas vérifiée pour tous les mesurables. Si $\mathcal{P}$ désigne un ensemble de parties mesurables de X , on définit la version restreinte de la borne supérieure suivante :

(BS $_{\mathcal{P}}$) pour tout $A \in \mathcal{P}$,

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) \leq \sup_{\bar{A}} s$$

En particulier, si D est une partie de X , on notera :

(BS $_{b,D}$) pour tout $K \in \mathcal{F}$ tel que $K \cap D$ soit relativement compact,

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(K) \leq \sup_{\bar{K}} s$$

Si $(\mu_n, v_n)_{n \geq 1}$ vérifie (BS $_{b,X}$), on dit que $(\mu_n, v_n)_{n \geq 1}$ vérifie un *principe de grandes déviations faible* (PGD faible). Notons que, si $D_1 \subset D_2$, alors (BS $_{b,D_1}$) entraîne (BS $_{b,D_2}$). De plus, si D est une partie de X et ν une probabilité sur $\mathcal{F}$, on dira que ν est portée par D si

$$\forall (A, B) \in \mathcal{F}^2 \quad A \cap D = B \cap D \Rightarrow \nu(A) = \nu(B)$$

Définissons la tribu trace $\mathcal{F}|_D = \{A \cap D; A \in \mathcal{F}\}$. Si ν est portée par D , alors

$$\nu|_D(A \cap D) = \nu(A)$$

définit une mesure de probabilité sur $\mathcal{F}|_D$. En particulier, on voit ainsi que ν est la loi d'une variable aléatoire à valeurs dans D : il suffit de considérer l'inclusion

$$(D, \mathcal{F}|_D, \mu|_D) \hookrightarrow (X, \mathcal{F})$$

Soit D une partie de X telle que, pour tout $n \geq 1$, μ_n soit portée par D . Si $(\mu_n, v_n)_{n \geq 1}$ vérifie (BS $_{b,D}$), alors, sur l'espace D muni de la tribu $\mathcal{F}|_D$ et de la topologie trace $\tau|_D = \{U \cap D; U \in \tau\}$, $(\mu_n|_D, v_n)_{n \geq 1}$ vérifie (BS $_{b,D}$) (sachant que, pour tout $x \in D$, $s|_D(x) = s(x)$).

Remarque : Nous venons de voir que, si, pour tout $n \geq 1$, μ_n est portée par D , (BS $_{b,D}$) sur X est un résultat plus fort que (BS $_{b,X}$) sur X et que (BS $_{b,D}$) sur D . De manière générale, il n'y a pas de réciproque. Par exemple :

- sur $X = \mathbb{R}$, muni de la topologie standard τ et de la tribu $\mathcal{F} = \{\emptyset,]-\infty, 0],]0, +\infty[, \mathbb{R}\}$, soient, pour tout $n \geq 1$, $\mu_n = \delta_{(-1)^n}$ et $v_n = n$. Alors, $\underline{s} = -\infty$. Pourtant, comme tous les ensembles mesurables non vides sont non bornés, donc non relativement compacts, le PGD faible est trivialement vrai, autrement dit on a (BS $_{b,X}$) sur X . En revanche, si $D = \{-1, 1\}$, on n'a ni (BS $_{b,D}$) sur X , ni (BS $_{b,D}$) sur D .

- sur $X = \mathbb{R}$, muni de la topologie standard τ et de la tribu $\mathcal{F} = \sigma(\{]-\infty, -2],]-2, 0],]0, 2],]2, +\infty[\})$, soient, pour tout $n \geq 1$, $\mu_n = \delta_{(-1)^n}$, $v_n = n$ et $D =]-1, 1[$. Alors,

$\underline{s} = -\infty$. Cette fois, $(BS_{b,D})$ sur D est trivialement vérifiée : les adhérences, dans D , des ensembles mesurables, à savoir $] -1, 0]$, $]0, 1[$ et $] -1, 1[$, ne sont pas compactes. En revanche, on n'a ni $(BS_{b,D})$ sur X , ni $(BS_{b,X})$ sur X .

Les deux exemples peuvent paraître pathologiques. Il n'en est rien. Autre exemple pour le premier point : sur $X = \mathbb{R}^{\mathbb{R}}$, muni de la tribu cylindrique $\mathcal{F}$ et de la topologie faible τ , soient $x \in X \setminus \{0\}$ et, pour tout $n \geq 1$, $\mu_n = \delta_{(-1)^n x}$. Alors, $\underline{s} = -\infty$. Pourtant, comme tous les ensembles mesurables non vides sont non bornés, donc non relativement compacts, le PGD faible est trivialement vrai, autrement dit on a $(BS_{b,X})$ sur X . En revanche, si $D = \{-x, x\}$, on n'a ni $(BS_{b,D})$ sur X , ni $(BS_{b,D})$ sur D . Toutefois :

- si, pour tout $n \geq 1$, μ_n est portée par D et si les compacts de X sont mesurables (en particulier si $\mathcal{F}$ est la tribu borélienne), alors

$$(BS_{b,D}) \text{ sur } X \iff (BS_{b,X}) \text{ sur } X$$

- si, pour tout $n \geq 1$, μ_n est portée par D et si D est fermé, alors

$$(BS_{b,D}) \text{ sur } X \iff (BS_{b,D}) \text{ sur } D$$

Dans le cas de Sanov, où $X = \mathcal{M}_b(E)$ est muni de la tribu cylindrique et de la topologie produit, et $D = \mathcal{M}_1^+(E)$, aucune des deux conditions n'est vérifiée.

On peut aussi définir

$(BS_{c,D})$ pour tout $C \in \mathcal{F}$ tel que $C \cap D$ soit convexe,

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(C) \leq \sup_{\overline{C}} s$$

Dans le cas i.i.d., il existe des contre-exemples à (BS) déjà dans $X = \mathbb{R}^3$. En revanche, on montre que $(BS_{c,X})$ est vérifiée dès que μ_1 est convexe-tendue (en particulier, si X est un espace de Banach séparable).

4.3.2 Bornes inférieures

Si φ est une fonction de X dans $[-\infty, +\infty]$, on dit que φ est *semi-continue inférieurement* si, pour tout $t \in [-\infty, +\infty]$, l'ensemble

$$\{x \in X ; \varphi(x) > t\}$$

est un ouvert de X . Pour toute fonction $\varphi : X \rightarrow [-\infty, +\infty]$, on note $\varphi_\bullet$ sa régularisée semi-continue inférieurement, autrement dit la plus grande fonction semi-continue inférieurement qui soit inférieure à φ .

Théorème 4.3.2.1. *Pour tout $x \in X$ et pour toute fonction $\varphi : X \rightarrow [-\infty, +\infty]$ mesurable, on a :*

$$\sup_{x \in X} (\varphi_\bullet(x) + \underline{s}(x)) \leq \underline{p}(\varphi) \quad \text{et} \quad \sup_{x \in X} (\varphi_\bullet(x) + \overline{s}(x)) \leq \overline{p}(\varphi)$$

Démonstration : Démontrons la première inégalité, la deuxième se montrant de façon similaire. Soient $\varphi \in \mathcal{M}$, $x \in X$, $\delta > 0$ et $M > 0$. L'ensemble

$$V = \{y \in X; \varphi(y) \geq \min(\varphi_\bullet(x) - \delta, M)\}$$

est mesurable et contient $\{y \in X; \varphi_\bullet(y) \geq \min(\varphi_\bullet(x) - \delta, M)\}$, donc est un voisinage de x . On a alors, pour tout $n \geq 1$:

$$\begin{aligned} \frac{1}{v_n} \log \int e^{v_n \varphi} d\mu_n &\geq \frac{1}{v_n} \log \int_V e^{v_n \varphi} d\mu_n \\ &\geq \min(\varphi_\bullet(x) - \delta, M) + \frac{1}{v_n} \log \mu_n(V) \end{aligned}$$

Prenant la limite inférieure en n , il vient

$$\liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{v_n \varphi} d\mu_n \geq \min(\varphi_\bullet(x) - \delta, M) + \underline{s}(x)$$

On conclut en faisant tendre δ vers 0 et M vers $+\infty$, puis en prenant le supremum en $x \in X$. $\square$

Ce résultat est ce qu'on a appelé lemme de Varadhan ouvert dans le cas i.i.d. : il est vrai en toute généralité. On peut aussi le montrer avec l'inégalité de Tchebychev dans le cas $\varphi = \lambda$ forme linéaire continue mesurable. Remarquons que la définition des fonctions $\underline{s}$ et $\overline{s}$ peut se récrire

$$\underline{s}(x) = \inf_{A \in \mathcal{F}} \left(\underline{p}(\delta_A) \dot{+} (-(\delta_A)_\bullet(x)) \right) \quad \text{et} \quad \overline{s}(x) = \inf_{A \in \mathcal{F}} \left(\overline{p}(\delta_A) \dot{+} (-(\delta_A)_\bullet(x)) \right)$$

où $(-\infty) \dot{+} (+\infty) = +\infty$ et, pour toute partie A de X , on note

$$\delta_A := -\infty \cdot 1_{X \setminus A}$$

Corollaire 4.3.2.2. *Pour tout $x \in X$,*

$$\underline{s}(x) = \inf_{\varphi \in \mathcal{M}} \left(\underline{p}(\varphi) \dot{+} (-\varphi_\bullet(x)) \right) \quad \text{et} \quad \overline{s}(x) = \inf_{\varphi \in \mathcal{M}} \left(\overline{p}(\varphi) \dot{+} (-\varphi_\bullet(x)) \right)$$

Ecrites sous cette forme, les bornes inférieures font penser à la fameuse égalité

$$s(x) = \inf_{\lambda \in X^*} (p(\lambda) - \lambda(x)) = -p^*(x)$$

qu'on a quand X est un espace vectoriel et sous certaines conditions. On notera toutefois qu'on n'a pas de lien intéressant entre entropie $s = \underline{s}$ et pression $p = \overline{p}$, pour le moment. Cela pourra se faire si $\underline{s} = \overline{s}$ ou $\underline{p} = \overline{p}$. Remarquons que, dans le cadre du chapitre 5, $\underline{s} = \overline{s}$ et $\underline{p}(\lambda) = \overline{p}(\lambda)$ pour λ forme linéaire continue et mesurable. D'autre part, la concavité de s est une condition nécessaire pour obtenir l'égalité $s = -p^*$. Si s n'est pas concave, le dernier corollaire suggère d'étendre l'ensemble X^* à d'autres fonctions, par exemple des formes quadratiques (ce qui donne l'ensemble gaussien ; cf. les travaux de Costeniuc, Ellis, Haven, Touchette et Turkington, par exemple [EHT00] et [CETT05]).

4.3.3 Borne supérieure faible

Si φ est une fonction de X dans $[-\infty, +\infty]$, on dit que φ est *semi-continue supérieurement* si, pour tout $t \in [-\infty, +\infty]$, l'ensemble

$$\{x \in X; \varphi(x) < t\}$$

est un ouvert de X . Pour toute fonction $\varphi : X \rightarrow [-\infty, +\infty]$, on note $\varphi^\bullet$ sa régularisée semi-continue supérieurement, autrement dit la plus petite fonction semi-continue supérieurement qui soit supérieure à φ .

Théorème 4.3.3.1. *Soit D une partie de X telle que, pour tout $n \geq 1$, μ_n soit portée par D . Alors, pour tout $x \in X$ et pour toute fonction $\varphi : X \rightarrow [-\infty, +\infty]$ mesurable telle que $K = \{\varphi > -\infty\} \cap D$ soit relativement compact, on a :*

$$\bar{p}(\varphi) \leq \sup_{x \in \bar{K}} (\varphi^\bullet(x) + \bar{s}(x))$$

Démonstration : Soient $\varphi \in \mathcal{M}$, $\delta > 0$ et $M > 0$. On suppose que $K = \{\varphi > -\infty\} \cap D$ est relativement compact. Par définition de $\varphi^\bullet$, pour tout $x \in \bar{K}$, l'ensemble

$$U(x) = \{y \in X; \varphi(y) \leq \max(\varphi^\bullet(x) + \delta, M)\}$$

est un voisinage mesurable de x . Par définition de $\bar{s}(x)$, il existe un voisinage $A(x) \in \mathcal{F}$ de x tel que

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A(x)) \leq \max(\bar{s}(x) + \delta, M)$$

Notons $V(x) = U(x) \cap A(x)$. Du recouvrement de $\bar{K}$ par les $V(x)$ avec $x \in \bar{K}$, on peut extraire un sous-recouvrement fini, noté $\{V(x_i); i \in \{1, \dots, r\}\}$. Pour tout $n \geq 1$, le fait que μ_n soit la loi d'une variable aléatoire à valeurs dans D , puis la définition de $V(x)$ donnent

$$\begin{aligned} \frac{1}{v_n} \log \int e^{v_n \varphi} d\mu_n &\leq \frac{1}{v_n} \log \sum_{i=1}^r \int_{V(x_i)} e^{v_n \varphi} d\mu_n \\ &\leq \frac{1}{v_n} \log \sum_{i=1}^r e^{v_n \max(\varphi^\bullet(x_i) + \delta, M)} \mu_n(V(x_i)) \end{aligned}$$

Prenant la limite supérieure en n et utilisant le lemme 4.6.2.1, on obtient :

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{v_n \varphi} d\mu_n &\leq \max_{1 \leq i \leq r} \left(\max(\varphi^\bullet(x_i) + \delta, M) + \max((\bar{s}(x_i) + \delta), M) \right) \\ &\leq \sup_{x \in \bar{K}} \left(\max(\varphi^\bullet(x) + \delta, M) + \max((\bar{s}(x) + \delta), M) \right) \end{aligned}$$

ce qui donne le résultat attendu, en faisant tendre δ vers 0 et M vers $+\infty$. $\square$

4.3.4 Condition suffisante de PGD faible

Combinant les bornes inférieure et supérieure faible, on obtient une version du lemme de Varadhan :

Théorème 4.3.4.1 (Lemme de Varadhan). *Supposons $\underline{s} = \overline{s}$. Alors, pour toute fonction $\varphi : X \rightarrow]-\infty, +\infty[$ mesurable et continue, en notant*

$$d\tilde{\mu}_n = e^{n\varphi} d\mu_n$$

la suite $(\tilde{\mu}_n, v_n)_{n \geq 1}$ vérifie un PGD faible, l'entropie associée étant $\varphi + s$.

Remarque : Nous n'avons défini la notion de PGD faible que pour des suites de mesures de probabilité. Etendre la définition à une suite de mesures quelconques ne pose pas de problème.

En particulier, on obtient des conditions suffisantes pour qu'une suite vérifie un PGD faible. Pour tout $x \in X$, on note $\mathcal{V}_x$ un système fondamental de voisinages mesurables de x . On définit les propriétés (la notation est pour « sous-additivité » et « contrôle » du lemme 4.6.1.2) :

(SAC) Pour tout $x \in X$ et pour tout $A \in \mathcal{V}_x$,

$$\liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) = \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A)$$

(SAC_b) Pour tout $x \in X$ et pour tout $B \in \mathcal{V}_x$, il existe $A \in \mathcal{V}_x$ tel que

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) \leq \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(B)$$

Proposition 4.3.4.2. *Soit D une partie de X telle que, pour tout $n \geq 1$, μ_n soit portée par D . On a les implications (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (6) où :*

- (1) (SAC) ;
- (2) (SAC_b) ;
- (3) $\underline{s} = \overline{s}$;
- (4) pour toute fonction $\varphi \in \mathcal{M}$ telle que $\{\varphi > -\infty\} \cap D$ soit relativement compact,

$$p(\varphi) \leq \sup_{x \in X} (\varphi^\bullet(x) + s(x))$$

- (5) $(\mu_n, v_n)_{n \geq 1}$ vérifie (BS_{b,D}) ;
- (6) $(\mu_n, v_n)_{n \geq 1}$ vérifie un PGD faible.

Démonstration : La première implication est immédiate. Pour la deuxième implication, il suffit, pour tout $x \in X$, de passer, dans (SAC_b), à l'infimum en A voisinage mesurable de x , puis en B voisinage mesurable de x . La troisième implication est conséquence immédiate du théorème 4.3.3.1. Pour la quatrième implication, il suffit de prendre $\varphi = \delta_K$ pour $K \in \mathcal{F}$

tel que $K \cap D$ soit relativement compact. Enfin, la dernière implication a été vue dans la section Cadre et définition de cette partie. $\square$

C'est sur cette proposition que repose la théorie de Cramér pour les PGD faibles. Le lemme sous-additif du cas a.d.i. (chapitre 5) permet de montrer (SAC_b) avec les voisinages convexes mesurables. Dans le cas i.i.d. (chapitre 3), on montre même la limite pour tous les convexes ouverts, et on a donc (SAC) . Profitons de cette remarque pour énoncer un principe de contraction. De manière générale, le principe de contraction ne s'applique pas directement aux PGD faibles car l'image réciproque, même par une application raisonnable, d'un relativement compact n'est pas relativement compacte. En revanche, la propriété (SAC) se transporte bien.

Théorème 4.3.4.3 (Principe de contraction linéaire). *Supposons que X soit muni d'une structure d'espace vectoriel et que $(\mu_n, v_n)_{n \geq 1}$ vérifie (SAC) avec, pour tout $x \in X$, $\mathcal{V}_x = \mathcal{C}_x$ l'ensemble des convexes ouverts mesurables contenant x . Soit Y un espace vectoriel, muni d'une tribu $\mathcal{G}$ et d'une topologie σ telles que, pour tout $y \in Y$, l'ensemble $\mathcal{C}_y$ des voisinages de y convexes ouverts mesurables soit un système fondamental de voisinages. Soit $f : X \rightarrow Y$ mesurable, continue et linéaire. Alors $(\mu_n \circ f^{-1}, v_n)_{n \geq 1}$ vérifie (SAC) . En particulier, la suite $(\mu_n \circ f^{-1}, v_n)_{n \geq 1}$ vérifie un PGD faible.*

Démonstration : Soient $y \in Y$ et $C \in \mathcal{C}_y$. Les hypothèses sur f assurent que $f^{-1}(C)$ est un convexe ouvert mesurable de X . La propriété (SAC) permet alors de conclure que

$$\liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n \circ f^{-1}(C) = \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n \circ f^{-1}(C)$$

Passant à l'infimum en $C \in \mathcal{C}_y$, on obtient le résultat. $\square$

4.3.5 Condition nécessaire et suffisante de PGD faible

On étend ici, à notre cadre, une caractérisation des PGD faibles à l'aide de l'entropie, mentionnée dans [LP95]. Cela requiert simplement une hypothèse raisonnable de séparation avec des ensembles mesurables.

Définition 4.3.5.1. *Soient X un ensemble, $\mathcal{F}$ une tribu sur X et τ une topologie sur X . On dit que X vérifie l'axiome de séparation T_2^m si, pour tout $(x_1, x_2) \in X^2$ avec $x_1 \neq x_2$, il existe $(A_1, A_2) \in \mathcal{F}^2$ tel que $x_1 \in \overset{\circ}{A}_1$, $x_2 \in \overset{\circ}{A}_2$ et $A_1 \cap A_2 = \emptyset$.*

On définit³

$$\bar{s}_b(x) = \inf_{\substack{U \in \tau \\ U \ni x}} \sup_{\substack{K \in \mathcal{F} \\ K \subset \subset U}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(K)$$

³ On aurait pu aussi penser définir

$$\bar{s}_{b,1}(x) = \inf_{\substack{A \in \mathcal{F} \\ \overset{\circ}{A} \ni x}} \sup_{\substack{K \in \mathcal{F} \\ K \subset \subset \overset{\circ}{A}}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) \quad \text{ou} \quad \bar{s}_{b,2}(x) = \inf_{\substack{A \in \mathcal{F} \\ A \ni x}} \sup_{\substack{K \in \mathcal{F} \\ \overline{K} \subset \subset A}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A)$$

mais on se rend compte que ce ne sont pas les bonnes fonctions.

La notation $\overline{K} \subset U$ signifie que $\overline{K}$ est un compact inclus dans U . On notera que $\overline{s}_b \leq \overline{s}$.

Proposition 4.3.5.2. *Supposons que X vérifie T_2^m . Alors $(\mu_n, v_n)_{n \geq 1}$ vérifie un PGD faible si et seulement si*

$$\overline{s}_b \leq \underline{s}$$

Remarque : Sans supposer que X vérifie T_2^m , si $(\mu_n, v_n)_{n \geq 1}$ vérifie un PGD faible, alors $\overline{s}_b \leq \underline{s}$.

Démonstration : Pour l'implication directe, soient $A \in \mathcal{F}$ et $K \in \mathcal{F}$ tels que $\overline{K} \subset \overset{\circ}{A}$. Alors, le PGD faible donne

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(K) \leq \sup_{\overline{K}} s \leq \sup_{\overset{\circ}{A}} s \leq \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A)$$

On en déduit, pour tout $x \in X$,

$$\overline{s}_b(x) = \inf_{\substack{U \in \tau \\ U \ni x}} \sup_{\substack{K \in \mathcal{F} \\ \overline{K} \subset U}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(K) \leq \inf_{\substack{A \in \mathcal{F} \\ \overset{\circ}{A} \ni x}} \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) = \underline{s}(x)$$

Pour l'autre implication, il suffit de voir que $\overline{s}_b$ vérifie (BS_b) . Pour cela, il suffit de vérifier que

$$\mathfrak{s}(U) = \sup_{\substack{K \in \mathcal{F} \\ \overline{K} \subset U}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(K)$$

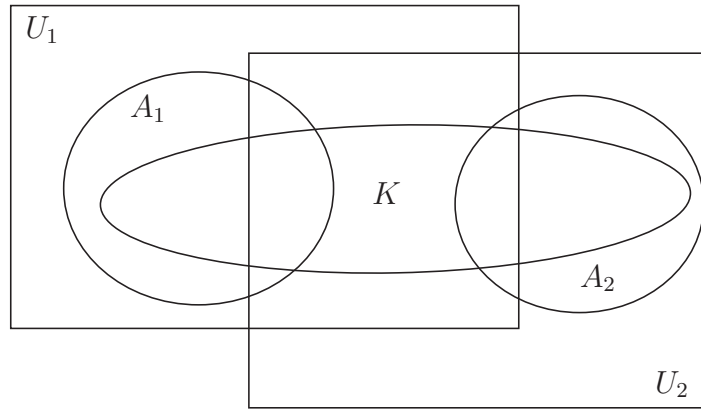
vérifie le *principe of the largest term* :

$$\overline{s}_b(U_1 \cup U_2) = \overline{s}_b(U_1) \vee \overline{s}_b(U_2)$$

En vertu du lemme 4.6.2.1, cela revient à montrer que

$$\sup_{\substack{K \in \mathcal{F} \\ \overline{K} \subset U_1 \cup U_2}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(K) = \sup_{\substack{K_1 \in \mathcal{F} \\ \overline{K}_1 \subset U_1}} \sup_{\substack{K_2 \in \mathcal{F} \\ \overline{K}_2 \subset U_2}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(K_1 \cup K_2)$$

Si $\overline{K} \subset U_1 \cup U_2$, $\overline{K} \setminus U_2$ et $\overline{K} \setminus U_1$ sont deux compacts disjoints. L'hypothèse T_2^m donne alors l'existence de $A_1, A_2 \in \mathcal{F}$ tels que $A_1 \cap A_2 = \emptyset$, $\overline{K} \setminus U_2 \subset \overset{\circ}{A}_1$ et $\overline{K} \setminus U_2 \subset \overset{\circ}{A}_2$. Posant $K_1 = K \setminus A_2$ et $K_2 = K \setminus A_1$, on a $\overline{K}_1 \subset U_1$, $\overline{K}_2 \subset U_2$ et $K = K_1 \cup K_2$.



On en déduit l'égalité voulue. $\square$

Au passage, cela permet d'affiner certains résultats, par exemple la borne supérieure. La démonstration suivante est un peu différente de la première.

Proposition 4.3.5.3. *Si X vérifie T_2^m et si φ est une fonction mesurable majorée telle que $K = \{\varphi > -\infty\} \cap D$ soit relativement compact, alors*

$$p(\varphi) \leq \sup_{x \in \overline{K}} (\varphi(x) + \overline{s}_b(x))$$

Démonstration : Soit $\delta > 0$. On définit, pour $i \in \mathbb{Z}$,

$$K(\delta, i) = K \cap \{i \cdot \delta \leq \varphi \leq (i+1) \cdot \delta\}$$

Comme φ est mesurable, $K(\delta, i) \in \mathcal{F}$. De plus, notons que $\overline{K(\delta, i)} \subseteq \overline{K}$ est compact. Comme φ est majorée, il existe $i_0 \in \mathbb{Z}$ tel que

$$\forall i > i_0 \quad K(\delta, i) = \emptyset$$

Fixons $M \geq i_0$. On montre que :

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{n\varphi} d\mu_n &\leq \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \left(\sum_{|i| \leq M} e^{n(i+1)\delta} \mu_n(K(\delta, i)) + e^{-nM\delta} \right) \\ &= \left(\bigvee_{|i| \leq M} (i+1) \cdot \delta + \overline{s}(K(\delta, i)) \right) \vee (-M\delta) \\ &\leq \sup_{x \in \overline{K}} (\varphi(x) + \delta + \overline{s}_b(x)) \vee (-M\delta) \end{aligned}$$

Reste à prendre l'infimum en $M \geq i_0$, puis à prendre l'infimum en $\delta > 0$. $\square$

Notons que la condition φ majorée est en fait relative à la suite $(\mu_n, v_n)_{n \geq 1}$: il suffit que

$$\liminf_{a \rightarrow +\infty} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \int_{|\varphi| \geq a} e^{n\varphi} d\mu_n = -\infty$$

En découle une autre version du lemme de Varadhan :

Théorème 4.3.5.4 (Lemme de Varadhan). *Supposons que X vérifie T_2^m et que $(\mu_n, v_n)_{n \geq 1}$ vérifie un PGD faible. Alors, pour toute fonction $\varphi : X \rightarrow]-\infty, +\infty[$ mesurable, semi-continue inférieurement et majorée, en notant*

$$d\tilde{\mu}_n = e^{n\varphi} d\mu_n$$

la suite $(\tilde{\mu}_n, v_n)_{n \geq 1}$ vérifie un PGD faible, l'entropie associée étant $\varphi + s$.

4.4 Espaces topologiques localement mesurables

On introduit enfin une condition reliant tribu et topologie. Non seulement cette hypothèse permet de montrer que l'entropie est semi-continue supérieurement, mais elle se comporte bien vis à vis des limites projectives que nous introduisons ensuite.

4.4.1 Définition

Soient X un ensemble, $\mathcal{F}$ une tribu sur X et τ une topologie sur X . On suppose que (ETLM) tout point x de X admet un système fondamental de voisinages mesurables $\mathcal{V}_x$. On dit alors que $(X, (\mathcal{V}_x)_{x \in X}, \mathcal{F}, \tau)$ est un *espace topologique localement mesurable*.

Proposition 4.4.1.1. *Sur un espace topologique localement mesurable, les fonctions $\underline{s}$ et $\overline{s}$ sont semi-continues supérieurement et ont pour expressions, pour tout $x \in X$,*

$$\underline{s}(x) = \inf_{V \in \mathcal{V}_x} \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(V) \quad \text{et} \quad \overline{s}(x) = \inf_{V \in \mathcal{V}_x} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(V)$$

Démonstration : Les expressions données de $\underline{s}$ et $\overline{s}$ découlent de la croissance de μ_n , pour tout n . Montrons que $\underline{s}$ est semi-continue supérieurement. Soient $t \in \mathbb{R}$ et $x \in X$ tels que $s(x) < t$. Par définition de $s(x)$, il existe $V \in \mathcal{V}_x$ tel que

$$\liminf_{n \geq 1} \frac{1}{v_n} \log \mu_n(V) < t$$

Alors, pour tout $y \in V$, il existe $W \in \mathcal{V}_y$ tel que $W \subset V$, et ainsi

$$s(y) \leq \liminf_{n \geq 1} \frac{1}{v_n} \log \mu_n(W) \leq \liminf_{n \geq 1} \frac{1}{v_n} \log \mu_n(V) < t$$

Donc $\{x \in X ; s(x) < t\}$ est ouvert. La démonstration est analogue pour $\overline{s}$. $\square$

4.4.2 Limites projectives d'espaces topologiques localement mesurables

Soient X un ensemble muni d'une tribu $\mathcal{F}$ et

$$\overleftarrow{\mathcal{X}} = (X_i, f_i, f_{ij})_{i \leq j}$$

une famille telle que

(PROJ₁) les indices i et j décrivent un ensemble $(J, \leq)$ préordonné filtrant à droite ;

(PROJ₂) pour tout $i \in J$, $(X_i, (\mathcal{V}_{x,i})_{x \in X}, \mathcal{F}_i, \tau_i)$ est un espace topologique localement mesurable, f_i une application mesurable de X dans X_i et, pour tout $(i, j) \in J^2$ tel que $i \leq j$, f_{ij} est une application mesurable de X_j dans X_i telle que, pour tous $x \in X_j$ et $V \in \mathcal{V}_{f_{ij}(x), i}$,

$$f_{ij}^{-1}(V) \in \mathcal{V}_{x,j}$$

(en particulier, f_{ij} est continue) ;

(PROJ₃) pour tout $(i, j, k) \in J^3$ tel que $i \leq j \leq k$, on a $f_{ii} = id_{X_i}$,

$$f_i = f_{ij} \circ f_j \quad \text{et} \quad f_{ik} = f_{ij} \circ f_{jk}$$

Pour tout $x \in X$, on définit

$$\mathcal{V}_x = \{f_i^{-1}(V_i) ; i \in J, V_i \in \mathcal{V}_{f_i(x), i}\}$$

et τ la topologie initiale sur X pour la famille $(f_i)_{i \in J}$. Alors, $(X, (\mathcal{V}_x)_{x \in X}, \mathcal{F}, \tau)$ est un espace topologique localement mesurable. On dit que $\overleftarrow{\mathcal{X}}$ est un *système projectif d'espaces topologiques localement mesurables* et que X est sa *limite projective*⁴.

Théorème 4.4.2.1. Soient $\overleftarrow{\mathcal{X}} = (X_i, f_i, f_{ij})_{(i,j) \in J^2}$ un système projectif d'espaces topologiques localement mesurables et $(X, (\mathcal{V}_x)_{x \in X}, \mathcal{F}, \tau)$ sa limite projective. Soient $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité sur $\mathcal{F}$ et $(v_n)_{n \geq 1}$ une suite de réels strictement positifs.

- Si, pour tout $i \in J$, $(\mu_n \circ f_i^{-1}, v_n)_{n \geq 1}$ vérifie (SA_b), alors $(\mu_n, v_n)_{n \geq 1}$ vérifie (SA_b).
- Pour tout $(i, j) \in J^2$, si $i \leq j$, on a :

$$\underline{s}_i \circ f_i \geq \underline{s}_j \circ f_j \quad \text{et} \quad \overline{s}_i \circ f_i \geq \overline{s}_j \circ f_j$$

De plus,

$$\underline{s} = \inf_{i \in J} (\underline{s}_i \circ f_i) \quad \text{et} \quad \overline{s} = \inf_{i \in J} (\overline{s}_i \circ f_i)$$

En particulier, si, pour tout $i \in J$, $\underline{s}_i = \overline{s}_i$, alors $\underline{s} = \overline{s}$.

- Soit D une partie de X telle que, pour tout $n \geq 1$, μ_n soit portée par D . Si, pour tout $i \in J$, les compacts de X_i sont mesurables et $(\mu_n \circ f_i^{-1}, v_n)_{n \geq 1}$ vérifie un PGD faible, alors $(\mu_n, v_n)_{n \geq 1}$ vérifie (BS_{b,D}).

Démonstration : Pour le premier point, soient $x \in X$ et $B \in \mathcal{F}$ un voisinage de x . Soit $V \in \mathcal{V}_x$ tel que $V \subset B$. Il existe $i \in J$ et $V_i \in \mathcal{V}_{f_i(x), i}$ tels que $V = f_i^{-1}(V_i)$. L'hypothèse (SA_b) appliquée à $(\mu_n \circ f_i^{-1}, v_n)_{n \geq 1}$ donne l'existence d'un voisinage $A_i \in \mathcal{F}_i$ tel que

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n \circ f_i^{-1}(A_i) \leq \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n \circ f_i^{-1}(V_i)$$

Enfin, il existe $W_i \in \mathcal{V}_{f_i(x), i}$ tel que $W_i \subset A_i$, de sorte que $A = f_i^{-1}(W_i) \in \mathcal{V}_x$ vérifie

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) \leq \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(B)$$

Pour le deuxième point, nous ferons la démonstration pour les entropies inférieures. Soit $(i, j) \in J^2$ avec $i \leq j$. On a

$$\underline{s}_i \circ f_i = \underline{s}_i \circ f_{ij} \circ f_j$$

⁴ Au niveau des ensembles, cette définition de limite projective d'ensembles est un peu plus générale que celle de [Bou66, III.51], les structures d'espaces mesurables ne se comportant pas bien vis à vis des limites projectives.

Pour montrer le premier résultat, il suffit donc de montrer que $\underline{s}_i \circ f_{ij} \geq \underline{s}_j$. Et, en effet, pour tout $x \in X_j$,

$$\begin{aligned} \inf_{V_i \in \mathcal{V}_{f_{ij}(x), i}} \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n \circ f_i^{-1}(V_i) &= \inf_{V_i \in \mathcal{V}_{f_{ij}(x), i}} \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n \circ f_j^{-1} \circ f_{ij}^{-1}(V_i) \\ &\geq \inf_{V_j \in \mathcal{V}_{x, j}} \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n \circ f_j^{-1}(V_j) \end{aligned}$$

car $f_{ij}^{-1}(V_i) \in \mathcal{V}_{x, j}$. Puis, pour tout $x \in X$, on a :

$$\begin{aligned} \underline{s}(x) &= \inf_{V \in \mathcal{V}_x} \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(V) \\ &= \inf_{i \in J} \inf_{V_i \in \mathcal{V}_{f_i(x), i}} \liminf_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(f_i^{-1}(V_i)) = \inf_{i \in J} \underline{s}_i(f_i(x)) \end{aligned}$$

Pour le troisième point, soient $K \in \mathcal{F}$ tel que $K \cap D$ soit relativement compact et $i \in J$. Etant donné que, pour tout $n \geq 1$, μ_n est la loi d'une variable aléatoire à valeurs dans D , on a :

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(K) \leq \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log(\mu_n \circ f_i^{-1})(f_i(\overline{K \cap D})) \leq \sup_{f_i(\overline{K \cap D})} s_i = \sup_{\overline{K \cap D}} s_i \circ f_i$$

Il reste à montrer que

$$\inf_{i \in J} \sup_{\overline{K \cap D}} s_i \circ f_i \leq \sup_{\overline{K \cap D}} \inf_{i \in J} s_i \circ f_i$$

Soient $\delta > 0$ et $M < 0$. Le point précédent permet de voir que, pour tout $x \in \overline{K \cap D}$, il existe $i(x) \in J$ tel que

$$s_{i(x)} \circ f_{i(x)}(x) \leq (s(x) + \delta/2) \wedge (M - \delta/2)$$

Etant donné que $s_{i(x)} \circ f_{i(x)}$ est s.c.s., il existe $V(x) \in \mathcal{V}_x$ tel que

$$\sup_{V(x)} s_{i(x)} \circ f_{i(x)} \leq (s_{i(x)} \circ f_{i(x)}(x) + \delta/2) \wedge M$$

On en déduit que

$$\sup_{V(x)} s_{i(x)} \circ f_{i(x)} \leq (s(x) + \delta) \wedge M$$

Du recouvrement de $\overline{K \cap D}$ par les $V(x)$, pour $x \in \overline{K \cap D}$, on peut extraire un sous-recouvrement fini, noté $\{V(x_k); k \in \{1, \dots, r\}\}$. Comme J est filtrant à droite, il existe $i \in J$ majorant $\{i(x_1), \dots, i(x_r)\}$. Alors, en vertu du premier point,

$$\forall k \in \{1, \dots, r\} \quad s_i \circ f_i \leq s_{i(x_k)} \circ f_{i(x_k)}$$

On en déduit

$$\sup_{\overline{K \cap D}} s_i \circ f_i \leq \max_{1 \leq k \leq r} (s(x_k) + \delta) \wedge M \leq \sup_{\overline{K \cap D}} (s(x) + \delta) \wedge M$$

On conclut en prenant l'infimum en $i \in J$, puis l'infimum en $\delta > 0$ et en $M < 0$. $\square$

On notera que les deux premiers points seraient encore vrais si on ne supposait que : pour tout $x \in X$, pour tout voisinage mesurable V de X , il existe $i \in J$ et un voisinage mesurable V_i de $f_i(x)$ tel que

$$V \supset f_i^{-1}(V_i)$$

Pour le dernier point en revanche, on se sert vraiment de la structure d'espace topologique localement mesurable. En ce qui concerne l'hypothèse que, pour tout $i \in J$, les compacts de X_i sont mesurables, il suffit de supposer que : pour tout $K \in \mathcal{F}$ tel que $K \cap D$ soit relativement compact et pour tout $i \in J$, il existe un ensemble mesurable compris entre $f_i(K \cap D)$ et $f_i(\overline{K})$. La démonstration s'adapte alors très bien, en remarquant que, comme K est relativement compact, $\overline{f_i(K)} = f_i(\overline{K})$. Notons enfin que, sans structure d'espace topologique localement mesurable, seuls subsistent les résultats suivants :

$$\underline{s}_i \circ f_i \geq \underline{s}_j \circ f_j \quad \text{et} \quad \overline{s}_i \circ f_i \geq \overline{s}_j \circ f_j$$

et

$$\underline{s} \leq \inf_{i \in J} (\underline{s}_i \circ f_i) \quad \text{et} \quad \overline{s} \leq \inf_{i \in J} (\overline{s}_i \circ f_i)$$

4.5 Espaces vectoriels topologiques mesurables

4.5.1 Transformation de Fenchel-Legendre

Soient X un espace vectoriel, $\mathcal{F}$ une tribu sur X et τ une topologie sur X . On note X^{*m} l'ensemble des formes linéaires continues et mesurables sur X . On définit, pour tout $x \in X$,

$$-\underline{p}^*(x) := \inf_{\lambda \in X^{*m}} (\underline{p}(\lambda) - \lambda(x)) \quad \text{et} \quad -\overline{p}^*(x) = \inf_{\lambda \in X^{*m}} (\overline{p}(\lambda) - \lambda(x))$$

Proposition 4.5.1.1. *Les fonctions $-\underline{p}^*$ et $-\overline{p}^*$ sont concaves et semi-continues supérieurement.*

Démonstration : Ce sont des infimums de fonctions affines continues. $\square$

Sous cette forme, les définitions de $-\underline{p}^*$ et $-\overline{p}^*$ rappellent fortement le corollaire 4.3.2.2, et on voit immédiatement que

$$\underline{s} \leq -\underline{p}^* \quad \text{et} \quad \overline{s} \leq -\overline{p}^*$$

La question de savoir s'il y a égalité est centrale dans la théorie de Cramér. Plus précisément, on s'intéresse à l'égalité reliant entropie et pression : l'égalité

$$\underline{s} = -\overline{p}^*$$

est-elle vérifiée ? Pour qu'elle le soit, plusieurs conditions doivent être réunies : les remarques précédentes montrent qu'il faut que $\underline{s} = \overline{s}$ et que $\underline{s}$ soit concave. Dans le cas a.d.i. du

chapitre 5, ces deux premières propriétés sont conséquences du lemme sous-additif. Ensuite, il est nécessaire que l'ensemble X^{*m} soit assez riche dans $\mathcal{M}$. Il s'agit là d'une hypothèse reliant la tribu et la topologie (cf. espaces vectoriels topologiques mesurables). De manière générale, ces conditions ne sont pas suffisantes. Notons que, s'il existe une partie K de X relativement compacte telle que, pour tout $n \geq 1$, μ_n soit portée par K , alors la partie sur la borne supérieure faible permet de conclure. Plus généralement, si la suite $(\mu_n)_{n \geq 1}$ est exponentiellement tendue, on a le résultat. Dans le cas a.d.i., le concept de convexe-tension est moins restrictif et plus pertinent pour aboutir au résultat.

4.5.2 Espaces vectoriels topologiques mesurables

Soient X un espace vectoriel, $\mathcal{F}$ une tribu sur X et τ une topologie sur X . On suppose que (EVTM₁) (X, τ) est un espace vectoriel topologique ;

(EVTM₂) tout point x de X a un système fondamental de voisinages mesurables $\mathcal{V}_x$;

(EVTM₃) toute forme linéaire continue est mesurable.

On dit alors que $(X, (\mathcal{V}_x)_{x \in X}, \mathcal{F}, \tau)$ est un *espace vectoriel topologique mesurable*. Si on note X^* (resp. X^{*m}) l'ensemble des formes linéaires continues (resp. continues et mesurables) sur X , on a $X^{*m} = X^*$.

En particulier, $(X, (\mathcal{V}_x)_{x \in X}, \mathcal{F}, \tau)$ est un espace topologique localement mesurable. Les deux autres hypothèses sont motivées par deux raisons. La première est qu'ainsi $\underline{p}^*$ et $\overline{p}^*$ sont des transformées de Fenchel-Legendre au sens habituel. En particulier, on a le théorème d'inversion (qui dit essentiellement que X^* est suffisamment riche dans l'ensemble des fonctions convexes semi-continues inférieurement). La seconde est qu'alors les fonctions $-\underline{p}^*$ et $-\overline{p}^*$ se comportent bien vis à vis des limites projectives. Remarquons que les *espaces localement convexes mesurables* introduits au chapitre 3 sont des espaces vectoriels topologiques mesurables. Les convexes y jouent un rôle crucial en lien avec la sous-additivité.

Remarque : En revanche, les fonctions $(-\underline{s})^*$ et $(-\overline{s})^*$ définies pas la transformée de Fenchel-Legendre inverse se comportent mal vis à vis des limites projectives. On montre que

$$(-\underline{s})^*(\lambda_i \circ f_i) = \sup_{x \in X} \inf_{j \geq i} \left(\lambda_i \circ f_{ij}(f_j(x)) + \underline{s}_j \circ f_j(x) \right)$$

et on ne peut pas intervertir le supremum et l'infimum, en général. On mentionnera simplement qu'on saurait intervertir si x décrivait un ensemble relativement compact (cf. la démonstration du théorème relatif aux limites projectives d'espaces topologiques localement mesurables).

4.5.3 Limites projectives d'espaces vectoriels topologiques mesurables

Soient X un espace vectoriel, muni d'une tribu $\mathcal{F}$, et

$$\overleftarrow{\mathcal{X}} = (X_i, f_i, f_{ij})_{i \leq j}$$

une famille telle que

(PROJ₁) les indices i et j décrivent un ensemble $(J, \leq)$ préordonné filtrant à droite ;

(PROJ₂) pour tout $i \in J$, $(X_i, (\mathcal{V}_{x,i})_{x \in X}, \mathcal{F}_i, \tau_i)$ est un espace vectoriel topologique mesurable, f_i une application linéaire mesurable de X dans X_i et, pour tout $(i, j) \in J^2$ tel que $i \leq j$, f_{ij} est une application linéaire mesurable de X_j dans X_i telle que, pour tous $x \in X_j$ et $V \in \mathcal{V}_{f_{ij}(x), i}$,

$$f_{ij}^{-1}(V) \in \mathcal{V}_{x,j}$$

(en particulier, f_{ij} est continue) ;

(PROJ₃) pour tout $(i, j, k) \in J^3$ tel que $i \leq j \leq k$, on a $f_{ii} = id_{X_i}$,

$$f_i = f_{ij} \circ f_j \quad \text{et} \quad f_{ik} = f_{ij} \circ f_{jk}$$

Pour tout $x \in X$, on définit

$$\mathcal{V}_x = \{f_i^{-1}(V_i) ; i \in J, V_i \in \mathcal{V}_{f_i(x), i}\}$$

et τ la topologie initiale sur X pour la famille $(f_i)_{i \in J}$. Alors, le lemme ci-après permet de voir que $(X, (\mathcal{V}_x)_{x \in X}, \mathcal{F}, \tau)$ est un espace vectoriel topologique mesurable. On dit que $\overleftarrow{\mathcal{X}}$ est un *système projectif d'espaces vectoriels topologiques mesurables* et que X est sa *limite projective*.

Lemme 4.5.3.1. Soient $\overleftarrow{\mathcal{X}} = (X_i, f_i, f_{ij})_{(i,j) \in J^2}$ un système projectif d'espaces vectoriels topologiques mesurables et $(X, (\mathcal{V}_x)_{x \in X}, \mathcal{F}, \tau)$ sa limite projective. Alors

$$X^* = \{\lambda_i \circ f_i ; i \in J, \lambda_i \in X_i^*\}$$

Démonstration : Soit $\lambda \in X^*$. Comme λ est continue, il existe $V \in \mathcal{V}_0$ tel que

$$V \subset \{x \in X ; \lambda(x) < 1\}$$

Il existe $i \in J$ et $V_i \in \mathcal{V}_{0,i}$ tels que $V = f_i^{-1}(V_i)$. On voit alors que

$$\forall x \in X \quad f_i(x) = 0 \Rightarrow \lambda(x) = 0$$

En effet, si $\lambda(x) \neq 0$, il existe $t \in \mathbb{R}$ tel que $\lambda(tx) > 1$, donc $f_i(tx) \notin V_i$, d'où $f_i(x) \neq 0$. On en déduit que l'on peut passer au quotient dans

$$\begin{array}{ccc} x \in X & \longrightarrow & \lambda(x) \in \mathbb{R} \\ & \searrow & \nearrow \\ & f_i(x) \in X_i & \end{array}$$

L'application quotient est une forme linéaire continue sur X_i .

□

Théorème 4.5.3.2. Soient $\overleftarrow{\mathcal{X}} = (X_i, f_i, f_{ij})_{(i,j) \in J^2}$ un système projectif d'espaces vectoriels topologiques mesurables et $(X, (\mathcal{V}_x)_{x \in X}, \mathcal{F}, \tau)$ sa limite projective. Soient $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité sur $\mathcal{F}$ et $(v_n)_{n \geq 1}$ une suite de réels strictement positifs. Alors, pour tout $i \in J$, on a :

$$\underline{p}_i(\lambda_i) = \underline{p}(\lambda_i \circ f_i) \quad \text{et} \quad \overline{p}_i(\lambda_i) = \overline{p}(\lambda_i \circ f_i)$$

et, pour tout $(i, j) \in J^2$, si $i \leq j$, on a :

$$-\underline{p}_i^* \circ f_i \geq -\underline{p}_j^* \circ f_j \quad \text{et} \quad -\overline{p}_i^* \circ f_i \geq -\overline{p}_j^* \circ f_j$$

De plus,

$$-\underline{p}^* = \inf_{i \in J} (-\underline{p}_i^* \circ f_i) \quad \text{et} \quad -\overline{p}^* = \inf_{i \in J} (-\overline{p}_i^* \circ f_i)$$

En particulier, si, pour tout $i \in J$, $s_i = -p_i^*$, alors $s = -p^*$.

Démonstration : On traitera à chaque fois le cas des pressions supérieures, l'autre cas étant analogue. Pour tout $i \in J$,

$$\overline{p}_i(\lambda_i) = \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{v_n \lambda_i} d(\mu_n \circ f_i^{-1}) = \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{v_n \lambda_i \circ f_i} d\mu_n = \overline{p}(\lambda_i \circ f_i)$$

Soit $(i, j) \in J^2$ avec $i \leq j$. Etant donné que $f_i = f_{ij} \circ f_j$, il suffit de montrer que $-\overline{p}_i^* \circ f_{ij} \geq -\overline{p}_j^*$. Soit $x \in X_j$. On a, en utilisant le point précédent et l'égalité $f_i = f_{ij} \circ f_j$,

$$\begin{aligned} -\overline{p}_i^* \circ f_{ij}(x) &= \inf_{\lambda_i \in X_i^*} (\overline{p}(\lambda_i \circ f_{ij} \circ f_j - \lambda_i \circ f_{ij}(x))) \\ &\geq \inf_{\lambda_j \in X_j^*} (\overline{p}(\lambda_j \circ f_j) - \lambda_j(x)) \\ &= -\overline{p}_j^*(x) \end{aligned}$$

Enfin, pour tout $x \in X$,

$$-\overline{p}^*(x) = \inf_{\lambda \in X^*} (\overline{p}(\lambda) - \lambda(x)) = \inf_{i \in J} \inf_{\lambda_i \in X_i^*} (\overline{p}(\lambda_i \circ f_i) - \lambda_i \circ f_i(x)) = \inf_{i \in J} (-\overline{p}_i^*(f_i(x))) \quad \square$$

Pour résumer les résultats précédents, on peut énoncer une version linéaire du théorème de Dawson-Gärtner dans le cadre des espaces vectoriels topologiques mesurables. On laisse au lecteur le soin d'adapter l'énoncé au cas où, pour tout $n \geq 1$, μ_n est portée par une partie D de X .

Théorème 4.5.3.3 (Dawson-Gärtner linéaire). Soient $\overleftarrow{\mathcal{X}} = (X_i, f_i, f_{ij})_{(i,j) \in J^2}$ un système projectif d'espaces vectoriels topologiques mesurables et $(X, (\mathcal{V}_x)_{x \in X}, \mathcal{F}, \tau)$ sa limite projective. Soient $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité sur $\mathcal{F}$ et $(v_n)_{n \geq 1}$ une suite de réels strictement positifs.

- Si, pour tout $i \in J$, $\underline{s}_i = \overline{s}_i$, alors $\underline{s} = \overline{s}$ et $(\mu_n, v_n)_{n \geq 1}$ vérifie un PGD faible.
- Si, pour tout $i \in J$, les compacts de X_i sont mesurables et $(\mu_n \circ f_i^{-1}, v_n)_{n \geq 1}$ vérifie un PGD faible, alors $(\mu_n, v_n)_{n \geq 1}$ vérifie un PGD faible.
- Si, pour tout $i \in J$, $s_i = -p_i^*$, alors $s = -p^*$.

4.6 Les trois fondements de la théorie de Cramér

Dans cette section, nous dégagons trois résultats généraux sur lesquels reposent la démonstration du cas i.i.d., résultats qui s'étendent au cas asymptotiquement découplé (seul le lemme sous-additif doit être adapté).

4.6.1 Lemme sous-additif

Le premier résultat de ce type remonte à un article de M. Fekete [Fek23].

Lemme 4.6.1.1. *Soit $(u(n))_{n \geq 1}$ une suite à valeurs dans $[0, +\infty]$. On suppose que (SA) u est sous-additive, i.e.*

$$\forall m, n \geq 1 \quad u(m+n) \leq u(m) + u(n)$$

Alors,

$$\liminf_{n \rightarrow \infty} \frac{u(n)}{n} = \inf_{n \geq 1} \frac{u(n)}{n}$$

Démonstration : Par sous-additivité, pour tous $d, m \geq 1$,

$$\frac{u(dm)}{dm} \leq \frac{u(m)}{m}$$

En faisant tendre d vers ∞ , on obtient

$$\liminf_{n \rightarrow \infty} \frac{u(n)}{n} \leq \liminf_{d \rightarrow \infty} \frac{u(dm)}{dm} \leq \frac{u(m)}{m}$$

d'où le résultat en passant à l'infimum en $m \geq 1$. □

Lemme 4.6.1.2. *Soit $(u(n))_{n \geq 1}$ une suite à valeurs dans $[0, +\infty]$. On suppose que (SA) u est sous-additive ;*

(C) u est contrôlée, i.e. il existe $N \geq 1$ tel que

$$\forall n \geq N \quad u(n) < +\infty$$

Alors, la suite $(u(n)/n)_{n \geq 1}$ converge vers

$$\inf_{n \geq 1} \frac{u(n)}{n}$$

Démonstration : Soient $n \geq m \geq N$. La division euclidienne de n par m s'écrit $n = mq + r$ avec $q \geq 1$ et $r \in \{0, \dots, m-1\}$; ainsi, la sous-additivité permet d'écrire

$$u(n) = u(mq + r) = u(m(q-1) + m + r) \leq (q-1)u(m) + u(m+r)$$

puis

$$\frac{u(n)}{n} \leq \frac{u(m)}{m} + \frac{1}{n} \max_{0 \leq i < m} u(m+i)$$

D'où, comme u est contrôlée, en faisant tendre n , puis m vers ∞ , on obtient

$$\limsup_{n \rightarrow \infty} \frac{u(n)}{n} \leq \liminf_{m \rightarrow \infty} \frac{u(m)}{m}$$

autrement dit la suite $(u(n)/n)_{n \geq 1}$ converge. D'après le lemme 4.6.1.1, sa limite est

$$\inf_{n \geq 1} \frac{u(n)}{n}$$

□

4.6.2 Intersion infimum-supremum

Il est question ici du fameux *principe of the largest term* de [LPS94a].

Lemme 4.6.2.1. *Si $(u_1(n))_{n \geq 1}, \dots, (u_r(n))_{n \geq 1}$ sont r suites à valeurs dans $[0, +\infty]$, on a l'égalité*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^r u_i(n) = \max_{1 \leq i \leq r} \limsup_{n \rightarrow \infty} \frac{1}{n} \log u_i(n)$$

Démonstration : De l'encadrement (on rappelle que les suites sont positives)

$$\max_{1 \leq i \leq r} u_i(n) \leq \sum_{i=1}^r u_i(n) \leq r \max_{1 \leq i \leq r} u_i(n)$$

on déduit que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^r u_i(n) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log \max_{1 \leq i \leq r} u_i(n)$$

Puis

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \max_{1 \leq i \leq r} u_i(n) &= \lim_{n \rightarrow \infty} \sup_{k \geq n} \frac{1}{k} \log \max_{1 \leq i \leq r} u_i(k) \\ &= \lim_{n \rightarrow \infty} \max_{1 \leq i \leq r} \left[\sup_{k \geq n} \frac{1}{k} \log u_i(k) \right] \\ &= \max_{1 \leq i \leq r} \lim_{n \rightarrow \infty} \left[\sup_{k \geq n} \frac{1}{k} \log u_i(k) \right] \\ &= \max_{1 \leq i \leq r} \limsup_{n \rightarrow \infty} \frac{1}{n} \log u_i(n) \end{aligned}$$

La troisième égalité vient du fait facile à vérifier que

$$\max : [-\infty, +\infty]^r \rightarrow [-\infty, +\infty]$$

est continue.

□

4.6.3 Transformation de Fenchel-Legendre

Pour une présentation plus large, on renvoie à [Mor67] et [Roc70]. La preuve présentée ici est reprise de [Cer07, proposition 12.2.]. Soit (X, τ) un espace vectoriel localement convexe. Notons X^* son dual topologique. Définissons, pour $f : X \rightarrow [-\infty, +\infty]$, sa *transformée de Fenchel-Legendre* par

$$\forall \lambda \in X^* \quad f^*(\lambda) := \sup_{x \in X} (\langle \lambda | x \rangle - f(x))$$

La transformée de Fenchel-Legendre de $g : X^* \rightarrow [-\infty, +\infty]$ se définit de façon analogue. Etendons l'addition à $[-\infty, +\infty]$ via

$$\forall a \in \mathbb{R} \quad \pm \infty + a = \pm \infty \quad \text{et} \quad -\infty + (+\infty) = -\infty$$

ainsi que la multiplication par un réel via

$$\begin{aligned} \forall \alpha > 0 \quad \alpha \cdot (\pm \infty) &= \pm \infty \\ \forall \alpha < 0 \quad \alpha \cdot (\pm \infty) &= \mp \infty \end{aligned} \quad \text{et} \quad 0 \cdot (\pm \infty) = 0$$

On dit que $f : X \rightarrow [-\infty, +\infty]$ est *convexe* si

$$\forall (x, y) \in X^2 \quad \forall \alpha \in [0, 1] \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

On notera que la seule fonction convexe prenant la valeur $-\infty$ est la fonction constante de valeur $-\infty$. Les fonctions convexes autres que les deux constantes $\pm \infty$ sont habituellement dénommées *fonctions convexes propres*. On dit que $f : X \rightarrow [-\infty, +\infty]$ est *concave* si $-f$ est convexe. On dit que $q : X \rightarrow [-\infty, +\infty]$ est *affine* si q est convexe et concave. Les fonctions affines à valeurs dans $[-\infty, +\infty]$ sont alors les fonctions affines habituelles à valeurs dans $] -\infty, +\infty[$ et les deux fonctions constantes $\pm \infty$.

Le seul résultat qui nous intéresse ici est le suivant :

Proposition 4.6.3.1. *Soit (X, τ) un espace vectoriel localement convexe et $f : X \rightarrow [-\infty, +\infty]$. Alors*

$$f^{**} = f$$

si et seulement si f est convexe et semi-continue inférieurement relativement à la topologie faible $\sigma(X, X^)$.*

Remarque : Plus précisément, la transformation de Fenchel-Legendre réalise une bijection des fonctions convexes $\sigma(X, X^*)$ -s.c.i. sur X sur les fonctions convexes $\sigma(X^*, X)$ -s.c.i. sur X^* , de sorte que, si f est convexe et $\sigma(X, X^*)$ -s.c.i., sa transformée de Fenchel-Legendre f^* prend le nom de fonction *convexe-conjuguée* de f .

Démonstration : L'implication directe découle du fait que, pour toute fonction $g : X^* \rightarrow [-\infty, +\infty]$, la fonction g^* est convexe et $\sigma(X, X^*)$ -s.c.i. Montrons la réciproque. On remarque que, si $\lambda \in X^*$, $\langle \lambda | \cdot \rangle - f^*(\lambda)$ est la plus grande fonction affine (y compris les deux fonctions affines $\pm \infty$) dirigée par λ et inférieure à f . Il s'agit donc de voir que f est la borne

supérieure de l'ensemble des fonctions affines continue plus petites que f (y compris les deux fonctions affines $\pm\infty$). Si $f = \pm\infty$, le résultat est immédiat. Sinon, f est une fonction convexe propre. Définissons l'épigraphe de f par $\text{epi}(f) = \{(x, t) \in X \times \mathbb{R}; f(x) \leq t\}$. Le fait que f soit convexe et $\sigma(X, X^*)$ -s.c.i. assure que $\text{epi}(f)$ est convexe et fermé relativement à τ . Soit $(x, t) \in X \times \mathbb{R} \setminus \text{epi}(f)$. Le théorème de Hahn-Banach dans l'espace localement convexe $X \times \mathbb{R}$ donne l'existence d'un hyperplan fermé séparant strictement (x, t) et $\text{epi}(f)$. Si cet hyperplan n'est pas vertical, il correspond à une fonction affine plus petite que f . Sinon, l'hyperplan est de la forme $H \times \mathbb{R}$ où H est un hyperplan affine fermé de X et $f(x) = +\infty$. Soit alors p une fonction affine continue sur X telle que $p(x) > 0$ et $p(y) = 0$ pour tout $y \in H$. Comme f est une fonction convexe propre, il existe une fonction affine continue et finie q inférieure à f : en effet, il existe $x \in X$ tel que $f(x) \in]-\infty, +\infty[$ et un hyperplan fermé H séparant $(x, f(x) - 1)$ de $\text{epi}(f)$; cet hyperplan H n'est pas vertical et correspond à une fonction affine continue et finie q inférieure à f . Ainsi, pour tout $\alpha > 0$, $q + \alpha p$ est toujours une fonction affine continue inférieure à f . Il suffit alors de choisir α tel que $q(x) + \alpha p(x) > t$. $\square$

4.7 Suppléments sur le PGD

4.7.1 Lemme de Varadhan fermé

Proposition 4.7.1.1. *Si φ est une fonction mesurable majorée et si $(\mu_n, v_n)_{n \geq 1}$ vérifie un PGD, alors :*

$$p(\varphi) \leq \sup_{x \in X} (\varphi(x) + \underline{s}(x))$$

Démonstration : Soit $\delta > 0$. On définit, pour $i \in \mathbb{Z}$,

$$A(\delta, i) = \{i \cdot \delta \leq \varphi \leq (i+1) \cdot \delta\}$$

Comme φ est mesurable, $A(\delta, i) \in \mathcal{F}$. Comme φ est majorée, il existe $i_0 \in \mathbb{Z}$ tel que

$$\forall i > i_0 \quad A(\delta, i) = \emptyset$$

Fixons $M \geq i_0$. On a :

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \int e^{n\varphi} d\mu_n &\leq \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \left(\sum_{|i| \leq M} e^{n(i+1)\delta} \mu_n(A(\delta, i)) + e^{-nM\delta} \right) \\ &= \left(\bigvee_{|i| \leq M} (i+1) \cdot \delta + \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A(\delta, i)) \right) \vee (-M\delta) \\ &\leq \sup_{x \in X} (\varphi(x) + \delta + \underline{s}(x)) \vee (-M\delta) \end{aligned}$$

Reste à prendre l'infimum en $M \geq i_0$, puis en $\delta > 0$. □

Notons que la condition φ majorée est en fait relative à la suite $(\mu_n, v_n)_{n \geq 1}$: il suffit que

$$\liminf_{a \rightarrow +\infty} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \int_{|\varphi| \geq a} e^{n\varphi} d\mu_n = -\infty$$

4.7.2 Conditions de PGD

On définit

$$\bar{s}_\#(x) = \inf_{\substack{U \in \tau \\ U \ni x}} \sup_{\substack{F \in \mathcal{F} \\ F \subset U}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(F)$$

On notera que $\bar{s}_\# \leq \bar{s}$. Contrairement aux PGD faibles, pour lesquels le *principle of the largest term* permettait de montrer une borne pour les compacts, on n'a facilement qu'une condition nécessaire de PGD fort.

Théorème 4.7.2.1 (Condition nécessaire de PGD fort). *Soit (μ_n) une suite de mesures localement bornées. On a :*

$$(\mu_n) \text{ vérifie un PGD fort} \implies \forall x \in X \quad \bar{s}_\#(x) \leq \underline{s}(x)$$

En s'inspirant de la preuve de (BS_b) , on peut donner une condition suffisante de PGD fort.

Théorème 4.7.2.2 (Condition suffisante de PGD fort). *Soit (μ_n) une suite de mesures localement bornées. Si $\bar{s}_\# \leq \underline{s}$ et si, pour tout $A \in \mathcal{F}$,*

$$\sup_{x \in \bar{A}} \inf_{\substack{U \in \tau \\ U \ni x}} \sup_{\substack{F \in \mathcal{F} \\ F \subset U}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(F) = \inf_{\substack{U \in \tau \\ U \supset \bar{A}}} \sup_{\substack{F \in \mathcal{F} \\ F \subset U}} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(F)$$

alors (μ_n) vérifie un PGD fort.

4.7.3 Identification de $\bar{s}_\#$ et de $\bar{s}_b$

Les résultats donnés ici sont adaptés de [LP95]

Définition 4.7.3.1. Soient X un ensemble, $\mathcal{F}$ une tribu sur X et τ une topologie sur X . On dit que X vérifie l'axiome de séparation T_3^m si, pour tout $x \in X$ et pour tout fermé F avec $x \notin F$, il existe $A \in \mathcal{F}$ tel que $x \in \mathring{A}$ et $F \cap \bar{A} = \emptyset$.

Sous l'hypothèse T_3^m , on montre que $\bar{s}_\# = \bar{s}$.

Définition 4.7.3.2. Soient X un ensemble, $\mathcal{F}$ une tribu sur X et τ une topologie sur X . On dit que X vérifie l'axiome LC^m si tout point de X admet une base de voisinages mesurables d'adhérences compactes.

Sous l'hypothèse LC^m , on montre que $\bar{s}_b = \bar{s}$. Notons que T_2 et LC^m entraînent T_2^m .

4.7.4 Tension exponentielle

Soient X un ensemble, $\mathcal{F}$ une tribu sur X et τ une topologie sur X . Soient $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité sur $\mathcal{F}$ et $(v_n)_{n \geq 1}$ une suite de réels strictement positifs. On dit que la suite $(\mu_n, v_n)_{n \geq 1}$ est *exponentiellement tendue* si, pour tout $M < 0$, il existe $K \in \mathcal{F}$ relativement compact tel que

$$\bar{p}(\delta_{X \setminus K}) = \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(X \setminus K) \leq M$$

Théorème 4.7.4.1. Si $(\mu_n, v_n)_{n \geq 1}$ est exponentiellement tendue et vérifie un PGD faible, alors $(\mu_n, v_n)_{n \geq 1}$ vérifie un PGD et, pour tout $t \in \mathbb{R}$, l'ensemble de niveau

$$\{x \in X ; s(x) \geq t\}$$

de l'entropie associée est relativement compact.

Démonstration : Soit $t \in \mathbb{R}$. Il existe $K \in \mathcal{F}$ relativement compact tel que $\bar{p}(\delta_{X \setminus K}) \leq t$. La borne inférieure pour $X \setminus K$ donne

$$\sup_{X \setminus \bar{K}} s \leq t$$

donc $\{x \in X; s(x) > t\} \subset \bar{K}$ et s est coercive. Soit maintenant $A \in \mathcal{F}$. Soit $M < 0$. Il existe $K \in \mathcal{F}$ relativement compact tel que $\bar{p}(\delta_{X \setminus K}) \leq M$. Alors

$$\mu_n(A) \leq \mu_n(A \cap K) + \mu_n(X \setminus K)$$

La borne supérieure faible pour $A \cap K$ et le lemme 4.6.2.1 montrent que

$$\limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(A) \leq \max \left(M, \sup_{A \cap K} s \right) \leq \max \left(M, \sup_A s \right)$$

On obtient la borne supérieure en faisant tendre M vers $-\infty$. $\square$

4.7.5 Condition suffisante de tension exponentielle

Soient (X, τ) un espace vectoriel topologique et $\mathcal{F}$ une tribu sur X invariante par dilatation. Soient $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilité sur $\mathcal{F}$ et $(v_n)_{n \geq 1}$ une suite de réels strictement positifs. Pour toute partie A de X , on définit sa fonctionnelle de Minkowski M_A par

$$\forall x \in X \quad M_A(x) = \inf \{t > 0; x \in tA\}$$

Théorème 4.7.5.1. *S'il existe $K \in \mathcal{F}$ relativement compact tel que*

$$\bar{p}(M_K) < +\infty$$

alors la suite $(\mu_n, v_n)_{n \geq 1}$ est exponentiellement tendue.

Démonstration : Soit $M < 0$ et $t = \bar{p}(M_K) - M$. L'inégalité de Tchebychev permet d'obtenir

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(X \setminus (tK)) &\leq \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \mu_n(v_n M_K > v_n t) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{v_n} \log \left(e^{-v_n t} \int e^{v_n M_K} d\mu_n \right) = \bar{p}(M_K) - t = M \quad \square \end{aligned}$$

5

Chapitre 5

THÉORIE DE CRAMÉR POUR DES CHAMPS ASYMPTOTIQUEMENT DÉCOUPLÉS

Sommaire

5.1	Introduction	108
5.2	Cadre	108
5.2.1	Espace des configurations	108
5.2.2	Découplage asymptotique	110
5.2.3	Contrôle local	111
5.2.4	Espaces vectoriels localement convexes mesurables	112
5.3	Théorie de Cramér pour des champs asymptotiquement découplés	113
5.3.1	Entropie et pression	113
5.3.2	Énoncé des résultats principaux	114
5.3.3	Exemples d'applications	115
5.4	Retour sur les hypothèses et compléments	115
5.4.1	Partition adaptée de $\mathbb{Z}^d$	115
5.4.2	Convexe-tension locale	118
5.5	Démonstration des résultats principaux	118
5.5.1	Entropie	118
5.5.2	Pression	123
5.5.3	Pression et entropie	124
5.6	Compléments	126
5.6.1	PGD fort	126
5.6.2	Cas d'égalité des entropies inférieure et supérieure	126
5.6.3	Propriétés de la pression	127
5.6.4	Champs contrôlés localement	128

5.1 Introduction

Ce chapitre étend la théorie de Cramér à l'autre généralisation importante : relaxer l'hypothèse d'indépendance. Nous importons ici les idées du chapitre 3 dans le cadre proposé par Pfister [Pfi02]. Son approche des théorèmes de Sanov (au niveau 3, dans la terminologie de [Ell85]) pour les mesures de Gibbs met en lumière la sous-additivité sous-jacente, donnant un cadre général englobant les résultats précédents ([Geo93], mais aussi [DV75] pour les chaînes de Markov). Voyant, comme [Cer07, chapitre 24], le théorème de Sanov comme une conséquence du théorème de Cramér, nous généralisons donc la théorie de Cramér pour les variables indépendantes au cadre asymptotiquement découplé de [Pfi02].

Le cadre contient celui du chapitre 3, mais aussi, cette fois, ceux de [BZ79] et de [Cer07]. La question de la mesurabilité de l'addition vis à vis de la tribu produit ne se posait que parce que nous considérions des variables indépendantes, dont les lois jointes étaient naturellement définies sur des tribus produits. Ici, nous rendons l'addition mesurable en considérant des tribus initiales (qui sont égales aux tribus produits dans le cas du chapitre 3). Ensuite, nous complétons l'hypothèse de *découplage asymptotique* de Pfister (assurant un analogue de la propriété (SA) du lemme sous-additif 4.6.1.2) par une hypothèse de *contrôle local* (pour avoir l'analogue de la propriété (C) de 4.6.1.2) et par la *convexe-tension locale* (analogue de la convexe-tension dans le cas i.i.d.). Les deux premières hypothèses permettent donc d'établir un lemme sous-additif (ou plutôt pseudo-sous-additif) duquel découle le PGD faible, comme dans le cas indépendant. La suite de la démonstration est analogue à celle du chapitre 3. On notera toutefois qu'on n'a plus (BS_c) mais une forme affaiblie.

5.2 Cadre

Après quelques préliminaires, nous introduisons deux notions qui permettront d'établir un lemme sous-additif pour des champs non indépendants analogue au lemme clef sur lequel reposait la démonstration dans le cas i.i.d. : le découplage asymptotique inférieur et le contrôle local. Ces deux notions correspondent, dans le cas i.i.d., aux propriétés de sous-additivité (SA) et de contrôle (C). D'autre part, l'invariance par translation sur le réseau $\mathbb{Z}^d$ remplacera l'identique distribution.

5.2.1 Espace des configurations

Soient X un espace vectoriel réel, $\mathcal{F}$ une tribu sur X stable par dilatation et translation et d un entier strictement positif. On appelle *boîte* un sous-ensemble cubique de $\mathbb{Z}^d$, autrement dit un translaté de

$$\Lambda(n) := [0, n]^d \cap \mathbb{Z}^d$$

pour un entier $n \geq 1$. On notera $\mathbb{B}$ l'ensemble des boîtes de $\mathbb{Z}^d$. Si $\sigma \in X^{\mathbb{Z}^d}$ et $z \in \mathbb{Z}^d$, on notera $\sigma(z)$ la coordonnée suivant z de σ . Pour tout $\Lambda \in \mathbb{B}$, on définit l'application

$$\mathbf{m}_\Lambda : \sigma \in X^{\mathbb{Z}^d} \mapsto \frac{1}{|\Lambda|} \sum_{z \in \Lambda} \sigma(z) \in X$$

Pour tout S sous-ensemble fini de $\mathbb{Z}^d$, on définit, sur $X^{\mathbb{Z}^d}$, la tribu $\mathcal{F}_S$ initiale pour les applications

$$\mathbf{m}_\Lambda : X^{\mathbb{Z}^d} \rightarrow X$$

où Λ décrit l'ensemble des boîtes incluses dans S . La tribu $\mathcal{F}_S$ est donc la plus petite tribu sur $X^{\mathbb{Z}^d}$ qui rende ces applications mesurables. Notons que, si $S_1 \subset S_2$, alors $\mathcal{F}_{S_1} \subset \mathcal{F}_{S_2}$. D'autre part, si $S \in \mathbb{B}$ et $z \in \mathbb{Z}^d$, alors $\mathcal{F}_{S+z} = \{A + z; A \in \mathcal{F}_S\}$ où, pour tout $A \subset X^{\mathbb{Z}^d}$, $A + z = \{\sigma(\cdot - z); \sigma \in A\}$.

Remarque : Pour tout S sous-ensemble fini de $\mathbb{Z}^d$, on a toujours l'inclusion $\mathcal{F}^{\otimes S} \subset \mathcal{F}_S$. Etant donné que $\mathcal{F}$ est stable par dilatation, l'égalité a lieu si et seulement si l'addition vectorielle

$$(x, y) \in (X^2, \mathcal{F}^{\otimes S}) \mapsto x + y \in (X, \mathcal{F})$$

est mesurable. Nous avons vu (cf. cas i.i.d. du chapitre 3) que c'est le cas si $\mathcal{F}$ est la tribu initiale pour une famille $(N_i, f_i)_{i \in I}$ où, pour tout $i \in I$, N_i est un espace vectoriel normé séparable muni de sa tribu borélienne. Mais, étant donné que, pour tout $\Lambda \in \mathbb{B}$, nous allons nous donner une loi jointe sur $\mathcal{F}_\Lambda$, et non simplement, comme dans le cas i.i.d., une loi sur $\mathcal{F}$ qui permet de construire un $|\Lambda|$ -uplet i.i.d., nous n'avons pas besoin de l'égalité $\mathcal{F}^{\otimes \Lambda} = \mathcal{F}_\Lambda$.

Enfin, pour tout $\Lambda \in \mathbb{B}$, on se donne une mesure de probabilité μ_Λ sur $\mathcal{F}_\Lambda$. On dira que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est un *système compatible de probabilités invariantes par translation (s.c.i.t.)* sur $X^{\mathbb{Z}^d}$ si :

(COMP) pour toutes boîtes Λ_1 et Λ_2 de $\mathbb{Z}^d$ avec $\Lambda_1 \subset \Lambda_2$, on a :

$$\forall A \in \mathcal{F}_{\Lambda_1} \quad \mu_{\Lambda_2}(A) = \mu_{\Lambda_1}(A)$$

(INV) pour toute boîte Λ de $\mathbb{Z}^d$ et pour tout $z \in \mathbb{Z}^d$, on a :

$$\forall A \in \mathcal{F}_\Lambda \quad \mu_{\Lambda+z}(A + z) = \mu_\Lambda(A)$$

Remarque : Dans de nombreux cas étudiés (mesures de Gibbs, chaînes de Markov...), il existe une mesure en volume infini décrivant les interactions, *i.e.* une probabilité $\mathbb{P}$ sur $X^{\mathbb{Z}^d}$ (c'est dans ce cadre que se place [Pfi02]). Dans ce cas, le s.c.i.t. $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est simplement la famille des lois marginales de $\mathbb{P}$ sur les X^Λ , pour $\Lambda \in \mathbb{B}$ (les deux conditions (COMP) et (INV) sont bien vérifiées). Cependant, il n'existe pas toujours de mesure en volume infini : dans le cas indépendant, le théorème d'extension de Kolmogorov ne s'applique que si l'espace X est polonais, ce que nous ne supposons pas ici. Et nous avons vu, dans

le chapitre 3, que l'étude des grandes déviations ne nécessitait que les marginales finidimensionnelles $\mu^{\otimes n}$. Nous avons donc introduit ici la notion de s.c.i.t. pour englober dans le présent cadre la théorie du chapitre 3. Toutefois, la construction habituelle de mesures en volume infini (les mesures de Gibbs, par exemple) repose sur un système de probabilités conditionnelles compatible au sens de Dobrushin, Lanford et Ruelle (DLR), et non au sens de Kolmogorov comme les systèmes que nous avons dénommés s.c.i.t. Ainsi, au lieu de construire une mesure en volume infini avec un système DLR, puis de définir le s.c.i.t. des marginales associées, il est sans doute possible d'obtenir un principe de grandes déviations directement sur le système DLR de départ. D'autant que, dans le cas indépendant, le système DLR et le s.c.i.t. coïncident. Ce sera l'objet d'un travail ultérieur.

Etant donné un tel s.c.i.t. $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$, on définit, pour tout $\Lambda \in \mathbb{B}$ et pour tout $A \in \mathcal{F}_\Lambda$,

$$\mathbb{P}(A) = \mu_\Lambda(A)$$

L'hypothèse (COMP) assure que $\mathbb{P}(A)$ est bien défini pour tout A dans l'union des $\mathcal{F}_\Lambda$ où $\Lambda \in \mathbb{B}$ (c'est une mesure additive sur le clan engendré par les $\mathcal{F}_\Lambda$, pour $\Lambda \in \mathbb{B}$). On définit également, pour $\Lambda \in \mathbb{B}$, $A \in \mathcal{F}_\Lambda$ et $\mathcal{G}$ une sous-tribu de $\mathcal{F}_\Lambda$,

$$\mathbb{P}(A|\mathcal{G}) = \mu_\Lambda(A|\mathcal{G})$$

et, pour $\Lambda \in \mathbb{B}$ et $Z : X^{\mathbb{Z}^d} \rightarrow [-\infty, +\infty]$ une application $\mathcal{F}_\Lambda$ -mesurable dont l'intégrale contre μ_Λ a un sens,

$$\mathbb{E}(Z) = \int Z(\sigma) d\mu_\Lambda(\sigma)$$

Si D est une partie de $X^{\mathbb{Z}^d}$, on dira que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est *porté par* D si, pour tout $\Lambda \in \mathbb{B}$, pour tout $(A, B) \in (\mathcal{F}_\Lambda)^2$,

$$A \cap D = B \cap D \Rightarrow \mu_\Lambda(A) = \mu_\Lambda(B)$$

C'est le cas si et seulement si, pour tout $\Lambda \in \mathbb{B}$, μ_Λ est la loi d'un champ σ à valeurs dans D (cf. cas i.i.d.). Comme dans le cas i.i.d., on fera souvent appel à cette description. Remarquons que σ dépend de Λ , mais nous ne préciserons pas cette notation.

5.2.2 Découplage asymptotique

Soient X un espace vectoriel réel, $\mathcal{F}$ une tribu sur X stable par dilatation et translation, d un entier strictement positif et $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ un s.c.i.t. sur $X^{\mathbb{Z}^d}$. Nous reprenons ici la définition de découplage de [Pfi02]. Elle donne une condition de dépendance faible entre ce qui se passe dans une boîte et loin de cette boîte. Explicitons. On notera dist la distance associée à la norme $|\cdot|_\infty$ sur $\mathbb{Z}^d$. On dit que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est *asymptotiquement découplé inférieurement* (*a.d.i.*) s'il existe deux applications g et c de $\mathbb{N}^*$ dans $[0, +\infty[$ telles que

$$\frac{g(m)}{m} \rightarrow 0 \quad \text{et} \quad \frac{c(m)}{|\Lambda(m)|} \rightarrow 0$$

et telles que : pour tout $m \geq 1$, pour tout sous-ensemble fini S de $\mathbb{Z}^d$, pour tous $A \in \mathcal{F}_{\Lambda(m)}$ et $B \in \mathcal{F}_S$,

$$\text{dist}(S, \Lambda(m)) > g(m) \Rightarrow \mathbb{P}(A \cap B) \geq e^{-c(m)} \mathbb{P}(A) \mathbb{P}(B)$$

On dit que (g, c) est un *paramètre de découplage* de $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$. On verra par la suite que nous suffirait l'inégalité pour des A et B convexes mesurables particuliers. En effet, on aura seulement besoin de : pour tout $k \geq 1$, pour toutes boîtes $\Lambda_1, \dots, \Lambda_k$ de $\mathbb{Z}^d$ de taille m et telles que, pour tous $1 \leq i < j \leq k$, $\text{dist}(\Lambda_i, \Lambda_j) > g(n)$, pour tous $C_1, \dots, C_k$ translatés de convexes internes,

$$\mathbb{P} \left(\bigcap_{i=1}^k \{\mathbf{m}_{\Lambda_i} \in C_i\} \right) \geq e^{-(k-1)c(n)} \prod_{i=1}^k \mathbb{P}(\mathbf{m}_{\Lambda_i} \in C_i)$$

Mais, en pratique, l'inégalité de découplage sera vérifiée pour tous les A et B mesurables.

5.2.3 Contrôle local

On reprend les notations de la section précédente. Si C est un convexe de X , sa jauge (ou fonctionnelle de Minkowski) est l'application $M_C : X \rightarrow [0, +\infty]$ définie par

$$\forall x \in X \quad M_C(x) = \inf\{t \geq 0; x \in tC\}$$

On dit que C est *interne* si sa jauge M_C est finie partout et si

$$C = \{x \in X; M_C(x) < 1\}$$

L'ensemble des convexes internes est l'ensemble des convexes, voisinages ouverts de 0 pour une certaine topologie localement convexe sur X . On renvoie au cas i.i.d. (chapitre 3) pour plus de détails sur les convexes internes. On note $\mathcal{C}_0(\mathcal{F})$ l'ensemble des convexes internes mesurables de X . On veut maintenant définir une condition de contrôle de chaque site conditionnellement au reste de la configuration. Soit $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ un s.c.i.t. sur $X^{\mathbb{Z}^d}$. On dit que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est *contrôlé localement (c.l.)* s'il existe deux applications t et α de $\mathcal{C}_0(\mathcal{F})$ dans $]0, +\infty[$ telles que : pour tout $C \in \mathcal{C}_0(\mathcal{F})$ et pour tout $\Lambda \in \mathbb{B}$,

$$\mathbb{P}(\sigma(0) \in t(C) \cdot C | \mathcal{F}_{\Lambda \setminus \{0\}}) \geq \alpha(C)$$

On dit que (t, α) est un *paramètre de contrôle* de $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$. Comme pour le découplage asymptotique inférieur, on pourrait se contenter de l'inégalité pour des conditionnements par des ensembles A de la forme

$$\bigcap_{i=1}^k \{\mathbf{m}_{\Lambda_i} \in C_i\}$$

avec $\Lambda_1, \dots, \Lambda_k$ boîtes disjointes de $\mathbb{Z}^d \setminus \{0\}$ et $C_1, \dots, C_k$ translatés de convexes internes. Encore une fois, en pratique, l'inégalité de contrôle sera vérifiée pour tous les conditionnements A mesurables.

5.2.4 Espaces vectoriels localement convexes mesurables

Etant donné que l'on n'a pas de problème de mesurabilité des applications $\mathbf{m}_\Lambda$ (contrairement au cas i.i.d.), on étend la notion d'espace vectoriel localement convexe mesurable précédemment définie. Soient X un espace vectoriel réel, $\mathcal{F}$ une tribu sur X stable par dilatation et translation. On suppose qu'il existe une famille $\mathcal{C}_0$ de parties de X telle que :
 (EVL $\mathcal{C}\mathcal{M}_1$) pour tout $C \in \mathcal{C}_0$, C est un convexe interne mesurable (*i.e.* $\mathcal{C}_0 \subset \mathcal{C}_0(\mathcal{F})$) et symétrique ;

(EVL $\mathcal{C}\mathcal{M}_2$) $\mathcal{C}_0$ est stable par intersection finie et par dilatation de rapport non nul, et l'intersection des éléments de $\mathcal{C}_0$ est réduite à $\{0\}$;

Alors, $\mathcal{C}_0$ est un système fondamental de voisinages de 0 pour une unique topologie localement convexe (séparée) τ . Si, de plus

(EVL $\mathcal{C}\mathcal{M}_3$) toute forme linéaire continue sur X est mesurable,

on dit que $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est un *espace vectoriel localement convexe mesurable* (e.v.l.c.m.).

Remarque : On a supprimé deux hypothèses : on n'impose plus la séparabilité de la topologie localement convexe engendrée par $C \in \mathcal{C}_0$; et on ne demande pas que $\mathcal{F}$ soit engendrée par les translatés des éléments de $\mathcal{C}_0$.

On étend en conséquence la notion de limite projective d'e.v.l.c.m. évoquée au chapitre 3. Soient X un espace vectoriel réel, $\mathcal{F}$ une tribu sur X stable par dilatation et translation, et

$$\overleftarrow{\mathcal{X}} = (X_i, f_i, f_{ij})_{i \leq j}$$

une famille telle que

(PROJ $_1$) les indices i et j décrivent un ensemble $(J, \leq)$ préordonné filtrant à droite ;

(PROJ $_2$) pour tout $i \in J$, $(X_i, \mathcal{C}_{0,i}, \mathcal{F}_i, \tau_i)$ est un e.v.l.c.m., f_i une application linéaire mesurable de X dans X_i et, pour tout $(i, j) \in J^2$ tel que $i \leq j$, f_{ij} est une application linéaire, continue et mesurable de X_j dans X_i ;

(PROJ $_3$) pour tout $(i, j, k) \in J^3$ tel que $i \leq j \leq k$, on a $f_{ii} = id_{X_i}$,

$$f_i = f_{ij} \circ f_j \quad \text{et} \quad f_{ik} = f_{ij} \circ f_{jk}$$

On dit alors que $\overleftarrow{\mathcal{X}}$ est un *système projectif d'espaces vectoriels localement convexes mesurables*. De plus, si l'on note

$$\mathcal{C}_0 = \{f_i^{-1}(C_i) ; i \in J, C_i \in \mathcal{C}_0(\mathcal{F}_i)\}$$

et τ la topologie localement convexe engendrée par $\mathcal{C}_0$, $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est un e.v.l.c.m., appelé *limite projective*¹ du système projectif d'e.v.l.c.m. $\overleftarrow{\mathcal{X}}$.

¹ Au niveau des ensembles, cette définition de limite projective d'ensembles est un peu plus générale que celle de [Bou66, III.51.]

5.3 Théorie de Cramér pour des champs asymptotiquement découplés

Dans toute cette partie, $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ désigne un e.v.l.c.m., d un entier strictement positif et $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ un s.c.i.t. sur $X^{\mathbb{Z}^d}$. Pour tout $x \in X$, on note

$$\mathcal{C}_x = \{x + C; C \in \mathcal{C}_0\}$$

et, pour tout $n \geq 1$,

$$\mu_n = \mu_{\Lambda(n)} \circ (\mathbf{m}_{\Lambda(n)})^{-1}$$

On dira que $(\mu_n)_{n \geq 1}$ est la *suite de Cramér* associée au s.c.i.t. $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$. Si D est une partie convexe de X et si $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est porté par $D^{\mathbb{Z}^d}$, alors, pour tout $n \geq 1$, μ_n est une probabilité portée par D , au sens où, pour tout $(A, B) \in \mathcal{F}^2$,

$$A \cap D = B \cap D \Rightarrow \mu_n(A) = \mu_n(B)$$

5.3.1 Entropie et pression

On appelle *entropie* de $(\mu_n)_{n \geq 1}$ la fonction s définie par : pour tout $x \in X$,

$$s(x) := \inf_{C \in \mathcal{C}_x} \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(C)$$

Par construction, l'entropie s est la plus grande fonction vérifiant la borne inférieure :

(BI) pour tout $A \in \mathcal{F}$,

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(A) \geq \sup_{\overset{\circ}{A}} s$$

On dit que $(\mu_n)_{n \geq 1}$ vérifie un *principe de grandes déviations* (PGD) si la borne supérieure suivante est vérifiée :

(BS) pour tout $A \in \mathcal{F}$,

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(A) \leq \sup_{\overline{A}} s$$

De manière générale, (BS) n'est pas vérifiée pour tous les mesurables. Si $\mathcal{P}$ désigne un ensemble de parties mesurables de X , on définit la version restreinte de la borne supérieure suivante :

(BS $_{\mathcal{P}}$) pour tout $A \in \mathcal{P}$,

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(A) \leq \sup_{\overline{A}} s$$

En particulier, si D est une partie de X , on notera :

$(BS_{b,D})$ pour tout $K \in \mathcal{F}$ tel que $K \cap D$ soit relativement compact,

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(K) \leq \sup_K s$$

Si $(\mu_n)_{n \geq 1}$ vérifie $(BS_{b,X})$, on dit que $(\mu_n)_{n \geq 1}$ vérifie un *principe de grandes déviations faible* (PGD faible). On appelle *pression* de $(\mu_n)_{n \geq 1}$ la fonction p définie par : pour tout $\lambda \in X^*$ (mesurable, par hypothèse),

$$p(\lambda) := \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \int e^{|\Lambda(n)| \langle \lambda | x \rangle} d\mu_n(x)$$

5.3.2 Énoncé des résultats principaux

Si $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est un s.c.i.t. et si D est une partie de X , on dit que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est *convexe-tendu localement sur D* (c.t.l. sur D) si, pour tout $\gamma > 0$, il existe $K(\gamma) \in \mathcal{F}$ tel que $K(\gamma) \cap D$ soit convexe relativement compact et tel que, pour toute boîte Λ contenant 0,

$$\mathbb{P}(\sigma(0) \in K(\gamma) | \mathcal{F}_{\Lambda \setminus \{0\}}) \geq 1 - \gamma$$

Théorème 5.3.2.1. *Soient $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ un e.v.l.c.m., d un entier strictement positif, D une partie convexe de X , $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ un s.c.i.t. sur $X^{\mathbb{Z}^d}$ porté par $D^{\mathbb{Z}^d}$, et $(\mu_n)_{n \geq 1}$ la suite de Cramér associée. Si $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est a.d.i. et c.l., alors*

- $(\mu_n)_{n \geq 1}$ vérifie $(BS_{b,D})$; en particulier, $(\mu_n)_{n \geq 1}$ et $(\mu_n|_D)_{n \geq 1}$ vérifient un PGD faible ;
- la pression de $(\mu_n)_{n \geq 1}$ est définie par

$$\forall \lambda \in X^* \quad p(\lambda) = \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \int e^{|\Lambda(n)| \langle \lambda | x \rangle} d\mu_n(x)$$

et vérifie

$$\forall \lambda \in X^* \quad \forall x \in X \quad p(\lambda) - s(x) \geq \langle \lambda | x \rangle$$

- si, de plus, $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est c.t.l. sur D , alors l'entropie est l'opposée de la convexe-conjuguée de la pression, autrement dit

$$\forall x \in X \quad s(x) = \inf_{\lambda \in X^*} (p(\lambda) - \langle \lambda | x \rangle)$$

Le théorème précédent s'étend aux limites projectives en utilisant le théorème de Dawson-Gärtner linéaire 4.5.3.3 énoncé au chapitre 4. Si f est une application de X dans un ensemble E , on définit

$$f^{\mathbb{Z}^d} : (\sigma(z))_{z \in \mathbb{Z}^d} \in X^{\mathbb{Z}^d} \mapsto (f(\sigma(z)))_{z \in \mathbb{Z}^d} \in E^{\mathbb{Z}^d}$$

Théorème 5.3.2.2 (Limites projectives). *Soient $\overleftarrow{\mathcal{X}} = (X_i, f_i, f_{ij})_{(i,j) \in J^2}$ un système projectif d'e.v.l.c.m. et $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ sa limite projective. Soient d un entier strictement positif, D une partie convexe de X , $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ un s.c.i.t. sur $X^{\mathbb{Z}^d}$ porté par $D^{\mathbb{Z}^d}$ et $(\mu_n)_{n \geq 1}$ la suite*

de Cramér associée. Si, pour tout $i \in J$, $(\mu_\Lambda \circ (f_i^{\mathbb{Z}^d})^{-1})_{\Lambda \in \mathbb{B}}$ est a.d.i. et c.l., alors

- $(\mu_n)_{n \geq 1}$ vérifie $(BS_{b,D})$; en particulier, $(\mu_n)_{n \geq 1}$ et $(\mu_n|_D)_{n \geq 1}$ vérifient un PGD faible ;
- la pression de $(\mu_n)_{n \geq 1}$ est définie par

$$\forall \lambda \in X^* \quad p(\lambda) = \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \int e^{|\Lambda(n)| \langle \lambda | x \rangle} d\mu_n(x)$$

et vérifie

$$\forall \lambda \in X^* \quad \forall x \in X \quad p(\lambda) - s(x) \geq \langle \lambda | x \rangle$$

- Si, de plus, pour tout $i \in J$, $(\mu_\Lambda \circ (f_i^{\mathbb{Z}^d})^{-1})_{\Lambda \in \mathbb{B}}$ est c.t.l. sur $f_i(D)$, alors l'entropie est l'opposée de la convexe-conjuguée de la pression, autrement dit

$$\forall x \in X \quad s(x) = \inf_{\lambda \in X^*} (p(\lambda) - \langle \lambda | x \rangle)$$

5.3.3 Exemples d'applications

Le cadre défini ici contient la théorie de Cramér indépendante (du chapitre 3) et le cadre de [Pfi02]. En particulier, on a donc le PGD faible et l'égalité $s = -p^*$ dans le cadre de la théorie de Cramér dans les espaces de Banach séparables, ainsi qu'en topologie faible, et dans le cadre du théorème de Sanov. D'autre part, pour les champs non indépendants, la théorie développée dans ce chapitre s'applique immédiatement aux exemples mentionnés dans [Pfi02] (champs de Gibbs, chaînes de Markov, *etc.*) pour les principes de grandes déviations de niveau 2 ou 3 (dans la terminologie de [Ell85]), car la relative compacité assure le contrôle local et la convexe-tension locale. Pour les résultats de niveau 1, le contrôle local est vérifié notamment si η prend ses valeurs dans un compact ou si η est a.d.i. avec $g(1) = 0$.

5.4 Retour sur les hypothèses et compléments

Dans toute cette partie, $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ désigne un e.v.l.c.m., d un entier strictement positif et $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ un s.c.i.t. sur $X^{\mathbb{Z}^d}$. Pour tout $x \in X$, on note

$$\mathcal{C}_x = \{x + C ; C \in \mathcal{C}_0\}$$

et, pour tout $n \geq 1$,

$$\mu_n = \mu_{\Lambda(n)} \circ (\mathbf{m}_{\Lambda(n)})^{-1}$$

la suite de Cramér associée.

5.4.1 Partition adaptée de $\mathbb{Z}^d$

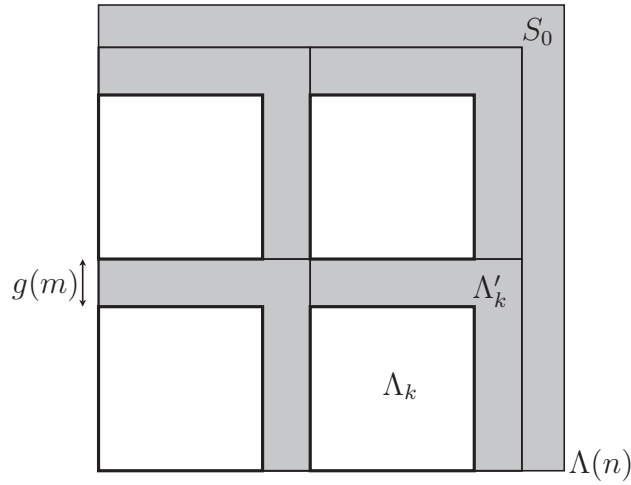
Nous décrivons, dans cette section, une partition des sites de $\mathbb{Z}^d$ adaptée au découplage asymptotique (nous reprenons la construction de Pfister). Soient m et n deux entiers tels que $n \geq m \geq 1$. La division euclidienne de n par $m + g(m)$ s'écrit

$$n = q(m + g(m)) + r$$

On pave (pas complètement) $\Lambda(n)$ avec q^d boîtes isométriques à $\Lambda(m+g(m))$, que l'on note Λ'_k , pour $k \in [1, q^d]$. Puis, pour chaque k , on note Λ_k la boîte isométrique à $\Lambda(m)$ dont le plus petit élément pour l'ordre lexicographique usuel de $\mathbb{Z}^d$ — le “coin en bas à gauche” — coïncide avec le coin en bas à gauche de Λ'_k . Soit enfin

$$S_0 = \Lambda(n) \setminus \bigcup_{k=1}^{q^d} \Lambda_k$$

l'ensemble des sites marginaux. Les boîtes Λ_k dépendent de n et m : ces indices seront sous-entendus.



On notera que la distance entre deux boîtes Λ_k distinctes est supérieure à $g(m)$, ce qui permet d'utiliser le découplage asymptotique et d'énoncer :

Proposition 5.4.1.1. *Pour tout $k \in \{1, \dots, q^d\}$, soit $f_k : X \rightarrow [-\infty, +\infty]$ mesurable. On a :*

$$\mathbb{E} \left(\prod_{k=1}^{q^d} e^{f_k(\mathbf{m}_{\Lambda_k}(\sigma))} \right) \geq e^{-q^d c(m)} \prod_{k=1}^{q^d} \mathbb{E} \left(e^{f_k(\mathbf{m}_{\Lambda_k}(\sigma))} \right)$$

Démonstration : On applique $q^d - 1$ fois l'hypothèse de découplage asymptotique inférieur :

$$\begin{aligned} \mathbb{E} \left(\prod_{k=1}^{q^d} e^{f_k(\mathbf{m}_{\Lambda_k}(\sigma))} \right) &= \int_{t_1, \dots, t_{q^d} \geq 0} \mathbb{P} \left(\forall k \in \{1, \dots, q^d\} \quad e^{f_k(\mathbf{m}_{\Lambda_k}(\sigma))} > t_k \right) \\ &\geq \int_{t_1, \dots, t_{q^d} \geq 0} (e^{-c(m)})^{q^d-1} \prod_{k=1}^{q^d} \mathbb{P} \left(e^{f_k(\mathbf{m}_{\Lambda_k}(\sigma))} > t_k \right) \\ &\geq e^{-q^d c(m)} \prod_{k=1}^{q^d} \mathbb{E} \left(e^{f_k(\mathbf{m}_{\Lambda_k}(\sigma))} \right) \end{aligned} \quad \square$$

Proposition 5.4.1.2. *Soit $f : X \rightarrow [-\infty, +\infty]$ une application mesurable. On suppose qu'il existe $\beta \in \mathbb{R}$ et $C \in \mathcal{C}_0(\mathcal{F})$ tels que*

$$\{x \in X ; f(x) \geq \beta\} \supset t(C) \cdot C$$

Alors, en notant $S = \Lambda_1 \cup \dots \cup \Lambda_{q^d}$,

$$\mathbb{E} \left(\prod_{z \in S_0} e^{f(\sigma(z))} \middle| \mathcal{F}_S \right) \geq (e^\beta \alpha(C))^{|S_0|}$$

Si, de plus, $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est porté par $D^{\mathbb{Z}^d}$ et c.t.l. sur D , et si $\gamma \leq \alpha(C)/2$, alors

$$\mathbb{E} \left(\prod_{z \in S_0} e^{f(\sigma(z))} 1_{\sigma(z) \in K(\gamma)} \middle| \mathcal{F}_S \right) \geq (e^\beta \alpha(C)/2)^{|S_0|}$$

Démonstration : Soient $z_0 \in S_0$ et $S_1 = S \cup (S_0 \setminus \{z_0\})$. On a :

$$\mathbb{E} \left(\prod_{z \in S_0} e^{f(\sigma(z))} \middle| \mathcal{F}_S \right) = \mathbb{E} \left(\mathbb{E}(e^{f(\sigma(z_0))} \middle| \mathcal{F}_{S_1}) \prod_{z \in S_0 \setminus \{z_0\}} e^{f(\sigma(z))} \middle| \mathcal{F}_S \right)$$

Or, l'inégalité de Tchebychev donne

$$\mathbb{E}(e^{f(\sigma(z_0))} \middle| \mathcal{F}_{S_1}) \geq e^\beta \mathbb{P}(f(\sigma(z_0)) \geq \beta \middle| \mathcal{F}_{S_1}) \geq e^\beta \mathbb{P}(\sigma(z_0) \in t(C) \cdot C \middle| \mathcal{F}_{S_1}) \geq e^\beta \alpha(C)$$

Poursuivant par récurrence, on obtient le premier résultat. Quant au second, il suffit d'adapter la preuve précédente ainsi :

$$\mathbb{E}(e^{f(\sigma(z_0))} 1_{\sigma(z_0) \in K(\gamma)} \middle| \mathcal{F}_{S_1}) \geq e^\beta \mathbb{P}(\sigma(z_0) \in (t(C) \cdot C) \cap K(\gamma) \middle| \mathcal{F}_{S_1}) \geq e^\beta \alpha(C)/2$$

et de continuer par récurrence. □

Notons

$$\rho_{m,n} := \frac{|S_0|}{|\Lambda(n)|}$$

la proportion de sites marginaux dans $\Lambda(n)$. Un petit calcul donne

$$\rho_{m,n} = 1 - \left(1 - \frac{r}{n}\right)^d \left(1 - \frac{g(m)}{m + g(m)}\right)^d \lesssim d \cdot \left(\frac{m}{n} + \frac{g(m)}{m + g(m)}\right) \xrightarrow[m \geq m]{m \rightarrow \infty} 0$$

Remarquons que, pour que $\rho_{m,n}$ puisse être rendu arbitrairement petit, il faut que le second terme de la dernière parenthèse tende vers 0 avec m , ce qui équivaut à

$$\frac{g(m)}{m} \rightarrow 0$$

L'hypothèse faite sur g permet donc que la proportion de sites marginaux soit asymptotiquement négligeable. Ainsi l'hypothèse de contrôle local s'avérera suffisante pour contrôler les sites marginaux.

5.4.2 Convexe-tension locale

On rappelle que, si $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est un s.c.i.t. et si D est une partie de X , on dit que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est *convexe-tendue localement sur D* (c.t.l. sur D) si, pour tout $\gamma > 0$, il existe $K(\gamma) \in \mathcal{F}$ tel que $K(\gamma) \cap D$ soit convexe relativement compact et tel que, pour toute boîte Λ contenant 0,

$$\mathbb{P}(\sigma(0) \in K(\gamma) | \mathcal{F}_{\Lambda \setminus \{0\}}) \geq 1 - \gamma$$

En particulier, passant aux espérances, on voit que la loi μ_1 de $\sigma(0)$ est convexe-tendue sur D (au sens entendu dans le cas i.i.d.). Plus généralement :

Proposition 5.4.2.1. *Si $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est c.t.l. sur D , alors, pour tout $n \geq 1$, μ_n est convexe-tendue sur D . Plus précisément,*

$$\lim_{\gamma \rightarrow 0^+} \mu_n(K(\gamma)) = 1$$

Démonstration : Soient $n \in \mathbb{N}$ et $\gamma > 0$. Soit $K = K(\gamma/n^d)$. Alors

$$\mu_n(X \setminus K) \leq \mu_{\Lambda(n)}(\exists z \in \Lambda(n) \quad \sigma(z) \in X \setminus K) \leq n^d \mu_1(X \setminus K) < \gamma \quad \square$$

5.5 Démonstration des résultats principaux

On garde les notations introduites dans la partie précédente. Sachant que le second théorème de la partie Énoncé des résultats principaux est conséquence du premier et du théorème 4.5.3.3, on donne uniquement la démonstration du premier.

5.5.1 Entropie

Théorème 5.5.1.1 (lemme sous-additif). *Si $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est a.d.i. et c.l., alors, pour tout $x \in X$, pour tout $C \in \mathcal{F}$ convexe contenant un convexe $B \in \mathcal{C}_0(\mathcal{F})$ et pour tout $\varepsilon > 0$, il existe $M(\varepsilon) \in \mathbb{N}$ tel que, pour tous $n \geq m \geq M(\varepsilon)$,*

$$\frac{1}{|\Lambda(n)|} \log \mu_n(x + (1 + \varepsilon)C) \geq \frac{1}{|\Lambda(m)|} \log \mu_m(x + C) - \frac{c(m)}{\Lambda(m)} + \rho_{m,n} \log \alpha(B)$$

Démonstration : Quitte à travailler sur les champs $(\sigma(z) - x)_{z \in \mathbb{Z}^d}$, on se ramène à $x = 0$. Soient $C \in \mathcal{F}$ convexe contenant un convexe $B \in \mathcal{C}_0(\mathcal{F})$ et $m, n \in \mathbb{N}^*$ avec $m \leq n$. La sous-additivité et la $\mathbb{R}_+$ -linéarité de M_C montrent que

$$\begin{aligned} M_C(\mathbf{m}_{\Lambda(n)}(\sigma)) &\leq \sum_{z \in S_0} \frac{1}{|\Lambda(n)|} M_C(\sigma(z)) + \sum_{k=1}^{q^d} \frac{|\Lambda(m)|}{|\Lambda(n)|} M_C(\mathbf{m}_{\Lambda_k}(\sigma)) \\ &\leq \sum_{z \in S_0} \frac{1}{|S_0|} \rho_{m,n} M_C(\sigma(z)) + \sum_{k=1}^{q^d} \frac{1}{q^d} M_C(\mathbf{m}_{\Lambda_k}(\sigma)) \end{aligned}$$

D'où, pour $\varepsilon > 0$,

$$\begin{aligned}
& \frac{1}{|\Lambda(n)|} \log \mu_n((1 + \varepsilon)C) \\
& \geq \frac{1}{|\Lambda(n)|} \log \mathbb{P}\left(M_C(\mathbf{m}_{\Lambda(n)}(\sigma)) < 1 + \varepsilon\right) \\
& \geq \frac{1}{|\Lambda(n)|} \log \mathbb{P}\left(\forall z \in S_0 \quad \rho_{m,n} M_C(\sigma(z)) < \varepsilon; \right. \\
& \quad \left. \forall k \in [1, q^d] \quad M_C(\mathbf{m}_{\Lambda_k}(\sigma)) \leq 1\right) \\
& \geq \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(\mathbb{E} \left(\prod_{z \in S_0} \exp(\delta_{B_1}(\sigma(z))) \middle| \mathcal{F}_S \right) \prod_{k=1}^{q^d} \exp(\delta_C(\mathbf{m}_{\Lambda_k}(\sigma))) \right)
\end{aligned}$$

où on a noté $S = \Lambda_1 \cup \dots \cup \Lambda_{q^d}$, $B_1 = (\varepsilon/\rho_{m,n})B$ et $\delta_{B_1} = -\infty 1_{X \setminus B_1}$. Compte tenu de la majoration de $\rho_{m,n}$ donnée dans la section précédente, il existe $M(\varepsilon) \in \mathbb{N}$ tel que

$$\forall (m, n) \in \mathbb{N}^2 \quad n \geq m \geq M(\varepsilon) \Rightarrow \frac{\varepsilon}{\rho_{m,n}} \geq t(B)$$

Alors, pour $n \geq m \geq M(\varepsilon)$, la proposition 5.4.1.2, avec $\beta = 0$, permet de minorer l'espérance conditionnelle et d'obtenir

$$\frac{1}{|\Lambda(n)|} \log \mu_n((1 + \varepsilon)C) \geq \rho_{m,n} \log \alpha(B) + \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(\prod_{k=1}^{q^d} \exp(\delta_C(\mathbf{m}_{\Lambda_k}(\sigma))) \right)$$

Puis la proposition 5.4.1.1 donne

$$\begin{aligned}
\frac{1}{|\Lambda(n)|} \log \mu_n((1 + \varepsilon)C) & \geq \rho_{m,n} \log \alpha(B) - \frac{q^d c(m)}{|\Lambda(n)|} + \frac{1}{|\Lambda(n)|} \log \prod_{k=1}^{q^d} \mathbb{E} \left(\exp(\delta_C(\mathbf{m}_{\Lambda_k}(\sigma))) \right) \\
& \geq \rho_{m,n} \log \alpha(B) - \frac{c(m)}{|\Lambda(m)|} + \frac{1}{|\Lambda(m)|} \log \mu_m(C)
\end{aligned}$$

où on a utilisé le fait que

$$\frac{q^d}{|\Lambda(n)|} = \frac{1 - \rho_{m,n}}{|\Lambda(m)|} \leq \frac{1}{|\Lambda(m)|}$$

et l'invariance par translation. □

Remarque : Intuitivement, on a d'abord fait en sorte que l'écartement entre les petites boîtes soit suffisamment petit devant la taille de ces boîtes. Puis, on a pris suffisamment de petites boîtes pour que la proportion occupée par elles dans une grande boîte soit assez grande. Au final, on obtient une sous-additivité asymptotique entre les petites boîtes et la grande.

Corollaire 5.5.1.2. *Si $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est a.d.i. et c.l., on a l'égalité $\underline{s} = \bar{s}$.*

Démonstration : Prenant, dans l'inégalité du théorème précédent, et pour $C \in \mathcal{C}_0$, la limite inférieure en n , puis la limite supérieure en m , on obtient

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(x + (1 + \varepsilon)C) \geq \limsup_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \log \mu_m(x + C)$$

Reste à prendre l'infimum en $C \in \mathcal{C}_0$ pour conclure. $\square$

Proposition 5.5.1.3. *L'entropie s est semi-continue supérieurement et concave.*

Démonstration : Elle est analogue à celle dans le cas i.i.d. Montrons la semi-continuité supérieure. Soient $t \in \mathbb{R}$ et $x \in X$ tels que $s(x) < t$. Il existe $C \in \mathcal{C}_x$ tel que

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(C) < t$$

Alors, pour tout $y \in C$,

$$s(y) \leq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(C) < t$$

donc $x \in C \subset \{s < t\}$ et $\{s < t\}$ est ouvert. Pour la concavité, il suffit alors de prouver que

$$s\left(\frac{x+y}{2}\right) \geq \frac{s(x) + s(y)}{2}$$

La semi-continuité supérieure permet alors de passer des inégalités

$$s((1-u)x + uy) \geq (1-u)s(x) + us(y)$$

pour u dyadique à celles pour $u \in [0, 1]$. La preuve est un raffinement de la démonstration du lemme sous-additif (5.5.1.1). Soit $C \in \mathcal{C}_0$. Par continuité des applications d'espace vectoriel, pour tout $\varepsilon \in]0, 1[$, il existe $C_x \in \mathcal{C}_x$ et $C_y \in \mathcal{C}_y$ tels que

$$\frac{1}{2}C_x + \frac{1}{2}C_y \subset \frac{1}{2}(x+y) + (1-\varepsilon)C$$

Puis

$$\begin{aligned} & \frac{1}{|\Lambda(n)|} \log \mu_n\left(\frac{1}{2}(x+y) + C\right) \\ & \geq \frac{1}{|\Lambda(n)|} \log \mathbb{P}\left(\forall z \in S_0 \quad \rho_{m,n} M_C(\sigma(z)) < \varepsilon; \right. \\ & \quad \left. \bigcap_{q \text{ pair}} \{\mathbf{m}_{\Lambda_q} \sigma \in C_x\} \cap \bigcap_{q \text{ impair}} \{\mathbf{m}_{\Lambda_q} \sigma \in C_y\}\right) \end{aligned}$$

En poursuivant comme précédemment, on obtient

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n \left(\frac{1}{2}(x+y) + C \right) \\ \geq \frac{1}{2} \left(\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(C_x) + \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(C_y) \right) \\ \geq \frac{1}{2}(s(x) + s(y)) \end{aligned}$$

puis l'inégalité désirée en passant à l'infimum en $C \in \mathcal{C}_0$. $\square$

Pour achever la démonstration du PGD faible, on montre un résultat un peu plus fort. On étend l'addition à $[-\infty, +\infty]$ via

$$-\infty \dot{+} (+\infty) = +\infty$$

Si f est une fonction de X dans $[-\infty, +\infty]$, on dit que f est *semi-continue supérieurement* si, pour tout $t \in [-\infty, +\infty]$, l'ensemble

$$\{x \in X ; f(x) < t\}$$

est un ouvert de X . Pour toute fonction $f : X \rightarrow [-\infty, +\infty]$, on note $f^\bullet$ sa régularisée semi-continue supérieurement, autrement dit la plus petite fonction semi-continue supérieurement qui soit supérieure à f .

Théorème 5.5.1.4 (Varadhan compact). *Soit D une partie convexe de X . On suppose $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ a.d.i., c.l. et porté par $D^{\mathbb{Z}^d}$. Pour toute fonction mesurable $f : X \rightarrow [-\infty, +\infty]$ et pour tout $K \in \mathcal{F}$ tel que $K \cap D$ soit relativement compact,*

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(e^{|\Lambda(n)|f(\mathbf{m}_{\Lambda(n)}(\sigma))} 1_{\mathbf{m}_{\Lambda(n)}(\sigma) \in K} \right) \leq \sup_{\overline{K \cap D}} [f^\bullet \dot{+} s]$$

Démonstration : Soient $K \in \mathcal{F}$ relativement compact, $\delta > 0$ et $M < 0$. Par définition de $f^\bullet$, pour tout $x \in X$, il existe $C(x) \in \mathcal{C}_x$ tel que, pour tout $y \in C(x)$,

$$f(y) \leq \max(f^\bullet(x) + \delta, M)$$

Par définition de $s(x) = \overline{s}(x)$, quitte à réduire un peu $C(x)$, on peut supposer que

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(C(x)) \leq \max(s(x) + \delta, M)$$

Du recouvrement de $\overline{K \cap D}$ par les $C(x)$ avec $x \in \overline{K \cap D}$, on peut extraire un sous-recouvrement fini, noté $\{C(x_i); i \in \{1, \dots, r\}\}$. Pour tout $n \geq 1$, ayant choisi σ à valeurs

dans $D^{\mathbb{Z}^d}$, on a :

$$\begin{aligned} & \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(e^{|\Lambda(n)| f(\mathbf{m}_{\Lambda(n)}(\sigma))} 1_{\mathbf{m}_{\Lambda(n)}(\sigma) \in K} \right) \\ & \leq \frac{1}{|\Lambda(n)|} \log \sum_{i=1}^r \mathbb{E} \left(e^{|\Lambda(n)| f(\mathbf{m}_{\Lambda(n)}(\sigma))} 1_{\mathbf{m}_{\Lambda(n)}(\sigma) \in C(x_i)} \right) \\ & \leq \frac{1}{|\Lambda(n)|} \log \sum_{i=1}^r e^{|\Lambda(n)| \max(f^\bullet(x_i) + \delta, M)} \mu_n(C(x_i)) \end{aligned}$$

Prenant la limite supérieure en n et utilisant le lemme 4.6.2.1, on obtient :

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(e^{|\Lambda(n)| f(\mathbf{m}_{\Lambda(n)}(\sigma))} 1_{\mathbf{m}_{\Lambda(n)}(\sigma) \in K} \right) \\ & \leq \max_{1 \leq i \leq r} \left(\max(f^\bullet(x_i) + \delta, M) + \max((s(x_i) + \delta), M) \right) \\ & \leq \sup_{x \in K \cap D} \left(\max(f^\bullet(x) + \delta, M) + \max((s(x) + \delta), M) \right) \end{aligned}$$

ce qui donne le résultat attendu, en faisant tendre δ vers 0 et M vers $-\infty$. $\square$

Corollaire 5.5.1.5 (Borne supérieure faible). *Pour tout $K \in \mathcal{F}$ tel que $K \cap D$ soit relativement compact,*

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(K) \leq \sup_{K \cap D} s$$

En particulier, $(\mu_n)_{n \geq 1}$ vérifie un PGD faible.

Démonstration : Il suffit d'appliquer le lemme de Varadhan compact 5.5.1.4 à $f = 0$. $\square$

Terminons cette section avec un résultat proche de la borne supérieure pour les convexes dans le cas où $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est convexe-tendu localement.

Théorème 5.5.1.6. *Soit D une partie convexe de X . On suppose que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est a.d.i., c.l., porté par D et c.t.l. sur D . Alors, pour tout $C \in \mathcal{F}$ convexe contenant un convexe $B \in \mathcal{C}_0(\mathcal{F})$, on a :*

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(C) \leq \inf_{\varepsilon > 0} \sup_{(1+\varepsilon)\overline{C}} s$$

Remarque : On sait conclure qu'on a bien (BS_c) dans le cas où s a ses ensembles de niveau compacts.

Démonstration : Soit $\gamma \leq \alpha(B)/2$. On adapte la démonstration du lemme sous-additif 5.5.1.1 en appliquant la seconde partie de la proposition 5.4.1.2 et on obtient, pour tous $n \geq m \geq M(\varepsilon)$,

$$\frac{1}{|\Lambda(m)|} \log \mu_m(K(\gamma) \cap C) - \frac{c(m)}{|\Lambda(m)|} + \rho_{m,n} \log \frac{\alpha(B)}{2} \leq \frac{1}{|\Lambda(n)|} \log \mu_n(K(\gamma) \cap (1 + \varepsilon)C)$$

Puis, la borne supérieure faible donne

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(K(\gamma) \cap (1 + \varepsilon)C) \leq \sup_{\overline{K(\gamma) \cap (1 + \varepsilon)C}} s \leq \sup_{(1 + \varepsilon)\overline{C}} s$$

car $\overline{(1 + \varepsilon)C} = (1 + \varepsilon)\overline{C}$. Combinant les deux inégalités et prenant le supremum en $\gamma \leq \alpha(B)/2$, il vient, sachant que $\mu_m(K(\gamma)) \rightarrow 1$ lorsque $\gamma \rightarrow 0^+$ (cf. 5.4.2.1),

$$\frac{1}{|\Lambda(m)|} \log \mu_m(C) - \frac{c(m)}{|\Lambda(m)|} + \limsup_{n \rightarrow \infty} \rho_{m,n} \log \frac{\alpha(B)}{2} \leq \sup_{(1 + \varepsilon)\overline{C}} s$$

Prenant alors la limite supérieure en m , puis l'infimum en $\varepsilon > 0$, on obtient la borne supérieure voulue. $\square$

5.5.2 Pression

On rappelle que la pression de la suite $(\mu_n)_{n \geq 1}$ est définie par : pour tout $\lambda \in X^*$,

$$p(\lambda) := \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \int e^{|\Lambda(n)|\langle \lambda | x \rangle} d\mu_n(x)$$

Théorème 5.5.2.1. *Si $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est a.d.i. et c.l., on a*

$$p(\lambda) = \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \int e^{|\Lambda(n)|\langle \lambda | x \rangle} d\mu_n(x)$$

Démonstration : De même que pour l'entropie, nous allons montrer une inégalité sous-additive. Soient $m, n \in \mathbb{N}$ avec $m \leq n$. On écrit

$$\begin{aligned} & \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(\exp (|\Lambda(n)|\langle \lambda | \mathbf{m}_{\Lambda(n)}(\sigma) \rangle) \right) \\ &= \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(\mathbb{E} \left(\prod_{z \in S_0} \exp (\langle \lambda | \sigma(z) \rangle) \middle| \mathcal{F}_S \right) \prod_{k=1}^{q^d} \exp (|\Lambda(m)|\langle \lambda | \mathbf{m}_{\Lambda_k}(\sigma) \rangle) \right) \end{aligned}$$

avec $S_0 = \Lambda_1 \cup \dots \cup \Lambda_{q^d}$. Etant donné que $\lambda \in X^*$, il existe $C \in \mathcal{C}_0$ tel que $C \subset \{x \in X; \langle \lambda | x \rangle \geq -1\}$. On peut donc utiliser, comme pour l'entropie, les propositions 5.4.1.2 (avec $\beta = -t(C)$) puis 5.4.1.1 pour obtenir

$$\begin{aligned} & \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(e^{|\Lambda(n)|\langle \lambda | \mathbf{m}_{\Lambda(n)}(\sigma) \rangle} \right) \\ & \geq \rho_{m,n} (\log \alpha(C) - t(C)) - \frac{c(m)}{|\Lambda(m)|} + (1 - \rho_{m,n}) \frac{1}{|\Lambda(m)|} \log \mathbb{E} \left(e^{|\Lambda(m)|\langle \lambda | \mathbf{m}_{\Lambda(m)}(\sigma) \rangle} \right) \end{aligned}$$

On conclut alors en prenant la limite inférieure en n , puis la limite supérieure en m . $\square$

5.5.3 Pression et entropie

Théorème 5.5.3.1. *Pour tout $x \in X$ et pour tout $\lambda \in X^*$, on a :*

$$p(\lambda) - s(x) \geq \langle \lambda | x \rangle$$

Démonstration : Soient $x \in X$ et $\lambda \in X^*$. Définissons, pour $\varepsilon > 0$, le demi-espace ouvert (mesurable)

$$H = \{y \in X; \langle \lambda | y \rangle > \langle \lambda | x \rangle - \varepsilon\}$$

L'inégalité de Tchebychev donne

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(H) &= \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(|\Lambda(n)|\lambda - |\Lambda(n)|\langle \lambda | x \rangle + |\Lambda(n)|\varepsilon > 0) \\ &\leq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \int \exp(|\Lambda(n)|\lambda - |\Lambda(n)|\langle \lambda | x \rangle + |\Lambda(n)|\varepsilon) d\mu_n \\ &= p(\lambda) - \langle \lambda | x \rangle + \varepsilon \end{aligned}$$

Etant donné que H est un voisinage mesurable de x , il vient $s(x) \leq p(\lambda) - \langle \lambda | x \rangle + \varepsilon$. Puis, passant à l'infimum en $\varepsilon > 0$, on obtient

$$p(\lambda) - s(x) \geq \langle \lambda | x \rangle \quad \square$$

Dans le cas où $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est convexe-tendu localement, nous allons montrer que $s = -p^*$.

Lemme 5.5.3.2. *Soit D une partie convexe de X . On suppose que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est a.d.i., c.l., porté par D et c.t.l. sur D . Alors la pression est la fonction convexe-conjuguée de l'opposée de l'entropie, i.e.*

$$\forall \lambda \in X^* \quad p(\lambda) = \sup_{x \in X} (\langle \lambda | x \rangle + s(x))$$

Démonstration : L'inégalité $p \geq (-s)^*$ ne nécessite pas la convexe-tension : elle repose sur l'inégalité de Tchebychev. Il suffit de passer au supremum en $x \in X$ dans 5.5.3.1 :

$$\forall \lambda \in X^* \quad p(\lambda) \geq \sup_{x \in X} (\langle \lambda | x \rangle + s(x))$$

Montrons l'autre inégalité. Soient $\lambda \in X^*$ et $m, n \in \mathbb{N}$ avec $m \leq n$. Soit $C \in \mathcal{C}_0$ tel que $C \subset \{x \in X; \langle \lambda | x \rangle \geq -1\}$. Soit $\gamma \leq \alpha(C)/2$. On adapte la démonstration du théorème 5.5.2.1 en utilisant la seconde partie de la proposition 5.4.1.2 et on obtient

$$\begin{aligned} \frac{1}{|\Lambda(m)|} \log \mathbb{E} \left(e^{|\Lambda(m)|\langle \lambda | m_{\Lambda(m)}(\sigma) \rangle} 1_{m_{\Lambda(m)}(\sigma) \in K(\gamma)} \right) - \frac{c(m)}{|\Lambda(m)|} + \rho_{m,n} \left(\log \frac{\alpha(C)}{2} - t(C) \right) \\ \leq \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(e^{|\Lambda(n)|\langle \lambda | m_{\Lambda(n)}(\sigma) \rangle} 1_{m_{\Lambda(n)}(\sigma) \in K(\gamma)} \right) \end{aligned}$$

Puis, le lemme de Varadhan compact 5.5.1.4 donne

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(e^{|\Lambda(n)| \langle \lambda | \mathbf{m}_{\Lambda(n)}(\sigma) \rangle} 1_{\mathbf{m}_{\Lambda(n)}(\sigma) \in K(\gamma)} \right) \leq \sup_{x \in X} (\langle \lambda | x \rangle + s(x))$$

Combinant les deux dernières inégalités et passant au supremum en $\gamma \leq \alpha(C)/2$, il vient, sachant que $\mu_m(K(\gamma)) \rightarrow 1$ lorsque $\gamma \rightarrow 0^+$ (cf. 5.4.2.1) et par un argument d'uniforme intégrabilité (cf. lemme 3.6.5.2),

$$\frac{1}{|\Lambda(m)|} \log \mathbb{E} \left(e^{|\Lambda(m)| \langle \lambda | \mathbf{m}_{\Lambda(m)}(\sigma) \rangle} \right) - \frac{c(m)}{|\Lambda(m)|} + \limsup_{n \rightarrow \infty} \rho_{m,n} \left(\log \frac{\alpha(C)}{2} - t(C) \right) \leq \sup_X (\lambda + s)$$

Reste alors à prendre la limite en m . □

Théorème 5.5.3.3. *Soit D une partie convexe de X . On suppose que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ est a.d.i., c.l., porté par D et c.t.l. sur D . Alors l'entropie est l'opposée de la fonction convexe-conjuguée de la pression, i.e.*

$$\forall x \in X \quad s(x) = \inf_{\lambda \in X^*} (p(\lambda) - \langle \lambda | x \rangle)$$

Démonstration : Avec les notations de la proposition 4.6.3.1, le théorème précédent s'écrit $p = (-s)^*$. Comme $-s$ est convexe et s.c.i. (cf. la proposition 5.5.1.3), donc $\sigma(X, X^*)$ -s.c.i., la proposition 4.6.3.1 donne

$$p^* = (-s)^{**} = -s$$

□

5.6 Compléments

5.6.1 PGD fort

Il existe des conditions analogues à celles du cas i.i.d. donnant le PGD fort à partir du PGD faible. La plus simple est le cas où il existe un ensemble convexe mesurable K d'adhérence compacte tel que $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ soit porté par $K^{\mathbb{Z}^d}$. Voici une condition plus faible (et classique dans le cas i.i.d.) :

Théorème 5.6.1.1 (PGD fort). *Soit $(\mu_\Lambda)_{\Lambda \in \mathbb{B}}$ un s.c.i.t. a.d.i. et c.l. On suppose qu'il existe K convexe mesurable relativement compact de X tel que, en notant M_K sa jauge,*

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left(e^{|\Lambda(n)| M_K(\mathfrak{m}_{\Lambda(n)}(\sigma))} \right) < +\infty$$

alors $(\mu_n)_{n \geq 1}$ vérifie un PGD fort.

Démonstration : On renvoie aux théorèmes 4.7.4.1 et 4.7.5.1 du chapitre de résultats généraux. $\square$

5.6.2 Cas d'égalité des entropies inférieure et supérieure

En récrivant le corollaire 5.5.1.2, on a montré que, pour tout $x \in X$, pour tout $C \in \mathcal{C}_x$ et pour tout $\varepsilon \in]0, 1[$,

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(x + C) \geq \limsup_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \log \mu_m(x + (1 - \varepsilon)C)$$

On donne ici une condition suffisante pour que

$$\lim_{\varepsilon \rightarrow 0^+} \limsup_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \log \mu_m(x + (1 - \varepsilon)C) = \limsup_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \log \mu_m(x + C)$$

Proposition 5.6.2.1. *On suppose qu'il existe $y \in C$, $\beta \in [0, 1[$ et $\alpha_0 > 0$ tels que, pour tout sous-ensemble fini S de $\mathbb{Z}^d \setminus \{0\}$ et pour tout $D \in \mathcal{C}(X^S)$,*

$$\mathbb{P}(\sigma(0) \in y + \beta(C - y); \eta_S \in D) \geq \alpha_0 \mathbb{P}(\sigma_S \in D)$$

Alors

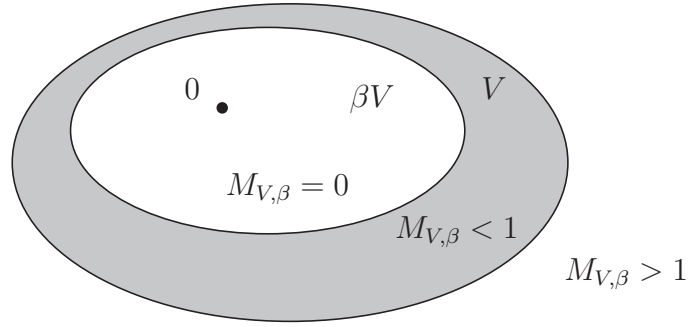
$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_n(x + C) = \limsup_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \log \mu_m(x + C)$$

Remarque : La condition qui apparaît ici est plus forte que l'hypothèse de contrôle local pour V . Elle impose, essentiellement, que la loi de $\sigma(0)$ donne de la masse à l'ouvert convexe C , conditionnellement au reste de la configuration.

Démonstration : On définit des jauges tronquées de V : pour tout $\beta \in [0, 1[$,

$$M_{V,\beta} = \frac{(\beta \vee M_V) - \beta}{1 - \beta}$$

et on note ϕ_β la fonction convexe $M_{V,\beta}(\cdot - y)$.



Par un cheminement analogue à celui de la démonstration du lemme sous-additif 5.5.1.1, on obtient

$$\begin{aligned}
 & \frac{1}{|\Lambda(n)|} \log \mu_m(x + (1 - \varepsilon)C) \\
 &= \frac{1}{|\Lambda(n)|} \log \mathbb{P} \left(\phi_\beta(\mathbf{m}_{\Lambda(n)}(\sigma)) < 1 - \frac{\varepsilon}{1 - \beta} \right) \\
 &\geq \frac{1}{|\Lambda(m)|} \log \mathbb{P}(\phi_\beta(\mathbf{m}_{\Lambda(m)}(\sigma)) < 1 - \frac{2\varepsilon}{1 - \beta}) - \frac{c(m)}{|\Lambda(m)|} + \rho_{m,n} \log \alpha_0
 \end{aligned}$$

En faisant tendre n vers ∞ , puis ε vers 0^+ , et enfin m vers ∞ , on obtient

$$\lim_{\varepsilon \rightarrow 0^+} \limsup_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \log \mu_m(x + (1 - \varepsilon)C) = \limsup_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \log \mu_m(x + C)$$

ce qu'on voulait. $\square$

La démonstration s'adapte aux limites projectives en considérant l'indice $i \in J$ tel que $C = f_i^{-1}(C_i)$.

5.6.3 Propriétés de la pression

Etendons l'addition à $[-\infty, +\infty]$ via

$$\forall a \in \mathbb{R} \quad \pm \infty + a = \pm \infty \quad \text{et} \quad -\infty + (+\infty) = -\infty$$

ainsi que la multiplication par un réel via

$$\begin{aligned}
 & \forall \alpha > 0 \quad \alpha \cdot (\pm \infty) = \pm \infty \\
 & \forall \alpha < 0 \quad \alpha \cdot (\pm \infty) = \mp \infty \quad \text{et} \quad 0 \cdot (\pm \infty) = 0
 \end{aligned}$$

On dit que $f : X \rightarrow [-\infty, +\infty]$ est *convexe* si

$$\forall (x, y) \in X^2 \quad \forall \alpha \in [0, 1] \quad f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$$

Proposition 5.6.3.1. *La pression est une fonction convexe.*

Démonstration : Pour tout $n \geq 1$, l'inégalité de Hölder permet de montrer que

$$\lambda \in X^{*m} \mapsto \frac{1}{v_n} \log \int e^{v_n \lambda} d\mu_n \in [-\infty, +\infty]$$

est convexe. On conclut en remarquant qu'une limite supérieure de fonctions convexes est convexe. $\square$

Comme dans le cas indépendant, on montre :

Proposition 5.6.3.2. *Si $\mu_{\Lambda(1)}$ est convexe-tendue localement, alors la pression est semi-continue inférieurement relativement à $\sigma(X, X^*)$.*

5.6.4 Champs contrôlés localement

Le modèle d'Ising (et plus généralement les mesures de Gibbs à espace d'états borné) est a.d.i. avec $g(1) = 0$. On montre par exemple que, pour toute boîte $\Lambda \subset \mathbb{Z}^d \setminus \{0\}$ et pour tout $A \in \mathcal{F}_\Lambda$,

$$\mathbb{P}(\sigma(0) = 1; \sigma_\Lambda \in A) \geq \gamma \mathbb{P}(\sigma(0) = 1) \mathbb{P}(\sigma_\Lambda \in A)$$

avec

$$\gamma = \frac{1}{(1 + e^{8\beta}) \mathbb{P}(\sigma(0) = 1)} > 0$$

6

Chapitre 6

UNE PREUVE DE $s = -p^*$ À PARTIR DU THÉORÈME DE MOSCO

Sommaire

6.1	Introduction	130
6.2	Lemme sous-additif	130
6.2.1	Découplage asymptotique convexe	130
6.2.2	Contrôle local	131
6.2.3	Enoncé du lemme sous-additif	132
6.2.4	Démonstration du lemme sous-additif	133
6.3	PGD faible	135
6.3.1	Limite projective	136
6.3.2	PGD faible	136
6.4	Pression	138
6.4.1	Pression en volume infini	139
6.5	Pression et entropie	142
6.5.1	Inégalité de Tchebychev	142
6.5.2	Égalité $s = -p^*$: cas borné	142
6.5.3	Égalité $s = -p^*$: cas c.a.d.i. et c.l.	143
6.5.4	Égalité $s = -p^*$: cas général	151

6.1 Introduction

La démonstration exposée ici de l'égalité

$$s = -p^*$$

dans le cas asymptotiquement découplé et pour un espace de Fréchet séparable (plus précisément pour une limite projective d'espaces de Fréchet séparables dans lesquels il y a le découplage asymptotique) a été rédigée bien avant les chapitres 3 et 5. Elle généralise, au cas asymptotiquement découplé, l'idée de [Cer07, chapitre 26] (reprise de [Zab92a] et [Zab92b]) d'utiliser le théorème de Mosco (cf. [Mos71]). La démonstration est plus laborieuse que celle apparaissant au chapitre précédent. Nous nous servons de l'égalité vraie si la coordonnée $\sigma(0)$ du champ σ en 0 prend ses valeurs dans un ensemble compact ; puis, pour l'étendre à une mesure quelconque, nous approximons σ par une suite de champs dont la coordonnée en 0 prend ses valeurs dans un ensemble compact.

Le cadre et les notations n'ont pas été repris pour être conformes à ceux du chapitre précédent. On pourra noter quelques variations de définitions. Par exemple, on considère d'emblée une mesure sur $X^{\mathbb{Z}^d}$, et non un système compatible de mesures. Nous refaisons la démonstration du lemme sous-additif dans le cas asymptotiquement découplé, cette fois pour une mesure, non plus invariante par toutes les translations sur le réseau $\mathbb{Z}^d$, mais seulement invariante par les translations d'un sous-réseau $(\ell\mathbb{Z})^d$ de $\mathbb{Z}^d$ (car les mesures que nous manipulerons pour l'approximation ne seront pas invariantes par translation).

On montre le résultat dans X , limite projective d'espaces Y_i . Il est raisonnable d'imposer aux Y_i d'être des espaces de Fréchet séparables. Voici en effet les hypothèses apparaissant au cours de la démonstration : pour tous z , la loi de $\eta(z) = f_i(\sigma(z))$ est tendue, les ouverts de Y_i sont des réunions dénombrables de pavés ouverts mesurables (c'est le cas si $\tau(Y_i)$ est à base dénombrable d'ouverts), Y_i est un e.v.l.c. séparable et métrisable (pour appliquer le théorème de Mosco), Y_i est tonnelé (pour appliquer le théorème de Banach-Steinhaus et pour l'argument final faisant intervenir un tonneau).

6.2 Lemme sous-additif

Soient (Y, τ) un espace de Fréchet séparable, $\mathcal{B}$ sa tribu borélienne et $d \in \mathbb{N}^*$.

6.2.1 Découplage asymptotique convexe

Pour tout sous-ensemble fini S de $\mathbb{Z}^d$, on notera $\mathcal{C}(Y^S)$ l'ensemble des ouverts convexes de Y^S relativement à la topologie produit. Les éléments de $\mathcal{C}(Y^S)$ sont mesurables relativement à la tribu produit. On entendra par *boîte* un sous-ensemble cubique de $\mathbb{Z}^d$, autrement dit un ensemble de la forme

$$\Lambda(z; n) := z + [0, n]^d \cap \mathbb{Z}^d$$

où $n \in \mathbb{N}^*$. On notera $\Lambda(n) = \Lambda(0; n)$. On notera dist la distance associée à la norme $|\cdot|_\infty$ sur $\mathbb{Z}^d$. Si $\xi \in Y^{\mathbb{Z}^d}$ et $z \in \mathbb{Z}^d$, on notera $\xi(z)$ la coordonnée suivant z de ξ . Plus généralement, pour tout sous-ensemble fini S de $\mathbb{Z}^d$, on notera

$$\xi_S := (\xi(z))_{z \in S}$$

La définition fondamentale suivante est reprise de [Pfi02]. On se donne un champ η à valeurs dans $(Y^{\mathbb{Z}^d}, \mathcal{B}^{\otimes \mathbb{Z}^d})$ et deux applications, g et $c : \mathbb{N}^* \rightarrow [0, +\infty[$, telles que

$$\frac{g(n)}{n} \rightarrow 0 \quad \text{et} \quad \frac{c(n)}{|\Lambda(n)|} \rightarrow 0$$

On dit que η est *asymptotiquement découplé inférieurement pour les convexes* (c.a.d.i.) de paramètre (g, c) (ou encore (g, c) -c.a.d.i.) si, pour tous $z \in \mathbb{Z}^d$ et $n \in \mathbb{N}$, pour tout ensemble fini S de $\mathbb{Z}^d$, pour tous ouverts convexes mesurables $C \in \mathcal{C}(Y^{\Lambda(z;n)})$ et $D \in \mathcal{C}(Y^S)$,

$$\text{dist}(S, \Lambda(z; n)) > g(n) \implies \mathbb{P}(\eta_{\Lambda(z;n)} \in C, \eta_S \in D) \geq e^{-c(n)} \mathbb{P}(\eta_{\Lambda(z;n)} \in C) \mathbb{P}(\eta_S \in D)$$

De plus, si $\mu \in \mathcal{M}_1^+(Y^{\mathbb{Z}^d}, \mathcal{B}^{\otimes \mathbb{Z}^d})$, on dit que μ est c.a.d.i. de paramètre (g, c) si c'est la loi d'un champ η c.a.d.i. de paramètre (g, c) . Si $f : Y \rightarrow Z$, on notera sans ambiguïté

$$f(\eta) := (f(\eta(z)))_{z \in \mathbb{Z}^d}$$

Proposition 6.2.1.1 (Stabilité affine). *Soit η un champ c.a.d.i. de paramètre (g, c) . Si Z est un espace de Fréchet séparable et $f : Y \rightarrow Z$ une application affine, continue et mesurable, alors $f(\eta)$ est c.a.d.i. également, de paramètre (g, c) .*

Démonstration : Il suffit de remarquer que, pour toute boîte Λ , l'application $\phi : \xi_\Lambda \rightarrow (f(\xi(z)))_{z \in \Lambda}$ est continue et mesurable.

6.2.2 Contrôle local

Introduisons aussi une hypothèse pour contrôler chaque site conditionnellement au reste de la configuration. On note $\mathcal{C}_0(Y)$ une base de voisinages convexes ouverts de 0 dans Y . On se donne un champ η à valeurs dans $(Y^{\mathbb{Z}^d}, \mathcal{B}^{\otimes \mathbb{Z}^d})$ et deux applications, $t : \mathcal{C}_0(Y) \rightarrow]0, +\infty[$ et $\alpha : \mathcal{C}_0(Y) \rightarrow]0, 1]$. On dit que η est *contrôlé localement* (c.l.) de paramètre (t, α) (ou encore (t, α) -c.l.) si, pour tout $V \in \mathcal{C}_0(Y)$, pour tout $z \in \mathbb{Z}^d$ et tout sous-ensemble fini S de $\mathbb{Z}^d$, pour tout $D \in \mathcal{C}(Y^S)$,

$$\mathbb{P}(\eta(z) \in t(V) \cdot V; \eta_S \in D) \geq \alpha(V) \mathbb{P}(\eta_S \in D)$$

De plus, si $\mu \in \mathcal{M}_1^+(Y^{\mathbb{Z}^d}, \mathcal{B}^{\otimes \mathbb{Z}^d})$, on dit que μ est c.l. de paramètre (t, α) si c'est la loi d'un champ η c.l. de paramètre (t, α) . Si V est un voisinage convexe de 0, sa jauge (ou fonctionnelle de Minkowski) est l'application $M_V : Y \rightarrow [0, +\infty]$ définie par

$$\forall y \in Y \quad M_V(y) = \inf\{t \geq 0; y \in tV\}$$

Proposition 6.2.2.1 (Stabilité affine). *Soit η un champ c.l. de paramètre (t, α) . Si Z est un espace de Fréchet séparable et $f : Y \rightarrow Z$ une application affine, continue et mesurable, alors $f(\eta)$ est c.l. également. Plus précisément :*

- si $\lambda : Y \rightarrow Z$ est linéaire, $\lambda(\eta)$ est c.l. de paramètre $V \mapsto (t(\lambda^{-1}(V)), \alpha(\lambda^{-1}(V)))$;
- si $y_0 \in Y$, $\eta - y_0$ est c.l. de paramètre $V \mapsto (t(V) + M_V(y_0), \alpha(V))$.

6.2.3 Enoncé du lemme sous-additif

Dans cette partie, on se donne un champ η à valeurs dans $(Y^{\mathbb{Z}^d}, \mathcal{B}^{\otimes \mathbb{Z}^d})$. On note μ sa loi. On suppose qu'il existe un entier ℓ tel que μ soit invariante par les translations du sous-réseau $R_\ell = (\ell\mathbb{Z})^d$ de $\mathbb{Z}^d$. On dira que η (ou μ) est R_ℓ -invariant. On suppose aussi que η est c.a.d.i. et c.l. La notion de champ c.a.d.i. R_ℓ -invariant est une généralisation naturelle de celle de suite de v.a. i.i.d. Rappelons que le théorème de Cramér, pour les moyennes de v.a. i.i.d., est conséquence du lemme sous-additif de Fekete. On montre ici que l'hypothèse de découplage asymptotique, jointe à celle de contrôle local, donne elle aussi une forme de sous-additivité, suffisante pour énoncer un résultat de type Cramér.

Pour toute boîte Λ , on définit l'opérateur moyenne sur Λ , noté $\mathbf{m}_\Lambda$, par

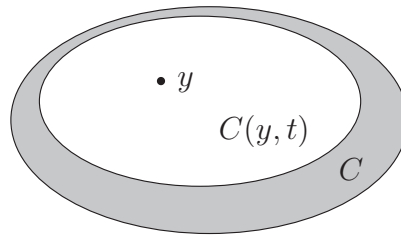
$$\forall \xi \in Y^{\mathbb{Z}^d} \quad \mathbf{m}_\Lambda \xi = \frac{1}{|\Lambda|} \sum_{z \in \Lambda} \xi(z)$$

Remarquons que, pour toute boîte Λ , $\mathbf{m}_\Lambda$ est une application mesurable pour la tribu produit $\mathcal{B}^{\otimes \mathbb{Z}^d}$ (cf. chapitre 3). Donnons-nous un ouvert convexe de la forme

$$C = y + V$$

où $V \in \mathcal{C}_0(Y)$. Pour $\varepsilon \in]0, 1[$, on définit

$$C(y, \varepsilon) = y + (1 - \varepsilon)V$$



On rappelle que, si V est un convexe de Y , sa jauge (ou fonctionnelle de Minkowski) est l'application $M_V : Y \rightarrow [0, +\infty]$ définie par

$$\forall y \in Y \quad M_V(y) = \inf\{t \geq 0; y \in tV\}$$

De plus, si $V \in \mathcal{C}_0(Y)$, alors (cf. 3.2.2.1)

$$V = \{y \in Y; M_V(y) < 1\}$$

Théorème 6.2.3.1 (lemme sous-additif). *Soit η un champ c.a.d.i., c.l. et R_ℓ -invariant. Pour tout $\delta > 0$,*

$$\begin{aligned} \exists M(\varepsilon, \delta) \quad \forall m \geq M \quad \exists N(m, \varepsilon, \delta) \quad \forall n \geq N \\ \frac{1}{|\Lambda(n)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(n)} \eta \in C) \geq \frac{1}{|\Lambda(m)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(m)} \eta \in C(y, \varepsilon)) - \delta \end{aligned}$$

6.2.4 Démonstration du lemme sous-additif

La démonstration suit les étapes de celle de Pfister [Pfi02] en l'adaptant à ce nouveau cadre. Si m et $n \in \mathbb{N}^*$, la division euclidienne de n par $m + g(m) + \ell$ s'écrit

$$n = k(m + g(m) + \ell) + r$$

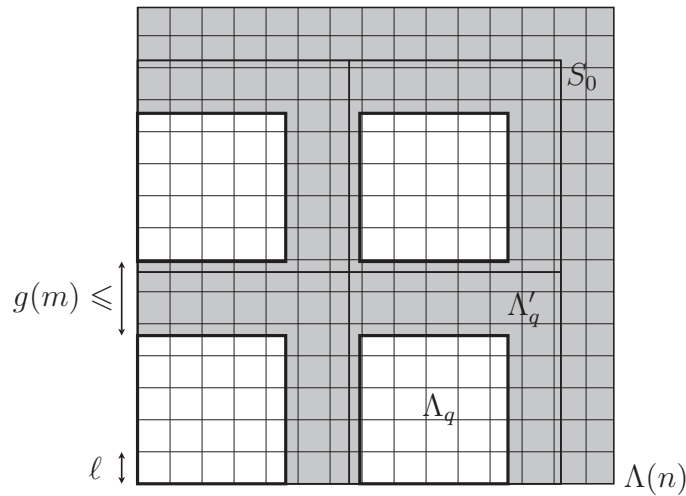
On pave (pas complètement) $\Lambda(n)$ avec k^d boîtes isométriques à $\Lambda(m + g(m) + \ell)$, que l'on note Λ'_q , pour $q \in \{1, \dots, k^d\}$. Puis, pour chaque q , on note Λ_q la boîte isométrique à $\Lambda(m)$ dont le plus petit élément pour l'ordre lexicographique usuel de $\mathbb{Z}^d$ — le “coin en bas à gauche” — est le plus petit élément, pour ce même ordre, de $R_\ell \cap \Lambda'_q$. On notera que la distance entre deux boîtes Λ_q distinctes est supérieure à $g(m)$. Soit enfin

$$S_0 = \Lambda(n) \setminus \bigcup_{q=1}^{k^d} \Lambda_q$$

l'ensemble des sites marginaux. Pour justifier cette dénomination, notons que

$$\rho_{m,n} := \frac{|S_0|}{|\Lambda(n)|} \xrightarrow[m \geq m]{m \rightarrow \infty} 0$$

Les boîtes Λ_q dépendent de n et m , mais les indices seront sous-entendus.



Remarque : On a l'ordre de grandeur suivant :

$$\rho_{m,n} \lesssim d \cdot \left(\frac{g(m) + \ell}{m + g(m) + \ell} + \frac{r}{n} \right)$$

Ainsi, pour que $\rho_{m,n}$ puisse être rendu arbitrairement petit, il faut que le premier terme tende vers 0 avec m , ce qui équivaut à

$$\frac{g(m)}{m} \rightarrow 0$$

L'hypothèse mise sur g est donc utilisée ici.

Notant ϕ la fonction convexe $M_V(\cdot - y)$, on a :

$$\begin{aligned} \phi(\mathbf{m}_{\Lambda(n)}\eta) &\leq \sum_{z \in S_0} \frac{1}{|\Lambda(n)|} \phi(\eta(z)) + \sum_{q=1}^{k^d} \frac{|\Lambda(m)|}{|\Lambda(n)|} \phi(\mathbf{m}_{\Lambda_q}\eta) \\ &\leq \sum_{z \in S_0} \frac{1}{|S_0|} \rho_{m,n} \phi(\eta(z)) + \sum_{q=1}^{k^d} \frac{1}{k^d} \phi(\mathbf{m}_{\Lambda_q}\eta) \end{aligned}$$

D'où

$$\begin{aligned} &\frac{1}{|\Lambda(n)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(n)}\eta \in C) \\ &= \frac{1}{|\Lambda(n)|} \log \mathbb{P}(\phi(\mathbf{m}_{\Lambda(n)}\eta) < 1) \\ &\geq \frac{1}{|\Lambda(n)|} \log \mathbb{P}\left(\forall z \in S_0 \quad \rho_{m,n} \phi(\eta(z)) < \varepsilon; \forall q \in \{1, \dots, k^d\} \quad \phi(\mathbf{m}_{\Lambda_q}\eta) < 1 - \varepsilon\right) \\ &= \frac{1}{|\Lambda(n)|} \log \mathbb{P}\left(\forall z \in S_0 \quad \rho_{m,n} \phi(\eta(z)) < \varepsilon; \forall q \in \{1, \dots, k^d\} \quad \mathbf{m}_{\Lambda_q}\eta \in C(y, \varepsilon)\right) \end{aligned}$$

Tirons profit de l'hypothèse de contrôle local de η : il existe $M(\varepsilon)$ et $N(m, \varepsilon)$ tels que

$$\forall m \geq M(\varepsilon) \quad \forall n \geq N(m, \varepsilon) \quad \frac{\varepsilon}{\rho_{m,n}} > t(V)$$

Dès lors, on obtient par récurrence sur le cardinal de S_0 ,

$$\begin{aligned} &\mathbb{P}\left(\forall z \in S_0 \quad \rho_{m,n} \phi(\eta(z)) < \varepsilon; \forall q \in \{1, \dots, k^d\} \quad \mathbf{m}_{\Lambda_q}\eta \in C(y, \varepsilon)\right) \\ &\geq \alpha(V)^{|S_0|} \mathbb{P}\left(\forall q \in \{1, \dots, k^d\} \quad \mathbf{m}_{\Lambda_q}\eta \in C(y, \varepsilon)\right) \end{aligned}$$

Ainsi, pour $m \geq M(\varepsilon, \delta)$ et $n \geq N(m, \varepsilon, \delta)$

$$\begin{aligned} &\frac{1}{|\Lambda(n)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(n)}\eta \in C) \\ &\geq \rho_{m,n} \log \alpha(V) + \frac{1}{|\Lambda(n)|} \log \mathbb{P}(\forall q \in \{1, \dots, k^d\} \quad \mathbf{m}_{\Lambda_q}\eta \in C(y, \varepsilon)) \end{aligned} \quad (6.1)$$

Reste à exploiter l'hypothèse de découplage asymptotique :

$$\mathbb{P}(\forall q \in \{1, \dots, k^d\} \quad \mathbf{m}_{\Lambda_q} \eta \in C(y, \varepsilon)) \geq (e^{-c(m)})^{k^d-1} \prod_{q=1}^{k^d} \mathbb{P}(\mathbf{m}_{\Lambda_q} \eta \in C(y, \varepsilon))$$

Puis, utilisant le fait que η est R_ℓ -invariant et que $(1 - \rho_{m,n})|\Lambda(n)| = k^d|\Lambda(m)|$, il vient

$$\begin{aligned} \frac{1}{|\Lambda(n)|} \log \mathbb{P}(\forall q \in \{1, \dots, k^d\} \quad \mathbf{m}_{\Lambda_q} \eta \in C(y, \varepsilon)) \\ \geq \frac{1}{|\Lambda(m)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(m)} \eta \in C(y, \varepsilon)) - \frac{c(m)}{|\Lambda(m)|} \end{aligned} \quad (6.2)$$

Quitte à grandir M et N , on peut faire en sorte que

$$\forall m \geq M(\varepsilon, \delta) \quad \forall n \geq N(m, \varepsilon, \delta) \quad \rho_{m,n} \log \alpha(V) - \frac{c(m)}{|\Lambda(m)|} \geq -\delta \quad (6.3)$$

On obtient alors le résultat annoncé en combinant (6.1), (6.2) et (6.3). $\square$

Intuitivement, on a d'abord fait en sorte que l'écartement entre les petites boîtes soit suffisamment petit devant la taille de ces boîtes. Puis, on a pris suffisamment de petites boîtes pour que la proportion occupée par elles dans une grande boîte soit assez grande. Au final, on obtient une sous-additivité asymptotique entre les petites boîtes et la grande.

Remarque : En combinant simplement (6.1) et (6.2), on obtient le résultat plus précis suivant :

$$\begin{aligned} \exists M(\varepsilon) \quad \forall m \geq M \quad \exists N(m, \varepsilon) \quad \forall n \geq N \\ \frac{1}{|\Lambda(n)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(n)} \eta \in C) \geq \frac{1}{|\Lambda(m)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(m)} \eta \in C(y, \varepsilon)) \\ - \frac{c(m)}{|\Lambda(m)|} + \rho_{m,n} \log \alpha(V) \end{aligned} \quad (6.4)$$

Ici apparaissent tous les paramètres. On se souviendra que g est sous-entendu dans $\rho_{m,n}$ et t dans le choix de M et N .

6.3 PGD faible

On pourrait dès maintenant énoncer un principe de grandes déviations pour un champ η c.a.d.i., c.l. et R_ℓ -invariant, mais on se place dans le cadre plus général d'une limite projective. Cela ne complexifie pas la démonstration et s'applique à d'autres cas (notamment, $\sigma(z) = \delta_{\theta_z \omega}$, qui n'est pas c.a.d.i.).

6.3.1 Limite projective

Soient X un espace vectoriel réel et

$$\overleftarrow{\mathcal{X}} = (Y_i, f_i, f_{ij})_{i \leq j}$$

un système projectif d'espaces de Fréchet séparables, autrement dit une famille telle que

(PROJ₁) les indices i et j décrivent un ensemble $(J, \leq)$ préordonné filtrant à droite ;

(PROJ₂) pour tout $i \in J$, Y_i est un espace de Fréchet (*i.e.* un espace vectoriel localement convexe métrisable complet) séparable, f_i une application linéaire de X dans Y_i et, pour tout $(i, j) \in J^2$ tel que $i \leq j$, f_{ij} est une application linéaire continue de Y_j dans Y_i ;

(PROJ₃) pour tout $(i, j, k) \in J^3$ tel que $i \leq j \leq k$, on a $f_{ii} = id_{Y_i}$,

$$f_i = f_{ij} \circ f_j \quad \text{et} \quad f_{ik} = f_{ij} \circ f_{jk}$$

On note, pour tout $i \in J$, $\mathcal{C}_0(Y_i)$ un système fondamental de voisinages convexes ouverts de 0 dans Y_i et on définit

$$\mathcal{C}_0(X) = \{f_i^{-1}(t_i C_i) ; i \in J, t_i \in]0, +\infty[, C_i \in \mathcal{C}_0(Y_i)\}$$

On munit alors X de la tribu $\mathcal{F}$ (resp. de la topologie τ) engendrée par les translatés des éléments de $\mathcal{C}_0(X)$. On dit que X est la *limite projective* du système projectif $\overleftarrow{\mathcal{X}}$.

Remarque : La famille $\mathcal{C}_0$ satisfait aux trois axiomes (EVLCM₁), (EVLCM₂) et (EVLCM₃) du chapitre 3. On vérifie que, si $(X, \mathcal{C}_0, \mathcal{F}, \tau)$ est l'e.v.l.c.m. associé à $\mathcal{C}_0$, alors τ (resp. $\mathcal{F}$) est la topologie (resp. tribu) initiale pour la famille $(F_i, f_i)_{i \in J}$.

Exemples :

- si X est déjà un e.v.l.c.m et si $\mathcal{Y}$ est engendré par $f_0 = Id$ ($Y_0 = X$), on récupère la topologie sur X ;
- si X est un e.v.l.c.m et si $\mathcal{Y}$ est engendré par X' ($Y_i = \mathbb{R}$ pour $i \in X'$), on tombe sur la topologie $\sigma(X, X')$;
- pour X' dual d'un e.v.l.c.m, si $\mathcal{Y}$ est engendrée par X ($Y_i = \mathbb{R}$ pour $i \in X$), on a la topologie $\sigma(X', X)$.

6.3.2 PGD faible

Le décor est maintenant planté. On peut énoncer¹ :

¹ En termes de limite projective, on peut penser exploiter le théorème de Dawson-Gärtner (cf. [DZ93]) : on obtient alors un PGD pour σ , à condition que les s_η soient de bonnes fonctions de taux, ce qui est moins bon que le résultat obtenu. Il n'est pas étonnant qu'on n'obtienne pas ainsi un résultat optimal étant donné que le théorème de Dawson-Gärtner ne se sert pas de la linéarité des f_i .

Proposition 6.3.2.1 (PGD). *Soit σ un champ à valeurs dans $(X^{\mathbb{Z}^d}, \mathcal{F}^{\otimes \mathbb{Z}^d})$. On suppose que, pour tout $i \in J$, le champ $f_i(\sigma)$ est c.a.d.i., c.l. et R_ℓ -invariant. Alors*

$$\left(\frac{1}{|\Lambda(n)|} \sum_{z \in \Lambda(n)} \sigma(z) \right)_{n \geq 0}$$

vérifie un PGD faible d'entropie

$$s(x) = \inf_{\substack{A \in \mathcal{C} \\ A \ni x}} \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mathbb{P}(m_{\Lambda(n)} \sigma \in A)$$

Démonstration : Donnons-nous $A \in \mathcal{C}$; il est de la forme

$$A = x + f_i^{-1}(V)$$

avec $V \in \mathcal{C}_0(Y_i)$. Notons $y = f_i(x)$ et définissons

$$C = y + V$$

puis, pour $\varepsilon \in]0, 1[$,

$$A(x, \varepsilon) = x + (1 - \varepsilon)U \quad \text{et} \quad C(y, \varepsilon) = y + (1 - \varepsilon)V$$

La famille $\eta = f_i(\sigma)$ vérifie les hypothèses du lemme sous-additif 6.2.3.1 : on en déduit que, pour tout $\delta > 0$,

$$\begin{aligned} \exists M(\varepsilon, \delta) \quad \forall m \geq M \quad \exists N(m, \varepsilon, \delta) \quad \forall n \geq N \\ \frac{1}{|\Lambda(n)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(n)} \eta \in C) \geq \frac{1}{|\Lambda(m)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(m)} \eta \in C(y, \varepsilon)) - \delta \end{aligned}$$

Puis, en faisant tendre n , puis m , vers l'infini, et enfin δ vers 0^+ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(n)} \eta \in C) \geq \limsup_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(m)} \eta \in C(y, \varepsilon))$$

Pour toute boîte finie Λ , la linéarité de f_i assure

$$\{\mathbf{m}_\Lambda \sigma \in A\} = \{f_i(\mathbf{m}_\Lambda \sigma) \in C\} = \{\mathbf{m}_\Lambda \eta \in C\}$$

et les égalités analogues avec $A(x, \varepsilon)$ et $C(y, \varepsilon)$. On en déduit

$$\underline{s}(A) \geq \overline{s}(A(x, \varepsilon))$$

On conclut alors grâce au lemme 4.3.4.2. □

Proposition 6.3.2.2. *L'entropie s est s.c.s. et concave.*

Démonstration : Elle se fait comme dans le cas indépendant. La semi-continuité est immédiate : si $u \in \mathbb{R}$ et $x \in \{s < u\}$, il existe $A \in \mathcal{C}$ contenant x tel que $\underline{s}(A) < u$. Alors, pour tout $y \in A$, $s(y) \leq \underline{s}(A) < u$, donc $x \in A \subset \{s < u\}$ et $\{s < u\}$ est ouvert. Pour la concavité, il suffit alors de prouver que

$$s\left(\frac{x+y}{2}\right) \geq \frac{s(x) + s(y)}{2}$$

La semi-continuité permet alors de passer des inégalités

$$s((1-u)x + uy) \geq (1-u)s(x) + us(y)$$

pour u dyadique à celles pour $u \in [0, 1]$. La preuve est un raffinement de la démonstration du lemme sous-additif (6.2.3.1). Soit $A \in \mathcal{C}$ contenant $\frac{1}{2}(x+y)$. Par continuité des applications d'espace vectoriel, pour tout $\varepsilon \in]0, 1[$, il existe A_x et $A_y \in \mathcal{C}$, contenant respectivement x et y , tels que

$$\frac{1}{2}A_x + \frac{1}{2}A_y \subset A\left(\frac{1}{2}(x+y), \varepsilon\right)$$

Puis

$$\begin{aligned} & \frac{1}{|\Lambda(n)|} \log \mathbb{P}(\mathbf{m}_{\Lambda(n)}\sigma \in A) \\ & \geq \frac{1}{|\Lambda(n)|} \log \mathbb{P}\left(\forall z \in \Lambda_0 \quad \rho_{m,n}\phi(\eta(z)) < \varepsilon ; \right. \\ & \quad \left. \bigcap_{q \text{ pair}} \{\mathbf{m}_{\Lambda_q}\sigma \in A_x\} \cap \bigcap_{q \text{ impair}} \{\mathbf{m}_{\Lambda_q}\sigma \in A_y\}\right) \end{aligned}$$

En poursuivant comme précédemment, on obtient

$$\underline{s}(A) \geq \frac{1}{2}(\underline{s}(A_x) + \underline{s}(A_y)) \geq \frac{1}{2}(s(x) + s(y))$$

puis l'inégalité désirée en passant à l'infimum sur les $A \ni \frac{1}{2}(x+y)$.

6.4 Pression

Soient $\overleftarrow{\mathcal{X}} = (Y_i, f_i, f_{ij})_{i \leq j}$ un système projectif d'espaces de Fréchet séparables et X sa limite projective. Soit σ un champ à valeurs dans $(X^{\mathbb{Z}^d}, \mathcal{F}^{\otimes \mathbb{Z}^d})$ tel que, pour tout $i \in J$, le champ $f_i(\sigma)$ soit c.a.d.i., c.l. et R_ℓ -invariant. On définit, pour toute boîte $\Lambda \subset \mathbb{Z}^d$,

$$\forall \lambda \in X^* \quad p_\Lambda(\lambda) = \frac{1}{|\Lambda|} \log \mathbb{E} \left[\exp \left\langle \lambda \middle| \sum_{z \in \Lambda} \sigma(z) \right\rangle \right]$$

Notons X^* le dual topologique de X . Soit $\lambda \in X^*$. Il existe un voisinage convexe de 0 sur lequel $|\lambda| < 1$, voisinage de la forme $f_i^{-1}(C_i)$. Par linéarité, on remarque que

$$\forall x, x' \in X \quad f_i(x) = f_i(x') \Rightarrow \lambda(x) = \lambda(x')$$

Ainsi, on peut définir la section

$$\lambda_i \in Y_i^* : y \mapsto \lambda(f_i^{-1}(y))$$

de sorte que $\lambda = \lambda_i \circ f_i$. Donc

$$X^* = \{\lambda_i \circ f_i; i \in I, \lambda_i \in Y_i^*\}$$

En particulier, pour tout $\lambda \in X^*$, écrivant λ sous la forme $\lambda_i \circ f_i$, on voit que le champ $\eta = \langle \lambda | \sigma \rangle = \langle \lambda_i | f_i(\sigma) \rangle$ est c.a.d.i., c.l. (cf. 6.2.1.1 et 6.2.2.1) et R_ℓ -invariant.

6.4.1 Pression en volume infini

On veut maintenant définir la pression p en volume infini. Tout se passe à merveille ainsi que le montre la proposition suivante :

Proposition 6.4.1.1. *Pour tout $\lambda \in X^*$, la suite $(p_{\Lambda(n)}(\lambda))_{n \in \mathbb{N}}$ converge dans $[-\infty, +\infty]$. On note $p(\lambda)$ sa limite.*

Démonstration : La preuve est analogue à celle du lemme sous-additif 6.2.3.1. On définit $\eta(z) = \langle \lambda | \sigma(z) \rangle$; on note (g, c) (resp. (t, α)) le paramètre de découplage (resp. de contrôle local) de η . On écrit alors

$$p_{\Lambda(n)}(\lambda) = \frac{1}{|\Lambda(n)|} \log \mathbb{E}[P_{m,n} \cdot R_{m,n}]$$

avec un terme principal

$$P_{m,n} = \prod_{q=1}^{k^d} \exp \left(\sum_{z \in \Lambda_q} \eta(z) \right)$$

et un terme résiduel

$$R_{m,n} = \mathbb{E} \left[\exp \left(\sum_{z \in S_0} \eta(z) \right) \middle| \prod_{q=1}^{k^d} \exp \left(\sum_{z \in \Lambda_q} \eta(z) \right) \right]$$

Pour qu'il y ait convergence, il suffit de s'assurer que
– l'on a une minoration du terme résiduel de la forme

$$R_{m,n} \geq e^{-\delta |\Lambda(n)|}$$

pour m et n assez grands.

– l’hypothèse de découplage s’applique bien pour traiter le terme principal.

Terme résiduel : Encore une fois, on va utiliser le contrôle local.

Lemme 6.4.1.2. *Il existe $\beta > 0$ tel que, pour toute boîte finie $\Lambda \subset \mathbb{Z}^d \setminus \{0\}$,*

$$\mathbb{E}[e^{\eta(0)} | \eta_\Lambda] \geq \beta$$

Démonstration du lemme : On note que, pour tous Λ finie et $t > 0$,

$$\begin{aligned} \mathbb{E}[e^{\eta(0)} | \eta_\Lambda] &\geq \mathbb{E}[e^{\eta(0)} 1_{|\eta(0)| < t} | \eta_\Lambda] \\ &\geq e^{-t} \mathbb{P}(|\eta(0)| < t | \eta_\Lambda) \end{aligned}$$

En utilisant l’hypothèse de contrôle local pour η et $V_1 =]-1, 1[$, on note $t = t(V_1)$ et $\alpha = \alpha(V_1)$, de sorte que

$$\begin{aligned} \mathbb{P}_{\eta_\Lambda} - p.s. \quad \mathbb{P}(|\eta(0)| < t | \eta_\Lambda = y) &= \lim_{r \rightarrow 0^+} \mathbb{P}(|\eta(0)| < t | \eta_\Lambda \in B(y, r)) \\ &\geq \alpha \end{aligned}$$

L’égalité de la première ligne est due au fait que l’ensemble de Lebesgue d’une fonction de $L^1(\mathbb{R}^\Lambda; \mathbb{R})$ est de mesure pleine (cf. par exemple [Rud66, théorème 8.8]). Ainsi, $\beta = e^{-t}\alpha > 0$ convient. $\square$

On déduit du lemme, par récurrence, que

$$R_{m,n} \geq \beta^{|S_0|} \geq e^{-\delta|\Lambda(n)|}$$

pour $m \geq M(\delta)$ et $n \geq N(m, \delta)$.

Terme principal : On a

$$\left\{ \exp \sum_{z \in \Lambda_q} \eta(z) > t \right\} \in \mathcal{C}(\mathbb{R}^{|\Lambda_q|})$$

(il s'agit en fait d'un demi-espace ouvert). Ainsi,

$$\begin{aligned}
& \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left[\prod_{q=1}^{k^d} \exp \left(\sum_{z \in \Lambda_q} \eta(z) \right) \right] \\
&= \frac{1}{|\Lambda(n)|} \log \mathbb{E} \left[\prod_{q=1}^{k^d} \int_{t_q \geq 0} 1_{\exp \left(\sum_{z \in \Lambda_q} \eta(z) \right) > t_q} \right] \\
&= \frac{1}{|\Lambda(n)|} \log \int_{t_1, \dots, t_{k^d} \geq 0} \mathbb{E} \left[\prod_{q=1}^{k^d} 1_{\exp \left(\sum_{z \in \Lambda_q} \eta(z) \right) > t_q} \right] \\
&\geq \frac{1}{|\Lambda(n)|} \log \int_{t_1, \dots, t_{k^d} \geq 0} (e^{-c(m)})^{k^d-1} \prod_{q=1}^{k^d} \mathbb{E} \left[1_{\exp \left(\sum_{z \in \Lambda_q} \eta(z) \right) > t_q} \right] \\
&\geq \frac{k^d}{|\Lambda(n)|} \log \mathbb{E} \left[\exp \left(\sum_{z \in \Lambda(m)} \eta(z) \right) \right] - \frac{c(m)}{|\Lambda(m)|}
\end{aligned}$$

On obtient finalement que, quitte à grandir M et N (pour absorber le terme $c(m)/|\Lambda(m)|$),

$$\begin{aligned}
\forall m \geq M(\delta) \quad \forall n \geq N(m, \delta) \\
p_{\Lambda(n)}(\lambda) &\geq \frac{1}{|\Lambda(n)|} \log \mathbb{E} [P_{m,n} \cdot e^{-\delta|\Lambda(n)|}] \\
&\geq (1 - \rho_{m,n}) p_{\Lambda(m)}(\lambda) - \delta
\end{aligned} \tag{6.5}$$

D'où, en prenant la limite inférieure en n , puis la limite supérieure en m et, enfin, la limite quand $\delta \rightarrow 0^+$,

$$\liminf_{n \rightarrow \infty} p_{\Lambda(n)}(\lambda) \geq \limsup_{m \rightarrow \infty} p_{\Lambda(m)}(\lambda)$$

autrement dit $p_{\Lambda(n)}(\lambda)$ converge. □

Remarque : Voici, comme pour s , la version de l'inégalité sous-additive avec tous les paramètres :

$$\begin{aligned}
\forall n \geq N(m) \\
p_{\Lambda(n)}(\lambda) &\geq (1 - \rho_{m,n}) p_{\Lambda(m)}(\lambda) - \frac{c(m)}{|\Lambda(m)|} - \rho_{m,n} (t(V_1) - \log \alpha(V_1))
\end{aligned} \tag{6.6}$$

On se souviendra que les paramètres intervenant ici sont ceux de $\langle \lambda | \sigma \rangle$, et dépendent donc de λ .

Proposition 6.4.1.3 (Convexité de p). *La pression p est convexe.*

Démonstration : L'inégalité de Hölder assure que les p_Λ sont convexes. Puis, sachant qu'une limite supérieure de fonctions convexes est convexe, p l'est également.

6.5 Pression et entropie

6.5.1 Inégalité de Tchebychev

L'inégalité de Tchebychev exponentielle (cf. [Cer07, lemme 12.6]) prend ici la forme suivante :

Proposition 6.5.1.1 (Inégalité de Tchebycheff exponentielle). *Pour toute boîte $\Lambda \subset \mathbb{Z}^d$ et $A \in \mathcal{B}$,*

$$\mathbb{P}(\mathbf{m}_\Lambda \sigma \in A) \leq \exp \left[-|\Lambda| \sup_{\lambda \in X^*} \left(\inf_{x \in A} \langle \lambda | x \rangle - p_\Lambda(\lambda) \right) \right]$$

On en déduit, en considérant $\Lambda = \Lambda(n)$ et en faisant tendre n vers ∞ :

$$\underline{s}(A) \leq \liminf_{\lambda \in X^*} \inf_{\lambda \in X^*} \left(p_{\Lambda(n)}(\lambda) - \inf_{a \in A} \langle \lambda | a \rangle \right) \leq \inf_{\lambda \in X^*} \left(p(\lambda) - \inf_{a \in A} \langle \lambda | a \rangle \right)$$

Puis,

$$\begin{aligned} s(x) &= \inf_{\substack{A \in \mathcal{C} \\ A \ni x}} \underline{s}(A) \\ &\leq \inf_{\substack{A \in \mathcal{C}_w \\ A \ni x}} \underline{s}(A) = \underline{s}_m(x) \\ &\leq \inf_{\substack{A \in \mathcal{C}_w \\ A \ni x}} \inf_{\lambda \in X^*} \left(p(\lambda) - \inf_{a \in A} \langle \lambda | a \rangle \right) \\ &= \inf_{\lambda \in X^*} \inf_{\substack{A \in \mathcal{C}_w \\ A \ni x}} \left(p(\lambda) - \inf_{a \in A} \langle \lambda | a \rangle \right) \\ &= - \sup_{\lambda \in X^*} \left(\langle \lambda | x \rangle - p(\lambda) \right) \\ &= -p^*(x) \end{aligned}$$

Résumons :

Théorème 6.5.1.2. *On a toujours l'inégalité $s \leq -p^*$.*

On voudrait maintenant savoir dans quels cas l'égalité est vérifiée.

6.5.2 Égalité $s = -p^*$: cas borné

Théorème 6.5.2.1. *Si $\sigma(0)$ est à valeurs dans un ensemble faiblement borné, on a l'égalité*

$$s(\sigma; x) = -p^*(\sigma; x)$$

Démonstration : L'égalité $p = (-s)^*$ est conséquence du lemme de Varadhan fermé 4.7.1.1 en remarquant que

$$p_{\Lambda(n)} = \frac{1}{|\Lambda(n)|} \log \int e^{|\Lambda(n)| \langle \lambda | x \rangle} d\nu_n(x)$$

où ν_n est la loi de $m_{\Lambda(n)}\sigma$. L'égalité duale découle alors de 4.6.3.1, sachant que s est concave et semi-continue supérieurement. $\square$

6.5.3 Égalité $s = -p^*$: cas c.a.d.i. et c.l.

Théorème 6.5.3.1. *Soit (Y, τ) un espace de Fréchet séparable et $\mathcal{B}$ sa tribu borélienne. Soit η un champ à valeurs dans $(Y^{\mathbb{Z}^d}, \mathcal{B}^{\otimes \mathbb{Z}^d})$ c.a.d.i., c.l. et R_ℓ -invariant. Alors,*

$$s(\eta; y) = -p^*(\eta; y)$$

Démonstration : L'idée de la démonstration est de se ramener au théorème 6.5.2.1 en conditionnant les sites de η à être dans des compacts. Comme il y a une infinité de sites, on est obligé de prendre des précautions pour ne pas manipuler des conditionnements dégénérés. C'est l'objet de la définition suivante :

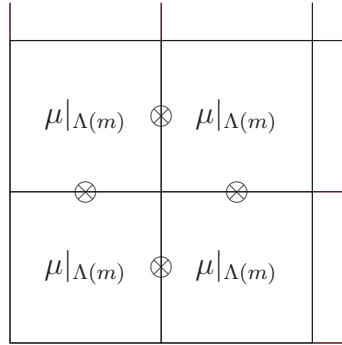
Définition 6.5.3.2. *Si μ est une loi sur $Y^{\mathbb{Z}^d}$ et $m \in \mathbb{N}$, on note $\mu_{\Lambda(m)}$ la loi sur $Y^{\mathbb{Z}^d}$ définie par*

$$\mu_{\Lambda(m)} := (\mu|_{\Lambda(m)})^{\otimes \mathbb{Z}^d}$$

où $\mu|_{\Lambda(m)}$ désigne la marginale de μ sur $Y^{\Lambda(m)}$. On définit aussi, pour K compact de Y , $\mu_{\Lambda(m)}^K$ la loi sur $Y^{\mathbb{Z}^d}$ donnée par

$$\mu_{\Lambda(m)}^K := \left(\mu|_{\Lambda(m)}(\cdot \mid \forall z \in \Lambda(m) \quad \xi(z) \in K) \right)^{\otimes \mathbb{Z}^d}$$

(quand le conditionnement a un sens).



Remarque : Comme nous utiliserons ces notations, nous préférons l'usage de μ (resp. ξ) à celui de $\mathbb{P}$ (resp. η).

Soit μ la loi de η , (g, c) et (t, α) ses paramètres respectifs de découplage et de contrôle local. On reprend les notations du lemme sous-additif (cf. p. 132) :

$$C = y + V \quad \text{et} \quad C(y, \varepsilon) = y + (1 - \varepsilon)V$$

où $V \in \mathcal{C}_0(Y)$ et $\varepsilon \in]0, 1[$. On note aussi $\varphi = M_V(\cdot - y)$. La démonstration sera faite en deux étapes :

Étape 1 : Si $(K_m)_{m \in \mathbb{N}}$ est une suite de compacts vérifiant

$$\lim_{m \rightarrow \infty} \mu(K_m^{\Lambda(m+g(m)+\ell)}) = 1 \tag{6.7}$$

alors

$$\forall y \in Y \quad \forall \varepsilon \in]0, 1[$$

$$s(\mu; y) \geq \inf_{\substack{C \in \mathcal{C}(Y) \\ C \ni y}} \limsup_{m \rightarrow \infty} \sup_{x \in C(y, \varepsilon)} \left[-p^*(\mu_{\Lambda(m+g(m)+\ell)}^{K_m}; x) \right]$$

Étape 2 : Grâce au théorème de Mosco, on exhibe une suite de compacts $(K_m)_{m \in \mathbb{N}}$ vérifiant (6.7) et telle que

$$\liminf_{m \rightarrow \infty} \sup_{x \in C(y, \varepsilon)} \left[-p^*(\mu_{\Lambda(m+g(m)+\ell)}^{K_m}; x) \right] \geq \sup_{x \in C(y, \varepsilon)} \left[-p^*(\mu; x) \right]$$

On conclut alors ainsi :

$$s(\mu; y) \geq \inf_{\substack{C \in \mathcal{C}(Y) \\ C \ni y}} \sup_{x \in C(y, \varepsilon)} \left[-p^*(\mu; x) \right] \geq -p^*(\mu; y)$$

Étape 1 : Par un raisonnement analogue à la démonstration du théorème 6.2.3.1, on montre que

$$\begin{aligned} & \frac{1}{|\Lambda(n)|} \log \mu(\mathbf{m}_{\Lambda(n)} \xi \in C) \\ & \geq \frac{1}{|\Lambda(n)|} \log \mu \left((1 - \rho_{m,n}) \varphi(\mathbf{m}_{\cup \Lambda_q} \xi) < 1 - \frac{\varepsilon}{2} \right) + \rho_{m,n} \log \alpha(V) \\ & \geq \frac{1}{|\Lambda(n)|} \log \mu_{\Lambda(m+g(m)+\ell)} \left((1 - \rho_{m,n}) \varphi(\mathbf{m}_{\cup \Lambda_q} \xi) < 1 - \frac{\varepsilon}{2} \right) \\ & \quad - \frac{c(m)}{\Lambda(m)} + \rho_{m,n} \log \alpha(V) \end{aligned}$$

En effet, utilisant le fait que Y est à base dénombrable d'ouverts et adaptant les idées de la démonstration de [Cer07, lemme 11.2.], on peut écrire l'ensemble

$$\left\{ (y_1, \dots, y_{k^d}) \in Y^{k^d}; \frac{y_1 + \dots + y_{k^d}}{k^d} \in (1 - \frac{\varepsilon}{2})V \right\}$$

comme union dénombrable de produits d'ouverts convexes

$$\bigsqcup_{i \in \mathbb{N}} V_1^i \times \dots \times V_{k^d}^i$$

et on en déduit :

$$\begin{aligned}
& \mu\left((1-\rho_{m,n})\varphi(\mathbf{m}_{\bigcup \Lambda_q}\xi) < 1 - \frac{\varepsilon}{2}\right) \\
&= \mu\left((1-\rho_{m,n})(\mathbf{m}_{\bigcup \Lambda_q}\xi - y) \in (1 - \frac{\varepsilon}{2})V\right) \\
&= \mu\left(\bigcup_{i \in \mathbb{N}} \prod_{q=1}^{k^d} \{(1-\rho_{m,n})(\mathbf{m}_{\Lambda_q}\xi - y) \in V_q^i\}\right) \\
&\geq e^{-(k^d-1)c(m)} \mu_{\Lambda(m+g(m)+\ell)}\left(\bigcup_{i \in \mathbb{N}} \prod_{q=1}^{k^d} \{(1-\rho_{m,n})(\mathbf{m}_{\Lambda_q}\xi - y) \in V_q^i\}\right) \\
&= e^{-(k^d-1)c(m)} \mu_{\Lambda(m+g(m)+\ell)}\left((1-\rho_{m,n})(\mathbf{m}_{\bigcup \Lambda_q}\xi - y) \in (1 - \frac{\varepsilon}{2})V\right) \\
&= e^{-(k^d-1)c(m)} \mu_{\Lambda(m+g(m)+\ell)}\left((1-\rho_{m,n})\varphi(\mathbf{m}_{\bigcup \Lambda_q}\xi) < 1 - \frac{\varepsilon}{2}\right)
\end{aligned}$$

Enfin, soit $(K_m)_{m \in \mathbb{N}}$ une suite de compacts mesurables de Y . On a

$$\begin{aligned}
& \frac{1}{|\Lambda(n)|} \log \mu_{\Lambda(m+g(m)+\ell)}\left((1-\rho_{m,n})\varphi(\mathbf{m}_{\bigcup \Lambda_q}\xi) < 1 - \frac{\varepsilon}{2}\right) \\
&\geq \frac{1}{|\Lambda(n)|} \log \mu_{\Lambda(m+g(m)+\ell)}^{K_m}\left(\mathbf{m}_{\Lambda(n)}\xi \in C(y, \varepsilon); \rho_{m,n}\varphi(-\mathbf{m}_{\Lambda_0}\xi) < \frac{\varepsilon}{2}\right) \\
&\quad + \frac{1}{|\Lambda(n)|} \log \mu_{\Lambda(m+g(m)+\ell)}(\forall z \in \Lambda(n) \quad \xi(z) \in K_m)
\end{aligned}$$

Comme $-y - K_m$ est absorbé par V ,

$$\begin{aligned}
& \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_{\Lambda(m+g(m)+\ell)}^{K_m}\left(\mathbf{m}_{\Lambda(n)}\xi \in C(y, \varepsilon); \rho_{m,n}\varphi(-\mathbf{m}_{\Lambda_0}\xi) < \frac{\varepsilon}{2}\right) \\
&= \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_{\Lambda(m+g(m)+\ell)}^{K_m}\left(\mathbf{m}_{\Lambda(n)}\xi \in C(y, \varepsilon)\right)
\end{aligned}$$

D'autre part,

$$\begin{aligned}
& \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_{\Lambda(m+g(m)+\ell)}(\forall z \in \Lambda(n) \quad \xi(z) \in K_m) \\
&\geq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \mu_{\Lambda(m+g(m)+\ell)}(\forall z \in \Lambda((k+1) \cdot (m+g(m)+\ell)) \quad \xi(z) \in K_m) \\
&\geq \frac{1}{|\Lambda(m+g(m)+\ell)|} \log \mu(\forall z \in \Lambda(m+g(m)+\ell) \quad \xi(z) \in K_m)
\end{aligned}$$

Pour tout $z \in \mathbb{Z}^d$, la loi ν de $\eta(z)$ est tendue, donc il existe une suite croissante $(K_m)_{m \geq 1}$ de compacts de Y telle que, pour tout $m \geq 1$,

$$\nu(K_m) \geq 1 - \frac{1}{m|\Lambda(m+g(m)+\ell)|}$$

Alors,

$$\lim_{m \rightarrow \infty} \mu(K_m^{\Lambda(m+g(m)+\ell)}) = 1$$

Introduisons la notation

$$\mu_m := \mu_{\Lambda(m+g(m)+\ell)}^{K_m}$$

On obtient finalement, en faisant tendre n , puis m , vers l'infini :

$$\begin{aligned} \forall \varepsilon > 0 \quad \underline{s}(\mu; C) &\geq \limsup_{m \rightarrow \infty} \underline{s}(\mu_m; C(y, \varepsilon)) \\ &\geq \limsup_{m \rightarrow \infty} \sup_{x \in C(y, \varepsilon)} s(\mu_m; x) \\ &= \limsup_{m \rightarrow \infty} \sup_{x \in C(y, \varepsilon)} [-p^*(\mu_m; x)] \end{aligned}$$

L'égalité finale est conséquence du théorème 6.5.2.1. On conclut cette première étape en passant à l'infimum sur $C \ni y$. $\square$

Étape 2 : Montrons maintenant que

$$\limsup_{m \rightarrow \infty} \inf_{x \in C(y, \varepsilon)} p^*(\mu_m; x) \leq \inf_{x \in C(y, \varepsilon)} p^*(\mu; x) \quad (6.8)$$

Nous allons voir que cette inégalité est conséquence du théorème suivant, dont on trouvera une démonstration dans [Zab92b].

Théorème 6.5.3.3 (Mosco). *Soit Y un e.v.l.c. séparable et métrisable. On se donne $f : [-\infty, +\infty] \rightarrow Y^*$ et $(f_m)_{m \in \mathbb{N}} \in [-\infty, +\infty]^{Y^*}$ une suite de fonctions convexes s.c.i., uniformément propre. On suppose que $f_m \xrightarrow[m \rightarrow \infty]{M_2} f$, i.e.*

$$\forall \lambda \in Y^* \quad \forall \lambda_m \xrightarrow[m \rightarrow \infty]{\sigma(Y^*, Y)} \lambda \quad \liminf_{m \rightarrow \infty} f_m(\lambda_m) \geq f(\lambda)$$

Alors $f_m^* \xrightarrow[m \rightarrow \infty]{M_1} f^*$, i.e.

$$\forall y \in Y \quad \exists y_m \xrightarrow[m \rightarrow \infty]{\tau(Y^*, Y)} y \quad \limsup_{m \rightarrow \infty} f_m^*(y_m) \leq f^*(y)$$

Conclusion de l'étape 2 (en admettant que $f = p(\mu; \cdot)$ et $f_m = p(\mu_m; \cdot)$ vérifient les hypothèses du théorème 6.5.3.3) : Soit $x_0 \in C(y, \varepsilon)$. Il existe une suite $y_m \rightarrow x_0$ telle que

$$\limsup_{m \rightarrow \infty} p^*(\mu_m; y_m) \leq p^*(\mu; x_0)$$

Or, pour tout m assez grand, $y_m \in C(y, \varepsilon)$, donc

$$\exists m_0 \quad \forall m \geq m_0 \quad \inf_{x \in C(y, \varepsilon)} p^*(\mu_m; x) \leq p^*(\mu_m; y_m)$$

D'où

$$\limsup_{m \rightarrow \infty} \inf_{x \in C(y, \varepsilon)} p^*(\mu_m; x) \leq \limsup_{m \rightarrow \infty} p^*(\mu_m; y_m) \leq p^*(\mu; x_0)$$

On obtient (6.8) en passant à l'infimum sur $x_0 \in C(y, \varepsilon)$. $\square$

Vérifions donc que $f = p(\mu; \cdot)$ et $f_m = p(\mu_m; \cdot)$ satisfont les hypothèses du théorème de Mosco 6.5.3.3. Les fonctions f_m sont convexes d'après la proposition 6.4.1.3. Pour voir qu'elles sont s.c.i. (pour la topologie faible, par exemple, ou la topologie sur X ; cela ne change rien puisqu'elles sont convexes), il suffit de constater (cf. la proposition 6.5.3.4 ci-dessous) que

$$p(\mu_{\Lambda(m+g(m)+\ell)}^{K_m}; \lambda) = p_{\Lambda(m+g(m)+\ell)}(\mu_{\Lambda(m+g(m)+\ell)}^{K_m}; \lambda)$$

ce qui permet de se ramener à [Cer07, lemme 12.1.].

Proposition 6.5.3.4. *Pour tout $j \in \mathbb{N}$, on a :*

$$p(\mu_{\Lambda(j)}; \lambda) = p_{\Lambda(j)}(\mu_{\Lambda(j)}; \lambda)$$

Démonstration : En effet, on peut écrire :

$$\begin{aligned} p(\mu_{\Lambda(j)}; \lambda) &= \lim_{n \rightarrow \infty} p_{\Lambda(n)}(\mu_{\Lambda(j)}; \lambda) \\ &= \lim_{n=kj \rightarrow \infty} p_{\Lambda(n)}(\mu_{\Lambda(j)}; \lambda) \\ &= \lim_{n=kj \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \int \exp \left\langle \lambda \left| \sum_{z \in \Lambda(n)} \xi(z) \right. \right\rangle d\mu_{\Lambda(j)}(\xi) \\ &= \lim_{n=kj \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \prod_{q=1}^{k^d} \int \exp \left\langle \lambda \left| \sum_{z \in \Lambda'_q} \xi(z) \right. \right\rangle d\mu_{\Lambda(j)}(\xi) \\ &= p_{\Lambda(j)}(\mu_{\Lambda(j)}; \lambda) \end{aligned}$$

où les Λ'_q sont les boîtes translatées de $\Lambda(j)$ qui pavent naturellement $\Lambda(kj) = \Lambda(n)$. $\square$

Définition 6.5.3.5. *Soit $(f_m)_{m \in \mathbb{N}}$ une famille de fonctions convexes propres (i.e. pour tout m , $f_m > -\infty$ et il existe $\lambda_m \in Y^*$ tel que $f_m(\lambda_m) < \infty$). On dit que $(f_m)_{m \in \mathbb{N}}$ est uniformément propre s'il existe une suite $(\lambda_m)_{m \in \mathbb{N}}$ $\sigma(Y^*, Y)$ -relativement compacte telle que*

$$\sup_{m \in \mathbb{N}} f_m(\lambda_m) < \infty$$

Il est alors immédiat que la famille de fonctions $(p(\mu_m; \cdot))_{m \in \mathbb{N}}$ est uniformément propre, car leur valeur en 0 est 0 ($\lambda_m = 0$ convient). Reste à montrer la convergence (M_2) . Pour ce faire, soient $\lambda \in Y^*$ et $(\lambda_m)_{m \in \mathbb{N}}$ tels que λ_m converge vers λ pour la topologie $\sigma(Y^*, Y)$. On procède en deux sous-étapes :

Sous-étape 1 : Grâce au théorème de Banach-Steinhaus, on montre que

$$\liminf_{j \rightarrow \infty} \liminf_{m \rightarrow \infty} p_{\Lambda(j)}(\mu_m; \lambda_m) \geq p(\mu; \lambda)$$

Sous-étape 2 : On inverse ensuite les limites pour obtenir

$$\liminf_{m \rightarrow \infty} p(\mu_{\Lambda(m+g(m)+\ell)}^{K_m}; \lambda_m) \geq \limsup_{j \rightarrow \infty} \liminf_{m \rightarrow \infty} p_{\Lambda(j)}(\mu_{\Lambda(m+g(m)+\ell)}^{K_m}; \lambda_m)$$

Cela est vrai pour une certaine suite $(K_m)_{m \in \mathbb{N}}$ vérifiant bien la condition (6.7). On utilise ici un argument d'uniformité dans la sous-additivité définissant les $p(\mu_{\Lambda(m+g(m)+\ell)}; \lambda_m)$.

Sous-étape 1 : Si $n \in \mathbb{N}$, le théorème de Banach-Steinhaus montre que les fonctions

$$x \in K_n \mapsto \exp\langle \lambda_m | x \rangle$$

convergent uniformément vers $\exp\langle \lambda | \cdot \rangle$ sur K_n . Ainsi, pour m assez grand

$$\int \exp \left\langle \lambda_m \left| \sum_{z \in \Lambda(j)} \xi(z) \right. \right\rangle d\mu_m(\xi) \geq \int_{K_n^{\Lambda(j)}} \exp \left\langle \lambda \left| \sum_{z \in \Lambda(j)} \xi(z) \right. \right\rangle d\mu_m(\xi) - \frac{1}{n}$$

En passant à la limite inférieure en m , puis à la limite en n , on obtient²

$$\liminf_{m \rightarrow \infty} \int \exp \left\langle \lambda_m \left| \sum_{z \in \Lambda(j)} \xi(z) \right. \right\rangle d\mu_m(\xi) \geq \int \exp \left\langle \lambda \left| \sum_{z \in \Lambda(j)} \xi(z) \right. \right\rangle d\mu(\xi)$$

En prenant enfin la limite inférieure en j , on obtient

$$\liminf_{j \rightarrow \infty} \liminf_{m \rightarrow \infty} p_{\Lambda(j)}(\mu_m; \lambda_m) \geq p(\mu; \lambda)$$

Sous-étape 2 : Notons $\mu'_m = \mu_{\Lambda(m+g(m)+\ell)}$. On va se ramener à montrer que

$$\liminf_{m \rightarrow \infty} p(\mu'_m; \lambda_m) \geq \limsup_{j \rightarrow \infty} \liminf_{m \rightarrow \infty} p_{\Lambda(j)}(\mu'_m; \lambda_m)$$

En effet, on a d'une part, pour tout j ,

$$\liminf_{m \rightarrow \infty} p_{\Lambda(j)}(\mu'_m; \lambda_m) \geq \liminf_{m \rightarrow \infty} p_{\Lambda(j)}(\mu_m; \lambda_m)$$

car

$$\liminf_{m \rightarrow \infty} \frac{1}{|\Lambda(j)|} \log \mu_{\Lambda(m+g(m)+\ell)}(K_m^{\Lambda(m+g(m)+\ell)}) = 0$$

² Pour $m \geq j \vee n$, on a

$$\int_{\{\xi_{\Lambda(j)} \in K_n^{\Lambda(j)}\}} \exp \left\langle \lambda \left| \sum_{z \in \Lambda(j)} \xi(z) \right. \right\rangle d\mu_m(\xi) = \int_{\{\xi_{\Lambda(j)} \in K_n^{\Lambda(j)}\}} \exp \left\langle \lambda \left| \sum_{z \in \Lambda(j)} \xi(z) \right. \right\rangle d\mu(\xi)$$

D'autre part, en utilisant la proposition 6.5.3.4,

$$p(\mu_m; \lambda_m) = \frac{1}{|\Lambda(m + g(m) + \ell)|} \log \int_{K_m^{\Lambda(m+g(m)+\ell)}} \exp \left\langle \lambda \middle| \sum_{z \in \Lambda(m+g(m)+\ell)} \xi(z) \right\rangle d\mu(\xi) \\ - \frac{1}{|\Lambda(m + g(m) + \ell)|} \log \mu(K_m^{\Lambda(m+g(m)+\ell)})$$

et

$$p(\mu'_m; \lambda_m) = \frac{1}{|\Lambda(m + g(m) + \ell)|} \log \int \exp \left\langle \lambda \middle| \sum_{z \in \Lambda(m+g(m)+\ell)} \xi(z) \right\rangle d\mu(\xi)$$

Or, pour tout m ,

$$\lim_{n \rightarrow \infty} \int_{K_n^{\Lambda(m)}} \exp \left\langle \lambda \middle| \sum_{z \in \Lambda(m)} \xi(z) \right\rangle d\mu(\xi) = \int \exp \left\langle \lambda \middle| \sum_{z \in \Lambda(m)} \xi(z) \right\rangle d\mu(\xi)$$

Ainsi, par un procédé diagonal, quitte à extraire une sous-suite de (K_m) , on obtient

$$\liminf_{m \rightarrow \infty} p(\mu_m; \lambda_m) = \liminf_{m \rightarrow \infty} p(\mu'_m; \lambda_m)$$

Reste donc à montrer que

$$\liminf_{m \rightarrow \infty} p(\mu'_m; \lambda_m) \geq \limsup_{j \rightarrow \infty} \liminf_{m \rightarrow \infty} p_{\Lambda(j)}(\mu'_m; \lambda_m)$$

Pour ce faire, on va montrer une forme d'uniformité sur l'inégalité sous-additive définissant les $p(\mu'_m; \lambda_m)$; celle-ci est conséquence de la proposition suivante :

Proposition 6.5.3.6. *Si μ est c.a.d.i. et c.l., et $m \in \mathbb{N}$, alors $\mu_{\Lambda(m)}$ est aussi c.a.d.i. et c.l., avec les mêmes paramètres.*

Démonstration : Montrons par exemple le contrôle local. Notons (t, α) le paramètre de μ et fixons $z \in \mathbb{Z}^d$. On note Λ_z l'unique boîte de la forme $am + \Lambda(m)$ ($a \in \mathbb{Z}$) qui contient z . Alors, pour tout $\Lambda \subset \mathbb{Z}^d \setminus \{z\}$,

$$\begin{aligned} & \mu_{\Lambda(m)}(\xi(z) \in t(V) V; \xi_\Lambda \in C_\Lambda) \\ &= \sum_i \mu_{\Lambda(m)}(\xi(z) \in t(V) V; \xi_{\Lambda \cap \Lambda_z} \in C_{\Lambda \cap \Lambda_z}^i; \xi_{\Lambda \setminus \Lambda_z} \in C_{\Lambda \setminus \Lambda_z}^i) \\ &= \sum_i \mu_{\Lambda(m)}(\xi(z) \in t(V) V; \xi_{\Lambda \cap \Lambda_z} \in C_{\Lambda \cap \Lambda_z}^i) \mu_{\Lambda(m)}(\xi_{\Lambda \setminus \Lambda_z} \in C_{\Lambda \setminus \Lambda_z}^i) \\ &\geq \sum_i \alpha(V) \mu_{\Lambda(m)}(\xi_{\Lambda \cap \Lambda_z} \in C_{\Lambda \cap \Lambda_z}^i) \mu_{\Lambda(m)}(\xi_{\Lambda \setminus \Lambda_z} \in C_{\Lambda \setminus \Lambda_z}^i) \\ &= \sum_i \alpha(V) \mu_{\Lambda(m)}(\xi_{\Lambda \cap \Lambda_z} \in C_{\Lambda \cap \Lambda_z}^i; \xi_{\Lambda \setminus \Lambda_z} \in C_{\Lambda \setminus \Lambda_z}^i) \\ &= \alpha(V) \mu_{\Lambda(m)}(\xi_\Lambda \in C_\Lambda) \end{aligned}$$

Le raisonnement est analogue pour le découplage. $\square$

Notons ζ_m un champ de loi μ'_m . En vertu de la proposition précédente, ainsi que des propositions 6.2.1.1 et 6.2.2.1, $\langle \lambda_m | \zeta_m \rangle$ est (g, c) -c.a.d.i. et $(t \circ \lambda_m^{-1}, \alpha \circ \lambda_m^{-1})$ -c.l. Ainsi, l'inégalité (6.6) se récrit

$$\forall n \geq N(m)$$

$$p_{\Lambda(n)}(\mu'_m; \lambda_m) \geq (1 - \rho_{j,n})p_{\Lambda(j)}(\mu'_m; \lambda_m) - \frac{c(j)}{|\Lambda(j)|} - \rho_{j,n}(t(\lambda_m^{-1}(V_1)) - \log \alpha(\lambda_m^{-1}(V_1)))$$

Plus précisément, comme Y est tonnelé et comme les λ_m sont bornées en tout point (étant donné que $\lambda_m \rightarrow \lambda$ faiblement), le tonneau

$$T = \bigcap_{m=0}^{\infty} \lambda_m^{-1}([-\frac{1}{2}, \frac{1}{2}])$$

est un voisinage de 0. Soit $V \in \mathcal{C}_0(Y)$ contenu dans T . La démonstration pour obtenir 6.6 peut être refaite pour aboutir à

$$\forall n \geq N(m)$$

$$p_{\Lambda(n)}(\mu'_m; \lambda_m) \geq (1 - \rho_{j,n})p_{\Lambda(j)}(\mu'_m; \lambda_m) - \frac{c(j)}{|\Lambda(j)|} - \rho_{j,n}(t(V) - \log \alpha(V))$$

Prenant la limite en n , on obtient

$$\forall m \in \mathbb{N} \quad \forall j \in \mathbb{N} \quad p(\mu'_m; \lambda_m) \geq p_{\Lambda(j)}(\mu'_m; \lambda_m) - \frac{c'(j)}{|\Lambda(j)|}$$

Puis, pour tout $\delta > 0$, il existe $j_\delta \in \mathbb{N}$ tel que

$$\forall m \in \mathbb{N} \quad p(\mu'_m; \lambda_m) \geq \sup_{j \geq j_\delta} p_{\Lambda(j)}(\mu'_m; \lambda_m) - \delta$$

Alors

$$\begin{aligned} \liminf_{m \rightarrow \infty} p(\mu'_m; \lambda_m) &\geq \liminf_{m \rightarrow \infty} \sup_{j \geq j_\delta} p_{\Lambda(j)}(\mu'_m; \lambda_m) - \delta \\ &\geq \sup_{j \geq j_\delta} \liminf_{m \rightarrow \infty} p_{\Lambda(j)}(\mu'_m; \lambda_m) - \delta \\ &\geq \limsup_{j \rightarrow \infty} \liminf_{m \rightarrow \infty} p_{\Lambda(j)}(\mu'_m; \lambda_m) - \delta \end{aligned}$$

Prenant enfin la limite quand $\delta \rightarrow 0^+$, on obtient le résultat annoncé. $\square$

6.5.4 Égalité $s = -p^*$: cas général

Enfin, on peut énoncer :

Théorème 6.5.4.1. *Soient $\overleftarrow{\mathcal{X}} = (Y_i, f_i, f_{ij})_{i \leq j}$ un système projectif d'espaces de Fréchet séparables et X sa limite projective. Soit σ un champ à valeurs dans $(Y^{\mathbb{Z}^d}, \mathcal{B}^{\otimes \mathbb{Z}^d})$ tel que, pour tout $i \in J$, le champ $f_i(\sigma)$ soit c.a.d.i., c.l. et R_ℓ -invariant. Alors,*

$$s(\sigma; x) = -p^*(\sigma; x)$$

Démonstration : Le théorème 6.5.3.1 assure que

$$\forall i \in I \quad \forall y_i \in Y_i \quad s(f_i(\sigma); y_i) = -p^*(f_i(\sigma); y_i)$$

Soit maintenant $x \in X$; on a :

$$\begin{aligned} s(\sigma; x) &= \inf_{\substack{A \in \mathcal{C} \\ A \ni x}} \underline{s}(\sigma; A) \\ &= \inf_{i \in I} \left[\inf_{\substack{V_i \in \mathcal{C}_i \\ V_i \ni f_i(x)}} \underline{s}(f_i(\sigma); V_i) \right] \\ &= \inf_{i \in I} s(f_i(\sigma); f_i(x)) \\ &= -\sup_{i \in I} p^*(f_i(\sigma); f_i(x)) \end{aligned}$$

la dernière égalité étant conséquence du théorème 6.5.3.1. Puis

$$\begin{aligned} &\sup_{i \in I} p^*(f_i(\sigma); f_i(x)) \\ &= \sup_{i \in I} \sup_{\lambda_i \in Y_i^*} \left[\langle \lambda_i \circ f_i | x \rangle - \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \int \exp \left\langle \lambda_i \circ f_i \left| \sum_{z \in \Lambda(n)} \sigma(z) \right. \right\rangle d\mathbb{P} \right] \\ &= \sup_{\lambda \in X^*} \left[\langle \lambda | x \rangle - \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \log \int \exp \left\langle \lambda \left| \sum_{z \in \Lambda(n)} \sigma(z) \right. \right\rangle d\mathbb{P} \right] \\ &= p^*(\sigma; x) \end{aligned}$$

D'où le résultat attendu. □

BIBLIOGRAPHIE

- [Aze80] R. AZENCOTT : *Grandes déviations et applications*. Ecole d'Eté de Probabilités de Saint-Flour VIII-1978, Lecture Notes in Mathematics 774. Springer-Verlag, 1980.
- [Bah71] R. R. BAHADUR : *Some Limit Theorems in Statistics*. SIAM, 1971.
- [Bal88] P. BALDI : Large deviations and stochastic homogenization. *Ann. Math. Pura Appl.*, 151:161–177, 1988.
- [Ban32] S. BANACH : *Théorie des opérations linéaires*. Chelsea, 1932. Deuxième édition chez AMS Bookstore en 1978.
- [Bár78] P. BÁRTFAI : Large deviations of the sample mean in euclidean spaces. *Mimeograph Series No. 78–13*, Statist. Depart., Purdue Univ., 1978.
- [Bil68] P. BILLINGSLEY : *Convergence of Probability Measures*. John Wiley & Sons, New York, 1968.
- [Bil79] P. BILLINGSLEY : *Probability and Measures*. John Wiley & Sons, Chichester/New York, 1979.
- [Bol98] L. BOLTZMANN : *Vorlesungen über Gastheorie (Bd I & II)*. Johann Ambrosius Barth, Leipzig, 1896–98. Traduction française par M. Brillouin : *Leçons sur la théorie des gaz (vol. I & II)*, Gauthier-Villars, 1902–05.
- [Bol77] L. BOLTZMANN : Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärme Gleichgewicht. *Wien. Akad. Sitz.*, 76:373–435, 1877.
- [Bou66] N. BOURBAKI : *Théorie des ensembles*. Hermann, Paris, 1966.
- [Bou81] N. BOURBAKI : *Espaces vectoriels topologiques*. Masson, 1981.
- [BZ79] R. R. BAHADUR et S. L. ZABELL : Large deviations of the sample mean in general vector spaces. *Ann. Prob.*, 7(4):587–621, 1979.
- [Cer07] R. CERF : *On Cramér's Theory in Infinite Dimensions*. Panoramas et Synthèses 23. Société Mathématique de France, Paris, 2007.
- [CETT05] M. COSTENIUC, R. S. ELLIS, H. TOUCHETTE et B. TURKINGTON : The generalized canonical ensemble and its universal equivalence with the microcanonical ensemble. *J. Stat. Phys.*, 119(5/6):1283–1329, 2005.
- [Che52] H. CHERNOFF : A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23:493–507, 1952.

- [Cla67] R. CLAUSIUS : *Abhandlungen über die mechanische Wärmetheorie (Bd I & II)*. F. Vieweg & Sohn, Brunswick, 1864–67. Traduction française : *Théorie mécanique de la chaleur (vol. I & II)*, Lacroix, Paris, 1868–69.
- [Com86] F. COMETS : Grandes déviations pour des champs de Gibbs sur $\mathbb{Z}^d$. *C. R. Acad. Sc. Paris*, 303, sér. I(11):511–513, 1986.
- [Cra36] H. CRAMÉR : Über eine Eigenschaft der normalen Verteilungsfunktion. *Math. Z.*, 41(1):405–414, 1936.
- [Cra38] H. CRAMÉR : Sur un nouveau théorème limite de la théorie des probabilités. *Actual. Sci. Indust.*, 736:5–23, 1938.
- [DA85] A. DE ACOSTA : On large deviations of sums of independent random vectors. *Probability in Banach spaces V, Lecture Notes in Math.*, 1153:1–14, 1985.
- [DG87] D. A. DAWSON et J. GÄRTNER : Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20:247–308, 1987.
- [DGLR89] B. DIU, C. GUTHMANN, D. LEDERER et B. ROULET : *Éléments de physique statistique*. Hermann, 1989.
- [Din91] I. H. DINWOODIE : A note on the upper bound for i.i.d. large deviations. *Ann. Prob.*, 19(4):1732–1736, 1991.
- [DS89] J.-D. DEUSCHEL et D. W. STROOCK : *Large Deviations*. Academic Press, 1989.
- [DSZ91] J.-D. DEUSCHEL, D. W. STROOCK et H. ZESSIN : Microcanonical distributions for lattice gases. *Commun. Math. Phys.*, 139:83–101, 1991.
- [DT73] R. L. DOBRUSHIN et B. TIROZZI : The central limit theorem and the problem of equivalence of ensembles. *Commun. Math. Phys.*, 54:173–, 1973.
- [DV75] M. D. DONSKER et S. R. S. VARADHAN : Asymptotic evaluation of certain Markov process expectations for large time I. *Comm. Pure Applied Math.*, 28:1–47, 1975. Première partie d’une série de quatre articles sortis en 1975 et 1976.
- [DZ93] A. DEMBO et O. ZEITOUNI : *Large Deviations Techniques and Applications*. Jones and Bartlett, 1993. Deuxième édition chez Springer-Verlag en 1998.
- [EHT00] R. S. ELLIS, K. HAVEN et B. TURKINGTON : Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles. *J. Stat. Phys.*, 101(5/6):999–1064, 2000.
- [Ell84] R. S. ELLIS : Large deviations for a general class of random vectors. *Ann. Prob.*, 12(1):1–12, 1984.
- [Ell85] R. S. ELLIS : *Entropy, Large Deviations and Statistical Mechanics*. Springer-Verlag, 1985.
- [Fek23] M. FEKETE : Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Mathematische Zeitschrift*, 17: 228–249, 1923.

- [Fen49] W. FENCHEL : On conjugate convex functions. *Canad. J. Math.*, 1:73–77, 1949.
- [Föl73] H. FÖLLMER : On entropy and information gain in random fields. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 26:207–217, 1973.
- [FO88a] H. FÖLLMER et S. OREY : Large deviations for the empirical field of a Gibbs measure. *Ann. Prob.*, 16(3):961–977, 1988.
- [FO88b] H. FÖLLMER et M. ORT : Large deviations and surface entropy for Markov fields. *Société Maathématique de France, Astérisque*, 157–158:173–190, 1988.
- [FW84] M. I. FREIDLIN et A. D. WENTZELL : *Random Perturbations of Dynamical Systems*. Springer-Verlag, 1984.
- [Geo93] H.-O. GEORGII : Large deviations and maximum entropy principle for interacting random fields on $\mathbb{Z}^d$. *Ann. Prob.*, 21(4):1845–1875, 1993.
- [Geo94] H.-O. GEORGII : Large deviations and the equivalence of ensembles for Gibbsian particule systems with superstable interaction. *Probab. Theory Relat. Fields*, 99:171–195, 1994.
- [Geo95] H.-O. GEORGII : The equivalence of ensembles for classical systems of particles. *J. Stat. Phys.*, 80(5–6):1341–1378, 1995.
- [Gib02] W. GIBBS : *Elementary principles of statistical mechanics*. Charles Scribner’s sons, New York, 1902.
- [GOR79] P. GROENEBOOM, J. OOSTERHOFF et F. H. RUYMGAART : Large deviation theorems for empirical probability measures. *Ann. Prob.*, 7(4):553–586, 1979.
- [Gär77] J. GÄRTNER : On large deviations from the invariant measure. *Th. Prob. Appl.*, 22(1):24–39, 1977.
- [Ham74] J. M. HAMMERSLEY : Postulates for subadditive processes. *Ann. Prob.*, 2(4):652–680, 1974.
- [Hua63] K. HUANG : *Statistical Mechanics*. Wiley, New York, 1963.
- [Khi43] A. Ya. KHINSHIN : *Matematicheskie Osnovaniya Statisticheskoi Mekhaniki*. Gostekhizdat, Moscow-Leningrad, 1943. Traduction anglaise : *Mathematical Foundations of Statistical Mechanics*, Dover, New York, 1949.
- [Kin76] J. F. C. KINGMAN : *Subadditive Processes*. Ecole d’Eté de Probabilités de Saint-Flour V-1975, Lecture Notes in Mathematics 539. Springer, Berlin/Heidelberg, 1976.
- [KL51] S. KULLBACK et R. A. LEIBLER : On information and sufficiency. *Ann. Math. Stat.*, 22(1):79–86, 1951.
- [Lan73] O. E. LANFORD : *Entropy and Equilibrium States in Classical Statistical Mechanics*. Lecture Notes in Physics 20. Springer, 1973.
- [LC86] L. LE CAM : The central limit theorem around 1935. *Statist. Sc.*, 1(1):78–96, 1986.
- [LP95] J. T. LEWIS et C.-E. PFISTER : Thermodynamic probability theory. *Russian Math. Surveys*, 50:279–317, 1995.

- [LPS94a] J. T. LEWIS, C.-E. PFISTER et W. G. SULLIVAN : *Large deviations and the thermodynamic formalism : a new proof of the equivalence of ensembles*. Plenum Press, 1994.
- [LPS94b] J. T. LEWIS, C.-E. PFISTER et W. G. SULLIVAN : The equivalence of ensembles for lattice systems : some examples and a counterexample. *J. Stat. Phys.*, 77:397–419, 1994.
- [LPS95] J. T. LEWIS, C.-E. PFISTER et W. G. SULLIVAN : Entropy, concentration of probability and conditional limit theorems. *Markov Processes Relat. Fields*, 1(3):319–386, 1995.
- [LU34] Z. ŁOMNICKI et S. ULAM : Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités I. Variables indépendantes. *Fund. Math.*, 23:237–278, 1934.
- [Mor67] J.-J. MOREAU : *Fonctionnelles convexes*. Séminaire sur les Equations aux Dérivées Partielles, Collège de France, 1966–67.
- [Mos71] U. MOSCO : On the continuity of the Young–Fenchel transform. *J. Math. Anal. Applic.*, 35:518–535, 1971.
- [Oll88] S. OLLA : Large deviations for Gibbs random fields. *Probab. Theory Relat. Fields*, 77:343–357, 1988.
- [OP88] S. OREY et S. PELIKAN : Large deviation principles for stationary processes. *Ann. Prob.*, 16(4):1481–1495, 1988.
- [Pfi02] C.-E. PFISTER : *Thermodynamical aspects of classical lattice systems. Lectures delivered at the 4th brazilian school of probability, Mambucaba, RJ (2000), In and Out of Equilibrium, Probability with a Physics Flavor*. V. Sidoravicius, Birkhäuser, 2002.
- [Pól20] G. PÓLYA : Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.*, 8:171–181, 1920.
- [PS54] G. PÓLYA et G. SZEGÖ : *Aufgaben und Lehrsätze aus der Analysis I (Anh. 2)*. Springer-Verlag, Berlin, 1954. Dernière édition anglaise : *Problems and Theorems in Analysis I*, Springer, 2004.
- [PS75] D. PLACHKY et J. STEINEBACH : A theorem about probabilities of large deviations with an application to queuing theory. *Period. Math. Hungarica*, 6(4):343–345, 1975.
- [Roc70] R. Tyrell ROCKAFELLAR : *Convex Analysis*. Princeton Univ. Press, 1970.
- [RR64] A. P. ROBERTSON et W. J. ROBERTSON : *Topological vector spaces*. Cambridge University press, 1964.
- [Rud66] W. RUDIN : *Real and Complex Analysis*. McGraw–Hill, New Dehli, 1966. Référence dans la deuxième édition de 1974.
- [Rue65] D. RUELLLE : Correlation functionals. *J. Math. Phys.*, 6(2):201–220, 1965.

- [Rue69] D. RUELLE : *Statistical Mechanics – Rigorous Results*. W. A. Benjamin, Inc., 1969. Deuxième édition en 1974.
- [SAJ48] E. SPARRE ANDERSEN et B. JESSEN : On the introduction of measures in infinite product sets. *Det. Kgl. Dan. Vid. Sel., Mat.–Fys. Medd.*, 25(4), 1948.
- [San57] I. N. SANOV : О вероятности больших отклонений случайных величин. *Mat. Sb.*, 42(1):11–44, 1957. Traduction anglaise : On the probability of large deviations of random variables, *Sel. Transl. Math. Statist. Prob. I* : 213–244, 1961.
- [Sha48] C. E. SHANNON : A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.
- [Sla88] M. SLABY : On the upper bound for large deviations of sums of i.i.d. random vectors. *Ann. Prob.*, 16(3):978–990, 1988.
- [Smi33] N. SMIRNOV : О вероятностях больших уклонений (Über Wahrscheinlichkeiten grosser Abweichungen). *Mat. Sb.*, 40(4):443–454, 1933.
- [TM71] M. TRIBUS et E. C. MCIRVINE : Energy and information. *Scientific American*, 225:179–188, 1971.
- [Var66] S. R. S. VARADHAN : Asymptotic probabilities and differential equations. *Comm. Pure Applied Math.*, 19:261–286, 1966.
- [vH49] L. van HOVE : Quelques propriétés générales de l'intégrale de configuration d'un système de particules avec interaction. *Physica*, 15(11–12):951–961, 1949.
- [vN35] J. von NEUMANN : Functional Operators. Mimeographed lectur notes. Institute for Advanced Study, Princeton, N.J., 1935. Publié dans Ann. Math. Studies 21, Functional Operators I, Measures and Integrals, Princeton University Press, 1950.
- [YL52] C. N. YANG et T. D. LEE : Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Physical Review*, 87(3):404–409, 1952.
- [Zab92a] S. L. ZABELL : Mosco convergence and large deviations. *Probability in Banach spaces 8, Progr. Probab.* 30, pages 245–252, 1992.
- [Zab92b] S. L. ZABELL : Mosco convergence in locally convex spaces. *J. Funct. Anal.*, 110(1):226–246, 1992.
- [ZRM09] R. K. P. ZIA, E. F. REDISH et S. R. MCKAY : Making sense of the Legendre transform. *Amer. J. Phys.*, 77:616–622, 2009.