

Deuxième partie

Grandes déviations et analyse
gaussienne

Chapitre 4

Introduction aux principes de grandes déviations

4.1 Introduction

Considérons X_1, \dots, X_n une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans \mathbb{R}, \mathbb{R}^d avec $d \geq 2$ ou, plus généralement, un espace de Banach E . Par la suite, nous noterons par S_n la quantité suivante : $S_n = X_1 + \dots + X_n$ pour $n \geq 1$. Sous l'hypothèse $\mathbb{E}[|X_1|] < \infty$ (il est possible de remplacer la norme euclidienne $|\cdot|$ par n'importe quelle norme $\|\cdot\|$ de \mathbb{R}^d ou de E), la loi forte des grands nombres nous assure que

$$\frac{S_n}{n} \rightarrow \mathbb{E}[X_1] \quad p.s. \quad \text{lorsque } n \rightarrow +\infty$$

En particulier, si F est un ensemble fermé ne contenant pas $\mathbb{E}[X_1]$, nous avons

$$\mathbb{P}\left(\frac{S_n}{n} \in F\right) \rightarrow 0 \quad \text{lorsque } n \rightarrow +\infty$$

La théorie des grandes déviations s'attache à quantifier la vitesse de convergence, vers 0, de cette probabilité. Dans les années 20, il a été mis en évidence par S. Varadhan que cette vitesse de convergence était quantifiée par une fonction dépendant de la loi de la variable aléatoire X_1 . Une première partie de ce chapitre se focalisera sur le Théorème de Cramer ainsi que celui de Gärtner-Ellis

Voici un autre exemple pour lequel une étude de grandes déviations peut-être faite. Considérons (E, d) un espace métrique (éventuellement fini) ainsi que X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées (*i.i.d.* en abrégé) à valeurs dans E .

Observons que, dans un tel cadre, l'absence de structure d'espace vectoriel ne permet pas de donner un sens à la quantité $X_1 + X_2$. A partir de cette famille de variable aléatoires, il est possible de construire, sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, des mesures empiriques de la façon suivante : pour chaque élément $\omega \in \Omega$ nous définissons l'application $L_n : \Omega \rightarrow \mathcal{P}(E)$ par

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

Observons que, à ω fixé, $L_n(\omega)$ est un élément de l'ensemble des mesures de probabilités sur E (noté $\mathcal{P}(E)$). Considérons une fonction $f : E \rightarrow \mathbb{R}$ telle que $\mathbb{E}[|f(X_1)|] < +\infty$. Si μ désigne la loi de X_1 , la loi forte des grands nombres nous assure à nouveau que

$$\int_E f dL_n = \frac{1}{n} \sum_{i=1}^n f(X_i) \rightarrow \mathbb{E}[f(X_1)] = \int_E f d\mu \quad \text{p.s. lorsque } n \rightarrow +\infty \quad (4.1.1)$$

D'un point de vue heuristique, la convergence précédente semble indiquer que, *p.s.* $L_n \Rightarrow \mu$ lorsque $n \rightarrow +\infty$. Nous nous interrogerons à nouveau quant à la vitesse de convergence de L_n vers μ observée dans l'équation (4.1.1). Ceci sera le contenu du Théorème de Sanov dont nous présenterons une version affaiblie.

La méthode de Laplace est une méthode célèbre (cf. [47]) permettant de déterminer un développement asymptotique d'une intégrale dépendant d'un paramètre. Le mathématicien Varadhan a obtenu un résultat analogue dans le cadre des grandes déviations. Bien que ce résultat soit important nous ne l'aborderons pas dans ce cours. La plupart des éléments exposés ci-dessous proviennent des ouvrages [37, 1] auxquels nous renvoyons le lecteur pour plus de détails.

4.2 Théorème de grandes déviations de Cramer dans \mathbb{R}

Par soucis de pédagogie nous débutons par le cas réel qui permettra de mettre en valeur les points clés permettant d'établir un principe de grande déviations dans des cadres plus complexes.

Par la suite, nous considérons une variable aléatoire $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et noterons sa loi μ (en abrégé $\mathcal{L}(X) = \mu$). Etant donnée une telle variable aléatoire nous noterons par $M(\theta)$ sa transformée de Laplace. Celle-ci est définie par

$$M(\theta) = \mathbb{E}[e^{\theta X}] = \int_{\mathbb{R}} e^{\theta x} d\mu(x) \quad \text{avec } \theta \in \mathbb{R};$$

nous noterons ensuite par $\Lambda(\theta) = \log M(\theta)$ le logarithme de la transformée de Laplace. La fonction Λ est usuellement désignée sous l'appellation « log-Laplace ». Observons au passage que $\Lambda(0) = 0$ et que $\Lambda(\theta) \in]-\infty, +\infty]$.

Soient X_1, \dots, X_n des copies i.i.d. de X . Rappelons la notation suivante $S_n = X_1 + \dots + X_n$, avec $n \geq 1$, et désignons par μ_n la loi de $\frac{S_n}{n}$.

Supposons que $\mathbb{E}[|X|] < +\infty$ et considérons $F \subset \mathbb{R}$ un ensemble fermé ne contenant pas $\mathbb{E}[X]$, la loi forte des grands nombres implique que

$$\mu_n(F) = \mathbb{P}\left(\frac{S_n}{n} \in F\right) \rightarrow 0 \quad \text{lorsque } n \rightarrow +\infty.$$

Nous allons chercher à quantifier cette vitesse de convergence à l'aide de la fonction Λ .

4.2.1 Transformée de Fenchel-Legendre (1ère partie)

Il apparaîtra par la suite que la fonction $\Lambda^*(x) = \sup_{\theta \in \mathbb{R}} [\theta x - \Lambda(\theta)]$, appelée transformée de Fenchel-Legendre de la fonction convexe Λ , aura un rôle important dans la démonstration du Théorème de Cramer. Nous dédions les quelques lignes qui vont suivre pour, brièvement, décrire certaines de ses propriétés qui serviront dans la démonstration de la première assertion de ce théorème. Bien que nous l'omettrons par la suite, les fonctions M , Λ et Λ^* dépendent toutes de la loi μ .

Suivant les notations introduites au début de cette section, voici quelques propriétés du logarithme de la transformée de Laplace Λ (de la variable X) et de sa transformée de Fenchel-Legendre Λ^* .

Lemme 4.2.1 (Quelques propriétés des fonctions Λ et Λ^*). 1. Λ est une fonction convexe et $\Lambda(0) = 0$.

2. Λ^* est une fonction convexe positive.

3. Λ^* est une fonction semi-continue inférieurement. Autrement dit, pour tout $\alpha \geq 0$, l'ensemble $\{\Lambda^* \leq \alpha\}$ est fermé.

Démonstration. Le premier point est une conséquence de l'inégalité de Hölder. Le second point est évident puisque Λ^* s'exprime comme le supremum de fonctions affines en θ . Quant à la dernière assertion, \mathbb{R} étant un espace métrique, il est possible de démontrer sa véracité de manière séquentielle. Pour cela il suffit de montrer que

$$\text{si } \lim_{n \rightarrow +\infty} x_n = x \text{ alors } \liminf_{n \rightarrow +\infty} \Lambda^*(x_n) \geq \Lambda^*(x)$$

où, rappelons le, $\liminf_{n \rightarrow +\infty}$ est une abréviation désignant la limite inférieure lorsque $n \rightarrow +\infty$. De manière similaire, $\limsup_{n \rightarrow +\infty}$ désigne la limite supérieure lorsque $n \rightarrow +\infty$. Nous laissons la vérification de ce dernier point en exercice. \square

Remarque. Lorsque Λ est dérivable, le supremum $\sup_{\theta \in \mathbb{R}} [\theta x - \Lambda(\theta)]$ intervenant dans la transformée de Fenchel-Legendre $\Lambda^*(x)$ est atteint.

Avant d'énoncer, puis de démontrer, le Théorème de Cramer, voici quelques exemples de formules explicites pour les fonctions Λ et Λ^* associées à des lois usuelles.

Exemple 4.2.1. 1. (loi normale) Si $X \sim \mathcal{N}(m, \sigma^2)$ avec $m \in \mathbb{R}$ et $\sigma \geq 0$ alors $\Lambda(\theta) = m\theta + \frac{\sigma^2 \theta^2}{2}$, $\theta \in \mathbb{R}$ et $\Lambda^* = \frac{(x-m)^2}{2\sigma^2}$, $x \in \mathbb{R}$.

2. (loi de Poisson) Si $X \sim \mathcal{P}(\lambda)$ avec $\lambda > 0$ alors $\Lambda(\theta) = \lambda(e^\theta - 1)$, $\theta \in \mathbb{R}$ et

$$\Lambda^*(x) = \begin{cases} x \log(x/\lambda) - x + \lambda & \text{si } x \geq 0 \\ +\infty & \text{sinon.} \end{cases}$$

Observons que la fonction Λ^* est discontinue.

3. (loi de Bernoulli) Si $X \sim \mathcal{B}e(p)$ avec $p \in [0, 1]$ alors $\Lambda(\theta) = \log [p(e^\theta - 1) + 1]$, $\theta \in \mathbb{R}$ et

$$\Lambda^*(x) = \begin{cases} x \log(x/p) - x + (1-x) \log [(1-x)/(1-p)] & \text{si } x \in [0, 1] \\ +\infty & \text{sinon.} \end{cases}$$

4. (loi exponentielle) Si $X \sim \mathcal{E}xp(\lambda)$ avec $\lambda > 0$ alors

$$\Lambda(\theta) = \begin{cases} \log [\lambda/(\lambda - \theta)] & \text{si } \theta < \lambda \\ +\infty & \text{sinon.} \end{cases}$$

et

$$\Lambda^*(x) = \begin{cases} \lambda x - 1 - \log(\lambda x) & \text{si } x \geq 0 \\ +\infty & \text{sinon.} \end{cases}$$

Après ces quelques exemples, voici l'énoncé du Théorème de Cramer.

4.2.2 Théorème de Cramer sur \mathbb{R}

Le théorème suivant permet de quantifier la vitesse de convergence de la loi forte des grands nombres.

Théorème 4.2.1 (Cramer). *Dans le cadre précédent, les deux assertions suivantes sont satisfaites :*

1. Pour tout fermé $F \subset \mathbb{R}$, $\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x \in F} \Lambda^*(x) := -\Lambda^*(F)$.
2. Pour tout ouvert $G \subset \mathbb{R}$, $\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_{x \in G} \Lambda^*(x) := -\Lambda^*(G)$.

Remarque. Observons la similarité (vis-à-vis des ensembles et du sens des inégalités) de l'énoncé précédent avec celui du Théorème du portemanteau 2.4.1.

La démonstration de ce Théorème se fera sous les hypothèses supplémentaires suivantes :

- 1.

$$\mathbb{E}[e^{\theta X}] = \int_{\mathbb{R}} e^{\theta x} d\mu(x) < +\infty \quad \text{pour tout } \theta \in \mathbb{R}. \quad (4.2.1)$$

- 2.

$$.\mu \quad \text{ne charge pas les points.} \quad (4.2.2)$$

Au prix d'efforts techniques, ces hypothèses peuvent être affaiblies (cf. [37]). La proposition suivante présente les conséquences de telles suppositions.

Lemme 4.2.2. *Sous les hypothèses précédentes, les assertions suivantes sont vérifiées.*

1. $\mathbb{E}[e^{\theta X}] < +\infty$ pour tout $\theta \geq 0$ et $\mathbb{E}[|X|^p e^{\theta X}] < +\infty$ pour tout $\theta \in \mathbb{R}$ et tout $p \in \mathbb{N}$.
2. L'application $\theta \mapsto \Lambda(\theta)$ est de classe $C^\infty(\mathbb{R})$.

Remarque. Observons que la régularité de la fonction $\theta \mapsto \Lambda(\theta)$ permet d'obtenir les formules suivantes :

$$\Lambda'(\theta) = \int_{\mathbb{R}} x e^{\theta x} \frac{d\mu(x)}{M(\theta)} \quad \text{et} \quad \Lambda''(\theta) = \int_{\mathbb{R}} x^2 e^{\theta x} \frac{d\mu(x)}{M(\theta)} - \left(\int_{\mathbb{R}} x e^{\theta x} \frac{d\mu(x)}{M(\theta)} \right)^2$$

où $M(\theta) = \int_{\mathbb{R}} e^{\theta x} d\mu(x)$.

Par la suite, il sera utile de considérer la famille de mesure exponentielle (par rapport à la mesure μ) définie par

$$d\mu_{\theta}(x) = \frac{e^{\theta x}}{M(\theta)} d\mu(x) \quad \text{avec} \quad \theta \in \mathbb{R}.$$

En particulier, si Z_{θ} admet pour loi μ_{θ} nous avons

$$\Lambda'(\theta) = \mathbb{E}[Z_{\theta}] \quad \text{et} \quad \Lambda''(\theta) = \text{Var}(Z_{\theta})$$

Si $\theta = 0$ nous obtenons une expression de l'espérance et de la variance de X via les dérivées première et seconde (en zéro) de la fonction Λ . Observons de plus que le dernier point rend évident la convexité de $\theta \mapsto \Lambda(\theta)$ puisque $\Lambda'' \geq 0$.

Par ailleurs, si nous supposons que μ ne charge pas les points, nous obtenons que $\Lambda''(\theta) = \text{Var}(Z_{\theta}) > 0$ signifiant que la fonction Λ est strictement convexe.

Enfin, il est facile de montrer que $\Lambda^*(x) = 0$ si et seulement si $x = \mathbb{E}[X]$. En effet, d'une part Λ^* est une fonction positive car, pour tout $x \in \mathbb{R}$, $\Lambda^*(x) \geq 0x - \Lambda(0) = 0$. D'autre part, pour tout $\theta \in \mathbb{R}$, d'après l'inégalité de Jensen, nous avons

$$\Lambda(\theta) = \log \mathbb{E}[e^{\theta X}] \geq \theta \mathbb{E}[X].$$

En conséquence, $\Lambda^*(\mathbb{E}[X]) = \sup_{\theta \in \mathbb{R}} [\theta \mathbb{E}[X] - \Lambda(\theta)] \leq 0$. Le résultat annoncé s'ensuit.

Démonstration. Débutons la démonstration du Théorème de Cramer sous les hypothèses (4.2.1) et (4.2.2). Pour alléger les notations, nous désignerons par m l'espérance de X . En particulier, puisque μ n'est pas une masse de Dirac, $\Lambda^*(m) = 0$.

1. **Démonstration de l'inégalité pour les ensembles fermés :** soit $F \subset \mathbb{R}$ un ensemble fermé, nous devons montrer l'inégalité suivante

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(F) \leq -\Lambda^*(F) := -\inf_{x \in F} \Lambda^*(x).$$

Rappelons que $\mu_n(F) = \mathbb{P}\left(\frac{S_n}{n} \in F\right)$. Soit $n \geq 1$ fixé.

- Supposons que $F = [a, +\infty[$ pour un certain $a > m$. Pour $\theta > 0$ nous avons, d'après l'inégalité de Markov,

$$\begin{aligned}\mu_n(F) = \mu_n([a, +\infty[) &= \mathbb{P}\left(\frac{S_n}{n} \in F\right) \leq e^{-\theta na} \mathbb{E}[e^{\theta S_n}] \\ &= e^{-\theta na} \mathbb{E}[e^{\theta X}]^n\end{aligned}$$

où la dernière égalité provient du fait que les variables aléatoires $(X_i)_{i=1, \dots, n}$ composant S_n sont indépendantes et identiquement distribuées. En reformulant l'inégalité précédente, nous avons donc montré que, pour tout $\theta \geq 0$ (le cas $\theta = 0$ étant trivial)

$$\mu_n(F) \leq e^{-n[\theta a - \Lambda(\theta)]} \iff \frac{1}{n} \log \mu_n(F) \leq -[\theta a - \Lambda(\theta)].$$

Ainsi, $\frac{1}{n} \log \mu_n(F) \leq -\sup_{\theta \geq 0} [\theta a - \Lambda(\theta)]$. Il n'est pas utile de traiter le cas de figure $\theta < 0$ car

$$\sup_{\theta \in \mathbb{R}} [\theta a - \Lambda(\theta)] = \sup_{\theta \geq 0} [\theta a - \Lambda(\theta)].$$

En effet, si $\theta < 0$, nous avons

$$\theta a - \Lambda(\theta) \leq \theta m - \Lambda(\theta) \leq \Lambda^*(m) = 0$$

Finalement, nous venons donc de prouver que, pour tout $n \geq 1$,

$$\frac{1}{n} \log \mu_n(F) \leq -\Lambda^*(a) \leq -\Lambda^*(F)$$

Supposons à présent que $F =]-\infty, a]$ avec $a < m$. De manière similaire, en appliquant ce qui précède à la variable $-X$ avec $-m$, nous obtenons à nouveau que

$$\frac{1}{n} \log \mu_n(F) \leq -\Lambda^*(a) \leq -\Lambda^*(F)$$

En résumé, nous venons de démontrer la première assertion du Théorème lorsque F est une **demi droite (fermée)**. Voyons de quelle manière ce résultat nous permet de démontrer le cas général.

- Soit $F \subset \mathbb{R}$ un ensemble fermé (non vide). Pour démontrer

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x \in F} \Lambda^*(x),$$

nous pouvons supposer que $\Lambda^*(F) > 0$ (autrement l'inégalité est triviale). Si c'est le cas cela signifie que $m \notin F$ (rappelons que $\Lambda^*(x) > 0$ si et seulement si $x \neq m$). Plusieurs cas de figures se proposent alors à nous.

- **(1er cas)** : $F \subset]m, +\infty[$. Puisque F est un ensemble fermé, il est possible de définir $a \in F$ tel que $d(m, F) = d(m, a)$. En particulier, $F \subset [a, +\infty[$. En conséquence, selon supra, pour tout $n \geq 1$ nous avons

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(F) &\leq \frac{1}{n} \log \mu_n([a, +\infty[) \\ &\leq -\Lambda^*(a) \leq -\Lambda^*(F) \end{aligned}$$

puisque $a \in F$.

- **(2ème cas)** : $F \subset]-\infty, m[$. Par symétrie il est possible d'appliquer le point précédent.
- **(3ème cas)** : $F \subset]-\infty, a_1] \cup [a_2, +\infty[$ avec $a_1 < m < a_2$ et $a_1, a_2 \in F$. Pour tout $n \geq 1$, nous avons les inégalités suivantes :

$$\begin{aligned} \mu_n(F) &\leq \mu_n(]-\infty, a_1] \cup [a_2, +\infty[) \\ &\leq \mu_n(]-\infty, a_1]) + \mu_n([a_2, +\infty[) \\ &\leq e^{-n\Lambda^*(a_1)} + e^{-n\Lambda^*(a_2)} \\ &\leq 2e^{-n\Lambda^*(F)} \quad (\text{puisque } a_1, a_2 \in F) \end{aligned}$$

Ainsi,

$$\frac{1}{n} \log \mu_n(F) \leq \frac{1}{n} \log(2) - \Lambda^*(F)$$

et donc $\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(F) \leq -\Lambda^*(F)$.

2. Tournons nous vers la deuxième partie du Théorème portant sur les ensembles ouverts. Pour tout ensemble $G \subset \mathbb{R}$ ouvert, nous souhaitons montrer que

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(G) \geq -\Lambda^*(G)$$

- (Preliminaires) Quelques détails techniques sont nécessaires avant d'entamer la démonstration de ce résultat. Faisons l'hypothèse supplémentaire que le support de μ soit \mathbb{R} tout entier. Autrement dit, pour tout $a > 0$,

$$\min \left(\mu([a, +\infty[), \mu(]-\infty, a]) \right) > 0 \quad (4.2.3)$$

Lemme 4.2.3. *Sous l'hypothèse (4.2.3), les assertions suivantes sont satisfaites :*

(a) $\lim_{\theta \rightarrow \pm\infty} \frac{\Lambda(\theta)}{\theta} = \pm\infty$.

(b) $\Lambda' : \mathbb{R} \rightarrow \mathbb{R}$ est une bijection continue.

(c) L'application réciproque de Λ' est $(\Lambda^*)'$.

Démonstration. Débutons par la première assertion et montrons que l'hypothèse (4.2.3) entraîne que $\lim_{\theta \rightarrow \pm\infty} \frac{\Lambda(\theta)}{\theta} = \pm\infty$. En effet, pour tout $a > 0$ et tout $\theta > 0$, nous avons

$$\begin{aligned} \frac{\Lambda(\theta)}{\theta} &= \frac{1}{\theta} \log \int_{\mathbb{R}} e^{\theta x} d\mu(x) \\ &\geq \frac{1}{\theta} \log \int_a^{+\infty} e^{\theta x} d\mu(x) \\ &\geq a + \frac{1}{\theta} \log \mu([a, +\infty[) \end{aligned}$$

Ainsi, puisque $\mu([a, +\infty[) > 0$, nous en déduisons que $\liminf_{\theta \rightarrow +\infty} \frac{\Lambda(\theta)}{\theta} \geq a$ pour tout $a > 0$. Ceci entraîne que

$$\lim_{\theta \rightarrow +\infty} \frac{\Lambda(\theta)}{\theta} = +\infty. \quad (4.2.4)$$

Par symétrie, en remplaçant θ par $-\theta$, nous avons aussi

$$\lim_{\theta \rightarrow -\infty} \frac{\Lambda(\theta)}{\theta} = -\infty. \quad (4.2.5)$$

A présent, attachons nous à la démonstration du deuxième point et montrons que Λ' est une bijection continue de \mathbb{R} dans \mathbb{R} . A cet effet, rappelons que Λ est convexe, $\Lambda(0) = 0$ et Λ est dérivable. Ainsi, par convexité,

$$\Lambda(\theta) - \Lambda(\theta') \leq (\theta - \theta')\Lambda'(\theta)$$

En particulier, $\Lambda(\theta) \leq \theta\Lambda'(\theta)$. Donc, à l'aide de (4.2.4), nous obtenons que $\lim_{\theta \rightarrow +\infty} \Lambda'(\theta) = +\infty$. La limite de Λ' est $-\infty$ est obtenu similairement en remplaçant θ par $-\theta$.

En résumé, nous avons Λ une fonction strictement convexe (puisque μ n'est pas une masse de Dirac) et Λ' est une fonction continue strictement croissante de \mathbb{R} dans \mathbb{R} . Ceci entraîne que Λ' est une bijection continue de \mathbb{R} dans \mathbb{R} .

Voyons ce qu'il est possible de préciser sur la fonction réciproque de Λ' .

Fixons $x \in \mathbb{R}$ et considérons $\alpha_x : \theta \mapsto \theta x - \Lambda(\theta)$. Il s'agit d'une application dérivable et strictement concave. De plus, en utilisant les limites (4.2.4) et (4.2.5) de la fonction $\theta \mapsto \frac{\Lambda(\theta)}{\theta}$, nous en déduisons que $\lim_{\theta \rightarrow \pm\infty} \alpha_x(\theta) = -\infty$. C'est pourquoi α_x admet un unique maximum que nous noterons $\tilde{\theta}_x$.

En particulier, l'observation précédente combinée à la définition de Λ^* nous assure que $\Lambda^*(x) = \tilde{\theta}_x x - \Lambda(\tilde{\theta}_x)$. De plus, en utilisant le fait que $\tilde{\theta}_x$ est le point pour lequel la fonction α_x atteint son maximum, nous avons

$$\left. \frac{d}{d\theta} [\alpha_x(\theta)] \right|_{\theta=\tilde{\theta}_x} = \left. \frac{d}{d\theta} [\theta x - \Lambda(\theta)] \right|_{\theta=\tilde{\theta}_x} = 0$$

Ce qui entraîne l'identité suivante : $\Lambda'(\tilde{\theta}_x) = x$. Autrement dit, $x \mapsto \tilde{\theta}_x$ est l'application réciproque de $\theta \mapsto \Lambda'(\theta)$.

En outre, en utilisant la relation $\Lambda^*(x) = \tilde{\theta}_x x - \Lambda(\tilde{\theta}_x)$ nous en déduisons que

$$(\Lambda^*)'(x) = \tilde{\theta}'_x x + \tilde{\theta}_x - \Lambda'(\tilde{\theta}_x)\tilde{\theta}'_x = \tilde{\theta}_x$$

puisque, par définition de l'application réciproque, $\Lambda'(\tilde{\theta}_x) = x$.

En résumé, nous venons de montrer que Λ' est une bijection dont l'inverse est $(\Lambda^*)'$. Il est également possible de montrer que Λ^* est strictement convexe et s'annule uniquement au point $m = \mathbb{E}[X]$. \square

Après cet intermède, nous pouvons achever la démonstration du Théorème de Cramer.

- **(1ère étape)** Démontrons à présent que, pour tout ouvert $G \subset \mathbb{R}$, nous avons

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(G) \geq -\Lambda^*(G).$$

Observons qu'il suffit pour cela de montrer que, pour tout $x_0 \in \mathbb{R}$ et tout $\delta > 0$, l'inégalité suivante est satisfaite :

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(] - x_0 - \delta, x_0 + \delta]) \geq -\Lambda^*(x_0) \quad (4.2.6)$$

En effet, puisque $] - x_0 - \delta, x_0 + \delta[\subset G$, nous avons

$$\mu_n(G) \geq \mu_n(] - x_0 - \delta, x_0 + \delta])$$

et donc, si (4.2.6) est vérifiée,

$$\begin{aligned} \liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(G) &\geq \liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(] - x_0 - \delta, x_0 + \delta]) \\ &\geq -\Lambda^*(x_0). \end{aligned}$$

Il suffit ensuite d'optimiser en $x_0 \in G$ pour conclure.

Attachons nous donc à démontrer l'inégalité (4.2.6). Fixons $x_0 \in \mathbb{R}$ et désignons par $\theta_0 = \tilde{\theta}_{x_0}$ l'unique réel tel que $\Lambda^*(x_0) = \theta_0 x_0 - \Lambda(\theta_0)$. L'argument clé de cette partie de la démonstration est de faire un changement de probabilité en considérant la famille de mesure exponentielle suivante

$$\mu \mapsto d\mu_{\theta_0}(x) = \frac{e^{\theta_0 x}}{M(\theta_0)} d\mu(x)$$

Ceci fournit alors $\Lambda_{\mu_{\theta_0}}(\tau) = \Lambda_{\mu}(\tau + \theta_0) - \Lambda_{\mu}(\theta_0)$, avec $\tau \in \mathbb{R}$. De plus, si $Z = Z_{\theta}$ est une variable aléatoire de loi μ_{θ_0} alors $\mathbb{E}[Z] = \Lambda'(\theta_0) = \Lambda'(\tilde{\theta}_{x_0}) = x_0$ par choix du réel $\tilde{\theta}_{x_0}$ (rappelons que, d'après le Lemme précédent, $\tilde{\theta} = (\Lambda')^{-1}$).

En particulier, si Z_1, \dots, Z_n sont des copies indépendantes de Z , la loi forte des grands nombres implique que

$$\frac{1}{n}T_n = \frac{1}{n}(Z_1 + \dots, Z_n) \rightarrow \mathbb{E}[Z_1] = x_0 \quad p.s. \quad \text{lorsque } n \rightarrow +\infty$$

Supposons que $\theta_0 \geq 0$ et introduisons un paramètre $\eta \in \mathbb{R}$ tel que $0 < \eta < \delta$. Pour tout $n \geq 1$, nous avons

$$\mu_n([x_0 - \delta, x_0 + \delta]) \geq \mu_n([x_0 - \eta, x_0 + \eta]) = \mathbb{P}\left(\frac{S_n}{n} \in [x_0 - \eta, x_0 + \eta]\right)$$

Réécrivons le membre de gauche de l'inégalité précédente sous forme intégrale pour obtenir une expression équivalente sous la loi μ_{θ_0} .

$$\mu_n([x_0 - \delta, x_0 + \delta]) = \int_{A_{\delta, n, x_0}} d\mu^{\otimes n}(x) = \int_{A_{\delta, n, x_0}} e^{-\theta_0 \sum_{i=1}^n x_i + n\Lambda(\theta_0)} d\mu_{\theta_0}^{\otimes n}(x)$$

avec $A_{\delta, n, x_0} = \left\{x \in \mathbb{R}^n; \frac{1}{n} \sum_{i=1}^n x_i \in [x_0 - \delta, x_0 + \delta]\right\}$ où $x = (x_1, \dots, x_n)$.

Par choix de η , observons que $A_{\eta, n, x_0} \subset A_{\delta, n, x_0}$. De plus, l'implication suivante est satisfaite

$$\text{si } \sum_{i=1}^n x_i \in [n(x_0 - \eta), n(x_0 + \eta)] \quad \text{alors } \sum_{i=1}^n x_i \leq n(x_0 + \eta).$$

D'où,

$$\begin{aligned} \mu_n([x_0 - \delta, x_0 + \delta]) &\geq \int_{A_{\eta, n, x_0}} e^{-n[\theta_0(x_0 + \eta) - \Lambda(\theta_0)]} d\mu_{\theta_0}^{\otimes n}(x) \\ &= e^{-n[\theta_0(x_0 + \eta) - \Lambda(\theta_0)]} \mu_{\theta_0}^{\otimes n}(A_{\eta, n, x_0}) \\ &= e^{-n[\theta_0(x_0 + \eta) - \Lambda(\theta_0)]} \mu_{\theta_0}^{\otimes n}\left(\frac{1}{n} \sum_{i=1}^n x_i \in [x_0 - \eta, x_0 + \eta]\right) \end{aligned}$$

En passant au logarithme et en utilisant la définition de θ_0 (i.e. $\theta_0 x_0 - \Lambda(\theta_0) = \Lambda^*(x_0)$), nous en déduisons que

$$\begin{aligned} \frac{1}{n} \log \mu_n([x_0 - \delta, x_0 + \delta]) &\geq \Lambda(\theta_0) - \theta_0(x_0 + \eta) + \frac{1}{n} \log \mu_{\theta_0}^{\otimes n}\left(\frac{1}{n} \sum_{i=1}^n x_i \in [x_0 - \eta, x_0 + \eta]\right) \\ &\geq -\Lambda^*(x_0) - \theta_0 \eta + \frac{1}{n} \log \mathbb{P}\left(\frac{Z_1 + \dots + Z_n}{n} \in [x_0 - \eta, x_0 + \eta]\right) \end{aligned}$$

En outre, rappelons que

$$\frac{1}{n}(Z_1 + \dots, Z_n) \rightarrow x_0 \quad p.s. \quad \text{lorsque } n \rightarrow +\infty$$

En conséquence, $\mathbb{P}\left(\frac{Z_1+\dots+Z_n}{n} \in]x_0 - \eta, x_0 + \eta[\right) \rightarrow 1 > 0$ lorsque $n \rightarrow +\infty$. Nous en déduisons alors que

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(]x_0 - \delta, x_0 + \delta[) \geq -\Lambda^*(x_0) - \theta_0 \eta$$

Il suffit ensuite de faire tendre η vers 0 pour conclure.

- **(2ième étape)** : Nous allons voir qu'il est possible de supprimer l'hypothèse (4.2.3) portant sur le support de la mesure μ . A cet effet, nous allons procéder à un argument de régularisation par convolution par un noyau gaussien.

Considérons une variable aléatoire gaussienne $G \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 > 0$ indépendante de X et désignons par G_1, \dots, G_N des copies indépendantes de G (également indépendantes de X).

Posons $Y = X + G$ et notons par ν la loi de Y . Observons ensuite que ν satisfait l'hypothèse de support de l'étape précédente (grâce aux propriétés de la variable aléatoire G). De plus, par indépendance,

$$\mathbb{E}[e^{\theta Y}] = \mathbb{E}[e^{\theta X}] \mathbb{E}[e^{\theta G}] = \mathbb{E}[e^{\theta X}] e^{\frac{\sigma^2 \theta^2}{2}} \quad \forall \theta \in \mathbb{R}$$

C'est pourquoi

$$\Lambda_\nu \geq \Lambda_\mu \quad \text{et} \quad \Lambda_\nu^* \leq \Lambda_\mu^*$$

Donc, en utilisant la première étape de la preuve, pour tout $x_0 \in \mathbb{R}$ et tout $\delta > 0$, nous avons

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n Y_i \in]x_0 - \delta, x_0 + \delta[\right) \geq -\Lambda_\nu^*(x_0) \geq -\Lambda_\mu^*(x_0) \quad (4.2.7)$$

avec $Y_i = X_i + G_i$ pour $i = 1, \dots, n$. De plus, pour tout $n \geq 1$,

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i \in]x_0 - 2\delta, x_0 + 2\delta[\right) \geq \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n (X_i + G_i) \in]x_0 - \delta, x_0 + \delta[\right) - \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n G_i\right| \geq \delta \right)$$

En effet, il suffit d'utiliser l'inégalité suivante $\mathbb{P}(A) \leq \mathbb{P}(A \cap B^c) + \mathbb{P}(B)$ avec les évènements

$$A = \left\{ \frac{1}{n} \sum_{i=1}^n (X_i + G_i) \in]x_0 - \delta, x_0 + \delta[\right\} \quad \text{et} \quad B = \left\{ \left| \frac{1}{n} \sum_{i=1}^n G_i \right| \geq \delta \right\}$$

En combinant cette remarque avec la minoration (4.2.7), nous obtenons que, pour n suffisamment grand et pour tout $\epsilon > 0$, nous avons

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i \in]x_0 - 2\delta, x_0 + 2\delta[\right) \geq e^{-n[\Lambda^*(x_0) + \epsilon]} - \mathbb{P}(|G_1| \geq \delta\sqrt{n})$$

Or, $\mathbb{P}(|G_1| \geq t) \leq e^{-t^2/4\sigma^2}$ pour tout t suffisamment grand. Nous en déduisons que

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \in]x_0 - 2\delta, x_0 + 2\delta[\right) \geq e^{-n[\Lambda^*(x_0) + \epsilon]} \times \left(1 - e^{-n\left[\frac{\delta^2}{4\sigma^2} - \Lambda^*(x_0) - \epsilon\right]}\right).$$

Il suffit à présent de choisir, σ suffisamment proche de 0 de sorte que $\frac{\delta^2}{2\sigma^2} > \Lambda^*(x_0) + \epsilon$. un tel choix de σ entraîne alors

$$\lim_{n \rightarrow +\infty} e^{-n\left[\frac{\delta^2}{2\sigma^2} - \Lambda^*(x_0) - \epsilon\right]} = 0$$

Finalement, nous obtenons donc

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \in]x_0 - 2\delta, x_0 + 2\delta[\right) \geq -\Lambda^*(x_0) - \epsilon$$

et il ne reste plus qu'à faire tendre ϵ vers 0. Ce qui conclut la démonstration du Théorème de Cramer.

□

Dans la section suivante nous allons voir de quelle manière le Théorème de Cramer s'étend au cas multidimensionnel. Avant cela, nous allons proposer une application du théorème de Cramer en statistiques.

4.2.3 Application en statistiques

Pour simplifier, supposons que nous ayons à disposition un échantillon (Y_1, \dots, Y_n) tel que $\mathcal{L}(Y_1) = \mu$ une mesure de probabilité (inconnue) admettant une densité, strictement positive, par rapport à la mesure de Lebesgue. Etant donné deux mesures de probabilités μ_0 et μ_1 (admettant également une mesure de densité par rapport à la mesure de Lebesgue), il n'est pas rare de faire face au problème de test suivant :

$$H_0 : \mu = \mu_0 \quad \text{ou} \quad H_1 : \mu = \mu_1$$

Les statisticiens ont développé de nombreuses méthodes permettant d'obtenir une règle de décision (à partir de l'échantillon (Y_1, \dots, Y_n)) permettant de rejeter ou non l'hypothèse H_0 . Ces méthodes consistent à construire une fonction T telle que

- si $T = 1$ alors l'hypothèse H_0 est rejetée,
- si $T = 0$ alors l'hypothèse H_0 n'est pas rejetée.

Ce genre de décision n'est pas exempt d'erreur. Nous parlons alors d'erreur de 1ère espèce lorsque H_0 est rejetée à tort et d'erreur de 2ème espèce lorsque H_0 n'est pas rejetée alors qu'elle devrait l'être. Pour quantifier la qualité du test, il faut calculer le niveau du test $= \mu_0(T = 1)$ et $\mu_1(T = 0)$ la puissance du test.

Dans ce cadre, le test du rapport de vraisemblance est une règle de décision bien connue. Ce test consiste à choisir un niveau α pour déterminer un seuil γ afin de poser

$$T = 1_{\{V_1 > \gamma V_0\}}$$

où V_i (pour $i = 1, 2$) désigne la densité de l'échantillon (Y_1, \dots, Y_n) sous la loi μ_1 . Autrement dit, le test rejette l'hypothèse H_0 lorsque le rapport de vraisemblance $\frac{V_1}{V_0}$ dépasse un seuil γ .

Par ailleurs, il est bien connu (toujours dans le cadre décrit plus haut) que ce genre de test, de niveau α , est sans biais et est le plus puissant parmi les tests sans biais de niveau au plus α : il s'agit du Lemme de Neyman-Pearson. Pour plus de détails à ce sujet, nous renvoyons le lecteur vers le formidable cours de Boucheron *et al.* [25].

Nous allons présenter une application du théorème de Cramer permettant d'estimer le niveau et la puissance du test T lorsque le seuil $\tilde{\gamma}$ est choisi dans un certain intervalle. A cet effet, posons

$$X_j = \log \frac{V_1}{V_0}(Y_j), \quad \bar{x}_0 = \mathbb{E}_{\mu_0}[X_1] \quad \text{et} \quad \bar{x}_1 = \mathbb{E}_{\mu_1}[X_1]$$

Par convention, \mathbb{E}_μ indique que l'espérance doit être calculée par rapport à la mesure μ . Avec ces notations, il est possible de montrer que le test de Neyman-Pearson repose sur le fait que le rapport de log-vraisemblance $\frac{1}{n} \sum_{j=1}^n X_j$ dépasse ou non un certain seuil $\tilde{\gamma}$. Il est alors possible de montrer (cf. [37]) le fait suivant : si $\tilde{\gamma} \in]\bar{x}_0, \bar{x}_1[$ alors

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log \mu_0(T = 1) = -\Lambda_0^*(\tilde{\gamma}) < 0 \quad \text{et} \quad \lim_{n \rightarrow +\infty} \frac{1}{n} \log \mu_1(T = 0) = \tilde{\gamma} - \Lambda_0^*(\tilde{\gamma}) < 0$$

où Λ_0^* désigne la transformée de Fenchel-Legendre de $\Lambda_0(\lambda) = \log \mathbb{E}_{\mu_0}[e^{\lambda X_1}]$.

4.3 Théorème de Cramer dans \mathbb{R}^d avec $d \geq 1$

La section précédente a servi à introduire quelques idées essentielles de la théorie des grandes déviations. Cependant, la démonstration du Théorème de Cramer reposait énormément sur la topologie de \mathbb{R} . Nous allons voir de quelle manière nous pouvons modifier la démonstration de ce résultat pour obtenir un Théorème analogue en dimension $d \geq 1$.

Tout d'abord, voici quelques notations. Dans ce qui va suivre X désigne une variable aléatoire

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \quad \text{de loi} \quad \mu = \mathcal{L}(X).$$

A nouveau, $\Lambda(\theta) = \log M(\theta) = \log \mathbb{E}[e^{\langle \theta, X \rangle}]$ avec $\theta \in \mathbb{R}^d$, désignera le logarithme de la transformée de Laplace de X . De manière analogue, pour tout $x \in \mathbb{R}^d$, la transformée de Fenchel-Legendre de Λ sera notée

$$\Lambda^*(x) = \sup_{\theta \in \mathbb{R}^d} [\langle \theta, x \rangle - \Lambda(\theta)]$$

La suite X_1, \dots, X_n désignera des copies indépendantes de la variable X et nous noterons la suite des sommes partielles par

$$S_n = X_1 + \dots, X_n \quad \text{et désignerons par } \mu_n = \mathcal{L}\left(\frac{S_n}{n}\right) \quad \text{pour } n \geq 1.$$

Voici, l'énoncé du Théorème de Cramer en dimension d .

Théorème 4.3.1 (Cramer). *Supposons que $\Lambda(\theta) < +\infty$ pour tout $\theta \in \mathbb{R}^d$. Alors*

1. *Pour tout fermé $F \subset \mathbb{R}^d$, $\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(F) \leq -\Lambda^*(F)$.*
2. *Pour tout ouvert $G \subset \mathbb{R}^d$, $\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(G) \geq -\Lambda^*(G)$.*

Remarque. 1. Au niveau de la terminologie, il est d'usage de dire que la suite $(\mu_n)_{n \geq 0}$ satisfait un principe de grandes déviations (P.G.D. en abrégé) avec Λ^* pour fonction de taux.

2. Notons par $D_\Lambda = \{\theta \in \mathbb{R}^d, \Lambda(\theta) < +\infty\}$ le domaine de l'application Λ . Dans l'énoncé nous supposons que $D_\Lambda = \mathbb{R}^d$, cette hypothèse peut-être affaiblie en $0 \in D_\Lambda$ (cf. [37]).

4.3.1 Transformée de Fenchel-Legendre (2ième partie)

Certaines propriétés des fonctions Λ et Λ^* nous seront utiles par la suite et correspondent à une extensions des résultats obtenues dans la section précédente en dimension un.

Proposition 4.3.1. *Sous l'hypothèse $D_\Lambda = \mathbb{R}^d$, les assertions suivantes sont satisfaites*

1. *Λ est une application convexe et différentiable.*
2. *Λ^* est une application convexe et positive. De plus, il s'agit d'une application semi-continue inférieurement et les ensembles $\{\Lambda^* \leq a\}$, $a \geq 0$ sont des compacts de \mathbb{R}^d .*
3. *Si $x = \nabla \Lambda(\theta)$ alors $\Lambda^*(x) = \langle \theta, x \rangle - \Lambda(\theta)$.*

Démonstration. Voici les arguments permettant de démontrer la proposition ci-dessus.

1. L'inégalité de Hölder permet de justifier la convexité de l'application Λ . La différentiabilité de cette même fonction découle du Théorème de dérivation sous l'intégrale de Lebesgue.
2. Puisque $\langle 0, x \rangle - \Lambda(0) = 0$ pour tout $x \in \mathbb{R}^d$, il est évident que $\Lambda^*(x) \geq 0$. Il s'agit d'une fonction convexe puisque Λ^* est définie comme le supremum de fonctions affines. Soit à présent $(x_n)_{n \geq 1}$ une suite convergente de \mathbb{R}^d et montrons que

$$\liminf_{n \rightarrow +\infty} \Lambda^*(x_n) \geq \Lambda^*(x) \quad \text{où} \quad \lim_{n \rightarrow +\infty} x_n = x.$$

Ceci justifiera la propriété de semi-continuité (inférieurement) de l'application Λ^* . Par continuité de l'application $y \mapsto \langle \theta, y \rangle - \Lambda(\theta)$, $\theta \in \mathbb{R}^d$, nous avons

$$\liminf_{n \rightarrow +\infty} \Lambda^*(x_n) \geq \liminf_{n \rightarrow +\infty} [\langle x_n, \theta \rangle - \Lambda(\theta)] = \langle x, \theta \rangle - \Lambda(\theta)$$

Il suffit ensuite de prendre le supremum en θ pour conclure.

D'un point de vu topologique, une manière équivalente de décrire la propriété de semi-continuité inférieure de la fonction Λ^* est de dire que les ensembles $\{\Lambda^* \leq a\}$, avec $a \geq 0$, sont fermés. Ainsi, il nous suffit de prouver que ces ensembles sont bornés pour démontrer qu'il s'agit de compacts de \mathbb{R}^d .

Observons alors que, pour tout $x \in \mathbb{R}^d$, $x \neq 0$,

$$\Lambda^*(x) = \sup_{\theta \in \mathbb{R}^d} [\langle \theta, x \rangle - \Lambda(\theta)] \geq \|x\|_2 - \Lambda\left(\frac{x}{\|x\|_2}\right) \geq \|x\|_2 - \sup_{\|\theta\|_2=1} \Lambda(\theta) \quad (4.3.1)$$

où la première minoration est obtenue en choisissant $\theta = \frac{x}{\|x\|_2}$. De plus, l'inégalité reste vraie si $x = 0$. Alors, pour tout $a \geq 0$,

$$\{\Lambda^* \geq a\} \subset \left\{ x \in \mathbb{R}^d, \quad \|x\|_2 \leq a + \sup_{\|\theta\|_2=1} \Lambda(\theta) \right\} := A$$

Puisque $\theta \mapsto \Lambda(\theta)$ est une fonction continue et $\|\theta\|_2 = 1$ un ensemble compact, le supremum apparaissant dans l'ensemble A est alors atteint. En conséquence, l'ensemble A est donc borné. Par suite, $\{\Lambda^* \leq a\}$ est un ensemble compact.

3. Soit $x \in \mathbb{R}^d$ de la forme $x = \nabla \Lambda(\theta)$, $\theta \in \mathbb{R}^d$. Fixons $\lambda \in \mathbb{R}^d$ et considérons la fonction $\phi : \mathbb{R} \rightarrow \mathbb{R}$ définie par

$$\phi(\alpha) = \langle \alpha\lambda + (1-\alpha)\theta, x \rangle - \Lambda(\alpha\lambda + (1-\alpha)\theta), \quad \alpha \in]0, 1[$$

Il s'agit d'une fonction concave et $|\phi(0)| < +\infty$ puisque, par hypothèse, $D_\Lambda = \mathbb{R}^d$. Ainsi, par concavité, l'inégalité suivante est satisfaite :

$$\phi(1) - \phi(0) \leq \phi'(0)(1-0).$$

Autrement dit,

$$\langle \lambda, x \rangle - \Lambda(\lambda) - [\langle \theta, x \rangle - \Lambda(\theta)] \leq \langle \lambda - \theta, x - \nabla \Lambda(\theta) \rangle = 0$$

car $x = \nabla \Lambda(\theta)$ par hypothèse. Donc nous avons finalement montré que, pour tout $\lambda \in \mathbb{R}^d$,

$$\langle \lambda, x \rangle - \Lambda(\lambda) \leq \langle \theta, x \rangle - \Lambda(\theta) \leq \sup_{\theta \in \mathbb{R}^d} [\langle \theta, x \rangle - \Lambda(\theta)] = \Lambda^*(x).$$

En prenant le supremum en $\lambda \in \mathbb{R}^d$, nous en déduisons que

$$\Lambda^*(x) \leq \langle \theta, x \rangle - \Lambda(\theta) \leq \Lambda^*(x)$$

□

Ces propriétés des fonctions Λ et Λ^* étant établies, nous pouvons débiter la démonstration du Théorème de Cramer dans \mathbb{R}^d . Nous allons séparer la démonstration en deux parties : l'une concernant l'inégalité sur les ensembles fermés, l'autre sur les ensembles ouverts.

Démonstration. 1. Soit $F \subset \mathbb{R}^d$ un fermé, nous souhaitons montrer que

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(F) \leq -\Lambda^*(F)$$

- **(1ère étape).** Supposons, dans un premier temps, que l'ensemble F est **compact**. Fixons $\delta > 0$ et, pour tout $x \in F \subset \mathbb{R}^d$, choisissons $\theta_x \in \mathbb{R}^d$ tel que

$$\langle \theta_x, x \rangle - \Lambda(\theta_x) \geq \min \left(\Lambda^*(x) - \delta, \frac{1}{\delta} \right) := \Lambda_\delta(x).$$

L'introduction de cette nouvelle fonction Λ_δ permet d'incorporer le cas de figure $\Lambda^*(x) = +\infty$ dans la démonstration. Fixons également $\rho_x > 0$ tel que $\rho_x \|\theta_x\|_2 \leq \delta$ et considérons $B(x, \rho_x)$ la boule ouverte (pour la norme euclidienne) de centre x et de rayon $\rho_x > 0$. Par la suite, ces boules serviront à recouvrir par l'ensemble compact F .

A présent, pour tout $n \geq 1$, tout $\theta \in \mathbb{R}^d$ et tout $A \subset \mathbb{R}^d$, observons que l'inégalité suivante est satisfaite :

$$\mu_n(A) = \mathbb{P} \left(\frac{S_n}{n} \in A \right) \leq e^{-\inf_{y \in A} \langle \theta, y \rangle} \mathbb{E} \left[e^{\langle \theta, \frac{S_n}{n} \rangle} \right]$$

Il s'agit d'une adaptation de l'inégalité classique de Markov et se démontre via l'inégalité suivante :

$$1_{\{Z \in A\}} \leq e^{-\inf_{y \in A} \langle \theta, y \rangle} e^{\langle \theta, Z \rangle} \quad \text{pour tout } \theta \in \mathbb{R}^d.$$

En appliquant ceci à l'ensemble $A = B(x, \rho_x)$, avec $x \in F$, et $\theta = n\theta_x$ nous obtenons

$$\mu_n(B(x, \rho_x)) \leq e^{-\inf_{y \in B(x, \rho_x)} n \langle \theta_x, y \rangle} \mathbb{E} \left[e^{n \langle \theta_x, \frac{S_n}{n} \rangle} \right]$$

Nous en déduisons ensuite, puisque les variables X_1, \dots, X_n sont i.i.d., que

$$\mathbb{E} \left[e^{n \langle \theta_x, \frac{S_n}{n} \rangle} \right] = \mathbb{E} \left[e^{\langle \theta_x, S_n \rangle} \right] = \mathbb{E} \left[e^{\langle \theta_x, X_1 \rangle} \right]^n = e^{n \Lambda(\theta_x)}$$

En outre, pour tout $y \in B(x, \rho_x)$, par définition de ρ_x , l'inégalité suivante est vérifiée :

$$\langle \theta_x, x - y \rangle \leq \|\theta_x\|_2 \|x - y\|_2 \leq \|\theta_x\|_2 \rho_x \leq \delta.$$

Ce qui permet d'obtenir,

$$\begin{aligned} \inf_{y \in B(x, \rho_x)} \langle \theta_x, y \rangle &\geq \langle \theta_x, x \rangle - \rho_x \|\theta_x\|_2 \\ &\geq \langle \theta_x, x \rangle - \delta. \end{aligned}$$

Pour résumer, nous avons obtenu, pour tout $n \geq 1$,

$$\mu_n(B(x, \rho_x)) \leq e^{-n(\langle \theta_x, x \rangle - \Lambda(\theta_x)) - \delta}$$

Il est évident que l'ensemble F peut-être recouvert par l'ensemble $\cup_{x \in F} B(x, \rho_x)$. C'est pourquoi, puisque F est un ensemble compact, il est possible d'extraire un sous-recouvrement fini ; c'est à dire : il existe $N \geq 1$ et $x_1, \dots, x_N \in F$ tels que

$$F \subset \cup_{i=1}^N B(x_i, \rho_{x_i})$$

Ainsi, en combinant les arguments précédents, nous avons

$$\begin{aligned} \mu_n(F) &\leq \sum_{i=1}^N \mu_n(B(x_i, \rho_{x_i})) \\ &\leq N e^{-n \left(\min_{i=1, \dots, N} [\langle \theta_{x_i}, x_i \rangle - \Lambda(x_i)] - \delta \right)} \end{aligned}$$

et donc

$$\begin{aligned} \frac{1}{n} \log \mu_n(F) &\leq \frac{\log N}{n} - \min_{i=1, \dots, N} [\langle \theta_{x_i}, x_i \rangle - \Lambda(x_i)] + \delta \\ &\leq \frac{1}{n} \log N - \min_{i=1, \dots, N} \Lambda_\delta^*(x_i) + \delta \quad (\text{par choix des } \theta_{x_i}) \\ &\leq \frac{1}{n} \log N - \Lambda_\delta^*(F) + \delta \quad (\text{puisque } x_i \in F) \end{aligned}$$

Finalement, nous avons montré que

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(F) \leq -\Lambda_\delta^*(F) + \delta$$

Pour conclure, il suffit d'observer que $\lim_{\delta \rightarrow 0} \Lambda_\delta^*(F) + \delta = \Lambda^*(F)$.

- **(2ième étape)** Soient $F \subset \mathbb{R}^d$ un ensemble fermé et $\delta > 0$. Observons dans un premier temps que l'hypothèse $D_\Lambda = \mathbb{R}^d$ portant sur la log-Laplace entraîne que $\mathbb{E}[e^{\|X\|_2}] < +\infty$ (il suffit de choisir $\theta = \frac{X}{\|X\|_2}$). En conséquence, pour tout $t > 0$,

$$\begin{aligned} \mathbb{P}\left(\left\|\frac{S_n}{n}\right\|_2 \geq t\right) &\leq e^{-nt} \mathbb{E}[e^{\|S_n\|_2}] \\ &\leq e^{-nt} \mathbb{E}[e^{\|X_1\|_2 + \dots + \|X_n\|_2}] \\ &\leq e^{-nt} \mathbb{E}[e^{\|X_1\|_2}]^n \end{aligned}$$

En conclusion, pour tout $M > 0$, il existe $t > 0$ tel que

$$\mathbb{P}\left(\left\|\frac{S_n}{n}\right\|_2 \geq t\right) \leq e^{-nM}, \quad n \geq 1$$

Lorsqu'une telle propriété est satisfaite, il est usuel de parler de tension exponentielle. Cette propriété va nous permettre de démontrer l'inégalité de grandes déviations pour

des ensembles fermés à partir de celle obtenue pour des ensembles compacts.

Désignons par K la boule euclidienne fermée de centre 0 et de rayon $t > 0$. L'ensemble $F \cap K$ étant compact, la première étape de la démonstration nous assure que, pour tout $n \geq n_0$ assez grand,

$$\mu_n(F \cap K) \leq e^{-n\Lambda_\delta^*(F \cap K)} \leq e^{-n\Lambda_\delta^*(F)}$$

Ainsi, toujours pour $n \geq n_0$, par choix de K nous avons, pour tout $M > 0$,

$$\begin{aligned} \mu_n(F) &\leq \mu_n(F \cap K) + \mu_n(K^c) \\ &\leq e^{-n\Lambda_\delta^*(F)} + e^{-Mn} \end{aligned}$$

puisque, par choix de K , $\mu_n(K^c) = \mathbb{P}\left(\left\|\frac{S_n}{n}\right\|_2 \geq t\right)$. A partir de l'ensemble F et de $\delta > 0$, il suffit donc d'ajuster la constante M afin que $M > \Lambda_\delta^*(F)$ pour conclure. C'est pourquoi

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(F) \leq -\Lambda_\delta^*(F)$$

et le résultat s'ensuit lorsque $\delta \rightarrow 0$.

2. Soit $G \subset \mathbb{R}^d$ un ouvert, montrons que $\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(G) \geq -\Lambda^*(G)$. A cet effet, il suffit de montrer que pour tout $x_0 \in \mathbb{R}^d$ et tout $\delta > 0$ l'inégalité suivante est satisfaite :

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(b(x_0, \delta)) \geq -\Lambda^*(x_0)$$

A nouveau, l'argument principal sera de **changer de mesure de probabilité**.

- Tout d'abord, considérons deux paramètres $0 < \eta < \delta$ et supposons que $x_0 = \nabla \Lambda(\theta_0)$ avec $\theta_0 \in \mathbb{R}^d$. Rappelons ensuite que

$$d\mu_{\theta_0, n}(x) = e^{\langle \theta_0, x \rangle - \Lambda(\theta_0)} d\mu(x)$$

est une mesure de probabilité. D'ailleurs, si les variables aléatoires Z_i , $i = 1, \dots, n$ sont des variables aléatoires indépendantes de loi commune μ_{θ_0} et $T_n = \frac{Z_1 + \dots + Z_n}{n}$ pour tout $n \geq 1$ alors $\mu_{\theta_0, n} = \mathcal{L}\left(\frac{T_n}{n}\right)$.

En procédant à ce changement de mesure, il n'est pas difficile de montrer que

$$\begin{aligned} \frac{1}{n} \log \mu_n(B(x_0, \delta)) &= \Lambda(\theta_0) - \langle \theta_0, x_0 \rangle + \frac{1}{n} \log \int_{B(x_0, \delta)} e^{-n\langle \theta_0, x_0 - x \rangle} d\mu_{\theta_0, n}(x) \\ &\geq \Lambda(\theta_0) - \langle \theta_0, x_0 \rangle + \frac{1}{n} \log \int_{B(x_0, \eta)} e^{-n\langle \theta_0, x_0 - x \rangle} d\mu_{\theta_0, n}(x) \end{aligned}$$

En outre, pour tout $x \in B(x_0, \eta)$, nous avons $|\langle \theta_0, x - x_0 \rangle| \leq \|\theta_0\|_2 \eta$. Ainsi,

$$\frac{1}{n} \log \mu_n(B(x_0, \delta)) \geq \Lambda(\theta_0) - \langle \theta_0, x_0 \rangle - \eta \|\theta_0\|_2 + \frac{1}{n} \log \mu_{\theta_0, n}(B(x_0, \eta))$$

Observons de plus que, par construction, l'espérance de Z_1 vaut $\mathbb{E}[Z_1] = \nabla \Lambda(\theta_0) = x_0$. D'où, par la loi faible des grands nombres,

$$\lim_{n \rightarrow +\infty} \mu_{\theta_0, n}(B(x_0, \eta)) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\frac{Z_1 + \dots + Z_n}{n} \in B(x_0, \eta)\right) = 1 > 0$$

C'est pourquoi nous avons,

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mu_n(B(x_0, \delta)) \geq \Lambda(\theta_0) - \langle \theta_0, x_0 \rangle - \eta \|\theta_0\|_2$$

Pour conclure, il suffit d'observer que $\Lambda(\theta_0) - \langle \theta_0, x_0 \rangle = \Lambda^*(x_0)$ (d'après la Proposition 4.3.1) puis de faire tendre $\eta \rightarrow 0$ et enfin d'optimiser en x_0 .

- **(Cas général).** Pour traiter le cas où $x_0 \notin \{\nabla \Lambda(\theta_0); \theta_0 \in \mathbb{R}^d\}$ il faut procéder à un argument de régularisation. Pour cela, débutons avec X une variable aléatoire de loi μ et considérons G un vecteur gaussien standard de \mathbb{R}^d indépendant de X . Posons ensuite $Y = X + \sigma G$ avec $\sigma > 0$ et désignons par ν la loi de Y . De plus le Lemme suivant est satisfait.

Lemme 4.3.1. *Pour tout $x_0 \in \mathbb{R}^d$, il existe $\theta_0 \in \mathbb{R}^d$ tel que $x_0 = \nabla \Lambda_\nu(\theta_0)$*

Démonstration. Ce résultat technique est laissé en exercice. Sa démonstration peut-être trouvée dans [37] (page 40). \square

Grâce à ce lemme, il est alors possible d'appliquer la première étape à Y pour enfin conclure en faisant tendre σ vers 0 comme nous l'avons fait en dimension une. \square

4.3.2 Application : Théorème de Sanov pour des mesures empiriques

Soient $E = \{a_1, \dots, a_d\}$ un ensemble fini muni de la tribu \mathcal{T} (désignant l'ensemble des parties de E) et $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{T})$ une variable aléatoire de loi μ . Considérons alors X_1, \dots, X_n des variables aléatoires, indépendantes de loi commune μ . A l'aide de celles-ci, il est possible de définir une mesure empirique $L_n : \Omega \rightarrow \mathcal{P}(E)$ (où $\mathcal{P}(E)$ désigne, rappelons le, l'ensemble des mesures de probabilités sur E) par

$$L_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)} \quad \text{avec } \omega \in \Omega.$$

Si $f : E \rightarrow \mathbb{R}$ est intégrable alors, par la loi forte des grands nombres,

$$\lim_{n \rightarrow +\infty} \int f dL_n(\omega) = \frac{1}{n} \sum_{i=1}^n f(X_i(\omega)) = \mathbb{E}[f(X)] = \int_E f d\mu \quad p.s.$$

puisque dans ce cas $(Y_i)_{i=1, \dots, n}$, avec $Y_i = f(X_i)$ pour $i = 1, \dots, n$, est une suite de variables réelles intégrables, *i.i.d.*. Une manière heuristique de résumer ce qui précède est la suivante :

p.s. $L_n \Rightarrow \mu$ dans $\mathcal{P}(E)$ lorsque $n \rightarrow +\infty$

Supposons à présent que $\Gamma \subset \mathcal{P}(E)$ soit un ensemble ne contenant pas μ (i.e. $\mu \notin \Gamma$). D'après ce qui précède $\lim_{n \rightarrow +\infty} \mathbb{P}(L_n \in \Gamma) = 0$. Le Théorème de Sanov s'attache alors à quantifier la vitesse de cette convergence.

Théorème 4.3.2 (Sanov (version faible)). *Supposons que E soit un ensemble fini. Alors, dans le cadre décrit précédemment, nous avons*

1. Si $\Gamma \subset \mathcal{P}(E)$ est fermé

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}(L_n \in \Gamma) \leq - \inf_{\nu \in \Gamma} H(\nu|\mu)$$

2. Si $\Gamma \subset \mathcal{P}(E)$ est un ouvert

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}(L_n \in \Gamma) \geq - \inf_{\nu \in \Gamma} H(\nu|\mu)$$

où $H(\nu|\mu)$ désigne l'entropie relative de ν par rapport à μ . Cette application est définie par

$$H(\nu|\mu) = \begin{cases} \sum_{k=1}^d \nu(a_k) \log \frac{\nu(a_k)}{\mu(a_k)} & \text{si } \nu \ll \mu \\ +\infty & \text{sinon.} \end{cases}$$

où la notation $\nu \ll \mu$ signifie que la mesure ν est absolument continue par rapport à la mesure μ .

Il est important de faire quelques remarques vis-à-vis du théorème précédent.

Remarque. 1. Tout d'abord, il est primordial de munir $\mathcal{P}(E)$ d'une topologie afin de définir la notion d'ensembles ouverts ou fermés. Une première possibilité consiste à utiliser la distance en variation totale d_{VT} définie par :

$$d_{VT}(\nu, \mu) = \|\nu - \mu\|_{VT} = \sup_{A \subset E} |\nu(A) - \mu(A)| \quad \text{pour } \mu, \nu \in \mathcal{P}(E).$$

Une autre option, lorsque $E = \{a_1, \dots, a_d\}$ est un ensemble fini, est de plonger $\mathcal{P}(E)$ dans \mathbb{R}^d de la manière suivante

$$\mu = (\mu(a_1), \dots, \mu(a_d)), \quad \mu \in \mathcal{P}(E)$$

afin d'utiliser la topologie induite par \mathbb{R}^d .

2. Voici quelques remarques concernant l'entropie relative $H(\nu|\mu)$. Nous utiliserons les conventions suivantes :

$$0 \log 0 = 0 \quad \text{et} \quad 0 \log(0/0) = 0$$

Observons aussi que, pour tout $k = 1, \dots, d$, $\mu(a_k) = 0$ entraîne $\nu(a_k) = 0$ puisque $\nu \ll \mu$. Notons également que $H(\nu, |\mu) \geq 0$ pour toutes mesures ν et μ . En effet,

$$\begin{aligned}
H(\nu|\mu) &= \sum_{k=1}^d \nu(a_k) \log \frac{\nu(a_k)}{\mu(a_k)} \\
&= \sum_{k=1}^d \frac{\nu(a_k)}{\mu(a_k)} \log \frac{\nu(a_k)}{\mu(a_k)} \mu(a_k) \\
&\geq \phi \left(\sum_{k=1}^d \nu(a_k) \frac{\nu(a_k)}{\mu(a_k)} \right) = 0
\end{aligned}$$

en utilisant l'inégalité de Jensen pour la fonction convexe $\phi(x) = x \log x$.

3. Il est possible de généraliser le Théorème de Sanov en supposant que E est un espace polonais. Dans ce cas, l'énoncé du théorème reste le même avec toutefois une expression légèrement différente de l'entropie relative. Celle-ci devient

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

où $\frac{d\nu}{d\mu}$ désigne la dérivée de Radon-Nykodym de ν par rapport à μ .

Démonstration. (du Théorème de Sanov) Comme annoncé, lorsque E est de cardinal fini, la **preuve s'effectue en plongeant $\mathcal{P}(E)$ dans \mathbb{R}^d** . Nous pouvons alors appliquer le Théorème de Cramer. A cet effet, considérons l'application $\psi : \mathcal{P}(E) \rightarrow \mathbb{R}^d$ définie par

$$\psi(\nu) = (\nu(a_1), \dots, \nu(a_d))$$

où, pour tous $k = 1, \dots, d$, $\nu(a_k)$ est un abus de notation afin de désigner $\nu(\{a_k\})$.

A nouveau, considérons la mesure empirique $L_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_i}$ obtenue à partir de la suite de variable aléatoires i.i.d $(X_i)_{i=1, \dots, n}$ de loi commune μ (E étant discret, il est muni de la tribu des parties de E).

Posons ensuite $\chi_i = \psi(\delta_{X_i}) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Il est essentiel d'observer ensuite que, pour tout $k = 1, \dots, d$,

$$\chi_i = e_k \quad \text{avec probabilité } \mu(a_k)$$

où e_k désigne le k -ième vecteur de la base canonique de \mathbb{R}^d . Ainsi, nous avons

$$L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \in \Gamma \iff \frac{1}{n} \sum_{i=1}^n \chi_i \in \psi(\Gamma)$$

Nous sommes à présent en position pour appliquer le Théorème de Cramer aux variables χ_i . Celui-ci nous fournit simultanément :

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}(L_n \in \Gamma) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \chi_i \in \psi(\Gamma)\right) \leq - \inf_{x \in \psi(\Gamma)} \Lambda^*(x)$$

et

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}(L_n \in \Gamma) = \liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \chi_i \in \psi(\Gamma)\right) \geq - \inf_{x \in \psi(\Gamma)} \Lambda^*(x)$$

où, bien entendu, Λ^* désigne la transformée de Fenchel-Legendre de la log-Laplace des variables χ_i . Il reste à présent à identifier Λ^* . Ici, nous avons

$$\begin{aligned} \Lambda^*(x) &= \sup_{\theta \in \mathbb{R}^d} \left[\langle \theta, x \rangle - \log \mathbb{E}[e^{\langle \theta, \chi_1 \rangle}] \right] \\ &= \sup_{\theta \in \mathbb{R}^d} \left[\langle \theta, x \rangle - \log \sum_{k=1}^d e^{\theta_k} \mu(a_k) \right] \end{aligned}$$

Soit alors $x = \psi(\nu) = (\nu(a_1), \dots, \nu(a_d))$ avec $\nu \in \mathcal{P}(E)$.

- Supposons, dans un premier temps, que $\nu(a_1), \dots, \nu(a_d) > 0$ ainsi

$$\begin{aligned} \Lambda^*(x) &= \sup_{\theta \in \mathbb{R}^d} \left[\sum_{i=1}^d \theta_k \nu(a_k) - \log \left(\sum_{k=1}^d e^{\theta_k} \mu(a_k) \right) \right] \\ &= \sup_{\theta \in \mathbb{R}^d} \left[\sum_{i=1}^d \theta_k \nu(a_k) - \log \left(\sum_{k=1}^d e^{\theta_k} \frac{\mu(a_k)}{\nu(a_k)} \nu(a_k) \right) \right] \\ &\leq \sup_{\theta \in \mathbb{R}^d} \left[\sum_{k=1}^d \theta_k \nu(a_k) + \sum_{k=1}^d \nu(a_k) \left(-\log \left(e^{\theta_k} \frac{\mu(a_k)}{\nu(a_k)} \right) \right) \right] \end{aligned}$$

où la dernière inégalité est obtenue en utilisant l'inégalité de Jensen pour la fonction convexe $x \mapsto -\log x$ sous la mesure de probabilité ν . En outre, la quantité suivante ne dépend pas de θ

$$\sum_{k=1}^d \theta_k \nu(a_k) + \sum_{k=1}^d \nu(a_k) \left(-\log \left(e^{\theta_k} \frac{\mu(a_k)}{\nu(a_k)} \right) \right) = - \sum_{k=1}^d \nu(a_k) \log \left(\frac{\mu(a_k)}{\nu(a_k)} \right)$$

C'est pourquoi

$$\Lambda^*(x) \leq - \sum_{k=1}^d \nu(a_k) \log \left(\frac{\mu(a_k)}{\nu(a_k)} \right) = H(\nu|\mu)$$

où la dernière égalité provient de la définition de l'entropie relative de ν par rapport à μ . L'inégalité entre $\Lambda^*(x)$ et $H(\nu|\mu)$ est en fait une égalité. Pour le constater, il suffit de choisir, pour tout $k = 1, \dots, d$, $\theta_k = \log \left(\frac{\mu(a_k)}{\nu(a_k)} \right)$ dans les calculs qui précèdent. Autrement dit

$$\Lambda^*(x) = H(\nu|\mu).$$

- Les mêmes arguments peuvent être développés lorsque $\nu \ll \mu$ en restreignant la somme aux entiers $k \in \{1, \dots, d\}$ tels que $\nu(a_k) > 0$.

- Lorsque ν n'est pas absolument continue par rapport à la mesure μ , il existe $k \in \{1, \dots, d\}$ tel que $\mu(a_k) = 0$ mais $\nu(a_k) > 0$. Il suffit alors de choisir $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$ tel que $\theta_j = 0$ lorsque $j \neq k$ puis de faire tendre θ_k vers $+\infty$ pour voir que dans ce cas de figure $\Lambda^*(x) = +\infty$ (puisque alors, par définition, $H(\nu|\mu) = +\infty$).

□

4.4 Théorème de Gärtner-Ellis

Si nous résumons les sections précédentes nous avons considéré une variable aléatoire

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}^d))$$

de loi μ avec sa log-Laplace

$$\Lambda(\theta) = \log \mathbb{E}[e^{\langle \theta, X \rangle}] \quad \text{avec } \theta \in \mathbb{R}^d$$

et

$$\Lambda^*(x) = \sup_{\theta \in \mathbb{R}^d} [\langle \theta, x \rangle - \Lambda(\theta)],$$

avec $x \in \mathbb{R}^d$, la transformée de Fenchel-Legendre associée.

Ensuite, nous avons considéré X_1, \dots, X_n des copies indépendantes de X et désigné par S_n la quantité suivante $S_n = \sum_{i=1}^n X_i$, $n \geq 1$. Sous certaines hypothèses portant sur Λ , le Théorème de Cramer nous assure essentiellement que, pour n suffisamment grand,

$$\mathbb{P}\left(\frac{S_n}{n} \in A\right) \simeq e^{-n\Lambda^*(A)}$$

avec A un ensemble ouvert ou fermé de \mathbb{R}^d et $\Lambda^*(A) = \inf_{x \in A} \Lambda^*(x)$.

L'hypothèse d'indépendance fut un point essentiel de la démonstration. Comme nous allons le voir, il est possible d'obtenir un résultat un peu plus général et plus souple que celui obtenu par Cramer. Il s'agit du Théorème de Gärtner-Ellis. Pour l'établir, nous devons au préalable modifier légèrement notre cadre d'étude.

Sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, considérons $(Z_n)_{n \geq 1}$ une suite de variables aléatoires à valeurs dans \mathbb{R}^d (nous noterons sa loi μ_n) ainsi que $(a_n)_{n \geq 1}$ une suite de nombres réels strictement positifs tels que

$$\lim_{n \rightarrow +\infty} a_n = +\infty.$$

Définissons ensuite l'application Λ_n par

$$\Lambda_n(a_n \theta) = \log \mathbb{E}[e^{\langle a_n \theta, Z_n \rangle}] \quad \text{avec } \theta \in \mathbb{R}^d$$

Cette application est l'analogue (pour la suite $(Z_n)_{n \geq 0}$) de la log-Laplace que nous avons utilisé dans la section précédente. Observons que $\Lambda_n(\theta) \in]-\infty, \infty]$ pour tout $\theta \in \mathbb{R}^d$. Par la suite, nous ferons les hypothèses suivantes sur cette nouvelle fonction :

Hypothèses 4.4.1. 1. Pour tout $\theta \in \mathbb{R}^d$, la limite suivante existe

$$\lim_{n \rightarrow +\infty} \frac{\Lambda_n(a_n \theta)}{a_n} := \Lambda(\theta) \in]-\infty, +\infty]. \quad (4.4.1)$$

2. De plus, $0 \in \overset{\circ}{D}_\Lambda$. Ceci signifiant qu'il existe $\epsilon > 0$ tel que

$$\text{si } \|\theta\|_2 \leq \epsilon \text{ alors } \Lambda(\theta) < +\infty.$$

Remarque. Le Théorème de Cramer est contenu dans ce nouveau cadre de travail. En effet, il suffit de choisir $Z_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$, pour $n \geq 1$ avec $(X_i)_{i=1, \dots, n}$ des variables aléatoires i.i.d. et $a_n = n$. Dans ce cas, nous avons alors

$$\frac{\Lambda_n(n\theta)}{n} = \log \mathbb{E}[e^{\langle \theta, X_1 \rangle}].$$

En général, ceci n'est pas satisfait : i.e. $\Lambda(\theta) = \lim_{n \rightarrow +\infty} \frac{\Lambda_n(n\theta)}{n} \neq \log \mathbb{E}[e^{\langle \theta, X_1 \rangle}]$.

4.4.1 Transformée de Fenchel-Legendre (3ième partie)

Avant d'énoncer le Théorème de Gärtner-Ellis, nous allons voir que la fonction Λ (définie plus haut par (4.4.1)) satisfait des propriétés analogues à la log-Laplace (utilisée dans le Théorème de Cramer).

Proposition 4.4.1. (*Propriétés de Λ*) Dans le cadre décrit précédemment, l'application Λ vérifie les propriétés suivantes :

1. Λ est convexe.
2. Sa transformée de Fenchel-Legendre $\Lambda^*(x) = \sup_{\theta \in \mathbb{R}^d} [\langle \theta, x \rangle - \Lambda(\theta)]$, avec $x \in \mathbb{R}^d$, est une application positive et convexe.
3. L'application Λ^* est aussi semi-continue inférieurement et ses lignes de niveau $\{\Lambda^* \leq a\}$, pour $a \geq 0$, sont des ensembles compacts.

Démonstration. La démonstration de ce résultat (pouvant se trouver dans [37]) est laissée en exercice. Les arguments sont semblables à ceux utilisés lorsque Λ désigne la log-Laplace d'un vecteur aléatoire de \mathbb{R}^d dont les coordonnées sont indépendantes et de même loi. \square

Pour obtenir une extension du théorème de Cramer dans un cadre autorisant de la dépendance, nous devons introduire la notion de point exposé.

Définition 4.4.1. Un point $x \in \mathbb{R}^d$ tel que $\Lambda^*(x) < +\infty$ est dit exposé s'il est possible de trouver $\theta \in \mathbb{R}^d$ tel que

$$\langle \theta, x \rangle - \Lambda^*(x) > \langle \theta, y \rangle - \Lambda^*(y) \quad \text{pour tout } y \in \mathbb{R}^d, y \neq x.$$

Un tel $\theta \in \mathbb{R}^d$ est appelé hyperplan exposant associé à x .

Exemple 4.4.1. Dans le cas gaussien, en dimension 1, nous avons $\Lambda^*(x) = \frac{x^2}{2}$ et il n'est pas difficile de vérifier que le choix $\theta = x$ convient.

4.4.2 Théorème de Gärtner-Ellis

Énonçons à présent ce nouveau résultat.

Théorème 4.4.1 (Gärtner-Ellis). *Soient $(Z_n)_{n \geq 1}$ et $(a_n)_{n \geq 1}$ des suites satisfaisant les hypothèses (4.4.1) décrites un peu plus haut. Rappelons que μ_n désigne la loi de Z_n , $n \geq 1$. Alors*

1. pour tout ensemble fermé $F \subset \mathbb{R}^d$

$$\limsup_{n \rightarrow +\infty} \frac{1}{a_n} \log \mu_n(F) \leq -\Lambda^*(F)$$

2. pour tout ensemble ouvert $G \subset \mathbb{R}^d$

$$\liminf_{n \rightarrow +\infty} \frac{1}{a_n} \log \mu_n(F) \geq -\Lambda^*(G \cap \mathcal{F})$$

où \mathcal{F} est l'ensemble des points exposés $x \in \mathbb{R}^d$ tels que l'hyperplan exposant $\theta \in D_\Lambda^\circ$.

3. Si $\theta \mapsto \Lambda$ est finie et différentiable sur tout \mathbb{R}^d alors la minoration de l'assertion précédente est satisfaite avec $\Lambda^*(G)$ au lieu de $\Lambda^*(G \cap \mathcal{F})$.

Avant d'entamer la preuve de ce résultat nous allons l'illustrer sur différents exemples, ceci permettra notamment de préciser l'importance de certaines hypothèses ainsi que de mettre en évidence les différences avec le Théorème de Cramer. Pour plus de détails à ce sujet nous renvoyons le lecteur vers [37].

Exemple 4.4.2. 1. Débutons par un exemple pour lequel il n'est pas possible de supprimer la présence de l'ensemble \mathcal{F} . A cet effet, considérons la suite de variables aléatoires définie par :

$$\mathbb{P}(Z_n = 1) = \mathbb{P}(Z_n = -1) = \frac{1}{2}$$

Dans ce cas, pour le choix $a_n = n$, nous avons $\Lambda_n(n\theta) = \log \left(\frac{e^{\theta n} + e^{-\theta n}}{2} \right)$ avec $\theta \in \mathbb{R}$ et

$$\lim_{n \rightarrow +\infty} \frac{\Lambda_n(n\theta)}{a_n} = |\theta| = \Lambda(\theta) \quad \text{avec } \theta \in \mathbb{R}$$

Ainsi, il est alors possible de calculer la transformée de Fenchel-Legendre Λ^* de Λ : pour tout $x \in \mathbb{R}$

$$\Lambda^*(x) = \sup_{\theta \in \mathbb{R}} [\theta x - |\theta|] = \begin{cases} 0 & \text{si } |x| \leq 1 \\ +\infty & \text{sinon.} \end{cases}$$

Observons ensuite que $\Lambda(\theta) < +\infty$ pour tout $\theta \in \mathbb{R}$ (i.e. $D_\Lambda = \mathbb{R}$). Cependant Λ n'est pas différentiable en 0. En particulier, la dernière assertion du Théorème de Gärtner-Ellis n'est pas vérifiée. Ceci est notamment visible en choisissant l'ensemble $G =]-\frac{1}{n}, \frac{1}{n}[$ et en observant que

$$\mu_n(G) = \mathbb{P}\left(\frac{Z_n}{n} \in G\right) = \mathbb{P}(Z_n \in]-1, 1]) = 0.$$

Ainsi, si la dernière assertion était vérifiée, ce choix d'ouvert G aboutirait à une contradiction. En revanche, il est possible de montrer que $\mathcal{F} = \{-1, 1\}$.

2. Dans cet exemple, $0 \in \overset{\circ}{D}_\Lambda$ mais la dernière assertion n'est toujours pas vérifiée, car la transformée de Fenchel-Legendre n'est pas différentiable sur tout \mathbb{R} . Soit $Z_n = \frac{1}{n}(S_n + \mathcal{E})$ avec \mathcal{E} une variable aléatoire de loi exponentielle $\mathcal{E}(1)$ (de densité de probabilité $x \mapsto e^{-x}1_{x \geq 0}$) et $S_n = \sum_{i=1}^n X_i$ avec $(X_i)_{i=1, \dots, n}$ des variables i.i.d. gaussiennes standards indépendantes de \mathcal{E} .

Nous avons alors, pour $a_n = n$,

$$\frac{\Lambda_n(n\theta)}{n} = \frac{1}{n} \times (\log \mathbb{E}[e^{\theta X_1}]^n + \log \mathbb{E}[e^{\theta \mathcal{E}}]) = \begin{cases} \frac{\theta^2}{2} + \frac{1}{n(1-\theta)} & \text{si } \theta < 1 \\ +\infty & \text{sinon.} \end{cases}$$

C'est pourquoi

$$\lim_{n \rightarrow +\infty} \frac{\Lambda_n(n\theta)}{a_n} = \begin{cases} \frac{\theta^2}{2} & \text{si } \theta < 1 \\ +\infty & \text{sinon.} \end{cases}$$

3. Enfin, présentons un exemple qui ne peut-être pas traité par le Théorème de Cramer. Soit

$$W : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

une variable aléatoire dont la log-Laplace est finie sur \mathbb{R} : $\Lambda_W(\theta) < +\infty$, $\theta \in \mathbb{R}$. Considérons ensuite $(W_n)_{n \geq 1}$ une suite de copies indépendantes de W . Ensuite, pour $|\alpha| < 1$, définissons par récurrence la suite $(X_n)_{n \geq 1}$ de la manière suivante

$$\begin{cases} X_0 = 0 \\ X_n = \alpha X_{n-1} + W_n, n \geq 1 \end{cases}$$

Cette nouvelle suite permet de construire le processus auto-regressif $Y_n = X_1 + \dots + X_n$ avec $n \geq 1$. Si $Z_n = \frac{Y_n}{n}$ et $a_n = n$, nous avons alors $\Lambda_n(n\theta) = \log \mathbb{E}[e^{\theta Y_n}]$, $\theta \in \mathbb{R}$. Il n'est pas difficile de montrer que

$$X_n = \frac{1 - \alpha^n}{1 - \alpha} W_1 + \frac{1 - \alpha^{n-1}}{1 - \alpha} W_2 + \dots + W_n \quad \text{avec } n \geq 1.$$

C'est pourquoi

$$\Lambda_n(n\theta) = \sum_{i=1}^n \Lambda_W\left(\frac{1 - \alpha^i}{1 - \alpha} \theta\right)$$

Enfin, d'après le Théorème de Césaro, nous avons

$$\lim_{n \rightarrow +\infty} \Lambda_n(n\theta) = \Lambda_W\left(\frac{\theta}{1 - \alpha}\right) = \Lambda(\theta).$$

Il est ensuite possible de calculer Λ^* afin d'obtenir

$$\Lambda^*(x) = \Lambda_W^*((1-\alpha)x), \quad x \in \mathbb{R}$$

Dans un tel cadre, il est alors clair que toutes les assertions du Théorème de Gärtner-Ellis sont vérifiées alors que nous n'aurions pas pu appliquer le Théorème de Cramer à cause l'absence d'indépendance.

4. Le Théorème de Gärtner-Ellis peut également être utile pour étudier des déviations modérées. Plus précisément, considérons X une variable aléatoire de loi μ et supposons que sa transformée de Laplace est finie sur tout \mathbb{R} .

Considérons ensuite X_1, \dots, X_n des copies indépendantes de X et posons à nouveau

$$S_n = X_1 + \dots + X_n \quad \text{avec} \quad n \geq 1.$$

Choisissons ensuite $\alpha \in]\frac{1}{2}, 1[$ et étudions $Z_n = \frac{S_n}{n^\alpha}$. Nous supposons également, par souci de clarté, que $\mathbb{E}[X] = 0$ et $\mathbb{E}[X^2] = 1$. Avec ces conventions et pour le choix $a_n = n^{2\alpha-1}$, nous obtenons

$$\Lambda_n(a_n\theta) = n\Lambda\left(\frac{\theta}{n^{1-\alpha}}\right)$$

et, en utilisant un développement de Taylor d'ordre 2 en 0, nous avons

$$\begin{aligned} \frac{\Lambda_n(\theta)}{a_n} &= n^{2-2\alpha}\Lambda\left(\frac{\theta}{n^{1-\alpha}}\right) \\ &= n^{2-2\alpha}\left[\Lambda(0) + \frac{\theta}{n^{1-\alpha}}\Lambda'(0) + \frac{\theta^2}{2n^{2-2\alpha}}\Lambda''(0) + o(n^{2\alpha-2})\right] \\ &= \frac{\theta^2}{2} + o(1) \end{aligned}$$

D'où,

$$\lim_{n \rightarrow +\infty} \frac{\Lambda_n(\theta)}{a_n} = \frac{\theta^2}{2} \quad \text{et} \quad \Lambda^*(x) = \frac{x^2}{2} \quad \text{avec} \quad x \in \mathbb{R}.$$

Il est alors possible d'appliquer le Théorème de Gärtner-Ellis afin d'obtenir

$$\lim_{n \rightarrow +\infty} \frac{1}{n^{2\alpha-1}} \log \mathbb{P}\left(\frac{S_n}{n^\alpha} \geq t\right) = -\frac{t^2}{2}$$

en choisissant $F = [t, +\infty[$ et $G =]t, +\infty[$ pour $t \in \mathbb{R}$. Il est à noter que la limite obtenue ne dépend pas de la mesure de probabilité μ . Ce résultat est également à comparer avec celui fourni par le Théorème de Cramer.

Après ces quelques exemples d'applications, nous pouvons maintenant nous attacher à la démonstration du Théorème de Gärtner-Ellis.

Démonstration. 1. La première assertion est très similaire à celle du Théorème de Cramer. Pour la démontrer il faut d'abord considérer des ensembles compacts puis utiliser de la tension exponentielle. Nous laissons cette partie de la démonstration à titre d'exercice.

2. Le dernier point ressemble également beaucoup à la minoration dans le Théorème de Cramer. La démonstration s'effectue par convolution avec une gaussienne. Plus précisément, le Lemme suivant est nécessaire (cf. [37] pour une démonstration)

Lemme 4.4.1. (a) Si $x = \nabla\Lambda(\theta)$ avec $\theta \in \overset{\circ}{D}_\Lambda$ alors $\Lambda^*(x) = \langle \theta, x \rangle - \Lambda(\theta)$.

(b) De plus, $x \in \mathcal{F}$ et $\theta \in \mathbb{R}^d$ est un hyperplan exposant associé à x .

En s'appuyant sur le précédent Lemme, la preuve s'ensuit par un argument de régularisation (via une convolution avec une gaussienne indépendante) permettant de montrer que tout les points sont exposés et qu'ainsi $\Lambda^*(G \cap \mathcal{F}) = \Lambda^*(G)$.

3. La seconde assertion du Théorème requiert de nouvelles idées. Rappelons que nous souhaitons prouver l'inégalité suivante :

$$\liminf_{n \rightarrow +\infty} \frac{1}{a_n} \log \mu_n(F) \geq -\Lambda^*(G \cap \mathcal{F})$$

où G est un ouvert de \mathbb{R}^d et \mathcal{F} l'ensemble des points exposés. Il suffit de montrer que pour tout $x \in \mathcal{F}$

$$\lim_{\delta \rightarrow 0} \liminf_{n \rightarrow +\infty} \frac{1}{a_n} \log \mu_n(B(x, \delta)) \geq -\Lambda^*(x)$$

En effet, si $x \in G \cap \mathcal{F}$ il existe $\epsilon > 0$ tel que $B(x, \epsilon) \subset G$. En particulier,

$$\begin{aligned} \underline{\lim}_{n \rightarrow +\infty} \frac{1}{a_n} \log \mu_n(G) &\leq \liminf_{n \rightarrow +\infty} \frac{1}{a_n} \log \mu_n(B(x, \epsilon)) \\ &\leq \lim_{\delta \rightarrow 0} \liminf_{n \rightarrow +\infty} \frac{1}{a_n} \log \mu_n(B(x, \delta)) \leq -\Lambda^*(x) \end{aligned}$$

puisque l'application $\epsilon \mapsto \mu_n(B(x, \epsilon))$ est décroissante.

Considérons donc $x \in \mathcal{F}$, il existe donc $\theta \in \mathbb{R}^d$ un hyperplan exposant associé à x tel que $\Lambda(\theta) < +\infty$. En particulier, puisque $\Lambda(\theta) = \lim_{n \rightarrow +\infty} \frac{\Lambda_n(\theta)}{a_n}$, $\Lambda_n(\theta) < +\infty$ pour $n \geq N_0$ suffisamment grand. Procédons à un changement de mesure de probabilité

$$d\mu_{\theta,n}(z) = e^{a_n \langle \theta, z \rangle - \Lambda_n(a_n \theta)} d\mu_n(z) \quad \text{où} \quad \mu_n = \mathcal{L}(Z_n)$$

Exactement comme dans la démonstration du Théorème de Cramer, nous obtenons

$$\frac{1}{a_n} \log \mu_n(B(x, \delta)) \geq \frac{1}{a_n} \Lambda_n(a_n \theta) - \langle \theta, x \rangle - \delta \|\theta\|_2 + \frac{1}{a_n} \log \mu_{\theta,n}(B(x, \delta))$$

En conséquence,

$$\begin{aligned} \lim_{\delta \rightarrow 0} \liminf_{n \rightarrow +\infty} \frac{1}{a_n} \log \mu_n(B(x, \delta)) &\geq \Lambda(\theta) - \langle \theta, x \rangle + \lim_{\delta \rightarrow 0} \frac{1}{a_n} \log \mu_{\theta, n}(B(x, \delta)) \\ &\geq -\Lambda^*(x) + \lim_{\delta \rightarrow 0} \frac{1}{a_n} \log \mu_{\theta, n}(B(x, \delta)) \end{aligned}$$

par définition de $\Lambda^*(x)$.

Il ne reste plus qu'à montrer que le dernier terme de cette minoration est positif. Dans le Théorème de Cramer nous avons utilisé la loi faible des grands nombres, ce n'est pas envisageable ici car nous n'avons plus l'hypothèse d'indépendance.

Pour résoudre cette difficulté, nous allons utiliser la majoration (obtenue dans la première assertion du Théorème de Gärtner-Ellis) avec l'ensemble fermé $F = B(x, \delta)^c$. Dans un premier temps, nous devons donc vérifier que les hypothèses du Théorème sont vérifiées par la mesure $\mu_{\theta, n}$ pour tout $n \geq N_0$.

Il faut donc calculer la log-Laplace associée. Ici, nous avons, pour tout $\lambda \in \mathbb{R}$

$$\begin{aligned} \frac{1}{a_n} \Lambda_{\theta, n}(a_n \lambda) &= \frac{1}{a_n} \log \left(\int_{\mathbb{R}^d} e^{a_n \langle \lambda, z \rangle} d\mu_{\theta, n}(z) \right) \\ &= -\frac{\Lambda_n(a_n \theta)}{a_n} + \frac{1}{a_n} \log \left(\int_{\mathbb{R}^d} e^{a_n \langle \lambda + \theta, z \rangle} d\mu(z) \right) \end{aligned}$$

Ceci mène alors à $\lim_{n \rightarrow +\infty} \frac{\Lambda_{\theta, n}(a_n \lambda)}{a_n} = \Lambda(\lambda + \theta) - \Lambda(\theta)$. Par hypothèse sur θ , cette quantité est finie à condition que $|\lambda|$ soit petit. Les hypothèses du Théorème de Gärtner-Ellis sont donc satisfaites.

Pour de tels λ , désignons par $\Lambda_\theta(\lambda) := \Lambda(\lambda + \theta) - \Lambda(\theta)$ et définissons $\Lambda_\theta^*(x)$ par

$$\Lambda_\theta^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{ \langle \lambda, x \rangle - \Lambda_\theta(\lambda) \}$$

La majoration du Théorème de Gärtner-Ellis nous assure que

$$\limsup_{n \rightarrow +\infty} \frac{1}{a_n} \log \mu_{\theta, n}(F) \leq -\Lambda_\theta^*(F)$$

Si jamais $\Lambda_\theta^*(F) > 0$ cela signifie que $\lim_{n \rightarrow +\infty} \mu_{\theta, n}(F) = 0$, en conséquent $\lim_{n \rightarrow +\infty} \mu_{\theta, n}(F^c) = 1$ et nous pourrions conclure.

Vérifions alors que $\Lambda_{\mu, \theta}^*(F) > 0$. Deux cas de figures s'imposent à nous :

- (a) Si $\Lambda_\theta^*(F) = +\infty$, il n'y a rien à faire.

- (b) Si $\Lambda_\theta^*(F) < +\infty$, il est possible de montrer que $\Lambda_\theta^*(F) = \Lambda_\theta^*(x_0)$ avec $x_0 \in F$. Pour obtenir ce résultat il faut utiliser le fait que les lignes de niveau de $x \mapsto \Lambda_\theta^*(x)$ sont compactes (cf. [37]).

Soit un tel $x_0 \in F$ et d'observons que, pour tout $y \in \mathbb{R}^d$,

$$\Lambda_\theta^*(y) = \Lambda^*(y) - \langle \theta y \rangle + \Lambda(\theta)$$

En particulier,

$$\Lambda_\theta^*(x_0) = \Lambda^*(x_0) - \langle \theta x_0 \rangle + \Lambda(\theta)$$

D'où, pour tout $x \in \mathbb{R}^d$,

$$\Lambda_\theta^*(x_0) \geq (\Lambda^*(x_0) - \langle \theta, x_0 \rangle) - (\Lambda^*(x) - \langle \theta, x \rangle)$$

puisque, par définition du supremum, $\Lambda^*(x) \geq \langle \theta, x \rangle - \Lambda(\theta)$.

Enfin, puisque x est un point exposé d'hyperplan associé θ et $x_0 \neq x$ (car $F = B(x, \delta)^c$ donc $x \in F^c$),

$$\langle \theta, x \rangle - \Lambda^*(x) > \langle \theta, x_0 \rangle - \Lambda^*(x_0) \iff (\Lambda^*(x_0) - \langle \theta, x_0 \rangle) - (\Lambda^*(x) - \langle \theta, x \rangle) > 0$$

Autrement dit, $\Lambda_\theta^*(x_0) > 0$ ce qui conclut la démonstration. □

4.5 Principes de grandes déviations

Comme nous avons pu le voir dans les sections précédentes, les démonstrations des Théorèmes de grandes déviations s'obtiennent par un schéma de preuve assez identique. Dans cette section, nous allons mettre en évidence certains principes qui peuvent être mis en oeuvre dans le cadre, plus général, espaces métriques (E, d) (par exemple \mathbb{R}^d , $C^0[0, 1], \dots$).

Par la suite nous considérons $(\mu_\epsilon)_{\epsilon > 0}$ une famille de mesures de probabilités sur les boréliens d'un espace métrique (E, d) . Pour $\epsilon > 0$ fixé, nous désignerons par Z_ϵ une variable aléatoire

$$Z_\epsilon : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{B}(E))$$

telle que $\mathcal{L}(Z_\epsilon) = \mu_\epsilon$.

Dans ce qui va suivre nous adopterons la notation suivante : si $f : E \rightarrow \overline{\mathbb{R}}_+$ nous noterons par D_f l'ensemble suivant

$$D_f = \{x \in E \ ; \ f(x) < +\infty\}$$

4.5.1 Fonctions de taux et principes de grandes déviations

Voyons ce qu'il est possible de dire sur la fonction décrivant le comportement des grandes déviations.

Définition 4.5.1. 1. Une fonction de taux est une application $I : E \rightarrow [0, +\infty]$ semi-continue inférieurement.

2. De plus, si les lignes de niveaux $\{I \leq \alpha\}$, avec $\alpha \geq 0$, sont compactes alors I est une bonne fonction de taux.

Remarque. Il est aussi possible de montrer (cf. [37]) qu'une bonne fonction de taux atteint ses minimums sur des ensembles fermés. Autrement dit, si F est un fermé de E alors $I(F) := \inf_{x \in F} I(x) = I(x_0)$ pour un certain $x_0 \in F$. Par convention, si $A = \emptyset$, $I(A) = +\infty$.

Définition 4.5.2. La famille $(\mu_\epsilon)_{\epsilon > 0}$ satisfait un principe de grandes déviations (P.G.D. en abrégé) de vitesse ϵ avec pour fonction de taux $I : E \rightarrow [0, +\infty]$ si, pour tout boréliens $A \subset E$, nous avons

$$-I(\overset{\circ}{A}) \leq \liminf_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(A) \leq \limsup_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(A) \leq -I(\bar{A})$$

De manière équivalente, cela s'exprime comme suit

1. Pour tout ensemble $F \subset E$ fermé, nous avons

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(F) \leq -I(F).$$

2. Pour tout ouvert $G \subset E$, nous avons

$$\liminf_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(F) \geq -I(G).$$

Bien entendu, les Théorèmes de Gärtner-Ellis et de Cramer s'inscrivent dans un tel cadre.

Exemple 4.5.1. Dans les sections précédentes nous avons $E = \mathbb{R}^d$, $\epsilon = \epsilon_n = \frac{1}{n}$, avec $n \geq 1$ et $\mu_n = \mathcal{L}\left(\frac{S_n}{n}\right)$. Le Théorème de Cramer montrait que la famille $(\mu_n)_{n \geq 1}$ vérifiant un P.G.D. de vitesse $\frac{1}{n}$ avec pour bonne fonction de taux $I = \Lambda^*$.

4.5.2 Schéma de preuve et aspects théoriques des grandes déviations

Un oeil attentif aux démonstrations faites dans les sections précédentes permet d'établir un schéma de preuve récurrent. En effet, voici ce que nous avons mis en oeuvre pour démontrer le Théorème de Cramer afin d'établir un P.G.D. :

- Pour établir la majoration sur les ensembles fermés, il suffit de montrer que l'inégalité suivante est vérifiée pour tout $0 < r < I(F)$. C'est-à-dire,

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(F) \leq -r$$

pour ensuite faire tendre r vers $I(F)$. Alternativement, il est aussi possible de montrer que pour tout $\delta > 0$ l'inégalité suivante est satisfaite :

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(F) \leq -I^\delta(F)$$

avec $I^\delta(x) = \min(I(x) - \delta, \frac{1}{\delta})$ pour enfin faire tendre δ vers 0.

- Pour établir la minoration sur les ouverts, il suffit de montrer que, pour tout $x \in D_I$, nous avons

$$\lim_{\delta \rightarrow 0} \liminf_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(B(x, \delta)) \geq -I(x)$$

Voici à présent quelques propriétés, plus théoriques, concernant les principes de grandes déviations.

Proposition 4.5.1 (Unicité de la fonction de taux). *Si $(\mu_\epsilon)_{\epsilon > 0}$ satisfait un P.G.D., la fonction de taux associée est alors unique.*

Démonstration. Supposons que la famille $(\mu_\epsilon)_{\epsilon > 0}$ satisfasse un P.G.D. avec deux fonctions de taux différentes I_1 et I_2 . Alors, il existe $x_0 \in E$ tel que $I_1(x_0) > I_2(x_0)$ (par exemple). Supposons ensuite que $I_1(x_0) < +\infty$ et considérons, pour tout $\delta > 0$ l'ensemble suivant

$$O_\delta = \{y \in E, I_1(y) > I_1(x_0) - \delta\}$$

Pour $\delta > 0$ fixé, il s'agit d'un ensemble ouvert (puisque I_1 est semi-continue inférieurement). Ainsi, pour tout $x_0 \in O_\delta$, il existe un voisinage ouvert A (de x_0) tel que $\bar{A} \subset O_\delta$. En appliquant le P.G.D. pour I_1 et I_2 , nous obtenons alors

$$\begin{aligned} -I_2(x_0) \leq -I_2(A) &\leq \liminf_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(A) \\ &\leq \liminf_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(\bar{A}) \quad (\text{puisque } A \subset \bar{A}) \\ &\leq -I_1(\bar{A}) \leq -I_1(O_\delta) \quad (\text{par définition de } A) \\ &\leq -I_1(x_0) + \delta \end{aligned}$$

Finalement, lorsque $\delta \rightarrow 0$, nous en déduisons que $I_1(x_0) \leq I_2(x_0)$ ce qui est impossible. \square

Il peut-être parfois difficile d'établir un P.G.D., c'est pourquoi nous introduisons une notion plus faible et plus simple à obtenir.

Définition 4.5.3. $(\mu_\epsilon)_{\epsilon > 0}$ satisfait un P.G.D. faible si la majoration n'est valable que pour les compacts de E . Autrement dit, pour tout ensemble $K \subset E$ compact,

$$\liminf_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(K) \leq -I(K). \quad (4.5.1)$$

La proposition suivante montre de quelle manière il est possible d'établir un *P.G.D.* à partir d'un *P.G.D.* faible.

Proposition 4.5.2. 1. Supposons que $(\mu_\epsilon)_{\epsilon>0}$ vérifie les hypothèses suivantes :

- (a) $(\mu_\epsilon)_{\epsilon>0}$ satisfait un *PGD* faible ;
- (b) la famille de probabilités $(\mu_\epsilon)_{\epsilon>0}$ est exponentiellement tendue : pour tout $M > 0$, il existe un compact K de (E, d) tel que

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(K^c) \leq -M.$$

alors la majoration (4.5.1) est valable pour les ensembles fermés de E .

- 2. Si de plus la minoration sur les ouverts est satisfaite alors la fonction de taux est bonne.

Démonstration. Soit F un fermé de E . Si $I(F) = 0$, il n'y a rien à prouver. Alors, soit r , un réel tel que $0 < r < I(F)$. Pour $M = 2r > 0$, considérons $K = K_M$ le compact fourni par la tension exponentielle. Le *P.G.D.* faible, appliqué à l'ensemble compact $F \cap K$ nous fournit

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(F \cap K) \leq -I(F \cap K) \leq -I(F) < -r$$

puisque $\inf_{x \in F \cap K} I(x) \geq \inf_{x \in F} I(x)$. De plus, en utilisant la tension exponentielle,

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon(K_M^c) \leq -2r.$$

En outre, $\mu_\epsilon(F) \leq \mu_\epsilon(F \cap K) + \mu_\epsilon(K^c)$. Pour conclure la démonstration, nous aurons besoin du Lemme suivant :

Lemme 4.5.1. Soit $N \geq 1$ étant fixé, pour $k = 1, \dots, N$ soient $(a_\epsilon^k)_{\epsilon>0}$ tels que $a_\epsilon^k > 0$. Alors

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \left(\sum_{k=1}^N a_\epsilon^k \right) = \max_{k=1, \dots, N} \limsup_{\epsilon \rightarrow 0} \epsilon \log a_\epsilon^k$$

Démonstration. (du Lemme) En effet, $\sum_{k=1}^N a_\epsilon^k \leq N \max_{k=1}^N a_\epsilon^k$. Ainsi, puisque $x \mapsto \log x$ est une fonction croissante,

$$\epsilon \log \left(\sum_{k=1}^N a_\epsilon^k \right) \leq \epsilon \log N + \max_{k=1, \dots, N} \epsilon \log(a_\epsilon^k)$$

□

Pour terminer la preuve de la proposition, il suffit alors de choisir $a_\epsilon^1 = \mu_\epsilon(F \cap K)$ et $a_\epsilon^2 = \mu_\epsilon(K^c)$. En effet, par ce qui précède, nous obtenons

$$\begin{aligned} \limsup_{\epsilon \rightarrow 0} \epsilon \log_{\epsilon}(F) &\leq \limsup_{\epsilon \rightarrow 0} \epsilon \log (\mu_{\epsilon}(F \cap K) + \mu_{\epsilon}(K^c)) \\ &\leq \max_{k=1,2} (-2r, -r) = -r \quad (\text{d'après le Lemme}) \end{aligned}$$

□

4.5.3 Principes de contraction

Voici un autre façon d'établir des *P.G.D.* : le principe de contraction.

Théorème 4.5.1. *Soient E et F des espaces métriques et $f : E \rightarrow F$ une application continue. Soit I est une bonne fonction de taux, nous définissons alors l'application J par :*

$$J(y) = \inf\{I(x), x \in E \mid f(x) = y\} \quad \text{avec } y \in F. \quad (4.5.2)$$

Alors, J est bonne fonction de taux sur F . Si de plus, $(\mu_{\epsilon})_{\epsilon > 0}$ satisfait un *P.G.D.* sur E avec I pour fonction de taux alors la famille $(\mu_{\epsilon} \circ f^{-1})_{\epsilon > 0}$ satisfait un *P.G.D.* sur F avec pour fonction de taux J

Remarque. Rappelons que la famille de mesure de probabilités $(\mu_{\epsilon} \circ f^{-1})_{\epsilon > 0}$ est définie comme suit : pour tout $A \in \mathcal{B}(F)$,

$$\mu_{\epsilon} \circ f^{-1}(A) = \mu_{\epsilon}(f^{-1}(A))$$

Démonstration. Tout d'abord observons que J est une application positive et semi-continue inférieurement. Montrons que J est une bonne fonction de taux. A cet effet, soit $\alpha \geq 0$, alors $\{J \leq \alpha\} = \{y \in F, J(y) \leq \alpha\} = \{f(x) \mid x \in E, I(x) \leq \alpha\}$. Puisque f est continue l'ensemble $f^{-1}(\{y\})$ est fermé. De plus, I étant une bonne fonction de taux nous avons

$$J(y) = \inf_{x \in f^{-1}(\{y\})} I(x) = I(x_0) \quad \text{avec } x_0 \in f^{-1}(\{y\})$$

En outre, notons que $\{f(x) \mid x \in E \text{ et } I(x) \leq \alpha\} = f(\{I \leq \alpha\})$. I étant une bonne fonction de taux, l'ensemble $\{I \leq \alpha\}$ est un compact de E . La continuité de f nous assure alors que l'ensemble $\{J \leq \alpha\}$ est également compact.

Enfin, montrons que $(\mu_{\epsilon} \circ f^{-1})_{\epsilon > 0}$ satisfait un *P.G.D.* avec J pour fonction de taux. Pour cela, il suffit d'observer que $\mu_{\epsilon} \circ f^{-1}(A) = \mu_{\epsilon}(f^{-1}(A))$. C'est pourquoi, en utilisant la continuité de f , si A un fermé (respectivement un ouvert) de F alors $f^{-1}(A)$ est un fermé (respectivement un ouvert) de E . En conclusion

$$J(A) = I(f^{-1}(A))$$

□

Voici une illustration de ce principe de contraction.

Exemple 4.5.2. Reprenons l'exemple proposé dans la section concernant le Théorème de Gärtner-Ellis. Nous avons

$$S_n = X_1 + \dots + X_n, n \geq 1$$

avec $X_i \sim \mathcal{N}(0, 1)$, $i = 1, \dots, n$ des variables aléatoires i.i.d. Soit également, \mathcal{E} une variable aléatoire exponentielle (de densité de probabilité $x \mapsto e^{-x}1_{\{x \geq 0\}}$ indépendante de la famille $(X_i)_{i=1, \dots, n}$ et posons $Y_n = \mathcal{E} + S_n$. Comme nous l'avions observé, le Théorème de Gärtner-Ellis permettait d'obtenir la majoration sur les fermés (apparaissant dans un *P.G.D.*) mais était inefficace vis à vis de la minoration sur les ouverts.

Nous présentons ci-dessous une stratégie potentielle permettant d'affiner ce résultat. Tout d'abord, il est possible de montrer via le Théorème de Cramer que $(\frac{S_n}{n})_{n \geq 1}$ satisfait un *P.G.D.* et, avec un peu de travail, que $(\frac{\mathcal{E}}{n})_{n \geq 1}$ également. Il suffit ensuite d'utiliser le Lemme suivant.

Lemme 4.5.2. Soient $(\mu_n)_{n \geq 0}$ et $(\nu_n)_{n \geq 0}$ deux familles de mesures de probabilités satisfaisant chacune un *PGD*, de vitesse n , avec, respectivement pour bonne fonction de taux I (respectivement J) sur \mathbb{R} .

Si les familles $(\mu_n)_{n \geq 0}$ et $(\nu_n)_{n \geq 0}$ sont exponentiellement tendues alors la famille de mesures de probabilités $(\mu_n \otimes \nu_n)_{n \geq 1}$ satisfait un *P.G.D.* sur \mathbb{R}^2 , de vitesse n avec pour fonction de taux $(x, y) \mapsto I(x) + J(y)$.

Remarque. L'énoncé précédent peut se généraliser en remplaçant $\frac{1}{n}$ par ϵ .

Voici des variations envisageables du principe de contraction :

Théorème 4.5.2 (Variations autour du principe de contraction). Soient (E, d) et (F, δ) deux espaces métriques. Supposons que $(\mu_\epsilon)_{\epsilon > 0}$ vérifie un *PGD* sur E avec I pour bonne fonction de taux et que J soit définie comme précédemment (cf. (4.5.2)) pour $f : \mathbb{E} \rightarrow F$ une application continue.

Soient $f_\epsilon : E \rightarrow F$, $\epsilon > 0$ une famille de fonctions mesurables telles que, pour tout $\eta > 0$,

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon \left(x \in E, \delta(f_\epsilon(x), f(x)) > \eta \right) = -\infty$$

alors $(\mu_\epsilon \circ f_\epsilon^{-1})_{\epsilon > 0}$ vérifie un *P.G.D.* sur F avec pour fonction de taux J .

Remarque. Le résultat précédent permettait de traiter un *P.G.D.* pour une famille de fonction. La remarque suivante permet d'enlever l'hypothèse de continuité portant sur f en procédant par approximation.

Supposons cette fois-ci $f : E \rightarrow F$ seulement mesurable et que $(\mu_\epsilon)_{\epsilon > 0}$ vérifie un *P.G.D.* sur E avec une bonne fonction de taux.

Supposons que nous ayons à disposition $f_n : E \rightarrow F$, $n \geq 1$ une suite de fonctions continues telles que, pour tout $\alpha \geq 0$,

$$\limsup_{n \rightarrow +\infty} \sup_{x \in \{I \leq \alpha\}} \delta(f_n(x), f(x)) = 0$$

et

$$\forall \eta > 0 \quad \lim_{n \rightarrow +\infty} \limsup_{\epsilon \rightarrow 0} \epsilon \log \mu_\epsilon \left(x \in E, \delta(f_n(x), f(x)) \right) = -\infty$$

alors $(\mu_\epsilon \circ f^{-1})_{\epsilon > 0}$ vérifie un *P.G.D.* avec J (définie comme précédemment) comme bonne fonction de taux.

4.6 Versions générales des Théorèmes de Cramer, Gärtner-Ellis et Sanov

Décrivons brièvement les énoncés des Théorèmes précédents en dimension infinie.

4.6.1 Théorème de Cramer

Soit $(E, \|\cdot\|)$ un espace de Banach réel séparable (par exemple, $E = C^0[0, 1]$ muni de $\|\cdot\|_\infty$). Considérons alors

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (E, \mathcal{B}(E))$$

ainsi que X_1, \dots, X_n des copies indépendantes de X et posons à nouveau $S_n = X_1 + \dots + X_n$ avec $n \geq 1$.

Théorème 4.6.1 (Loi forte des grands nombres). *Dans le cadre précédent nous avons le résultat suivant :*

$$\text{Si } \mathbb{E}[\|X\|] < +\infty \text{ alors } \frac{S_n}{n} \rightarrow \mathbb{E}[X] \text{ p.s. lorsque } n \rightarrow +\infty.$$

Il est alors naturel de s'interroger quant à l'existence d'un *P.G.D.* pour $(\frac{S_n}{n})_{n \geq 1}$. Les définitions de la log-Laplace Λ et de sa transformée de Fenchel-Legendre Λ^* s'étendent à ce nouveau cadre comme suit

$$\Lambda(\xi) = \log \mathbb{E}[e^{\langle \xi, X \rangle}] \quad \text{avec } \xi \in E^*, \quad \text{et } \Lambda^*(x) = \sup_{\xi \in E^*} [\langle \xi, x \rangle - \Lambda(\xi)] \quad \text{avec } x \in E$$

L'analogue du Théorème de Cramer s'énonce alors comme suit :

Théorème 4.6.2 (Cramer). *Si, pour tout $\alpha > 0$, $\mathbb{E}[e^{\alpha \|X\|}] < +\infty$ alors $(\frac{S_n}{n})_{n \geq 1}$ vérifie un *P.G.D.* avec pour bonne fonction de taux Λ^**

Remarque. 1. La partie délicate de la preuve consiste à établir la tension exponentielle.

2. Notons que si X est gaussien l'hypothèse d'intégrabilité du Théorème est satisfaite. De plus $\mathbb{P}(\frac{S_n}{n} \in A) = \mathbb{P}(\frac{1}{\sqrt{n}}X \in A)$ pour tout borélien A de E . Alors, au moins pour $\epsilon = \frac{1}{n}$, le Théorème de Cramer permet de retrouver le *P.G.D.* de Schilder pour des vecteurs gaussiens. Celui-ci fait intervenir la fonction

$$\Lambda^*(x) = \begin{cases} \frac{1}{2}\|x\|_{\mathcal{H}}^2 & \text{si } x \in \mathcal{H} \\ +\infty & \text{sinon.} \end{cases}$$

où $\mathcal{H} = \{x \in E; x = \mathbb{E}[\phi X] \text{ avec } \phi \in L^2\}$. Dans ce cas, si x est un élément de \mathcal{H} , pour tout $\xi \in E'$, $\langle \xi, X \rangle$ est une variable aléatoire gaussienne. C'est pourquoi, nous avons

$$\Lambda(\xi) = \log \mathbb{E}[e^{\langle \xi, X \rangle}] = \frac{\mathbb{E}[\langle \xi, X \rangle^2]}{2}$$

et, en conséquence,

$$\Lambda^*(x) = \sup_{\xi \in E'} [\mathbb{E}[\phi \langle \xi, X \rangle] - \frac{1}{2}\mathbb{E}[\langle \xi, X \rangle^2]] = \frac{1}{2}\mathbb{E}[\phi^2] = \frac{\|h\|_{\mathcal{H}}^2}{2}$$

Nous reviendrons sur ce résultat dans le chapitre suivant. Le Théorème de Schilder sera démontré à l'aide d'inégalités de concentration.

4.6.2 Théorème de Gärtner-Ellis

L'énoncé du théorème de Gärtner-Ellis, dans sa version générale, est similaire à sa version en dimension finie. Il faut néanmoins faire la supposition additionnelle de la présence de tension exponentielle. Nous renvoyons le lecteur à l'ouvrage [37] pour plus de détails.

4.6.3 Théorème de Sanov

Soit (E, d) un espace polonais et rappelons que $\mathcal{P}(E)$ désigne les mesures de probabilités sur E . Nous avons étudié au chapitre un les notions de convergence faible dans cet espace, ainsi que sa topologie. En particulier, nous avons caractérisé les compacts de $\mathcal{P}(E)$.

Soit X une variable aléatoire à valeurs dans E de loi $\mu \in \mathcal{P}(E)$. Considérons X_1, \dots, X_n des copies indépendantes de X . La mesure empirique $L_n; (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow \mathcal{P}(E)$ associée à μ est définie par

$$L_n(\omega) = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}(\omega)$$

Observons le fait suivant : si $f \in C_b^0(\Gamma, E)$ alors

$$\langle f, L_n \rangle = \frac{1}{n} \sum_{k=1}^n f(X_k) \rightarrow \mathbb{E}[f(X)] = \int_E f d\mu \quad \text{lorsque } n \rightarrow +\infty$$

d'après la loi forte des grands nombres (à conditions que $X \in L^1$). Le Théorème de Sanov correspond à un *P.G.D.* pour $(L_n)_{n \geq 1}$ et s'énonce comme suit :

Théorème 4.6.3 (Sanov). $(L_n)_{n \geq 1}$ satisfait un P.G.D. avec pour bonne fonction de taux

$$H(\nu|\mu) = \begin{cases} \int_E \log \frac{d\nu}{d\mu} d\nu & \text{si } \nu \ll \mu \\ +\infty & \text{sinon.} \end{cases}$$

$H(\nu|\mu)$ est appelée l'entropie relative de ν par rapport à μ .

Démonstration. La preuve repose sur le Théorème de Cramer/Gärtner-Ellis et nécessite l'identification de la transformée de Fenchel-Legendre avec l'entropie relative définie ci-dessus. Un point délicat est d'établir la tension exponentielle : $\forall M > 0$, il existe K un ensemble compact dans $\mathcal{P}(E)$ tel que

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}(L_n \notin K) \leq -M$$

La présence d'ensemble compacts de $\mathcal{P}(E)$ implique la présence du Théorème de Prokhorov. Celui-ci affirme, rappelons le, que de tels ensembles sont caractérisés par : K est un compact de $\mathcal{P}(E)$ si cet ensemble est fermé et uniformément tendu. Autrement dit, pour tout $\epsilon > 0$ il existe \tilde{K} un compact de E tel que

$$\nu(\tilde{K}) \geq 1 - \epsilon \quad \forall \nu \in K$$

□

4.7 Références historiques

Les notes historiques présentées ci-dessous proviennent des ouvrages [37, 1] auxquels nous renvoyons le lecteur pour plus de détails.

L'histoire des grandes déviations n'a pas suivi la présentation que nous venons de faire. Le théorème de Cramer fut étudié pour des variables aléatoires particulières par les statisticiens (dont Khinchin en 1929) de l'époque. En démontrant son théorème dans \mathbb{R} en 1938, Cramer fut celui qui introduit l'argument de changement de mesure. A l'origine, Cramer s'intéressait à ce phénomène de grande déviations pour modéliser des problèmes d'assurances. L'extension de son théorème à \mathbb{R}^d en 1952 est le fait de Chernoff; l'inégalité exponentielle permettant de traiter le cas d'ensembles fermés porte désormais son nom.

L'extension du théorème de Cramer à un cadre dépendant est obtenu (dans \mathbb{R}) par Plachky et Steinebach en 1975. Gärtner prolonge leurs travaux en 1977 pour traiter le cas $\mathcal{D}_{\mathbb{R}^d}$ et ce fut Ellis qui, en 1984, proposa la version que nous avons proposé dans ce cours.

La théorie moderne des grandes déviations et ses différentes applications doit essentiellement sa paternité aux travaux de Donsker/Varadhan et Freidlin/Wentzell. Néanmoins, ce sujet est beaucoup plus ancien et remonte aux débuts des années 1900. Il existe de nombreux ouvrages traitant de ce sujet, chacun d'entre eux se focalisant sur des thématiques particulières :

- En 1984, Varadhan propose un cours qui résumait de manière claire et concise les travaux obtenus jusqu'aux années 1984 dans cette branche des mathématiques

- Freidlein et Wentzell mettent en avant l'utilisation des P.G.D. en lien avec l'étude de petites perturbations aléatoires de systèmes dynamiques dans leur ouvrage « Random perturbations of Dynamical Systems » en 1984.
- En 1985, Ellis propose une application systématique des P.G.D. à des problèmes de mécaniques statistiques. A ce propos, dans ses travaux mêlant probabilités et mécanique statistique, Boltzman en 1870 introduit la notion d'entropie. Quant à elle, l'entropie relative (apparaissant dans le théorème de Sanov) est apparue dans les travaux de Kullback et Leibler en 1951 ; à cette époque, elle porte le nom de *nombre d'information de Kulback-Leibler* et s'avérant d'une grande utilité en statistiques.
- Les liens entre P.G.D. pour des mesures empiriques et les propriétés analytiques de semi groupes Markovien ont été développé dans l'ouvrage, publié en 1989, par Stroock et Deuschel. C'est d'ailleurs ce dernier qui introduit le terme de *tension exponentielle* de manière explicite.

En 1989, Donsker et Varadhan utilisent le principe de contraction pour étudier les chaînes de Markov de mesures empiriques. Bien longtemps avant l'existence d'une « théorie » générale des P.G.D. Sanov démontre son théorème en 1961. Le cadre abstrait (dans des espaces métriques polonais) du théorème de Cramer provient des travaux de Bahadur et Zabell en 1979. Dans le même article, ils obtiennent un P.G.D. pour la mesure empirique associée à une suite de variable aléatoire *i.i.d.* à valeurs dans un espace de Banach. Ce cadre d'étude est également considéré dans les travaux de Donsker/Varadhan (en 1975), ceux de Bolthausen (en 1984 et 1986 – 1987)) mais aussi ceux de de Acosta en 1997.

Exercice 1. Soit μ une mesure gaussienne centrée sur \mathbb{R}^d de matrice de covariance Γ et désignons par $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ un vecteur aléatoire tel que $\mathcal{L}(X) = \mu$.

Pour tout $\epsilon > 0$, nous définissons μ_ϵ comme étant l'image de μ par l'application $x \mapsto \sqrt{\epsilon}x$, autrement dit $\mu_\epsilon = \mathcal{L}(\sqrt{\epsilon}X)$.

1. Montrer que μ_ϵ vérifie un PGD sur \mathbb{R}^d . Décrire la fonction de taux et montrer qu'il s'agit d'une bonne fonction de taux.
2. Montrer que $\lim_{t \rightarrow +\infty} \frac{1}{t^2} \log \mathbb{P}(\|X\|_2 \geq t)$ existe. Nous noterons une telle limite par $-b$.
3. Si Γ est inversible, montrer que $\frac{1}{2b} = \sup_{\|x\|_2=1} \{\langle x, \Gamma x \rangle\}$.
4. Montrer que le résultat précédent reste satisfait si Γ n'est pas inversible.

