Thèse présentée

pour obtenir le titre de

docteur de l'Université Paris-10 Nanterre

Spécialité: probabilités

Grandes déviations pour certaines mesures empiriques

Jamal Najim

Directeur de thèse: Christian Léonard

soutenue le 20 décembre 2001

devant le jury composé de:

Patrick Cattiaux,	Président
Amir Dembo,	Rapporteurs
Alain Rouault,	
Francis Comets,	Examinateurs
Guy Fayolle,	
Fabrice Gamboa,	
Christian Léonard,	Directeur de thèse.

Remerciements

J'ai passé avec Christian Léonard trois années très stimulantes. Son optimisme indéfectible a été pour moi une constante source de motivation. Christian a été un directeur de thèse disponible et attentionné. Je l'en remercie.

It was a great honour for me to have Amir Dembo as one of my "rapporteurs". I would like to thank him for this and for the interest he showed in my work.

Alain Rouault a partagé, avec Amir Dembo, la lourde tâche d'être rapporteur. Je tiens à l'en remercier très sincèrement. Sa lecture très attentive du manuscript a permis d'en améliorer substantiellement la rédaction.

Merci a Patrick Cattiaux, Francis Comets, Guy Fayolle et Fabrice Gamboa d'avoir accepté de participer à mon jury de thèse.

Mon temps s'est partagé, lors de mes trois années de thèse, entre le laboratoire Modal'X de l'université de Nanterre et le laboratoire TSI de l'Ecole Nationale Supérieure des Télécommunications.

J'aimerais remercier toute l'équipe de Nanterre pour l'accueil qui m'a été réservé. Merci tout particulièrement à Rida Laraki, Christian Léonard, Luc Miller, Boyan Sirakov et Sylvain Sorin pour l'animation du groupe de travail "Solutions de viscosité". Merci également à Christian, Laurent Mesnager et Maximilien Rzepka pour avoir monté le groupe de lecture "A Weak Approach to Large Deviations".

Eric Moulines et Suleyman Üstünel, professeurs à l'ENST, sont deux personnes qui ont énormément compté pour moi. Je tiens à les remercier pour le temps qu'ils m'ont consacré et les conseils qu'ils m'ont prodigués tout au long de ces trois années. Merci tout particulièrement à Suleyman Üstünel pour avoir animé un "read & course" de luxe en présence de Laurent Veilex et moi-même, sur le thème du Calcul de Malliavin.

Vincent Buchoux, Randal Douc, Stéphanie Dubost, Gilles Faÿ, Gersende Fort, Denis Matignon, Mila Nikolova, François Roueff, Olivier Thomas et Cyril Touzé ont été de forts sympathiques compagnons de route, que je remercie pour leur disponibilité.

Last but not least, Olivier Cappé et Jean-François Cardoso ont été d'un secours inestimable pour la chose informatique. Qu'ils en soient grandement remerciés.

J'aimerais dédier cette thèse à Gaëlle, Sami et Yassine et à mes parents, Annie et Mao.

Table des matières

Introduction

En 1987, Lynch et Sethuraman [39] ont établi un principe de grandes déviations (PGD) pour la famille de processus $\left(\frac{1}{\lambda}X(\lambda\,\cdot)\right)_\lambda$ où $X(t)$ est un processus stationnaire, réel, à accroissements positifs et indépendants, sous des hypothèses d'intégrabilité très faibles.

Parler de grandes déviations, c'est dire qu'il existe une fonctionnelle de taux I telle que pour un ensemble A de fonctions données, on ait, pour $\lambda \to \infty$:

$$\mathbb{P}\left(\frac{1}{\lambda}X(\lambda\,\cdot) \in A\right) \approx_\lambda e^{-\lambda I_A} \qquad \text{où} \quad I_A = \inf\{I(x),\ x \in A\}.$$

Dans le cas où $I_A > 0$, la terminologie prend tout son sens : A est un événement *déviant* au sens où sa probabilité de réalisation tend vers zéro. De plus, la *déviation* est *grande* puisque la vitesse de convergence vers zéro est exponentielle au *taux* I_A. En fait, derrière le symbole $\approx$ se cache une définition un peu moins intuitive qui requiert une topologie pour pouvoir parler d'adhérence $\bar{A}$ et d'intérieur $\mathring{A}$:

$$\begin{aligned} \limsup_\lambda \frac{1}{\lambda}\log \mathbb{P}(X(\lambda\,\cdot)/\lambda \in A) &\leq -\inf\{I(x),\ x \in \bar{A}\}, \\ -\inf\{I(x),\ x \in \mathring{A}\} &\leq \liminf_\lambda \frac{1}{\lambda}\log \mathbb{P}(X(\lambda\,\cdot)/\lambda \in A). \end{aligned}$$

On se réfèrera à la première inégalité comme à la majoration des grandes déviations et à la seconde comme à la minoration.

Lynch et Sethuraman ont démontré que si le processus admet quelques moments exponentiels, *i.e.*

$$\mathbb{E}e^{\alpha|X(1)|} < \infty \quad \text{pour un } \alpha > 0,$$

la fonction de taux a pour forme

$$I(x) = \int_{[0,1]} \Lambda^*(\dot{x}_a(t))\,dt + Cx_s([0,1]). \tag{0.1}$$

Autrement dit, si x est une fonction à variation bornée (assimilable *de facto* à une mesure signée) et si $x_a + x_s$ est sa décomposition de Lebesgue avec $x_a \ll dt$ et $x_s \perp dt$ alors la valeur de I en x est donnée par (0.1). Sous l'hypothèse plus forte d'existence de tous les moments exponentiels,

$$\mathbb{E}e^{\alpha|X(1)|} < \infty \quad \text{pour tout } \alpha > 0,$$

les techniques de Varadhan [55] permettent d'établir un principe de grandes déviations pour la famille $\frac{1}{\lambda}X(\lambda\,\cdot)$, gouverné cette fois-ci par la fonctionnelle de taux

$$\tilde{I}(x) = \int_{[0,1]} \Lambda^*(\dot{x}(t))\,dt$$

si x est absolument continue (et admet donc une dérivée $\dot{x}$ presque partout) et $+\infty$ sinon. Cette fonctionnelle est bien connue puisqu'elle apparaît en 1966 dans les travaux de Schilder sur les grandes déviations du Brownien normalisé (dans le cas particulier où $\Lambda^*(x) = \frac{|x|^2}{2}$ est la transformée de Cramér de la gaussienne) puis sous cette forme dans les articles de Borovkov [8] et Mogul's-kii [40].

Notons que si Mogul'skii [40] aborde dès 1976 l'étude des grandes déviations de processus empiriques sous l'hypothèse d'existence de quelques moments exponentiels, ses résultats ne mettent pas en évidence la forme tout à fait particulière de la fonctionnelle donnée par (0.1). En effet, Mogul'skii établit un principe de grandes déviations pour des processus empiriques gouverné par une fonctionnelle de taux abstraite, qui est définie comme régularisée semi-continue inférieure de $\tilde{I}$.

Et c'est le mérite de Lynch et Sethuraman d'avoir établi la corrélation entre existence de quelques moments exponentiels et apparition d'un terme supplémentaire dans la fonction de taux. Plus encore, ils ont décrit précisément cette contribution singulière au moyen de la direction de récession de Λ^* :

$$C = \lim_{x\to\infty} \frac{\Lambda^*(x)}{x}.$$

La découverte de ce phénomène a entrainé de nombreux travaux dont le point commun est l'apparition d'un terme singulier dans la fonction de taux lors de l'étude d'un phénomène de grandes déviations en l'absence de tous les moments exponentiels. Autour des fonctionnelles de processus gaussiens, citons, sans être exhaustif, les travaux de Bryc et Dembo [10], ainsi que ceux de Bercu, Gamboa, Lavielle, Rouault et Zani [5, 6, 30]. Les mesures empiriques ont été étudiées par Van den Berg, Dorlas, Ellis, Gough, Lewis et Pulé [28, 54], Gamboa et Gassiat [29] et Léonard [36] (mesures de Poisson). Enfin, l'étude des processus à accroissements indépendants, en filiation directe avec l'article de Lynch et Sethuraman, a été menée par Mogul'skii [41], De Acosta [17] et Léonard [36].

Outre la fonction de taux et sa forme inhabituelle qu'il faut déterminer, l'absence de tous les moments exponentiels soulève des difficultés importantes pour établir la borne inférieure d'un principe de grandes déviations. Essayons de les décrire et rappelons pour cela la technique de changement exponentiel de probabilité.

Soit $\overline{X}_n = \frac{1}{n}\sum_1^n X_i$ une moyenne empirique de variables réelles indépendantes et identiquement μ-distribuées et soit $\Lambda(\lambda) = \log Ee^{\lambda X}$ la log-Laplace associée. Introduisons les variables $\tilde{X}_i$ dont la loi est donnée par

$$\frac{d\tilde{\mu}}{d\mu}(x) = e^{\lambda_y x - \Lambda(\lambda_y)}$$

où λ_y est choisi de sorte que

$$\Lambda'(\lambda_y) = y \quad \text{soit} \quad \frac{\mathbb{E}Xe^{\lambda_y X}}{\mathbb{E}e^{\lambda_y X}} = \mathbb{E}Xe^{\lambda_y X - \Lambda(\lambda_y)} = y. \tag{0.2}$$

La minoration devient ainsi

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{\frac{1}{n}\sum_1^n X_i \in [y-\delta, y+\delta]\right\} \\
&= \lim_{n\to\infty} \frac{1}{n} \log \int_{\{\overline{x}_n \in [y-\delta, y+\delta]\}} e^{\lambda_y n \overline{x}_n - n\Lambda(\lambda_y)} \\
&\qquad\qquad e^{-\lambda_y n \overline{x}_n + n\Lambda(\lambda_y)} \mu(dx_1)\cdots\mu(dx_n) \\
&\geq \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{\frac{1}{n}\sum_1^n \tilde{X}_i \in [y-\delta, y+\delta]\right\} - (\lambda_y y - \Lambda(\lambda_y)) - \delta|y|.
\end{aligned}$$

Cette "translation exponentielle" permet de "transformer" un événement déviant en un événement typique. En effet, la définition même de λ_y (0.2) entraine que les variables $\tilde{X}_i$ sont centrées en y. Par suite, la loi des grands nombres nous assure que

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{1}{n}\sum_1^n \tilde{X}_i \in [y-\delta, y+\delta]\right\} = 1$$

et l'on obtient une minoration :

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{\frac{1}{n}\sum_1^n X_i \in [y-\delta, y+\delta]\right\} \geq -(\lambda_y y - \Lambda(\lambda_y)) - \delta|y| = -\Lambda^*(y) - \delta|y|$$

FIG. 1 – Cas d'une log-Laplace non-escarpée

qui entraine immédiatement la forme générale de la minoration de grandes déviations, associée à la fonction de taux Λ^* (conjuguée convexe de la log-Laplace). Retenons que le point crucial de cette minoration est l'obtention d'une solution à l'équation

$$\Lambda'(\lambda) = y$$

pour tout réel y. En d'autres termes, les pentes de la log-Laplace doivent "balayer" l'ensemble des directions possibles. On parle alors d'escarpement (*steepness*) de la log-Laplace.

Cette technique délicate et *subtle* qui, par comparaison, fait passer la majoration de grandes déviations pour une étape *rather mundane*[1] est due à Cramér. Elle s'est avérée extrêmement fructueuse et a donné lieu à de nombreuses extensions parmi lesquelles les travaux de Gärtner [31] et Ellis [27].

Il apparaît néanmoins dans la description rapide que nous en avons faite que le changement de probabilité ne peut s'opérer sur l'exemple simple de la moyenne empirique lorsque la log-Laplace associée n'est pas escarpée (voir FIG. 1). Ce cas de figure peut arriver en l'absence de tous les moments exponentiels sans pour autant remettre en cause l'existence d'un principe de grandes déviations.

Et c'est un fait que les travaux précédemment cités sont caractérisés par une grande variété de techniques visant à passer outre cette difficulté, c'est à dire essentiellement à établir des principes de grandes déviations en l'absence du théorème de Gärtner-Ellis. Approximation discrète de la fonction de taux chez Lynch et Sethuraman, changement exponentiel de probabilité dépendant du temps chez Bryc et Dembo, extension *ad hoc* du théorème de Gärtner-Ellis(-Baldi) chez Bercu, Gamboa et Rouault, etc.

Cette thèse s'inscrit dans la problématique précédente et son objet est l'étude des grandes déviations de différentes mesures empiriques : la probabilité empirique

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

et la mesure empirique MEM associée à la méthode du Maximum d'Entropie

[1] Stroock [53], Remarque 1.3.15, p.31

en Moyenne
$$\frac{1}{n}\sum_{i=1}^{n} Z_i\delta_{x_i}.$$
Les variables aléatoires (Z_i) considérées sont (sauf au dernier chapitre) indépendantes et identiquement distribuées (i.i.d.), et l'accent est mis ici sur l'existence de quelques moments exponentiels :

$$\exists \alpha > 0, \quad \mathbb{E}e^{\alpha|Z_i|} < \infty.$$

Ainsi, nous nous sommes intéressés au théorème de Sanov dans le cas où les fonctions-test satisfont de faibles conditions d'intégrabilité (à l'échelle des grandes déviations) :

$$\exists\, \alpha_f > 0, \quad \mathbb{E}e^{\alpha_f|f(Z_i)|} < \infty.$$

Sans surprise, un terme s'ajoute à l'entropie relative $H(\cdot|\mu)$, fonction de taux habituelle :
$$I(\ell) = H(\ell^a \mid \mu) + I^s(\ell^s),$$
ℓ^a étant une probabilité et ℓ^s, une forme linéaire continue qui n'est pas σ-additive.

Notre contribution à l'étude de la mesure empirique MEM a consisté à établir un principe de grandes déviations dans un contexte multidimensionnel et en s'affranchissant de l'hypothèse d'escarpement de la log-Laplace. Dès lors, l'impossibilité d'utiliser le théorème de Gärtner-Ellis dans cette situation nous a conduit à mettre en place une technique particulière, fondée sur la sous-additivité et sur la notion d'approximation exponentielle, nous permettant ainsi d'établir un principe de grandes déviations intermédiaire pour la moyenne pondérée :
$$\frac{1}{n}\sum_{1}^{n} f(x_i)Z_i.$$
Nous avons illustré ces résultats par différentes applications -principe de conditionnement de Gibbs, grandes déviations pour les processus- et enfin nous avons étudié de nombreux exemples, de manière à saisir la nature particulière des contributions singulières qui apparaissent. Rappelons quelques définitions fondamentales :

Principe de grandes déviations On dit qu'une famille de probabilités $\mathbb{P}_n$, définies sur un espace mesurable $(\mathcal{X}, \mathcal{E})$ muni d'une topologie, satisfait un principe de grandes déviations de fonction de taux I s'il existe une fonction I :

$\mathcal{X} \to \mathbb{R}^+$ semi-continue inférieurement telle que pour tout ensemble mesurable A,

$$\begin{aligned} -\inf\{I(x),\ x \in \mathring{A}\} &\leq \liminf_n \frac{1}{n} \log \mathbb{P}_n(A), \\ \limsup_n \frac{1}{n} \log \mathbb{P}_n(A) &\leq -\inf\{I(x),\ x \in \bar{A}\}. \end{aligned}$$

Ici, $\mathring{A}$ et $\bar{A}$ désignent respectivement l'intérieur et l'adhérence de A. On parle de bonne fonction de taux lorsque I est inf-compacte, c'est-à-dire lorsque ses ensembles de niveau $\{I \leq a\}$ (a réel) sont compacts. Dire qu'une famille de variables aléatoires L_n satisfait un principe de grandes déviations, c'est dire que la famille de probabilités $\mathbb{P}(L_n \in \cdot)$ satisfait un principe de grandes déviations au sens de la définition précédente. Rappelons enfin le point de vue plus intuitif évoqué plus haut : si L_n satisfait un principe de grandes déviations et si l'ensemble A vérifie

$$\inf\{I(x),\ x \in \mathring{A}\} = \inf\{I(x),\ x \in \bar{A}\} \stackrel{\triangle}{=} I_A,$$

alors

$$\lim_n \frac{1}{n} \log \mathbb{P}(L_n \in A) = -I_A \quad \text{soit} \quad \mathbb{P}(L_n \in A) \approx_n e^{-nI_A}.$$

Borne inférieure et escarpement de la log-Laplace Une des techniques les plus utilisées (Cramér [12], Gärtner [31], Ellis [27]) pour établir la borne inférieure d'un principe de grandes déviations est fondée sur le changement exponentiel de probabilité précédemment évoqué. Cette technique s'appuie sur une propriété particulière de la log-Laplace $\Lambda(\lambda) = \log Ee^{\lambda Z}$. Celle-ci doit être une fonction **escarpée**, c'est à dire :
- $\mathcal{D}_\Lambda = \{\lambda,\ \Lambda(\lambda) < \infty\}$ est d'intérieur non vide.
- $\Lambda(\lambda)$ est partout différentiable à l'intérieur de $\mathcal{D}_\Lambda$.
- $\lim_{n\to\infty} |\nabla\Lambda(\lambda_n)| = \infty$ lorsque λ_n est une suite appartenant à l'intérieur de $\mathcal{D}_\Lambda$ et convergeant vers un point de la frontière de $\mathcal{D}_\Lambda$.

Sous ces hypothèses, un principe de grandes déviations a lieu sous des conditions relativement générales (théorème de Gärtner-Ellis).
Néanmoins, il existe des résultats de grandes déviations établis sous des hypothèses moins restrictives.

Une approche alternative : la sous-additivité La généralisation du théorème de Cramér due à Bahadur et Zabell [4] est fondée sur une idée dite de sous-additivité attribuée à Ruelle et Lanford[2]. Elle permet d'établir un principe

[2]Bahadur et Zabell, [4], introduction.

de grandes déviations pour la moyenne empirique d'une famille de variables aléatoires indépendantes et identiquement distribuées $\overline{X}_n = \frac{1}{n}\sum_1^n X_i$ dès lors que les variables aléatoires X_i ont quelques moments exponentiels :

$$\mathbb{E}e^{\alpha|X_i|} < \infty \quad \text{pour un } \alpha > 0.$$

Ainsi, la moyenne empirique associée aux variables positives et i.i.d., distribuées selon la loi $P(dx) = C\frac{e^{-x}}{1+x^3}\,dx$, satisfait un PGD en vertu du théorème de Bahadur et Zabell bien que la log-Laplace associée ne soit pas escarpée (voir FIG. 1).
L'idée de la sous-additivité est la suivante : considérons la moyenne empirique $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ d'un échantillon de variables aléatoires (X_i) i.i.d. et à valeurs dans $\mathbb{R}^d$ et soit $B(x,\epsilon)$ la boule centrée en x et de rayon ϵ. Alors

$$\overline{X}_{n+m} = \frac{n}{n+m}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) + \frac{m}{n+m}\left(\frac{1}{m}\sum_{i=1}^m X_{n+i}\right).$$

Par conséquent, comme $B(x,\epsilon)$ est convexe, on a

$$\left\{\overline{X}_n \in B(x,\epsilon)\right\} \cap \left\{1/m\sum_1^m X_{n+i} \in B(x,\epsilon)\right\} \subset \left\{\overline{X}_{n+m} \in B(x,\epsilon)\right\}$$

et la fonction $f(n) = -\log\mathbb{P}\{\overline{X}_n \in B(x,\epsilon)\}$ est sous-additive : $f(n+m) \leq f(n)+f(m)$, ce qui assure la convergence de $\frac{f(n)}{n}$ ([20], section 6.1). La convergence d'une telle quantité pour toutes les boules assure, via un argument abstrait ([20], section 4.1), l'existence d'un principe de grandes déviations. Notons que cette méthode, pour puissante qu'elle est, s'appuie très fortement sur la structure d'indépendance et d'équidistribution du modèle.

Approximation exponentielle Il s'agit d'établir un principe de grandes déviations pour un objet L_n au moyen d'approximations L_n^ϵ : si L_n^ϵ réalise une approximation de L_n en un certain sens et si L_n^ϵ satisfait (à ϵ fixé) un principe de grandes déviations, est-il possible d'en déduire un principe de grandes déviations pour L_n ?
On dit que L_n^ϵ, à valeurs dans $\mathbb{R}^d$, est une approximation exponentielle de L_n (à valeurs dans $\mathbb{R}^d$ également) si

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{|L_n^\epsilon - L_n| > \delta\right\} = -\infty, \quad \forall\delta > 0.$$

Dans ce cas, si L_n^ϵ satisfait un principe de grandes déviations de fonction de taux I^ϵ, L_n satisfait un principe de grandes déviations faible (la majoration des grandes déviations n'est valide que pour les ensembles compacts) de fonction de taux

$$\begin{aligned} I(z) &= \sup_{\delta>0} \liminf_{\epsilon\to 0} \inf_{y\in B(z,\epsilon)} I^\epsilon(y) \\ &= \sup_{\delta>0} \limsup_{\epsilon\to 0} \inf_{y\in B(z,\epsilon)} I^\epsilon(y). \end{aligned}$$

La forme peu avenante de la fonction de taux I illustre le fait que l'application de cette technique s'accompagne souvent d'un travail d'identification de la fonction de taux.

Structure du document et résumé des chapitres

Chapitre 1 : Une généralisation du théorème de Sanov. Par théorème de Sanov, on entend PGD pour la mesure empirique $\frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$ d'un échantillon (Z_i) de variables i.i.d., distribuées selon une loi μ. Nous généralisons dans ce chapitre les résultats de Groeneboom, Ooesterhoff et Ruymgaart [33] et de Eichelsbacher et Schmock [26] en établissant un PGD pour la mesure empirique sous une topologie plus fine que celles précédemment considérées. Ce résultat est à rapprocher des travaux de Csiszár [13, 14] et de Schied [51].

Si f est une fonction mesurable qui admet quelques moments exponentiels :

$$\mathbb{E}e^{\alpha|f(Z)|} < \infty \quad \text{pour un } \alpha > 0,$$

et si L_τ désigne l'espace vectoriel de ces fonctions, alors la probabilité empirique $\frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$ satisfait un principe de grandes déviations dans un certain espace $\mathcal{Q}$ muni de la topologie $*$-faible $\sigma(\mathcal{Q}, L_\tau)$ avec pour bonne fonction de taux :

$$I(\ell) = H(\ell^a \mid \mu) + I^s(\ell^s),$$

où ℓ^a est une probabilité, ℓ^s est une forme linéaire continue sur L_τ qui n'est pas σ-additive. En particulier, ℓ^s **n'est pas** une mesure singulière par rapport à μ (pour l'apparition de mesures singulières dans un contexte différent, voir chapitre 4). Ici, $H(\cdot|\mu)$ désigne l'entropie relative habituelle par rapport à μ :

$$H(\nu \mid \mu) = \int_\Sigma \log\left(\frac{d\nu}{d\mu}\right) d\nu \quad \text{si} \quad \nu \ll \mu.$$

Ce résultat est établi au **théorème 1.9**. Notons que la topologie considérée est beaucoup plus riche que la topologie traditionnelle de la convergence étroite et que le PGD a lieu dans un ensemble plus grand que l'ensemble des probabilités, ce qui est illustré par l'apparition d'une contribution singulière $I^s(\ell^s)$

dans la fonction de taux. Ces résultats sont établis grâce au cadre fonctionnel particulier des espaces d'Orlicz, dont les principaux résultats sont rappelés au paragraphe 1.1.

Enfin, un exemple illustrant l'existence de parties singulières et faisant le lien avec les "projections généralisées" de Csiszár [13, 14] est traité au paragraphe 1.3.

Les résultats de ce chapitre et du suivant font l'objet d'un article [38] écrit en collaboration avec Christian Léonard et actuellement soumis pour publication.

Chapitre 2 : Autour du principe de conditionnement de Gibbs. Les grandes déviations sont intimement liées aux problèmes de mécanique statistique. Ainsi, Stroock et Zeitouni [52] ont établi un principe de conditionnement de Gibbs à l'aide du théorème de Sanov ([20], théorème 7.3.3, corollaire 7.3.5). Dans ce chapitre, nous généralisons leurs travaux en suivant leur stratégie, munis cette fois de la généralisation du théorème de Sanov établie au chapitre précédent. Si $L_n^Y = \frac{1}{n}\sum_1^n \delta_{Y_i}$ désigne la mesure empirique associée aux variables Y_i et si A_0 est un ensemble convexe mesurable pouvant être approximé par des fermés $(A_\delta)_{\delta>0}$ (voir les hypothèses au paragraphe 2.1), alors

$$\lim_{n\to\infty} \mathbb{P}\Big((Y_1,\ldots,Y_k) \in \cdot \mid L_n^Y \in A_0\Big) = \nu_*^k(\cdot), \tag{0.3}$$

où ν_* est une probabilité reliée (il s'agit de la projection généralisée introduite par Csiszár dans [14]) aux minimisants de I, fonction de taux associée au théorème de Sanov généralisé, sous la contrainte A_0. La convergence (0.3) a lieu pour la topologie de la convergence étroite :

$$\mathbb{E}[f(Y_1,\cdots,Y_k)|L_n^Y \in A_\delta] \to \int_{\mathcal{X}^k} f(y_1,\cdots,y_k)\nu_*(dy_1)\cdots\nu_*(dy_k)$$

pour $n\to\infty$ suivi de $\delta\to 0$ et si f est une fonction-test à valeurs réelles continue bornée (**théorème 2.5**). Si f a tous ses moments exponentiels, *i.e.* $Ee^{\alpha|f(Y_i)|}<\infty$ pour tout α réel, alors on a également

$$\mathbb{E}[f(Y_1)|L_n^Y \in A_\delta] \to \int_{\mathcal{X}} f(y_1)\nu_*(dy_1).$$

Ce résultat est démontré au **théorème 2.11**. Enfin, si f n'admet que quelques moments exponentiels, *i.e.* $Ee^{\alpha|f(Y_i)|}<\infty$ pour un $\alpha>0$, alors la convergence précédente n'a en général plus lieu (phénomène de butée). Des résultats précisant ce point sont établis au paragraphe 2.5.3 (**théorème 2.16**).

On notera que tous les résultats précédemment cités ne nécessitent pas la condition suivante de loi des grands nombres sous-jacente :

$$\lim_{n\to\infty} \nu_*^n(\{L_n^Y \in A_\delta\}) = 1, \quad \forall\delta > 0.$$

Cette condition peut même ne pas être vérifiée comme en témoigne l'exemple étudié au paragraphe 2.4.

Chapitre 3 : Un théorème de Cramér pour une moyenne empirique pondérée. Dans ce chapitre, nous étudions les grandes déviations de la moyenne empirique pondérée

$$\langle L_n, f\rangle = \frac{1}{n}\sum_1^n f(x_i^n)Z_i,$$

où $(Z_i)_{i\in\mathbb{N}}$ est une suite de variables aléatoires i.i.d. à valeurs $\mathbb{R}^d$ satisfaisant :

$$\mathbb{E}e^{\alpha|Z|} < +\infty \quad \text{pour un} \quad \alpha > 0,$$

et $\frac{1}{n}\sum_1^n \delta_{x_i^n}$ converge étroitement vers la probabilité R. Comme on s'affranchit de l'hypothèse d'escarpement de la log-Laplace, le principe de grandes déviations considéré ne découle pas des théorèmes classiques. En particulier, les hypothèses du théorème de Gärtner-Ellis ne sont pas forcément vérifiées. Par suite, une partie importante de ce chapitre est consacrée à la preuve du principe de grandes déviations. On remarquera que, du fait de l'utilisation d'une technique d'approximation exponentielle, l'identification de la fonction de taux est assez délicate (section 3.2.2). Le résultat principal du chapitre est le **théorème 3.2**.
Deux contre-exemples sont construits au paragraphe 3.1.3 lorsque la probabilité R ne satisfait pas la condition suivante de stricte positivité :

$$\text{si } U \text{ est un ouvert non vide, alors} \quad R(U) > 0.$$

Les résultats de ce chapitre et du suivant font l'objet d'un article [42] accepté pour publication dans *Electronic Journal of Probability.*

Chapitre 4 : Principe de grandes déviations pour le maximum d'entropie en moyenne. Nous récoltons ici les fruits du travail réalisé au chapitre précédent. On établit dans un premier temps un principe de grandes déviations pour la mesure empirique associée au Maximum d'entropie en moyenne (qu'on appelle dans la suite mesure MEM) :

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

Ce résultat est l'objet du **théorème 4.1**. Il constitue une généralisation des travaux de Gamboa et Gassiat (voir [29]). De ce PGD découlent des principes de grandes déviations pour certains processus empiriques (**théorèmes 4.7 et 4.8**).

Chapitre 5 : Grandes déviations pour des variables indépendantes mais non identiquement distribuées On s'intéresse à nouveau dans ce chapitre à la moyenne empirique $\frac{1}{n}\sum_1^n f(x_i)Z_i$ dans le cas où les variables Z_i sont indépendantes mais ne sont plus identiquement distribuées. Il s'agit plus précisément, d'autoriser la distribution de la variable Z_i à dépendre du "site" x_i. De tels objets apparaissent naturellement en estimation par la technique du Maximum d'entropie en moyenne. On introduit une distance entre probabilités, que l'on nomme distance d'**Orlicz-Wasserstein**, pour formaliser clairement cette dépendance :

$$d_{OW}(P,Q) = \inf_{\eta} \inf \left\{ a > 0; \int_{\mathbb{R}^d \times \mathbb{R}^d} \tau \left(\frac{z - z'}{a} \right) \eta(dzdz') \leq 1 \right\},$$

l'infimum étant pris sur l'ensemble des probabilités η dont les marginales sont P et Q. Cette distance, qui permet de mesurer la "proximité exponentielle" de deux distributions, semble bien adaptée à l'étude des grandes déviations. Elle nous permet d'exprimer la dépendance de la loi de Z_i au site x_i en termes de continuité de l'application qui à x_i associe la loi de Z_i. Après avoir étudié la distance d'Orlicz-Wasserstein et donné quelques exemples au paragraphe 5.1, on établit le principe de grandes déviations pour la moyenne empirique :

$$\frac{1}{n} \sum_{1}^{n} f(x_i) Z_i$$

au **théorème 5.3** en suivant de près la technique développée au chapitre 3. Malheureusement, le procédé d'approximation est plus complexe dans ce cas et ne nous permet pas (encore) de donner une forme explicite à la fonction de taux.

Chapitre 1

Une extension du théorème de Sanov

Sommaire

Soit $(Y_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.), de loi μ sur un espace mesurable $(\Sigma, \mathcal{A})$. Les mesures empiriques

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

(δ_a est la mesure de Dirac en a) sont des éléments aléatoires de l'ensemble $\mathcal{P}$ des probabilités sur $(\Sigma, \mathcal{A})$.

Le théorème de Sanov

Le théorème de Sanov décrit le comportement asymptotique de

$$\frac{1}{n}\log \mathbb{P}(L_n^Y \in \cdot)$$

lorsque n tend vers l'infini, au moyen d'un principe de grandes déviations (PGD). La fonction de taux associée est donnée, pour tout $\nu \in \mathcal{P}$, par l'entropie relative par rapport à μ :

$$H(\nu \mid \mu) = \int_\Sigma \log\left(\frac{d\nu}{d\mu}\right) d\mu \quad \text{si} \quad \nu \ll \mu$$

et $+\infty$ sinon. Pour ce PGD, la topologie sur $\mathcal{P}$ est $\sigma(\mathcal{P}, B)$. C'est la topologie la moins fine rendant les fonctionnelles $\nu \in \mathcal{P} \mapsto \int_\Sigma f\, d\nu \in \mathbb{R}$ continues pour toute fonction f mesurable et bornée ($\in B$) sur Σ.
Dans ce chapitre, ce PGD est étendu en considérant une topologie plus forte. Les fonctions-test définissant la topologie ne sont plus bornées. On considère maintenant l'espace $\mathcal{L}_\tau$ de toutes les fonctions f qui ont quelques moments exponentiels par rapport à μ :

$$\int_\Sigma e^{a|f|}\, d\mu < \infty, \quad \text{pour un } a > 0. \tag{1.1}$$

L'espace d'états dans lequel évolue la mesure empirique $L_n^Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ n'est plus l'ensemble des probabilités $\mathcal{P}$. C'est l'ensemble plus grand $\mathcal{Q}$ de toutes les formes linéaires positives sur $\mathcal{L}_\tau$ et de masse totale unité. La topologie sur $\mathcal{Q}$ est $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ et la fonction de taux I est de la forme suivante (pour tout $\ell \in \mathcal{Q}$ vérifiant $I(\ell) < \infty$) :

$$I(\ell) = H(\ell^a \mid \mu) + \sup\left\{\langle \ell^s, f\rangle;\ f, \int_\Sigma e^f\, d\mu < \infty\right\},$$

où $\ell = \ell^a + \ell^s$ se décompose de manière unique en la somme d'une mesure de probabilité ℓ^a, qui est absolument continue par rapport à μ, et d'une forme linéaire positive ℓ^s sur $\mathcal{L}_\tau$ qui n'est plus σ-additive. En particulier, sa masse est $\langle \ell^s, \mathbf{1}\rangle = 0$, sans pour autant que $\ell^s \geq 0$ soit nulle ! Pour l'énoncé précis du théorème de Sanov, voir le théorème 1.9 ci-dessous. L'existence de telles formes linéaires est mise en évidence sur un exemple particulier au paragraphe 1.3.

La littérature

Sanov [50] a démontré le PGD pour L_n^Y dans le cadre où $\Sigma = \mathbb{R}$ et pour la topologie de la convergence étroite sur $\mathcal{P}$. Ce théorème a été étendu au cas où Σ est un espace polonais par Donsker et Varadhan [22] et par Bahadur et Zabell [4] dans le cas où on considère la topologie de la convergence étroite. Groeneboom, Oosterhoff et Ruymgaart [33] se sont affranchis de l'hypothèse d'espace polonais et ont obtenu un théorème de Sanov pour la τ-topologie $\sigma(\mathcal{P}, B)$. De Acosta a amélioré ce résultat et simplifié la preuve dans [18]. Dans [14], Csiszár démontre un théorème de Sanov dans un cadre général via une approche alternative fondée sur la projection en information. Dans [25], Eichelsbacher et Schmock considèrent les mesures U-empiriques

$$L_n^{Y,k} = \frac{1}{n\cdots(n-k+1)} \sum_{(i_1,\ldots,i_k)} \delta_{(X_{i_1},\ldots,X_{i_k})},$$

où la sommation est effectuée sur tous les k-uplets $(i_1,\ldots,i_k)$ pour lesquels les indices i_j sont à valeurs $\{1,\ldots,n\}$ et deux à deux distincts. Le cas particulier où $k=1$ est une extension du théorème de Sanov. En effet, la topologie considérée est plus forte que la τ-topologie habituelle. La fonction de taux reste cependant l'entropie relative habituelle. La borne inférieure est obtenue pour une topologie qui est un peu plus faible que $\sigma(\mathcal{P}_\tau,\mathcal{L}_\tau)$. Quant à la borne supérieure, elle est obtenue pour une topologie sur $\mathcal{P}$ qui est plus faible que $\sigma(\mathcal{P}_\tau,\mathcal{M}_\tau)$, où $\mathcal{M}_\tau$ est l'espace des fonctions f qui ont tous leurs moments exponentiels par rapport à μ :

$$\int_\Sigma e^{a|f|}\,d\mu < \infty, \quad \text{pour tout } a > 0 \tag{1.2}$$

et $\mathcal{P}_\tau$ représente l'ensemble des probabilités qui intègrent toutes les fonctions de $\mathcal{L}_\tau$ ou $\mathcal{M}_\tau$. En résumé, notre résultat généralise les travaux d'Eichelsbacher et Schmock dans le cas où $k=1$. En revanche, lorsque $k\geq 2$, le PGD pour $L_n^{Y,k}$ dans [25] n'est pas une conséquence de nos résultats et fait intervenir des techniques de preuves différentes.

1.1 Introduction aux espaces d'Orlicz

Dans ce paragraphe, nous rappelons quelques éléments concernant les espaces d'Orlicz et leurs espaces duals. Il s'agit en effet du cadre fonctionnel que nous utiliserons pour proposer une extension du théorème de Sanov.

1.1.1 Définitions et résultats élémentaires

Une fonction de Young θ est une fonction à valeurs dans $[0,+\infty]$, paire, convexe et satisfaisant :

$$\begin{aligned} \theta(0) &= 0 \\ \lim_{s\to+\infty}\theta(s) &= +\infty \\ \exists s_0>0,\quad \theta(s_0) &< +\infty. \end{aligned}$$

Etant donnée une probabilité μ sur un espace mesurable $(\Sigma, \mathcal{A})$, on considère les espaces vectoriels suivants :

$$\begin{aligned} \mathcal{L}_\theta &= \left\{ f : \Sigma \to \mathbb{R}, f \text{ mesurable}, \ \exists a > 0, \int_\Sigma \theta\Big(\frac{f}{a}\Big)\, d\mu < \infty \right\}, \\ \mathcal{M}_\theta &= \left\{ f : \Sigma \to \mathbb{R}, f \text{ mesurable}, \ \forall a > 0, \int_\Sigma \theta\Big(\frac{f}{a}\Big)\, d\mu < \infty \right\}. \end{aligned}$$

Conformément à l'usage, L_θ et M_θ correspondent aux espaces $\mathcal{L}_\theta$ et $\mathcal{M}_\theta$ quand les fonctions μ-presque sûrement égales sont identifiées. On munit L_θ de la norme de Luxemburg suivante :

$$\|f\|_\theta = \inf \left\{ a > 0, \quad \int_\Sigma \theta\Big(\frac{f}{a}\Big)\, d\mu \leq 1 \right\}. \tag{1.3}$$

$(L_\theta, \|.\|_\theta)$ est alors un espace de Banach appelé espace d'Orlicz associé à θ. De manière évidente, M_θ est un sous-espace de L_θ. Si θ est une fonction partout finie, alors M_θ est l'adhérence de l'ensemble des fonctions en escalier $\sum_{i=1}^n a_i \mathbf{1}_{A_i}$ pour la norme $\|\cdot\|_\theta$. On pourra se référer aux ouvrages [2], [43].

Ce formalisme permet en particulier de définir les espaces $L^p(\mu)$ pour $1 \leq p \leq \infty$. En effet, introduisons, dans le cas où $p < \infty$:

$$\theta_p(x) = \frac{|x|^p}{p}.$$

Ainsi, $\mathcal{L}_{\theta_p} = \{f \ , \int |f|^p \, d\mu < \infty\}$ et $\|f\|_{\theta_p} = (\int |f|^p \, d\mu / p)^{1/p}$, norme équivalente à la norme usuelle $\|f\|_p$. Pour définir $L^\infty(\mu)$, introduisons

$$\theta_\infty(x) = \left\{ \begin{array}{ll} |x| & \text{si } |x| \leq 1 \\ +\infty & \text{sinon.} \end{array} \right. .$$

Dans ce cas, $\mathcal{L}_{\theta_\infty}$ est l'ensemble des fonctions μ-presque sûrement bornées et $\|f\|_{\theta_\infty} = \|f\|_\infty$.

1.1.2 Dualité dans les espaces d'Orlicz

Considérons maintenant θ^*, conjuguée convexe de la fonction θ :

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

Il est immédiat de constater que θ^* est également une fonction de Young. Ainsi, on peut considérer l'espace d'Orlicz L_{θ^*}. Par exemple,

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

où $\frac{1}{p} + \frac{1}{q} = 1$. Les relations entre le dual topologique de L_θ et L_{θ^*} sont explicitées ci-dessous.

Soient $f \in L_\theta$ et $g \in L_{\theta^*}$, alors l'inégalité suivante, dite de Young, généralise l'inégalité de Hölder habituelle :

$$fg \in L^1(\mu) \quad \text{et} \quad \int_\Sigma |fg|\, d\mu \leq 2\|f\|_\theta \|g\|_{\theta^*}. \tag{1.4}$$

On dit qu'une fonction de Young θ satisfait la condition Δ_2 s'il existe une constante $K > 0$ et un réel $s_0 \geq 0$ tels que pour tout $s \geq s_0$ on ait :

$$\theta(2s) \leq K\theta(s)$$

(pour plus de détails, voir par exemple [43], Corollaire 5, p. 77). Ainsi, pour $1 \leq p < \infty$, la fonction θ_p satisfait la condition Δ_2.
Grâce à (1.4), toute fonction $g \in L_{\theta^*}$ définit une forme linéaire continue sur L_θ pour le crochet de dualité $\langle f, g\rangle = \int fg\, d\mu$. Le dual topologique de $(L_\theta, \|.\|_\theta)$ peut, dans le cas général être plus grand que L_{θ^*}. Cependant, le résultat suivant est toujours vrai.

Théorème 1.1 *Considérons une fonction de Young θ partout finie et sa conjuguée convexe θ^*. Le dual topologique de M_θ peut être identifié à L_{θ^*} :*

$$M'_\theta \simeq L_{\theta^*},$$

ce qui signifie qu'à toute forme linéaire continue $\ell \in M'_\theta$, on peut associer une fonction $g_\ell \in L_{\theta^*}$ telle que

$$\langle \ell, f\rangle = \int f g_\ell\, d\mu, \quad \forall f \in M_\theta.$$

Pour une démonstration de ce résultat, voir par exemple [43] ou ([37], Section 4).

Théorème 1.2 *Si θ satisfait la condition Δ_2, alors $L_\theta = M_\theta$ et par conséquent $L'_\theta \simeq L_{\theta^*}$*

On retrouve ainsi, lorsque $\theta_p(x) = \frac{|x|^p}{p}$, le résultat de dualité habituel dans les espaces L^p : $L'_{\theta_p} = L_{\theta_q}$ pour $1 \leq p < \infty$ et $\frac{1}{p} + \frac{1}{q} = 1$.
Comme L_θ est un espace de Riesz (voir par exemple [9] pour la définition), on peut définir, pour toute forme linéaire continue $\ell \in L'_\theta$, sa valeur absolue $|\ell|$.

Définition 1.3 *Etant donnée une forme linéaire continue $\ell \in L'_\theta$, on dit que ℓ est singulière par rapport à μ s'il existe une suite décroissante (au sens de l'inclusion) d'ensembles mesurables (A_k) satisfaisant :*

$$\lim_k \mu(A_k) = 0 \quad et \quad \langle |\ell|, \mathbf{1}_{\Sigma \setminus A_k} \rangle = 0, \quad \forall k \geq 1.$$

On note par L^s_θ le sous-espace de tous les éléments de L'_θ singuliers par rapport à μ.

Théorème 1.4 *Si θ est une fonction de Young partout finie, alors le dual topologique L'_θ de $(L_\theta, \|\cdot\|_\theta)$ est la somme directe :*

$$L'_\theta \simeq (L_{\theta^*} \cdot \mu) \oplus L^s_\theta.$$

De ce fait, toute forme linéaire continue ℓ sur L_θ se décompose de manière unique en $\ell = \ell^a + \ell^s$ où ℓ^a et ℓ^s sont $\|\cdot\|_\theta$-continues , ℓ^a est une mesure absolument continue par rapport à μ dont la dérivée de Radon-Nikodym satisfait $\frac{d\ell^a}{d\mu} \in L_{\theta^*}$ et ℓ^s est une forme linéaire singulière par rapport à μ (qui, en général, n'est pas *une mesure).*

Pour une preuve de ce résultat, on pourra se référer à ([35], théorème 2.2) ou ([37], théorème 4.3). ℓ^a est appelée partie absolument continue de ℓ et ℓ^s sa partie singulière.

Proposition 1.5 *Si θ est une fonction de Young partout finie, alors pour toute fonction $f \in M_\theta$ et $\ell^s \in L^s_\theta$, on a $\langle \ell^s, f \rangle = 0$.*

Pour une preuve de ce résultat, on pourra se référer à [35] ou ([37], Proposition 4.2). Le lemme qui suit est un lemme de structure détaillant la contribution de la partie singulière et de la partie absolument continue de toute forme linéaire continue $\ell \in L'_\theta$.

Lemme 1.6 *Soit $\ell \in L'_\theta$. Alors, pour toute fonction $f \in L_\theta$,*

1. $\lim_{n\to\infty} \langle \ell, f_n \rangle = \langle \ell^a, f \rangle$ *où (f_n) est n'importe quelle suite de fonctions mesurables bornées qui converge ponctuellement vers f et telle que $|f_n| \leq |f|$, pour tout $n \geq 1$,*
2. $\lim_{n\to\infty} \langle \ell, \mathbf{1}_{\{|f|>n\}} f \rangle = \langle \ell^s, f \rangle$.

Preuve. 1. Comme f_n est bornée, $\langle \ell^s, f_n \rangle = 0$ (voir proposition 1.5). Par conséquent

$$\langle \ell, f_n \rangle = \langle \ell^a, f_n \rangle = \int f_n \frac{d\ell^a}{d\mu} \, d\mu,$$

avec $\frac{d\ell^a}{d\mu} \in L_{\theta^*}$. La limite découle du théorème de convergence dominée.
2. On a

$$\langle \ell, \mathbf{1}_{\{|f|>n\}} f \rangle = \langle \ell^s, \mathbf{1}_{\{|f|>n\}} f \rangle + \langle \ell^a, \mathbf{1}_{\{|f|>n\}} f \rangle.$$

Le théorème de convergence dominée implique que $\lim_{n\to\infty} \langle \ell^a, \mathbf{1}_{\{|f|>n\}} f \rangle = 0$ et comme $\mathbf{1}_{\{|f|\leq n\}} f$ est bornée, $\langle \ell^s, f \rangle = \langle \ell^s, \mathbf{1}_{\{|f|>n\}} f \rangle$ (voir proposition 1.5). □

1.1.3 Espaces d'Orlicz et conditions de moments exponentiels

Les espaces d'Orlicz sont particulièrement adaptés à l'expression de conditions portant sur les moments exponentiels de variables aléatoires. En effet, considérons

$$\gamma(s) = e^s - s - 1 \quad \text{et} \quad \tau(s) = \gamma(|s|). \tag{1.5}$$

Alors τ est une fonction de Young et les équivalences suivantes sont immédiates :

$$\begin{aligned} \left(\exists a > 0,\ \int_\Sigma e^{a|f(y)|}\, \mu(dy) < +\infty \right) &\Leftrightarrow \quad f \in \mathcal{L}_\tau, \\ \left(\forall a > 0,\ \int_\Sigma e^{a|f(y)|}\, \mu(dy) < +\infty \right) &\Leftrightarrow \quad f \in \mathcal{M}_\tau. \end{aligned}$$

Dans le cas où $f \in \mathcal{L}_\tau$, on dit que f admet *quelques* moments exponentiels. Si $f \in \mathcal{M}_\tau$, on dit que f a *tous* ses moments exponentiels.

1.2 Une extension du théorème de Sanov

Dans ce paragraphe, nous énonçons et démontrons une extension du théorème de Sanov. Nous nous attachons dans un premier temps à préciser le cadre fonctionnel dans lequel nous allons travailler.

1.2.1 Enoncé du théorème de Sanov étendu

Considérons $\mathcal{L}_\tau$ et son dual algébrique $\mathcal{L}_\tau^*$. On remarquera que les fonctions égales presque sûrement ne sont pas identifiées quand on travaille dans $\mathcal{L}_\tau$. On munit $\mathcal{L}_\tau^*$ de la topologie $*$-faible (notée $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$), c'est-à-dire de la topologie la plus faible rendant continues les applications linéaires $G_f : \ell \in \mathcal{L}_\tau^* \mapsto \langle \ell, f \rangle$,

$f \in \mathcal{L}_\tau$. On munit d'autre part $\mathcal{L}_\tau^*$ d'une tribu $\mathcal{E}$ qui est la plus petite tribu rendant les applications G_f mesurables. On s'attache ici à démontrer un principe de grandes déviations pour la mesure empirique

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

vue comme un élément de $\mathcal{L}_\tau^*$. Les $\{Y_i\}_{i\geq 1}$ sont des variables aléatoires i.i.d. selon la loi μ :

$$Y_i : (\Omega, \mathbb{P}) \to (\Sigma, \mu).$$

Ici, $\mathbb{E}$ dénote l'espérance par rapport à $\mathbb{P}$.

Une identification fonctionnelle. On fera, dans le reste du chapitre, l'identification fonctionnelle suivante :

$$L_{\tau^*} \subset L'_\tau \subset L_\tau^* \subset \mathcal{L}_\tau^*, \tag{1.6}$$

où

- L_{τ^*} est l'espace d'Orlicz associé à la fonction de Young τ^*
- L'_τ (resp. L^*_τ) est le dual topologique (resp. algébrique) de L_τ et
- $\mathcal{L}_\tau^*$ est le dual algébrique de $\mathcal{L}_\tau$.

La première inclusion se comprend de la manière suivante : soit $f \in L_{\tau^*}$, alors $f\mu \in L'_\tau$ au sens où

$$g \in L_\tau \longrightarrow \int gf\, d\mu$$

est une forme linéaire et continue du fait de l'inégalité de Hölder pour les espaces d'Orlicz (voir 1.4). La seconde inclusion est évidente. Pour ce qui est de la troisième inclusion, soit $\ell \in L_\tau^*$. Considérons $\tilde{\ell}$ définie sur $\mathcal{L}_\tau^*$ par $\langle \tilde{\ell}, f\rangle = \langle \ell, \dot{f}\rangle$ où $f \in \mathcal{L}_\tau$ et $\dot{f} \in L_\tau$ est la classe d'équivalence de f par rapport à l'égalité μ-presque sûrement. L'application $\tilde{\ell}$ est bien définie sur $\mathcal{L}_\tau$, elle est linéaire. C'est donc un élément de $\mathcal{L}_\tau^*$.

L'espace d'états. Considérons

$$\mathcal{Q} \stackrel{\triangle}{=} \{\ \ell \in \mathcal{L}_\tau^*;\ \ell \geq 0,\ \langle \ell, \mathbf{1}\rangle = 1\ \}.$$

On le munit de la tribu induite par $\mathcal{E}$ sur $\mathcal{Q}$, notée $\mathcal{E}_\mathcal{Q}$. Remarquons que $\mathcal{L}_\tau$, $\mathcal{Q}$, $\mathcal{E}$ et $\mathcal{E}_\mathcal{Q}$ dépendent de μ.

L'entropie relative étendue. Considérons la fonction suivante, définie sur $\mathcal{L}_\tau^*$:

$$I(\ell) = \begin{cases} \int_\Sigma \log\left(\frac{d\ell^a}{d\mu}\right) d\ell^a + \sup_{f \in D_\mu} \langle \ell^s, f \rangle & \text{si } \ell \in \mathcal{Q} \cap L'_\tau \\ +\infty & \text{sinon} \end{cases},$$

où $\ell = \ell^a + \ell^s$ est la décomposition énoncée au théorème 1.4, et

$$D_\mu = \{f \in \mathcal{L}_\tau;\ \mathbb{E}\, e^{f(Y)} < \infty\}.$$

Remarque 1.7 On notera que, du fait de l'identification (1.6), l'ensemble $\mathcal{Q} \cap L'_\tau$ est bien défini.

Définition 1.8 *On appelle la fonction $I(\ell)$* l'entropie relative étendue *de ℓ par rapport à μ.*

Si $\ell \in L'_\tau$ alors $\ell = \ell^a + \ell^s$ (cf. théorème 1.4) et on notera $I(\ell) = I_a(\ell^a) + I_s(\ell^s)$ où :

$$\begin{aligned} I_a(\ell) &= \int_\Sigma \log\left(\frac{d\ell^a}{d\mu}\right) d\ell^a, \\ I_s(\ell) &= \sup\{\langle \ell^s, f\rangle;\ f, \mathbb{E}\, e^{f(Y)} < \infty\}. \end{aligned}$$

Le principe de grandes déviations. Nous avons désormais tous les éléments pour énoncer le principe de grandes déviations.

Théorème 1.9 *La famille de mesures empiriques $\{L_n^Y\}_{n\geq 1}$ satisfait un principe de grandes déviations dans $\mathcal{Q}$ muni de la tribu $\mathcal{E}_\mathcal{Q}$ et de la topologie $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$, de fonction de taux I. Cela signifie que*

1. *Pour tout fermé mesurable F de $\mathcal{Q}$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n^Y \in F) \leq -\inf_{\ell \in F} I(\ell).$$

2. *Pour tout ouvert mesurable G de $\mathcal{Q}$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n^Y \in G) \geq -\inf_{\ell \in G} I(\ell).$$

De plus, I est convexe et c'est une bonne fonction de taux : ses ensembles de niveau $\{I \leq \alpha\}$ sont compacts.

Le théorème est démontré au paragraphe 1.2.4.

Remarque 1.10 Supposons que $I(\ell) < \infty$, alors $\ell \in L'_\tau$, $\ell^a \geq 0$ et du fait de la proposition 1.5, $\langle \ell^a, \mathbf{1} \rangle = 1$. Ainsi $\frac{d\ell^a}{d\mu}$ est une densité de probabilité et I_a est proche de l'entropie relative habituelle $H(\cdot \mid \mu)$. La différence provient du fait que I_a est définie sur $\mathcal{L}^*_\tau$ alors que H est définie sur $\mathcal{P}$.

Remarque 1.11 La trace de la décomposition de L'_τ en composante absolument continue et en composante singulière (théorème 1.4) sur $\mathcal{Q}$ est $\mathcal{Q} \cap L'_\tau = \Big(L_{\tau^*} \cap \mathcal{P} \Big) \oplus (L^s_\tau \cap \{\ell \geq 0\})$.

Remarque 1.12 On ne peut pas espérer un PGD avec la bonne fonction de taux $H(\cdot \mid \mu)$. En effet, A. Schied a prouvé dans [51] que si la topologie est trop fine (il considère une topologie τ_ϕ où ϕ admet quelques moments exponentiels), alors $\{H \leq \alpha\}$ n'est plus compact. Le même argument est valable dans notre contexte :

$$\begin{aligned} &\{\ell \in \mathcal{Q}; H(\ell^a) \leq \alpha\} \\ &\quad = \{f\mu; f \in L_{\tau^*}, f \geq 0, \int_\Sigma f \, d\mu = 1, \int_\Sigma f \log f \, d\mu \leq \alpha\} + (L^s_\tau \cap \{\ell \geq 0\}), \end{aligned}$$

qui n'est évidemment pas compact. En effet, L^s_τ est un espace vectoriel et $\{\ell \geq 0\}$, un cône. Par conséquent, si $\jmath \in L^s_\tau \cap \{\ell \geq 0\}$, il en va de même pour tout $\alpha \, \jmath$ où $\alpha > 0$. L'ensemble considéré n'est, de fait, pas compact.

On note $\mathcal{P}_{\tau,\mu}$ l'ensemble des probabilités qui intègrent toutes les fonctions de $\mathcal{M}_\tau(\mu)$:

$$\mathcal{P}_{\tau,\mu} = \{\nu \in \mathcal{P}; \int_\Sigma |f| \, d\nu < \infty, \forall f \in \mathcal{M}_\tau\}.$$

On le munit de la tribu $\sigma \left(\nu \mapsto \int_\Sigma f \, d\nu; f \in \mathcal{M}_\tau\right)$ et de la topologie $\sigma(\mathcal{P}_{\tau,\mu}, \mathcal{M}_\tau)$.

Corollaire 1.13 *La famille de mesures empiriques $\{L^Y_n\}_{n\geq 1}$ satisfait un principe de grandes déviations dans $(\mathcal{P}_{\tau,\mu}, \sigma(\mathcal{P}_{\tau,\mu}, \mathcal{M}_\tau))$ avec pour bonne fonction de taux :*

$$H(\nu \mid \mu) = \begin{cases} \int_\Sigma \log \left(\frac{d\nu}{d\mu}\right) d\nu & \text{si } \nu \ll \mu \\ +\infty & \text{sinon.} \end{cases}.$$

Nous retrouvons ainsi un résultat de Eichelsbacher et Schmock ([25], théorème 1.8).

Preuve. C'est une conséquence directe du principe de contraction appliqué à la transformation $\ell \in \mathcal{Q} \to \ell_{|\mathcal{M}_\tau} \in \mathcal{M}^*_\tau$. En effet, la proposition 1.1 implique que $\ell_{|\mathcal{M}_\tau} = \ell^a_{|\mathcal{M}_\tau} = \ell^a$ (ici, la dernière égalité est une identification). De ce fait, $\inf\{I(\ell); \ell_{|\mathcal{M}_\tau} = \nu\} = \inf\{I_a(\nu) + I_s(\ell^s); \ell^a = \nu\} = I_a(\nu) = H(\nu \mid \mu)$. Le caractère continu de la transformation considérée étant évident, on en déduit le résultat. □

1.2.2 Preuve du PGD

Le lemme 1.14 ci-dessous énonce le PGD avec la fonction de taux Θ^* exprimée comme la conjuguée convexe de

$$\Theta(f) = \log \mathbb{E}\, e^{f(Y)} = \log \int_\Sigma e^f \, d\mu \in (-\infty, +\infty], \;\; f \in \mathcal{L}_\tau.$$

Lemme 1.14 *La famille de mesures empiriques $\{L_n^Y\}_{n\geq 1}$ satisfait un PGD (au sens du théorème 1.9) dans $\mathcal{L}_\tau^*$ muni de la tribu $\mathcal{E}$ et de la topologie $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ avec la bonne fonction de taux $\Theta^*(\ell) = \sup_{f\in\mathcal{L}_\tau}\{\langle \ell, f\rangle - \Theta(f)\}$.*

Preuve. La preuve est basée sur une technique de limite projective. Du fait du corollaire 4.6.9 [20] du théorème de Dawson-Gärtner, il est suffisant de vérifier que pour tout $d \geq 1$ et $f_1, \ldots, f_d \in \mathcal{L}_\tau$, $(\langle L_n^Y, f_1\rangle, \ldots, \langle L_n^Y, f_d\rangle)$ satisfait un PGD.
Mais $(\langle L_n^Y, f_1\rangle, \ldots, \langle L_n^Y, f_d\rangle) = \frac{1}{n}\sum_{i=1}^n (f_1(Y_i), \ldots, f_d(Y_i)) = \frac{1}{n}\sum_{i=1}^n \vec{f}(Y_i)$ où $\vec{f}(x) = \big(f_1(x), \ldots, f_d(x)\big)$ est une fonction à valeurs $\mathbb{R}^d$. De ce fait, $\{\vec{f}(Y_i)\}$ apparaît comme une suite de variables aléatoires i.i.d., à valeurs $\mathbb{R}^d$. Comme d'autre part $f_1, \cdots, f_d \in \mathcal{L}_\tau$, $\vec{f}(Y_i)$ admet quelques moments exponentiels. On peut donc appliquer le théorème de Cramér dans $\mathbb{R}^d$ ([20], théorème 6.1.3 et corollaire 6.1.6) et $\frac{1}{n}\sum_{i=1}^n \vec{f}(Y_i)$ satisfait un PGD dans $\mathbb{R}^d$ avec la bonne fonction de taux

$$I_d(x) = \sup_{\lambda\in\mathbb{R}^d}\{\lambda\cdot x - \log \mathbb{E}\, e^{\lambda\cdot\vec{f}(Y)}\}, \qquad x \in \mathbb{R}^d.$$

Le théorème de Dawson-Gärtner entraine que $L_n^Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ satisfait un PGD de bonne fonction de taux (définie pour tout $\ell \in \mathcal{L}_\tau^*$,) donnée par :

$$\sup\{\sum_{i=1}^d \lambda_i \langle \ell, f_i\rangle - \log \mathbb{E}\, e^{\sum \lambda_i f_i(Y)};\; d\geq 1,\; \lambda \in \mathbb{R}^d,\; f_1, \cdots, f_d \in \mathcal{L}_\tau\}$$
$$= \sup_{f\in\mathcal{L}_\tau}\{\langle \ell, f\rangle - \log \mathbb{E}\, e^{f(Y)}\} = \Theta^*(\ell),$$

ce qui est le résultat désiré. □

1.2.3 Identification de la fonction de taux

Nous établissons dans ce paragraphe l'identité $I = \Theta^*$ (proposition 1.18). Les lemmes 1.15 et 1.16 sont des résultats préliminaires utiles dans la suite. On considère

$$J(\ell) = \sup_{f \in \mathcal{L}_\tau} \left\{ \langle \ell - \mu, f \rangle - \int_\Sigma \gamma(f)\, d\mu \right\},$$

où γ est donnée par (1.5).

Lemme 1.15 *Soit $\ell \in \mathcal{L}_\tau^*$, alors*

1. $\Theta^*(\ell) < \infty \Rightarrow \ell \in L_\tau^*$,
2. $J(\ell) \leq \Theta^*(\ell)$,
3. $\Theta^*(\ell) < \infty \Rightarrow \ell \in \mathcal{Q} \cap L'_\tau$.

Preuve. 1. Il suffit de montrer que ℓ est constante sur les classes d'équivalence de $\mathcal{L}_\tau$ ou encore que $\langle \ell, f \rangle = 0$ si $f \in \mathcal{L}_\tau$ est nulle μ-p.p. Considérons donc $f \in \mathcal{L}_\tau$, nulle μ-p.s. Alors, pour tout réel λ, $\Theta(\lambda f) = 0$ et $\lambda \langle \ell, f \rangle \leq \Theta^*(\ell) + \Theta(\lambda f) = \Theta^*(\ell)$. De ce fait, $\Theta^*(\ell) < \infty$ implique que $\langle \ell, f \rangle = 0$. Ainsi ℓ est constante sur les classes d'équivalence et $\ell \in L_\tau^*$.
2. Comme pour tout $t \geq 0$, $\log t \leq t - 1$, on a $-\mathbb{E}\, e^{f(Y)} + 1 \leq -\log \mathbb{E}\, e^{f(Y)}$ et

$$J(\ell) = \sup_{f \in \mathcal{L}_\tau} \left\{ \langle \ell, f \rangle - \int_\Sigma (e^f - 1)\, d\mu \right\} \leq \Theta^*(\ell).$$

3. Soit ℓ tel que $\Theta^*(\ell) < \infty$. D'après 1., $\ell \in L_\tau^*$. Pour tout $f \in L_\tau$,

$$\langle \ell - \mu, f \rangle \leq J(\ell) + \int_\Sigma \gamma(f)\, d\mu.$$

Comme $\gamma(s) \leq \gamma(|s|) = \tau(s)$, on a

$$\langle \ell - \mu, f \rangle \leq J(\ell) + \int_\Sigma \tau(f)\, d\mu.$$

En choisissant $\eta = \pm 1/\|f\|_\tau$ quand $f \neq 0$, la définition même de la norme de Luxemburg (1.3) entraine que $\int_\Sigma \tau(\eta f)\, d\mu = 1$, ce qui implique que

$$|\langle \ell - \mu, f \rangle| \leq (J(\ell) + 1)\|f\|_\tau.$$

Cette inégalité reste vraie avec $\|f\|_\tau = 0$. Du fait que

$$J(\ell) \leq \Theta^*(\ell) < \infty,$$

$\ell - \mu$, élément de L_τ^*, est continue pour la norme $\|\cdot\|_\tau$, soit $\ell - \mu \in L'_\tau$. Remarquons maintenant que μ est un élément de L'_τ :

$$|\int f\, d\mu| \leq 2\|f\|_\tau \|1\|_{\tau^*}, \quad \text{pour tout } f \in L_\tau.$$

Par conséquent, $\ell \in L'_\tau$, *i.e.* ℓ est une forme linéaire continue sur L_τ.
Supposons que $\langle \ell, \mathbf{1}\rangle = a \neq 1$. Alors $\Theta^*(\ell) \geq \langle \ell, \lambda\mathbf{1}\rangle - \log \mathbb{E}\, e^{\lambda\mathbf{1}} = \lambda(a-1)$ qui tend vers l'infini quand λ tend vers l'infini avec le signe de $a-1$ si $a \neq 1$. Par conséquent, $\langle \ell, \mathbf{1}\rangle = 1$ si $\Theta^*(\ell) < \infty$.
Supposons maintenant qu'il existe $f > 0$ avec $\langle \ell, f\rangle < 0$. Si $\lambda \geq 0$, alors $\Theta^*(\ell) \geq \langle \ell, -\lambda f\rangle - \log \mathbb{E}\, e^{-\lambda f} \geq \langle \ell, -\lambda f\rangle$. Cette dernière quantité tend vers $+\infty$ quand λ tend vers $+\infty$. De ce fait, $\Theta^*(\ell) < \infty$ implique que $\ell \geq 0$. Le lemme 1.15 est démontré. □

Lemme 1.16 *Pour tout élément ℓ de L'_τ, $\Theta^*(\ell) = \Theta^*(\ell^a) + \sup\{\langle \ell^s, f\rangle;\ f \in \operatorname{dom}\Theta\}$ où $\operatorname{dom}\Theta = \{f \in \mathcal{L}_\tau, \Theta(f) < \infty\}$ est le domaine effectif de Θ.*

Remarque 1.17 Rappelons que $D_\mu = \{f \in \mathcal{L}_\tau;\ \mathbb{E}\, e^{f(Y)} < \infty\}$. Il est alors immédiat de constater que $\operatorname{dom}\Theta = D_\mu$.

Preuve. Remarquons dans un premier temps que si $\ell \in L'_\tau$,

$$\begin{aligned}\Theta^*(\ell) &= \sup_{f \in \mathcal{L}_\tau} \{\langle \ell, f\rangle - \Theta(\ell)\} \\ &= \sup_{f \in L_\tau} \{\langle \ell, f\rangle - \Theta(\ell)\}\end{aligned}$$

Cela provient du fait que ℓ est constante sur les classes d'équivalence de $\mathcal{L}_\tau$.
On introduit d'abord quelques notations usuelles en analyse convexe. Soit A un sous-ensemble convexe de L_τ et soit ℓ un élément de L'_τ. L'indicatrice convexe de A est

$$\delta(f \mid A) = \begin{cases} 0 & \text{si } f \in A, \\ +\infty & \text{sinon.} \end{cases}$$

Sa conjuguée convexe

$$\delta^*(\ell|A) = \sup_{f \in L_\tau} \{\langle \ell, f\rangle - \delta(f \mid A)\} = \sup_{f \in A} \langle \ell, f\rangle$$

est appelée fonction de support de A. Pour tout $\ell \in L'_\tau$, on a

$$\begin{aligned}\Theta^*(\ell) &= \sup_{f \in L_\tau} \{\langle \ell^a, f\rangle - \Theta(f) + \langle \ell^s, f\rangle - \delta(f|\operatorname{dom}\Theta)\} \\ &\leq \Theta^*(\ell^a) + \delta^*(\ell^s|\operatorname{dom}\Theta).\end{aligned}$$

Pour montrer l'inégalité inverse, considérons u, v éléments de L_τ et soit $n \geq 1$. On considére $f_{u,v,n} = \mathbf{1}_{\{|v|\leq n\}}(-n \vee u \wedge n) + \mathbf{1}_{\{|v|>n\}}v$. Alors,

$$\begin{aligned}\Theta^*(\ell) &\geq \langle \ell, f_{u,v,n}\rangle - \Theta(f_{u,v,n}) \\ &\geq \langle \ell, \mathbf{1}_{\{|v|\leq n\}}(-n \vee u \wedge n)\rangle + \langle \ell, \mathbf{1}_{\{|v|>n\}}v\rangle - \Theta(f_{u,v,n}).\end{aligned}$$

Le lemme de structure (lemme 1.6) entraine que les deux éléments du membres de droite de l'inégalité convergent vers $\langle \ell^a, f\rangle + \langle \ell^s, f\rangle$. Le théorème de convergence dominée entraine la convergence de $\Theta(f_{u,v,n})$ vers $\Theta(f)$. Cela complète la preuve de la proposition. □

Proposition 1.18 *L'identité $\Theta^* = I$ a lieu sur $\mathcal{L}_\tau^*$.*

Preuve. Le lemme 1.15, 3. entraine que le domaine effectif de Θ^* est inclus dans $\mathcal{Q} \cap L'_\tau$. Le lemme 1.16 entraine, quant à lui, que :

$$\forall \ell \in L'_\tau, \quad \Theta^*(\ell) = \Theta^*(\ell^a) + I_s(\ell^s).$$

En tenant compte de la remarque 1.11, il reste à prouver que pour tout $\ell \in \mathcal{P} \cap L_{\tau^*}$, $\Theta^*(\ell) = H(\ell \mid \mu)$. Soit $\ell = h\mu$ appartenant à $\mathcal{P} \cap L_{\tau^*}$, c'est-à-dire $h \in L_{\tau^*}$, $h \geq 0$, $\int_\Sigma h\, d\mu = 1$. Si $f \in \mathrm{dom}\ \Theta$, on obtient par un calcul direct :

$$\Theta(f) = \inf_{\lambda\in\mathbb{R}} \left\{-\lambda - 1 + e^\lambda \int_\Sigma e^f\, d\mu\right\}. \tag{1.7}$$

Par conséquent,

$$\begin{aligned}\Theta^*(h\mu) &= \sup_{f\in\mathrm{dom}\ \Theta} \{\langle h\mu, f\rangle - \Theta(f)\} \\ &= \sup_{\lambda\in\mathbb{R}, f\in\mathrm{dom}\ \Theta} \left\{\langle h\mu, f\rangle + \lambda + 1 - e^\lambda \int_\Sigma e^f\, d\mu\right\} \\ &\overset{(a)}{=} \sup_{\lambda\in\mathbb{R}, f\in\mathrm{dom}\ \Theta} \left\{\langle h\mu, \lambda + f\rangle - \int_\Sigma (e^{\lambda+f} - 1)\, d\mu\right\} \\ &= \sup_{g\in\mathrm{dom}\ \Theta} \left\{\int_\Sigma hg\, d\mu - \int_\Sigma (e^g - 1)\, d\mu\right\} \\ &\overset{(b)}{=} \int_\Sigma (h\log h - h + 1)\, d\mu \\ &\overset{(c)}{=} \int_\Sigma h\log h\, d\mu = H(h\mu \mid \mu),\end{aligned}$$

où (a) et (c) proviennent du fait que μ et $h\mu$ sont des mesures de probabilité et (b) provient d'un résultat général de R. T. Rockafellar ([45], théorème 2), en remarquant que la conjuguée convexe de $e^s - 1$ est $t\log t - t + 1$. Cela complète la preuve de la proposition. □

1.2.4 Fin de la preuve

Preuve. Le PGD dans $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ est démontré au lemme 1.14 avec pour bonne fonction de taux Θ^* qui est convexe en tant que conjuguée. La proposition 1.18 établit l'identité $\Theta^* = I$. Enfin, comme le domaine de I est inclus dans $\mathcal{Q}$ (voir pour cela le lemme 1.15), le PGD a lieu dans $\mathcal{Q}$ ([20], lemme 4.1.5(b)). Ceci achève la preuve du théorème. □

1.3 Application à l'exemple de Csiszár

Dans ce paragraphe, nous mettons en évidence l'existence d'un minimisant de l'entropie relative étendue sous contrainte linéaire avec une partie singulière non nulle. Nous travaillons avec une mesure de probabilité qui apparaît déjà dans ([14], exemple 3.2) et dans ([20], exercice 7.3.11). Soit μ la probabilité définie sur $\Sigma = [0, +\infty)$ par

$$\mu(dy) = c\frac{e^{-y}}{1+y^3}\,dy.$$

Soit $\{Y_i\}$ une suite de variables aléatoires i.i.d. à valeurs $[0, +\infty)$ et de distribution μ. On considère :

$$L_n^Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i} \quad \text{et} \quad \hat{S}_n = \frac{1}{n}\sum_{i=1}^n Y_i.$$

Le théorème de Sanov habituel entraine que L_n^Y satisfait un PGD de bonne fonction de taux $H(\cdot \mid \mu)$ dans $(\mathcal{P}, \sigma(\mathcal{P}, B))$. Le théorème de Cramér assure, lui, que $\hat{S}_n$ satisfait également un PGD de bonne fonction de taux Λ^*, où Λ^* est la conjuguée convexe de $\Lambda(\lambda) = \log \int_{[0,\infty)} c\frac{e^{(\lambda-1)y}}{1+y^3}\,dy$. On peut se demander si le principe de contraction a lieu entre L_n^Y et $\hat{S}_n$. Soit $u : [0, +\infty) \to [0, \infty)$, $u(y) = y$. De cette façon,

$$\langle L_n^Y, u\rangle = \hat{S}_n.$$

Comme u n'est pas bornée, on ne peut pas appliquer le principe de contraction au théorème de Sanov usuel pour obtenir :

$$\inf\{H(\nu \mid \mu),\, \nu \in \mathcal{P},\, \langle \nu, u\rangle = x\} = \Lambda^*(x), \quad x \geq 0. \tag{1.8}$$

Il s'avère néanmoins que cette égalité est vraie (voir proposition 1.20 ci-dessous) mais que l'infimum peut ne pas être atteint dans $\mathcal{P}$ quand x est grand. Constatons d'abord que u appartient à $\mathcal{L}_\tau(\mu)$. En effet,

$$Ee^{|u(x)|} = \int_{[0,\infty)} \frac{c}{1+x^3}\,dx < \infty.$$

De ce fait, $G_u : \mathcal{L}_\tau^*(\mu) \to \mathbb{R}$, $G_u(\ell) = \langle \ell, u \rangle$ est une forme linéaire continue pour la topologie $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$. On pourra noter ici l'avantage d'utiliser la topologie $\sigma(\mathcal{L}_\tau^*, \mathcal{L}_\tau)$ qui est plus "riche" que la topologie $\sigma(\mathcal{P}, B)$. Le théorème 1.9 et le principe de contraction assurent cette fois-ci que $G_u(L_n^Y) = \frac{1}{n}\sum_{i=1}^n Y_i = \hat{S}_n$ satisfait un PGD de bonne fonction de taux

$$I'(x) = \inf\{I(\ell); \ell \in \mathcal{Q}; \langle \ell, u \rangle = x\}.$$

Comme I est une bonne fonction de taux, il existe au moins un élément minimisant $\ell_x \in \mathcal{Q}$ tel que $I(\ell_x) = I'(x)$. En outre, du fait de l'unicité de la fonction de taux associée au PGD de $\hat{S}_n$, on peut identifier I' et Λ^* pour obtenir, pour tout $x \in [0, +\infty)$:

$$\Lambda^*(x) = \inf\{I(\ell); \ell \in \mathcal{Q}; \langle \ell, u \rangle = x\} = I(\ell_x), \tag{1.9}$$

pour un minimisant $\ell_x \in \mathcal{Q}$ vérifiant $\langle \ell_x, u \rangle = x$.

Proposition 1.19 *On note $x^* = \Lambda'(1)$.*

1. *Pour tout $0 < x < x^*$, il existe un unique minimisant ℓ_x réalisant l'égalité (1.9). Il est donné par $\ell_x = \nu_x$ où*

$$\nu_x(dy) = \exp(\lambda_x y - \Lambda(\lambda_x))\mu(dy)$$

et λ_x est l'unique solution de $\Lambda'(\lambda) = x$.

2. *Pour $x = x^*$, l'énoncé précédent reste reste vrai avec $\lambda_{x^*} = 1$, $\Lambda'(1^-) = x^*$ et $\nu_{x^*} = \nu_*$ qui est donnée par :*

$$\nu_*(dy) = e^{y - \Lambda(1)}\,\mu(dy) = \frac{c'}{1+y^3}\,dy.$$

3. *Pour tout $x \geq x^*$ et tout minimisant ℓ_x de l'équation (1.9), on a $\ell_x^a = \nu_*$. De plus, $\langle \nu_*, u \rangle = x^*$, $\langle \ell_x^s, u \rangle = x - x^*$ et*

$$I(\ell_x) = \int_{[0,\infty)} \log\left(\frac{d\nu_*}{d\mu}\right) d\nu_* + I_s(\ell_x^s), \tag{1.10}$$

avec $\int_{[0,\infty)} \log\left(\frac{d\nu_*}{d\mu}\right) d\nu_* = \Lambda^*(x^*)$ et $I_s(\ell_x^s) = x - x^*$.

Cette proposition signifie que lorsque $x > x^*$, les minimisants de (1.9) ne peuvent pas être des probabilités. La contribution de la partie absolument continue est stoppée à x^* : $\langle \nu_*, u\rangle = x^*$. C'est la partie singulière qui permet de "combler l'espace" entre x^* et x : $\langle \ell^s_x, u\rangle = x - x^*$. De plus, la contribution de la partie singulière apparaît dans la fonction de taux (voir (1.10)). Finalement, le fait que $I_s(\ell^s) = \sup_{f\in D_\mu}\langle \ell^s, f\rangle > 0$ implique que cette partie singulière est non nulle lorsque $x > x^*$.

Preuve de la proposition 1.19. Preuve de 1. et 2. Nous avons clairement, pour tout $0 < x \leq x^*$, $\langle \nu_x, u\rangle = x$. Soit ℓ telle que $\langle u, \ell\rangle = x$ et (sans perte de généralité) $I(\ell) < \infty$. Alors, $\ell = \ell^a + \ell^s$ avec $\ell^a \in \mathcal{P}$ et $\ell^s \geq 0$. On note : $\langle \ell^a, u\rangle = x'$. Nous avons alors $\langle \ell^s, u\rangle = x - x' \geq 0$ et

$$\begin{aligned}
I(\ell) - I(\nu_x) &= H(\ell^a \mid \mu) - H(\nu_x \mid \mu) + I_s(\ell^s) \\
&= H(\ell^a \mid \nu_x) + \int \log\left(\frac{d\nu_x}{d\mu}\right) d(\ell^a - \nu_x) + I_s(\ell^s) \\
&= H(\ell^a \mid \nu_x) + I_s(\ell^s) - \lambda_x(x - x') \\
&\overset{(a)}{\geq} H(\ell^a \mid \nu_x) + \langle \ell^s, u\rangle - \lambda_x(x - x') \\
&\overset{(b)}{\geq} H(\ell^a \mid \nu_x) \\
&\geq 0,
\end{aligned}$$

où (a) découle de $I_s(\ell^s) = \sup\{\langle \ell^s, v\rangle; v \in D_\mu\} \geq \langle \ell^s, u\rangle$ et (b) découle de $\langle \ell^s, u\rangle - \lambda_x(x - x') = (1 - \lambda_x)(x - x^*) \geq 0$.

Pour que l'égalité ait lieu, il est nécessaire que $\ell^a = \nu_x$. Ainsi $I_s(\ell^s) = 0$, ce qui implique que $\ell^s = 0$. Finalement, ν_x est l'unique minimisant de (1.9).

Preuve de 3. Pour tout $\lambda \leq 1$, $\Lambda(\lambda)$ et $\Lambda'(\lambda)$ sont finis, alors que $\Lambda(\lambda) = \infty$ si $\lambda > 1$. Comme $\Lambda'(1^-) = x^*$ est finie, Λ n'est pas escarpée. Des arguments de convexité habituels entrainent que $\Lambda^*(x^*) = x^* - \Lambda(1)$ et un calcul élémentaire donne

$$\Lambda^*(x) = \Lambda^*(x^*) + x - x^*,\ x \geq x^*, \tag{1.11}$$

(on pourra se reporter à l'annexe A). Le reste de la preuve est divisé en trois étapes :

Etape 1. Soit ℓ_x (resp. ℓ_y) un minimisant de $I'(x)$ (resp. $I'(y)$) : $I(\ell_x) = I'(x) = \inf\{I(\ell), \langle \ell, u\rangle = x\}$. Nous allons montrer que pour tout $0 \leq \alpha, \beta \leq 1$, $\alpha + \beta = 1$, l'égalité suivante a lieu :

$$\forall x, y \geq x^*, \quad I(\alpha\ell_x + \beta\ell_y) = \alpha I(\ell_x) + \beta I(\ell_y). \tag{1.12}$$

Par définition de ℓ_x et ℓ_y, on a $\langle \ell_x, u\rangle = x$ et $\langle \ell_y, u\rangle = y$. Par ailleurs, (1.11) entraine $I(\ell_x) = \Lambda^*(x) = (x - x^*) + \Lambda^*(x^*)$. De manière similaire, $I(\ell_y) =$

$(y - x^*) + \Lambda^*(x^*)$. La convexité de I implique que

$$I(\alpha\ell_x + \beta\ell_y) \leq \alpha I(\ell_x) + \beta I(\ell_y) = \Lambda^*(x^*) + \alpha x + \beta y - x^* = I'(\alpha x + \beta y).$$

Mais $I'(\alpha x + \beta y) = \inf\{I(\ell); \ell \in L'_\tau; \langle \ell, u\rangle = \alpha x + \beta y\}$ et $\alpha\ell_x + \beta\ell_y$ satisfait la contrainte $\langle \alpha\ell_x + \beta\ell_y, u\rangle = \alpha x + \beta y$. Ainsi $I'(\alpha x + \beta y) \leq I(\alpha\ell_x + \beta\ell_y)$ et (1.12) a lieu.

ETAPE 2. On montre maintenant que pour tous $x, y \geq x^*$ et ℓ_x (resp. ℓ_y) minimisant de $I'(x)$ (resp. $I'(y)$), on a $\ell_x^a = \ell_y^a \stackrel{\triangle}{=} \nu$ où $\ell_x = \ell_x^a + \ell_x^s$ resp. $\ell_y = \ell_y^a + \ell_y^s$).
Par définition de I, on a :

$$\begin{array}{lcl} I(\alpha\ell_x + \beta\ell_y) & = & I_a(\alpha\ell_x^a + \beta\ell_y^a) + I_s(\alpha\ell_x^s + \beta\ell_y^s) \\ \alpha I(\ell_x) + \beta I(\ell_y) & = & \alpha I_a(\ell_x^a) + \beta I_a(\ell_y^a) + \alpha I_s(\ell_x^s) + \beta I_s(\ell_y^s) \end{array}.$$

La convexité de I_a et I entraine que

$$\begin{cases} I_a(\alpha\ell_x^a + \beta\ell_y^a) & \leq \alpha I_a(\ell_x^a) + \beta I_a(\ell_y^a) \\ I(\alpha\ell_x^s + \beta\ell_y^s) & \leq \alpha I(\ell_x^s) + \beta I(\ell_y^s) \end{cases}.$$

Compte tenu de (1.12), $I(\alpha\ell_x+\beta\ell_y) = \alpha I(\ell_x)+\beta I(\ell_y)$. Par conséquent, l'égalité doit avoir lieu dans les deux inéqualités précédentes. Mais du fait de la stricte convexité de I_a, l'égalité $I_a(\alpha\ell_x^a + \beta\ell_y^a) = \alpha I_a(\ell_x^a) + \beta I_a(\ell_y^a)$ implique que $\ell_x^a = \ell_y^a = \nu$.

ETAPE 3. Montrons maintenant que pour tout $x \geq x^*$, $I_a(\ell_x^a) = \Lambda^*(x^*)$ et $I_s(\ell_x^s) = x - x^*$.
Considérant $\nu_*(dy) = e^{y-\Lambda(1)}\mu(dy)$, on montre que $\langle \nu_*, u\rangle = x^*$, $I(\nu_*) = I_a(\nu_*) = \Lambda^*(x^*)$. Ainsi, ν_* satisfait (1.9) en x^*. Par conséquent, $\nu_* = \nu$ et pour tout $x \geq x^*$, $\ell_x^a = \nu_*$. Il s'ensuit que $I_a(\ell_x^a) = \Lambda^*(x^*)$ et $I(\ell_x^s) = x - x^*$. Ceci achève la démonstration de la proposition. □

Proposition 1.20 *L'égalité (1.8) :*

$$\inf\{H(\nu \mid \mu),\, \nu \in \mathcal{P},\, \langle \nu, u\rangle = x\} = \Lambda^*(x)$$

a lieu pour tout $x \geq 0$.

Preuve. Dans le cas où $x = 0$, $\Lambda^*(0) = \infty$. Comme il n'existe pas de $\nu \in \mathcal{P}$ tel que $\nu \ll \mu$ et $\langle \nu, u\rangle = 0$, l'égalité est établie. Si $0 < x \leq x^*$, l'égalité (1.8) est une conséquence de la proposition 1.19 1. et 2. On considère maintenant le cas où $x > x^*$. Remarquons dans un premier temps que

$$\begin{array}{ll} \multicolumn{2}{l}{\inf\{H(\nu \mid \mu),\, \nu \in \mathcal{P},\, \langle \nu, u\rangle = x\}} \\ = & \inf\{I(\nu),\, \nu \in \mathcal{P},\, \langle \nu, u\rangle = x\} \\ \geq & \inf\{I(\ell),\ \ell \in \mathcal{Q},\, \langle \ell, u\rangle = x\} = \Lambda^*(x). \end{array}$$

Il nous suffit donc d'exhiber une suite minimisante de probabilités (ν_n) vérifiant $\langle \nu_n, u\rangle = x$ et telle que $\lim_{n\to\infty} H(\nu_n \mid \mu) = \Lambda^*(x)$. On la prend de la forme suivante :

$$\nu_n = (1-\frac{1}{n})\nu_* + \frac{1}{n}\frac{\mathbf{1}_{I_n}}{\mu(I_n)}\mu,$$

où I_n doit être déterminé de sorte que $\langle \nu_n, u\rangle = x$. Comme $\langle \nu_*, u\rangle = x^*$, I_n doit vérifier :

$$\frac{\int_{I_n} u\,d\mu}{\mu(I_n)} = x^* + n(x-x^*).$$

Considérons $I_n(t) = [x^* + n(x-x^*) - t, x^* + n(x-x^*) + 1]$ et

$$\phi(t) = \frac{\int_{I_n(t)} u\,d\mu}{\mu(I_n(t))} = \frac{\int_{I_n(t)} yf(y)\,dy}{\mu(I_n(t))},$$

où f est la densité de μ. Alors

$$\phi(0) > x^* + n(x-x^*) \quad \text{et} \quad \phi(1) < x^* + n(x-x^*).$$

La première inégalité est élémentaire et la seconde s'obtient par une intégration par parties où on introduit F, fonction de répartition de μ (on notera $z_n = x^* + n(x-x^*)$) :

$$\begin{aligned}\int_{z_n-1}^{z_n+1} yf(y)\,dy &= z_n[F(z_n+1) - F(z_n-1)] \\ &\quad +F(z_n+1) + F(z_n-1) - \int_{z_n-1}^{z_n+1} F(y)\,dy, \\ &< z_n[F(z_n+1) - F(z_n-1)] = (x^* + n(x-x^*))\mu(I_n(1)).\end{aligned}$$

Du fait que ϕ est continue, il existe $\alpha_n \in [0,1]$ tel que $\phi(\alpha_n) = x^* + n(x-x^*)$ et on pose $I_n \stackrel{\triangle}{=} I_n(\alpha_n)$. L'égalité (1.9) indique que $H(\nu_n \mid \mu) = I(\nu_n) \geq \Lambda^*(x)$. Nous avons également

$$\begin{aligned}H(\nu_n \mid \mu) &= \int_{[0,\infty)\setminus I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n + \int_{I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n \\ &\leq H(\nu_* \mid \mu) + \int_{I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n = \Lambda^*(x^*) + \int_{I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n,\end{aligned}$$

où la dernière égalité provient de l'identité $H(\nu_* \mid \mu) = \Lambda^*(x^*)$. Notons enfin que

$$\frac{d\nu_n}{d\mu}(y) = e^{y-\Lambda(1)}(1+\frac{1}{n}) + \frac{1_{I_n}(y)}{n\mu(I_n)}. \tag{1.13}$$

On utilisera l'inégalité suivante :

$$(a+b)\log(a+b) \leq (a\log a)\left(1+\frac{b}{a}\right)^2, \quad \forall a \geq e, b > 0, \tag{1.14}$$

Du fait de (1.13) et (1.14), on obtient

$$\begin{aligned}
&\int_{I_n} \log\left(\frac{d\nu_n}{d\mu}\right) d\nu_n \\
&\leq \int_{I_n} \frac{1}{n\mu(I_n)} \log\left(\frac{1}{n\mu(I_n)}\right)\left(1+e^{y-\Lambda(1)}(1+\frac{1}{n})n\mu(I_n)\right)^2 d\mu(y)
\end{aligned}$$

Remarquons maintenant que :

- si $y \in I_n$, alors $e^{y-\Lambda(1)}(1+\frac{1}{n})n\mu(I_n) \leq \epsilon_n \xrightarrow[n\to\infty]{} 0$,
- d'autre part, $\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{n\mu(I_n)} = x - x^*$.

Par conséquent,

$$\lim_{n\to\infty} \int_{I_n} \frac{1}{n\mu(I_n)} \log\left(\frac{1}{n\mu(I_n)}\right)\left(1+e^{y-\Lambda(1)}(1+\frac{1}{n})n\mu(I_n)\right)^2 d\mu(y) = x - x^*,$$

ce qui achève la démonstration. □

Chapitre 2

Autour du principe de conditionnement de Gibbs

Sommaire

Le principe de conditionnement de Gibbs

Le principe de conditionnement de Gibbs (PCG) décrit le comportement asymptotique, lorsque n tend vers l'infini, de la loi de k particules $Y_1, \dots, Y_k$ sous la contrainte que L_n^Y appartienne à un certain ensemble de probabilités A_0 avec $\mathbb{P}(L_n^Y \in A_0)$ strictement positif pour tout $n \geq 1$. Typiquement, le résultat attendu est de la forme :

$$\lim_{n\to\infty} \mathbb{P}\Big((Y_1, \dots, Y_k) \in \cdot \mid L_n^Y \in A_0\Big) = \nu_*^k(\cdot), \tag{2.1}$$

où ν_* minimise $\nu \mapsto H(\nu|\mu)$ sous la contrainte $\nu \in A_0$. Cette question est importante en mécanique statistique. En effet, la loi conditionnelle s'interprète comme une distribution canonique dans ce cadre-là. Un ensemble de conditionnement typique est :

$$A_0 = \left\{\nu \in \mathcal{P}; \int_\Sigma \varphi \, d\nu = E\right\}, \tag{2.2}$$

où φ est une fonction d'énergie.
Les liens existant entre le théorème de Sanov et le Principe de Conditionnement de Gibbs sont bien connus. On pourra se référer aux travaux de Csiszár [13, 14], Van Campenhout et Cover [11] et Stroock et Zeitouni [52]. Comme dans [14] et dans [52], nous ne traiterons pas le cas important et difficile où $\mathbb{P}(L_n^Y \in A_0) = 0$ pour $n \geq 1$. Nous nous inspirons de [52] en introduisant des grossissements A_δ de $A_0 = \cap_{\delta>0} A_\delta$ de manière à ce que $\mathbb{P}(L_n^Y \in A_\delta) > 0$ pour tout $n \geq 1$ et $\delta > 0$.
Les résultats de ce chapitre sont basés sur l'extension du théorème de Sanov (théorème 1.9) que nous avons établie au chapitre précédent. Notre théorème étendu nous permet de présenter une preuve plus directe du principe de conditionnement de Gibbs. Cela est principalement dû au fait que nous disposons de la borne inférieure du PGD pour une plus grande gamme d'ensembles (voir le théorème 1.9) que dans le cas du théorème de Sanov classique. Cela nous permet également d'affaiblir de manière notable les hypothèses sous lesquelles le PCG a lieu. On considère d'une part des ensembles de conditionnement A_0 (voir 2.2) qui sont naturellement construits à partir de fonctions d'énergie φ appartenant à $\mathcal{L}_\tau$, c'est à dire telles que

$$Ee^{\alpha|\varphi(Z)|} < \infty \quad \text{pour un } \alpha > 0.$$

On s'affranchit d'autre part d'une hypothèse du type loi des grands nombres sous-jacente :

$$\lim_{n\to\infty} \nu_*^n \{L_n^Y \in A_\delta\} = 1 \quad \text{pour tout } \delta > 0, \tag{2.3}$$

présente dans [52]. Une conséquence intéressante est que le PCG a lieu dans des situations où la "pression"

$$\beta \mapsto \log \int_\Sigma e^{\beta\varphi}\, d\mu$$

n'est pas escarpée (*steep*) (voir par exemple l'exemple de Csiszár au paragraphe 2.4). Le principe de conditionnement de Gibbs suivant :

$$\lim_{\delta\to 0} \lim_{n\to\infty} \mathbb{P}\Big((Y_1, \dots, Y_k) \in \cdot \mid L_n^Y \in A_\delta\Big) = \nu_*^k(\cdot).$$

est énoncé au théorème 2.5, qui est le principal résultat de cette section.

Rappelons que sous les hypothèses classiques, la loi limite ν^* apparaissant dans (2.1) minimise l'entropie relative $H(\cdot|\mu)$ sous une certaine contrainte. Dans notre cas, la fonctionnelle n'est plus H mais

$$I(\ell) = H(\ell^a|\mu) + I_s(\ell^s).$$

Il apparaît néanmoins que sous une contrainte convexe, l'ensemble $\mathcal{M}$ des minimisants de I est de la forme

$$\mathcal{M} = \nu^* + \mathcal{S},$$

où $\mathcal{S}$ est une famille de formes linéaires singulières qui ne sont pas des mesures (voir le paragraphe 2.2). C'est la loi ν^*, commune à tous les minimisants, qui jouera le rôle de loi limite dans le PCG. Notre présentation s'inspire fortement du paragraphe 7.3 de l'ouvrage de Dembo et Zeitouni [20].

Littérature

Comme nous l'avons précédemment mentionné, le principe de conditionnement de Gibbs a été étudié par Csiszár dans [14], Van Campenhout et Cover [11] et Stroock et Zeitouni [52]. Une présentation didactique de la dernière approche est disponible dans le livre de Dembo et Zeitouni ([20], Section 7.3). Dans [7], Bolthausen et Schmock ont prouvé un PCG pour les mesures d'occupation de chaines de Markov uniformément ergodiques. Se basant sur [14], Aboulalaâ [1] a obtenu un PCG pour les mesures empiriques de processus de sauts markoviens. Avec un PGD établi pour $L_n^{Y,k}$, Eichelsbacher et Schmock [25] obtiennent un PCG, via l'approche de [52]. Ils l'obtiennent pour k particules avec une fonction d'énergie φ appartenant à $\mathcal{M}_\tau$ (voir (2.2) et (1.2)) et pour une topologie sur $\mathcal{P}_\tau$ qui est légèrement plus faible que $\sigma(\mathcal{P}_\tau, \mathcal{M}_\tau)$.

Dans le cadre, différent, de mesures empiriques intervenant dans des problèmes de maximum d'entropie en moyenne, Gamboa et Gassiat [29] établissent des résultats de convergence asymptotique de lois conditionnées sans loi des grands nombres sous-jacente. Enfin, notons que Csiszár [14] a obtenu des résultats alternatifs via une puissante approche entropique. En particulier, il prouve la convergence en information des lois conditionnelles, ce qui implique leur convergence en variation. Csiszár introduit la notion de I-projection, ce qui lui permet d'établir des PCG basés sur des fonctions d'énergie satisfaisant (1.1).

2.1 Notations et hypothèses

Ce chapitre reprend les notations introduites aux paragraphes 1.1 et 1.2 du chapitre précédent. Ainsi considérons, comme précédemment $L_n^Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i} \in \mathcal{L}_\tau^*$ où $\{Y_i\}_{i\geq 1}$ est une suite de variables aléatoires i.i.d. de loi μ et à valeur Σ. On appelle μ^n la mesure produit induite par μ sur Σ^n et Q_n la probabilité induite par L_n^Y sur ($\mathcal{Q}$, $\mathcal{E}_\mathcal{Q}$), où $\mathcal{Q}$ est muni de la topologie $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ et de la tribu $\mathcal{E}_\mathcal{Q}$:

$$Q_n(A) = \mu^n\{L_n^y \in A\}, \quad A \in \mathcal{E}_\mathcal{Q}.$$

On s'intéresse ici au comportement-limite de la distribution de $(Y_1, \cdots, Y_k)$ sous le conditionnement $\{L_n^Y \in A_\delta\}$ lorsque $n \to \infty$ puis $\delta \to 0$. On note cette distribution

$$\mu^n_{Y^k|A_\delta}(\cdot) = \mu^n\Big((y_1, \dots, y_k) \in \cdot \mid L_n^y \in A_\delta\Big). \tag{2.4}$$

Dans le cas où $k = 1$, on écrit simplement $\mu^n_{Y|A_\delta}$. Nous suivons ici Stroock et Zeitouni dans [52] en considérant l'ensemble de contraintes $\{L_n^Y \in A_\delta\}$, où A_δ est un grossissement de A_0, plutôt que l'ensemble $\{L_n^Y \in A_0\}$. Du fait de l'hypothèse (H-1) ci-dessous, A_δ doit satisfaire $Q_n(A_\delta) > 0$ alors que A_0 peut être un ensemble Q_n-négligeable.

On adoptera, dans la suite de ce chapitre, les conventions suivantes :

$$\Gamma_n = \{L_n^Y \in \Gamma\}, \quad A_{\delta,n} = \{L_n^Y \in A_\delta\},$$

$\mathring{A}$ est l'intérieur de A pour la topologie $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ et

$$I(A) = \inf\{I(\ell); \ell \in A\} \text{ pour tout } A \subset \mathcal{Q}.$$

Hypothèse H-0 *L'ensemble A_0 peut s'écrire $A_0 = \cap_{\delta>0} A_\delta$, où $(A_\delta)_{\delta>0}$ est une famille décroissante d'ensembles mesurables, fermés pour la topologie $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ et satisfaisant*

$$I(A_\delta^\circ) \leq I(A_0), \ \textit{pour tout } \delta > 0. \tag{2.5}$$

Voici deux cas importants où (H-0) est satisfaite :

1. $A_0 \subset A_\delta^\circ$, pour tout $\delta > 0$
2. $A_\delta = A_0$ pour tout $\delta > 0$, et $I(A_0^\circ) = I(A_0)$.

Remarque 2.1 La topologie $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$ et la tribu $\mathcal{E}$ qui apparaissent dans l'énoncé de l'hypothèse (H-0) sont toutes les deux plus "riches" que la topologie $\sigma(\mathcal{P}, B)$, qui est la topologie la moins fine rendant continues les applications

$$G_f : \mathcal{P} \to \mathbb{R}, \quad \mu \mapsto \langle f, \mu \rangle,$$

f étant mesurable et bornée et la tribu $\mathcal{B}^{cy}$, qui est la tribu la moins fine rendant ces applications mesurables. De ce fait, plus d'ensembles ouverts mesurables sont disponibles. Pour illustrer cette remarque, considérons l'exemple suivant :

$$A_\delta = \{\ell \in \mathcal{Q};\ |\langle \ell, u \rangle - 1| \leq \delta\}, \qquad A_0 = \{\ell \in \mathcal{Q};\ \langle \ell, u \rangle = 1\},$$

où u satisfait la condition (1.1). Cette famille d'ensembles satisfait l'hypothèse (H-0).

L'hypothèse suivante correspond à l'hypothèse (A-1) dans ([20], section 7.3).

Hypothèse H-1 $I(A_0) < +\infty$ *et pour tous* $\delta > 0$, $n \geq 1$, $Q_n(A_\delta) > 0$.

Remarque 2.2 L'équation (2.3), qui fait partie des hypothèses dans [20] (voir également [25], Condition 1.16), présuppose une loi des grands nombres sous la loi minimisante ν_*. On remarquera que cela n'est pas nécessaire dans notre approche. En effet, l'hypothèse (H-0) nous permet d'appliquer directement la borne inférieure du théorème de Sanov (voir la preuve du lemme 2.6 ci-dessous), ce qui rend superflu toute loi des grands nombres sous-jacente. Il existe de plus des cas où (2.3) n'est pas vérifée alors que (H-1) est satisfaite (voir la remarque 2.9 ci-dessous).

2.2 Contraintes convexes

On dénote l'ensemble des minimisants par :

$$\mathcal{M} \stackrel{\triangle}{=} \{\ell \in A_0;\ I(\ell) = I(A_0)\}.$$

Le résultat suivant établit que $\mathcal{M}$ a une forme particulière quand la contrainte A_0 est convexe.

Lemme 2.3 *Supposons que* A_0 *soit convexe, alors*

$$\mathcal{M} = \nu_* + \mathcal{S},$$

où ν_* *est une probabilité et* $\mathcal{S}$ *est un ensemble de parties singulières qui ne sont pas des mesures. Autrement dit, si* $\ell \in \mathcal{M}$, *alors* $\ell = \nu_* + \ell^s$ *où* $\ell^s \in \mathcal{S}$ *est la partie singulière de* ℓ *et* ν_*, *sa partie absolument continue, commune à tous les éléments de* $\mathcal{M}$.

Remarque 2.4 Dans ce cas, on peut montrer que ν_* est la I-projection généralisée de μ sur l'ensemble des contraintes A_0 au sens de Csiszár (voir [14]).

Preuve du lemme 2.3. Soient ℓ et $\tilde{\ell}$ deux éléments de $\mathcal{M}$. Du fait de la convexité de A_0 et de I, on a

$$I(A_0) \leq I(\alpha\ell + \beta\tilde{\ell}) \leq \alpha I(\ell) + \beta I(\tilde{\ell}) = I(A_0)$$

pour tous $\alpha, \beta \geq 0$ tel que $\alpha + \beta = 1$. De manière similaire, comme I_a et I_s sont convexes, on obtient :

$$\left\{ \begin{array}{lcl} I_a(\alpha\ell^a + \beta\tilde{\ell}^a) & \leq & \alpha I_a(\ell^a) + \beta I_a(\tilde{\ell}^a) \\ I_s(\alpha\ell^s + \beta\tilde{\ell}^s) & \leq & \alpha I_s(\ell^s) + \beta I_s(\tilde{\ell}^s) \end{array} \right. .$$

Supposons qu'au moins une de ces inégalités soit stricte. Nous obtiendrions en sommant

$$I(A_0) < \alpha I(\ell) + \beta I(\tilde{\ell}) = I(A_0),$$

ce qui est absurde. Par suite, $I_a(\alpha\ell^a+\beta\tilde{\ell}^a) = \alpha I_a(\ell^a)+\beta I_a(\tilde{\ell}^a)$ et $I_s(\alpha\ell^s+\beta\tilde{\ell}^s) = \alpha I_s(\ell^s) + \beta I_s(\tilde{\ell}^s)$. Comme I_a est strictement convexe, on obtient $\ell^a = \tilde{\ell}^a \stackrel{\triangle}{=} \nu_*$. Comme I_s n'est pas strictement convexe, ℓ^s et $\tilde{\ell}^s$ ne sont pas forcément égales. □

On verra au paragraphe 2.4 que dans le cadre de l'exemple de Csiszár, $\mathcal{M} = \nu_* + \mathcal{S}$ où $\mathcal{S}$ contient plusieurs éléments distincts.

2.3 Convergence de $\mu^n_{Y^k|A_\delta}$ vers une probabilité

On suppose dans ce paragraphe que Σ est un espace métrique séparable muni de sa tribu borélienne. On note $C_b(\Sigma^k)$ l'ensemble des fonctions continues et bornées sur Σ^k. Le théorème 2.5 ci-dessous est l'équivalent du corollaire 7.3.5 dans [20]. C'est un corollaire du lemme 2.6.

Théorème 2.5 *Supposons que les hypothèses (H-0) et (H-1) sont vérifiées, que Σ est un espace métrique séparable et que la contrainte A_0 est convexe. Alors, pour toute fonction $f \in C_b(\Sigma^k)$, on a :*

$$\langle \mu^n_{Y^k|A_\delta}, f\rangle \to \langle \nu^k_*, f\rangle$$

lorsque $n \to \infty$ puis $\delta \to 0$. Ici, ν_* est la partie absolument continue commune à tous les éléments de $\mathcal{M}$ (voir le lemme 2.3).

Le théorème 2.5 améliore le corollaire 7.3.5 de [20] dans deux directions :

1. Les ensembles de contrainte A_δ peuvent être construits à partir de fonctions qui n'ont pas tous leurs moments exponentiels, par exemple :

   $$A_\delta = \{\ell \in \mathcal{Q}; |\langle \ell, u\rangle - 1| \le \delta\} \quad \text{avec} \quad u \in \mathcal{L}_\tau.$$

   De telles fonctions peuvent croître assez rapidement.

2. Il a été noté précédemment que le jeu d'hypothèses (H-0) et (H-1) ne nécessite pas de loi des grands nombres sous-jacente. Pour une illustration du gain, voir le paragraphe 2.4 ci-dessous et en particulier la remarque 2.9.

On aura besoin du lemme suivant qui joue le rôle du théorème 7.3.3 dans [20].

Lemme 2.6 *Supposons que les hypothèses (H-0) et (H-1) sont vérifiées. Alors $\mathcal{M}$ est un ensemble compact non vide pour la topologie $\sigma(\mathcal{Q}, \mathcal{L}_\tau)$. $\mathcal{M}$ est inclus dans $\mathcal{Q}$ et pour tout $\Gamma \in \mathcal{E}$ vérifiant $\mathcal{M} \subset \overset{\circ}{\Gamma}$, on a :*

$$\limsup_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mu^n\big(L_n^y \notin \Gamma | L_n^y \in A_\delta\big) < 0.$$

Preuve. Des arguments standards entrainent que $\mathcal{M} \neq \emptyset$ et que $\mathcal{M} = A_0 \cap \{I \leq I(A_0)\}$. Comme I est une bonne fonction de taux, $\{I \leq I(A_0)\}$ est un ensemble compact et A_0 étant fermé, (voir (H-0)), il s'ensuit que $\mathcal{M}$ est compact. Enfin, comme $(A_\delta)_{\delta>0}$ est une famille décroissante d'ensembles mesurables, on obtient l'inégalité suivante :

$$\begin{aligned} &\limsup_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mu^n\big(L_n^y \notin \Gamma | L_n^y \in A_\delta\big) \\ &\qquad \leq \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log Q_n\big(\Gamma^c \cap A_\delta\big) - \lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log Q_n\big(A_\delta\big). \end{aligned} \tag{2.6}$$

Le même argument que dans [20] (borne supérieure dans le théorème de Sanov) entraine :

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log Q_n\big(\Gamma^c \cap A_\delta\big) < -I(A_0). \tag{2.7}$$

Par ailleurs, la borne inférieure du théorème 1.9 du paragraphe précédent entraine que pour tout $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n\big(A_\delta\big) \geq -\inf\{I(\ell), \ell \in A_\delta^\circ\}. \tag{2.8}$$

Finalement, en combinant ces arguments avec (H-0), on obtient :

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log Q_n\big(A_\delta\big) \geq -I(A_0).$$

Enfin, on complète la preuve du lemme en utilisant cette inégalité combinée avec (2.7) dans (2.6). □

Preuve du théorème 2.5. Comme A_0 est convexe, $\mathcal{M} = \nu_* + \mathcal{S}$ (lemme 2.3). Considérons la fonction $f(x) = \prod_{i=1}^k f_i(x_i)$ où chaque $f_i \in C_b(\Sigma)$. La définition de $\mu^n_{Y^k|A_\delta}$ (voir (2.4)) entraine :

$$\frac{d\mu^n_{Y^k|A_\delta}}{d\mu^k}(y_1, \ldots, y_k) = \int_{\Sigma^{n-k}} \frac{\mathbf{1}_{A_{\delta,n}}(y_1, \ldots, y_n)}{Q_n(A_\delta)} \mu(dy_{k+1}) \cdots \mu(dy_n). \tag{2.9}$$

Considérons

$$\Gamma(\eta) = \cap_{i=1}^k \{\ell \in \mathcal{L}_\tau^*; \ |\langle \ell, f_i\rangle - \langle \nu_*, f_i\rangle| \leq \eta\}$$

et soit $\Gamma_n(\eta) = \{L_n^Y \in \Gamma(\eta)\}$. Alors $\Gamma(\eta)$ satisfait les hypothèses du lemme 2.6, c'est à dire $\Gamma^\circ(\eta) \subset \mathcal{M}$. En effet, si $\ell \in \mathcal{M}$, alors $\ell = \nu_* + \ell^s$ en vertu du lemme 2.3. Comme $f_i \in M_\tau$ pour $1 \leq i \leq k$, $\langle \ell^s, f_i\rangle = 0$ (proposition 1.5), on obtient $\langle \ell, f_i\rangle = \langle \nu_*, f_i\rangle$ et $\mathcal{M} \subset \Gamma^\circ(\eta)$. Par conséquent, le lemme 2.6 entraine :

$$\lim_{\delta\to 0}\lim_{n\to\infty} \mu^n\big(L_n^y \notin \Gamma | L_n^y \in A_\delta\big) = 0.$$

Le reste de la preuve suit pas à pas la preuve du corollaire 7.3.5 dans [20]. Le théorème 2.5 est donc prouvé. □

2.4 Retour à l'exemple de Csiszár

En reprenant le cadre du paragraphe 1.3 du chapitre précédent, nous nous intéressons ici au comportement asymptotique de $\mu^n_{Y|A_\delta(x)}$ dans le cas où

$$\begin{array}{rcl} A_\delta(x) & = & \{\ell \in \mathcal{L}_\tau^*;\quad |\langle \ell, u\rangle - x| \leq \delta\} \quad \text{avec} \quad u(y) = y \\ A_0(x) & = & \{\ell \in \mathcal{L}_\tau^*;\quad \langle \ell, u\rangle = x\}. \end{array}$$

Les ensembles de contrainte sont

$$\{L_n^Y \in A_\delta(x)\} = \{(Y_1, \cdots, Y_n);\ |\frac{1}{n}\sum_1^n Y_i - x| \leq \delta\}$$

et $\mu^n_{Y|A_\delta(x)}$ représente la loi de Y_1 sous la contrainte que la moyenne $\frac{1}{n}\sum_1^n Y_i$ est proche de x. On dénote par $\mathcal{M}_x$ l'ensemble des minimisants de l'entropie généralisée sous $A_0(x)$, soit

$$\mathcal{M}_x = \inf\{I(\ell);\ \ell \in \mathcal{Q};\ \langle \ell, u\rangle = x\}.$$

Comme la contrainte est convexe, on sait, en vertu du lemme 2.3, que

$$\mathcal{M}_x = \nu_*(x) + \mathcal{S}_x.$$

Proposition 2.7 *Pour tout $x \geq x_*$, ℓ appartient à $\mathcal{M}_x$ si et seulement si*

1. $\ell^a = \nu_*$ *avec* $\nu_*(dy) = e^{y-\Lambda(1)}\,\mu(dy)$
2. $\langle \ell^s, u\rangle = x - x_*$ *où* $u(y) = y$, $y \geq 0$
3. $\sup\{\langle \ell^s, f\rangle; f, \int_{[0,\infty)} e^f\, d\mu < \infty\} = \langle \ell^s, u\rangle$

En particulier, pour tout $x > x_*$, il existe une infinité d'éléments dans $\mathcal{M}_x$.

Remarque 2.8 La proposition précédente montre en particulier que ν_* ne dépend pas de x : $\mathcal{M}_x = \nu_* + \mathcal{S}_x$ pour tout $x > x_*$. D'autre part, ν_* ne minimise pas la fonction de taux I puisque

$$I(\nu_*) = H(\nu_*|\mu) < I(A_0(x)) = H(\nu_*|\mu) + x - x_*.$$

Preuve. L'équivalence est déjà prouvée dans la preuve de la proposition 1.19. Montrons maintenant qu'il existe une infinité de minimisants lorsque $x > x_*$. Du fait du point 3. de la proposition, il est suffisant de prouver que la fonction de jauge $p(g) = \inf\{\lambda > 0;\ g/\lambda \in D_\mu\}$ de $D_\mu = \{f \in L_\tau; \int_{[0,\infty)} e^f\, d\mu < \infty\}$ n'est pas Gâteaux-différentiable en u (on pourra se reporter, pour cet argument, à [32], p. 123). Il s'agit donc d'exhiber une fonction $f \in L_\tau$ vérifiant

$$\lim_{t\to 0,\ t>0} \frac{p(u+tf)-p(u)}{t} \neq \lim_{t\to 0,\ t<0} \frac{p(u+tf)-p(u)}{t}. \tag{2.10}$$

Considérons

$$f(y) = \begin{cases} ay & \text{si} \quad y \in \cup_{n\geq 0}[2n, 2n+1) \\ -by & \text{si} \quad y \in \cup_{n\geq 0}[2n+1, 2n+2) \end{cases} \quad \text{où} \quad a \neq b,\ a > 0,\ b > 0.$$

Par un calcul immédiat, nous obtenons

$$\lim_{t\to 0,\ t>0} \frac{p(u+tf)-p(u)}{t} = a \quad \text{et} \quad \lim_{t\to 0,\ t<0} \frac{p(u+tf)-p(u)}{t} = -b.$$

Par suite (2.10) est vérifiée et la proposition est prouvée. □

On pourra se reporter à ([37], proposition 10.5) pour une démonstration différente et pour plus de détails.

Appliquons le théorème 2.5 à $\mu^n_{Y|A_\delta(x)}$. On constate que pour $x > x_*$: $\mathcal{M}_x = \nu_* + \mathcal{S}_x$. Par conséquent, comme $f \in C_b([0,\infty))$, $\langle \mu^n_{Y|A_\delta(x)}, f\rangle$ tend vers $\langle \nu_*, f\rangle$ lorsque $n \to \infty$ puis $\delta \to 0$.

Remarque 2.9 Dans cet exemple, on pourra vérifier très facilement que :

$$\begin{aligned} \lim_{n\to\infty} \nu_*^n\{L_n^Y \in A_\delta(x_*)\} &= 1, \quad \delta > 0 \\ \lim_{n\to\infty} \nu_*^n\{L_n^Y \in A_\delta(x)\} &= 0, \quad \delta > 0, \quad \text{pour} \quad x > x_*. \end{aligned}$$

Par conséquent, l'équation (2.3), qui implique une loi des grands nombres sous ν_*, n'est pas satisfaite quand $x > x_*$. La convergence de $\mu^n_{Y|A_\delta(x)}$ vers ν_* a néanmoins lieu.

On remarquera que l'approche développée par I. Csiszár dans [14] et qui est basée sur une convergence en information ne repose pas non plus sur une hypothèse restrictive de loi des grands nombres sous-jacente.

2.5 Résultats complémentaires de convergence

On a vu dans le paragraphe précédent que la convergence étroite ne permet pas de déceler les parties singulières. On a vu également, lors de l'étude de l'exemple de Csiszár, que ces parties singulières nous renseignent sur la contrainte que l'on impose à notre mesure empirique. L'objet de ce paragraphe est de répondre aux questions suivantes : "Peut-on considérer une topologie plus fine que la topologie de la convergence étroite ? Cette topologie nous permettrait-elle de déceler les parties singulières ?"
Comme on le verra, les réponses que nous apportons à ces questions sont partielles. Les principaux résultats de ce paragraphe sont les théorèmes 2.11 et 2.16. On suppose à nouveau dans ce paragraphe que Σ est un espace métrique séparable muni de sa tribu borélienne.

2.5.1 Une estimée préliminaire

Dans cette section, nous établissons dans le lemme 2.10 une estimée qui nous permettra d'obtenir un résultat de convergence pour $\{\mu^n_{Y^k|A_\delta}\}$ plus précis que celui obtenu au théorème 2.5. L'ensemble des minimisants est à nouveau noté :

$$\mathcal{M} = \{\ell \in A_0;\ I(\ell) = I(A_0)\}.$$

Lemme 2.10 *On suppose que les hypothèses (H-0) et (H-1) sont vérifiées. On suppose également que $\Gamma \in \mathcal{E}$ vérifie les mêmes propriétés qu'au lemme 2.6, c'est à dire $\mathcal{M} \subset \Gamma^\circ$. Alors, pour toute fonction $f \in \mathcal{L}_\tau$, on a :*

$$\lim_{\delta\to 0}\lim_{n\to\infty} \mathbb{E}\left[|\langle f, L_n^Y\rangle| \mathbf{1}_{\Gamma^c}(L_n^Y) \big| L_n^Y \in A_\delta\right] = 0.$$

Preuve. Pour tout $n \geq 1$, $\delta > 0$,

$$\begin{aligned}
&\mathbb{E}\left[|\langle f, L_n^Y\rangle| \mathbf{1}_{\Gamma^c}(L_n^Y) \big| A_{\delta,n}\right] \\
&\quad\leq\ \mathbb{E}\left[|\langle f, \delta_{Y_1}\rangle| \mathbf{1}_{\Gamma^c}(L_n^Y) \big| A_{\delta,n}\right] \\
&\quad=\ \mathbb{E}\left[|f(Y_1)| u_{n,\delta}(Y_1)\right],
\end{aligned}$$

où

$$u_{n,\delta}(y_1) = \frac{1}{Q_n(A_\delta)} \mathbb{P}\{L_n^Y \in A_\delta \cap \Gamma^c | Y_1 = y_1\}.$$

Par suite,

$$\mathbb{E}\left[|\langle f, L_n^Y\rangle| \mathbf{1}_{\Gamma^c}(L_n^Y) \big| A_{\delta,n}\right] \leq \mathbb{E}\left[|f(Y_1)| u_{n,\delta}(Y_1)\right] \leq 2\|f\|_\tau \|u_{n,\delta}\|_{\tau^*},$$

la dernière inégalité étant simplement l'inégalité de Young (voir (1.4), chap. 1).
Il nous reste maintenant à prouver que $\|u_{n,\delta}\|_{\tau^*} \to 0$ pour $n \to \infty$ suivi de $\delta \to 0$. Par définition de la norme de Luxemburg, $\int_\Sigma \tau^*\big(\frac{u}{\epsilon}\big)\, d\mu \leq 1$ implique que $\|u\|_{L_{\tau^*}} \leq \epsilon$. Par conséquent, il est suffisant de prouver que pour tout $\epsilon > 0$

$$\limsup_{\delta\to 0} \limsup_{n\to\infty} \mathbb{E}\, \tau^*\Big(\frac{u_{n,\delta}(Y_1)}{\epsilon}\Big) < 1. \tag{2.11}$$

Comme τ^* est convexe, l'inégalité de Jensen entraine que

$$\begin{aligned}
&\mathbb{E}\, \tau^*\Big(\frac{u_{n,\delta}(Y_1)}{\epsilon}\Big)\\
&= \mathbb{E}\, \tau^*\Big(\frac{\mathbb{P}\{L_n^Y \in A_\delta \cap \Gamma^c | Y_1 = y_1\}}{\epsilon\, Q_n(A_\delta)}\Big)\\
&\leq \mathbb{E}\,\mathbb{E}\left[\tau^*\Big(\frac{\mathbf{1}_{\{L_n^Y \in A_\delta \cap \Gamma^c\}}}{\epsilon\, Q_n(A_\delta)}\Big)\Big| Y_1\right]\\
&= \mathbb{E}\left[\tau^*\Big(\frac{\mathbf{1}_{\{L_n^Y \in A_\delta \cap \Gamma^c\}}}{\epsilon\, Q_n(A_\delta)}\Big)\right]
\end{aligned}$$

On rappelle que $\tau^*(t) = (|t|+1)\log(|t|+1) - |t|$. Comme $\tau^*(0) = 0$,

$$\tau^*\Big(\frac{\mathbf{1}_{\{L_n^y \in A_\delta \cap \Gamma^c\}}(y_1, \cdots, y_n)}{\epsilon\, Q_n(A_\delta)}\Big) = \mathbf{1}_{\Gamma_n^c \cap A_{\delta,n}}(y_1, \cdots, y_n)\ \tau^*\Big(\frac{1}{\epsilon Q_n(A_\delta)}\Big).$$

Par suite,

$$\begin{aligned}
\mathbb{E}\, \tau^*\left(\frac{\mathbf{1}_{\{L_n^Y \in A_\delta \cap \Gamma^c\}}}{\epsilon\, Q_n(A_\delta)}\right) &= Q_n(\Gamma^c \cap A_\delta)\tau^*\Big(\frac{1}{\epsilon Q_n(A_\delta)}\Big)\\
&= \frac{1}{\epsilon}\mu^n(L_n^y \notin \Gamma | L_n^y \in A_\delta)\ \epsilon\ Q_n(A_\delta)\tau^*\Big(\frac{1}{\epsilon Q_n(A_\delta)}\Big).
\end{aligned}$$

Comme $t\tau^*(1/t) = (1+t)\log(1+1/t) - 1 \leq 2\log(2/t)$ pour $0 < t \leq 1$, on obtient :

$$\begin{aligned}
&\mathbb{E}\, \tau^*\left(\frac{\mathbf{1}_{\{L_n^Y \in A_\delta \cap \Gamma^c\}}}{\epsilon\, Q_n(A_\delta)}\right)\\
&\leq \mu^n(L_n^y \notin \Gamma | L_n^y \in A_\delta)\frac{2}{\epsilon}\log\Big(\frac{2}{\epsilon Q_n(A_\delta)}\Big)\\
&= \mu^n(L_n^y \notin \Gamma | L_n^y \in A_\delta)\frac{2}{\epsilon}\log\Big(\frac{2}{\epsilon}\Big)\\
&\quad + \mu^n(L_n^y \notin \Gamma | L_n^y \in A_\delta)\frac{2}{\epsilon}|\log Q_n(A_\delta)|,
\end{aligned} \tag{2.12}$$

où $0 < \epsilon \leq 1$. Le premier terme de la partie droite de l'équation (2.12) tend vers 0 grâce au lemme 2.6. Quant au second terme, il peut se réécrire

$$\frac{2}{\epsilon}\mu^n(L_n^y \notin \Gamma | L_n^y \in A_\delta) |\log Q_n(A_\delta)| = \frac{2}{\epsilon}\ n\mu^n(L_n^y \notin \Gamma | L_n^y \in A_\delta) \left| \frac{1}{n}\log Q_n(A_\delta)\right|.$$

Du fait de (2.8) et (H-0), on a :

$$\limsup_{\delta\to 0}\limsup_{n\to\infty} |\frac{1}{n}\log Q_n(A_\delta)| \leq I(A_0)$$

et du fait du lemme 2.6 on obtient :

$$\limsup_{\delta\to 0}\limsup_{n\to\infty} n\mu^n(\Gamma_n^c | A_{\delta,n}) = 0.$$

Finalement, pour tout $\epsilon > 0$, le second terme de la partie droite de (2.12) tend vers 0. Par conséquent, (2.11) est établi et la démonstration du lemme est achevée. $\square$

2.5.2 Convergence de $\mu^n_{Y^k|A_\delta}$ vers une probabilité

On définit l'ensemble de fonctions-test :

$$\mathcal{F} = \big\{ f : \Sigma^k \to \mathbb{R}; \\ f(x_1, \cdots, x_k) = f_1(x_1) g(x_2, \cdots, x_k),\ f_1 \in M_\tau(\Sigma),\ g \in C_b(\Sigma^{k-1}) \big\}.$$

La caractéristique principale est qu'une fonction-coordonnée f_i appartient à M_τ, les autres étant bornées. On choisit la première par simplicité mais il est évident, du fait de l'échangeabilité, que nos résultats seraient valables si $\mathcal{F}$ était remplacé par l'espace des fonctions symétrisées. Notons que $\mathcal{F}$ est un sous-ensemble de $M_\tau(\Sigma^k)$.

Enfin, comme la dérivée de Radon-Nikodym $\frac{d\mu^n_{Y^k|A_\delta}}{d\mu^n}$ est bornée sous l'hypothèse (H-1), le crochet de dualité $\langle \mu^n_{Y^k|A_\delta}, f\rangle$, $n \geq 1, \delta > 0$, est bien défini pour tout $f \in \mathcal{F}$. Le théorème 2.11 est à rapprocher du corollaire 7.3.5 dans [20]. C'est un corollaire des lemmes 2.6 et 2.10.

Théorème 2.11 *Supposons que les hypothèses (H-0) et (H-1) sont vérifiées, que Σ est un espace métrique séparable et que la contrainte A_0 est convexe. Alors, pour toute fonction-test f appartenant à $\mathcal{F}$, on a :*

$$\lim_{\delta\to 0}\lim_{n\to\infty} \langle \mu^n_{Y^k|A_\delta}, f\rangle = \langle \nu_*^k, f\rangle,$$

où ν_* est la partie absolument continue commune à tous les éléments de $\mathcal{M}$ (voir le lemme 2.3).

Remarque 2.12 Si $k=1$, alors $\mathcal{F}=M_\tau(\Sigma)$ et $\mu^n_{Y|A_\delta}\to\nu_*$ pour la topologie $\sigma(M'_\tau, M_\tau)$. Notons que M_τ ne sépare pas les points de L'_τ. Mais en identifiant les éléments non séparés, on obtient $\sigma(L'_\tau, M_\tau)=\sigma(M'_\tau, M_\tau)=\sigma(L_{\tau^*}, M_\tau)$, la dernière égalité provenant du théorème 1.1.

Outre les améliorations apportées au corollaire 7.3.5 dans [20] par le théorème 2.5, le théorème 2.11 permet de considérer une convergence un peu plus générale que la topologie de la convergence étroite. En effet, ce théorème implique le théorème 2.5 (notre choix de l'ordre de présentation des deux théorèmes nous a néanmoins amené à donner deux preuves distinctes).

Preuve. Comme A_0 est supposé convexe, $\mathcal{M}=\nu_*+\mathcal{S}$ (lemme 2.3). Fixons une fonction $f(x)=\prod_{i=1}^k f_i(x_i)$ où $f_1\in M_\tau(\Sigma)$ et $f_i\in B,\ i\geq 2$. On contrôle d'abord la croissance de f_1 grâce au lemme 2.10 en dehors d'un certain ensemble Γ puis on utilise la même technique combinatoire que dans la preuve du corollaire 7.3.5 dans [20] pour travailler à l'intérieur de Γ. Considérons

$$\Gamma(\eta)=\cap_{i=1}^k\{\ell\in\mathcal{L}^*_\tau;\ |\langle\ell,f_i\rangle-\langle\nu_*,f_i\rangle|\leq\eta\}\cap\{\ell;\ |\langle\ell,|f_1|\rangle-\langle\nu_*,|f_1|\rangle|\leq\eta\}.$$

Alors $\Gamma^\circ(\eta)\subset\mathcal{M}$ (mêmes arguments que dans la preuve du théorème 2.5). On notera dans la suite $\Gamma_n(\eta)=\{L^Y_n\in\Gamma(\eta)\}$ et on utilisera fréquemment Γ et Γ_n à la place de $\Gamma(\eta)$ et $\Gamma_n(\eta)$. Comme $f:\Sigma^k\to\mathbb{R}$ peut être considérée comme une fonction de Σ^n dans $\mathbb{R}$, (2.9) et une application immédiate du théorème de Fubini entrainent que :

$$\langle f,\mu^n_{Y^k|A_\delta}\rangle=\langle f,\mu^n_{Y^n|A_\delta}\rangle=\langle f\mathbf{1}_{\Gamma_n},\mu^n_{Y^n|A_\delta}\rangle+\langle f\mathbf{1}_{\Gamma_n^c},\mu^n_{Y^n|A_\delta}\rangle.$$

Notre but est de démontrer que $\lim_{\delta\to 0}\lim_{n\to\infty}\langle f,\mu^n_{Y^k|A_\delta}\rangle=\langle\nu_*^k,f\rangle$. Le même calcul que précédemment entraine que :

$$\begin{aligned}\langle f,\mu^n_{Y^k|A_\delta}\rangle-\langle f,\nu_*^k\rangle=&\ \langle f\mathbf{1}_{\Gamma_n^c},\mu^n_{Y^n|A_\delta}\rangle-\langle f,\nu_*^k\rangle\mu^n\big(L^y_n\notin\Gamma|L^y_n\in A_\delta\big)\\&+\langle f\mathbf{1}_{\Gamma_n},\mu^n_{Y^n|A_\delta}\rangle-\langle f,\nu_*^k\rangle\mu^n\big(L^y_n\in\Gamma|L^y_n\in A_\delta\big).\end{aligned}\qquad(2.13)$$

Occupons-nous d'abord des deux premiers termes de la partie droite de l'équation (2.13). Le lemme 2.6 entraine que

$$\lim_{\delta\to 0}\lim_{n\to\infty}\mu^n\big(L^y_n\notin\Gamma|L^y_n\in A_\delta\big)=0$$

et

$$
\begin{aligned}
&|\langle f\mathbf{1}_{\Gamma_n^c}, \mu^n_{Y^n|A_\delta}\rangle| \\
&= \Big| \int_{\Sigma^n} \prod_{i=1}^k f_i(y_i) \frac{\mathbf{1}_{A_{\delta,n}\cap\Gamma_n^c}(y_1,\dots,y_n)}{Q_n(A_\delta)}\, \mu(dy_1)\cdots\mu(dy_n)\Big| \\
&\le \Big(\prod_{i=2}^k \|f_i\|_\infty\Big) \int_{\Sigma^n} |f_1(y_1)| \frac{\mathbf{1}_{A_{\delta,n}\cap\Gamma_n^c}(y_1,\dots,y_n)}{Q_n(A_\delta)}\, \mu(dy_1)\cdots\mu(dy_n),
\end{aligned}
$$

où la dernière intégrale converge vers 0 grâce au lemme 2.10. Par suite, pour tout $\eta > 0$,

$$
\lim_{\delta\to 0}\lim_{n\to\infty} \Big\{ \langle f\mathbf{1}_{\Gamma_n^c(\eta)}, \mu^n_{Y^n|A_\delta}\rangle - \langle f, \nu_*^k\rangle \mu^n\big(L_n^y \notin \Gamma(\eta) | L_n^y \in A_\delta\big)\Big\} = 0. \tag{2.14}
$$

Considérons maintenant les deux derniers termes de la partie droite de (2.13). Il est utile de comparer $\langle f\mathbf{1}_{\Gamma_n}, \mu^n_{Y^n|A_\delta}\rangle$ avec $\mathbb{E}\,[\prod_1^k \langle f_i\ , L_n^Y\rangle\ \mathbf{1}_{\Gamma_n}|A_{\delta,n}]$. En effet, la définition de Γ_n nous donne des informations sur des quantités telles que $\langle f_i\ , L_n^Y\rangle$. En utilisant la même technique que dans la preuve du corollaire 7.3.5 dans [20], on obtient :

$$
\begin{aligned}
&|\langle f\mathbf{1}_{\Gamma_n}, \mu^n_{Y^n|A_\delta}\rangle - \mathbb{E}\,[\prod_{i=1}^k \langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n}|A_{\delta,n}]| \\
&\qquad \le 2\Big(1 - \frac{n!}{n^k(n-k)!}\Big)\Big(\prod_{i=2}^k \|f_i\|_\infty\Big)\cdots \\
&\qquad\qquad \int_{\Sigma^n} |f_1(y_1)| \mathbf{1}_{\Gamma_n}\, \mu^n_{Y^n|A_\delta}(dy_1\cdots dy_n).
\end{aligned} \tag{2.15}
$$

Il s'agit maintenant de contrôler la dernière intégrale :

$$
\begin{aligned}
&\int_{\Sigma^n} |f_1(y_1)|\, \mu^n_{Y^n|A_\delta}(dy_1\cdots dy_n) \\
&\qquad = \int_{\Sigma^n} \langle |f_1|, L_n^Y\rangle\, \mu^n_{Y^n|A_\delta}(dy_1\cdots dy_n) \le \langle |f_1|, \nu_*\rangle + \eta,
\end{aligned}
$$

où la dernière inégalité provient de la définition même de Γ (c'est précisément pour cette raison que le contrôle sur $|f_1|$ apparaît dans la définition de Γ). Par suite, pour tout $\delta > 0$,

$$
\big|\langle f\mathbf{1}_{\Gamma_n}, \mu^n_{Y^n|A_\delta}\rangle - \mathbb{E}\,[\prod_{i=1}^k \langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n}|A_{\delta,n}]\ \big| \le C\Big(1 - \frac{n!}{n^k(n-k)!}\Big) \xrightarrow[n\to\infty]{} 0
$$

et pour tout $\eta > 0$,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\left\{\langle \mu^n_{Y^k|A_\delta}, f\mathbf{1}_{\Gamma_n(\eta)}\rangle - \mathbb{E}\,[\prod_{i=1}^k \langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n(\eta)}|A_{\delta,n}]\right\} = 0. \tag{2.16}$$

Enfin, la définition de Γ entraine que

$$\begin{aligned}
&\left|\mathbb{E}\,[\prod_{i=1}^k \langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n(\eta)}|A_{\delta,n}] - \langle f, \nu_*^k\rangle \mu^n\big(L_n^y\in\Gamma(\eta)|L_n^y\in A_\delta\big)\right| \\
&= \ \left|\mathbb{E}\,[\prod_{i=1}^k \langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} - \prod_1^k \langle f_i, \nu_*\rangle \ \mathbf{1}_{\Gamma_n}|A_{\delta,n}]\ \right| \\
&= \ \Big|\mathbb{E}\,[\langle f_1, L_n^Y\rangle \prod_{i=2}^k \langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} - \langle f_1,\nu_*\rangle \prod_{i=2}^k \langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} \\
&\qquad + \langle f_1,\nu_*\rangle\langle f_2, L_n^Y\rangle \prod_{i=3}^k \langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} - \langle f_1,\nu_*\rangle\langle f_2,\nu_*\rangle \prod_{i=3}^k \langle f_i, L_n^Y\rangle \mathbf{1}_{\Gamma_n} + \cdots]\ \Big| \\
&\leq \ \eta \prod_{i=2}^k \|f_i\|_\infty + \eta|\langle f_1,\nu_*\rangle| \prod_{i=3}^k \|f_i\|_\infty + \cdots \\
&\leq \ C\eta.
\end{aligned} \tag{2.17}$$

Il reste à faire tendre η vers zéro pour conclure que (2.14), (2.16) et (2.17) entrainent que

$$\lim_{\delta\to 0}\lim_{n\to\infty}\left\{\langle \mu^n_{Y^k|A_\delta}, f\rangle - \langle f, \nu_*^k\rangle\right\} = 0.$$

La convergence est donc prouvée pour tout $f = \prod_1^k f_i$ où $f_1 \in M_\tau$ et $f_i \in B$, $i \geq 2$. En particulier, si $f_i \in C_b(\Sigma)$, $i \geq 2$, alors $\langle \mu^n_{Y^k|A_\delta}, f_1 \prod_2^k f_i\rangle \to \langle (\nu_*)^k, f_1 \prod_2^k f_i\rangle$, ce qui peut se réécrire

$$\langle m^n_{f_1}, g\rangle \xrightarrow[n\to\infty]{} x\langle \nu_*, f_1\rangle \langle (\nu_*)^{k-1}, g\rangle, \tag{2.18}$$

où $m^n_{f_1}(dy_2\cdots dy_k) = \int_{z\in\Sigma} f_1(z)\mu^n_{Y^k|A_\delta}(dzdy_2\cdots dy_k)$ est une mesure signée et $g = \prod_2^k f_i$. Comme Σ est séparable, $C_b(\Sigma)^{k-1}$ détermine la convergence pour $\mathcal{P}(\Sigma^k)$ et (2.18) est vrai pour tout $g \in C_b(\Sigma^{k-1})$. Cela complète la preuve du théorème 2.11. □

On pourrait se demander comment étendre ce résultat à une classe de fonctions-test plus large que $\mathcal{F}$, $M_\tau(\Sigma^k)$ par exemple. On serait alors tenté de remplacer, dans la preuve ci-dessus $\Gamma(\eta)$ par un ensemble du type $\Gamma_* =$

$\{\ell \in \mathcal{L}_\tau^*; |\langle \ell^{a\otimes k} - (\nu_*)^{\otimes k}, f\rangle| \leq \eta\}$ avec f dans $\mathcal{M}_\tau(\Sigma^k)$. Malheureusement, $\ell \mapsto \ell^{a\otimes k}$ n'est pas une application continue pour les topologies $\sigma(\mathcal{M}_\tau^*, \mathcal{M}_\tau)$ et $\sigma(\mathcal{M}_\tau(\Sigma^k)^*, \mathcal{M}_\tau(\Sigma^k))$ (pour un tel argument, voir par exemple [20], exercice 7.3.18). En conséquence, il n'est pas évident que $\mathcal{M} \subset \Gamma_*^\circ$ et que le lemme 2.6 soit utilisable dans un tel cadre.

2.5.3 Plus d'informations sur le comportement-limite de $\mu^n_{Y|A_\delta}$

Nous considérons ici le cas monodimensionnel où $k = 1$ et où la topologie est $\sigma(L'_\tau, L_\tau)$. On a vu dans les paragraphes 2.4 et 2.5.2 que

- aucune partie singulière n'apparaît quand on travaille avec la topologie $\sigma(L'_\tau, M_\tau)$ (voir la remarque 2.12) :

$$\langle \mu^n_{Y|A_\delta}, f\rangle \to \langle \ell^a, f\rangle \text{ pour tout } f \in M_\tau.$$

- les parties singulières présentent un intérêt.
- il peut exister, pour une contrainte donnée -même convexe-, plusieurs minimisants. (voir la proposition 2.7).

Par conséquent, plusieurs questions se posent :

- quand on considère la topologie $\sigma(L'_\tau, L_\tau)$, $\mu^n_{Y|A_\delta}$ admet-elle une unique limite ℓ, où ℓ est un minimisant ?
- la mesure $\mu^n_{Y|A_\delta}$ admet-elle plusieurs points d'accumulation ? Tous les minimisants sont-ils des points d'accumulation ?

Notre réponse à ces questions est partielle. On démontre juste, dans la section à venir, que l'ensemble des points d'accumulation de $\{\mu^n_{Y|A_\delta}\}_{n,\delta}$ est non vide et est un sous-ensemble de l'adhérence de l'enveloppe convexe de $\mathcal{M}$.

Remarque 2.13 La séparabilité d'un espace d'Orlicz est directement liée au fait que la fonction de Young associée θ satisfait la condition Δ_2, c'est à dire

$$\theta(2s) \leq K\theta(s).$$

Ici, τ croît trop vite pour satisfaire cette condition. Par suite, la boule unité (forte) de L'_τ n'admet pas de base de voisinages dénombrable pour la topologie $\sigma(L'_\tau, L_\tau)$. Ainsi, pour une suite de la boule unité de L'_τ, l'existence d'une sous-suite convergente n'est pas garantie.

Deux des outils les plus utilisés pour traiter les questions de convergence dans les espaces non séparables sont les filtres et les *nets* (anglais), que nous continuerons à appeler *nets* du fait de notre ignorance de toute traduction autorisée.

Bien que Cartan ait popularisé l'utilisation des filtres en France, nous utiliserons les *nets* dont la manipulation reste très proche de celle des suites. De bonnes références sur le sujet sont les ouvrages de Dudley [23], Kelley [34] ainsi que le livre de Reed et Simon [44].
Rappelons qu'un *subnet* $(y_\alpha)_{\alpha\in A}$ d'une suite $(u_n)_{n\geq 1}$ est une famille $(y_\alpha)_{\alpha\in A}$ où A est un ensemble filtrant à droite tel qu'il existe $p : A \to \mathbb{N}$ vérifiant :

- $\alpha \leq \beta \quad \Rightarrow \quad p(\alpha) \leq p(\beta)$.
- Pour tout $n \in \mathbb{N}$, il existe α_n tel que $p(\alpha_n) \geq n$.
- $y_\alpha = u_{p(\alpha)}$

Le *subnet* (y_α) converge vers y_∞ si pour tout voisinage V de y_∞, il existe α_0 tel que $\alpha \geq \alpha_0$ implique $y_\alpha \in V$. Voici deux différences fondamentales entre un *subnet* et une sous-suite : A peut être énorme par rapport à $\mathbb{N}$ et l'ordre $\leq$ est un ordre partiel sur A.

Comme la convergence est d'un type particulier (d'abord $n \to \infty$ puis $\delta \to 0$), on précise d'abord ce qu'est un point d'accumulation.
On dit que $\nu \in L'_\tau$ est un (n,δ)-point d'accumulation de $\{\mu^n_{Y|A_\delta}\}_{n,\delta}$ s'il existe une famille $\{\nu_\delta\}_{\delta>0} \subset L'_\tau$ telle que

- pour tout $\delta > 0$, il existe un *subnet* de la suite $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$ convergent vers ν_δ pour la toplogie $\sigma(L'_\tau, L_\tau)$.
- il existe une suite croissante $\delta_n \to 0$ telle que il existe un *subnet* de $\{\nu_{\delta_n}\}_{n\geq 1}$ convergent vers ν pour la topologie $\sigma(L'_\tau, L_\tau)$.

La proposition suivante nous sera utile pour démontrer le théorème 2.16, principal résultat de ce paragraphe.

Proposition 2.14 *Soit $\{\mu^{p(\beta)}_{Y|A_\delta}\}_{\beta\in B}$ un subnet convergent vers ν_δ. Alors, pour tout $f \in L_\tau$, il existe une sous-suite $\{\mu^{\phi_\delta(n)}_{Y|A_\delta}\}$ vérifiant*

$$\lim_{n\to\infty} \langle \mu^{\phi_\delta(n)}_{Y|A_\delta}, f\rangle = \langle \nu_\delta, f\rangle. \tag{2.19}$$

Preuve. Considérons le voisinage suivant de ν_δ, $V_n = \{\xi \in L'_\tau, |\langle \xi - \nu_\delta, f\rangle| < \frac{1}{n}\}$.
La définition même du *subnet* entraine que : $\exists \beta_n \in B,\ p(\beta_n) \geq n,\ \mu^{p(\beta_n)}_{Y|A_\delta} \in V_n$.
Choisissons une suite croissante de β_n, alors

$$\langle \mu^{p(\beta_n)}_{Y|A_\delta}, f\rangle \to \langle \nu_\delta, f\rangle$$

et la proposition est démontrée. □

Remarque 2.15 On remarquera qu'il est nécessaire, dans la proposition précédente, de travailler avec des *nets*. En effet, le théorème de Bolzano-Weierstass nous permet d'extraire une sous-suite convergente de $\langle \mu^n_{Y|A_\delta}, f\rangle$. Néanmoins, nous ne sommes pas sûrs que cette sous-suite converge vers $\langle \nu_\delta, f\rangle$. Or cette information nous importe puisqu'on a besoin du contrôle suivant :

$$|\langle f, \nu_\delta\rangle| \le \limsup_{n\to\infty} |\langle f, \mu^n_{Y|A_\delta}\rangle|.$$

Théorème 2.16 *Supposons que (H-0) et (H-1) sont vérifiées. Soit $\mathcal{C}$ l'ensemble des (n,δ)-points d'accumulation de $\{\mu^n_{Y|A_\delta}\}_{n,\delta}$, alors $\mathcal{C}$ est un sous-ensemble non vide de $\overline{\mathrm{co}}(\mathcal{M})$, l'adhérence de l'enveloppe convexe de $\mathcal{M}$.*

La preuve du théorème 2.16 est basée sur le lemme suivant.

Lemme 2.17 *Pour tout $\delta > 0$, $\{\mu^n_{Y|A_\delta}\}_{n\ge 1}$ est relativement compact pour la topologie $\sigma(L'_\tau, L_\tau)$.*

Dans ce lemme, δ est fixé et on s'intéresse à un résultat valide uniformément en f. Par suite, les lemmes 2.6 et 2.10 ne sont pas directement exploitables. Néanmoins, les résultats suivants se démontrent très facilement en reprenant leurs preuves. Soit $\Gamma_{a,f} \stackrel{\triangle}{=} \{\ell \in L'_\tau, |\langle \ell, f\rangle| \le a\}$, où $f \in L_\tau$ et $a \ge 0$. Comme $\mathcal{M}$ est compact, il existe pour tout f un a tel que $\mathcal{M} \subset \Gamma_{a,f}$. De plus, on peut démontrer que pour tout $\delta > 0$ et $f \in L_\tau$, on a

$$\lim_{a\to\infty} \limsup_n \frac{1}{n} \log \mu^n\big(L^y_n \notin \Gamma_{a,f} | L^y_n \in A_\delta\big) = -\infty \tag{2.20}$$

et il existe $a > 0$ tel que

$$\mathbb{E}\left[|\langle f, L^Y_n\rangle| \mathbf{1}_{\Gamma^c_{a,f}}(L^Y_n)\big| L^Y_n \in A_\delta\right] \to_{n\to\infty} 0. \tag{2.21}$$

Ces deux énoncés sont analogues aux lemmes 2.6 et 2.10 pour un δ fixé.

Preuve du lemme 2.17. Soit f appartenant à L_τ. Choisissons a de manière à ce que (2.21) soit vérifié. Alors

$$\begin{aligned}
&|\langle f, \mu^n_{Y|A_\delta}\rangle| \\
&\quad = \ |\mathbb{E}\left[\langle f, L^Y_n\rangle \big| L^Y_n \in A_\delta\right]| \\
&\quad \le \ \mathbb{E}\left[|\langle f, L^Y_n\rangle| \mathbf{1}_{\Gamma_{a,f}}(L^Y_n)\big| L^Y_n \in A_\delta\right] + \mathbb{E}\left[|\langle f, L^Y_n\rangle| \mathbf{1}_{\Gamma^c_{a,f}}(L^Y_n)\big| L^Y_n \in A_\delta\right] \\
&\quad \le \ a + \mathbb{E}\left[|\langle f, \delta_{Y_1}\rangle| \mathbf{1}_{\Gamma^c_{a,f}}(L^Y_n)\big| L^Y_n \in A_\delta\right].
\end{aligned}$$

Grâce à (2.21), $\lim_n \mathbb{E}\left[|\langle f, \delta_{Y_1}\rangle| \mathbf{1}_{\Gamma^c_{a,f}}(L^Y_n)\big| L^Y_n \in A_\delta\right] = 0$. Il s'ensuit que :

$$\sup_n |\langle f, \mu^n_{Y|A_\delta}\rangle| < \infty, \text{ pour tout } f \in L_\tau \text{ et } \delta > 0.$$

Les théorèmes de Banach-Steinhaus et Banach-Alaoglu entrainent que $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$ est relativement compact et qu'il existe un *subnet* de $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$ convergent vers $\nu_\delta \in L'_\tau$. $\square$

Preuve du théorème 2.16. Montrons d'abord que $\mathcal{C}$ n'est pas vide. Fixons $f \in L_\tau$. Comme $\ell \mapsto |\langle \ell, f\rangle|$ est continue et que $\mathcal{M}$ est un ensemble compact, il existe $a > 0$ tel que $\mathcal{M} \subset \{\ell \in L'_\tau, |\langle \ell, f\rangle| \ < \ a\ \}$. Soit $\Gamma = \{\ell \in L'_\tau, |\langle \ell, f\rangle| \leq a\}$ alors $\mathcal{M} \subset \overset{\circ}{\Gamma}$ et

$$\begin{aligned}
|\langle f, \mu^n_{Y|A_\delta}\rangle| &= \left|\mathbb{E}\left[\langle f, L^Y_n\rangle \middle| A_{\delta,n}\right]\right| \\
&\leq \mathbb{E}\left[|\langle f, L^Y_n\rangle| \mathbf{1}_{\Gamma_n} \middle| A_{\delta,n}\right] + \mathbb{E}\left[|\langle f, L^Y_n\rangle| \mathbf{1}_{\Gamma^c_n} \middle| A_{\delta,n}\right] \\
&\leq a + \mathbb{E}\left[|\langle f, \delta_{Y_1}\rangle| \mathbf{1}_{\Gamma^c_n} \middle| A_{\delta,n}\right].
\end{aligned}$$

Du fait du lemme 2.10, $\lim_\delta \lim_n \mathbb{E}\left[|\langle f, \delta_{Y_1}\rangle| \mathbf{1}_{\Gamma^c}(L^Y_n) \middle| L^Y_n \in A_\delta\right] = 0$. Par conséquent, pour tout $f \in L_\tau$

$$\limsup_\delta \limsup_n |\langle f, \mu^n_{Y|A_\delta}\rangle| < \infty. \tag{2.22}$$

Prenons $\delta > 0$ et ν_δ un point d'accumulation de $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$. Le lemme 2.17 et la proposition 2.14 entrainent l'existence d'une sous-suite $\{\mu^{\phi_\delta(n)}_{Y|A_\delta}\}_{n\geq 1}$ de $\{\mu^n_{Y|A_\delta}\}_{n\geq 1}$ telle que $\langle \mu^{\phi_\delta(n)}_{Y|A_\delta}, f\rangle \to_n \langle \nu_\delta, f\rangle$. Il s'ensuit que

$$|\langle f, \nu_\delta\rangle| \leq \limsup_n |\langle f, \mu^n_{Y|A_\delta}\rangle|$$

et que

$$\limsup_\delta |\langle f, \nu_\delta\rangle| \leq \limsup_\delta \limsup_n |\langle f, \mu^n_{Y|A_\delta}\rangle|.$$

La propriété (2.22) entraine que pour toute sous-suite $\{\delta_n\}_{n\geq 1}$ convergeant vers zéro,

$$\limsup_n |\langle f, \nu_{\delta_n}\rangle| \leq \limsup_\delta |\langle f, \nu_\delta\rangle| < \infty.$$

Par conséquent, $\sup_n |\langle f, \nu_{\delta_n}\rangle| < \infty$ pour tout $f \in L_\tau$. Les théorèmes de Banach-Steinhaus et Banach-Alaoglu entrainent que $\{\nu_{\delta_n}\}_{n\geq 1}$ est relativement compact pour la topologie $\sigma(L'_\tau, L_\tau)$ et qu'il existe un *subnet* de $\{\nu_{\delta_n}\}_{n\geq 1}$ convergeant vers $\nu \in L'_\tau$. Par suite, $\mathcal{C} \subset L'_\tau$ n'est pas vide.

Montrons maintenant que $\mathcal{C} \subset \overline{\text{co}}(\mathcal{M})$. Supposons que $\xi \notin \overline{\text{co}}(\mathcal{M})$. Le théorème de Hahn-Banach assure l'existence de $f \in L_\tau$ telle que $\langle \xi, f\rangle > a$ et $\langle \ell, f\rangle < a$ pour tout $\ell \in \overline{\text{co}}(\mathcal{M})$. Soit $\Gamma \triangleq \{\ell \in L'_\tau, \langle \ell, f\rangle \leq a\}$, alors $\mathcal{M} \subset \Gamma^\circ$ et

$$\begin{aligned}
\langle \mu^n_{Y|A_\delta}, f\rangle - \langle \xi, f\rangle &= \mathbb{E}\left[\langle f, L^Y_n\rangle - \langle \xi, f\rangle | A_{\delta,n}\right] \\
&= \mathbb{E}\left[\langle L^Y_n, f\rangle \mathbf{1}_{\Gamma_n} | A_{\delta,n}\right] - \langle \xi, f\rangle \mu^n\big(L^y_n \in \Gamma | L^y_n \in A_\delta\big) + \epsilon(n, \delta),
\end{aligned}$$

où $\epsilon(n,\delta) = \mathbb{E}\left[\left\{\langle L_n^Y, f\rangle - \langle \xi, f\rangle\right\}\mathbf{1}_{\Gamma_n^c}|A_{\delta,n}\right]$. Grâce aux lemmes 2.6 et 2.10, $\lim_\delta \lim_n \epsilon(n,\delta) = 0$. La définition même de Γ entraine

$$\mathbb{E}\left[\langle L_n^Y, f\rangle \mathbf{1}_{\Gamma_n}|A_{\delta,n}\right] \leq a.$$

Comme $\overline{\mathrm{co}}(\mathcal{M}) \subset \Gamma$, on a :

$$\lim_{\delta\to 0}\lim_{n\to\infty} \mu^n\left(L_n^y \in \Gamma | L_n^y \in A_\delta\right) = 1,$$

ce qui entraine que

$$\lim_{\delta\to 0}\lim_{n\to\infty} \langle \xi, f\rangle \mu^n\left(L_n^y \in \Gamma | L_n^y \in A^\delta\right) > a.$$

Finalement, on obtient : $\liminf_\delta \liminf_n |\langle \mu^n_{Y|A_\delta}, f\rangle - \langle \xi, f\rangle| > 0$, soit $\xi \notin \mathcal{C}$. Ceci achève la démonstration du théorème . □

2.6 Bilan

Résumons brièvement l'ensemble des résultats de convergence établis dans ce chapitre.

– si f est mesurable bornée, alors :

$$\mathbb{E}[f(Y_1)|L_n^Y \in A_\delta] \to \int f\, d\nu_*,$$

où ν^* est la partie absolument continue commune à tous les minimisants de la fonction de taux I sur A_0 (théorème 2.5).

– si f a tous ses moments exponentiels ($f \in M_\tau$), *i.e.* :

$$\forall \alpha > 0, \quad \mathbb{E}e^{\alpha|f(Y)|} < \infty,$$

alors le résultat précédent demeure valide (théorème 2.11) :

$$\mathbb{E}[f(Y_1)|L_n^Y \in A_\delta] \to \int f\, d\nu_*.$$

– si f n'a que quelques moments exponentiels ($f \in L_\tau$), *i.e.* :

$$\exists \alpha > 0, \quad \mathbb{E}e^{\alpha|f(Y)|} < \infty,$$

alors nous sommes bien en peine de décrire le comportement asymptotique de $\mathbb{E}[f(Y_1)|L_n^Y \in A_\delta]$, si ce n'est qu'à priori, plusieurs minimisants sont "impliqués" (théorème 2.16).

Chapitre 3

Un théorème de Cramér pour une moyenne empirique pondérée

Sommaire

Soit $(Z_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires à valeurs $\mathbb{R}^d$, indépendantes et identiquement distribuées (i.i.d.) satisfaisant :

$$\mathbb{E}e^{\alpha|Z_i|} < +\infty \quad \text{pour un} \quad \alpha > 0. \tag{3.1}$$

Soit $(x_i^n;\ 1 \le i \le n;\ n \ge 1)$ une suite d'éléments à valeurs $\mathcal{X}$ satisfaisant :

$$\frac{1}{n}\sum_1^n \delta_{x_i^n} \xrightarrow[n\to\infty]{\text{conv. étroite}} R,$$

où R est une probabilité sur $\mathcal{X}$ vérifiant la condition suivante de stricte positivité :

$$R(U) > 0 \quad \text{pour tout } U \text{ ouvert non vide de } \mathcal{X}.$$

L'objet de ce chapitre est d'établir un principe de grandes déviations (PGD) pour la moyenne empirique pondérée :

$$\langle L_n, \mathbf{f} \rangle = \frac{1}{n} \sum_1^n \begin{pmatrix} f_1(x_i^n) \cdot Z_i \\ \vdots \\ f_m(x_i^n) \cdot Z_i \end{pmatrix} \quad \left(\stackrel{\triangle}{=} \frac{1}{n} \sum_1^n \mathbf{f}(x_i^n) \cdot Z_i \in \mathbb{R}^m \right),$$

où chaque f_k est une fonction continue bornée de $\mathcal{X}$ dans $\mathbb{R}^d$ et $\cdot$ représente le produit scalaire sur $\mathbb{R}^d$.
Des principes de grandes déviations similaires ont été étudiés dans un contexte de mécanique statistique par Ben Arous, Dembo et Guionnet [3], la mesure empirique $\frac{1}{n}\sum_1^n \delta_{x_i^n}$ étant cette fois-ci aléatoire. L'étude de formes quadratiques de processus gaussiens a également donné lieu à des travaux voisins, voir par exemple Bryc et Dembo [10] et les travaux de Bercu, Gamboa, Lavielle, Rouault et Zani [5, 6, 30, 59]. Le lien avec les processus gaussiens provient du fait que le périodogramme empirique de tels processus peut s'exprimer comme une moyenne pondérée de variables aléatoires i.i.d. et distribuées selon une loi du χ^2. Enfin, on notera que l'étude d'une telle moyenne empirique est le prélude à toute étude des grandes déviations de mesures empiriques du type $\frac{1}{n}\sum Z_i \delta_{x_i^n}$. Une telle étude est menée au chapitre 4 où on trouvera également les références bibliographiques appropriées.

Le résultat principal de ce chapitre est l'établissement d'un PGD sans faire l'hypothèse d'escarpement pour la log-Laplace (théorème 3.2). Sans cette hypothèse, on ne peut plus utiliser le théorème de Gärtner-Ellis. Plus précisement, on n'a plus accès à la technique de Cramér du changement de mesure exponentiel (cf. Introduction). Notre approche consiste à établir dans un premier temps le PGD pour des fonctions en escalier. Cela nous permet (schématiquement) d'utiliser le théorème de Cramér, valable sous la condition (3.1) (voir Bahadur et Zabell [4], et la section 6.1 dans [20]), sur chaque ensemble où la fonction en escalier est constante. Une approximation exponentielle nous permet enfin d'obtenir le PGD désiré.

3.1 Le PGD pour la moyenne empirique pondérée

3.1.1 Hypothèses et notations

Soit $\mathcal{X}$ un espace vectoriel topologique muni de sa tribu de Borel $\mathcal{B}(\mathcal{X})$ et soit R une probabilité sur $\mathcal{X}$. On appelle $C(\mathcal{X})$ (resp. $C_d(\mathcal{X})$) l'ensemble des fonctions continues bornées de $\mathcal{X}$ à valeurs $\mathbb{R}$ (resp. $\mathbb{R}^d$) ; $L_d^1(\mathcal{X})$ l'ensemble des fonctions R-intégrables sur $\mathcal{X}$ à valeurs $\mathbb{R}^d$ et $\mathcal{P}(\mathcal{X})$ l'ensemble des probabilités sur $\mathcal{X}$. On omettra de temps à autre $\mathcal{X}$ et on notera les espaces précédents simplement C, C_d, L_d^1. On désignera par $|\cdot|$ une norme sur un espace vectoriel de dimension finie (indifférement $\mathbb{R}$, $\mathbb{R}^d$ ou $\mathbb{R}^{m\times d}$). On note $\|\cdot\|$ la norme du supremum sur l'ensemble des fonctions continues bornées de $\mathcal{X}$ à valeurs dans un espace vectoriel de dimension finie (en général $\mathbb{R}$, $\mathbb{R}^d$ ou $\mathbb{R}^{m\times d}$), *i.e.* $\|f\| = \sup_{x\in\mathcal{X}} |f(x)|$. Comme d'habitude, δ_a est la mesure de Dirac en a. On fait les hypothèses suivantes :

Hypothèse H-2 *R est une probabilité sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ satisfaisant*

$$\textit{si } U \textit{ est un ouvert non vide, alors} \quad R(U) > 0.$$

Hypothèse H-3 *La famille $(\ x_i^n\ ; 1 \le i \le n,\ n \ge 1)\ \subset \mathcal{X}$ satisfait*

$$\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n} \xrightarrow[n\to\infty]{\textit{conv. étroite}} R \quad \textit{où} \quad R \in \mathcal{P}(\mathcal{X}). \tag{3.2}$$

Remarque 3.1 La combinaison des hypothèses (H-2) et (H-3) est habituelle (voir [3], [28] et [29]). On construit dans la section 3.1.3 deux contre-exemples dans le cas où l'hypothèse (H-2) n'est pas satisfaite.

Hypothèse H-4 *Soit $(Z_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires à valeurs $\mathbb{R}^d$, indépendantes et identiquement distribuées selon la loi $\mathbb{P}_Z$. On suppose que la condition suivante de moments exponentiels est vérifiée :*

$$\mathbb{E}e^{\alpha|Z|} < +\infty \quad \textit{pour un} \quad \alpha > 0. \tag{3.3}$$

Ici, $\mathbb{P}_Z$ est l'image de $\mathbb{P}$. C'est-à-dire que $\mathbb{P}\{Z_i \in A\} = \mathbb{P}_Z(A)$, où $Z_i : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
On note Λ la log-Laplace de Z et Λ^* sa conjuguée convexe :

$$\begin{aligned} \Lambda(\lambda) &= \log \mathbb{E}e^{\lambda\cdot Z}, \lambda \in \mathbb{R}^d, \\ \Lambda^*(z) &= \sup_{\lambda\in\mathbb{R}^d}\{\lambda\cdot z - \Lambda(\lambda)\},\ z \in \mathbb{R}^d, \end{aligned}$$

où $\cdot$ est le produit scalaire sur $\mathbb{R}^d$. On notera par $\mathcal{D}_\Lambda = \{\lambda \in \mathbb{R}^d;\ \Lambda(\lambda) < \infty\}$ le domaine effectif de Λ.
Soit $\mathbf{a}$ une matrice $m \times d$ et z un élément de $\mathbb{R}^d$. On notera $\mathbf{a}\cdot z$ le produit matriciel habituel, soit

$$\mathbf{a}\cdot z = \begin{pmatrix} a_1 \cdot z \\ \vdots \\ a_m \cdot z \end{pmatrix},$$

a_j étant la j$^{\text{e}}$ ligne de la matrice $\mathbf{a}$. Ainsi, selon le contexte $\cdot$ désigne le produit scalaire $\lambda \cdot z$ ou le produit matriciel $\mathbf{a} \cdot z$. Soit $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ une fonction (matricielle) continue bornée, alors toutes les composantes de la matrice $\mathbf{f}$ sont des fonctions continues bornées et

$$\mathbf{f}(x) \cdot z = \begin{pmatrix} f_1(x) \cdot z \\ \vdots \\ f_m(x) \cdot z \end{pmatrix},$$

où $f_j(x) \in C_d(\mathcal{X})$ est la j$^{\text{e}}$ ligne de la matrice $\mathbf{f}$. Si $u : \mathcal{X} \to \mathbb{R}^d$ est une fonction mesurable, on note

$$\int_{\mathcal{X}} \mathbf{f}(x) \cdot u(x)\, R(dx) = \begin{pmatrix} \int_{\mathcal{X}} f_1(x) \cdot u(x)\, R(dx) \\ \vdots \\ \int_{\mathcal{X}} f_m(x) \cdot u(x)\, R(dx) \end{pmatrix}.$$

Enfin, on introduit la moyenne pondérée dont on va étudier le PGD.

$$\langle L_n, \mathbf{f} \rangle \stackrel{\triangle}{=} \frac{1}{n} \sum_{i=1}^{n} \mathbf{f}(x_i^n) \cdot Z_i \quad (\in \mathbb{R}^m).$$

On utilisera la convention suivante : x sera à valeurs dans $\mathcal{X}$, y et θ seront des éléments de $\mathbb{R}^m$ et z et λ, des éléments de $\mathbb{R}^d$.

3.1.2 Le principe de grandes déviations

Théorème 3.2 *Soit $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ une fonction continue bornée. On suppose que (H-2), (H-3) et (H-4) sont vérifiées. Alors la famille*

$$\langle L_n, \mathbf{f} \rangle = \frac{1}{n} \sum_{i=1}^{n} \mathbf{f}(x_i^n) \cdot Z_i \ \in \mathbb{R}^m$$

satisfait le principe de grandes déviations dans $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ de bonne fonction de taux

$$I_{\mathbf{f}}(y) = \sup_{\theta \in \mathbb{R}^m} \big\{ \theta \cdot y - \int_{\mathcal{X}} \Lambda\big[\sum_{j=1}^{m} \theta_j f_j(x) \big]\, R(dx) \big\}, \quad y \in \mathbb{R}^m \tag{3.4}$$

où f_j représente la j$^{\text{e}}$ ligne de la matrice $\mathbf{f}$ ($f_j \in C_d(\mathcal{X})$).

La démonstration du théorème 3.2 est établie en section 3.2.

Remarque 3.3 On remarquera que la fonction de taux $I_{\mathbf{f}}$ s'exprime comme la conjuguée convexe de $\int_{\mathcal{X}} \Lambda[\sum_1^m \theta_j f_j(x)]\, R(dx)$ qui joue ici le rôle de la log-Laplace limite.

3.1.3 Deux contre-exemples

Considérons la probabilité empirique suivante, définie sur l'espace $[0,2]$:

$$P_n = \frac{1}{n}\sum_{i=1}^{n-1}\delta_{\frac{i}{n}} + \frac{\delta_2}{n}.$$

Alors P_n converge faiblement vers $\ell(dx)$ où ℓ représente la mesure de Lebesgue sur $[0,1]$. En tant que mesure sur $[0,2]$, ℓ ne satisfait pas l'hypothèse (H-2). On construit dans cette section une probabilité $\mathbb{P}_Z$ et une suite de variables aléatoires Z_i i.i.d. $\mathbb{P}_Z$-distribuées, satisfaisant (H-4) telles que
• si f est une fonction continue qui prend la valeur 0 sur $[0,1]$ et 1 en 2, alors les variables aléatoires

$$\frac{1}{n}\sum_{i=1}^{n-1} f\left(\frac{i}{n}\right) Z_i + \frac{f(2)}{n} Z_n = \frac{Z_n}{n}$$

ne satisfont pas de principe de grandes déviations.
• si f est une fonction continue qui prend la valeur 1 sur $[0,1]$ et -1 en 2, alors les variables aléatoires

$$\frac{1}{n}\sum_{i=1}^{n-1} f\left(\frac{i}{n}\right) Z_i + \frac{f(2)}{n} Z_n = \frac{1}{n}\sum_{i=1}^{n-1} Z_i - \frac{Z_n}{n}$$

ne satisfont pas de principe de grandes déviations.

Le premier contre-exemple montre comment une particule peut ne pas satisfaire de PGD et illustre l'effet régularisant de la moyennisation : bien que $\frac{Z}{n}$ ne satisfasse pas de PGD, $\frac{1}{n}\sum_1^n Z_i$ satisfait un PGD (théorème de Cramér). On montre également, au lemme 3.6, que $\frac{1}{n}\sum_1^{N(n)} Z_i$ satisfait un PGD sous réserve que $N(n)/n \to \rho > 0$.

Le second contre-exemple, dont l'étude est beaucoup plus délicate, apporte l'éclairage suivant : bien que $\frac{1}{n}\sum_1^{n-1} Z_i$ satisfasse un PGD, l'ajout d'une contribution $-\frac{Z_n}{n}$ brise le PGD, illustrant le fait qu' “en l'absence de tous les moments exponentiels, une simple particule peut modifier un phénomène de grandes déviations”.

Ces remarques nous permettent de bien comprendre l'hypothèse de stricte positivité de la probabilité R (hypothèse (H-2)). Supposons que la particule $\frac{Z}{n}$ ne satisfait pas de PGD et considérons $\frac{1}{n}\sum_1^n f(x_i)Z_i$. L'hypothèse (H-2) assure que, autour de n'importe quel point $x_0 \in \mathcal{X}$, il ne peut pas y avoir une unique particule $f(x_0)\frac{Z_i}{n}$ qui briserait le PGD, comme c'est le cas dans le second contre-exemple. Au contraire, il y a assez de particules (Z_i) autour de x_0 pour que le phénomène de moyennisation précédemment cité ait lieu. Expliquons qualitativement le phénomène. Soit V un voisinage autour de x_0 sur lequel f est presque constante (du fait de sa continuité) : $f \approx a$. Alors,

$$\begin{aligned} \frac{1}{n}\sum_1^n f(x_i)Z_i &= \frac{1}{n}\sum_{x_i\in V} f(x_i)Z_i + \frac{1}{n}\sum_{x_i\notin V} f(x_i)Z_i \\ &\approx \frac{a}{n}\sum_{x_i\in V} Z_i + \cdots . \end{aligned}$$

Comptons le nombre de x_i appartenant à V.

$$N_V(n) = \sharp\{x_i \in V\} = \sum_1^n 1_V(x_i) \quad \text{et} \quad \frac{N_V(n)}{n} \to R(V) > 0,$$

d'après l'hypothèse (H-2) (V est un ouvert non vide) et sous réserve que la frontière de V ait de bonnes propriétés ($R(\partial V) = 0$). Cela nous assure qu'autour de x_0, $\frac{1}{n}\sum_{x_i\in V} Z_i$ satisfait un PGD et que le phénomène mis en évidence lors du second contre-exemple ne peut pas avoir lieu.

On étudie d'abord le cas d'une "particule élémentaire" $\frac{Z}{n}$.

Exemple d'une particule qui ne satisfait pas de PGD. Considérons la variable aléatoire Z dont la distribution est donnée par

$$\mathbb{P}_Z(a_k) = \mathbb{P}\{Z = a_k\} = c\frac{e^{-a_k}}{1+a_k^3} \qquad k \geq 0, \tag{3.5}$$

où $a_k = 16^k$ et c est une constante de normalisation. On montre dans un premier temps que :

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left\{\frac{Z}{n} \in]1,2[\right\} = -\infty. \tag{3.6}$$

Pour cela, considérons la sous-suite $\phi(n) = 2a_n$ alors

$$\mathbb{P}\{Z/\phi(n) \in]1,2[\} = \mathbb{P}\{\phi(n) < Z < 2\phi(n)\} = 0,$$

du fait que $\phi(n) = 2a_n > a_n$ et $2\phi(n) = 4a_n < a_{n+1} = 16a_n$. Ainsi, (3.6) est prouvée. Montrons maintenant que :

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left\{\frac{Z}{n} \in [1+\eta, 2-\eta]\right\} \geq -\frac{8}{7}. \tag{3.7}$$

Etant donnée la sous-suite $\phi(n) = 14a_{n-1}$, on a

$$\mathbb{P}\left\{\frac{Z}{\phi(n)} \in [1+\eta, 2-\eta]\right\} = \mathbb{P}_Z\{a_n\} = c\frac{e^{-a_n}}{1+a_n^3}$$

et $\lim_n 1/\phi(n)\log\mathbb{P}\{Z/\phi(n) \in [1+\eta, 2-\eta]\} = -8/7$. Par suite, (3.7) est établie. Supposons maintenant que $\frac{Z}{n}$ satisfait un PGD de fonction de taux J, alors :

$$\begin{aligned} -8/7 &\leq \limsup_n \frac{1}{n}\log\mathbb{P}\{Z/n \in [1+\eta; 2-\eta]\} \leq -\inf_{u\in[1+\eta;2-\eta]} J(u) \\ -\inf_{u\in]1;2[} J(u) &\leq \liminf_n \frac{1}{n}\log\mathbb{P}\{Z/n \in]1;2[\} \leq -\infty. \end{aligned}$$

On devrait alors avoir $\inf_{u\in[1+\eta;2-\eta]} J(u) < \inf_{u\in]1;2[} J(u)$, ce qui est impossible. Par conséquent, $\frac{Z}{n}$ ne satisfait pas de PGD.

Second contre-exemple. On reprend la même famille de variables aléatoires (Z_i), $\mathbb{P}_Z$-distribuées (voir (3.5)). Le contre-exemple est basé sur l'idée suivante : considérons l'événement $\{\overline{Z}_n - Z/n \in [z-2\,\epsilon; z+2\,\epsilon]\}$ où z est négatif et où $\overline{Z}_n = 1/n\sum Z_i$. Comme $\overline{Z}_n$ est une quantité toujours positive, le seul moyen pour que l'événement précédent soit réalisé est que Z/n soit grand. Comme la "particule" Z/n ne satisfait pas de PGD, il y a de grandes chances que ce soit également le cas pour $\overline{Z}_n - Z/n$. On le démontre dans cette section.

Proposition 3.4 *Les variables aléatoires $(\overline{Z}_n - Z/n)$ ne satisfont pas de PGD.*

la proposition 3.4 repose sur la proposition 3.5. Soit $z_-^* = \Lambda'(-1)$, soit $\epsilon > 0$ fixé et soit $z < z_-^* - 2\epsilon$ alors :

Proposition 3.5 *Il existe un nombre réel fini $A > 0$ tel que*

$$\liminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}\{\overline{Z}_n - Z/n \in [z-2\,\epsilon; z+2\,\epsilon]\} \leq -A. \tag{3.8}$$

Il existe de plus des nombres réels $\alpha \in]0,1[$, $\delta \in]0,(1-\alpha)2\,\epsilon[$ et un nombre réel $B \in]0;A[$ tels que :

$$-B \leq \limsup_n \frac{1}{n}\log\mathbb{P}\left\{\overline{Z}_n - Z/n \in [z+\alpha\,2\,\epsilon-\delta; z+\alpha\,2\,\epsilon+\delta]\right\}. \tag{3.9}$$

La démonstration de la proposition 3.5 est établie dans la section 3.4.

Preuve de la proposition 3.4. Supposons que $(\overline{Z}_n - Z/n)_{n\geq 1}$ satisfasse un PGD de fonction de taux J et choisissons α et δ comme dans la proposition 3.5. Alors :

$$-\inf_{z\in(z-2\,\epsilon;z+2\,\epsilon)} J(z) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{\overline{Z}_n - Z/n \in (z-2\epsilon; z+2\epsilon)\} \overset{(a)}{\leq} -A,$$

où (a) provient de (3.8) et

$$\begin{aligned} -B \overset{(b)}{\leq} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{\overline{Z}_n - Z/n \in [z+\alpha\, 2\epsilon - \delta; z+\alpha\, 2\epsilon+\delta]\right\} \\ \leq -\inf_{[z+\alpha\, 2\epsilon-\delta;z+\alpha\, 2\epsilon+\delta]} J(z), \end{aligned}$$

où (b) provient de (3.9). Comme $B < A$, on devrait avoir

$$\inf_{u\in[z+\alpha\, 2\epsilon-\delta;z+\alpha\, 2\epsilon+\delta]} J(u) < \inf_{u\in(z-2\epsilon;z+2\epsilon)} J(u),$$

ce qui est impossible. □

3.2 Preuve du PGD

La preuve du théorème 3.2 comporte deux parties. Dans un premier temps, nous établissons le PGD (section 3.2.1) puis nous identifions la fonction de taux (section 3.2.2).

3.2.1 Le principe de grandes déviations

La démonstration repose sur plusieurs résultats intermédiaires.

A l'aide d'une version modifiée du théorème de Cramér (lemme 3.6), on établit le PGD pour des fonctions étagées prenant un nombre fini de valeurs (lemme 3.7) : soit $\mathbf{f}(x) = \sum_1^p \mathbf{a}_k 1_{A_k}(x)$ alors

$$\langle L_n, \mathbf{f}\rangle \overset{\mathcal{D}}{=} \frac{1}{n} \sum_{i=1}^{N_{A_1}(n)} \mathbf{a}_1 \cdot Z_i^{(1)} + \cdots + \frac{1}{n} \sum_{i=1}^{N_{A_p}(n)} \mathbf{a}_p \cdot Z_i^{(p)}$$

satisfait le PGD. Finalement, nous montrons à la proposition 3.8 que $(\langle L_n, \mathbf{f}^p\rangle)_{p\geq 1}$ est une approximation exponentielle de $\langle L_n, \mathbf{f}\rangle$ où $(\mathbf{f}^p)$ est une suite bien choisie de fonctions étagées prenant un nombre fini de valeurs. Cette étape est la clé de la preuve et entraine le PGD pour $\langle L_n, \mathbf{f}\rangle$.

Lemme 3.6 *Soit* $(Z_i)_{i\geq 1}$ *une suite de variables aléatoires satisfaisant l'hypothèse (H-4). Supposons de plus que* $(N_A(n))_{n\geq 1}$ *est une suite d'entiers satisfaisant :*

$$\lim_{n\to\infty} \frac{N_A(n)}{n} = R_A > 0.$$

Alors $\left(\frac{1}{N}\sum_1^{N_A(n)} Z_i\right)$ *satisfait le PGD de bonne fonction de taux*

$$I(z) = R_A\ \Lambda^*(\frac{z}{R_A}).$$

Preuve. On note $\Lambda^*(B) = \inf_{z\in B} \Lambda^*(z)$. Remarquons dans un premier temps que

$$\bar{Z}_n^A = \frac{1}{N_A(n)} \sum_{i=1}^{N_A(n)} Z_i$$

satisfait un PGD de bonne fonction de taux Λ^* à la vitesse $N_A(n)$, c'est-à-dire

$$\begin{aligned} -\Lambda^*(\overset{\circ}{B}) &\leq \liminf_{n\to\infty} \frac{1}{N_A(n)} \log \mathbb{P}\left\{\bar{Z}_n^A \in B\right\} \\ -\Lambda^*(\bar{B}) &\geq \limsup_{n\to\infty} \frac{1}{N_A(n)} \log \mathbb{P}\left\{\bar{Z}_n^A \in B\right\}, \end{aligned}$$

où $\overset{\circ}{B}$ (resp. $\bar{B}$) représente l'intérieur (resp. l'adhérence) de B. En effet, cela découle directement du théorème de Cramér dans $\mathbb{R}^d$, valable sous l'hypothèse (H-4) (voir par exemple [20], section 6).
Considérons d'autre part des variables aléatoires dégénérées α_n de loi $\mathbb{P}\{\alpha_n = \frac{N_A(n)}{n}\} = 1$ et indépendantes de $(\bar{Z}_n^A)$. Il est immédiat de vérifier que la suite (α_n) satisfait un PGD à la vitesse $N_A(n)$ de bonne fonction de taux

$$\delta(t|R_A) = \begin{cases} 0 & \text{si } t = R_A, \\ +\infty & \text{sinon.} \end{cases}$$

Par conséquent, le couple $(\alpha_n, \bar{Z}_n^A)$ satisfait un PGD à la vitesse $N_A(n)$, de bonne fonction de taux

$$\delta \oplus \Lambda^*(t,z) = \delta(t|R_A) + \Lambda^*(z)$$

(Lynch et Sethuraman [39], lemme 2.8). Enfin, le principe de contraction entraine que $(\alpha_n \bar{Z}_n^A)$ satisfait un PGD à la vitesse $N_A(n)$ de bonne fonction de taux

$$I(z) = \inf\{\delta \oplus \Lambda^*(t,z'),\ tz' = z\} = \Lambda^*\left(\frac{z}{R_A}\right),$$

soit

$$\begin{aligned} -\Lambda^*(\mathring{B}/R_A) &\le \liminf_{n\to\infty} \frac{1}{N_A(n)} \log \mathbb{P}\left\{ \frac{1}{n} \sum_{i=1}^{N_A(n)} Z_i \in B \right\} \\ -\Lambda^*(\bar{B}/R_A) &\ge \limsup_{n\to\infty} \frac{1}{N_A(n)} \log \mathbb{P}\left\{ \frac{1}{n} \sum_{i=1}^{N_A(n)} Z_i \in B \right\}, \end{aligned}$$

ce qui achève la démonstration. □

Lemme 3.7 *On suppose que (H-3) et (H-4) sont vérifiées. Soit $(A_k)_{1\le k\le p}$ une famille d'ensembles mesurables de $\mathcal{X}$ telle que $R(A_k) > 0$ et telle que chaque A_k soit un ensemble de continuité de R (i.e. $R(\partial A_k) = 0$ où $\partial A_k = \bar{A}_k - \mathring{A}_k$). Considérons la fonction étagée*

$$\mathbf{f}(x) = \sum_{k=1}^{p} \mathbf{a}_k 1_{A_k}(x),$$

les $\mathbf{a}_k$ étant des matrices $m \times d$. Alors

$$\begin{aligned} \langle L_n, \mathbf{f}\rangle &= \frac{1}{n} \sum_{1}^{n} \mathbf{f}(x_i^n) \cdot Z_i \\ &= \frac{1}{n} \sum_{1}^{n} (\mathbf{a}_1 \cdot Z_i)\ \ 1_{A_1}(x_i^n) + \cdots + \frac{1}{n} \sum_{1}^{n} (\mathbf{a}_p \cdot Z_i)\ \ 1_{A_p}(x_i^n) \end{aligned}$$

satisfait le PGD dans $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ avec pour bonne fonction de taux :

$$\begin{aligned} I_{\mathbf{f}}(y) &= \inf\Big\{ \sum_{1}^{p} R(A_k)\Lambda^*(u_k); \quad \sum_{1}^{p} \mathbf{a}_k \cdot u_k\, R(A_k) = y,\ u_k \in \mathbb{R}^d \Big\} \\ &= \inf_{u \in L_d^1} \Big\{ \int_{\mathcal{X}} \Lambda^*(u(x))\, R(dx); \ \int_{\mathcal{X}} \mathbf{f}(x) \cdot u(x)\, R(dx) = y \Big\}, \end{aligned} \tag{3.10}$$

pour $y \in \mathbb{R}^m$.

Preuve. • Soit $\mathbf{f}(x) = \sum_1^p \mathbf{a}_k 1_{A_k}(x)$. On peut réecrire $\langle L_n, \mathbf{f}\rangle$ sous la forme suivante :

$$\langle L_n, \mathbf{f}\rangle = \frac{1}{n} \sum_{i\in I_1(n)} \mathbf{a}_1 \cdot Z_i + \cdots + \frac{1}{n} \sum_{i\in I_p(n)} \mathbf{a}_p \cdot Z_i,$$

où $I_k(n) = \{i \leq n,\ x_i^n \in A_k\}$. Soit $N_{A_k}(n) = \sum_{i=1}^n 1_{A_k}(x_i)$. Il existe p familles indépendantes $\left(\tilde{Z}_i^{(k)}; i \geq 1\right)_{1\leq k\leq p}$ de variables aléatoires i.i.d. et distribuées selon la loi de Z telles que l'égalité suivante ait lieu en distribution :

$$\langle L_n, \mathbf{f}\rangle \overset{\mathcal{D}}{=} \frac{1}{n}\sum_{i=1}^{N_{A_1}(n)} \mathbf{a}_1 \cdot \tilde{Z}_i^{(1)} + \cdots + \frac{1}{n}\sum_{i=1}^{N_{A_p}(n)} \mathbf{a}_p \cdot \tilde{Z}_i^{(p)} \quad \big(\overset{\triangle}{=} \langle \tilde{L}_n, \mathbf{f}\rangle\big). \tag{3.11}$$

• Notons que

$$\lim_{n\to\infty} \frac{N_{A_k}(n)}{n} = \lim_{n\to\infty} \frac{\sum_{i=1}^n 1_{A_k}(x_i^n)}{n} = R(A_k),$$

du fait que A_k est un ensemble de continuité pour R. Par conséquent, on peut appliquer le lemme 3.6 à chaque suite $\frac{1}{n}\sum_{i=1}^{N_{A_k}(n)} \tilde{Z}_i^{(k)}$. De plus, la suite

$$(\zeta_n)_{n\geq 1} = \Big(\frac{1}{n}\sum_{i=1}^{N_{A_1}(n)} \tilde{Z}_i^{(1)}, \ldots, \frac{1}{n}\sum_{i=1}^{N_{A_p}(n)} \tilde{Z}_i^{(p)}\Big)_{n\geq 1}$$

satisfait le PGD de bonne fonction de taux

$$I(z_1, \ldots, z_p) = \sum_{k=1}^p R(A_k)\ \Lambda^*\big(\frac{z_k}{R(A_k)}\big), \qquad z_k \in \mathbb{R}^d$$

(Lynch et Sethuraman [39], lemme 2.8). Par conséquent, le principe de contraction entraine le PGD pour $\langle \tilde{L}_n, \mathbf{f}\rangle$ avec la bonne fonction de taux

$$\begin{aligned} I_{\mathbf{f}}(y) &= \inf\Big\{\sum_{k=1}^p R(A_k)\ \Lambda^*\Big(\frac{z_k}{R(A_k)}\Big),\ \sum_1^p \mathbf{a}_k \cdot z_k = y;\ z_k \in \mathbb{R}^d \text{ pour } 1 \leq k \leq p\Big\} \\ &= \inf\Big\{\sum_1^p R(A_k)\ \Lambda^*(u_k),\ \sum_1^p \mathbf{a}_k \cdot u_k\ R(A_k) = y;\ u_k \in \mathbb{R}^d \text{ pour } 1 \leq k \leq p\Big\}. \end{aligned}$$

Enfin, du fait de (3.11), le PGD a lieu pour $\langle L_n, \mathbf{f}\rangle$.
• On établit maintenant l'égalité suivante :

$$I_{\mathbf{f}}(y) = \inf_{u\in L_d^1}\{\int_{\mathcal{X}} \Lambda^*(u(x))\, R(dx),\ \int_{\mathcal{X}} \mathbf{f}(x)\cdot u(x)\, R(dx) = y\}. \tag{3.12}$$

Supposons que u_ϵ est un ϵ-minimisant, *i.e.*

$$\int_{\mathcal{X}} \Lambda^*(u_\epsilon(x))\, R(dx) \leq I_{\mathbf{f}}(y) + \epsilon, \quad \int \mathbf{f}\cdot u_\epsilon\, dR = y.$$

Considérons $u_k = \int_{A_k} u_\epsilon(x)\, R(dx)/R(A_k)$ alors $\sum_{k=1}^p \mathbf{a}_k \cdot u_k\; R(A_k) = y$ et

$$\sum_1^p R(A_k)\, \Lambda^*(u_k) \overset{\text{(Jensen)}}{\leq} \sum_1^p \int_{A_k} \Lambda^*(u_\epsilon(x)) \frac{R(dx)}{R(A_k)}\, R(A_k) = \int_{\mathcal{X}} \Lambda^*(u_\epsilon(x))\, R(dx).$$

Ainsi $I_{\mathbf{f}}(y) \leq \inf\{\int \Lambda^*(u)\, dR,\ \int \mathbf{f} \cdot u\, dR = y\}$. L'inégalité opposée est évidente. Par suite (3.12) est établi. Ceci achève la démonstration du lemme 3.7. □

Lemme 3.8 *On suppose que (H-2), (H-3) et (H-4) sont vérifiés.*
Soit $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ *une fonction continue bornée, alors il existe une suite* $(\mathbf{f}^p)_{p\geq 1}$ *de fonctions étagées prenant un nombre fini de valeurs et satisfaisant les hypothèses du lemme 3.7, telles que* $(\langle L_n, \mathbf{f}^p\rangle)_{p\geq 0}$ *soit une approximation exponentielle de* $\langle L_n, \mathbf{f}\rangle$*, i.e. :*

$$\lim_{p\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P(|\langle L_n, \mathbf{f}^p\rangle - \langle L_n, \mathbf{f}\rangle| > \delta) = -\infty \quad \textit{pour tout} \quad \delta > 0.$$

De plus, $(\langle L_n, \mathbf{f}\rangle)$ *satisfait le PGD de bonne fonction de taux :*

$$\Upsilon(y) = \sup_{\epsilon>0} \liminf_{p\to\infty} \inf_{y'\in B(y,\epsilon)} I_{\mathbf{f}^p}(y'), \quad y \in \mathbb{R}^m,$$

où $B(y,\epsilon) = \{y' \in \mathbb{R}^m, |y'-y| < \epsilon\}$

La preuve de ce lemme s'appuie sur le lemme précédent (hypothèses (H-3) et (H-4)), la proposition 3.12 et le lemme 3.17 (hypothèse (H-2)).
Preuve du lemme 3.8. • *Approximation de* $\mathbf{f}$ *par des "bonnes" fonctions étagées.*
$\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ est une fonction continue et bornée donc $\mathbf{f}(\mathcal{X})$ est relativement compact dans $\mathbb{R}^{m\times d}$. On peut donc appliquer la proposition 3.12. Celle-ci entraine l'existence d'un recouvrement de $\mathcal{X}$. Plus précisément, il existe $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p \in \mathbb{R}^{m\times d}$ et $\epsilon_1, \ldots, \epsilon_p \leq \epsilon$ tels que

$$\mathcal{X} \subset \cup_{k=1}^p \mathbf{f}^{-1} B(\boldsymbol{\alpha}_k, \epsilon_k) \quad \text{où} \quad R(\partial \mathbf{f}^{-1} B(\boldsymbol{\alpha}_k, \epsilon_k)) = 0.$$

Comme $B(\boldsymbol{\alpha}_k, \epsilon_k)$ est la boule de centre $\boldsymbol{\alpha}_k$ et de rayon ϵ_k, chaque ensemble $\mathbf{f}^{-1}B(\boldsymbol{\alpha}_k, \epsilon_k)$ est ouvert. Du fait de l'hypothèse (H-2), il est soit de mesure strictement positive, soit vide. En ne gardant que les ensembles non vides, on obtient un recouvrement de $\mathcal{X}$ par des ensembles R-continus et de mesure strictement positive.
Les hypothèses du lemme de partition (lemme 3.13) sont satisfaites. Par suite,

il existe une partition $(C_l)_{1\leq l\leq q}$ de $\mathcal{X}$ avec les propriétés énoncées au lemme 3.13. En particulier, pour chaque l, il existe un k tel que

$$C_l \subset \overline{\mathbf{f}^{-1}B(\boldsymbol{\alpha}_k, \epsilon_k)}. \tag{3.13}$$

Considérons une application qui à chaque l associe un unique $k(l)$ tel que (3.13) soit vérifié. On note

$$\mathbf{f}^\epsilon = \sum_{l=1}^{q} \boldsymbol{\alpha}_{k(l)}\ 1_{C_l}.$$

Il est alors immédiat de vérifier que $\|\mathbf{f}^\epsilon - \mathbf{f}\| \leq \epsilon$. De plus, $\mathbf{f}^\epsilon$ vérifie les hypothèses du lemme 3.7. En particulier, $\langle \mathbf{f}^\epsilon, L_n\rangle$ satisfait un PGD de bonne fonction de taux $I_{\mathbf{f}^\epsilon}$. Posons maintenant $p = [\epsilon^{-1}]$ (où $[\cdot]$ désigne la partie entière) et considérons la fonction étagée associée $\mathbf{f}^p$. On obtient ainsi la suite uniformément convergente vers $\mathbf{f}$.

• *L'approximation exponentielle et le PGD faible.* Soit $\mathbf{f}$ une fonction continue bornée à valeurs dans $\mathbb{R}^{m\times d}$ et soit $\mathbf{f}^p$ la fonction étagée approximante telle que nous l'avons précédemment construite. On fixe $\eta > 0$ et on considère $\{|\langle L_n, \mathbf{f}\rangle - \langle L_n, \mathbf{f}^p\rangle| > \eta\}$. Alors, il existe p_ϵ tel que pour $p \geq p_\epsilon$, $\|\mathbf{f}^p - \mathbf{f}\| < \epsilon$. Par suite,

$$\begin{aligned}\{|\langle L_n, \mathbf{f}\rangle - \langle L_n, \mathbf{f}^p\rangle| > \eta\} &= \left\{|1/n \sum_{1}^{n} (\mathbf{f} - \mathbf{f}^p)\cdot Z_i| > \eta\right\} \\ &\subset \left\{1/n \sum_{1}^{n} \epsilon |Z_i| > \eta\right\}\end{aligned}$$

et

$$\begin{aligned}&\limsup_{n\geq 1} \frac{1}{n} \log \mathbb{P}\{|\langle L_n, \mathbf{f}\rangle - \langle L_n, \mathbf{f}^p\rangle| > \eta\} \\ &\qquad\leq \limsup_{n\geq 1} \frac{1}{n} \log \mathbb{P}\left\{1/n \sum_{1}^{n} \epsilon|Z_i| > \eta\right\} \leq -\Lambda^*_{|Z|}\Big(\frac{\eta}{\epsilon}\Big).\end{aligned}$$

Du fait de l'hypothèse (H-4), on a $\lim_{\epsilon\to 0} \Lambda^*_{|Z|}(\eta/\epsilon) = +\infty$. Par suite,

$$\limsup_{p\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|\langle L_n, \mathbf{f}\rangle - \langle L_n, \mathbf{f}^p\rangle| > \eta\} = -\infty.$$

Par suite, $\langle L_n, \mathbf{f}^p\rangle$ est une approximation exponentielle de $\langle L_n, \mathbf{f}\rangle$. Comme chaque $\mathbf{f}^p$ vérifie les hypothèses du lemme 3.7, $\langle L_n, \mathbf{f}^p\rangle$ satisfait le PGD et le théorème 4.2.16 dans [20] entraine que $\langle L_n, \mathbf{f}\rangle$ satisfait un PGD faible de fonction de taux :

$$\Upsilon(y) = \sup_{\epsilon>0} \liminf_{p\to\infty} \inf_{y'\in B(y,\epsilon)} I_{\mathbf{f}^p}(y'), \quad y \in \mathbb{R}^m,$$

où $I_{\mathbf{f}^p}$ est donné par (3.12) et $B(y,\epsilon) = \{y' \in \mathbb{R}^m,\ |y'-y| < \epsilon\}$.

• *Tension exponentielle et le PGD fort.* Il nous reste à montrer que $\langle L_n, \mathbf{f}\rangle$ est exponentiellement tendue. Cela est immédiat du fait de l'inégalité suivante :

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|\langle L_n, \mathbf{f}\rangle| > K\} \\
&\qquad \le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{1/n \sum_1^n |Z_i| > \frac{K}{\|\mathbf{f}\|}\} \le -\Lambda^*_{|Z|}\Big(\frac{K}{\|\mathbf{f}\|}\Big) \xrightarrow[K\to\infty]{} -\infty,
\end{aligned}$$

où la dernière limite provient de l'hypothèse (H-4). Par suite, le PGD intégral est vérifié pour $(\langle L_n, \mathbf{f}\rangle)_{n\ge1}$ et Υ est une bonne fonction de taux. La preuve du lemme 3.8 est ainsi terminée. □

3.2.2 Identification de la fonction de taux

On rappelle que

$$\begin{aligned}
I_{\mathbf{f}}(y) &= \inf_{u\in L^1_d} \Big\{ \int_{\mathcal{X}} \Lambda^*(u(x))\, R(dx);\ \int \mathbf{f}\cdot u\, dR = y; \Big\}, \quad y \in \mathbb{R}^m, \\
\Upsilon(y) &= \sup_{\epsilon>0} \liminf_{p\to\infty} \inf_{y'\in B(y,\epsilon)} I_{\mathbf{f}^p}(y'), \quad y \in \mathbb{R}^m.
\end{aligned}$$

Dans le cas où $\mathbf{f}^p$ est une fonction satisfaisant les hypothèses du lemme 3.7, $I_{\mathbf{f}^p}$ est semi-continue inférieurement en tant que fonction de taux. Cette propriété n'est pas claire si $\mathbf{f} \in C_d(\mathcal{X})$. Pour cela, nous clarifions dans un premier temps le lien entre $I_{\mathbf{f}}$ et Υ (lemme 3.9). C'est un point-clé pour obtenir l'identité de dualité

$$\Upsilon(y) = \sup_{\theta\in\mathbb{R}^m} \{\theta\cdot y - \int_{\mathcal{X}} \Lambda(\sum_{j=1}^m \theta_j f_j(x))\, R(dx)\}.$$

Celle-ci est établie dans le lemme 3.11. Dans la suite, $\mathrm{Lscr}I_{\mathbf{f}}$ représentera la régularisée semi-continue inférieurement de $I_{\mathbf{f}}$, soit

$$\mathrm{Lscr}I_{\mathbf{f}}(y) = \sup_{\epsilon>0} \inf_{y'\in B(y,\epsilon)} I_{\mathbf{f}}(y').$$

Lemme 3.9 *Soit $\mathbf{f} \in C_d(\mathcal{X})$ et supposons que $(\mathbf{f}^p)_{p\ge1}$ satisfait les hypothèses du lemme 3.8. Alors*

$$\Upsilon = \mathrm{Lscr}I_{\mathbf{f}}.$$

L'inégalité suivante, qui nous permet de contrôler la norme L^1 de toute fonction u minimisant $I_{\mathbf{f}}(y)$, nous sera utile dans la suite.

Proposition 3.10 *Soit $u : \mathcal{X} \to \mathbb{R}^d$ une fonction mesurable satisfaisant $\int_{\mathcal{X}} \Lambda^*[u(x)]\, R(dx) \leq M+1$ alors $\int |u|\, dR \leq K_M < \infty$ où K_M dépend de M mais pas de u.*

Preuve de la proposition 3.10. Du fait de l'hypothèse (H-3), il existe $\epsilon > 0$ tel que $\{\lambda \in \mathbb{R}^d,\ |\lambda| = \epsilon\}$ est inclus dans l'intérieur de $\mathcal{D}_\Lambda$ et tel que $\sup_{|\lambda|=\epsilon} \Lambda(\lambda)$ est fini. Par conséquent,

$$\Lambda^*(z) + \sup_{|\lambda|=\epsilon} \Lambda(\lambda) \geq \epsilon |z| \quad \text{pour tout } z \in \mathbb{R}^d.$$

Finalement, on obtient une majoration indépendante de u en intégrant l'inégalité de part et d'autre. □

Preuve du lemme 3.9. On appelle contrôle toute fonction u qui vérifie $\int \mathbf{f} \cdot u\, dR = y$. On note

$$\begin{aligned}
\mathcal{C}_p(y,\epsilon) &\triangleq \left\{ u \in L^1_d(\mathcal{X}),\ |\int \mathbf{f}^p \cdot u\, dR - y| \leq \epsilon \right\}, \quad \mathcal{C}_p(y) \triangleq \mathcal{C}_p(y,0),\\
\mathcal{C}(y,\epsilon) &\triangleq \left\{ u \in L^1_d(\mathcal{X}),\ |\int \mathbf{f} \cdot u\, dR - y| \leq \epsilon \right\}, \quad \mathcal{C}(y) \triangleq \mathcal{C}(y,0).
\end{aligned}$$

• Montrons dans un premier temps que

$$I_{\mathbf{f}}(y) < \infty \quad \Rightarrow \quad \Upsilon(y) \leq I_{\mathbf{f}}(y). \tag{3.14}$$

Soit $I_{\mathbf{f}}(y) = M < \infty$ alors, grâce à la proposition 3.10,

$$I_{\mathbf{f}}(y) = \inf_{u \in L^1_d} \{ \int \Lambda^*(u)\, dR;\ \int \mathbf{f} \cdot u\, dR = y;\ \int |u|\, dR \leq K_M \}.$$

Fixons $\epsilon > 0$ alors il existe p^*_ϵ tel que $\|\mathbf{f}^p - \mathbf{f}\| \leq \epsilon$ pour $p \geq p^*_\epsilon$. Par conséquent, on obtient, pour $p \geq p^*_\epsilon$ et sous la condition $\int |u|\, dR < K_M$:

$$\mathcal{C}(y) \subset \mathcal{C}_p(y, \epsilon K_M).$$

Par suite, pour $p \geq p^*_\epsilon$,

$$\inf_{y' \in B(y, \epsilon K_M)} \left\{ \int \Lambda^*(u)\, dR,\ u \in \mathcal{C}_p(y') \right\} \leq I_{\mathbf{f}}(y).$$

Et pour tout $\epsilon > 0$,

$$\liminf_{p\to\infty} \inf_{y'\in B(y,\epsilon K_M)} \Big\{ \int \Lambda^*(u)\, dR,\ \int \mathbf{f}^p \cdot u\, dR = y' \Big\} \leq I_{\mathbf{f}}(y).$$

Finalement

$$\Upsilon(y) = \sup_{\epsilon>0} \liminf_{p\to\infty} \inf_{y'\in B(y,\epsilon K_M)} \Big\{ \int \Lambda^*(u)\, dR,\ \int \mathbf{f}^p \cdot u\, dR = y' \Big\} \leq I_{\mathbf{f}}(y).$$

Et (3.14) est établi.
• Montrons maintenant que

$$\Upsilon(y) < \infty \quad \Rightarrow \quad \mathrm{lscr} I_{\mathbf{f}}(y) \leq \Upsilon(y). \tag{3.15}$$

Si $\Upsilon(y) = M < \infty$, la proposition 3.10 entraine

$$\Upsilon(y) = \sup_{\epsilon>0} \liminf_{p\to\infty} \inf_{y'\in B(y,\epsilon)} \{ \int \Lambda^*(u)\, dR,\ u \in \mathcal{C}_p(y'),\ \int |u|\, dR \leq K_M \}.$$

Fixons $\epsilon > 0$. Il existe alors p^*_ϵ tel que $\|\mathbf{f}^p - \mathbf{f}\| \leq \epsilon$ pour $p \geq p^*_\epsilon$. Par conséquent, on obtient, pour $p \geq p^*_\epsilon$ et sous la condition $\int |u|\, dR < K_M$:

$$\mathcal{C}_p(y,\epsilon) \subset \mathcal{C}(y, \epsilon(1+K_M)).$$

Ainsi,

$$\inf\{ \int \Lambda^*(u)\, dR,\ u \in \mathcal{C}(y,(1+K_M)\epsilon)\} \leq \inf\{ \int \Lambda^*(u)\, dR,\ u \in \mathcal{C}_p(y,\epsilon)\}.$$

Et pour tout $\epsilon > 0$,

$$\inf\{ \int \Lambda^*(u)\, dR,\ u \in \mathcal{C}(y,(1+K_M)\epsilon)\} \leq \liminf_{p} \inf\{ \int \Lambda^*(u)\, dR,\ u \in \mathcal{C}_p(y,\epsilon)\}.$$

Finalement,

$$\sup_{(1+K_M)\epsilon>0} \inf_{y'\in B(y,(1+K_M)\epsilon)} I_{\mathbf{f}}(y') \leq \Upsilon(y).$$

C'est précisément la propriété désirée du fait que $\sup_{(1+K_M)\epsilon>0} \inf_{y'\in B(y,(1+K_M)\epsilon)} I_{\mathbf{f}}(y')$ est la régularisée semi-continue inférieurement de $I_{\mathbf{f}}$.
• Comme Υ est semi-continue inférieurement et $\mathrm{lscr} I_{\mathbf{f}} \leq \Upsilon \leq I_{\mathbf{f}}$, le lemme 3.9 est établi. □

Lemme 3.11 *Soit* $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ *une fonction continue bornée. On rappelle que*

$$I_{\mathbf{f}}(y) = \inf_{u\in L^1_d}\{\int \Lambda^*(u)\, dR; \quad \int \mathbf{f}\cdot u\, dR = y\},$$

où $y \in \mathbb{R}^m$*. Alors, l'identité suivante a lieu*

$$lscr\, I_{\mathbf{f}}(y) = \sup_{\theta\in\mathbb{R}^m}\{\theta\cdot y - \int \Lambda(\sum_1^m \theta_j f_j)\, dR\},$$

où f_j *représente la* j^e *ligne de la matrice* $\mathbf{f}$*. En particulier,*

$$\Upsilon(y) = \sup_{\theta\in\mathbb{R}^m}\{\theta\cdot y - \int \Lambda(\sum_1^m \theta_j f_j)\, dR\}.$$

Preuve. Introduisons l'espace des fonctions mesurables, essentiellement bornées et à valeurs $\mathbb{R}^d$, L^∞_d. Les fonctionnelles $\int \Lambda^*\, dR$ et $\int \Lambda\, dR$ sont conjuguées convexes pour la dualité (L^1_d, L^∞_d). (voir par exemple les articles de Rockafellar [47], [48] et la discussion intéressante au début du paragraphe 4, [47]). Cela signifie que les identités suivantes ont lieu :

$$\begin{aligned}
\int \Lambda^*(u)dR &= \sup_{g\in L^\infty_d}\{\int u\cdot g\ dR - \int \Lambda(g)dR\} \quad \text{pour tout } u\in L^1_d,\\
\int \Lambda(g)dR &= \sup_{u\in L^1_d}\{\int g\cdot u\ dR - \int \Lambda^*(u)dR\} \quad \text{pour tout } g\in L^\infty_d.
\end{aligned}$$

Rappelons que f_i (j$^{\mathrm{e}}$ ligne de la matrice $\mathbf{f}$) est un élément de $C_d(\mathcal{X})$. À ce titre, f_i peut être identifiée à un élément de L^∞_d. Considérons l'opérateur $A : L^1_d \to \mathbb{R}^m$ et son adjoint $A^* : \mathbb{R}^m \to L^\infty_d$ définis par :

$$Au = \int \mathbf{f}\cdot u\, dR\,,\ u\in L^1_d \quad \text{et} \quad A^*y = \sum_{j=1}^m y_j f_j\,,\ y\in\mathbb{R}^m.$$

En utilisant le fait que $\int \Lambda dR$ et $\int \Lambda^* dR$ sont conjuguées convexes, la première partie du lemme est une application directe du théorème 3 dans [48] :

$$\sup_{\theta\in\mathbb{R}^m}\left\{\theta\cdot y - \int \Lambda(A^*y)\, dR\right\} = \mathrm{lscr}\inf_{u\in L^1_d}\left\{\int \Lambda^*(u)\, dR;\ Au = y\right\}.$$

La seconde provient du lemme 3.9. $\square$

3.2.3 Preuve du théorème 3.2

Preuve. Le lemme 3.8 permet d'établir le PGD. Les lemmes 3.9 et 3.11 permettent l'identification de la fonction de taux. Cela achève la démonstration du théorème 3.2. □

3.3 Deux résultats techniques

On a vu au lemme 3.7 que les fonctions en escalier $\mathbf{f}(x) = \sum_{k=1}^{p} \mathbf{a}_k 1_{A_k}(x)$ vérifiant

$$\inf\{R(A_k),\ 1 \leq k \leq p\} > 0 \quad \text{et} \quad \sup\{R(\partial A_k),\ 1 \leq k \leq p\} = 0,$$

sont utiles pour obtenir des PGD. Les deux résultats qui suivent nous permettent de construire des fonctions en escalier ayant ces propriétés et approximant (au sens de la norme de la convergence uniforme) n'importe quelle fonction continue bornée. Ces résultats sont utilisés dans la preuve du lemme 3.8.

3.3.1 Comment obtenir de la R-continuité dans un recouvrement de $\mathcal{X}$?

Dans la proposition suivante, $\mathcal{X}$ est un espace topologique muni de sa tribu borélienne et d'une probabilité R et $\mathcal{Y}$, un espace métrique muni également de sa tribu borélienne. Comme d'habitude, $B(y, \epsilon)$ est la boule de centre $y \in \mathcal{Y}$ et de rayon $\epsilon > 0$. Rappelons qu'on dit qu'un ensemble A est R-continu si la mesure de sa frontière est nulle, *i.e.* :

$$R(\partial A) = 0 \quad \text{où} \quad \partial A = \bar{A} - \mathring{A}.$$

Ici, $\bar{A}$ représente l'adhérence de A et $\mathring{A}$ son intérieur. On utilisera dans la suite l'inclusion suivante

$$\partial f^{-1}(A) \subset f^{-1}(\partial A), \tag{3.16}$$

valable pour toute fonction $f : \mathcal{X} \to \mathcal{Y}$ continue.

Proposition 3.12 *Soit $f : \mathcal{X} \to \mathcal{Y}$ une fonction continue. Supposons de plus que l'image de $\mathcal{X}$ par f, $f(\mathcal{X})$, est relativement compacte dans $\mathcal{Y}$. Alors, pour tout $\epsilon > 0$, il existe $(y_k; 1 \leq k \leq p) \subset \mathcal{Y}$ et $\epsilon_1, \cdots, \epsilon_p \in (0, \epsilon]$ tels que*

$$\mathcal{X} \subset \cup_{k=1}^{p} f^{-1}B(y_k, \epsilon_k) \quad \textit{où} \quad R(\partial f^{-1}B(y_k, \epsilon_k)) = 0 \quad \textit{pour } k \in \{1, \ldots, p\}.$$

Preuve. On fixe $\epsilon > 0$ et on note

$$\begin{aligned} \mathcal{I}_\epsilon(y) &= \{\epsilon' \in (0,\epsilon],\ R(\partial f^{-1}B(y,\epsilon')) > 0\}, \\ \mathcal{J}_\epsilon(y) &= \{\epsilon' \in (0,\epsilon],\ R(\partial f^{-1}B(y,\epsilon')) = 0\}. \end{aligned}$$

Alors $\mathcal{I}_\epsilon(y)$ est au plus dénombrable. En effet, considérons

$$\begin{aligned} \phi : (0,\epsilon] &\to [0,1], \\ \epsilon' &\mapsto R \circ f^{-1}\ B(y,\epsilon'). \end{aligned}$$

ϕ est une application croissante et continue à gauche. Par conséquent, elle n'admet au plus qu'un nombre dénombrable de discontinuités. Montrons que si ϕ est continue en ϵ_0, alors :

$$R \circ f^{-1}\left[\partial B(y,\epsilon_0)\right] = 0. \tag{3.17}$$

En effet,

$$\lim_{\epsilon' \searrow \epsilon_0} R \circ f^{-1}B(y,\epsilon') = R \circ f^{-1}B(y,\epsilon_0)$$

par continuité. Mais

$$\begin{aligned} \lim_{\epsilon' \searrow \epsilon_0} R \circ f^{-1}B(y,\epsilon') &= R \circ f^{-1} \cap_{\epsilon' > \epsilon} B(y,\epsilon') \\ &= R \circ f^{-1}\bar{B}(y,\epsilon_0), \end{aligned}$$

où $\overline{B}(y,\epsilon_0)$ représente la boule fermée centrée en y et de rayon ϵ_0. Enfin,

$$R \circ f^{-1}\left[\partial B(y,\epsilon_0)\right] = R \circ f^{-1}\overline{B}(y,\epsilon_0) - R \circ f^{-1}B(y,\epsilon_0) = 0.$$

L'inclusion (3.16) entraine maintenant que pour un tel ϵ_0,

$$R\partial f^{-1}\overline{B}(y,\epsilon_0) \leq R \circ f^{-1}\left[\partial B(y,\epsilon_0)\right] = 0.$$

Par conséquent, $\mathcal{I}_\epsilon(y)$ est dénombrable. Comme $\mathcal{I}_\epsilon(y) \cup \mathcal{J}_\epsilon(y) = (0,\epsilon]$, $\mathcal{J}_\epsilon(y)$ n'est jamais vide et

$$f(\mathcal{X}) \subset \cup_{y \in f(\mathcal{X}),\ \epsilon' \in \mathcal{J}_\epsilon(y)} B(y,\epsilon').$$

Du fait de la relative compacité de $f(\mathcal{X})$, il existe $(y_k; 1 \leq k \leq p) \subset \mathcal{Y}$ et $(\epsilon_k; 1 \leq k \leq p) \subset (0,\infty)$ vérifiant

$$f(\mathcal{X}) \subset \cup_{k=1}^{p} B(y_k,\epsilon_k) \quad \Rightarrow \quad \mathcal{X} \subset \cup_{k=1}^{p} f^{-1}B(y_k,\epsilon_k).$$

De plus, comme $\epsilon_k \in \mathcal{J}_\epsilon(y_k)$, $\quad R(\partial f^{-1}B(y_k,\epsilon_k)) = 0$. La proposition 3.12 est ainsi démontrée. $\square$

3.3.2 Un lemme pour partitionner

Lemme 3.13 *Supposons que (H-2) est vérifiée. Soit* $(A_k; 1 \leq k \leq p)$ *un recouvrement mesurable de* $\mathcal{X}$ *satisfaisant* $R(A_1) > 0$ *et* $R(\partial A_k) = 0$ *pour* $1 \leq k \leq p$*. Alors il existe une partition* $(B_l; 1 \leq l \leq q)$ *de* $\mathcal{X}$ *satisfaisant* $B_1 = \overline{A}_1$*,* $R(B_l) > 0$ *et* $R(\partial B_l) = 0$ *pour* $1 \leq l \leq q$*. De plus, pour chaque* B_l*, il existe un* A_k *tel que* $B_l \subset \overline{A}_k$*.*

Preuve. Nous procédons par réccurence sur p. Si $p = 1$, on prend $B_1 = \overline{A}_1$ et le résultat est démontré. Soit $p > 1$. On peut toujours supposer que le recouvrement initial est constitué d'ensembles fermés. En effet, si $(A_k)_{1\leq k\leq p}$ est un recouvrement satisfaisant les hypothèses du lemme, il en va de même pour $(\overline{A}_k)_{1\leq k\leq p}$. Si $\mathcal{X} \subset \overline{A}_1$ alors la partition est réduite au seul élément $B_1 = \overline{A}_1$. Sinon, $\mathcal{X}\backslash\overline{A}_1$ est un ouvert non vide et $R(\mathcal{X}\backslash\overline{A}_1) > 0$ du fait de (H-2).
Il existe nécessairement k satisfaisant $R(\overline{A}_k\backslash\overline{A}_1) > 0$. En effet

$$0 < R(\mathcal{X}\backslash\overline{A}_1) \leq \sum_2^p R(\overline{A}_k\backslash\overline{A}_1).$$

Considérons maintenant la famille $\{\overline{A}_1 \cup \overline{A}_k;\ A_j,\ 2 \leq j \leq p, j \neq k\}$. C'est un recouvrement de $\mathcal{X}$ constitué de $p-1$ éléments satisfaisant les hypothèses du lemme (on rappelle que les ensembles R-continus forment une algèbre sur $\mathcal{X}$). Par conséquent, on peut appliquer l'hypothèse de récurrence et il existe une partition $(B_l; 1 \leq l \leq q)$ où $B_1 = \overline{A}_1 \cup \overline{A}_k$, $R(\partial B_l) = 0$ et $R(B_l) > 0$ pour $1 \leq l \leq q$. Séparons maintenant B_1 en $C_1 = \overline{A}_1$ et $C_2 = \overline{A}_k\backslash\overline{A}_1$ alors la partition $\{C_1;\ C_2;\ B_l,\ 2 \leq l \leq q\}$ fait l'affaire et le lemme 3.13 est démontré. □

3.4 Fin de la preuve du second contre-exemple

Nous nous appuyons sur deux résultats intermédiaires pour démontrer la proposition 3.5. Les démonstrations de ces résultats (Proposition 3.14 et 3.15), bien que reposant sur des arguments classiques, sont données intégralement. On introduit les notations suivantes :

$$\begin{aligned}
\delta(\lambda|_{[-1;+\infty)}) &= \begin{cases} 0 & \text{si } \lambda \in [-1;\infty), \\ +\infty & \text{sinon.} \end{cases}, \\
\delta^*(z) &= \sup_{\lambda\in\mathbb{R}}\{\lambda\, z - \delta(\lambda|_{[-1;+\infty)})\} = \begin{cases} -z & \text{si } z \leq 0, \\ +\infty & \text{sinon.} \end{cases}, \\
I(z) &= \inf\{\Lambda^*(x) + \delta^*(y),\ x + y = z\}.
\end{aligned}$$

On remarquera que δ^* aurait joué le rôle de la fonction de taux associée à la particule $-Z_n/n$ si celle-ci satisfaisait un PGD. De même, $I(z)$ aurait été la fonction de taux associée à $\bar{Z}_n - Z_n/n$.

Proposition 3.14 *On rappelle que $z_*^- = \Lambda'(-1)$.*
- *Si $z \geq z_*^-$ alors $I(z) = \Lambda^*(z)$.*
- *Si $z < z_*^-$ alors $I(z) = \Lambda^*(z_*^-) + z_*^- - z$.*

Dans le cas où $z < z_*^-$, l'infimum $\inf\{\Lambda^*(x) + \delta^*(y),\ x + y = z\}$ est atteint de manière unique pour $x = z_*^-$ et $y = z - z_*^-$.

Preuve. On remarque que :

$$I(z) = \sup_{\lambda \in \mathbb{R}}\{\lambda\ z - \Lambda(\lambda) - \delta(\lambda|_{[-1;+\infty)})\} = \sup_{\lambda \in [-1;1]}\{\lambda\ z - \Lambda(\lambda)\}.$$

On appelle $z_*^+ = \Lambda'(1)$.
•Si $z \geq z_*^+$ alors

$$\begin{aligned} I(z) &= \sup_{\lambda \in \mathbb{R}}\{\lambda\ (z - z_*^+) + \lambda\ z_*^+ - \Lambda(\lambda)\} \\ &\overset{(a)}{=} z - z_*^+ + \Lambda^*(z_*^+) \overset{(b)}{=} \Lambda^*(z), \end{aligned}$$

où (a) et (b) sont standards.
•Si $z_*^- \leq z \leq z_*^+$ alors il existe $\lambda_z \in [-1;1]$ tel que $z = \Lambda'(\lambda_z)$. Par suite $I(z) = z\lambda_z - \Lambda(\lambda_z) = \Lambda^*(z)$.
•Si $z \leq z_*^-$ alors

$$I(z) = \sup_{\lambda \in [-1;1]}\{\lambda\ (z - z_*^-) + \lambda\ z_*^- - \Lambda(\lambda)\} = z_*^- - z + \Lambda^*(z_*^-).$$

Pour montrer la dernière partie de la proposition, considérons la fonction $\Lambda^*(x) + \delta^*(z-x)$. Du fait de la deuxième partie de la proposition, cette fonction atteint son minimum pour $x = z_-^*$. Comme $\Lambda^*(x) + \delta^*(z - x)$ est strictement convexe dans un voisinage de z_-^*, ce minimum est unique et la proposition est démontrée. □

Proposition 3.15 *Fixons $\epsilon > 0$ et $z < z_-^* - 2\epsilon$. On considère les quantités suivantes :*

$$\begin{aligned} I_\epsilon &= \inf\{\Lambda^*(x) + \delta^*(y),\ x + y \in [z - 2\,\epsilon; z + 2\,\epsilon],\ x \notin (z_*^- - \epsilon; z_*^- + \epsilon)\}, \\ I_0 &= \inf\{I(u),\ u \in [z - 2\,\epsilon; z + 2\,\epsilon]\} = \Lambda^*(z_*^-) + z_*^- - z - 2\,\epsilon. \end{aligned}$$

Alors, $I_\epsilon > I_0$.

Preuve. Notons tout d'abord que I_0 et I_ϵ sont toujours finis et que $I_\epsilon \geq I_0$. Supposons que $I_\epsilon = I_0$. Soit (x_n, y_n) une suite d'éléments minimisants, c'est-à-dire

$$\begin{gathered} x_n + y_n \in [z - 2\,\epsilon; z + 2\,\epsilon], \\ x_n \notin (z_*^- - \epsilon; z_*^- + \epsilon), \\ \lim_{n\to\infty} [\Lambda^*(x_n) + \delta^*(y_n)] = I_\epsilon \quad = I_0. \end{gathered}$$

Du fait de la compacité des ensembles de niveaux de Λ^* et δ^*, on peut prouver qu'il existe un minimisant (x_*, y_*) satisfaisant :

$$\begin{gathered} x_* + y_* \in [z - 2\,\epsilon; z + 2\,\epsilon], \\ x_* \notin (z_*^- - \epsilon; z_*^- + \epsilon), \\ \Lambda^*(x_*) + \delta^*(y_*) = I_0 = \Lambda^*(z_*^-) + z_*^- - z - 2\epsilon. \end{gathered}$$

Le deuxième point de la proposition 3.14 entraine que $x_* = z_*^-$, ce qui contredit le fait que $x_* \notin (z_*^- - \epsilon; z_*^- + \epsilon)$. Nécessairement, $I_\epsilon > I_0$ et la proposition 3.15 est démontrée. □

On peut aborder maintenant la démonstration de la proposition 3.5.
Preuve de la proposition 3.5. • On démontre dans un premier temps qu'il existe une sous-suite $\phi(n)$ vérifiant

$$\limsup_{n\to\infty} \frac{1}{\phi(n)} \log \mathbb{P}\{\bar{Z}_{\phi(n)} - Z_{\phi(n)}/\phi(n) \in (z - 2\epsilon; z + 2\epsilon)\} \leq -I_\epsilon < -I_0, \quad (3.18)$$

où I_ϵ et I_0 sont définis dans la proposition 3.15. Cela entrainera la première partie de la proposition 3.5. On utilise les notations suivantes :

$$\begin{aligned} B_k^1 &= [z_*^- + 2k\epsilon - \epsilon; z_*^- + 2k\epsilon + \epsilon), \\ B_k^2 &= (z - z_*^- - 2k\epsilon - 3\epsilon; z - z_*^- - 2k\epsilon + 3\epsilon]. \end{aligned}$$

L'inclusion suivante est immédiate :

$$\left\{ \bar{Z}_n - \frac{Z_n}{n} \in [z - 2\epsilon; z + 2\epsilon] \right\} \subset \bigcup_{k\in\mathbb{Z}} \{\bar{Z}_n \in B_k^1\} \cap \left\{ \frac{-Z_n}{n} \in B_k^2 \right\}.$$

L'union précédente est une union d'ensembles disjoints. De plus, $\{\bar{Z}_n \in B_k^1\}$ est vide si k est négatif avec $|k|$ assez grand (disons $k \leq k^- < 0$) du fait que $\bar{Z}_n$ est une variable aléatoire positive. Par suite,

$$\mathbb{P}\left\{ \bar{Z}_n - \frac{Z_n}{n} \in [z - 2\epsilon; z + 2\epsilon] \right\} \leq \sum_{k\geq k^-} \mathbb{P}\left\{\bar{Z}_n \in B_k^1\right\} \mathbb{P}\left\{ \frac{-Z_n}{n} \in B_k^2 \right\}$$

Fixons $L > I_\epsilon$. Le théorème de Cramér entraine l'existence de k^+ tel que

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{\bar{Z}_n \geq z_-^* + k^+\,\epsilon\right\} \leq -L$$

et

$$\begin{aligned}\mathbb{P}&\left\{\bar{Z}_n - \frac{Z_n}{n} \in [z-2\epsilon; z+2\epsilon]\right\}\\ &\qquad\leq \sum_{k=k^-}^{k^+} \mathbb{P}\{\bar{Z}_n \in B_k^1\}\mathbb{P}\left\{\frac{-Z_n}{n} \in B_k^2\right\}\\ &\qquad\qquad + \mathbb{P}\left\{\bar{Z}_n \geq z_-^* + k^+\,\epsilon\right\}.\end{aligned}$$

Voici comment nous allons procéder : on sait grâce à la proposition 3.14 que l'infimum

$$\inf\{I(y); y \in [z-2\epsilon; z+2\epsilon]\} = I(z-2\epsilon)$$

est atteint de manière unique pour $I(z-2\epsilon) = \Lambda^*(z_*^-) + \delta^*(z - z_*^- - 2\epsilon)$. Considérons

$$\mathcal{T} = \{\bar{Z}_n \approx z_*^-\} \cap \{-Z_n/n \approx z - z_*^- - 2\epsilon\}.$$

On peut voir l'événement $\mathcal{T}$ comme étant le sous-ensemble le "plus typique" de $\{\bar{Z}_n - Z_n/n \in [z-2\epsilon; z+2\epsilon]\}$ au sens où l'infimum de la fonction de taux est "réalisé" sur $\mathcal{T}$. Si on choisit une sous-suite $\phi(n)$ telle que

$$\left\{-\frac{Z_{\phi(n)}}{\phi(n)} \approx z - z_*^- - 2\epsilon\right\} = \emptyset,$$

alors la borne supérieure du phénomène de grandes déviations va décroître et on devrait obtenir (3.18). Formalisons tout ça :
Considérons la sous-suite définie par $\phi(n) = \left[\frac{2a_n}{z_-^* + \epsilon - z}\right]$, où $[x]$ représente la partie entière de x. Il est immédiat de montrer que

$$\mathbb{P}\left\{-\frac{Z_{\phi(n)}}{\phi(n)} \in (z - z_*^- - 3\epsilon; z - z_*^- + 3\epsilon]\right\} = 0,$$

pour n assez grand. Par conséquent,

$$\begin{aligned}\mathbb{P}&\{\bar{Z}_{\phi(n)} - Z_{\phi(n)}/\phi(n) \in [z-2\epsilon; z+2\epsilon]\}\\ &\qquad\leq \sum_{k\neq 0; k=k^-}^{k^+} \mathbb{P}\{\bar{Z}_{\phi(n)} \in B_k^1\}\mathbb{P}\{-Z_{\phi(n)}/\phi(n) \in B_k^2\}\\ &\qquad\qquad + \mathbb{P}\left\{\bar{Z}_{\phi(n)} \geq z_-^* + k^+\,\epsilon\right\}.\end{aligned}$$

Soit $k \in \{k^-, \dots, k^+\}$, $k \neq 0$. Les techniques usuelles pour établir les bornes supérieures de grandes déviations entrainent :

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\phi(n)} \log \mathbb{P}\{\bar{Z}_{\phi(n)} \in B_k^1\}\mathbb{P}\{-Z_n/\phi(n) \in B_k^2\} \\
&\leq \ -\inf\{\Lambda^*(x),\ x \in B_k^1\} - \inf\{\delta^*(y),\ y \in B_k^2\} \\
&\leq \ -\inf\{\Lambda^*(x) + \delta^*(y),\ x + y \in [z - 2\epsilon; z + 2\epsilon],\ x \in B_k^1\}.
\end{aligned}$$

Enfin, du fait du lemme 1.2.15 de [20],

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\phi(n)} \log \mathbb{P}\{\bar{Z}_{\phi(n)} - Z_{\phi(n)}/\phi(n) \in [z - 2\epsilon; z + 2\epsilon]\} \\
&\leq \sup_{k\neq 0;\ k^-\leq k\leq k^+} -\inf\{\Lambda^*(x) + \delta^*(y),\ x + y \in [z - 2\epsilon; z + 2\epsilon],\ x \in B_k^1\} \\
&\qquad\qquad \vee (-L) \\
&\leq \ -\inf\{\Lambda^*(x) + \delta^*(y),\ x + y \in [z - 2\epsilon; z + 2\epsilon],\ x \notin (z_-^* - \epsilon; z_-^* + \epsilon)\} \\
&\leq \ -I_\epsilon < -I_0,
\end{aligned}$$

où $a \vee b = \sup(a; b)$ et où la dernière inégalité provient de la proposition 3.15. On a ainsi démontré la première partie de la proposition 3.5.

• Démontrons maintenant la seconde partie.

On considère

$$\left\{ \bar{Z}_n - \frac{Z_n}{n} \in [z + \alpha 2\epsilon - \delta; z + \alpha 2\epsilon + \delta] \right\},$$

où $\alpha < 1$ et $\delta < (1 - \alpha)2\,\epsilon$. Alors

$$\begin{aligned}
&\left\{\bar{Z}_n \in [z_*^- - \delta/2; z_*^- + \delta/2]\right\} \cap \left\{ \frac{-Z_n}{n} \in [z - z_*^- + \alpha 2\epsilon - \delta/2; z - z_*^- + \alpha 2\epsilon + \delta/2] \right\} \\
&\qquad\qquad \subset \left\{ \bar{Z}_n - \frac{Z_n}{n} \in [z + \alpha 2\epsilon - \delta; z + \alpha 2\epsilon + \delta] \right\}.
\end{aligned}$$

Choisissons une sous-suite d'entiers définie par les inégalités suivantes :

$$\begin{cases}
a_{n-1} & < \phi(n)(z_*^- - z - \alpha\, 2\,\epsilon - \delta/2) \quad \leq a_n, \\
a_n & \overset{(a)}{\leq} \phi(n)(z_*^- - z - \alpha\, 2\,\epsilon + \delta/2) \quad < a_{n+1},
\end{cases}$$

où $a_n = 16^n$. Une telle sous-suite existe toujours et

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{\phi(n)} \log \mathbb{P}\left\{ -\frac{Z_{\phi(n)}}{\phi(n)} \in [z - z_*^- + \alpha 2\epsilon - \delta/2; z - z_*^- + \alpha 2\,\epsilon + \delta/2] \right\} \\
&= \liminf_{n\to\infty} \frac{1}{\phi(n)} \log \mathbb{P}\{Z = a_n\} = \liminf_{n\to\infty} -\frac{a_n}{\phi(n)} \geq -(z_*^- - z - \alpha\, 2\,\epsilon + \delta/2),
\end{aligned}$$

où la dernière inégalité provient de (a). Le théorème de Cramér entraine

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{\bar{Z}_n \in [z_*^- - \delta/2; z_*^- + \delta/2]\} \geq -\Lambda^*(z_*^-).$$

Par suite, on obtient :

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{\bar{Z}_n - \frac{Z_n}{n} \in [z + \alpha 2\epsilon - \delta; z + \alpha 2\epsilon + \delta]\right\} \\ \geq -(\Lambda^*(z_*^-) + z_*^- - z - \alpha\, 2\, \epsilon + \delta/2).$$

Finalement, comme

$$\lim_{\alpha\to 1; \delta < (1-\alpha)\, 2\, \epsilon} (\Lambda^*(z_*^-) + z_*^- - z - \alpha\, 2\, \epsilon + \delta/2) = I_0$$

et $I_0 < I_\epsilon$, il existe $\alpha > 0,\ \delta > 0$ et $\gamma > 0$ vérifiant $\Lambda^*(z_*^-) + z_*^- - z - \alpha\, 2\, \epsilon + \delta/2 \leq I_\epsilon - \gamma$. La seconde partie de la proposition est ainsi démontrée. $\square$

Chapitre 4

Grandes déviations pour le maximum d'entropie en moyenne dans le cas i.i.d. et applications

Sommaire

Nous utilisons dans ce chapitre le PGD démontré au chapitre 3 pour établir plusieurs PGD dans un contexte infini-dimensionnel. On établit ainsi un PGD pour la suite de mesures empiriques suivante :

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

La fonction de taux associée à ce PGD est de la forme :

$$I(\mu) = \int_{\mathcal{X}} \Lambda^* \Big(\frac{d\mu_a}{dR} \Big)\, dR + I^s(\mu_s).$$

Ici, μ a la décomposition de Lebesgue $\mu = \mu_a + \mu_s$, μ_a étant absolument continue par rapport à R et les variables Z_i n'ont que quleques moments exponentiels :

$$Ee^{\alpha |Z_i|} < +\infty \quad \text{pour un } \alpha > 0. \tag{4.1}$$

Cette famille présente un intérêt du point de vue des applications. Elle a été étudiée par Dacunha-Castelle et Gamboa [16] et le PGD a été établi par Gamboa et Gassiat [29] dans le cas i.i.d. et par Csiszár, Gamboa et Gassiat [15] dans le cas de variables échangeables. Ces PGD s'avèrent extrêmement utiles pour prouver des résultats de convergence via le principe du Maximum d'entropie en moyenne (MEM) en théorie de l'estimation. De fait, nous appelerons L_n la mesure empirique MEM. Ellis *et al.* ont également établi un PGD similaire dans [28] pour étudier les propriétés de mécanique statistique du gaz de Bose. On améliore les travaux précédents dans deux directions. Nous considérons des variables aléatoires à valeurs $\mathbb{R}^d$ et nous nous affranchissons de l'hypothèse d'escarpement qui est présente dans dans ces deux articles. Nous montrons également que la stricte positivité de R est une condition nécessaire : deux contre-exemples ont été construits au chapitre précédent, dans le cas où cette hypothèse n'est pas vérifiée. Le PGD pour $(1/n\sum_1^n Z_i\delta_{x_i^n})_{n\geq 1}$ est également présent dans le livre de Dembo et Zeitouni ([20], Section 7.2) où il est utilisé pour établir plusieurs autres résultats parmi lesquels le théorème de Mogul'skii. Le PGD y est établi quand on a existence de tous les moments exponentiels :

$$Ee^{\alpha|Z|} < +\infty \quad \text{pour tout } \alpha > 0. \tag{4.2}$$

Il est intéressant de noter que sous cette condition, il n'y a pas de terme additionnel faisant intervenir μ_s dans la fonction de taux :

$$I(\mu) = \begin{cases} \int \Lambda^*(d\mu/dR)\, dR & \text{si } \mu \ll R, \\ +\infty & \text{sinon.} \end{cases}$$

Dans le cadre du théorème de Sanov, cette relation entre les conditions de moments exponentiels et l'apparition de "parties singulières" a été étudiée au chapitre 1 (voir également Léonard et Najim [38]).

Le PGD pour les mesures empiriques MEM est ensuite utilisé pour obtenir des résultats du type de ceux obtenus par Mogul'skii (voir [40]). Plus précisément, on établit des PGD pour les fonctions aléatoires suivantes :

$$t \mapsto \bar{Z}_n(t) = \frac{1}{n}\sum_{i=1}^{[nt]} Z_i \quad \text{and} \quad t \mapsto \tilde{Z}_n(t) = \bar{Z}_n(t) + \left(t - \frac{[nt]}{n}\right) Z_{[nt]+1}.$$

En nous inspirant de Lynch et Sethuraman [39] et de De Acosta dans [17], les PGD sont établis dans bv, l'espace des fonctions à variation bornée muni de la topologie $*$-faible $\sigma(bv, C([0,1];\mathbb{R}^d))$. Sous la condition (4.2), les PGD sont bien connus (voir par exemple [20] pour un état de l'art récent). Sous la condition (4.1), le PGD a été établi par Mogul'skii [40] et la forme particulière de la

fonction de taux apparaît dans l'article de Lynch et Sethuraman [39]. Pour des travaux voisins concernant les processus à accroissements indépendants sous la condition (4.1), on pourra se référer à Mogul'skii [41], De Acosta [17], Léonard [36]. Les résultats que nous présentons dans ce chapitre sont différents de ceux établis dans [40] (voir les commentaires au paragraphe 4.2.3).
Le chapitre est organisé de la manière suivante : le PGD pour la mesure empirique MEM est établi dans la section 4.1. Les théorèmes concernant les processus empiriques sont établis dans la section 4.2.
Nous nous servirons dans ce chapitre des notations et des hypothèses introduites au chapitre 3, paragraphe 3.1.1.

4.1 Le PGD pour les mesures empiriques du type MEM

4.1.1 Le principe de grandes déviations

Nous énonçons le principe de grandes déviations pour les mesures empiriques dans le théorème 4.1. Introduisons d'abord quelques notations. On note $C_d^*(\mathcal{X})$ le dual topologique de $C_d(\mathcal{X})$. On le munit de la topologie $*$-faible $\mathcal{E} = \sigma(C_d^*, C_d)$ et de la tribu borélienne associée $\mathcal{B}(C_d^*)$. Si $\xi \in C_d^*(\mathcal{X})$ et $f \in C_d(\mathcal{X})$, on notera par $\langle \xi, f \rangle$ le crochet de dualité.

Hypothèse H-5 *$\mathcal{X}$ est un espace compact.*

Dans le cas où $\mathcal{X}$ est compact, les formes linéaires continues sur $C_d(\mathcal{X})$ sont des mesures vectorielles. Par suite, on note par $M^d(\mathcal{X})$ l'ensemble des mesures vectorielles à valeurs $\mathbb{R}^d$, c'est à dire $\mu \in M^d(\mathcal{X})$ ssi $\mu = (\mu_1, \dots, \mu_d)$ où chaque $\mu_i \in M(\mathcal{X})$. Soit $f \in C_d(\mathcal{X})$ et soit $\mu \in M^d$. On note par

$$\int_{\mathcal{X}} f(x) \cdot \mu(\, dx) \triangleq \sum_{l=1}^{m} \int_{\mathcal{X}} f_l(x) \mu_l(\, dx). \tag{4.3}$$

Si $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m \times d}$ est une fonction (matricielle) continue bornée et si $f_j : \mathcal{X} \to \mathbb{R}^d$ représente la j$^{\text{e}}$ ligne de la matrice $\mathbf{f}$, alors on note

$$\int_{\mathcal{X}} \mathbf{f}(x) \cdot \mu(dx) \triangleq \begin{pmatrix} \int_{\mathcal{X}} f_1(x) \cdot \mu(dx) \\ \vdots \\ \int_{\mathcal{X}} f_m(x) \cdot \mu(dx) \end{pmatrix}. \tag{4.4}$$

On munit $M^d(\mathcal{X})$ de la topologie $*$-faible $\Sigma = \sigma(M^d, C_d)$, c'est à dire la topologie la moins fine rendant continues toutes les fonctionnelles $\Gamma_f : \mu \mapsto \int f \cdot d\mu$

et de la tribu borélienne associée $\mathcal{B}(M^d)$. Soit $\mu \in M^d(\mathcal{X})$, on note μ_a sa partie absolument continue par rapport à R et μ_s, sa partie singulière.

Théorème 4.1 *Supposons que (H-2), (H-3) et (H-4) sont vérifiées, alors la famille*

$$L_n = \frac{1}{n} \sum_1^n Z_i \delta_{x_i^n}$$

satisfait le principe de grandes déviations dans $(C_d^*(\mathcal{X}), \mathcal{E}, \mathcal{B}(C_d^*))$ *avec la bonne fonction de taux*

$$I(\xi) = \sup_{f \in C_d(\mathcal{X})} \{\langle \xi, f \rangle - \int_{\mathcal{X}} \Lambda[f(x)]\, R(dx)\}.$$

Supposons de plus que (H-5) est également vérifiée, alors le PGD a lieu dans $(M^d(\mathcal{X}), \Sigma, \mathcal{B}(M^d))$, *avec pour bonne fonction de taux*

$$\begin{aligned} I(\mu) &= \sup_{f \in C_d(\mathcal{X})} \{\int_{\mathcal{X}} f(x) \cdot \mu(dx) - \int_{\mathcal{X}} \Lambda[f(x)]\, R(dx)\} \\ &= \int_{\mathcal{X}} \Lambda^* \Big[\frac{d\mu_a}{dR}(x)\Big]\, R(dx) + \int_{\mathcal{X}} \rho \Big[\frac{d\mu_s}{d\theta}(x)\Big]\,, \theta(dx), \end{aligned} \tag{4.5}$$

où $\rho(z) = \sup\{\lambda \cdot z,\ \lambda \in \mathcal{D}_\Lambda\}$ *est la fonction de récession de* Λ^* *et* θ *est n'importe quelle mesure par rapport à laquelle* μ_s *est absolument continue.*

Remarque 4.2 (Compacité de $\mathcal{X}$) Sans l'hypothèse de compacité, les formes linéaires continues sur $C_d(\mathcal{X})$ ne sont plus des mesures. Ce sont des fonctions d'ensembles additives bornées et régulières (voir [24]). On notera de plus que l'hypothèse (H-5) est centrale pour identifier la fonction de taux (voir (4.5)) via une identité due à Rockafellar dans [48].

Remarque 4.3 Sous l'hypothèse (H-5) et dans le cas où $Ee^{\alpha|Z|} < \infty$ pour tout $\alpha > 0$, la fonction de récession est $\rho(z) = \sup\{\lambda \cdot z,\ \lambda \in \mathbb{R}^d\} = +\infty$ partout sauf en zéro où $\rho(0) = 0$. Par suite, $I(\mu)$ n'est finie que si μ est absolument continue par rapport à R et la fonction de taux devient :

$$I(\mu) = \int_{\mathcal{X}} \Lambda^* \Big[\frac{d\mu}{dR}(x)\Big]\, R(dx) \text{ si } \mu \ll R,\ +\infty \text{ sinon.}$$

Cela est cohérent avec le théorème 7.2.3 dans [20] où la partie singulière μ_s n'intervient pas.

Preuve. • *Le PGD.* Soit $C'_d(\mathcal{X})$ (resp. $C'(\mathcal{X})$) le dual algébrique de $C_d(\mathcal{X})$ (resp. $C(\mathcal{X})$). Le produit de dualité est noté $\langle\ ,\ \rangle$. Considérons l'application

$$p_{f_1,\dots,f_m} : C'_d(\mathcal{X}) \to \mathbb{R}^m; \quad \xi \mapsto \begin{pmatrix} \langle \xi, f_1\rangle \\ \vdots \\ \langle \xi, f_m\rangle \end{pmatrix},$$

où $f_i \in C_d(\mathcal{X})$. Alors $p_{f_1,\dots,f_m}(L_n) = \langle L_n, \mathbf{f}\rangle$ satisfait un PGD grâce au théorème 3.2. Le théorème de Dawson-Gärtner nous assure alors que L_n satisfait un PGD dans $C'_d(\mathcal{X})$ munit de la topologie $*$-faible de bonne fonction de taux

$$\begin{aligned} I(\xi) &= \sup_{m\geq 1} \sup_{f_1,\dots,f_m\in C_d(\mathcal{X})} \sup_{\theta\in\mathbb{R}^m} \{\sum_1^m \theta_i\langle\xi, f_i\rangle - \int_{\mathcal{X}} \Lambda[\sum_1^m \theta_i f_i(x)]\, R(dx)\} \\ &= \sup_{f\in C_d(\mathcal{X})} \{\langle \xi, f\rangle - \int_{\mathcal{X}} \Lambda[f(x)]\, R(dx)\}; \qquad \xi \in C'_d(\mathcal{X}). \end{aligned}$$

• *Restriction du PGD.* On montre ici que $I(\xi) < \infty$ implique que $\xi \in C^*_d(\mathcal{X})$. Supposons que $I(\xi) < \infty$ alors, pour tout $f \in C_d(\mathcal{X})$ vérifiant $f \neq 0$

$$\langle \xi, \frac{f}{a\|f\|}\rangle \leq I(\xi) + \int \Lambda\Big(\frac{f}{a\|f\|}\Big)\, dR \leq I(\xi) + \Lambda_{|Z|}\Big(\frac{1}{a}\Big), \tag{4.6}$$

où $\Lambda_{|Z|}$ représente la log-Laplace de $|Z|$. Pour a assez grand, $\Lambda_{|Z|}\Big(\frac{1}{a}\Big)$ est fini du fait de l'hypothèse (H-4) et $\langle \xi, f\rangle \leq K\|f\|$. Considérant $-f$, on obtient $|\langle\xi, f\rangle| \leq K\|f\|$. Par suite, ξ est une forme linéaire continue et le principe de grandes déviations a lieu dans le dual topologique grâce au lemme 4.1.5 de [20].
Si de plus l'hypothèse (H-5) a lieu, alors le théorème de représentation de Riesz nous assure que ξ se représente comme une mesure à valeurs $\mathbb{R}^d$ sur $\mathcal{X}$, *i.e.* $\xi \in M^d(\mathcal{X})$. On utilisera alors la notation plus familière μ. On peut dès lors à nouveau appliquer le lemme 4.1.5 de [20] pour obtenir le PGD dans $(M^d(\mathcal{X}), \Sigma, \mathcal{B}(M^d))$.
• *Représentation de la fonction de taux.* Du fait des hypothèses (H-2) et (H-5), on peut appliquer le théorème 5 de [48]. On obtient ainsi

$$I(\mu) = \int_{\mathcal{X}} \Lambda^*\Big[\frac{d\mu_a}{dR}(x)\Big]\, R(dx) + \int_{\mathcal{X}} \rho\Big[\frac{d\mu_s}{d\theta}\Big]\, d\theta,$$

où ρ est la fonction de récession (voir la définition dans [46]) de Λ^* et θ est n'importe quelle mesure réelle et positive par rapport à laquelle μ_s est absolument continue. Comme Λ est la conjuguée convexe de Λ^*, ρ est la fonction de support de Λ ([46], théorème 13.3), soit

$$\rho(z) = \sup\{\lambda \cdot z,\ \lambda \in \mathcal{D}_\Lambda\}.$$

Par suite, le théorème 4.1 est démontré. □

4.1.2 Effets de bord dans le cas non compact : un fait et une heuristique

Supposons que R soit une probabilité sur $\mathbb{R}^+$ satisfaisant (H-2), qu'il existe une suite (x_i^n) satisfaisant (H-3) et que (Z_i) soient des variables aléatoires i.i.d., distribuées selon $\mathbb{P}_Z$ où

$$\mathbb{P}_Z(\,dz) = 1_{[0,\infty)}(z)e^{-z}\,dz.$$

Dans ce cas, (H-4) est vérifiée et L_n satisfait un PGD dans $(C_d^*(\mathbb{R}^+), \mathcal{E}, \mathcal{B}(C_d^*))$ du fait du théorème 4.1. Comme d'habitude, $\Lambda(\lambda)$ est la log-Laplace.

Le fait considérons $\mu_n = a\,R + \delta_n$ alors $I(\mu_n) \leq 1 + \Lambda^*(a)$. En effet,

$$\begin{aligned} I(\mu_n) &= \sup_{f\in C(\mathbb{R}^+)} \left\{ \int a\,f\,dR + f(n) - \int \Lambda(f)\,dR \right\} \\ &= \sup_{f\in C(\mathbb{R}^+),\ f\leq 1} \left\{ \int a\,f\,dR + f(n) - \int \Lambda(f)\,dR \right\}, \end{aligned}$$

où la dernière égalité provient du fait que $\int \Lambda(f)\,dR = +\infty$ si $f > 1$ (on notera que $\Lambda(\lambda) = +\infty$ si $\lambda > 1$). Finalement, si $f \leq 1$ alors :

$$\begin{aligned} \int a\,f\,dR + f(n) - \int \Lambda(f)\,dR &\leq \int a\,f\,dR - \Lambda(\int f\,dR) + 1 \\ &\leq \Lambda^*(a) + 1. \end{aligned}$$

De ce fait, $I(\mu_n) \leq \Lambda^*(a) + 1$. Par suite, $\mu_n \in \{I \leq \Lambda^*(a) + 1\}$ qui est un ensemble compact. Par conséquent, μ_n admet des points d'accumulation (en tant que limites de sous-filtres du fait que $C_d^*(\mathbb{R}^+)$ n'est pas métrisable). De manière immédiate, ces limites ne sont pas des mesures.

L'heuristique Une interprétation de la remarque précédente est la suivante : dans un régime de grandes déviations, L_n peut se comporter asymptotiquement comme $a\,R + \delta_n$. Cela légitime l'apparence de formes linéaires qui ne sont plus des mesures dans le cas où $\mathcal{X}$ n'est pas compact. Pour voir cela, notons d'abord que la "particule" $\frac{Z}{n}$ satisfait un PGD de bonne fonction de taux :

$$\delta^*(z) = \begin{cases} z & \text{si } z \geq 0 \\ +\infty & \text{sinon} \end{cases} .$$

Par suite, la contribution d'une particule importe dans le phénomène de grandes déviations. Comme $\lim_n 1/n \sum \delta_{x_i^n} = R$ qui est une mesure dont le support est $\mathbb{R}^+$, il existe une sous suite de (x_i^n), disons x_n, qui satisfait $\lim_n x_n = +\infty$. Considérons maintenant

$$L_n = \frac{1}{n} \sum_{i=1}^{n-1} Z_i \ \delta_{x_i^n} + \frac{Z_n}{n} \delta_{x_n}.$$

Un comportement de grandes déviations peut avoir lieu, où $\frac{Z_n}{n}$ est grand, le reste de la mesure empirique ayant un "comportement non déviant". Cette heuristique nous donne $L_n \approx m\, R + \delta_{x_n}$ où $m = EZ$.

4.1.3 Exemple : quelques caractéristiques de la fonction de taux

Nous considérons dans cet exemple le cadre suivant : $\mathcal{X} = [0,1]$, $R(\,dx) = \ell(dx)$ où $\ell(dx)$ est la mesure de Lebesgue measure sur $[0,1]$, $(x_i^n) = (i/n)$ et $\mathbb{P}_Z(\,dz) = c.1_{[0,\infty)}(z)\frac{e^{-z}}{1+z^3}\,dz$. Ainsi, les (Z_i) sont des variables aléatoires positives à valeurs réelles On note $z^* = \Lambda'(1)$ et $m = EZ$. Un calcul simple entraine que Λ n'est pas escarpée et que Λ^* est linéaire pour $z > z^*$:

$$\Lambda^*(z) = z - z^* + \Lambda^*(z^*) \quad \text{pour} \quad z > z^*. \tag{4.7}$$

Dans ce contexte, la fonction de récession est $\rho(z) = z$ et on peut démontrer facilement que la fonction de taux du PGD pour les mesures empiriques MEM $I(\mu)$ est finie seulement si μ est une mesure positive. I a donc la forme suivante :

$$I(\mu) = \int_{[0,1]} \Lambda^* \left[\frac{d\mu_a}{d\ell}(x)\right] \ell(dx) + \mu_s[0,1].$$

Valeur de la fonction de taux I pour une mesure particulière On note κ la fonction de Cantor sur $[0,1]$. On rappelle que κ est continue, croissante de $[0,1]$ dans $[0,1]$ et que la dérivée de κ est ℓ-presque sûrement nulle. En particulier, κ est la fonction de répartition d'une probabilité μ^κ singulière par rapport à ℓ et

$$I(\ell + \mu^\kappa) = \Lambda^*(1) + \mu^\kappa[0,1] = \Lambda^*(1) + 1.$$

Existence de plusieurs minimisants sous contrainte convexe Considérons la contrainte convexe $\mathcal{C} = \{\mu \in M([0,1]),\ \langle \mu, 1\rangle = E\}$. On note $\mathcal{M}$ l'ensemble des minimisants de la fonction de taux I sous $\mathcal{C}$. Alors :

Proposition 4.4 *Soit E un nombre réel :*

1. *si $E \in [m, z^*)$, alors il existe un unique minimisant de I sous la contrainte $\mathcal{C}$. Ce minimisant μ est défini par $\frac{d\mu}{d\ell}(x) = E$ et $I(\mu) = \Lambda^*(E)$*
2. *si $E \geq z^*$, alors tout minimisant est une mesure positive satisfaisant $\mu = \mu_a + \mu_s$ où $g(x) = \frac{d\mu_a}{d\ell}(x)$ vérifie $g(x) \geq z^*\ \ell - p.s.$ et $\langle \mu, 1\rangle = E$, et réciproquement. De plus, $I(\mu) = E - z^* + \Lambda^*(z^*)$*

La preuve de la proposition 4.4 est établie dans la section 4.3.

Remarque 4.5 Au vu du second point de la proposition précédente, il apparaît que toutes les mesures suivantes appartiennent à $\mathcal{M}$ pour $E > z^*$:

$$\alpha\ \ell(dx) + \beta\ \delta_u(dx) + \gamma\ \mu^\kappa(dx),$$

avec $\alpha + \beta + \gamma = E$, $\alpha \geq z^*$ et $u \in [0,1]$. On remarque en particulier qu'il y a une infinité de minimisants.

Remarque 4.6 En particulier, la fonction I n'est pas strictement convexe sur son domaine effectif $\mathcal{D}_I = \{\mu,\ I(\mu) < \infty\}$. Un fait similaire [38] a été mis en évidence dans le contexte du théorème de Sanov.

4.2 Grandes déviations pour certains processus

Dans cette section, nous établissons des principes de grandes déviations fonctionels du type de ceux établis par Mogul'skii dans [40] pour les fonctions aléatoires suivantes :

$$t \mapsto \bar{Z}_n(t) = \frac{1}{n}\sum_{i=1}^{n} Z_i \quad \text{et} \quad t \mapsto \tilde{Z}_n(t) = \bar{Z}_n(t) + \left(t - \frac{[nt]}{n}\right) Z_{[nt]+1}.$$

Ces résultats sont essentiellement des corollaires du PGD pour les mesures empiriques MEM précédemment établi, dans le cas où $\mathcal{X} = [0,1]$, $R = dx$ est la mesure de Lebesgue sur $[0,1]$ et $(x_i^n) = \left(\frac{i}{n}\right)$.

Nous utilisons le cadre proposé par Lynch et Sethuraman [39] et De Acosta ([17], section 5). Pour cela, on introduit les notations suivantes : $bv([0,1], \mathbb{R}^d)$

(raccourci en bv) est l'espace des fonctions à variation bornée sur $[0,1]$. On identifie bv avec $M^d([0,1])$ de la manière habituelle : à chaque $f \in bv$ correspond μ^f caractérisée par $\mu^f([0,t]) = f(t)$. $C_d([0,1])$ est donc, via cette identification, le dual topologique de bv et on munit bv de la topologie $*$-faible $\sigma(bv, C_d([0,1]))$ (raccourci en σ_w) et de la tribu borélienne associée $\mathcal{B}_w$. Soit $f \in bv$ et μ^f sa mesure associée dans $M^d([0,1])$. Considérons la décomposition de Lebesgue de μ^f, $\mu^f = \mu_a^f + \mu_s^f$ où μ_a^f représente la partie absolument continue de μ^f par rapport à dx et μ_s^f sa partie singulière. On note $f_a(t) = \mu_a^f([0,t])$ et $f_s(t) = \mu_s^f([0,t])$.

4.2.1 Le PGD pour la fonction étagée $\bar{Z}_n(.)$

Théorème 4.7 *Les fonctions aléatoires* $\left(\bar{Z}_n(t)\right)_{t\in[0,1]}$ *satisfont un PGD dans* $(bv, \sigma_w, \mathcal{B}_w)$ *de bonne fonction de taux*

$$\Phi(f) = \int_{[0,1]} \Lambda^*(f_a'(t))\, dt + \int_{[0,1]} \rho(f_s'(t))\, d\theta(t),$$

où θ *est n'importe quelle mesure par rapport à laquelle* μ_s *est absolument continue et* $f_s' = df_s/d\theta$*.*

Preuve. considérons $\Pi : M^d \to bv$ où

$$\Pi(\mu) = \begin{pmatrix} \mu_1[0,t] \\ \vdots \\ \mu_d[0,t] \end{pmatrix}_{t\in[0,1]}.$$

Alors Π est une bijection continue et $\Pi(L_n) = \bar{Z}_n$. Comme L_n satisfait un PGD de bonne fonction de taux (4.5) du fait du théorème 4.1, le principe de contraction entraine le PGD. □

4.2.2 Le PGD pour la ligne polygonale $\tilde{Z}_n(.)$

Théorème 4.8 *Les fonctions aléatoires* $\left(\tilde{Z}_n(t)\right)_{t\in[0,1]}$ *satisfont un PGD dans* $(bv, \sigma_w, \mathcal{B}_w)$ *de bonne fonction de taux :*

$$\Phi(f) = \int_{[0,1]} \Lambda^*(f_a'(t))\, dt + \int_{[0,1]} \rho(f_s'(t))\, d\theta(t),$$

où θ *est n'importe quelle mesure par rapport à laquelle* μ_s *est absolument continue et* $f_s' = df_s/d\theta$*.*

Preuve. Soit $f \in C_d([0,1])$. Considérons la mesure empirique $\tilde{L}_n \in M_d$ caractérisée par

$$\langle \tilde{L}_n, f\rangle = \sum_{i=0}^{n-1} \int_{i/n}^{(i+1)/n} f(t)\,dt \cdot Z_i,$$

où l'intégrale $\int_{(i-1)/n}^{i/n} f(t)\,dt$ est à valeurs dans $\mathbb{R}^d$ et $\int_{i/n}^{(i+1)/n} f(t)\,dt{\cdot}Z_i$ représente le produit scalaire habituel. Alors $\Pi(\tilde{L}_n) = \tilde{Z}_n$ (où Π est définie comme dans la preuve précédente) et le théorème est démontré dès lors qu'on prouve le PGD pour $\tilde{L}_n$ avec la bonne fonction de taux donnée par (4.5). Comme précédemment, si $\mathbf{f} : [0,1] \to \mathbb{R}^{m\times d}$ est une fonction (matricielle) continue bornée et si f_j représente la j$^{\text{e}}$ ligne de la matrice, alors

$$\langle \tilde{L}_n, \mathbf{f}\rangle = \sum_{0}^{n-1} \begin{pmatrix} \int_{i/n}^{(i+1)/n} f_1(t)\,dt \cdot Z_i \\ \vdots \\ \int_{i/n}^{(i+1)/n} f_m(t)\,dt \cdot Z_i \end{pmatrix} \quad \left(\triangleq \sum_{1}^{n} \int_{(i-1)/n}^{i/n} \mathbf{f}(t)\,dt \cdot Z_i \right).$$

Montrons que $\langle \tilde{L}_n, \mathbf{f}\rangle$ et $\langle L_n, \mathbf{f}\rangle$ sont exponentiellement équivalents :

$$\begin{aligned} |\langle \tilde{L}_n, \mathbf{f}\rangle - \langle L_n, \mathbf{f}\rangle| &\le \sum_{i=1}^{n} |Z_i| \left| \int_{(i-1)/n}^{i/n} \mathbf{f}(t)\,dt - \frac{\mathbf{f}(i/n)}{n} \right| \\ &\le \sum_{i=1}^{n} |Z_i| \int_{(i-1)/n}^{i/n} |\mathbf{f}(t) - \mathbf{f}(i/n)|\,dt. \end{aligned}$$

Soit $\epsilon > 0$ donné. Comme $\mathbf{f}$ est uniformément continue sur $[0,1]$, $|\mathbf{f}(t) - \mathbf{f}(i/n)| \le \epsilon$ pour $t \in [(i-1)/n, i/n]$ et pour n assez grand. Par conséquent,

$$|\langle \tilde{L}_n, \mathbf{f}\rangle - \langle L_n, \mathbf{f}\rangle_n| \le \frac{1}{n} \sum_{1}^{n} |Z_i|\ \epsilon,$$

pour n assez grand. Par suite, pour tout $\epsilon > 0$,

$$\begin{aligned} &\limsup_n \frac{1}{n} \log P\{|\langle \tilde{L}_n, \mathbf{f}\rangle - \langle L_n, \mathbf{f}\rangle| > \delta\} \\ &\qquad \le \limsup_n \frac{1}{n} \log P\{1/n \sum_{1}^{n} |Z_i|\ \epsilon > \delta\} \le -\Lambda_{|Z|}^{*}\left(\frac{\delta}{\epsilon}\right). \end{aligned}$$

Comme, du fait de l'hypothèse (H-5), $\lim_{\epsilon\to 0} \Lambda_{|Z|}\left(\frac{\delta}{\epsilon}\right) = +\infty$, $\langle \tilde{L}_n, \mathbf{f}\rangle$ et $\langle L_n, \mathbf{f}\rangle$ sont exponentiellement équivalents. Ainsi, $\langle \tilde{L}_n, \mathbf{f}\rangle$ satisfait un PGD dans $\mathbb{R}^m$ de bonne fonction de taux donnée par (3.4). On établit enfin le PGD pour $\tilde{L}_n$ de manière similaire à celui établi au théorème 4.1. □

4.2.3 Commentaires bibliographiques

Mogul'skii [40], Djellout, Guillin et Wu [21] et Marguerite Zani [59] se sont -entre autres- intéressés aux grandes déviations des marches aléatoires dans un contexte où il n'existe que quelques moments exponentiels. Djellout *et al.* ainsi que Zani ont travaillé sur l'ensemble des fonctions à variation bornée dans un cadre unidimensionnel. Ainsi, Djellout *et al.* étudient les grandes déviations de l'estimateur de la variation quadratique

$$\Theta_n(t) = \sum_{i=0}^{[nt]} \left(X_{i/n} - X_{(i-1)/n} \right)^2$$

associé au processus de diffusion $dX_t = \sigma_t \, dB_t + b(t, X_t) \, dt$. M. Zani s'intéresse à la marche aléatoire

$$Z_n(t) = \frac{1}{n} \sum_{i=1}^{[nt]} X_i^2$$

associée au processus gaussien stationnaire (X_n). Dans les deux cas, la fonction de taux fait apparaître une partie singulière. Notons que les techniques d'établissement des PGD sont spécifiques à chaque modèle.

Mogul'skii s'est intéressé à la ligne brisée multidimensionnelle $\tilde{Z}_n$ telle que nous l'étudions. La topologie qu'il considère est relative à une métrique de Skorokhod pour laquelle l'espace des fonctions continues à valeurs dans $\mathbb{R}^d$ dans lequel vit la ligne brisée n'est cependant pas complet. Et des impératifs de compacité (pour établir la borne supérieure du PGD) l'obligent à considérer le complété de cet espace sous cette même métrique de Skorokhod. Les points-limite d'un tel complété ne s'expriment pas explicitement sous la forme fonction. Par ailleurs, la fonction de taux I est donnée sous forme abstraite. En effet, I est la régularisée semi-continue inférieure de la fonctionnelle habituelle

$$\tilde{I}(x) = \int_{[0,1]} \Lambda^*(\dot{x}(t)) \, dt.$$

Pour conclure, Mogul'skii établit un principe de grandes déviations pour une topologie plus fine que la topologie que nous considérons. Cela a un prix : l'espace d'états et la fonction de taux demeurent abstraits.

4.3 Absence de stricte convexité pour I -fin de la preuve-

Preuve. On peut vérifier facilement que la mesure définie par $d\gamma = E\ d\ell$ est toujours un minimisant de I sous $\mathcal{C}$ (Jensen). Par suite, si μ est un minimisant :

$$I(\mu) = \Lambda^*(E) = \begin{cases} \Lambda^*(E) & \text{si } E < z^*, \\ \Lambda^*(z^*) + E - z^* & \text{sinon.} \end{cases}$$

Dans le cas où $E < z^*$, prouvons que ce minimisant est unique. Supposons pour cela que μ et ν sont deux minimisants distincts, soit :

$$\begin{cases} \langle \mu, 1 \rangle = E \\ \langle \nu, 1 \rangle = E \end{cases} \quad \text{et} \quad I(\mu) = I(\nu) = \Lambda^*(E).$$

Notons que toute combinaison convexe $\alpha\mu + \beta\nu$ appartient à $\mathcal{C}$.

$$I(\alpha\mu + \beta\nu) = \int \Lambda^* \left(\alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} \right) d\ell + \alpha\mu_s[0;1] + \beta\nu_s[0;1].$$

Si $\ell \left\{ \alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} < z^* \right\} = 0$ alors

$$I(\alpha\mu + \beta\nu) \geq \int_{\left\{ \alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} \geq z^* \right\}} \Lambda^* \left(\alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} \right) d\ell \geq \Lambda^*(z^*) > \Lambda^*(E),$$

ce qui est impossible car $I(\alpha\mu + \beta\nu) \leq \alpha I(\mu) + \beta I(\nu) = \Lambda^*(E)$. Par suite supposons que $\ell \left\{ \alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} < z^* \right\} > 0$. Dans ce cas,

$$\int \Lambda^* \left(\alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} \right) d\ell$$
$$= \int_{\left\{ \alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} < z^* \right\}} \Lambda^* \left(\alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} \right) d\ell + \int_{\left\{ \alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} \geq z^* \right\}} \Lambda^* \left(\alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} \right) d\ell.$$

Comme Λ^* est strictement convexe sur l'intervalle $(-\infty; z^*]$,

$$\Lambda^* \left(\alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} \right) < \alpha\Lambda^* \left(\frac{d\mu_a}{d\ell} \right) + \beta\Lambda^* \left(\frac{d\nu_a}{d\ell} \right),$$

sur $\left\{ \alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} < z^* \right\}$. Cette inégalité demeure stricte en intégrant sur l'ensemble $\left\{ \alpha \frac{d\mu_a}{d\ell} + \beta \frac{d\nu_a}{d\ell} < z^* \right\}$. Par conséquent, $I(\alpha\mu + \beta\nu) < \alpha I(\mu) + \beta I(\nu) = \Lambda^*(E)$, ce qui est impossible. Nécessairement, le minimisant est unique.

Dans le cas où $E \geq z^*$, la seule chose à montrer est que $\ell\left\{\frac{d\mu_a}{d\ell} < z^*\right\} = 0$ si μ est un minimisant. Supposons que $\ell\left\{\frac{d\mu_a}{d\ell} < z^*\right\} > 0$. On notera $g(x) = \frac{d\mu_a}{dx}(x)$.

$$\int \Lambda^*(g)\, d\ell = \int_{\{g<z^*\}} \Lambda^*(g)\, d\ell + \int_{\{g\geq z^*\}} \Lambda^*(g)\, d\ell.$$

Un double application de l'inégalité de Jensen entraine que

$$\int \Lambda^*(g)\, d\ell \geq \ell\{g < z^*\}\Lambda^*\left(\frac{\int_{\{g<z^*\}} g\, d\ell}{\ell\{g<z^*\}}\right) + \ell\{g \geq z^*\}\Lambda^*\left(\frac{\int_{\{g\geq z^*\}} g\, d\ell}{\ell\{g\geq z^*\}}\right).$$

Comme $\frac{\int_{\{g<z^*\}} g\, dx}{\ell\{g<z^*\}} < z^*$ et Λ^* est strictement convexe sur l'intervalle $(-\infty, z^*]$, on obtient :

$$\ell\{g < z^*\}\Lambda^*\left(\frac{\int_{\{g<z^*\}} g\, d\ell}{\ell\{g<z^*\}}\right) + \ell\{g \geq z^*\}\Lambda^*\left(\frac{\int_{\{g\geq z^*\}} g\, d\ell}{\ell\{g\geq z^*\}}\right) > \Lambda^*\left(\int g\, d\ell\right)$$

et $I(\mu) > \Lambda^*(\langle\mu_a, 1\rangle) + \langle\mu_s, 1\rangle$. Notons $\alpha E = \langle\mu_a, 1\rangle$ où $\alpha \in (0; 1)$. Comme la pente maximale de la fonction convexe Λ^* est 1, l'inégalité suivante est vraie pour tout réel E :

$$\frac{\Lambda^*(E) - \Lambda^*(\alpha E)}{E - \alpha E} \leq 1.$$

Par suite, $I(\mu) > \Lambda^*(\alpha E) + E - \alpha E \geq \Lambda^*(E)$. Mais cela est impossible comme $I(\mu) = \Lambda^*(E)$ du fait que μ est un minimisant. Nécessairement, $\ell\left\{\frac{d\mu_a}{d\ell} < z^*\right\} = 0$ et la seconde partie de la proposition est démontrée. $\square$

Chapitre 5

Grandes déviations pour des variables indépendantes mais non identiquement distribuées

Sommaire

Nous poursuivons ici l'étude des grandes déviations de la moyenne empirique

$$\langle L_n, \mathbf{f} \rangle = \frac{1}{n} \sum_{1}^{n} \mathbf{f}(x_i^n) \cdot Z_i^n,$$

dans le cas où les variables aléatoires (Z_i^n) sont indépendantes mais ne sont plus, cette fois-ci, identiquement distribuées.
Soit $(Z_i^n)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes, à valeurs dans $\mathbb{R}^d$, et soit $(x_i^n;\ 1 \leq i \leq n;\ n \geq 1)$ une suite déterministe d'éléments à valeurs dans $\mathcal{X}$ et satisfaisant

$$\frac{1}{n} \sum_{1}^{n} \delta_{x_i^n} \xrightarrow[n \to \infty]{\text{conv. étroite}} R.$$

Dans ce chapitre, nous nous intéressons aux grandes déviations de la moyenne empirique :

$$\langle L_n, \mathbf{f} \rangle = \frac{1}{n} \sum_1^n \begin{pmatrix} f_1(x_i^n) \cdot Z_i^n \\ \vdots \\ f_m(x_i^n) \cdot Z_i^n \end{pmatrix} \quad \left(\triangleq \frac{1}{n} \sum_1^n \mathbf{f}(x_i^n) \cdot Z_i^n \right)..$$

où chaque f_k est une fonction continue bornée de $\mathcal{X}$ dans $\mathbb{R}^d$ et où $\cdot$ représente le produit scalaire sur $\mathbb{R}^d$, dans le cas où la distribution de chaque Z_i^n (que l'on notera $\mathcal{L}(Z_i^n)$) dépend de x_i^n.

Notre intérêt pour l'étude des grandes déviations de $\tilde{L}_n$ provient des travaux de Gamboa et Gassiat [29] ainsi que ceux de Csiszár, Gamboa et Gassiat [15]. Dans [29], il est démontré que l'estimateur bayésien $E[L_n|_{L_n \in \mathcal{C}}]$ (dans le cas où $L_n = \frac{1}{n} \sum Z_i \, \delta_{x_i}$, $Z_i \sim P$, Z_i i.i.d.) converge vers une mesure μ que l'on cherche à reconstruire. Du fait que $\mathcal{C}$ peut être un événement rare, l'estimateur précédent est difficile à calculer en pratique et les auteurs proposent l'alternative suivante :

$$\begin{array}{ll} \text{si} & H(Q^*|P^{\otimes n}) = inf_{Q \in \mathcal{C}} H(Q|P^{\otimes n}) \quad \text{alors } Q^* = \tilde{P}_1 \times \cdots \times \tilde{P}_n, \\ \text{où} & \dfrac{d\,\tilde{P}_i}{d\,P} = \dfrac{e^{\lambda^* \phi(x_i^n) z}}{\int e^{\lambda^* \phi(x_i^n) z} P(\,dz)} \end{array}$$

et la mesure empirique $\tilde{L}_n = \frac{1}{n} \sum \tilde{Z}_i^n \delta_{x_i^n}$ satisfait

$$E\tilde{L}_n \xrightarrow[n \to +\infty]{} \mu,$$

où $\tilde{Z}_i^n \sim \tilde{P}_i^n$. C'est l'estimateur du Maximum d'entropie en moyenne. L'étude des grandes déviations de $\langle \tilde{L}_n, \mathbf{f} \rangle$ apparaît ainsi comme le préalable à l'étude des grandes déviations de $\tilde{L}_n$.

Nous formalisons dans la suite la dépendance entre la loi de $\tilde{Z}_i^n$ et le site x_i^n, de manière à pouvoir considérer des variables qui n'ont pas tous leurs moments exponentiels et dont les log-Laplaces ne sont pas forcément escarpées. Considérons une famille $P(x,\, dz)$ de lois sur $\mathbb{R}^d$ indexées par $x \in \mathcal{X}$. L'idée est d'autoriser les lois $P(x,\, dz)$ et $P(x',\, dz)$ à légèrement différer quand les deux sites x et x' sont distincts mais néanmoins proches au sens de la topologie sur $\mathcal{X}$. Considérons l'ensemble des probabilités admettant quelques moments exponentiels :

$$\mathcal{P}_\tau(\mathbb{R}^d) = \{P \in M_1(\mathbb{R}^d),\ \exists\ \alpha > 0\ \int_{\mathbb{R}^d} e^{\alpha|z|}\, P(dz) < \infty\}$$

et introduisons sur cet ensemble une distance qu'on appelera distance d'Orlicz-Wasserstein :

$$d_{OW}(P,Q) = \inf_{\eta} \inf \left\{ a > 0; \int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{z-z'}{a}\right) \eta(dzdz') \leq 1 \right\}, \quad P, Q \in \mathcal{P}_\tau,$$

où η est une probabilité dont les marginales sont P et Q. La variation de la loi s'exprime alors naturellement en termes de continuité de l'application

$$\begin{array}{rcl} \mathcal{X} & \to & (\mathcal{P}_\tau, d_{OW}) \\ x & \mapsto & P(x,\cdot) \end{array}$$

et les variables aléatoires indépendantes Z_i^n sont distribuées selon la loi $P(x_i^n,\cdot)$.

La dépendance précédemment décrite entre $\mathcal{L}(Z_i^n)$ et x_i^n permet de considérer des variables aléatoires dont les log-Laplaces ne sont pas escarpées. De ce fait, le théorème de Gärtner-Ellis, qui est l'un des outils majeurs pour établir des principes de grandes déviations dans des cas indépendants mais non identiquement distribués, ne s'applique pas. En effet, la preuve de ce théorème repose sur un argument de changement de mesure exponentielle et, de fait, nécessite l'escarpement de la log-Laplace (voir l'introduction générale pour une discussion plus précise).

Nous établissons dans ce chapitre un principe de grandes déviations pour $\frac{1}{n}\sum \mathbf{f}(x_i^n)\cdot Z_i^n$ sans faire l'hypothèse d'escarpement. Très brièvement, la méthode employée consiste à utiliser à un argument de couplage pour se ramener au problème traité au chapitre 3. Malheureusement, la technique d'approximation exponentielle utilisée ici est nettement plus complexe qu'au chapitre 3 et nous ne sommes actuellement pas à même d'identifier complètement la fonction de taux. Le chapitre est organisé de la manière suivante : la définition de la distance d'Orlicz-Wasserstein, le jeu d'hypothèses précis sur la famille (Z_i^n) et quelques exemples de familles $(P(x,\cdot); x \in \mathcal{X})$ sont donnés dans la section 5.1. Le principe de grandes déviations est énoncé et démontré à la section 5.2. Enfin, le fait que d_{OW} est une distance est établi à la section 5.3.

Dans toute la suite du chapitre, on écrira indifféremment P_x ou $P(x,\cdot)$ et on se servira des notations introduites au chapitre 3, paragraphe 3.1.1.

5.1 Distribution des variables aléatoires

Dans cette section, on introduit une distance entre probabilités nous permettant de formaliser la "proximité exponentielle" requise pour pouvoir établir des principes de grandes déviations (paragraphe 5.1.1), on énonce le jeu d'hypothèses sur les lois $P(x,\cdot)$ et sur les variables Z_i^n au paragraphe 5.1.2 et enfin, on donne quelques exemples de telles familles (paragraphe 5.1.3).

5.1.1 La distance d'Orlicz-Wasserstein

Soit $\tau(z) = e^{|z|} - 1$, $z \in \mathbb{R}^d$. Considérons :

$$\begin{aligned}\mathcal{P}_\tau(\mathbb{R}^d) &= \{P \in M_1(\mathbb{R}^d),\ \exists\, a > 0\ \int_{\mathbb{R}^d} \tau\left(\frac{z}{a}\right) P(dz) < \infty\}\\ &= \{P \in M_1(\mathbb{R}^d),\ \exists\, \alpha > 0\ \int_{\mathbb{R}^d} e^{\alpha|z|}\, P(dz) < \infty\}.\end{aligned}$$

$\mathcal{P}_\tau$ représente l'ensemble des probabilités qui ont quelques moments exponentiels. On note par $M(P,Q)$ l'ensemble des probabilités appartenant à $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ de marginales respectives $P \in \mathcal{P}(\mathbb{R}^d)$ et $Q \in \mathcal{P}(\mathbb{R}^d)$. Considérons la métrique d'Orlicz-Wasserstein définie sur $\mathcal{P}_\tau(\mathbb{R}^d)$ par :

$$d_{OW}(P,Q) = \inf_{\eta \in M(P,Q)} \inf \left\{ a > 0; \int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{z - z'}{a}\right) \eta(dzdz') \le 1 \right\}.$$

d_{OW} apparaît comme une distance entre probabilités permettant de mesurer la "proximité exponentielle" de deux distributions.

Lemme 5.1 *d_{OW} est une distance sur $\mathcal{P}_\tau(\mathbb{R}^d)$.*

La preuve de ce lemme est très proche de la preuve habituelle où il est montré que que la distance de Wasserstein est bien une distance. L'aspect "jauge" ($\inf\{a > 0; \int_{\mathbb{R}^d\times\mathbb{R}^d} \tau(\frac{z-z'}{a})\, \eta(dzdz') \le 1\}$) de la distance d'Orlicz-Wasserstein peut néanmoins dérouter. De ce fait, une preuve complète est proposée à la section 5.3.

Dans la suite, on munit $\mathcal{P}_\tau(\mathbb{R}^d)$ de la topologie induite par d_{OW}.

5.1.2 Un jeu d'hypothèses sur l'échantillon (Z_i^n)

Nous rappelons dans cette section certaines des hypothèses utilisées aux chapitres 3 et 4. Nous introduisons également deux nouvelles hypothèses : les hypothèses (H-6) et (H-7).

Hypothèse H-2 *R est une probabilité sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ satisfaisant*

$$\textit{si } U \textit{ est un ouvert non vide, alors} \quad R(U) > 0.$$

Hypothèse H-3 *La famille $(x_i^n)_{1\le i\le n,\ 1\le n} \subset \mathcal{X}$ satisfait*

$$\frac{1}{n}\sum_1^n \delta_{x_i^n} \xrightarrow[n\to\infty]{\textit{conv. étroite}} R \quad \textit{où} \quad R \in \mathcal{P}(\mathcal{X}). \tag{5.1}$$

Hypothèse H-5 *$\mathcal{X}$ est un espace compact.*

Hypothèse H-6 *$(P(x,\cdot); x \in \mathcal{X})$ est une famille de probabilités incluse dans $\mathcal{P}_\tau(\mathbb{R}^d)$. On suppose que pour chaque borélien A de $\mathbb{R}^d$, l'application $x \mapsto P(x, A)$ est mesurable et que la fonction*

$$\begin{array}{rcl} \Gamma : \mathcal{X} & \to & \mathcal{P}_\tau(\mathbb{R}^d) \\ x & \mapsto & P(x,\cdot) \end{array}$$

est continue lorsque $\mathcal{P}_\tau(\mathbb{R}^d)$ est munie de la topologie induite par la distance d_{OW}.

Remarque 5.2 On remarquera que la conjonction des hypothèses (H-5) et (H-6) entraine l'existence d'un réel $\alpha^* > 0$ tel que

$$S(\alpha^*) = \sup_{x \in \mathcal{X}} \int_{\mathbb{R}^d} e^{\alpha^*|z|} P_x(dz) < \infty. \tag{5.2}$$

En effet, $x \mapsto d_{OW}(P_x, P_{x_0})$ est une application continue et donc est majorée sur le compact $\mathcal{X}$, soit $d_{OW}(P_x, P_{x_0}) \leq \delta^*$ pour tout x élément de $\mathcal{X}$. En particulier, il existe pour tout x une probabilité $\eta_x \in M(P_x, P_{x_0})$ telle que

$$\int e^{\frac{|z-z'|}{\delta^*}} \eta_x(dzdz') \leq 2.$$

Soient $\alpha + \beta = 1$. Choisissons δ assez grand pour que $\delta\alpha > \delta^*$ et $\int e^{\frac{|z'|}{\delta\beta}} P_{x_0}(dz') < \infty$ soient vérifiés. Alors

$$\begin{aligned} \int e^{\frac{|z|}{\delta}} P_x(dz) &= \int e^{\frac{|z-z'+z'|}{\delta}} \eta_x(dzdz') \\ &\leq \left[\int e^{\frac{|z-z'|}{\delta\alpha}} \eta_x(dzdz') \right]^\alpha \left[\int e^{\frac{|z'|}{\delta\beta}} P_{x_0}(dz') \right]^\beta \\ &\leq 2^\alpha \left[\int e^{\frac{|z'|}{\delta^*}} P_{x_0}(dz') \right]^\beta < \infty \end{aligned}$$

et (5.2) est établie en prenant $\alpha^* = (\delta)^{-1}$.

Hypothèse H-7 *Etant donnée une famille de probabilités $(P(x,\cdot); x \in \mathcal{X})$ et une famille $(x_i^n; 1 \leq i \leq n; 1 \leq n)$ à valeurs dans $\mathcal{X}$, on considère une suite de variables aléatoires (Z_i^n), indépendantes, à valeurs $\mathbb{R}^d$ et où chaque variable Z_i^n est distribuée selon $P_{x_i^n}$.*

L'hypothèse la plus contraignante à vérifier est (H-6). Une fois montré que la famille étudiée $(P(x,\cdot))$ satisfait cette hypothèse, on peut construire sans problèmes une famille de variables aléatoires indépendantes vérifiant (H-7). Pour cette raison, on étudie dans la section suivante quelques familles vérifiant l'hypothèse (H-6).

5.1.3 Exemples de familles satisfaisant l'hypothèse (H-6)

Nous nous intéressons ici à des familles de lois Gamma ainsi qu'à des familles de variables aléatoires du type $f(x_i, Z_i)$.

Lois Gamma On prend pour probabilité de référence une loi Gamma :

$$P(\,dz) = \frac{e^{-z}z^{\alpha-1}}{\Gamma(\alpha)}\,dz, \quad \alpha > 0.$$

On s'intéresse à la famille de probabilités perturbées

$$P_x(dz) = \frac{e^{xz}}{C_x}P(\,dz) \qquad \text{où} \quad C_x = \int_0^{+\infty} e^{xz}P(dz) = (1-x)^{\alpha}$$

Ici, le paramètre x satisfait $0 \leq x \leq x_0 < 1$. Alors la famille $(P_x; 0 \leq x \leq x_0)$ satisfait l'hypothèse (H-6). Montrons la continuité de $x \mapsto P_x$ pour la topologie induite par d_{OW}. Fixons $x \in [0, x_0]$ et $\epsilon > 0$. Il nous faut démontrer que pour x' proche de x, on a $d_{OW}(P_{x'}, P_x) \leq \epsilon$. Pour cela, on va construire un couple de variables aléatoires (Z, Z') de lois marginales P_x et $P_{x'}$. Nous montrerons ensuite que pour x' assez proche de x,

$$E\tau\left(\frac{Z - Z'}{\epsilon}\right) \leq 1,$$

ce qui entrainera $d_{OW}(P_{x'}, P_x) \leq \epsilon$.

Soit U une variable uniforme sur $[0,1]$ et F (resp. F_x et $F_{x'}$) la fonction de répartition associée à la loi de référence P (resp. P_x et $P_{x'}$). On pose $Z = F_x^{-1}(U)$ et $Z' = F_{x'}^{-1}(U)$. Alors,

$$F_x(z) = F(z(1-x)) \quad \text{et} \quad F_{x'}(z) = F(z(1-x')),$$
$$Z = F_x^{-1}(U) = \frac{1}{1-x}F^{-1}(U) \quad \text{et} \quad Z' = F_{x'}^{-1}(U) = \frac{1}{1-x'}F^{-1}(U).$$

Par conséquent

$$\begin{aligned} Z - Z' &= F_x^{-1}(U) - F_{x'}^{-1}(U) \\ &= \frac{F^{-1}(U)}{1-x} - \frac{F^{-1}(U)}{1-x'} \\ &= \frac{(x-x')F^{-1}(U)}{(1-x)(1-x')}. \end{aligned}$$

Comme

$$\lim_{x' \to x} \frac{x'-x}{\epsilon(x-1)(x'-1)} = 0,$$

il s'ensuit que

$$E\tau\left(\frac{Z-Z'}{\epsilon}\right) = \int_0^1 \left(\exp\left|\frac{(x'-x)F^{-1}(u)}{\epsilon(x-1)(x'-1)}\right| - 1\right) du \leq 1,$$

pour x' assez proche de x et la continuité de $x \mapsto P_x$ est établie.

Généralisation du modèle $f(x_i)Z_i$. Soit Z une variable aléatoire à valeurs dans $\mathbb{R}^d$ et de loi R . Introduisons l'espace d'Orlicz suivant :

$$L_\tau = \left\{ f : \mathbb{R}^d \to \mathbb{R}^d;\ \exists a > 0,\ \int_{\mathbb{R}^d} \tau\left(\frac{f(z)}{a}\right) P(dz) < \infty \right\}.$$

Munissons-le de la norme

$$\|f\|_\tau = \inf\left\{ a > 0;\ \int \tau\left(\frac{f}{a}\right) dP \leq 1 \right\}.$$

Alors $(L_\tau, \|\cdot\|_\tau)$ est un espace de Banach (on pourra se reporter à la section 1.1 du chapitre 1 pour plus d'informations sur les espaces d'Orlicz). On remarque de plus que :

$$f \in L_\tau \Leftrightarrow\ \exists \alpha > 0,\ Ee^{\alpha|f(Z)|} < \infty.$$

Soit $\mathcal{X}$ un espace métrique compact et considérons $C(\mathcal{X}; L_\tau)$ qui est l'ensemble des fonctions continues de $\mathcal{X}$ dans L_τ. Alors, la famille de probabilités $(P_x)_{x\in\mathcal{X}}$, P_x étant la loi de $\phi(x, Z)$, satisfait l'hypothèse (H-6). Pour démontrer que $d_{OW}(P_x, P_{x'}) \leq \epsilon$ pour x' assez proche de x, il suffit d'exhiber un couple de variables aléatoires (Y, Y') de lois marginales P_x et $P_{x'}$ telles que

$$E\tau\left(\frac{|Y-Y'|}{\epsilon}\right) \leq 1.$$

On choisit naturellement $Y = \phi(x, Z)$ et $Y' = \phi(x', Z)$. Par définition même de la norme d'Orlicz et de ϕ, on a, pour x' assez proche de x,

$$E\tau\left(\frac{|\phi(x,Z) - \phi(x',Z)|}{\epsilon}\right) \leq 1.$$

La seconde partie de l'hypothèse est ainsi établie. Voici un exemple non trivial d'une telle famille. Soient $g : \mathbb{R}^d \to \mathbb{R}$ et $f : \mathcal{X} \to \mathbb{R}$ ($\mathcal{X}$ compact) deux fonctions continues bornées et soit Z une variable aléatoire à valeurs dans $\mathbb{R}^d$ ayant quelques moments exponentiels :

$$\exists \alpha > 0, \quad E e^{\alpha |Z|} < \infty.$$

Alors la famille de probabilités associée à $Z_x = \frac{Z}{1+|f(x)g(Z)|}$ satisfait l'hypothèse (H-6).

5.2 Le principe de grandes déviations

Théorème 5.3 *Soit* $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ *une fonction continue bornée. Supposons que (H-2), (H-3), (H-5), (H-6) et (H-7) sont vérifiées. Alors, la famille*

$$\langle L_n, \mathbf{f}\rangle = \frac{1}{n}\sum_1^n \mathbf{f}(x_i^n)\cdot Z_i^n$$

satisfait un principe de grandes déviations dans $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ *gouverné par une (unique) fonction de taux* Υ.

Remarque 5.4 Considérons la log-Laplace

$$\Lambda(x,\lambda) = \log \int_{\mathbb{R}^d} e^{\lambda\cdot z} P(x, dz), \quad \lambda \in \mathbb{R}^d, x \in \mathcal{X}$$

(que l'on notera également Λ_x) et la fonction de taux du PGD Υ. Cette fonction nous est en fait donnée par la formule :

$$\Upsilon(y) = \sup_{\eta>0} \liminf_{\epsilon\to 0} \inf_{y'\in B(y,\eta)} I_{\mathbf{f}^\epsilon}(y'), \quad y \in \mathbb{R}^m,$$

où les $I_{\mathbf{f}^\epsilon}$ sont des fonctions de taux associées à une approximation exponentielle (voir les détails de la construction au lemme 5.6). Par analogie avec le résultat établi dans le cas i.i.d. au théorème 3.6, la tentation est grande d'affirmer que

$$\Upsilon(y) = I_{\mathbf{f}}(y),$$

où

$$I_{\mathbf{f}}(y) = \sup_{\theta \in \mathbb{R}^m} \big\{ \theta \cdot y - \int_{\mathcal{X}} \Lambda\big(x, \sum_{j=1}^{m} \theta_j f_j(x)\big)\, R(dx) \big\}, \quad y \in \mathbb{R}^m.$$

Ici f_j représente la j$^{\text{e}}$ ligne de la matrice $\mathbf{f}$. Néanmoins, l'approximation exponentielle est plus complexe dans ce cas là et nous n'arrivons pas à ce jour à établir l'identité $\Upsilon(y) = I_{\mathbf{f}}(y)$. L'inégalité $\Upsilon \le I_{\mathbf{f}}$ est démontrée à la proposition 5.7.

5.2.1 Preuve du PGD

Le schéma de la preuve est essentiellement le même qu'au paragraphe 3.2.1. Le principal changement réside dans la procédure d'approximation. Comme dans la section 3.2.1, on approxime $\mathbf{f}$ par une fonction étagée uniformément proche mais comme les variables aléatoires Z_i^n ne sont plus identiquement distribuées, on utilise un couplage fortement corrélé via la distance d'Orlicz-Wasserstein de manière à travailler avec un nombre fini de lois. Comme $(P_x; x \in \mathcal{X})$ est compact, on peut le recouvrir par un nombre fini de boules $B_{OW}(P_k, \epsilon)$. Si Z est tiré selon la loi $P_{x_i^n} \in B_{OW}(P_k, \epsilon)$ alors, on l'approxime par une variable aléatoire fortement dépendante $\tilde{Z}$ tirée selon P_k. Par suite, on approxime $\langle L_n, \mathbf{f} \rangle$ par

$$\frac{a_1}{n} \sum_{i=1}^{N_{A_1}(n)} \tilde{Z}_i^{(1)} + \cdots + \frac{a_p}{n} \sum_{i=1}^{N_{A_p}(n)} \tilde{Z}_i^{(p)},$$

où chaque famille $(\tilde{Z}_i^{(k)})_{1 \le i \le N_{A_k}(n)}$ est une famille de variables aléatoires i.i.d. distribuées selon P_k.

Lemme 5.5 *Supposons que (H-3) soit vérifiée et soit $(A_k; 1 \le k \le p)$ une partition d'ensembles mesurables de $\mathcal{X}$ telle que $R(A_k) > 0$ et $R(\partial A_k) = 0$. Soit $(Z_i^n; 1 \le i \le n, n \ge 1)$ une famille de variables aléatoires indépendantes, à valeurs dans $\mathbb{R}^d$, où Z_i^n a pour distribution $P_k \in \mathcal{P}_\tau(\mathbb{R}^d)$ si $x_i^n \in A_k$. Considérons la fonction (matricielle) étagée*

$$\begin{aligned} \mathbf{f} : \mathcal{X} &\longrightarrow \mathbb{R}^{m \times d} \\ x &\longmapsto \mathbf{f}(x) = \sum_{k=1}^{p} \mathbf{a}_k 1_{A_k}(x), \end{aligned}$$

les $\mathbf{a}_k$ *étant des matrices* $m \times d$. *Alors,*

$$\begin{aligned}\langle L_n, \mathbf{f}\rangle &= \frac{1}{n}\sum_1^n \mathbf{f}(x_i^n)\cdot Z_i^n \\ &= \frac{1}{n}\sum_1^n \mathbf{a}_1 \cdot Z_i^n\, 1_{A_1}(x_i^n) + \cdots + \frac{1}{n}\sum_1^n \mathbf{a}_p \cdot Z_i^n\, 1_{A_p}(x_i^n)\end{aligned}$$

satisfait un PGD dans $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ *de bonne fonction de taux*

$$\begin{aligned}I_{\mathbf{f}}(y) &= \inf\left\{\sum_{k=1}^p R(A_k)\Lambda_k^*(u_k);\quad \sum_{k=1}^p \mathbf{a}_k \cdot u_k\, R(A_k) = y;\ u_k \in \mathbb{R}^d, 1 \le k \le p\right\}, \quad y \in \mathbb{R}^m \\ &= \inf\left\{\int_{\mathcal{X}} \Lambda_p^*(x, u(x))\, dR;\ \int_{\mathcal{X}} \mathbf{f}\cdot u\, dR = y;\ u \in B(\mathcal{X}; \mathbb{R}^d)\right\}.\end{aligned} \tag{5.3}$$

Ici, $\Lambda_k^*(z)$ *est la log-Laplace de* P_k *et* $\Lambda_p^*(x,z) = \sum_{k=1}^p \Lambda_k^*(z) 1_{A_k}(x)$.

Preuve. se démontre de la même manière que le lemme 3.7. □

Lemme 5.6 *Supposons que (H-2), (H-3), (H-5) et (H-6) sont vérifiées et soit* $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ *une fonction continue bornée. Alors il existe :*
- *une famille de variables aléatoires* $(Z_i^n)_{i,n}$ *satisfaisant (H-7),*
- *une famille de variables aléatoires indépendantes* $(\tilde{Z}_i^{n,\epsilon})_{i,n}$ *distribuées selon un nombre fini de lois*
- *une suite de fonctions étagées* $(\mathbf{f}^\epsilon; \epsilon > 0)$ *telles que la famille* $\langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon\rangle$ *définie par*

$$\langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon\rangle = \frac{1}{n}\sum_{i=1}^n \mathbf{f}^\epsilon(x_i^n)\cdot \tilde{Z}_i^n$$

réalise une approximation exponentielle de $\langle L_n, \mathbf{f}\rangle$, *i.e. :*

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P\left\{|\langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon\rangle - \langle L_n, \mathbf{f}\rangle| > \delta\right\} = -\infty \quad \textit{pour tout} \quad \delta > 0.$$

De plus, $\langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon\rangle$ *satisfait un PGD de bonne fonction de taux* $I_{\mathbf{f}^\epsilon}$ *donnée par (5.3) et* $\langle L_n, \mathbf{f}\rangle$ *satisfait un PGD de bonne fonction de taux*

$$\Upsilon(y) = \sup_{\eta>0}\liminf_{\epsilon\to 0}\inf_{y'\in B(y,\eta)} I_{\mathbf{f}^\epsilon}(y'), \quad y \in \mathbb{R}^m,$$

où $B(y,\epsilon) = \{y' \in \mathbb{R}^m, |y'-y| < \epsilon\}$

Preuve. Soit $\epsilon > 0$ fixé et $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ continue bornée. La proposition 3.12 nous assure de l'existence de $\mathbf{a}_1, \ldots, \mathbf{a}_p \in \mathbb{R}^{m\times d}$ et de $\epsilon_1, \ldots, \epsilon_p \in (0, \epsilon]$ tels que les ouverts $B_k^{\mathbf{f}} = f^{-1}B(\mathbf{a}_k, \epsilon_k)$ sont des ensemble R-continus et recouvrent $\mathcal{X}$, soit

$$\mathcal{X} \subset \cup_{k=1}^m B_k^{\mathbf{f}}.$$

De même, comme $\Gamma : x \to P_x$ est continue et $\mathcal{X}$ compact (H-5), on peut à nouveau appliquer la proposition 3.12 : il existe $P_1, \ldots, P_{p'} \in \mathcal{P}_\tau(\mathbb{R}^d)$ et $\epsilon'_1, \ldots, \epsilon'_{p'} \in (0, \epsilon]$ tels que les ouverts $C_l^P = \Gamma^{-1}B_{OW}(P_l, \epsilon'_l)$ sont R-continus et recouvrent $\mathcal{X}$:

$$\mathcal{X} \subset \cup_{l=1}^{p'} C_l^P.$$

La famille $(B_k^{\mathbf{f}} \cap C_l^P)_{k,l}$ est un recouvrement d'ouverts R-continus de $\mathcal{X}$. L'hypothèse (H-2) implique que chacun des ensembles est soit vide, soit de mesure strictement positive. Gardons les ensembles non vides. La famille précédente satisfait alors les hypothèses du lemme de partition (lemme 3.13), à savoir que chacun des ensembles participant au recouvrement de $\mathcal{X}$ est R-continu et de mesure strictement positive. Par suite, il existe une partition $(D_\iota; 1 \le \iota \le q)$ vérifiant les propriétés du lemme 3.13 : chacun des D_ι est R-continu et de mesure strictement positive. Qui plus est, il existe, pour tout ι, k et l tels que

$$\begin{aligned} D_\iota &\subset \overline{B_k^{\mathbf{f}}} \cap \overline{C_l^P} \\ &\subset \overline{f^{-1}B(\mathbf{a}_k, \epsilon_k)} \cap \overline{\Gamma^{-1}B_{OW}(P_l, \epsilon'_l)}. \end{aligned} \tag{5.4}$$

Considérons un appariement qui associe à chaque ι un unique couple $(k, l) = (k(\iota), l(\iota))$ tel que (5.4) soit satisfaite. Notons

$$\mathbf{f}^\epsilon = \sum_{\iota=1}^q \mathbf{a}_{k(\iota)}\ 1_{D_\iota}.$$

Alors $\|\mathbf{f}^\epsilon - \mathbf{f}\| \le \epsilon$ et chaque D_ι est un ensemble de R-continuité de mesure strictement positive. D'autre part, si $x \in D_\iota$, alors

$$d_{OW}(P_x, P_{l(\iota)}) \le \epsilon.$$

On peut maintenant construire proprement $(Z_i^n)_{i,n}$ et $(\tilde{Z}_i^n)_{i,n}$.

• *Construction de la famille* $(Z_i^n)_{i,n}$ *et de la famille approximante* $(\tilde{Z}_i^n)_{i,n}$.

Soit $x_i^n \in \mathcal{X}$. Il existe un unique D_ι tel que $x_i^n \in D_\iota$. En particulier,

$$d_{OW}(P_{x_i^n}, P_{l(\iota)}) < \epsilon,$$

où $l(\iota)$ est associé à ι via l'appariement subséquent à (5.4). Par suite, il existe une mesure de probabilité $\eta^{i,n}$ sur $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$ telle que

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{x-y}{\epsilon}\right) \eta^{i,n}(dxdy) \leq 1.$$

Considérons maintenant $Z_i^n(x,y) = x$ et $\tilde{Z}_i^n(x,y) = y$ où $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ alors Z_i^n est distribuée selon la loi $P_{x_i^n}$, $\tilde{Z}_i^n$ selon la loi $P_{j(l)}$ et

$$E\tau\left(\frac{Z_i^n - \tilde{Z}_i^n}{\epsilon}\right) \leq 1 \quad \Rightarrow \quad Ee^{\frac{|Z_i^n - \tilde{Z}_i^n|}{\epsilon}} \leq 2. \tag{5.5}$$

L'extension dénombrable usuelle $\otimes_{i,n}\eta^{i,n}$ sur le produit cartésien $\Pi_{i,n}\mathbb{R}^d \times \mathbb{R}^d$ entraine l'existence d'une famille $(Z_i^n)_{i,n}$ qui satisfait (H-7). Une des conséquences de la construction précédente est que la famille $(\tilde{Z}_i^n;\ x_i^n \in B_l)$ est une suite de variables aléatoires indépendantes distribuées selon un nombre fini de lois $(P_{l(\iota)}; 1 \leq \iota \leq q)$. Par suite, $\langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon \rangle$ satisfait les hypothèses du lemme 5.5 et donc satisfait un PGD de bonne fonction de taux $I_{\mathbf{f}^\epsilon}$.

• *L'approximation exponentielle.*

Remarquons d'abord que

$$\begin{aligned}
&P\left\{\left|\langle L_n, \mathbf{f}\rangle - \langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon\rangle\right| > \delta\right\} \\
&\leq\ P\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i)\cdot(Z_i - \tilde{Z}_i)\right| > \frac{\delta}{2}\right\} + P\left\{\left|\frac{1}{n}\sum_1^n (\mathbf{f}(x_i) - \mathbf{f}^\epsilon(x_i))\cdot \tilde{Z}_i\right| > \frac{\delta}{2}\right\}
\end{aligned}$$

et

$$\begin{aligned}
&\limsup_n \frac{1}{n}\log P\left\{\left|\langle L_n, \mathbf{f}\rangle - \langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon\rangle\right| > \delta\right\} \\
&\leq\ \limsup_n \frac{1}{n}\log P\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i)\cdot(Z_i - \tilde{Z}_i)\right| > \frac{\delta}{2}\right\} \\
&\qquad \vee \limsup_n \frac{1}{n}\log P\left\{\left|\frac{1}{n}\sum_1^n (\mathbf{f}(x_i) - \mathbf{f}^\epsilon(x_i))\cdot \tilde{Z}_i\right| > \frac{\delta}{2}\right\},
\end{aligned}$$

où $a \vee b = \max(a,b)$. Considérons d'abord :

$$\begin{aligned}
P\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i)\cdot(Z_i - \tilde{Z}_i)\right| > \frac{\delta}{2}\right\} &\leq P\left\{\frac{1}{n}\sum_1^n |Z_i - \tilde{Z}_i| > \frac{\delta}{2\|\mathbf{f}\|}\right\} \\
\leq\ \exp\left(-\frac{n\delta}{2\|\mathbf{f}\|\epsilon}\right) \prod_1^n Ee^{\frac{|Z_i - \tilde{Z}_i|}{\epsilon}} &\leq \exp\left(-\frac{n\delta}{2\|\mathbf{f}\|\epsilon}\right) 2^n,
\end{aligned}$$

où la dernière inégalité provient de (5.5). Par suite,

$$\limsup_n \frac{1}{n} \log P\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i)\cdot(Z_i-\tilde{Z}_i)\right| > \frac{\delta}{2}\right\} \leq -\frac{n\delta}{2\|f\|\epsilon}. \tag{5.6}$$

Considérons maintenant

$$\begin{aligned} P\left\{\left|\frac{1}{n}\sum_1^n (\mathbf{f}(x_i)-\mathbf{f}^\epsilon(x_i))\cdot\tilde{Z}_i\right| > \frac{\delta}{2}\right\} &\leq P\left\{\frac{1}{n}\sum_1^n |\tilde{Z}_i| > \frac{\delta}{2\epsilon}\right\} \\ &\leq \exp\left(-\frac{\alpha^* n\delta}{2\epsilon}\right)\prod_1^n E e^{\alpha^*|\tilde{Z}_i|}, \end{aligned}$$

où $\alpha^* > 0$ est choisie de sorte que (5.2) est satisfaite. On obtient alors

$$\limsup_n \frac{1}{n}\log P\left\{\left|\frac{1}{n}\sum_1^n (\mathbf{f}(x_i)-\mathbf{f}^\epsilon(x_i))\cdot\tilde{Z}_i\right| > \frac{\delta}{2}\right\} \leq -\frac{\alpha^*\delta}{2\epsilon} + \log S(\alpha^*). \tag{5.7}$$

Finalement, en combinant (5.6) et (5.7) avec (5.5), on obtient

$$\lim_{\epsilon\to 0}\limsup_n \frac{1}{n}\log P\left\{\left|\langle L_n, \mathbf{f}\rangle - \langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon\rangle\right| > \delta\right\} = -\infty,$$

où $S(\alpha^*)$ est donnée par (5.5). Par conséquent, la famille $(\langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon\rangle; \epsilon > 0)$ réalise une approximation exponentielle de $\langle L_n, \mathbf{f}\rangle$. Comme $\langle \tilde{L}_n^\epsilon, \mathbf{f}^\epsilon\rangle$ satisfait un PGD du fait du lemme 5.5, le théorème 4.2.16 dans [20] entraine un PGD faible pour $\langle L_n, \mathbf{f}\rangle$ de fonction de taux

$$\Upsilon(y) = \sup_\eta \liminf_\epsilon \inf_{y'\in B(y,\eta)} I^\epsilon_{\mathbf{f}^\epsilon}(y').$$

- *Le PGD complet : tension exponentielle de* $\langle L_n, \mathbf{f}\rangle$.

$$P\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i)\cdot Z_i\right| > K\right\} \leq P\left\{\frac{1}{n}\sum_1^n |Z_i| > \frac{K}{\|\mathbf{f}\|}\right\} \leq \exp\left(-\frac{nK\alpha}{\|\mathbf{f}\|}\right)\prod_1^n E e^{\alpha Z_i|}.$$

Comme précédemment, si $\alpha > 0$ est choisie de sorte que (5.2) est satisfaite, alors

$$\limsup_n \frac{1}{n}\log P\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i)\cdot Z_i\right| > K\right\} \leq -\frac{K\alpha}{\|\mathbf{f}\|} + \log S(\alpha) \xrightarrow[K\to\infty]{} -\infty.$$

Par conséquent, $\langle L_n, \mathbf{f}\rangle$ est exponentiellement tendue et satisfait donc un PGD complet de bonne fonction de taux Υ. $\square$

5.2.2 Plus d'informations sur la fonction de taux

Introduisons d'abord quelques définitions usuelles en analyse convexe. On dit que J^ϵ **epi-converge** vers J et on écrit : $J = \mathrm{e}-\lim_{\epsilon\to 0} J^\epsilon$, si, pour tout $y \in \mathbb{R}^d$,

$$\begin{aligned} \liminf_{\epsilon\to 0} J^\epsilon(y^\epsilon) &\geq J(y) \quad \text{pour toute suite } y^\epsilon \xrightarrow[\epsilon\to 0]{} y, && (5.8)\\ \text{and} \quad \limsup_{\epsilon\to 0} J^\epsilon(y^\epsilon) &\leq J(y) \quad \text{pour une suite } y^\epsilon \xrightarrow[\epsilon\to 0]{} y. && (5.9) \end{aligned}$$

Cela est équivalent à :

$$\begin{aligned} \sup_{\eta>0} \liminf_{\epsilon\to 0} \inf_{y'\in B(y,\eta)} J^\epsilon &\geq J, && (5.10)\\ \text{et} \quad \sup_{\eta>0} \limsup_{\epsilon\to 0} \inf_{y'\in B(y,\eta)} J^\epsilon &\leq J. && (5.11) \end{aligned}$$

Pour les détails, on pourra se reporter au livre de Rockafellar et Wets [49], chapitre 7. Notons que certains auteurs appellent ce type de convergence la **Mosco-convergence** (voir Zabell [58]).
Ce type de convergence existe en théorie des grandes déviations. En effet, si U_n^ϵ satisfait un PGD de fonction de taux J^ϵ et si U_n^ϵ est une approximation exponentielle de U_n, alors U_n satisfait un principe de grandes déviations faibles de fonction de taux $J = e - \lim_{\epsilon\to 0} J^\epsilon$ (théorème 4.2.16 dans [20] et sa preuve).
Rappelons les notations introduites à la remarque 5.4.

Proposition 5.7

$$\Upsilon(y) \leq I_{\mathbf{f}}(y).$$

Preuve. On divise la preuve en plusieurs étapes. Rappelons dans un premier temps quelques caractéristiques essentielles de l'approximation exponentielle comsidérée dans le lemme précédent :
Pour tout $\epsilon > 0$, on a construit une partition $(D_\iota; 1 \leq \iota \leq q)$ de $\mathcal{X}$ telle que :
- $\mathbf{f}(x)$ est approximée par une fonction étagée $\mathbf{f}^\epsilon(x)$ au sens de la norme uniforme $\|\mathbf{f} - \mathbf{f}^\epsilon\| \leq \epsilon$.
- P_x est approximée par $P_{l(\iota)}$ dans le sens où $d_{OW}(P_x, P_{l(\iota)}) \leq \epsilon$,

et

$$I_{\mathbf{f}^\epsilon}(y) = \inf_{u\in L^1_d} \left\{ \int_{\mathcal{X}} \Lambda^*_\epsilon(x, u(x))\, dR;\ \int_{\mathcal{X}} \mathbf{f}^\epsilon \cdot u\, dR = y \right\} \quad y \in \mathbb{R}^m,$$

où $\Lambda^*_\epsilon(x, z)$ est la conjuguée convexe de $\Lambda_\epsilon(x, \lambda) = \log \int_{\mathbb{R}^d} e^{\lambda\cdot z} P^\epsilon_{l(\iota)}(\,dz)$. Enfin, nous avons vu que $\Upsilon = e - \lim_{\epsilon\to 0} I_{\mathbf{f}^\epsilon}$.

- *Soit $x \in \mathcal{X}$ fixé. Alors $\Lambda(x, \lambda) = e - \lim_{\epsilon\to 0} \Lambda_\epsilon(x, \lambda)$.*

En effet, pour tout $\epsilon > 0$, il existe $d_{OW}(P^\epsilon_{l(\iota)}, P_x) \leq \epsilon$. Considérons maintenant une famille de variables aléatoires $(\check{Z}^\epsilon_i)$ i.i.d. et distribuée selon $P^\epsilon_{l(\iota)}$. Alors $\frac{1}{n}\sum \check{Z}^\epsilon_i$ est une approximation exponentielle de $\frac{1}{n}\sum Z^x_i$. Comme cette dernière quantité satisfait un principe de grandes déviations de bonne fonction de taux $\Lambda^*(x,z)$ du fait du théorème de Cramér, on obtient :

$$\text{Soit } x \text{ fixé, } \Lambda^*(x,z) = e - \lim_{\epsilon\to 0} \Lambda^*_\epsilon(x,z).$$

Le théorème de Wijsman (théorème 11.34, [49]) entraine alors que :

$$\text{Soit } x \text{ fixé, } \Lambda(x,\lambda) = e - \lim_{\epsilon\to 0} \Lambda_\epsilon(x,\lambda).$$

- *Pour toute suite $\theta^\epsilon \to \theta$ à valeurs dans $\mathbb{R}^m$, on a :*

$$\liminf_{\epsilon\to 0} \int_{\mathbb{R}^d} \Lambda_\epsilon(x, \sum_{j=1}^m \theta^\epsilon_j f^\epsilon_j(x))\, R(dx) \geq \int_{\mathbb{R}^d} \Lambda(x, \sum_{j=1}^m \theta_j f_j(x))\, R(dx),$$

où f_j (respectivement f^ϵ_j) est la i^eligne de la matrice $\mathbf{f}$ (respectivement $\mathbf{f}^\epsilon$).
Plutôt que de travailler directement avec $\Lambda(x,\lambda)$, on utilisera, pour des questions de positivité, la décomposition suivante :

$$\begin{aligned}
\Lambda(x,\lambda) &= \log \int_{\mathbb{R}^d} e^{\lambda\cdot z}\, P(x,dz) \\
&= \log \int_{\mathbb{R}^d} e^{\lambda\cdot(z-m_x)+\lambda\cdot m_x}\, P(x,dz) \\
&= \tilde{\Lambda}(x,\lambda) + \lambda\cdot m_x,
\end{aligned}$$

où $m_x = \int_{\mathbb{R}^d} z\, P(x,dz)$ et $\tilde{\Lambda}$ est associée à la variable aléatoire centrée $Z^x - m_x$. En particulier, $\tilde{\Lambda}$ est positive. De manière similaire, $\Lambda^\epsilon(x,\lambda) = \tilde{\Lambda}^\epsilon(x,\lambda) + m^\epsilon_x$. On rappelle que $W(P,Q) \leq d_{OW}(P,Q)$ où W est la distance de Wasserstein. Une application immédiate du théorème de Kantorovich-Rubinstein (théorème 11.8.2,[23]) entraine que :

$$\lim_{\epsilon\to 0} m^\epsilon_x = m_x. \tag{5.12}$$

Par suite,

$$\tilde{\Lambda} = e - \lim_{\epsilon\to 0} \tilde{\Lambda}_\epsilon.$$

En particulier, si $\theta^\epsilon \to \theta$ alors, pour tout x, $\sum_{j=1}^m \theta^\epsilon_j f^\epsilon_j(x) \to \sum_{j=1}^m \theta_j f_j(x)$ et la borne inférieure (5.8) de l'épi-convergence entraine

$$\liminf_{\epsilon\to 0} \tilde{\Lambda}_\epsilon(x, \sum_{j=1}^m \theta^\epsilon_j f^\epsilon_j(x)) \geq \tilde{\Lambda}(x, \sum_{j=1}^m \theta_j f_j(x)).$$

Comme les $\tilde{\Lambda}$s sont positives, on peut appliquer le lemme de Fatou :

$$\liminf_{\epsilon\to 0}\int_{\mathbb{R}^d}\tilde{\Lambda}_\epsilon(x,\sum_{j=1}^m\theta_j^\epsilon f_j^\epsilon(x))\,R(dx)\geq\int_{\mathbb{R}^d}\tilde{\Lambda}(x,\sum_{j=1}^m\theta_j f_j(x))\,R(dx). \tag{5.13}$$

Il suffit, pour conclure, d'injecter (5.12) dans l'inégalité précédente.

• *L'inégalité* $\Upsilon\leq I_{\mathbf{f}}$. Rappelons que $\int\Lambda_\epsilon dR$ et $\int\Lambda_\epsilon^* dR$ sont conjuguées convexes. Une application directe du théorème 3 dans [48] entraine que :

$$\begin{aligned}&\sup_{\theta\in\mathbb{R}^m}\left\{\theta\cdot y-\int\Lambda_\epsilon(x,\sum_{j=1}^m\theta_j f_j^\epsilon(x))\,R(dx)\right\}\\ &=\inf_{u\in L_d^1}\left\{\int\Lambda_\epsilon^*(x,u(x))\,dR;\ \int_{\mathcal{X}}\mathbf{f}^\epsilon(x)u(x)\,R(dx)=y\right\}.\end{aligned}$$

En particulier, $I_{\mathbf{f}^\epsilon}(y)=\sup_{\theta\in\mathbb{R}^m}\left\{\theta\cdot y-\int\Lambda_\epsilon(x,\sum_{j=1}^m\theta_j f_j^\epsilon(x))\,R(dx)\right\}$. Il reste à appliquer une fois encore le théorème de Wijsman à (5.13) pour conclure que :

$$\Upsilon(y)\leq I_{\mathbf{f}}(y).$$

La proposition 5.7 est démontrée. □

5.3 Le caractère métrique de d_{OW}

Preuve. • Notons tout d'abord que si P et Q appartiennent à $\mathcal{P}_\tau(\mathbb{R}^d)$ alors $d_{OW}(P,Q)$ est finie. En effet, si $\eta(dzdz')=P(dz)\otimes Q(dz')$ alors

$$\int\tau\left(\frac{z-z'}{a}\right)\eta(dzdz')\leq\int e^{\frac{|z|}{a}}P(dz)\int e^{\frac{|z'|}{a}}Q(dz')-1.$$

Le théorème de convergence dominée entraine que

$$\lim_{a\to\infty}\int e^{\frac{|z|}{a}}P(dz)\int e^{\frac{|z|}{a}}Q(dz')=1.$$

Par suite, il existe a et η tels que $\int\tau\left(\frac{z-z'}{a}\right)\eta(dzdz')\leq 1$.

• Démontrons maintenant que d_{OW} satisfait l'inégalité triangulaire. On va pour cela adapter la preuve du lemme 11.8.3 de Dudley [23].

On rappelle que $\tau(z)=e^{|z|}-1$. On choisit un réel κ de manière à ce que

$\tau(\kappa) \leq 1$. Soient $P, Q \in \mathcal{P}_\tau$, alors il existe une probabilité jointe μ sur $\mathbb{R}^d \times \mathbb{R}^d$ de marginales P et Q et un réel $a > 0$ tels que

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{z - z'}{a}\right) \mu(dzdz') \leq 1, \tag{5.14}$$

avec $a \leq d_{OW}(P, Q) + \epsilon$. Considérons une partition mesurable $(S_n)_{n \geq 1}$ de $\mathbb{R}^d$ vérifiant :

$$\forall z, z' \in S_n, \quad |z - z'| \leq \kappa\epsilon.$$

Une telle partition existe sans problèmes. On définit la probabilité suivante sur $\mathbb{R}^d \times \mathbb{R}^d$.

$$\mu'(A \times B) = \sum_{m,n} \frac{\mu\left((A \cap S_m) \times S_n\right) \mu\left(S_m \times (B \cap S_n)\right)}{\mu(S_m \times S_n)},$$

où la sommation concerne les indices m et n tels que $\mu(S_m \times S_n) > 0$. Il est immédiat de vérifier que μ' est une probabilité de marginales P et Q. Montrons que

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{z - z'}{a + 2\epsilon}\right) \mu'(dzdz') \leq 1. \tag{5.15}$$

On a

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{z - z'}{a + 2\epsilon}\right) \mu'(dzdz') = \sum_{m,n} \int_{S_m \times S_n} \tau\left(\frac{z - z'}{a + 2\epsilon}\right) \mu'(dzdz').$$

Soit $z_n \in S_n$, alors

$$\begin{aligned}
&\int_{S_m \times S_n} \tau\left(\frac{z - z'}{a + 2\epsilon}\right) \mu'(dzdz') \\
&\qquad = \int_{S_m \times S_n} \tau\left[\left(\frac{a + \epsilon}{a + 2\epsilon}\right) \frac{z - z_n}{a + \epsilon} + \left(\frac{\epsilon}{a + 2\epsilon}\right) \frac{z_n - z'}{\epsilon}\right] \mu'(dzdz') \\
&\qquad \leq \frac{a + \epsilon}{a + 2\epsilon} \int_{S_m \times S_n} \tau\left(\frac{z - z_n}{a + \epsilon}\right) \mu'(dzdz') \\
&\qquad\quad + \frac{\epsilon}{a + 2\epsilon} \int_{S_m \times S_n} \tau\left(\frac{z_n - z'}{\epsilon}\right) \mu'(dzdz') \\
&\qquad \leq \frac{a + \epsilon}{a + 2\epsilon} \int_{S_m \times S_n} \tau\left(\frac{z - z_n}{a + \epsilon}\right) \mu'(dzdz') + \frac{\epsilon}{a + 2\epsilon} \tau(\kappa) \mu'(S_m \times S_n).
\end{aligned} \tag{5.16}$$

La première inégalité provient de la convexité de τ et la seconde de la définition même de κ. Remarquons maintenant que

$$\int_{S_m \times S_n} \tau\left(\frac{z-z_n}{a+\epsilon}\right)\mu'(dzdz') = \int_{S_m \times S_n} \tau\left(\frac{z-z_n}{a+\epsilon}\right)\mu(dzdz').$$

Enfin,

$$\begin{aligned}
&\int_{S_m \times S_n} \tau\left(\frac{z-z_n}{a+\epsilon}\right)\mu(dzdz') \\
&\qquad = \int_{S_m \times S_n} \tau\left[\left(\frac{a}{a+\epsilon}\right)\frac{z-z'}{a} + \left(\frac{\epsilon}{a+\epsilon}\right)\frac{z_n-z'}{\epsilon}\right]\mu(dzdz') \\
&\qquad \leq \frac{a}{a+\epsilon}\int_{S_m \times S_n} \tau\left(\frac{z-z'}{a}\right)\mu(dzdz') + \frac{\epsilon}{a+\epsilon}\int_{S_m \times S_n} \tau\left(\frac{z_n-z'}{\epsilon}\right)\mu(dzdz') \\
&\qquad \leq \frac{a}{a+\epsilon}\int_{S_m \times S_n} \tau\left(\frac{z-z'}{a}\right)\mu(dzdz') + \frac{\epsilon}{a+\epsilon}\tau(\kappa)\mu(S_m \times S_n).
\end{aligned} \tag{5.17}$$

En injectant l'équation (5.17) dans (5.16) et en utilisant la définition même de a (voir (5.14)) , on obtient bien

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{z-z'}{a+2\epsilon}\right)\mu'(dzdz') \leq 1.$$

Soit maintenant $Q, T \in \mathcal{P}_\tau$, alors il existe une loi jointe ν sur $\mathbb{R}^d \times \mathbb{R}^d$ et un réel b tels que

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{z'-z''}{b}\right)\nu(dz'dz'') \leq 1; \quad d_{OW}(Q,T) \leq b+\epsilon.$$

De même que précédemment, on construit ν' qui vérifie

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{z'-z''}{b+2\epsilon}\right)\nu'(dz'dz'') \leq 1. \tag{5.18}$$

On peut maintenant, à partir de μ' et de ν', définir une probabilité M sur $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$ telle que (z, z') a pour distribution μ', (z', z'') a pour distribution ν' et z et z'' sont indépendants conditionnellement à z'. (on pourra se reporter à la preuve du théorème 11.8.3, Dudley [23], pour les détails de la construction).

Evaluons maintenant :

$$\begin{aligned}
&\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \tau\left(\frac{z-z''}{a+b+4\epsilon}\right) M(dzdz'dz'') \\
&\qquad = \int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d} \tau\left[\left(\frac{a+2\epsilon}{a+b+4\epsilon}\right)\frac{z-z'}{a+2\epsilon}\right. \\
&\qquad\qquad \left. + \left(\frac{b+2\epsilon}{a+b+4\epsilon}\right)\frac{z'-z''}{b+2\epsilon}\right] M(dzdz'dz) \\
&\qquad \leq \frac{a+2\epsilon}{a+b+4\epsilon}\int_{\mathbb{R}^d\times\mathbb{R}^d} \tau\left(\frac{z-z'}{a+2\epsilon}\right)\mu'(dzdz') \\
&\qquad\qquad + \frac{b+2\epsilon}{a+b+4\epsilon}\int_{\mathbb{R}^d\times\mathbb{R}^d} \tau\left(\frac{z'-z''}{b+2\epsilon}\right)\nu'(dz'dz'') \\
&\qquad \leq 1,
\end{aligned}$$

où la dernière inégalité provient des équations (5.15) et (5.18). On a ainsi montré que

$$\begin{aligned}
&d_{OW}(P,T) \leq a+b+4\epsilon \quad \text{soit} \\
&d_{OW}(P,T) \leq d_{OW}(P,Q) + d_{OW}(Q,T) + 6\epsilon.
\end{aligned}$$

L'inégalité triangulaire est ensuite obtenue en faisant tendre ϵ vers zéro.
• Reste à prouver que $d_{OW}(P,Q) = 0$ implique $P = Q$. Considérons pour cela la métrique de Wasserstein habituelle :

$$W(P,Q) = \inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} |z-z'|\eta(dzdz'),\ \eta \in M(P,Q)\right\}.$$

Alors $W(P,Q) \leq d_{OW}(P,Q)$. En effet, $|z| \leq \tau(z)$ par suite

$$\int \tau\left(\frac{z-z'}{a}\right)\eta(dzdz') \leq 1 \Rightarrow \int |z-z'|\eta(dzdz') \leq a.$$

Ainsi, $d_{OW}(P,Q) = 0$ implique que $W(P,Q) = 0$ et comme W est une métrique, cela entraine $P = Q$ et le lemme 5.1 est démontré. $\square$

Conclusion

Nous avons, durant cette thèse, abordé l'étude des grandes déviations associées à certaines mesures empiriques :

$$\text{la probabilité empirique} \qquad \frac{1}{n}\sum_{i=1}^{n}\delta_{Z_i},$$
$$\text{La mesure MEM} \qquad \frac{1}{n}\sum_{i=1}^{n}Z_i\delta_{x_i},$$

quand les variables aléatoires sous-jacentes sont indépendantes, en général identiquement distribuées, et possèdent quelques moments exponentiels :

$$Ee^{\alpha|Z|}<\infty \quad \text{pour un } \alpha>0.$$

Cette étude a été motivée par la grande variété d'applications associées à ces modèles (principe de conditionnement de Gibbs, résultats de convergence via la méthode du Maximum d'entropie en moyenne, grandes déviations pour les processus, etc.). Une autre motivation, d'ordre plus technique, a résidé dans le fait que les outils habituels pour établir les bornes inférieures des principes de grandes déviations (changement de probabilité exponentielle, théorème de Gärtner-Ellis) n'étaient, en général, pas "transposables" à ce contexte.

Dans le cas de la probabilité empirique, nous avons établi un principe de grandes déviations (chapitre 1) et la fonction de taux a été identifiée grâce a l'introduction d'espaces d'Orlicz adaptés. Nous avons également établi un principe de grandes déviations pour la mesure MEM (chapitres 3, 4). La fonction de taux a été identifiée dans un cas multidimensionnel et a fait apparaître un terme que l'on appelle contribution singulière et dont l'existence a été mise en évidence par Lynch et Sethuraman [39] dans le cas de processus de Lévy.

Ces résultats nous ont permis d'étendre les conditions d'application du principe de conditionnement de Gibbs (chapitre 2) ainsi que certains résultats de grandes déviations pour des processus empiriques (chapitre 4).

La dernière partie de cette thèse a été consacrée à l'étude des grandes déviations de la moyenne empirique pondérée :

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i)Z_i,$$

dans le cas où les variables Z_i ne sont plus indépendantes et identiquement distribuées (chapitre 5). Celles-ci restent indépendantes, mais la loi de Z_i dépend de x_i. Cette dépendance vis-à-vis de x_i a été clairement formulée via l'introduction d'une distance ad hoc (distance d'Orlicz-Wasserstein). Enfin, le principe de grandes déviations a été établi mais la fonction de taux n'a pas été clairement identifiée et son expression demeure abstraite.

Perspectives

Autour du principe de conditionnement de Gibbs. Nous avons vu au chapitre 2 que lorsque f n'a que quelques moments exponentiels :

$$\exists \alpha > 0, \quad Ee^{\alpha|f(Y)|} < \infty,$$

il existe des cas où on ne sait rien dire sur le comportement de

$$E[f(Y_1)|L_n^Y \in A_\delta],$$

si ce n'est qu'on sait que la convergence

$$E[f(Y_1)|L_n^Y \in A_\delta] \to \int f\, d\nu_*,$$

n'a pas lieu (on pourra se reporter au bilan du chapitre 2, page 58). Nous pensons néanmoins qu'il est possible qu'un terme apparaisse dans la convergence :

$$E[f(Y_1)|L_n^Y \in A_\delta] \to \int f\, d\nu_*, + \int_{\mathcal{S}} \langle f, \ell^s\rangle Q(\, d\ell^s),$$

où Q est une espèce de distribution de probabilité sur l'ensemble des formes singulières $\mathcal{S}$. En quelque sorte, Q serait une probabilité limite qui pondèrerait chacune des contributions $\langle f, \ell^s\rangle$ à l'aune de son importance dans le phénomène limite.

C'est un problème intéressant mais qui nous semble vraiment difficile et qui mériterait d'être sérieusement motivé par des exemples venus de la mécanique statistique ou d'ailleurs.

Grandes déviations pour la mesure MEM en dimension infinie. Il semble intéressant, pour certains problèmes statistiques, de considérer les grandes déviations de la mesure MEM $\frac{1}{n}\sum Z_i\delta_{x_i}$ dans le cas où les Z_i sont à valeurs dans un espace de dimension infinie et n'ont que quelques moments exponentiels :

$$\exists\alpha > 0, \quad Ee^{\alpha\|Z\|} < \infty$$

Nous pensons que la méthode développée au chapitre 3 pour établir le PGD est susceptible d'être généralisée à ce cadre, le principal problème consistant à étudier la tension exponentielle du modèle. Cette tension exponentielle, contrairement au cas fini-dimensionnel, n'est pas automatiquement acquise.

D'autre part, l'existence d'un papier de Rockafellar [47] généralisant certaines égalités duales à un cadre infini-dimensionnel nous rend optimiste sur l'identification de la fonction de taux.

Grandes déviations dans le cas non identiquement distribué. Le premier problème en suspens est l'identification complète de la fonction de taux. C'est un problème qui nous semble assez difficile. Cette identification permettrait quasi-immédiatement, via la technique de la limite projective, d'obtenir les grandes déviations pour la mesure MEM dans un cadre d'indépendance mais où il n'y a plus equi-distribution, ainsi que les grandes déviations de certains processus (du type de la marche aléatoire de Mogul'skii) à accroissement indépendants mais pas stationnaires.

Plus généralement, il semblerait très intéressant de formaliser, en poursuivant les idées développées au chapitre 5, les variations des lois associées à deux incréments successifs d'un processus à accroissements indépendants. Cela permettrait d'envisager l'étude systématique des grandes déviations associées à des processus à accroissements indépendants, ou de différents statistiques associées (on pourra se référer aux travaux de Djellout *et al.* [21] pour l'étude des grandes déviations d'une statistique associée à un processus à accroissements indépendants en l'absence de tous les moments exponentiels).

Bibliographie

[1] A. Aboulalaâ. The Gibbs principle for Markov jump processes. *Stochastic Process. Appl.*, 64 :257–271, 1996.

[2] R. A. Adams. *Sobolev Spaces.* Academic Press, New York, 1975.

[3] G. Ben Arous, A. Dembo, and A. Guionnet. Aging of spherical spin glasses. *Probab. Theory Related Fields*, 120(1) :1–67, 2001.

[4] R.R. Bahadur and S. Zabell. Large deviations of the sample means in general vector spaces. *Ann. Probab.*, 7 :587–621, 1979.

[5] B. Bercu, F. Gamboa, and M. Lavielle. Sharp large deviations for gaussian quadratic forms with applications. *ESAIM Probab. Statist.*, 4 :1–24, 2000.

[6] B. Bercu, F. Gamboa, and A. Rouault. Large deviations for quadratic functionals of stationary Gaussian processes. *Stochastic Process. Appl.*, 71 :75–90, 1997.

[7] E. Bolthausen and U. Schmock. On the maximum entropy principle for uniformly ergodic Markov chains. *Stochastic Process. Appl.*, 33 :1–27, 1989.

[8] A. A. Borovkov. Boundary-value problems for random walks and large deviations in function spaces. *Theory Probab. Appl.*, 12 :575–595, 1967.

[9] N. Bourbaki. *Eléments de Mathématique - Intégration 1-4.* Hermann, Paris, second edition, 1961.

[10] W. Bryc and A. Dembo. Large deviations for quadratic functionals of Gaussian processes. *J. Theoret. Probab.*, 10 :307–332, 1997.

[11] J.M. Van Campenhout and T.M. Cover. Maximum entropy and conditionnal probability. *IEEE Trans. Inform. Theory*, IT-27(4) :483–489, 1981.

[12] H. Cramér. Sur un nouveau théorème limite de la théorie des probabilités. In Hermann, editor, *Colloque consacré à la théorie des Probabilités*, volume 736, pages 5–23, Paris, 1938.

[13] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, 3(1) :146–158, 1975.

[14] I. Csiszár. Sanov property, generalized I-projection and a conditionnal limit theorem. *Ann. Probab.*, 12(3) :768–793, 1984.

[15] I. Csiszár, F. Gamboa, and E. Gassiat. Mem pixel correlated solutions for generalized moment and interpolation problems. *IEEE Trans. Inform. Theory*, 45(7) :2253–2270, 1999.

[16] D. Dacunha-Castelle and F. Gamboa. Maximum d'entropie et problème des moments. *Ann. Inst. H. Poincaré Probab. Statist.*, 26 :567–596, 1990.

[17] A. de Acosta. Large deviations for vector-valued Lévy processes. *Stochastic Process. Appl.*, 51 :75–115, 1994.

[18] A. de Acosta. On large deviations of empirical measures in the τ-topology. *J. Appl. Probab.*, 31A (special volume) :41–47, 1994.

[19] A. de La Fortelle. *Contribution à la théorie des grandes déviations et applications*. PhD thesis, École Nationale des Ponts et Chaussées, 2000.

[20] A. Dembo and O. Zeitouni. *Large Deviations Techniques And Applications*. Springer Verlag, New York, second edition, 1998.

[21] H. Djellout, A. Guillin, and L. Wu. Large and moderate deviations for quadratic empirical processes. *Stat. Inference Stoch. Process.*, 2(3) :195–225, 1999.

[22] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov processes expectations for large time, III. *Comm. Pure Appl. Math.*, 29 :389–461, 1976.

[23] R. M. Dudley. *Real Analysis and Probability*. Wadsworth and Brooks/Cole, 1989.

[24] N. Dunford and J.T. Schwartz. *Linear Operators, Part I*. Interscience Publishers Inc., New York, 1958.

[25] P. Eichelsbacher and U. Schmock. Large deviations of products of empirical measures and U-empirical measures in strong topologies. Preprint, 1996.

[26] P. Eichelsbacher and U. Schmock. Exponential approximations in completely regular topological spaces and extensions of Sanov's theorem. *Stochastic Process. Appl.*, 77 :233–251, 1998.

[27] R. S. Ellis. Large deviations for a general class of random vectors. *Ann. Probab.*, 12 :1–12, 1984.

[28] R.S. Ellis, J. Gough, and J.V. Pulé. The large deviation principle for measures with random weights. *Rev. Math. Phys.*, 5(4) :659–692, 1993.

[29] F. Gamboa and E. Gassiat. Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann. Statist.*, 25(1) :328–350, 1997.

[30] F. Gamboa, A. Rouault, and M. Zani. A functional large deviations principle for quadratic forms of Gaussian stationary processes. *Statist. Probab. Lett.*, 43 :299–308, 1999.

[31] J. Gärtner. On large deviations from the invariant measure. *Theory Probab. Appl.*, 22 :24–39, 1977.

[32] J.R. Giles. Convex analysis with application in the differentiation of convex functions. In *Research Notes in Math. 58*. Pitman, London, 1982.

[33] P. Groeneboom, J. Ooesterhoff, and F.H. Ruymgaart. Large deviation theorems for empirical probability measures. *Ann. Probab.*, 7 :553–586, 1979.

[34] J. L. Kelley. *General Topology.* Springer Verlag, New York, 1955.

[35] A. Kozek. Convex integral functionals on Orlicz spaces. *Ann. Soc. Math. Polonae, Series I*, pages 109–135, 1979.

[36] C. Léonard. Large deviations for Poisson random measures and processes with independent increments. *Stochastic Process. Appl.*, 85 :93–121, 2000.

[37] C. Léonard. Minimizers of energy functionals under not very integrable constraints. Preprint, 2000.

[38] C. Léonard and J. Najim. An extention of Sanov's theorem. Application to the Gibbs conditionning principle. Preprint, 2000.

[39] J. Lynch and J. Sethuraman. Large deviations for processes with independent increments. *Ann. Probab.*, 15(2) :610–627, 1987.

[40] A.A. Mogul'skii. Large deviations for trajectories of multi-dimentional random walks. *Theory Probab. Appl.*, 21 :300–315, 1976.

[41] A.A. Mogul'skii. Large deviations for processes with independent increments. *Ann. Probab.*, 21 :202–213, 1993.

[42] J. Najim. A Cramér type theorem for weighted empirical means. preprint, 2001.

[43] M. Rao and Z. D. Ren. *Theory of Orlicz spaces.* Marcel Dekker Inc., New-York, 1991.

[44] M. Reed and B. Simon. *Methods of Modern Mathematical Physics I.* Academic Press, New York, 1980.

[45] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. Math.*, 24(3) :525–539, 1968.

[46] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1970.

[47] R. T. Rockafellar. Convex integral functionals and duality. In E. Zarantello, editor, *Contributions to Non Linear Functional Analysis*, pages 215–236. Academic Press, 1971.

[48] R. T. Rockafellar. Integrals which are convex functionals, II. *Pacific J. Math.*, 39(2) :439–469, 1971.

[49] R. T. Rockafellar and R. J-B. Wets. *Variational Analysis.* Springer, 1998.

[50] I. N. Sanov. On the probability of large deviations of random variables. *Selected transl. Math. Statist. and Prob.*, 1 :214–244, 1961.

[51] A. Schied. Cramér's condition and Sanov's theorem. *Statist. Probab. Lett.*, 39 :55–60, 1998.

[52] D. W. Stroock and O. Zeitouni. Microcanonical distributions, Gibbs states, and the equivalence of ensembles. In R. Durrett and H. Kesten, editors, *Festchrift in honour of F. Spitzer*, pages 399–424, Basel, Switzerland, 1991. Birkhäuser.

[53] D.W. Stroock. *Probability theory : an analytic view.* Cambridge University Press, revised edition, 1993.

[54] M. van den Berg, T. C. Dorlas, J. T. Lewis, and J. V. Pulé. A perturbed mean field model of an interacting boson gas and the large deviation principle. *Comm. Math. Phys.*, 127(1) :41–69, 1990.

[55] S.R.S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19 :261–286, 1966.

[56] V. Vinogradov. Large deviations for i.i.d. random sums when Cramér's condition is fulfilled only on a finite interval. *C. R. Math. Rep. Acad. Sci. Canada*, 15(5) :229–234, 1993.

[57] V. Vinogradov. *Refined large deviation limit theorems.* Longman Scientific & Technical, Harlow, 1994.

[58] S. L. Zabell. Mosco convergence in locally convex spaces. *J. Funct. Anal.*, 110(1) :226–246, 1992.

[59] M. Zani. *Grandes déviations pour des fonctionnelles issues de la statistique des processus.* PhD thesis, Université Paris-Sud, 1999.

Annexe A

Un exemple de non-escarpement et autres propriétés

On consigne dans cette annexe des propriétés liées à la mesure de probabilité

$$P(\,dz) = c\frac{e^{-z}}{1+z^3}\,dz. \quad z \geq 0.$$

Cette probabilité présente l'intérêt d'avoir une log-Laplace non-escarpée. C'est par conséquent une source de modèles simples ne satisfaisant pas les hypothèses du théorème de Gärtner-Ellis, ou plus généralement pour lesquels la technique de changement de probabilité exponentielle (cf. l'introduction) pour l'établissement de la minoration du PGD ne marche pas.

Cette probabilité nous a servi au chapitre 1 pour exhiber des projections de Csiszár généralisées. Nous avons illustré au chapitre 2, via cette probabilité, un principe de conditionnement de Gibbs dans le cas où l'hypothèse de loi des Grands Nombres sous-jacente n'était pas vérifiée. Une variante de la probabilité P est à l'origine des contre-exemples du chapitre 3. Enfin, cette probabilité apparaît lors d'exemples au chapitre 4.

Proposition A.1 *1.* **(non-escarpement)** *La fonction*

$$\Lambda(\lambda) = \log \int_{[0,\infty)} c\, e^{\lambda z} \frac{e^{-z}}{1+z^3}\, dz$$

n'est pas escarpée.

2. **(linéarité de la fonction de taux)** *Pour z assez grand, la conjuguée convexe $\Lambda^*(z) = \sup_{\lambda\in\mathbb{R}}\{\lambda z - \Lambda(\lambda)\}$ est linéaire :*

$$\Lambda^*(z) = \Lambda^*(z^*) + z - z^*,$$

Fig. A.1 – log-Laplace associée à P

où $z^* = \Lambda'(1)$.

Preuve. Voir Dembo-Zeitouni, exercice 2.3.17. □

Proposition A.2 (PGD pour une particule) *Soit Z une variable aléatoire de loi P, alors $\frac{Z}{n}$ satisfait un PGD de bonne fonction de taux :*

$$\delta^*(z) = \begin{cases} z & \text{si } z \geq 0, \\ +\infty & \text{sinon.} \end{cases}.$$

Preuve. La majoration ne pose pas de problèmes. Etablissons la minoration.

$$\begin{aligned} P\left\{\frac{Z}{n} \in (z-\epsilon, z+\epsilon)\right\} &= \int_{n(z-\epsilon)}^{n(x+\epsilon)} c\frac{e^{-u}}{1+u^3}\,du \\ &\geq \int_{n(z-\epsilon/k)}^{n(x+\epsilon/k)} c\frac{e^{-u}}{1+u^3}\,du \\ &\geq c\frac{e^{-n(z+\epsilon/k)}}{1+n^3(x+\epsilon/k)^3}2n(\epsilon/k). \end{aligned}$$

Soit

$$\liminf_{n\to\infty} \frac{1}{n} P\left\{\frac{Z}{n} \in (z-\epsilon, z+\epsilon)\right\} \geq -(z-\epsilon/k).$$

Comme l'inégalité précédente est valable pour tout k, il suffit de faire tendre k vers l'infini pour conclure. □

La proposition précédente met en lumière le fait que le comportement d'**une** particule importe dans le cas où la variable considérée n'a pas tous ses moments exponentiels. Bercu, Gamboa et Rouault ont mis en évidence un phénomène similaire dans [6] (théorème 1 et lemme 6).

Pour illustrer cette remarque, on établit ci-dessous la borne inférieure associée au PGD d'une moyenne empirique lorsque la famille (Z_i) est i.i.d. et P-distribuée.

Proposition A.3 *Soit (Z_i) une famille de variables aléatoires i.i.d. et P-distribuée. Soit $z > z^* = \Lambda'(1)$. Alors,*

$$\liminf_{n\to\infty} \frac{1}{n} \log P\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i \in [z-\epsilon, z+\epsilon]\right\} \geq -\Lambda^*(z)$$

L'intérêt réside "au-delà" de z^*, où on ne peut plus effectuer un changement de probabilité exponentielle.
Preuve. Remarquons dans un premier temps que

$$\begin{aligned} n(z-z^*) - n\epsilon/2 &\leq Z_n \leq n(z-z^*) + n\epsilon/2, \\ nz^* - n\epsilon/2 &\leq \sum_{i=1}^{n} Z_i \leq nz^* + n\epsilon/2, \end{aligned}$$

Par conséquent,

$$\begin{aligned} \{X_n/n \in [z-z^*-\epsilon/2, z-z^*+\epsilon/2]\} &\cap \left\{ \frac{1}{n}\sum_1^{n-1} Z_i \in [z^*-\epsilon/2, z^*+\epsilon/2] \right\} \\ &\subset \left\{ \frac{1}{n}\sum_1^{n} Z_i \in [z-\epsilon, z+\epsilon] \right\}. \quad (A.1) \end{aligned}$$

Du fait que $z^* = \Lambda'(1)$, on peut appliquer la technique de changement de probabilité exponentielle à $\frac{1}{n}\sum_1^{n-1} Z_i$ autour de z^* (même calcul que dans l'introduction) :

$$\liminf_{n\to\infty} \frac{1}{n} \log P \left\{ \frac{1}{n}\sum_1^{n-1} Z_i \in [z^*-\epsilon/2, z^*+\epsilon/2] \right\} \geq -\Lambda^*(z^*).$$

D'autre part, du fait que $\frac{Z_n}{n}$ satisfait un PGD (Proposition A.2), on a :

$$\liminf_{n\to\infty} \frac{1}{n} \log P \left\{ X_n/n \in [z-z^*-\epsilon/2, z-z^*+\epsilon/2] \right\} \geq -(z-z^*)$$

Enfin, on conclut en utilisant l'inclusion (A.1) et la linéarité de la fonction de taux (Proposition A.1) :

$$\liminf_{n\to\infty} \frac{1}{n} \log P \left\{ \frac{1}{n}\sum_1^{n} Z_i \in [z-\epsilon, z+\epsilon] \right\} \geq -(\Lambda^*(z^*) + z - z^*) = -\Lambda^*(z).$$

□

Tout se passe comme si la moyenne $\frac{1}{n}\sum_1^{n-1} Z_i$ réalisait une grande déviation autour de z^* et une particule Z_n réalisait un "grand saut" de l'ordre de $n(z-z^*)$. Ce calcul nous a été inspiré par les résultats de Vinogradov [56, 57] concernant les grandes déviations précises. Celui-ci a démontré dans ([57], chap. 5) que si la suite de variables aléatoires réelles et i.i.d. $(Z_n)_{n\in\mathbb{N}}$ a une fonction de répartition satisfaisant

$$1 - F(x) = C_a\, e^{-z_0\, x} x^{-a} + o(e^{-z_0\, x} x^{-s}),$$

où $a > 3$ et $s > a + 1$, alors,

$$P\left\{\frac{1}{n}\sum_{1}^{n} Z_i > u - u_0\right\} \sim n\,P\left\{\frac{Z_1}{n} > u - u_0\right\} e^{-(n-1)\Lambda^*(u_0)},$$

où $u_0 = \Lambda'(z_0)$ et $u > u_0$.