Grandes déviations autonormalisées pour des chaînes de Markov

Mathieu Faure

Université de Marne-La-Vallée
Laboratoire d'Analyse et de Mathématiques Appliquées
CNRS UMR 8050

Thèse

présentée

pour obtenir le titre de

DOCTEUR de l'UNIVERSITÉ de MARNE-LA-VALLÉE

Spécialité : Mathématiques

par

Mathieu Faure

Grandes déviations autonormalisées pour des chaînes de Markov

Soutenue le 20 décembre 2002
devant le jury composé des professeurs :

Paolo Baldi	Rapporteur
Bernard Bercu	Rapporteur
Marie Duflo	Directrice de thèse
Damien Lamberton	Examinateur
Christian Léonard	Examinateur
Alain Rouault	Examinateur

Remerciements

Il est temps pour moi d'exprimer toute ma gratitude à Marie DUFLO qui a encadré mon stage de DEA puis accepté de diriger cette thèse. Sa disponibilité et sa patience ont été des éléments décisifs à la réalisation de ce travail.

Je suis très reconnaissant à Paolo BALDI et Bernard BERCU d'avoir accepté de faire un rapport sur cette thèse. Je les remercie particulièrement de l'attention minutieuse qu'ils ont porté à mon travail.

Christian LÉONARD et Alain ROUAULT me font l'honneur de participer au jury ; pour cette raison, je les remercie chaleureusement.

Merci à Damien LAMBERTON pour son aide en DEA et pour m'avoir encouragé à entreprendre une thèse. Je tiens à le remercier également pour sa présence dans le jury.

Je voudrais également remercier Julien WORMS pour son aide ainsi que pour les précieux conseils qu'il m'a donnés lorsque j'étais encore en DEA puis tout au long de la thèse.

Je remercie tous les membres de l'équipe d'Analyse et de Mathématiques Appliquées de Marne-La-Vallée pour l'enseignement que j'ai reçu depuis 6 ans, l'accueil dans le cadre de cette thèse et pour m'avoir permis de faire mes premiers pas dans l'enseignement supérieur.

Enfin, je tiens à remercier mon entourage, familles et amis, pour m'avoir soutenu et encouragé tout au long de la réalisation de cette thèse.

Table des matières

Avant propos **i**

Présentation générale . . . i
Résumé des chapitres . . . iii
Points de vue non abordés et perspectives . . . vii
Notations et conventions . . . viii

1 PGD partiel, autonormalisation **1**

1.1 Introduction . . . 1
1.1.1 Objectifs de ce chapitre . . . 1
1.1.2 Plan . . . 1
1.2 Principes de grandes déviations et tension exponentielle . . . 2
1.2.1 Définitions . . . 2
1.2.2 Tension exponentielle . . . 2
1.2.3 PGD et tension exponentielle partiels . . . 3
1.3 Dualités . . . 8
1.3.1 PGD et principe de Laplace . . . 8
1.3.2 PGD vague-convexe . . . 10
1.3.3 Identification d'un taux convexe . . . 12
1.3.4 Exemples classiques i.i.d. . . . 13
1.4 Autonormalisation . . . 15
1.4.1 De Cramér à Shao . . . 15
1.4.2 Les classes de Chernov . . . 16
1.4.3 PGD autonormalisées . . . 17
1.4.4 Tension exponentielle partielle pour les classes de Chernov . . . 18
1.4.5 Preuves des résultats de 1.4.2 et 1.4.4 . . . 19

2 Autour de Donsker-Varadhan **22**

2.1 Introduction . . . 22
2.1.1 Objectifs de ce chapitre . . . 22
2.1.2 Plan . . . 22
2.1.3 Notations . . . 23
2.2 Taux markoviens . . . 23
2.2.1 Kullback et les principes variationnels . . . 23
2.2.2 Kullback sur un espace produit . . . 25
2.2.3 Taux de Donsker-Varadhan sur un espace polonais . . . 26
2.2.4 Transitions fellériennes et presque-fellériennes . . . 27

2.2.5 Taux spectral et taux de Donsker-Varadhan 28
2.2.6 Références 30
2.2.7 Preuves des résultats de 2.2.2 et 2.2.5 30
2.3 PGD markoviens supérieurs pour les lois empiriques 35
2.3.1 PGD vague supérieur uniforme 35
2.3.2 PGD étroit supérieur 36
2.3.3 Références 37
2.3.4 Preuves 37
2.4 Commentaires, exemples et contre-exemples 42
2.4.1 Taux spectral et oubli de la loi initiale 42
2.4.2 Sur la condition de rappel exponentiel et la stabilité 44

3 Transports de PGD **48**
3.1 Introduction 48
3.1.1 Position du problème 48
3.1.2 Plan de ce chapitre 48
3.1.3 Rappels et notations 49
3.2 Transport pour des couples équilibrés 50
3.2.1 Les couples équilibrés 51
3.2.2 Démonstration du théorème 3.1 51
3.3 Transport pour des couples exponentiellement équilibrés 52
3.3.1 Approximation exponentielle 52
3.3.2 Transport d'un PGD vague supérieur empirique pour un couple exponentiellement équilibré 54
3.3.3 Transport d'un PGD inférieur 54
3.3.4 Références 55
3.3.5 Démonstrations 55
3.4 Transports i.i.d. et markoviens 58
3.4.1 Transports i.i.d. 58
3.4.2 Transports markoviens 59
3.4.3 Transports autorégressifs 62
3.4.4 Références 63
3.4.5 Preuves des résultats de 3.3 63
3.5 PGD inférieurs markoviens 66
3.5.1 A propos de l'irréductibilité 66
3.5.2 Résultats connus sur les PGD inférieurs spectraux 67
3.5.3 Vers un principe de Laplace inférieur de fonctionnelle limite conjuguée au taux de Donsker-Varadhan 69
3.5.4 Conditions assurant la validité du principe de Laplace inférieur . . 72
3.5.5 Références 77

4 PGD autonormalisés **78**
4.1 Introduction 78
4.1.1 Objectifs de ce chapitre 78
4.1.2 Plan du chapitre 78
4.2 Méthode générale pour l'autonormalisation 79
4.2.1 PGD vague pour un couple exponentiellement équilibré 79

4.2.2 PGD autonormalisé et Chernov 80
4.2.3 PGD troué en 0 81
4.2.4 Preuves des résultats de 4.2.1 87
4.3 PGD empirique étroit et PGD autonormalisé 91
4.3.1 Processus dont la loi empirique satisfait un PGD étroit 91
4.3.2 Autonormalisation 92
4.3.3 Quelques exemples 94
4.4 PGD autonormalisés pour des modèles autorégressifs 95
4.4.1 Suite régressive à bruit dont le carré vérifie la condition de cramér renforcée 96
4.4.2 Application aux modèles AR(1) linéaires 98
4.4.3 Fonctionnelles additives martingales markoviennes 102

Avant-propos

Présentation générale

L'objectif de cette thèse est l'obtention de principes de grandes déviations autonormalisés, essentiellement pour des modèles markoviens (en particulier des modèles autorégressifs linéaires et fonctionnels).
Avant de rentrer dans les détails, il est essentiel de clarifier ce que l'on entend par résultats "autonormalisés" et je vais tenter préciser ce qui m'a motivé dans cette approche particulière de la théorie des grandes déviations.

Où se situe-t-on dans la recherche de théorèmes limites pour un modèle probabiliste ? Les résultats les plus fréquemment convoités sont des lois des grands nombres, des théorèmes de limite centrale (TLC), des lois du logarithme itéré et des principes de grandes déviations (PGD). La source de toute l'asymptotique probabiliste d'une suite de variables aléatoires réelles $(X_n)_n$ supposées indépendantes et identiquement distribuées (i.i.d) est l'étude de résultats de ce type pour la suite $(X_1 + ... + X_n)/n$ sous diverses conditions portant sur la loi de X_1.

- Loi des grands nombres si $\mathbb{E}|X_1| < \infty$;
- théorème de limite centrale si $\mathbb{E}|X_1|^2 < \infty$;
- principe de grandes déviations sous l'hypothèse de Cramér $\mathbb{E}\left(e^{t|X_1|}\right) < \infty$ pour un $t > 0$.

Naturellement, ceci a guidé un nombre important de recherches probabilistes en vue d'apporter des extensions à ces résultats. Une démarche naturelle est d'essayer d'affaiblir les hypothèses ; puis de chercher des résultats asymptotiques non plus pour la moyenne empirique $((X_1+...X_n)/n)_n$ mais pour une suite *autonormalisée* $((X_1 + ...X_n)/T_n(X_1, ..., X_n))_n$, où $(T_n(X_1, ..., X_n))_n$ est une suite de variables aléatoires positives et croissantes.
J'ai fait la découverte de ce type de résultats autonormalisés lors de mon stage de DEA, quelques temps après m'être initié à la théorie des grandes déviations. Q.M Shao avait alors montré tout récemment des PGD très peu exigeants en termes d'hypothèses dans le cadre i.i.d. A titre d'exemple, il obtenait pour une suite de variables aléatoires réelles $(X_n)_n$ i.i.d. avec, pour $p > 1$, $\mathbb{E}(X_1) = 0$ ou $\mathbb{E}(|X_1|^p) = \infty$, une limite du type suivant :

$$\forall r > 0,\ \lim_n \frac{1}{n} \log \mathbb{P}\left(\frac{X_1 + .. + X_n}{n^{1-1/p}\left(|X_1|^p + ... + |X_n|^p\right)^{1/p}} \geq r \right) = -K(r) < 0.$$

Si l'on compare cette relation au théorème de Chernov, il est intéressant de constater que le passage de l'étude de la suite des moyennes empiriques à une suite autonormalisée permet d'obtenir des résultats du même type mais sans l'hypothèse de Cramér. Cette dernière est une condition assez forte, qui n'est pas vérifiée dans bien des cas.

Dès lors, une motivation essentielle lors de l'élaboration de cette thèse a été d'étendre ce type de "bornes de Chernov" en sortant du contexte i.i.d, ce qui m'a conduit à travailler avec des modèles markoviens.
Dans la littérature, l'étude de suites autonormalisées est courante dans le cadre i.i.d, notamment pour l'introduction de statistiques telles que celle de Student.
Elle l'est aussi pour des modèles markoviens et des martingales ; en voici quelques exemples.
Une martingale de carré intégrable $(M_n)_n$ dont le processus croissant prévisible est $(< M >_n)_n$ conduit à des résultats portant sur la suite autonormalisée $(M_n / < M >_n)_n$.
De même, pour un modèle autorégressif linéaire d'ordre 1

$$X_{n+1} = \theta X_n + \varepsilon_{n+1},$$

l'estimateur des moindres carrés donné par $\left((\sum_{i=1}^n X_{i-1}X_i)/(\sum_{i=1}^n X_{i-1}^2)\right)_n$ est une suite autonormalisée qui converge presque-sûrement vers le paramètre θ dès que le bruit $(\varepsilon_n)_n$ est i.i.d. de carré intégrable et non nul, avec des vitesses en loi et presque-sûres bien connues.

Comme je l'ai évoqué plus haut, la démarche essentielle de cette thèse s'est orientée vers les chaînes de Markov sur un espace d'état polonais, ce cadre se rapprochant le plus du cas i.i.d. Pour cela, je me suis appuyé sur une analyse minutieuse des résultats de Donsker et Varadhan, avec des variantes venues de plusieurs autres auteurs et quelques ajouts originaux. On obtient sous des conditions très générales, un principe de grandes déviations vague ou étroit pour les lois empiriques qui généralise le théorème de Sanov au cas markovien.
Transposant la méthode de Cramér, plusieurs auteurs, notamment De Acosta, obtiennent aussi des PGD supérieurs ou inférieurs relatifs aux fonctionnelles additives des chaînes de Markov. Cependant, les "taux spectraux" associés à ces méthodes ne m'ont pas semblés propices à une transposition des travaux de Dembo et Shao qui s'appuyaient dans le cas i.i.d. sur la transformée de Cramér ; l'obtention par cette méthode d'un PGD vague assez général me paraît en effet hors de portée. J'ai donc choisi de m'appuyer sur les PGD relatifs aux lois empiriques et de tenter de les transposer vers les fonctionnelles additives en vue de PGD vagues et autonormalisés.

A ce stade, j'explore le cadre général d'une suite de probabilités aléatoires $(L_n)_n$ sur un espace polonais satisfaisant un PGD (vague ou étroit) et je propose un cadre général adapté à l'autonormalisation (en particulier propice au "transport" de PGD vagues), celui des "couples équilibrés" de fonctions. J'obtiens alors un PGD vague puis un PGD autonormalisé sous des conditions de tension exponentielle.

Finalement, cette thèse aboutit à deux cas de figure pour l'obtention de PGD autonormalisés.
Cas 1 Pour les processus dont la suite des lois empiriques satisfait un PGD étroit (ce qui correspond heuristiquement à un cadre "exponentiellement stable"), on transpose aux "couples équilibrés" des résultats analogues à ceux de Dembo et Shao.

Cas 2 Pour des processus dont la suite des lois empiriques satisfait seulement un PGD vague, la même méthode s'applique avec plus de difficulté. Cependant, l'étude des séries régressives associées à un modèle markovien autorégressif montre que notre méthode permet aussi d'obtenir des principes de grandes déviations autonormalisés sans hypothèse de stabilité.

Résumé des chapitres

Chapitre 1. PGD partiel, autonormalisation

Ce chapitre a pour objet d'introduire les notions spécifiques aux PGD partiels, indispensables pour pouvoir faire l'étude de PGD autonormalisés. Néanmoins, il met aussi en place la plupart des notions classiques liées à la théorie des grandes déviations dont nous aurons besoin dans la suite. E désigne un espace topologique, un taux étroit est une fonction définie sur E, à valeurs dans $[0, \infty]$ et à épigraphes $\{x \in E \mid I(x) \leq M\}$ ($M > 0$) compacts ; on note $I(A) = \inf_{x \in A} I(x)$. Une suite de probabilités $(\mu_n)_n$ sur E vérifie un PGD supérieur de vitesse $(v_n)_n$ et de taux étroit I si l'on a

$$-I(F) \geq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F), \text{ pour tout fermé } F \text{ de } E.$$

Elle vérifie un PGD inférieur de vitesse $(v_n)_n$ et de taux I si l'on a

$$-I(G) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G), \text{ pour tout ouvert } G \text{ de } E.$$

Un taux vague étant une fonction définie sur E, à valeurs dans $[0, \infty]$ et à épigraphes fermés, $(\mu_n)_n$ vérifie un PGD vague de vitesse $(v_n)_n$ et de taux vague I si elle vérifie la borne inférieure mais si la borne supérieure n'est valable que pour les parties compactes de E.
Enfin, $(\mu_n)_n$ vérifie un PGD partiel de vitesse $(v_n)_n$ et de taux vague I relativement à une partie $\mathcal{S}$ d'ensembles mesurables si elle vérifie un PGD inférieur et si

$$-I(\overline{A}) \geq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(A), \text{ pour tout élément } A \text{ de } \mathcal{S}.$$

Dans la section 1.2, il s'agit essentiellement d'adapter aux PGD partiels des méthodes usuelles pour les PGD, en présentant notamment le rôle de la tension exponentielle. Ainsi, un PGD vague, associé à une "tension exponentielle partielle" conduit à un PGD partiel. Ce schéma de démonstration sera le plus utilisé pour obtenir des PGD partiels.

La section 1.3 reprend les méthodes classiques des grandes déviations, notamment les dualités entre les fonctionnelles de Laplace et PGD ou entre fonctionnelles convexes et PGD vague-convexe. Ces méthodes sont le plus souvent présentées dans le cadre de "vrais" principes de grandes déviations associés à de bonnes fonctions de taux alors que nous parlerons ici de principes de grandes déviations "partiels", avec des "taux" ne possédant pas toutes les propriétés d'une bonne fonction de taux. Nous les rappellerons donc en tentant à chaque étape de les placer dans un cadre aussi large que possible.

La tension exponentielle partielle relative à la classe pour laquelle on veut avoir un PGD partiel est la clef et souvent la partie délicate pour obtenir un PGD autonormalisé. La section 1.4 est consacrée à des critères permettant d'assurer qu'une suite de lois est exponentiellement tendue. Suivant les traces de Dembo et Shao, on introduit une classe de boréliens particulièrement adaptée à l'étude de PGD autonormalisés que l'on appelle "classe de Chernov". On donne ensuite quelques critères de tension exponentielle partielle issus pour la plupart de ceux de Dembo et Shao ([DemSha98a] et [DemSha98b]).

Chapitre 2. Autour de Donkser-Varadhan

L'objectif de ce chapitre est de reprendre les travaux de Donsker-Varadhan (notamment [DonVar75a], DonVar76]), d'éclairer ou généraliser certains aspects et de les situer par rapport à des travaux ultérieurs, notamment ceux de De Acosta (comme [Aco85a], [Aco88], [Aco90] ou [AcoNey98])

En 2.2, on commence par rappeler les principes variationnels associés à l'information de Kullback $K(\nu \mid \mu)$ d'une probabilité ν relativement à une autre μ.
Puis, E étant un espace polonais, on considère une probabilité de transition p de E vers E. Lorsque ζ est une probabilité définie sur E^2, de marginales ζ^1 et ζ^2, une distance de ζ à p peut être mesurée par le taux "bidimensionnel"

$$I(\zeta) \stackrel{\triangle}{=} \begin{cases} K\left(\zeta \mid \zeta^1 \oslash p\right) & \text{si } \zeta^1 = \zeta^2 \\ \infty & \text{sinon} \end{cases}$$

(où $\zeta^1 \oslash p(\Gamma) = \int \mathbb{I}_\Gamma(x,y) p(x,dy)\zeta^1(dx)$). Le "taux de Donsker-Varadhan" usuel est le taux projeté défini pour une probabilité ν sur E par

$$\begin{aligned} I^1(\nu) &= \inf\left\{I(\zeta) \mid \zeta \text{ probabilité sur } E^2,\ \zeta^1 = \zeta^2 = \nu\right\} \\ &= \sup\left\{\int \left(g(x) - \log\left(\int p(x,dy)e^{g(y)}\right)\right)\nu(dx) \mid g \text{ continue, bornée sur } E\right\}. \end{aligned}$$

Cette dernière relation vient des travaux de Donsker-Varadhan ; nous en donnons une preuve plus simple, bénéficiant d'un résultat de Strassen. On reprend l'étude de ces taux sans hypothèse a priori sur la probabilité de transition p mais en faisant apparaître le rôle de $\mathcal{D}(p)$, ensemble de discontinuité de l'application $x \mapsto p(x, \bullet)$ pour la convergence étroite.
Une comparaison avec le taux de De Acosta achève cette partie.

Dans la section suivante, des PGD supérieurs relatifs à la suite des lois empiriques associées à une chaîne de Markov de transition p assez régulière sont établis avec les taux de Donsker-Varadhan. Dans les deux cas, on donne une version pour la loi empirique "bidimensionnelle" $\left(\left(\sum_{i=1}^n \delta_{X_{i-1},X_i}\right)/n\right)_n$ avec le taux I, puis une version unidimensionnelle pour $\left(\left(\sum_{i=1}^n \delta_{X_{i-1}}\right)/n\right)_n$ (resp. $\left(\left(\sum_{i=1}^n \delta_{X_i}\right)/n\right)_n$) avec le taux I^1. Le "taux bidimensionnel" étant, on l'a vu, plus intuitif que le taux projeté de Donsker-Varadhan, il est en effet naturel de se pencher d'abord sur les lois bidimensionnelles contrairement à ce qui est d'usage. Le PGD vague (théorème 2.3) ne nécessite rien de plus que la régularité de la transition p alors que le PGD étroit (théorème 2.4) utilise une "condition de rappel exponentiel" qui correspond heuristiquement à une certaine stabilité exponentielle du modèle. Divers changements ont été effectués comparativement aux énoncés originaux de Donsker et Varadhan. Pour le PGD étroit, l'hypothèse de régularité portant sur la transition est affaiblie de fellérienne à "presque fellérienne". De même on donne un PGD supérieur uniforme sur des ensembles de lois initiales, alors que cela avait été établi sur des ensembles d'états initiaux.

En section 2.4, quelques commentaires agrémentés d'exemples et contre-exemples concluent ce chapitre, afin d'éclaircir plusieurs points qui peuvent soulever des problèmes.
En particulier, l'obtention d'un PGD supérieur pose toujours la question de l'optimalité du taux quand il n'est pas accompagné d'un PGD inférieur de même taux. On montre que les PGD énoncés dans la section précédente ne sont pas toujours optimaux.
Puis, on discute de la condition de rappel exponentiel à l'aide d'un exemple autorégressif fonctionnel pour lequel on distingue plusieurs cas permettant d'établir que cette condition est vérifiée.

Chapitre 3. Transports de PGD : des lois empiriques vers les moyennes empiriques

Dans le chapitre 2, des PGD supérieurs ayant été prouvés pour la suite des lois empiriques d'une chaîne de Markov, on est tenté de les transporter vers des PGD portant sur des moyennes empiriques du type $((\sum_{i=1}^{n} f(X_i)/n)_n$. Ce chapitre propose d'en faire une étude dans un cadre plus général.
Soit $(L_n)_n$ une suite de variables aléatoires à valeurs dans l'espace des probabilités sur (E,d) polonais et f une fonction continue de E dans un espace de Banach B. Si J est un taux sur $\mathcal{P}(E)$, son image par f est le taux image défini de la manière suivante :

$$h_f(x) = \inf\left\{ J(\nu) \mid \ \nu \in \mathcal{P}(E), \ \int \|f\| d\nu < \infty \text{ et} \int f d\nu = x \right\}.$$

Tout au long du chapitre, on tente de répondre aux questions suivantes :

*(L_n) satisfaisant un PGD supérieur vague de taux J, peut on obtenir un PGD supérieur vague pour $(\int f dL_n)$ dont le taux serait h_f ou, si h_f n'est pas s.c.i, sa régularisée s.c.i ?
La réponse étant en général négative, peut on donner une réponse partielle à ce problème ?*

La même question pour la partie inférieure est prise en compte mais elle est moins importante dans le cadre markovien où les résultats inférieurs seront obtenus différemment. Bien sûr, par le principe de contraction, si la suite $(L_n)_n$ satisfait un PGD étroit avec une bonne fonction de taux J, la réponse aux deux questions est positive pour toute fonction $f \in \mathcal{C}_b(E)$. Il s'agit de trouver d'autres classes de fonctions pour lesquelles le transport fonctionne, en sortant du cadre des fonctions bornées et en n'utilisant qu'un PGD vague sur $(L_n)_n$.

Dans la section 3.2, on introduit une classe de fonctions naturellement adaptée au transport de PGD vague pour les lois empiriques au PGD vague pour les fonctionnelles additives. Il s'agit des "couples équilibrés" de fonctions (f, V) où V est une "fonction de Lyapounov" et f est dominée d'une certaine manière par V. On obtient des bornes supérieures du type suivant, pour tout fermé Γ,

$$\limsup_n \frac{1}{v_n} \log \mathbb{P}\left(\int f dL_n \in \Gamma,\ \int V dL_n \leq R\right) \leq -h_{f,V}\left(\Gamma \times [0,R]\right).$$

Des résultats du même type sont obtenus pour des classes de fonctions plus vastes ; les couples exponentiellement équilibrés. Il s'agit de couples de fonctions convenablement approchés par des familles de couples équilibrés. Une condition de "bonne" approximation permet de retrouver des résultats tout à fait identiques à ceux des couples équilibrés. De même, on a un résultat de transport d'un PGD inférieur pour des fonctions approchées par des familles de fonctions continues et bornées (par exemple les fonctions tronquées).

A titre d'applications, on utilise ces résultats pour retrouver des résultats bien connus dans les cadres i.i.d. ou markoviens, en section 3.4. Cela nous conduit aussi à un résultat sur le modèle autoregressif fonctionnel qui servira dans le chapitre 4.

Dans la section 3.5, on cherche à prouver un PGD inférieur markovien pour une suite de fonctionnelles additives. Cette partie sort quelque peu du cadre du chapitre 3 puisqu'il ne s'agit plus de transport. De Acosta et Ney [AcoNey98] obtiennent un PGD beaucoup plus général mais avec un taux spectral que l'on ne sait pas comparer avec le taux image. Les techniques utilisées pour prouver ces résultats inférieurs sont celles de Dupuis et Ellis ([DupEll97]). Aucune hypothèse n'est nécessaire sur la fonctionnelle mais des hypothèses d'irréductibilité sont indispensables pour obtenir les PGD inférieurs. On obtient finalement un PGD inférieur avec le taux image sous des hypothèses assez fortes ; une hypothèse introduite dans le cas des lois empiriques par Dupuis et Ellis dans le cas récurrent et une autre valable sans hypothèse de récurrence.

Chapitre 4. PGD autonormalisés

En suivant notre méthode, l'obtention de PGD autonormalisés passe par les PGD partiels relativement à des classes particulières, les "classes de Chernov" définies en 1.4, permettant le passage au quotient. Après avoir traduit les résultats du chapitre 3 en termes de PGD vagues pour les couples équilibrés, il reste à étudier la tension exponentielle partielle relative aux classes de Chernov. C'est ce que nous faisons dans diverses situations.

Dans la section 4.2, on traduit en premier lieu les résultats de transport obtenus dans les parties 3.2 et 3.3 en termes de PGD vagues supérieurs pour la suite $(\int f dL_n, \int V dL_n)_n$, lorsque (f, V) est un couple équilibré ou exponentiellement équilibré. Puis, on examine les conséquences d'un éventuel PGD partiel supérieur pour la suite $(\int f dL_n, \int V dL_n)_n$ relatif à une "classe de Chernov" associée à une fonction de Lyapounov U. Comme on l'espérait compte tenu de la manière dont les classes de Chernov ont été définies, on aboutit dans un "cadre favorable" à un PGD supérieur portant sur le quotient $\left(U\left(\int f dL_n\right) / \int V dL_n\right)_n$. Si $(\int f dL_n, \int V dL_n)_n$ vérifie aussi un PGD inférieur, il est possible d'obtenir un PGD pour le quotient.

En 4.3, on applique cette méthode au cadre général d'une suite de probabilités aléatoires satisfaisant un PGD étroit. C'est le cas, par exemple de la suite des lois empiriques d'une chaîne de Markov sous certaines conditions de "stabilité exponentielle". Dans ce cadre, il est facile d'obtenir un critère de tension exponentielle partielle de $(\int f dL_n, \int V dL_n)_n$ relative à une classe de Chernov et une fonction V bien choisies. Ainsi, quand la fonction f est à valeurs dans $\mathbb{R}^d$, on obtient des PGD partiels pour le quotient $\left(\| \int f dL_n \| / \left(\int \|f\|^p dL_n\right)^{1/p}\right)_n$.

Des modèles autorégressifs sont étudiés en 4.4. Puisque nous ne voulons pas faire d'hypothèse de stabilité, la partie précédente ne s'applique pas. On prouve cependant un PGD autonormalisé général pour certaines fonctionnelles additives martingales. En particulier, si l'on considère un modèle autorégressif linéaire

$$X_{n+1} = \theta X_n + \varepsilon_{n+1},$$

on peut appliquer ce qui précède à un estimateur de Yule-Walker (ou à un autre que nous proposons).

Enfin, nous présentons dans la section 4.5 le PGD de Heck-Maaouia ([HecMaa01]) relatif aux lois empiriques logarithmiques de $(M_n/\sqrt{n})_n$ ($(M_n)_n$ étant une fonctionnelle additive martingale d'une chaîne de Markov récurrente) et on propose de montrer comment les résultats de la section 4.3 s'appliquent dans ce cadre à l'obtention de PGD autonormalisés.

Points de vue non abordés et perspectives

Dans cette thèse, certains aspects importants de la théorie des grandes déviations n'ont pas été abordés. Entre autres, on ne parle pas de déviations exactes ou encore de moyennes déviations.

Déviations exactes ou précises Nous n'avons pas abordé les "inégalités de déviation", c'est à dire des majorations exactes avec des majorants décroissant exponentiellement vite. Sur un espace fini, les grandes déviations et les majorations exactes datent de Miller [Mil61] ; Lezaud obtient de bonnes majorations de type Chernov ([Lez98] et [Lez01]).
Par ailleurs, les inégalités de déviation entrent dans le vaste champ des inégalités de concentration exploré surtout dans le cadre i.i.d. Citons par exemple quelques articles qui rejoignent notre contexte : Bercu, Gassiat et Rio [BerGasRio02] ont prouvé des inégalités de déviation autonormalisées dans le cadre i.i.d ; Samson [Sam00] et Marton [Mar96] obtiennent des inégalités de déviations dans un cadre markovien récurrent.
Nous n'avons pas non plus abordé les grandes déviations précises, c'est à dire les PGD dont on sait évaluer la vitesse de convergence Dans le cadre markovien, Balaji et Meyn [BalMey] puis Meyn et Kontoyannis [KonMey02] obtiennent des PGD précis avec le taux spectral sous des hypothèses de récurrence très fortes dont nous cherchons justement à nous passer en abordant les PGD autonormalisés.

Moyennes déviations et moyennes déviations autonormalisées Dans le cas i.i.d., Dembo et Shao ont aussi exploré le domaine des moyennes déviations dans le cadre de l'autonormalisation [DemSha98a]. Il est naturel d'envisager des extensions markoviennes,

tout comme on l'a fait pour les résultats de grandes déviations. D'ailleurs, pour le modèle autoregressif linéaire AR(1) gaussien, Worms [Wor01] obtient un principe de déviations modérées supérieur pour l'estimateur des moindres carrés.
On connait, sous des hypothèses de stabilité, des principes de moyennes déviations sur les martingales et leur application à certaines fonctionnelles additives martingales de chaînes de Markov (voir [Wor99], [Wor00b] et [Wor01]). Pour un processus irréductible et sous une condition de Darling-Kac, Xia Chen [Chen01] obtient un résultat de moyennes déviations valable sans condition de stabilité. Des transpositions à des moyennes déviations autonormalisées mériteraient d'être considérées.

Extensions en perspective Dans le chapitre 4, nous obtenons un PGD autonormalisé valable pour une classe de modèles "exponentiellement stables" très générale. Cependant, nous n'avons bien exploré cette situation que pour les chaînes de Markov ; il reste à étudier de plus près la traduction de notre théorie aux processus de Markov à temps continu ou aux processus gaussiens.
Comme l'a montré Shao dans le cas i.i.d, les PGD autonormalisés sont bien adaptés à la construction de tests. Une étude précise de certains tests issus de nos résultats (notamment dans le cadre des modèles autorégressifs stables, instables ou explosifs) reste à faire.
Enfin, il serait intéressant d'obtenir d'autres PGD autonormalisés pour des fonctionnelles additives martingales de chaînes ou de processus de Markov, sans hypothèse de stabilité exponentielle.

Notations et conventions

Espaces mesurables, métriques, polonais

Etant donné un espace mesurable $(E, \mathcal{E})$ dont la tribu $\mathcal{E}$ sera souvent sous-entendue, on introduit les espaces suivants, relatifs à E : $\mathcal{B}_b(E)$ désigne l'ensemble des fonctions réelles boréliennes bornées définies sur E, $\mathcal{P}(E)$ l'ensemble des probabilités sur E et $\mathcal{M}_1(E)$ l'ensemble des mesures positives de masse totale inférieure ou égale à un sur cet espace. On munit l'espace $\mathcal{B}_b(E)$ de la norme infinie $\|f\|_\infty = \sup_x |f(x)|$.

Un espace métrique E métrisable par une distance d est noté (E, d) ; la tribu utilisée (sous-entendue) sera toujours la tribu borélienne. La boule ouverte (resp. fermée) de centre x et de rayon r est notée $B(x, r)$ (resp. $\overline{B}(x, r)$). Enfin, $\mathcal{C}_b(E)$ est l'ensemble des fonctions réelles continues bornées définies sur E, $\mathcal{LB}(E)$ l'ensemble des fonctions lipschitziennes bornées, muni de la norme

$$\|f\|_{\mathcal{LB}} = \sup(\|f\|_\infty, \|f\|_L) \;\text{ avec }\; \|f\|_L = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)},$$

et $\mathcal{B}_{maj}(E)$ (resp. $\mathcal{C}_{maj}(E)$) l'ensemble des fonctions boréliennes (resp. continues) définies sur E et majorées par une constante.

La théorie porte essentiellement sur un espace E polonais, c'est à dire pourvu d'une topologie pour laquelle il est séparable, et métrisable avec une distance d pour laquelle il est complet. L'expression "soit (E,d) espace polonais" sous-entend que d est une telle métrique. Selon la théorie de Prokhorov, on sait que l'espace $\mathcal{P}(E)$ est un espace séparable complet, par exemple pour les distances de Prokhorov et de Dudley. Rappelons la définition de la distance de Dudley qui sera parfois utilisée :

$$d_{Dudley}(\nu,\mu) = \sup\left\{\left|\int f d\nu - \int f d\mu\right| \mid f \in \mathcal{LB}(E),\ \|f\|_{\mathcal{LB}} \leq 1\right\}.$$

Lorsque l'on parle de $\mathbb{R}^d$ ou d'un espace de Banach B, il sera sous-entendu que ces espaces sont munis de leur tribu borélienne et de la distance associée à la norme $\|\cdot\|$ pour laquelle ils sont complets.

Sauf avis contraire, dans toute la thèse, on se place sur un espace polonais (E,d).

Variables aléatoires

Etant donné un espace de probabilité $(\Omega,\mathcal{A},\mathbb{P})$, une *variable aléatoire X à valeurs dans $(E,\mathcal{E})$ (resp. (E,d))* est une fonction mesurable de $(\Omega,\mathcal{A})$ dans $(E,\mathcal{E})$(resp. dans E muni de sa tribu borélienne). La loi de X est $X(\mathbb{P}) \in \mathcal{P}(E)$ mesure image de $\mathbb{P}$ par X. Dans la plupart des cas, les résultats évoqués ne portent que sur les lois. Peu importe dès lors de préciser sur quel espace de probabilité la variable aléatoire X est définie, il s'agit d'une réalisation parmi d'autres de sa loi. C'est pourquoi, (E,d) étant un espace métrique, on parle le plus souvent de X *variable aléatoire à valeurs dans* E en sous-entendant $(\Omega,\mathcal{A},\mathbb{P})$ et E *muni de sa tribu borélienne.*
Une suite $(Z_n)_n$ de variables aléatoires à valeurs dans (E,d) est une suite **indépendante et identiquement distribuée** (abrégé i.i.d.) si ces variables aléatoires sont indépendantes et s'il existe une probabilité μ sur (E,d) pour laquelle $\mu = X_n(\mathbb{P})$, quel que soit n. Un **bruit blanc de loi** μ est une suite i.i.d. de loi μ, supposée centrée.
Toute propriété Π portant sur les suites $(\mu_n)_n$ de probabilités sur E a son image ; on dira qu'une suite $(X_n)_n$ de variables aléatoires à valeurs dans E satisfait Π si la suite des lois $(X_n(\mathbb{P}))_n$ la satisfait.

Divers

- Soit une fonction $I : E \to [0,\infty]$. Pour tout $A \subset E$, on note
$$I(A) = \inf_{x\in A} I(x).$$
Et on associe à I sa **régularisée s.c.i.** I^{sci} :
$$I^{sci}(x) \stackrel{\triangle}{=} \lim_{r\to 0} \uparrow I\left(B\left(x,r\right)\right) = \lim_{r\to 0} \uparrow I\left(\overline{B}\left(x,r\right)\right).$$
- Une fonction $\rho : \mathbb{R}_+ \to [0,\infty]$ est dite **surlinéaire** si elle est borélienne et telle que $\lim_{t\to\infty} \rho(t)/t = \infty$.
- Une fonction de Lyapounov U définie sur (E,d) est une fonction borélienne positive dont les **épigraphes** $\{x \mid U(x) \leq R\}$ sont relativement compacts.
- On utilisera, tout au long de la thèse, les abréviations suivantes : i.i.d. pour indépendante et identiquement distribuée, PGD pour principe de grandes déviations et s.c.i. (resp. s.c.s.) pour semi-continue inférieurement (resp. semi-continue supérieurement)

Chapitre 1

PGD partiel, autonormalisation

1.1 Introduction

1.1.1 Objectifs de ce chapitre

Il s'agit d'abord (dans les sections 1.2 et 1.3) de mettre en place les principales notions sur lesquelles porte cette thèse ainsi que la boîte à outils à laquelle nous ferons sans cesse appel.
Les méthodes évoquées ne sont pas nouvelles et s'appuient sur les grands classiques des grandes déviations auxquels nous renverrons chemin faisant, notamment sur les livres de Dembo-Zeitouni [DemZei98], Dupuis-Ellis [DupEll97], Deuschel-Stroock [DeuStr89], Ellis [Ell85], Stroock [Str93], Varadhan [Var84]. Cependant, elles sont le plus souvent présentées dans le cadre de **vrais** principes de grandes déviations (PGD) associés à de **bonnes fonctions de taux** alors que nous parlerons ici de principes de grandes déviations **partiels**, avec des *taux* ne possédant pas toutes les propriétés d'une bonne fonction de taux. Nous les rappellerons donc en tentant à chaque étape de les placer dans un cadre aussi large que possible.
Simultanément, ce chapitre introduit et motive le thème des PGD autonormalisés, apparu dans le cadre indépendant et identiquement distribué avec les travaux de Shao, Dembo et Shao ([Sha97], [DemSha98a] et [DemSha98b]) et qui est au coeur de cette thèse. Cette étude a conduit à reprendre les notions de PGD partiel et de tension exponentielle partielle proposés par Dembo et Shao, puis à en faire l'étude en exhibant des critères suffisant dont certains apparaissent dans [DemSha98a] et [DemSha98b], mais de manière succinte dans le cadre i.i.d. moins général que celui que nous présentons.

1.1.2 Plan

La section 1.2 esquisse les définitions et le rôle clef de la tension exponentielle puis elle introduit le thème des PGD partiels.
La section 1.3 présente deux types de dualité :

a) la dualité entre PGD et fonctionnelle de Laplace qui fait bien apparaître le parallèle entre PGD et convergence étroite ;

b) la dualité entre PGD et fonctionnelles convexes (log-Laplace généralisées) lorsque l'espace étudié est un espace vectoriel localement convexe.

La section 1.4 introduit la notion de PGD autonormalisé et donne quelques exemples, ainsi que des outils appropriés à son étude.

Seuls les résultats relatifs aux PGD partiels ou autonormalisés seront prouvés. Selon la convention générale adoptée dans l'avant-propos, *dans tout ce chapitre, on se place sur* (E, d) *espace polonais.*

1.2 Principes de grandes déviations et tension exponentielle

1.2.1 Définitions

- Guidés par le parallèle qui existe entre la convergence étroite (resp. vague) des probabilités et les grandes déviations, nous adopterons le vocabulaire suivant. Un **taux étroit** est une application $I : E \to [0, \infty]$ dont les épigraphes $\{x \in E \mid I(x) \leq t\}$ (pour tout $t \geq 0$) sont compacts dans E. Une application $I : E \to [0, \infty]$ à épigraphes fermés dans E (autrement dit, semi-continue inférieurement (s.c.i.)) est appelée **taux vague**. Dans la littérature, les termes fréquemment utilisés sont respectivement "bonne fonction de taux" et "taux". Si A est une partie de E, on note $I(A) = \inf \{I(x) \mid x \in A\}$ (en adoptant la convention $I(\emptyset) = \infty$).
- Nous désignons par **vitesse** toute suite $(v_n)_n$ réelle, positive, croissant vers l'infini. Une suite $(\mu_n)_n$ de probabilités sur E satisfait un **principe de grandes déviations** (PGD) de taux étroit I et de vitesse $(v_n)_n$ si l'on a, pour tout $\Gamma \in \mathcal{E}$,

$$-I(\overset{\circ}{\Gamma}) \leq \liminf_{n\to\infty} \frac{1}{v_n} \log \mu_n(\Gamma) \leq \limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(\Gamma) \leq -I(\overline{\Gamma}), \tag{1.1}$$

ce qui est équivalent aux deux identités suivantes

$$-I(G) \leq \liminf_{n\to\infty} \frac{1}{v_n} \log \mu_n(G), \text{ pour tout ouvert } G \text{ de } E, \tag{1.2}$$

$$-I(F) \geq \limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(F), \text{ pour tout fermé } F \text{ de } E. \tag{1.3}$$

On dit que la suite $(\mu_n)_n$ vérifie un PGD inférieur (resp. supérieur) de vitesse $(v_n)_n$ si seule (**??**) (resp. seule (**??**)) est satisfaite.
- $(\mu_n)_n$ vérifie un **PGD vague supérieur** si (**??**) est vérifiée pour les compacts de E ; **vague** si (**??**) l'est aussi.

1.2.2 Tension exponentielle

Les résultats de cette sous-section rappelés pour la clarté de l'exposé ne seront pas démontrés ; les références y pourvoiront. Voici, pour les grandes déviations, l'analogue du théorème de Prokhorov sur la convergence étroite des probabilités. Rappelons tout d'abord les définitions suivantes.

Définition 1.1 (Tension exponentielle) *Une suite $(\mu_n)_n$ de probabilités sur (E,d) est **exponentiellement tendue** pour la vitesse $(v_n)_n$ si, pour tout réel $R>0$, il existe un compact K_R de E tel que*

$$\limsup_{n\to\infty}\frac{1}{v_n}\log\mu_n(K_R^c)\le -R.$$

Définition 1.2 (Compacité en grandes déviations) *Une suite $(\mu_n)_n$ de probabilités sur (E,d) est **relativement compacte en grandes déviations** pour la vitesse $(v_n)_n$ si, de toute suite, on peut extraire une sous-suite $(\mu_{u(n)})_n$ qui satisfait un PGD de vitesse $(v_{u(n)})_n$ et de taux I_u (qui dépend en général de la suite extraite $(u(n))_n$).*

Théorème 1.1 (Compacité et tension) *Une suite de $\mathcal{P}(E)$ est (v_n)-exponentiellement tendue sur (E,d) si, et seulement si, elle est (v_n)-relativement compacte en grandes déviations.*

Par le théorème précédent, l'unicité du taux I pour lequel un PGD peut être satisfait pour une suite extraite assure la validité du PGD pour la suite elle-même. Nous obtenons le schéma suivant :

$$\boxed{\text{Tension exponentielle + unicité du taux éventuel d'une sous-suite satisfaisant un PGD} \Longleftrightarrow \text{PGD pour la suite}}$$

On en déduit le critère suivant. Une suite $(\mu_n)_n$ de $\mathcal{P}(E)$ satisfait un PGD étroit supérieur de taux étroit I si, et seulement si, elle est exponentiellement tendue et vérifie un PGD vague supérieur de taux I.

$$\boxed{\text{Tension exponentielle + PGD vague supérieur de taux vague I} \Longleftrightarrow \text{PGD étroit supérieur de taux étroit I}}$$

La partie inférieure d'un PGD vague ou étroit étant la même, ce schéma est vrai sans les termes "supérieur". Cependant, il ne concerne que la partie supérieure.

Références : Le résultat énoncé dans le théorème 1.1 était plus ou moins présent dans la littérature depuis assez longtemps. Il a été solidement fondé plus récemment par Pukhalskii ([Puk91] ou [Puk94a], théorème (P)) ou bien, pour des topologies un peu différentes, par O'brien et Verwaat ([ObrVer91] ou [ObrVer95], théorème 2.3 et [Obr96]). Une autre preuve (par les fonctionnelles de Laplace) a été donnée par Dupuis et Ellis ([DupEll97], théorème 1.3.7). On peut voir aussi [DemZei98].

1.2.3 PGD et tension exponentielle partiels

♣ PGD partiel

Soit $\mathcal{S}$ une partie de $\mathcal{E}$. La suite (μ_n) vérifie un **principe de grandes déviations partiel** (PGDP), de taux vague I et de vitesse $(v_n)_n$ par rapport à $\mathcal{S}$ si (**??**) est vérifiée et que l'on a

$$-I(\overline{A})\ge\limsup_{n\to\infty}\frac{1}{v_n}\log\mu_n(A),\ \text{pour tout élément } A \text{ de } \mathcal{S}. \tag{1.4}$$

Ainsi un PGD vague est un PGD partiel par rapport à la classe des parties compactes de E. Lorsque E est une partie convexe fermée d'un espace vectoriel topologique, un PGD partiel par rapport à la classe des parties ouvertes convexes de E est appelé **PGD convexe**. La suite $(\mu_n)_n$ satisfait un PGD partiel supérieur si seul (**??**) est vérifiée.

♠ **Tension exponentielle partielle**

La notion suivante est naturellement liée à celle de PGD partiel.

Définition 1.3 *Une suite $(\mu_n)_n$ de probabilités sur E est (v_n)-**exponentiellement tendue par rapport à une partie** $\mathcal{S}$ de $\mathcal{E}$ si, pour tout $A \in \mathcal{S}$ et tout $R > 0$, il existe une partie compacte $K_{R,A}$ de E telle que :*

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(A \cap K_{R,A}^c) \le -R. \tag{1.5}$$

Voici quelques lemmes utiles dans la suite.

Lemme 1.1

*a) Si A et B satisfont (**??**), c'est aussi vrai pour $A \cup B$.*

*b) A et B satisfont (**??**) si, et seulement si, $A \cup B$ satisfait (**??**).*

Preuve

a) On a $\limsup_n \frac{1}{v_n} \log \mu_n(A \cup B) \le \sup\left\{-I(\overline{A}), -I(\overline{B})\right\} \le -I(\overline{A \cup B})$.

b) Pour $K_R = K_{R,A} \cup K_{R,B}$,

$$\begin{aligned}\limsup_n \frac{1}{v_n} \log \mu_n\left((A \cup B) \cap K_R^c\right) &= \sup\left\{\limsup_n \frac{1}{v_n} \log \mu_n\left(A \cap K_R^c\right),\ \limsup_n \frac{1}{v_n} \log \mu_n\left(B \cap K_R^c\right)\right\} \\ &\le -R. \ \diamondsuit\end{aligned}$$

Exemple 1.1 *Sur $\mathbb{R}\times\mathbb{R}_+$, on considère l'ensemble $\mathcal{S}_+$ des parties de la forme $\{(s,v) \mid s \ge rv\}$ et $\mathcal{S}_-$ l'ensemble des parties de la forme $\{(s,v) \mid s \le -rv\}$, pour $r > 0$. Vérifier la tension exponentielle relative à $\mathcal{S}_+$ et la tension exponentielle relative à $\mathcal{S}_-$ équivaut à la vérifier sur l'ensemble des parties de la forme $\{(s,v) \mid |s| \ge rv\}$.*

Lemme 1.2 *Soit E un espace de Banach séparable. La tension exponentielle partielle relative à $\mathcal{S}$ implique, pour tout $A \in \mathcal{S}$,*

$$\lim_{r\to\infty} \limsup_n \frac{1}{v_n} \log \mu_n\left(A \cap \{x \mid \|x\| > r\}\right) = -\infty;$$

cette propriété est nécessaire et suffisante lorsque $E = \mathbb{R}^d$.

Cela résulte facilement du fait que tout compact de E est contenu dans une boule fermée de rayon r, cette boule étant compacte si $E = \mathbb{R}^d$.

Lemme 1.3 *Soit F une fonction borélienne de $\mathbb{R}^d \times \mathbb{R}_+$ dans $\mathbb{R}_+$ telle que, pour tout $R > 0$,*

$$\lim_{\|s\|\to\infty} \inf_{v\le R} F(s,v) > 0.$$

On lui associe

$\mathcal{S}(F) \stackrel{\triangle}{=} \left\{ A \text{ borélien de } \mathbb{R}^d \times \mathbb{R}_+ \mid \lim_{R\to\infty} \inf\{F(s,v) \mid (s,v) \in A,\, v > R\} > 0 \right\},$

$\mathcal{T}(F) \stackrel{\triangle}{=} \left\{ A_\rho = \left\{ (s,v) \in \mathbb{R}^d \times \mathbb{R}_+ \mid F(s,v) > \rho \right\} \mid \rho > 0 \right\}.$

Alors, la tension exponentielle relative à $\mathcal{S}(F)$ équivaut à la tension exponentielle relative à $\mathcal{T}(F)$.

Preuve La classe $\mathcal{T}(F)$ est inclue dans $\mathcal{S}(F)$. Donc il sufit de montrer que la tension exponentielle relative à $\mathcal{T}(f)$ implique celle relative à $\mathcal{S}(F)$. Par l'hypothèse faite sur F, pour tout $R > 0$, il existe un $\rho > 0$ et un $r > 0$ tels que $\{(s,v) \mid v \le R,\, \|s\| > r\} \subset A_\rho$. On a donc, pour tout $R > 0$ et par le lemme 1.2,

$$\begin{aligned}
&\lim_{r\to\infty} \limsup_n \frac{1}{v_n} \log \mu_n \left(\{(s,v) \mid v \le R,\, \|s\| > r\} \right) \\
&\le \lim_{r\to\infty} \limsup_n \frac{1}{v_n} \log \mu_n \left(A_\rho \cap \{(s,v) \mid v \le R,\, \|s\| > r\} \right) = -\infty.
\end{aligned}$$

Soit $A \in \mathcal{S}(F)$. Il existe $\rho > 0$ et $R > 0$ tels que

$$\inf\{F(s,v) \mid (s,v) \in A,\, v > R\} > \rho, \text{ d'où } A \cap \{(s,v) \mid v > R\} \subset A_\rho;$$

$$\begin{aligned}
&\overline{\lim} \frac{1}{v_n} \log \mu_n \left(A \cap \{(s,v) \mid v + \|s\| > r\} \right) \\
\le\ & \sup \left\{ \overline{\lim} \frac{1}{v_n} \log \mu_n \left(\{(s,v) \mid v \le R,\, \|s\| > r - R\} \right), \overline{\lim} \frac{1}{v_n} \log \mu_n \left(A_\rho \cap \{(s,v) \mid v + \|s\| > r\} \right) \right\}.
\end{aligned}$$

On en déduit que

$$\lim_{r\to\infty} \limsup_n \frac{1}{v_n} \log \mu_n \left(A \cap \{(s,v) \mid v + \|s\| > r\} \right) = -\infty. \ \diamondsuit$$

$\diamondsuit$ Compacité exponentielle partielle

On a le résultat analogue au schéma présenté en 1.2.2 permettant d'établir un PGD.

Proposition 1.1 *Soient $(\mu_n)_n$ une suite de probabilités sur E qui vérifie un PGD vague supérieur de taux I et de vitesse (v_n) et $\mathcal{S}$ une partie de $\mathcal{E}$. Si $(\mu_n)_n$ est (v_n)-exponentiellement tendue par rapport à $\mathcal{S}$, elle vérifie un PGD partiel supérieur de taux I par rapport à $\mathcal{S}$.*

Tension exponentielle partielle + PGD vague supérieur de taux vague $I \Longrightarrow$ PGD partiel supérieur de taux vague I

Preuve Soit $A \in \mathcal{S}$. Pour tout $R > 0$, il existe un compact K_R de E tel que

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n (A \cap K_R^c) \le -R.$$

On a $\mu_n(A) \leq \mu_n(\overline{A} \cap K_R) + \mu_n(A \cap K_R^c)$, ce qui permet d'écrire

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(A) \leq \max\left\{\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(\overline{A} \cap K_R), \limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(A \cap K_R^c)\right\}.$$

Par la borne supérieure du PGD vague,

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(\overline{A} \cap K_R) \leq -I(\overline{A} \cap K_R) \leq -I(\overline{A}).$$

D'où :

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(A) \leq -\min\left\{I(\overline{A}), R\right\}.$$

Ceci étant vrai pour tout $R > 0$, la proposition est démontrée. $\diamondsuit$

Une majoration de type (**??**) ou (**??**) n'est intéressante que si le majorant est non nul. Un taux étroit atteint son infimum sur tout fermé, ce qui n'est pas vrai en général pour un taux vague qui l'atteint sur un compact. Malgré cela, nous avons le lemme suivant.

Lemme 1.4 *Soit un taux vague I et $(\mu_n)_n$ une suite de probabilités ayant la propriété de tension exponentielle partielle par rapport à $\mathcal{S}$. Soit F un fermé de $\mathcal{S}$*

1) *Supposons que la suite $(\mu_n)_n$ vérifie un PGD vague supérieur de vitesse (v_n) et de taux I ne s'annulant pas sur F, alors*

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(F) < 0.$$

2) *Si $(\mu_n)_n$ satisfait un PGD vague de taux I et si $I(F) = I(\overset{\circ}{F})$, alors il existe $x_F \in F$ tel que $I(F) = I(x_F)$.*

Preuve On prouve la partie 1). Soit un fermé $F \in \mathcal{S}$ sur lequel I ne s'annule pas. Pour tout $R > 0$, il existe un compact K_R de E tel que

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(F) \leq \max\{-R, -I(K_R \cap F)\}.$$

De plus, pour tout $x \in K_R \cap F$, $I(x) > 0$, ce qui implique que $I(K_R \cap F) > 0$ ($K_R \cap F$ étant compact). On en déduit que $\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(F) < 0$.
Pour ce qui est du second point, il suffit de considérer le cas $I(F) < R < \infty$. On a

$$I(F) = I(\overset{\circ}{F}) \geq \inf\{R, I(K_R \cap F)\}$$

donc $I(F) = I(K_R \cap F) = I(x_F)$ où $x_F \in K_R \cap F$ ($K_R \cap F$ étant compact). $\diamondsuit$

Nous avons vu en 1.2.2 que la tension exponentielle est équivalente à la relative compacité en grandes déviations. Voici une réciproque de la proposition 1.1, sous des conditions assez restrictives.

Définition 1.4 *Une partie $\mathcal{S}$ de $\mathcal{E}$ est dite stable par inclusion si, pour tout $A \in \mathcal{S}$, tout $B \in \mathcal{E}$ contenu dans A appartient à $\mathcal{S}$.*

Proposition 1.2 *Si une suite $(\mu_n)_n$ de probabilités sur E vérifie un PGD partiel supérieur de vitesse $(v_n)_n$ par rapport à une classe stable par inclusion $\mathcal{S}$, de taux vague I tel que, pour tout $A \in \mathcal{S}$ et tout $t \geq 0$, $\{x \in E \mid I(x) \leq t\} \cap \overline{A}$ soit compact, alors $(\mu_n)_n$ est (v_n)-exponentiellement tendue par rapport à $\mathcal{S}$.*

Preuve Soit $\{(x_j)_j\}$ une partie dénombrable dense dans E, $B > 0$ et $A \in \mathcal{S}$. Par hypothèse sur I, pour tout $M \geq 0$, l'ensemble $\Gamma_M \stackrel{\triangle}{=} \{x \mid I(x) \leq M\} \cap \overline{A}$ est relativement compact donc totalement borné :

$$\forall m \in \mathbb{N},\ \exists\, R(m) \in \mathbb{N} \text{ tel que } \Gamma_{B+m} \subset \bigcup_{j=1}^{R(m)} B(x_j, 2^{-m});$$

$\mathcal{S}$ est stable par inclusion et, pour tout $x \in (\cup_{j=1}^{R(m)} B(x_j, 2^{-m}))^c \cap \overline{A}$, on a $I(x) > B + m$ donc

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n \left\{ (\bigcup_{j=1}^{R(m)} B(x_j, 2^{-m}))^c \cap A \right\} \leq -I \left\{ (\bigcup_{j=1}^{R(m)} B(x_j, 2^{-m}))^c \cap \overline{A} \right\} \leq -(B+m).$$

Il existe un entier $Q(m)$ tel que $\forall n \geq Q(m)$ on ait :

$$\frac{1}{v_n} \log \mu_n \left\{ (\bigcup_{j=1}^{R(m)} B(x_j, 2^{-m}))^c \cap A \right\} \leq -(B + \frac{m}{2}).$$

Tout ensemble fini de probabilités est tendu donc l'inégalité ci-dessus est vraie pour tout $n \in \mathbb{N}$ quitte à prendre $R(m)$ plus grand. On introduit le compact suivant :

$$K_B \stackrel{\triangle}{=} \bigcap_{m\in\mathbb{N}} \bigcup_{j=1}^{R(m)} \overline{B(x_j, 2^{-m})}.$$

K_B est totalement borné et fermé dans E (complet). Par conséquent, il est compact et :

$$\mu_n(K_B^c \cap A) \leq e^{-Bv_n} \sum_{m=0}^{\infty} e^{\frac{mv_0}{2}},$$

$$\limsup_{n\to\infty} \frac{1}{v_n} \log\ \mu_n(A \cap K_B^c) \leq -B.\ \Diamond$$

Remarque 1.1 *Voici un résultat permettant d'obtenir des PGD vagues, étroits ou partiels supérieurs uniformes*
Soit A un espace métrique et $(\mu_{a,n})_{a\in A, n\in\mathbb{N}}$ une famille de probabilités sur E.

Soit $\Gamma \in \mathcal{E}$. On suppose que, pour toute suite $(a_n)_n$ convergente de A, la suite $(\mu_{a_n,n})_n$ vérifie

$$\limsup_n \frac{1}{v_n} \log \mu_{a_n,n}(\Gamma) \leq -I(\Gamma);$$

alors, pour toute partie compacte $\mathcal{H}$ de A,

$$\limsup_n \sup_{a \in \mathcal{H}} \frac{1}{v_n} \log \mu_{a,n}(\Gamma) \leq -I(\Gamma);$$

Références La notion de PGD partiel était implicite depuis longtemps, notamment pour les PGD vagues ou convexes. Elle a été explicitée par Dembo et Shao ([DemSha98a] et [DemSha98b]).

1.3 Dualités

1.3.1 PGD et principe de Laplace

Soient $(\mu_n)_n$ une suite de $\mathcal{P}(E)$ et $(v_n)_n$ une vitesse. Pour $f \in \mathcal{B}_b(E)$, on note

$$\Lambda_n(f) = \frac{1}{v_n} \log \int e^{-v_n f} d\mu_n.$$

Définition 1.5 *Une suite $(\mu_n)_n$ de $\mathcal{P}(E)$ satisfait un* ***principe de Laplace*** *de vitesse v et de fonctionnelle limite Λ baptisée* ***fonctionnelle de Laplace****, si, pour toute fonction $f \in \mathcal{C}_b(E)$, la limite suivante existe :*

$$\lim_{n\to\infty} \Lambda_n(f) = \Lambda(f).$$

On parle aussi de ***principe de Laplace supérieur*** *(resp.* ***inférieur****) si :*

$$\limsup_{n\to\infty} \Lambda_n(f) \leq \Lambda(f), \ (\textit{resp. si} \ \liminf_{n\to\infty} \Lambda_n(f) \geq \Lambda(f)).$$

Théorème 1.2 (Varadhan) *Soient une suite $(\mu_n)_n$ de $\mathcal{P}(E)$ et une fonction I de E dans $[0,\infty]$ arbitraire.*

1) *On suppose que, pour toute boule ouverte B,*

$$\liminf_{n\to\infty} \frac{1}{v_n} \log \mu_n(B) \geq -I(B).$$

Alors, on a un principe de Laplace inférieur de vitesse v : pour toute fonction $f \in \mathcal{C}_b(E)$,

$$\liminf_{n\to\infty} \Lambda_n(f) \geq -\inf_{x\in E} \{f(x) + I(x)\}.$$

2) *On suppose que, pour tout fermé F de E,*

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mu_n(F) \leq -I(F).$$

Alors, on a un principe de Laplace supérieur de vitesse v : pour toute fonction $f \in \mathcal{C}_b(E)$,

$$\limsup_{n\to\infty} \Lambda_n(f) \leq -\inf_{x\in E} \{f(x) + I(x)\}.$$

Corollaire 1.1 (Varadhan) *On considère une fonction f continue de E dans $\mathbb{R}$ arbitraire. Dans le cadre de la partie 1) du théorème 1.2, on a encore*

$$\liminf_{n\to\infty} \frac{1}{v_n} \log \int e^{v_n f} d\mu_n \geq \sup_x \{f(x) - I(x)\}.$$

Définition 1.6 *Si I est une fonction de E dans $[0, \infty]$, on appelle* ***conjuguée de I****, l'application Λ définie, pour toute fonction $f \in \mathcal{C}_b(E)$, par*

$$\Lambda(f) = -\inf_{x\in E} \{f(x) + I(x)\}.$$

Théorème 1.3 (Bryc-Varadhan supérieur) *Si $(\mu_n)_n$ est une suite de $\mathcal{P}(E)$ et si I est un taux étroit, les quatre propriétés suivantes sont équivalentes.*

(i) *Pour toute fonction $f \in \mathcal{LB}(E)$, on a*

$$\limsup_{n\to\infty} \Lambda_n(f) \leq \Lambda(f) = -\inf_x \{f(x) + I(x)\}.$$

(ii) *$(\mu_n)_n$ vérifie un principe de Laplace supérieur de fonctionnelle limite Λ, conjuguée de I.*

(iii) *Pour toute fonction f s.c.i. bornée,*

$$\limsup_{n\to\infty} \Lambda_n(f) \leq \Lambda(f) = -\inf_x \{f(x) + I(x)\}.$$

(iv) *$(\mu_n)_n$ satisfait un PGD supérieur de taux I.*

Soit $I : E \to [0, \infty]$ une fonction arbitraire. On rappelle la définition donnée en avant-propos de sa régularisée s.c.i. I^{sci} :

$$I^{sci}(x) = \lim_{r\to 0} \uparrow I(B(x, r)).$$

Théorème 1.4 (Bryc-Varadhan inférieur) *Si $(\mu_n)_n$ est une suite de $\mathcal{P}(E)$ et si I est une fonction $E \to [0, \infty]$ arbitraire, les propriétés suivantes sont équivalentes :*

(i) *Pour toute fonction $f \in \mathcal{LB}(E)$, on a*

$$\liminf_{n\to\infty} \Lambda_n(f) \geq \Lambda(f) = -\inf_x \{f(x) + I(x)\}.$$

(ii) *$(\mu_n)_n$ vérifie un principe de Laplace inférieur de fonctionnelle limite Λ, conjuguée de I.*

(iii) $(\mu_n)_n$ *satisfait un PGD inférieur de taux* I^{sci} *:*

$$\forall G \text{ ouvert },\ \liminf_n \frac{1}{v_n} \log \mu_n(G) \geq -I(G) = -I^{sci}(G).$$

Synthèse : Si $(\mu_n)_n$ est une suite de $\mathcal{P}(E)$ et si I est un taux étroit, les trois propriétés suivantes sont équivalentes :

(i) Pour toute fonction $f \in \mathcal{LB}(E)$, on a

$$\lim_{n\to\infty} \Lambda_n(f) = \Lambda(f) = -\inf_x \{f(x) + I(x)\}.$$

(ii) $(\mu_n)_n$ vérifie un principe de Laplace de fonctionnelle de Laplace limite Λ, conjuguée de I.

(iii) $(\mu_n)_n$ satisfait un PGD étroit de taux I.

Théorème 1.5 (Bryc - Conjugaison pour une suite exponentiellement tendue) *Soit une suite $(\mu_n)_n$ de $\mathcal{P}(E)$. On suppose que (E, d) est polonais et (μ_n)-exponentiellement tendue pour la vitesse v. Alors, dès que le principe de Laplace est satisfait pour toute fonction $f \in \mathcal{LB}(E)$, on a un PGD de taux étroit*

$$I(x) = -\inf\{f(x) + \Lambda(f) \mid f \in \mathcal{LB}(E)\}.$$

En outre, on a aussi, pour toute fonction $f \in \mathcal{C}_b(E)$,

$$\Lambda(f) = -\inf_x \{f(x) + I(x)\}$$

et, pour tout $x \in E$,

$$I(x) = -\inf\{f(x) + \Lambda(f) \mid f \in \mathcal{C}_b(E)\}.$$

Références Les sources principales sont Varadhan ([Var66] et [Var84]), Bryc [Bry90] et un exposé très clair de Dupuis et Ellis ([DupEll97], chapitre 1). Ainsi, nos interprétations des résultats de Varadhan et de Bryc reposent sur une lecture précise des démonstrations de Dupuis et Ellis qui les formulent avec des égalités et des vrais taux. Cette approche des PGD éclaire le lien entre "grandes déviations" et "convergence étroite" que Dupuis et Ellis ont su mettre en valeur.
L'ouvrage de Dupuis et Ellis prouve de nombreux PGD relatifs à des processus à l'aide d'une technique de programmation dynamique pour le calcul de $\Lambda_n(f)$ suivie d'une étude asymptotique. Nous aurons recours à cette méthode dans deux situations délicates (théorème 2.4 en 2.3.4, théorèmes 3.6 et 3.7 en 3.5.2). Cependant, elle conduit à des calculs assez compliqués qu'il est souvent agréable de contourner grâce à des méthodes de convexité évoquées dans la section suivante.

1.3.2 PGD vague-convexe

Voici un cadre de convexité adapté à l'étude des grandes déviations.

Axiome 1.1 *E est une partie convexe, fermée de G, espace vectoriel topologique séparable ; on suppose que la topologie trace sur E en fait un espace polonais. En outre, la topologie de E peut être définie par une distance d satisfaisant la condition de convexité suivante : pour tous x_1, x_2, y_1, y_2 de E et pour tout $t \in]0,1[$,*

$$d\left(tx_1 + (1-t)x_2, ty_1 + (1-t)y_2\right) \leq \sup\left\{d(x_1,y_1), d(x_2,y_2)\right\}.$$

Si G' est le dual topologique de G, on note E' l'ensemble des traces des éléments de G' sur E.

Remarque 1.2

a) La condition portant sur la distance implique la convexité des boules ouvertes $B(x,r)$ de centre x et de rayon r.

b) La condition portant sur la distance peut être remplacée par la condition plus large mais plus abstraite "l'enveloppe convexe de toute partie compacte est compacte" (voir [DemZei98] p. 252-253).

Ces conditions s'appliquent dans les deux situations suivantes.

Exemple 1.2 *E est un espace de Banach séparable de dual topologique E' ; bien souvent, $E = E' = \mathbb{R}^d$ dans les applications.*

Exemple 1.3 *$E = \mathcal{P}(H)$ muni de la distance de Dudley, pour un espace polonais H quelconque. Ici, $E' = \mathcal{C}_b(H)$ et la topologie faible rendant continues les applications $\mu \to \int f d\mu$ est la topologie étroite. E est une partie convexe de l'espace vectoriel des mesures signées bornées sur H. Pour la distance de Dudley, E est polonais et la condition de convexité de l'hypothèse est satisfaite.*

Définition 1.7 *Une suite de probabilités $(\mu_n)_n$ définie sur un espace E avec l'axiome 1.1. satisfait un* ***PGD vague convexe*** *de vitesse v et de taux I si*

a) I est s.c.i. et convexe,

b) $(\mu_n)_n$ satisfait un PGD vague de taux I,

c) pour tout ouvert convexe A de E,

$$\lim_{n\to\infty} \frac{1}{v_n} \log \mu_n(A) = -I(A).$$

Voici un outil important pour prouver un PGD vague-convexe.

Proposition 1.3 *On se place dans le cadre de l'axiome 1.1. et on suppose que, pour tout ouvert convexe A, la limite suivante existe :*

$$\lim_{n\to\infty} \frac{1}{v_n} \log \mu_n(A) = -L(A).$$

Alors, la suite $(\mu_n)_n$ satisfait un PGD vague convexe de taux vague

$$I(x) = \sup\left\{L(A) \mid \ A \ \textit{ouvert convexe contenant } x\right\} = \lim_{r\to 0^+} L\left(B(x,r)\right) = L^{sci}(x).$$

De plus, pour tout ouvert convexe, $I(A) = L(A)$.

Le PGD vague convexe le plus ancien est le suivant.

Théorème 1.6 (Bahadur-Zabell - PGD vague convexe pour une suite i.i.d.) *Soient $(X_n)_n$ une suite de variables aléatoires i.i.d. à valeurs dans un espace E vérifiant l'axiome 1.1. et $S_n = X_1 + ... + X_n$. La suite $(S_n/n)_n$ satisfait un PGD vague-convexe de vitesse n et de taux I vague et convexe.*

Références Le concept de PGD convexe et le théorème précédent proviennent de Bahadur et Zabell [BahZab79] et de Azencott [Aze80]. On pourra consulter aussi le chapitre 6 de [DemZei98] où l'on retrouve en outre la preuve de la proposition 1.3.

1.3.3 Identification d'un taux convexe

Soit E un espace vectoriel (sur $\mathbb{R}$) topologique et localement convexe ; son dual E' correspond à l'ensemble des applications linéaires continues de E dans $\mathbb{R}$, notées $\lambda : x \mapsto (\lambda \mid x)$. Si f est une fonction de E dans $]-\infty, \infty]$, la **transformée de Legendre** de f est la fonction définie sur E' par

$$f^*(\lambda) = \sup_{x \in E} \{(\lambda \mid x) - f(x)\}.$$

Le résultat suivant est prouvé dans tous les livres abordant l'analyse convexe.

Proposition 1.4 (Relation de dualité de la transformée de Legendre) *Si f est convexe et s.c.i, on a la formule duale*

$$f(x) = f^{**}(x) = \sup_{\theta \in E'} \{(\theta \mid x) - f^*(\theta)\}.$$

En adaptant le théorème de Varadhan, Dinwoodie obtient le résultat suivant.

Proposition 1.5 (Dinwoodie) *Soit $(\mu_n)_n$ une suite de $\mathcal{P}(E)$.*

1) *Lorsque l'on a un PGD inférieur de vitesse v et de taux vague I, on a, quels que soient $M < \infty$ et $\lambda \in E'$,*

$$\liminf_n \frac{1}{v_n} \int e^{v_n \inf((\lambda|y),M)} \mu_n(dy) \geq \sup_{x \in E} \{\inf((\lambda \mid x), M) - I(x)\};$$

et, I^* étant la conjuguée de I,

$$\lim_{M \to \infty} \liminf_n \frac{1}{v_n} \log \int e^{v_n \inf((\lambda|y),M)} \mu_n(dy) \stackrel{\triangle}{=} \Lambda_{\inf}(\lambda) \geq I^*(\lambda).$$

2) *On suppose que, pour une suite $(\mu_n)_n$ de probabilités sur E et un taux vague convexe I, on a, quel que soit l'ouvert convexe Γ de E,*

$$\limsup_n \frac{1}{v_n} \log \mu_n(\Gamma) \leq -I(\Gamma).$$

Alors, quels que soient $M < \infty$ et $\lambda \in E'$,

$$\limsup_n \frac{1}{v_n} \int e^{v_n \inf((\lambda|y),M)} \mu_n(dy) \leq \sup_{x \in E} \{\inf((\lambda \mid x), M) - I(x)\};$$

et, I^* étant la conjuguée de I,

$$\lim_{M \to \infty} \limsup_n \frac{1}{v_n} \log \int e^{v_n \inf((\lambda|y),M)} \mu_n(dy) \stackrel{\triangle}{=} \Lambda_{\sup}(\lambda) \leq I^*(\lambda).$$

3) *Si la suite $(\mu_n)_n$ satisfait un PGD vague convexe de taux convexe I, on a*

$$\Lambda_{\sup} = \Lambda_{\inf} = \Lambda, \quad d'o\grave{u}\ I = \Lambda^*.$$

Réciproquement, il y a diverses adaptations du théorème de Bryc au cas convexe, notamment le théorème de Gărtner étudié dans les divers livres généraux portant sur les PGD mais auquel nous ne ferons pas appel. Voici un résultat supérieur qui nous sera utile.

Proposition 1.6 (De Acosta) *On considère une suite $(\mu_n)_n$ de $\mathcal{P}(E)$ telle que, pour tout $\lambda \in E'$,*

$$\limsup_n \frac{1}{v_n} \log \int e^{v_n(\lambda|y)} \mu_n(dy) \leq \Lambda_{\sup}(\lambda).$$

a) *La fonction $\lambda \mapsto \Lambda_{sup}(\lambda)$ étant arbitraire, $(\mu_n)_n$ satisfait alors un PGD supérieur vague de taux $I = \Lambda^*_{\sup}$.*

b) *Si $\lambda \mapsto \Lambda_{\sup}(\lambda)$ est s.c.i et si I est un taux étroit, on a un PGD supérieur étroit.*

Références La proposition 1.5 a été obtenue par Dinwoodie [Din93]. La proposition 1.6 provient de De Acosta ([Aco85a] et [Aco90]). Le théorème de Gărtner (voir [Gar77] et [Ell84]), valable sur $\mathbb{R}^d$ et étendu à des espaces plus généraux par Baldi [Bal88] est une version convexe importante du théorème de Bryc ; mais il requiert des propriétés d'*escarpement* de la fonctionnelle convexe limite que nous ne vérifierons pas par la suite.

1.3.4 Exemples classiques i.i.d.

Voici deux exemples historiques de PGD pour des suites empiriques de variables aléatoires i.i.d. qui illustrent les résultats de 1.3.3.

♠ Moyennes empiriques d'une suite i.i.d. à valeurs dans un espace de Banach séparable

On considère une suite $(X_n)_n$ de variables aléatoires i.i.d. de loi μ, à valeurs dans un espace de Banach B (de dual topologique B'). La **log-Laplace** Λ_μ et **la transformée de Cramér** (que nous appelerons aussi **taux de Cramér**) Λ^*_μ de la loi μ sont définis par

$$\Lambda_\mu(\lambda) = \log \int e^{(\lambda|y)} \mu(dy) \text{ et } \Lambda^*_\mu(\lambda) = \sup_{\lambda \in B'} \{(\lambda \mid x) - \Lambda_\mu(\lambda)\}.$$

Soit $S_n \stackrel{\triangle}{=} X_1 + ... + X_n$. On a

$$\frac{1}{n} \log \mathbb{E}\left(e^{(\lambda|S_n)}\right) = \Lambda_\mu(\lambda).$$

Le théorème de Dinwoodie permet d'identifier le taux du PGD vague-convexe de Bahadur-Zabell : c'est le taux de Cramér.

Théorème 1.7 (Cramér)

1) *Pour une suite de variables aléatoires i.i.d. de loi μ dans un espace de Banach séparable, la suite des moyennes empiriques satisfait toujours un PGD vague-convexe dont le taux est la transformée de Cramér de μ.*
2) *Cette suite est exponentiellement tendue et le PGD est donc étroit sous la condition suivante :*
   a) *Pour $B = \mathbb{R}^d$, il existe $t > 0$ tel que $\int e^{t\|x\|}\mu(dx) < \infty$* ***(condition de Cramér)****.*
   b) *Pour un espace de Banach B quelconque, quel que soit $t > 0$, $\int e^{t\|x\|}\mu(dx) < \infty$* ***(condition de Cramér renforcée)****.*

♣ Lois empiriques d'une suite i.i.d. à valeurs dans un espace polonais

Soit $(Z_n)_n$ une suite i.i.d à valeurs dans (E, d) polonais. On étudie la suite des lois empiriques $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i}$ à valeurs dans $\mathcal{P}(E)$. Le théorème de Bahadur-Zabell s'applique encore et celui de Dinwoodie permet d'identifier le taux. Ici, le dual topologique de $\mathcal{P}(E)$ est $\mathcal{C}_b(E)$ et, pour $f \in \mathcal{C}_b(E)$, on a

$$(f \mid L_n) = \frac{1}{n}\sum_{i=1}^n f(Z_i),$$

$$\frac{1}{n}\log \mathbb{E}\left(e^{n(f|L_n)}\right) = \log \int e^f d\mu,$$

$$I(\nu) = \sup_{f\in\mathcal{C}_b(E)} \left\{\int f d\nu - \log \int e^f \mu\right\}.$$

Ce taux est l'information de Kullback de ν relative à μ, sur laquelle on reviendra dans le chapitre 2. Dans ce cadre, la tension exponentielle est toujours vérifiée donc le PGD étroit est toujours valable. Rappelons une preuve de la tension exponentielle. Soit A un borélien de E et $a > 0$;

$$\begin{aligned}\mathbb{P}\left(L_n(A) > r\right) \le \mathbb{E}\left(e^{na(L_n(A)-r)}\right) &= e^{-nar}\left(\mathbb{E}\left(e^{a\mathbb{I}_A(X)}\right)\right)^n \\ &= e^{-nar}\left(\mu(A^c) + e^a\mu(A)\right)^n.\end{aligned}$$

On peut associer à tout entier j un compact C_j de E tel que $\mu(C_j^c) \le e^{-2j^2}(e^j - 1)$. Prenant $A = C_j^c, r = 1/j$ et $a = 2j^2$, on a $\mathbb{P}\left(L_n(C_j^c) > 1/j\right) \le e^{-nj}$. Soit $K_J = \cap_{j\ge J}\left\{\nu \mid \nu(C_j^c) \le 1/j\right\}$. Comme chaque $\left\{\nu \mid \nu(C_j^c) \le 1/j\right\}$ est un convexe fermé de $\mathcal{P}(E)$, il en va de même pour K_J. De plus K_J est tendu, donc compact.

$$\mathbb{P}\left(L_n \in K_J^c\right) \le \sum_{j\ge J} e^{-nj} \le 2e^{-nJ},$$

$$\limsup_n \frac{1}{n}\log \mathbb{P}\left(L_n \in K_J^c\right) \le -J.$$

Théorème 1.8 (Sanov) *Pour toute suite i.i.d.* $(Z_n)_n$ *à valeurs dans un espace polonais, la suite des lois empiriques* $\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{Z_i}\right)_n$ *satisfait un PGD étroit et convexe dont le taux est l'information de Kullback relative à* μ, $\nu \mapsto K(\nu \mid \mu)$.

Références Le cas i.i.d. est, bien sûr, le plus ancien abordé dans la littérature relative aux grandes déviations. Pour $B = \mathbb{R}^d$ et sous la condition de Cramér sur $\mathbb{R}^d$, ces résultats remontent à Chernov [Che52] et à Cramér [Cra38], tandis que le théorème de Sanov vient de [San57] pour le cas réel. Les auteurs cités après l'énoncé du théorème 1.6. de Bahadur-Zabell pensaient entre autres à la situation i.i.d. décrite ci-dessus et ont identifié directement les taux.
La condition de Cramér pour la tension exponentielle est ancienne. Dans le cas des espaces de Banach, elle est établie par Donsker-Varadhan [DonVar76] et par De Acosta [Aco85b].
Les deux chapitres suivants présenteront diverses extensions de ces deux exemples i.i.d. aux chaînes de Markov. Les travaux de Dinwoodie et de De Acosta cités en 1.3.3 avaient précisément cet objectif.

1.4 Autonormalisation

1.4.1 De Cramér à Shao

Soit une fonction $g : E \to \mathbb{R}^d$ μ-intégrable. qui n'est pas nulle μ-presque-sûrement. On note, pour $p > 1$,

$$S_n(g) = \sum_{i=1}^{n} g(X_i) \quad \text{et} \quad V_{n,p}(g) = \left(\sum_{i=1}^{n} \|g(X_i)\|^p\right)^{1/p}.$$

Le résultat classique de Chernov (voir [Che52]) établit, dans le cas réel $d = 1$, la majoration suivante :

$$\mathbb{P}\left(\frac{S_n(g)}{n} \geq x\right) \leq e^{-n\Lambda_g^*(x)}, \text{ pour tout } x > \textstyle\int gd\mu,$$

où Λ_g^* est la transformée de Cramér de $g(\mu)$. L'inégalité n'a d'intérêt que si $\Lambda_g^*(x) > 0$ pour tout $x > \int gd\mu$, c'est à dire si la condition de Cramér unilatérale ($\exists \tau > 0$, $\int e^{\tau g} d\mu < \infty$) est satisfaite.
Sous cette condition, nous avons le corollaire suivant, conséquence du théorème de Cramér :

$$\forall x > \int gd\mu, \ \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{S_n(g)}{n} \geq x\right) = -\Lambda_g^*(x) < 0.$$

La condition de Cramér est trop forte dans certaines situations. D'où l'idée de Q.M. Shao de prouver des résultats de type Chernov, sans hypothèse de moments sur la fonction g (voir [Sha97]), non pas pour la moyenne empirique, mais pour une suite **autonormalisée** $R_{n,p}(g)$ (pour $p > 1$), définie dans le cas réel $d = 1$ par

$$R_{n,p}(g) = \frac{S_n(g)}{n^{1-1/p} V_{n,p}(g)}.$$

Il obtient le résultat suivant, pour tout $x > \frac{\int g d\mu}{\|g\|_{L^p(\mu)}}$ (en prenant $\frac{\int g d\mu}{\|g\|_{L^p(\mu)}} = 0$ si $\|g\|_{L^p(\mu)} = \infty$ et si $\int g d\mu = 0$),

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(R_{n,p}(g) \geq x\right) = -K(x) < 0.$$

Dans le cas où $d \geq 1$, une suite autonormalisée est donnée par

$$T_{n,p}(g) = \frac{\|S_n(g)\|}{n^{1-1/p} V_{n,p}(g)}$$

et on obtient, pour tout $x > \frac{\|\int g d\mu\|}{\|g\|_{L^p(\mu)}}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(T_{n,p}(g) \geq x\right) = -L(x) < 0. \tag{1.6}$$

Dans son papier, Q.M. Shao donne diverses applications statistiques à ce résultat, entre autres pour la statistique de Student (voir toujours [Sha97], [Sha97b] et [Sha98]) :

Si $\int g d\mu \geq 0$, on note $T_n \overset{\triangle}{=} \frac{\sum_{i=1}^n g(X_i)}{n^{1/2}\left(\frac{1}{n-1}\sum_{i=1}^n (g(X_i) - S_n(g)/n)^2\right)^{1/2}}$. Pour tout x tel que $x > \frac{\int g d\mu}{\|g\|_{L^2(\mu)}}$, on a

$$\lim_n \frac{1}{n} \log \mathbb{P}\left(T_n \geq n^{1/2} x\right) = \lim_n \frac{1}{n} \log \mathbb{P}\left(\frac{S_n(g)}{V_{n,2}(g)} \geq x \left(\frac{n}{n + nx^2 - 1}\right)^{1/2}\right) = -H(x) < 0.$$

Une application analogue aux tests de Hotelling est donnée par Dembo et Shao [DemSha98c].

Une possibilité, pour établir des résultats tels que (**??**), est l'obtention d'un PGD partiel pour le couple de moyennes empiriques $(\frac{1}{n} S_n(g), \frac{1}{n}(\sum_{i=1}^n \|g(X_i)\|^p))_n$, par rapport à une classe de boréliens particulière que nous appelerons "classe de Chernov".

1.4.2 Les classes de Chernov

Nous introduisons ici la notion de **classe de Chernov** associée à une fonction de Lyapounov. $(B, \mathcal{B})$ désigne un espace de Banach séparable, muni de sa tribu borélienne. On rappelle qu'une fonction de Lyapounov U définie sur B est une fonction borélienne positive dont les épigraphes $\{x \in B \mid U(x) \leq R\}$ $(R > 0)$ sont relativement compacts.

Définition 1.8 *Soit U une fonction de Lyapounov sur B. On définit les* ***classes de Chernov*** *associées à la fonction U suivantes :*

$\mathcal{S}(U) \overset{\triangle}{=} \left\{A \text{ borélien de } B \times \mathbb{R}_+ \mid \lim_{R\to\infty} \inf \left\{\frac{U(s)}{v} \mid (s,v) \in A,\ v > R\right\} > 0\right\};$

$\mathcal{T}(U) \overset{\triangle}{=} \left\{A_\rho = \{(s,v) \in B \times \mathbb{R}_+ \mid U(s) \geq \rho v\} \mid \rho > 0\right\}.$

Voici un premier critère de tension exponentielle partielle.

Proposition 1.7 *Dans le cadre de la définition précédente, on considère une suite de probabilités $(\mu_n)_n$ sur $B \times \mathbb{R}_+$.*

1) *La tension exponentielle partielle relative à $\mathcal{S}(U)$ est assurée par la propriété suivante ;*

$$\forall b > 0,\ \lim_{r\to\infty} \limsup \frac{1}{v_n} \log \mu_n \left(\{(s,v) \mid\ U(s) > r + bv\} \right) = -\infty. \tag{1.7}$$

2) *Lorsque $B = \mathbb{R}^d$, la tension exponentielle partielle de $(\mu_n)_n$ relative à $\mathcal{S}(U)$ équivaut à la tension exponentielle partielle de $(\mu_n)_n$ relative à $\mathcal{T}(U)$.*

1.4.3 PGD autonormalisées

Soit (E, d) un espace métrique polonais.

Définition 1.9 *Une variable aléatoire L à valeurs dans l'espace polonais $\mathcal{P}(E)$ sera appelée **probabilité aléatoire sur** (E, d)*

Le cadre dans lequel nous obtiendrons dans le chapitre 4 des PGD relatifs à une classe de Chernov est le suivant.

Axiome 1.2 *On suppose que*
- *$(L_n)_n$ est une suite de probabilités aléatoires définies sur l'espace polonais (E, d) ;*
- *V est une fonction positive définie sur (E, d) ;*
- *f est une fonction continue de (E, d) dans un espace de Banach séparable B ;*
- *U est une fonction de Lyapounov définie sur B.*

Voici un cas particulier fréquent.

Définition 1.10 (PGD autonormalisé) *Si $U \circ f$ est une fonction de Lyapounov et si la suite des couples $\left(\int f dL_n, \int U \circ f dL_n\right)_n$ satisfait un PGD partiel relatif à $\mathcal{S}(U)$, nous dirons que la suite $(\int f dL_n)_n$ vérifie un **PGD autonormalisé** relatif à U.*

Remarque 1.3 *Ce terme sera parfois utilisé lorsque $U \circ f \leq MV$ $(0 < M < +\infty)$ et la suite $\left(\int f dL_n, \int V dL_n\right)_n$ satisfait un PGD partiel relatif à $\mathcal{S}(U)$.*

On étudie les PGD *de type Chernov* autonormalisés du type :

$$\lim_n \frac{1}{n} \log \mathbb{P} \left(U \left(\int f dL_n \right) > \rho \int V dL_n \right) = -I(\rho).$$

Considérant la loi $(\mu_n)_n$ du couple $(\int f dL_n, \int V dL_n)$, il s'agit donc de prouver un PGD partiel relatif aux classes de Chernov associées à U. Cette preuve comportera deux étapes :
- la preuve d'un PGD vague *image* d'un PGD vague satisfait par la suite $(L_n)_n$,
- la preuve de la tension exponentielle partielle relative aux classes de Chernov.

La première étape sera l'objet des sections 3.2, 3.3 et 4.2. et nous abordons la seconde étape dans la section suivante.

1.4.4 Tension exponentielle partielle pour les classes de Chernov

En se situant dans le cadre de l'axiome 1.2 de 1.4.3, on donne quelques conditions suffisantes de tension exponentielle partielle de vitesse (v_n), relatives aux classes de Chernov.

Proposition 1.8 *Voici deux conditions suffisantes à la tension exponentielle de* $(\int f dL_n, \int V dL_n)_n$ *de vitesse* $(v_n)_n$*, relative à* $\mathcal{S}(U)$*.*

CS1. On a, pour tout $b > 0$*,*

$$\lim_{r\to\infty} \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) > r + b\int V dL_n\right) = -\infty.$$

CS2. Les deux conditions suivantes sont réalisées :

$$\forall R > 0,\ \lim_{r\to\infty} \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) > r,\ \int V dL_n < R\right) = -\infty,$$

$$\forall b > 0,\ \lim_{R\to\infty} \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) > b\int V dL_n \geq bR\right) = -\infty.$$

CS3. Dans le cas particulier où U *est une fonction de Lyapounov convexe, si* $U \circ f \leq \alpha V + \beta$ *pour des constantes* $\alpha > 0$ *et* $\beta > 0$*, la condition suivante est suffisante :*

$$\forall b > 0,\ \lim_{r\to\infty} \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) > b\int V dL_n \geq bR\right) = -\infty.$$

Voici maintenant deux corollaires utilisant la définition suivante.

Définition 1.11 (Troncature exponentielle) *Une fonction borélienne* $\phi : E \to B$ *satisfait la condition de* ***troncature exponentielle de vitesse*** $(v_n)_n$ *si :*

$$\forall r > 0,\ \lim_{a\to\infty} \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(\int \mathbb{I}_{\|\phi\|\geq a} dL_n \geq r\right) = -\infty. \tag{1.8}$$

Corollaire 1.2 *Pour* $B = \mathbb{R}^d$ *et* $p > 1$*, dès que* f *satisfait la condition de troncature exponentielle, la tension exponentielle de* $(\int f dL_n, \int \|f\|^p dL_n)$ *relative à* $\mathcal{S}(\|\cdot\|^p)$ *est assurée (avec la même vitesse).*

Le corollaire précédent s'interprète facilement et sa preuve est assez simple. Il aurait aussi bien pu être déduit du corollaire suivant.

Corollaire 1.3 *Pour* B *espace de Banach séparable, on considère* U*, fonction de Lyapounov convexe et s.c.i. définie sur* B*, et* $h : [0,\infty] \to [0,\infty]$ *convexe, croissante et finie sur* $[0,\infty[$*, telle que, si* t *et* z *tendent vers l'infini,*

$$\frac{h(tz)}{th(z)} \to \infty.$$

Alors, dès que f *satisfait la condition de troncature exponentielle, la tension exponentielle de la suite* $(\int f dL_n, \int h \circ U(f) dL_n)$ *relative à* $\mathcal{S}(h \circ U)$ *est assurée (avec la même vitesse).*

Références Les notions de classe de Chernov et de la tension exponentielle partielle correspondante, en vue du cas i.i.d, viennent de [DemSha98b]. Le corollaire 1.3 est dû, dans un cadre semblable, à Dembo et Shao [DemSha98b]. La condition CS1 apparaît dans [DemSha98a] et la condition CS3 dans [DemSha98b]. Il nous a semblé utile de reprendre ces résultats de manière plus détaillée tout en élargissant leur domaine d'application.

1.4.5 Preuves des résultats de 1.4.2 et 1.4.4

Preuve de la proposition 1.7

La partie 2) résulte du lemme 1.3, appliqué à la fonction $F(s,v) = U(s)/v$.
Prouvons la partie 1). L'adhérence $K_{r,b}$ de $\{(s,v) \mid U(s) \leq r + bv,\ v \leq \frac{r}{b}\}$ est compacte.
On aura donc montré le résultat si, pour tout $A \in \mathcal{S}(U)$, on peut choisir $b > 0$ et r_0 tels que pour tout $r \geq r_0$ on ait :

$$A \cap \left\{(s,v) \mid U(s) \leq r + bv,\ v \leq \frac{r}{b}\right\}^c \subset \{(s,v) \mid U(s) > r + bv\}.$$

Ceci est bien réalisé.
En effet, comme $A \in \mathcal{S}(U)$, il existe $\rho > 0$ et $R > 0$ tels que l'on ait $\inf \left\{\frac{U(s)}{v} \mid (s,v) \in A,\ v > R\right\} > \rho$. Pour $b > 0$ arbitraire et $r > bR = r_0$,

$$A \cap \left\{(s,v) \mid U(s) \leq r + bv,\ v > \frac{r}{b}\right\} \subset A \cap \{(s,v) \mid U(s) \leq r + bv,\ r < bv,\ U(s) > \rho v\}.$$

Il suffit de choisir $b < \rho/2$ pour que cet ensemble soit vide, ce qui implique, pour $r \geq r_0$,

$$A \cap \left\{(s,v) \mid U(s) \leq r + bv,\ v \leq \frac{r}{b}\right\}^c \subset \{(s,v) \mid U(s) > r + bv\} \ \diamondsuit$$

Preuve de la proposition 1.8.

a) CS1 résulte de la proposition 1.7.

b) Prouvons la condition CS2. Soit $A \in \mathcal{S}(U)$. Pour tous $r > 0$ et $R > 0$, on note $b(R) = \inf \{U(s)/v \mid (s,v) \in A,\ v > R\}$; $b(R)$ décroît vers $b > 0$, dépendant de A, quand $R \to \infty$. Pour le compact $K_{r,b}$ défini dans la preuve de la proposition 1.7, on a :

$$A \cap \{(s,v) \mid v > R\} \subset \{(s,v) \mid U(s) \geq vb(R) > Rb(R)\},$$
$$A \cap K_{r,R}^c \subset \{(s,v) \mid U(s) \geq vb \geq Rb\} \cup \{(s,v) \mid U(s) > r,\ v \leq R\}.$$

Donc,

$$\begin{aligned} & \limsup_n \frac{1}{n} \log \mathbb{P}\left(\left(\int f dL_n, \int V dL_n\right) \in A \cap K_{r,R}^c\right) \\ \leq\ & \max\left\{\limsup_n \frac{1}{n} \log \mathbb{P}\left(U\left(\int f dL_n\right) > b \int V dL_n \geq bR\right),\right. \\ & \left.\limsup_n \frac{1}{n} \log \mathbb{P}\left(U\left(\int f dL_n\right) > r,\ \int V dL_n \leq R\right)\right\}. \end{aligned}$$

c) Dans le cadre de CS3, on a $U\left(\int f dL_n\right) \leq \int U \circ f dL_n \leq \alpha \int V dL_n + \beta$. La première des deux conditions de la condition suffisante CS2) disparaît. $\diamondsuit$

Preuve des corollaires 1.2. et 1.3.

On peut, dans ces preuves, simplifier les notations en considérant les probabilités aléatoires $\Lambda_n = f(L_n)$.

♣ **Preuve du corollaire 1.2, à l'aide de CS1.** Pour tous $b > 0$ et $R < \infty$, l'hypothèse de troncature (**??**) implique qu'il existe $A > 0$ et N entier tels que l'on ait, pour tout $n \geq N$,

$$\mathbb{P}\left(\int \mathbb{I}_{\|x\|\geq A}\Lambda_n(dx) \geq b\right) \leq e^{-Rv_n}.$$

Soit q l'exposant conjugué de p. Par l'inégalité de Hőlder, en notant $V_n = \int \|x\|^p\Lambda_n(dx)$, pour $n \geq N$,

$$\int \|x\|\mathbb{I}_{\|x\|\geq A}\Lambda_n(dx) \leq V_n^{1/p}\left(\int \mathbb{I}_{\|x\|\geq A}\Lambda_n(dx)\right)^{1/q},$$

$$\mathbb{P}\left(\int \|x\|\mathbb{I}_{\|x\|\geq A}\Lambda_n(dx) \geq b^{1/q}V_n^{1/p}\right) \leq \mathbb{P}\left(\int \mathbb{I}_{\|x\|\geq A}\Lambda_n(dx) \geq b\right) \leq e^{-Rv_n}.$$

Or :

$$\left\{\int \|x\|\Lambda_n(dx) \geq r + b^{1/q}V_n^{1/p}\right\} \subset \left\{\int \|x\|\mathbb{I}_{\|x\|\leq A}\Lambda_n(dx) \geq r\right\} \cup \left\{\int \|x\|\mathbb{I}_{\|x\|\geq A}\Lambda_n(dx) \geq b^{1/q}V_n^{1/p}\right\}.$$

L'événement $\left\{\int \|x\|\mathbb{I}_{\|x\|\leq A}\Lambda_n(dx) \geq r\right\}$ étant vide pour $r > A$, on a

$$\forall r > A,\ \limsup_{n\to\infty} \frac{1}{v_n}\log\mathbb{P}\left(\int \|x\|\Lambda_n(dx) \geq r + b^{1/q}V_n^{1/p}\right) \leq -R.$$

D'autre part,

$$\begin{aligned}\left\{\left\|\int x\Lambda_n(dx)\right\|^p \geq 2^{p-1}r^p + 2^{p-1}b^{p/q}V_n\right\} &\subset \left\{\left\|\int x\Lambda_n(dx)\right\|^p \geq (r + b^{1/q}V_n^{1/p})^p\right\} \\ &\subset \left\{\int \|x\|\Lambda_n(dx) \geq r + b^{1/q}V_n^{1/p}\right\}\end{aligned}$$

d'où le résultat puisque $2^{p-1}b^{p/q}$ balaie toutes les valeurs possibles de $\mathbb{R}_+^*$. ♢

♠ **Preuve du corollaire 1.3, à l'aide de CS3.**

Notons $V = h \circ U$. La fonction h est minorée par une fonction linéaire de dérivée strictement positive, donc elle croît vers $+\infty$; ses épigraphes, contenus dans un épigraphe de U, sont relativement compacts.
De plus, h étant croissante et U étant convexe, on a, par l'inégalité de Jensen, $V\left(\int f dL_n\right) \geq h\left(U \circ f dL_n\right)$. Notons, pour $b > 0$, $r > 0$ et $n \in \mathbb{N}$,

$$U_n = \int U\circ f dL_n = \int U(x)\Lambda_n(dx),\ V_n = \int V\circ f dL_n = \int V d\Lambda_n \ \text{ et } A_{n,r,b} = \{h(U_n) \geq bV_n \geq br\}.$$

On aura prouvé que l'on est dans le cadre de CS3, si

$$\forall b > 0,\ \lim_{r\to\infty} \limsup_{n} \frac{1}{v_n}\log\mathbb{P}\left(A_{n,r,b}\right) = -\infty.$$

On pose $m_r = \inf\{z/4 \mid h(z) \geq br\}$. Sur $A_{n,r,b}$, $U_n \geq 4m_r$ et $h(U_n) \geq bV_n$ donc :

$$A_{n,r,b} \subset \left\{ \int \left(\frac{U(x)}{U_n} - \frac{bV(x)}{2h(U_n)} \right) \Lambda_n(dx) \geq 1 - \frac{1}{2} = \frac{1}{2} \right\}.$$

Posons $a_r \stackrel{\triangle}{=} \sup\left\{ \frac{U(x)}{z} - \frac{bV(x)}{2h(z)} \mid U(x) \geq m_r, z \geq m_r \right\}$ et $W_{n,r} \stackrel{\triangle}{=} \int b\mathbb{I}_{U(x)\geq m_r} \Lambda_n(dx)$. On a, sur $A_{n,r,b}$,

$$\begin{aligned} W_{n,r}a_r &\geq \int \mathbb{I}_{U(x)\geq m_r} \left(\frac{U(x)}{U_n} - \frac{bV(x)}{2h(U_n)} \right) \Lambda_n(dx) \\ &\geq \int \left(\frac{U(x)}{U_n} - \frac{bV(x)}{2h(U_n)} \right) \Lambda_n(dx) - \int \mathbb{I}_{U(x)\leq m_r} \left(\frac{U(x)}{U_n} - \frac{bV(x)}{2h(U_n)} \right) \Lambda_n(dx) \\ &\geq \frac{1}{2} - \frac{1}{4} = \frac{1}{4}. \end{aligned}$$

On aboutit à l'inclusion suivante : $A_{n,r,b} \subset \left\{ W_{n,r} \geq \frac{1}{4a_r} \right\}$.
Si $r \to \infty$, $m_r \to \infty$. Par conséquent, l'hypothèse de troncature exponentielle permet de conclure si, pour tout $b > 0$, il existe un $r(b)$ tel que $r \geq r(b)$ implique $a_r \leq a(b) < \infty$; en effet, on a alors

$$\mathbb{P}(A_{n,r,b}) \leq \mathbb{P}\left(W_{n,r} \geq \frac{1}{4a_r} \right) \leq \mathbb{P}\left(\int \mathbb{I}_{U\circ f \geq m_r} dL_n \geq \frac{1}{4ba(b)} \right).$$

Soit $b > 0$; prouvons donc qu'il existe un $r(b)$ tel que $r \geq r(b)$ implique $a(r) \leq a(b) < \infty$. Fixant $z \geq m_r$, quel que soit $t > 0$,

$$\begin{aligned} \sup\left\{ \frac{U(x)}{z} - \frac{bV(x)}{2h(z)} \mid U(x) \geq m_r \right\} &\leq \sup\left\{ u\left(1 - \frac{bh(zu)}{2uh(z)}\right) \mid u \geq m_r/z \right\} \\ &\leq \sup\left\{ t, \sup\left\{ u\left(1 - \frac{bh(zu)}{2uh(z)}\right) \mid u \geq t \right\} \right\}. \end{aligned}$$

Donc, quel que soit $t > 0$,

$$a_r \leq \sup\left\{ t, \sup\left\{ u\left(1 - \frac{bh(zu)}{2uh(z)}\right) \mid u \geq t,\ z \geq m_r \right\} \right\}.$$

Mais l'hypothèse sur h signifie qu'il existe un $r(b)$ et un $a(b)$ tels que $z \geq m_{r(b)}$ et $u \geq t \geq a(b)$, impliquent $h(zu)/uh(z) \geq 2/b$. Ainsi, pour $r > r(b)$, on a $a_r \leq a(b)$, ce qui achève la preuve. $\diamondsuit$

Chapitre 2

Autour de Donsker-Varadhan

2.1 Introduction

2.1.1 Objectifs de ce chapitre

L'objectif, dans ce chapitre, est de mieux situer les travaux de Donsker et Varadhan par rapport aux résultats plus récents obtenus dans le cadre de modèles markoviens, notamment ceux de De Acosta.
En premier lieu, on s'intéresse naturellement au taux utilisé par Donsker et Varadhan, ce qui nécessite une bonne connaissance des propriétés de l'information de Kullback et des principes variationnels qui lui sont associés. En effet, leur taux est défini par une formule variationnelle faisant intervenir la probabilité de transition. On essaye par conséquent d'étudier sa forme, suivant les hypothèses faites sur la probabilité de transition p, (en particulier son "domaine de continuité" $\mathcal{D}(p)$).
Dans bien des cas, le taux de Donsker-Varadhan n'est pas facilement manipulable. Pour cette raison, il peut être plus agréable de travailler avec une version bi-dimensionnelle moins abstraite qui apparaît comme une distance de la loi à la probabilité de transition. On montre que le taux de Donsker-Varadhan est égal à une certaine projection du taux bi-dimensionnel, ce qui rend les résultats compatibles. Ce résultat de projection est très important et est dû à Donsker et Varadhan à l'origine. Cependant, on en donne une preuve plus simple que l'on détaille.
On effectue enfin une comparaison des taux de Donsker-Varadhan avec le taux spectral, suivant ce qui a été supposé sur la probabilité de transition p.

Dans un second temps, on aborde des PGD supérieurs pour la suite des lois empiriques d'une chaîne de Markov. Là encore, l'étude est guidée essentiellement par les résultats de Donsker et Varadhan, avec quelques améliorations pour l'uniformité par rapport à des familles de lois initiales, ou encore sur l'hypothèse de régularité de la probabilité de transition.

2.1.2 Plan

Dans la section 2.2, on fait de multiples rappels autour du taux de Donsker-Varadhan

sur un espace polonais, puis on le compare au taux spectral usuel.
Un petit inventaire des PGD supérieurs vagues ou étroits pour les lois empiriques est fait dans la section 2.3, en mettant l'accent sur les résultats dus à Donsker et Varadhan .
Enfin, on donne dans la section 2.4 quelques commentaires sur les PGD obtenus pour les lois empiriques, agrémentés d'exemples pour illustrer nos propos.

2.1.3 Notations

Soient $(E_1, \mathcal{E}_1)$ et $(E_2, \mathcal{E}_2)$ deux espaces mesurables. Une probabilité de transition p de E_1 vers E_2 est notée $p : E_1 \hookrightarrow E_2$. Pour une probabilité α sur E_1,

$$\alpha \oslash p(dx, dy) \stackrel{\triangle}{=} \alpha(dx)p(x, dy) \text{ et } \alpha p(dy) = \int \alpha(dx)p(x, dy).$$

Pour une fonction g (resp. G) borélienne de E_2 (resp. de $E_1 \times E_2$) sur $\mathbb{R}^d$,

$$pg(x) \stackrel{\triangle}{=} \int p(x, dy)g(y) \text{ et } pG(x) \stackrel{\triangle}{=} \int p(x, dy)G(x, y),$$

quand ces expressions ont un sens.
Pour $\zeta \in \mathcal{P}(E_1 \times E_2)$, ζ^1 et ζ^2 désignent la première et la seconde loi marginale.

2.2 Taux markoviens

2.2.1 Kullback et les principes variationnels

Les premiers résultats variationnels sont valables dans un espace quelconque. Une structure d'espace polonais est ensuite requise pour affiner ces formules.
On considère un espace $(E, \mathcal{E})$ mesurable et $\mathcal{P}(E)$ l'espace des probabilités sur l'espace E.

Définition 2.1 *Pour μ et ν deux éléments de $\mathcal{P}(E)$, l'entropie (ou information de Kullback) de ν relative à μ est donnée par la formule suivante :*

$$K(\nu \mid \mu) \stackrel{\triangle}{=} \begin{cases} \int \log\left(\frac{d\nu}{d\mu}\right) d\nu & \text{si } \nu << \mu \\ \infty & \text{sinon.} \end{cases}$$

Pour $\mu \in \mathcal{P}(E)$, le principe variationnel le plus simple est le suivant.

Proposition 2.1
1) *Soit g une fonction borélienne telle que $\int e^g d\mu < \infty$. On a*

$$\log \int e^g d\mu = \sup\left\{\int g d\nu - K(\nu \mid \mu) \mid \nu << \mu\right\}.$$

2) *Soit ν une probabilité sur E absolument continue par rapport à μ. On a*

$$K(\nu \mid \mu) = \sup\left\{\int g d\nu - \log \int e^g d\mu \mid g \text{ borélienne telle que } \int e^g d\mu < \infty\right\}.$$

La partie 2) de la proposition 2.1 a été complétée par Donsker-Varadhan [DonVar76].

Proposition 2.2 (Principe variationnel de Donsker-Varadhan sur un espace quelconque)

Soient deux probabilités quelconques μ et ν sur E. On a

$$K(\nu \mid \mu) = \sup\left\{\int g d\nu - \log \int e^g d\mu \mid g \in \mathcal{B}_b(E)\right\}.$$

Remarque 2.1 *On obtient facilement, par passage à la limite monotone, la majoration suivante. Pour toute fonction borélienne g bornée inférieurement sur E,*

$$K(\nu \mid \mu) \geq \int g d\nu - \log \int e^g d\mu.$$

La formule variationnelle suivante est due à Donsker-Varadhan (voir [DonVar75a] et [DonVar76] pour la formulation originale).

Proposition 2.3 (Principe variationnel de Donsker-Varadhan sur un polonais) *Sur un espace polonais (E, d), l'entropie relative vérifie les propriétés suivantes.*

a) Pour tout μ et ν appartenant à $\mathcal{P}(E)$, on a

$$\begin{aligned} K(\nu \mid \mu) &= \sup_{g \in \mathcal{B}_b(E)} \left\{\int g d\nu - \log \int e^g d\mu\right\} \\ &= \sup_{g \in \mathcal{C}_b(E)} \left\{\int g d\nu - \log \int e^g d\mu\right\} \\ &= \sup_{g \in \mathcal{LB}(E)} \left\{\int g d\nu - \log \int e^g d\mu\right\}. \end{aligned}$$

b) $(\mu, \nu) \in \mathcal{P}(E) \times \mathcal{P}(E) \to K(\nu \mid \mu)$ est une application convexe et semi-continue inférieurement. De plus, à μ fixée, la fonction $\nu \mapsto K(\nu \mid \mu)$ est à épigraphes compacts et strictement convexe sur l'ensemble $\{\nu \in \mathcal{P}(E) \mid K(\nu \mid \mu) < \infty\}$.

Remarque 2.2 *Pour un espace polonais E, il existe une distance d_1 compatible avec sa topologie et pour laquelle E est totalement borné (voir par exemple [Dud89], théorème 2.8.2, p.54). Cela implique qu'il existe une partie $\mathcal{H}$ dénombrable dense, pour la norme uniforme, dans $\mathcal{LB}(E, d_1)$ (voir [Str93], lemme 3.1.4.). Donc on a*

$$\begin{aligned} K(\nu \mid \mu) &= \sup_{g \in \mathcal{LB}b(E,d_1)} \left\{\int g d\nu - \log \int e^g d\mu\right\} \\ &= \sup_{g \in \mathcal{H}} \left\{\int g d\nu - \log \int e^g d\mu\right\}. \end{aligned}$$

Proposition 2.4 *Soient (E,d) un espace polonais et $g : E \to \mathbb{R}$ une application mesurable qui est bornée inférieurement ou bornée supérieurement ; μ étant un élément de $\mathcal{P}(E)$, on a*

$$\log \int e^g d\mu = \sup \left\{ \int g d\nu - K(\nu \mid \mu) \mid \nu << \mu \right\}.$$

En conséquence, si g est une fonction borélienne réelle majorée par une constante sur l'espace polonais (E,d) on a, pour tout $\nu \in \mathcal{P}(E)$,

$$K(\nu \mid \mu) \geq \int g d\nu - \log \int e^g d\mu.$$

Donc, en notant $\mathcal{B}_{maj}(E)$ l'ensemble de ces fonctions,

$$K(\nu \mid \mu) = \sup_{g \in \mathcal{B}_{maj}(E)} \left\{ \int g d\nu - \log \int e^g d\mu \right\}.$$

De même pour les fonctions boréliennes réelles minorées.

2.2.2 Kullback sur un espace produit

♠ Kullback sur un produit d'espaces polonais
Lorsque (E_1,d_1) et (E_2,d_2) sont deux espaces polonais, on désigne par $D(p)$ l'ensemble des discontinuités de la fonction $x \mapsto p(x,\cdot)$ de E_1 dans $\mathcal{P}(E_2)$ muni de la topologie de la convergence étroite ; $\mathcal{D}(p)$ est un borélien de E_1.

Lemme 2.1
a) Si $g \in \mathcal{C}_b(E_2)$ (resp. $G \in \mathcal{C}_b(E_1 \times E_2)$), pg (resp pG) est continue en tout $x \notin D(p)$.
b) La fonction de $\mathcal{P}(E_1)$ dans $\mathcal{P}(E_1 \times E_2)$, $\alpha \mapsto \alpha \oslash p$ est continue sur l'ensemble $\{\alpha \mid \alpha(D(p)) = 0\}$.

Lemme 2.2 *Pour deux probabilités de transition p et $q : E_1 \hookrightarrow E_2$, on a les propriétés suivantes :*
a) $x \mapsto K(q(x,\cdot) \mid p(x,\cdot))$ est borélienne ;
b) pour ν et μ dans $\mathcal{P}(E_1)$,

$$K(\nu \oslash q \mid \mu \oslash p) = K(\nu \mid \mu) + \int K(q(x,\cdot) \mid p(x,\cdot))\, \nu(dx). \tag{2.1}$$

Si (E_1,d_1) et (E_2,d_2) sont deux espaces polonais, selon un résultat classique portant sur les probabilités conditionnelles régulières (voir, par exemple [Dud89] ou [Str93]), quel que soit $\zeta \in \mathcal{P}(E_1 \times E_2)$, il existe une probabilité de transition p telle que $\zeta = \zeta^1 \oslash p$ et $\zeta^2 = \zeta^1 p$. La partie b) du lemme précédent s'applique à tout couple (ζ,ξ) de probabilités de $\mathcal{P}(E_1 \times E_2)$.

♢ "Projection" pour des produits d'espaces polonais
Si $\mu \in \mathcal{P}(E_1)$ et $\nu \in \mathcal{P}(E_2)$, $\mathcal{C}(\mu,\nu)$ désigne l'ensemble des couplages de μ et de ν :

$$\mathcal{C}(\mu,\nu) = \left\{ \zeta \in \mathcal{P}(E_1 \times E_2) \mid \zeta^1 = \mu, \zeta^2 = \nu \right\}.$$

On suppose ici que (E_1, d_1) et (E_2, d_2) sont polonais. Alors μ et ν sont tendues, d'où il résulte facilement que $\mathcal{C}(\mu, \nu)$ est une partie tendue et fermée donc compacte de $\mathcal{P}(E_1 \times E_2)$.

Lemme 2.3 *La fonction $(\zeta, \mu) \mapsto K(\zeta \mid \mu \oslash p)$ est convexe sur $\mathcal{P}(E_1 \times E_2) \times \mathcal{P}(E_1)$. Elle est en outre s.c.i. en tout point (ζ, μ) tel que $\mu(D(p)) = 0$.*

Lemme 2.4 *Lorsque $\zeta^1 = \mu$, on a*

$$K(\zeta \mid \mu \oslash p) = \sup_{G \in \chi} \left\{ \int \Big(G(x,y) - \log p(e^G)(x) \Big) \zeta(dx, dy) \right\}$$

où $\chi = \mathcal{B}_{maj}(E_1 \times E_2)$, $\mathcal{B}_b(E_1 \times E_2)$, $\mathcal{C}_{maj}(E_1 \times E_2)$, $\mathcal{C}_b(E_1 \times E_2)$ ou $\mathcal{LB}(E_1 \times E_2)$.

Théorème 2.1 (Projection) *Soit*

$$\overline{I}(\mu, \nu) \stackrel{\triangle}{=} \inf \left\{ K(\zeta \mid \mu \oslash p) \mid \zeta \in \mathcal{C}(\mu, \nu) \right\}.$$

Lorsque $\mu(D(p)) = 0$, on a

$$\overline{I}(\mu, \nu) = \sup_{g \in \mathcal{C}_b(E_2)} \left\{ \int g d\nu - \int \log p(e^g) d\mu \right\} = \sup_{g \in \mathcal{B}_{maj}(E_2)} \left\{ \int g d\nu - \int \log p(e^g) d\mu \right\}.$$

2.2.3 Taux de Donsker-Varadhan sur un espace polonais

On introduit maintenant les taux de Donsker-Varadhan intervenant dans les PGD pour les lois empiriques d'une chaîne de Markov.

♣ Traduction des résultats de 2.2.

On se place sur un espace polonais (E, d) et $p : E \hookrightarrow E$ est une probabilité de transition. On donne ici la définition du taux de Donsker-Varadhan "bi-dimensionnel" qui est heuristiquement une distance de la loi ζ à la probabilité de transition p. Comme nous l'avons vu en introduction du chapitre, il est moins abstrait et d'un usage souvent plus agréable que le taux de Donsker-Varadhan unidimensionnel classiquement défini par la formule donnée dans le théorème 2.1

$$I^1(\nu) = \sup_{g \in \mathcal{C}_b(E)} \left\{ \int (g - \log p(e^g)) d\nu \right\}.$$

Définition 2.2

1) *Le taux de Donsker-Varadhan bi-dimensionnel est l'application définie sur $\mathcal{P}(E^2)$ par*

$$I(\zeta) \stackrel{\triangle}{=} \begin{cases} K(\zeta \mid \zeta^1 \oslash p) = \sup_{G \in \mathcal{C}_b(E^2)} \left\{ \int G d\zeta - \log \int e^G d\zeta^1 \oslash p \right\} & si\ \zeta^1 = \zeta^2 \\ \infty & sinon\ . \end{cases}$$

2) *Le taux de Donsker-Varadhan "projeté" est défini sur $\mathcal{P}(E)$ par*

$$I^1(\nu) \stackrel{\triangle}{=} \inf \left\{ I(\zeta) \mid \zeta \in \mathcal{C}(\nu, \nu) \right\}.$$

En corollaire du lemme 2.4 et du théorème 2.1, on obtient le corollaire suivant.

Corollaire 2.1

1) *Si* $\zeta^1 = \zeta^2$, *on a*

$$\begin{aligned} I(\zeta) &= \sup_{G \in \mathcal{B}_{maj}(E^2)} \left\{ \int \log \left(\frac{e^{G(x,y)}}{p(e^G)(x)} \right) \zeta(dx, dy) \right\} \\ &= \sup_{G \in \mathcal{LB}(E^2)} \left\{ \int \log \left(\frac{e^{G(x,y)}}{p(e^G)(x)} \right) \zeta(dx, dy) \right\}. \end{aligned}$$

Donc $\zeta \mapsto I(\zeta)$ *est une application convexe et s.c.i. sur* $\{\zeta \in \mathcal{P}(E^2) \mid \zeta^1(D(p)) = 0\}$.

2) *Quel que soit* ν *tel que* $\nu(D(p)) = 0$, *on a*

$$I^1(\nu) = \sup_{g \in \mathcal{C}_{maj}(E)} \left\{ \int (g - \log p(e^g)) d\nu \right\} = \sup_{g \in \mathcal{C}_b(E)} \left\{ \int (g - \log p(e^g)) d\nu \right\}.$$

Donc $\nu \mapsto I^1(\nu)$ *est s.c.i. sur* $\{\nu \mid \nu(D(p)) = 0\}$.

Remarque 2.3 *Si* $(X_n)_n$ *est une chaîne de Markov de transition* p, *la chaîne de Markov bi-dimensionnelle* $(X_n, X_{n+1})_n$ *a la transition* $\delta_x \oslash p$ *et on a* $I_p(\zeta) = I^1_{\delta_x \oslash p}(\zeta)$ *lorsque* $\zeta^1 = \zeta^2$.

2.2.4 Transitions fellériennes et presque-fellériennes

Voici la définition classique de transition fellérienne, ainsi que l'extension "presque fellérienne" proposée par Dupuis et Ellis (voir [DupEll97], chapitre 9).

Définition 2.3

- *Une probabilité de transition* $p : E \hookrightarrow E$ *est* ***fellérienne*** *si son domaine de discontinuité* $D(p)$ *est vide.*
- *Elle est* ***presque fellérienne*** *si, quels que soient* $x \in E$ *et* $\delta > 0$, *il existe un ouvert* G_δ *de* E *contenant* $D(p)$ *et un réel* $r(x) > 0$ *tels que, pour tout* $y \in E$,

$$d(y, x) \leq r(x) \Rightarrow p(y, G_\delta) < \delta.$$

En particulier, cela implique $p(x, D(p)) = 0$, *pour tout* x.

Remarque 2.4 *Une probabilité de transition* p *est fellérienne si* $p\mathcal{C}_b(E) \subset \mathcal{C}_b(E)$; *par le lemme 2.1, on a aussi* $p\mathcal{C}_b(E^2) \subset \mathcal{C}_b(E^2)$. *Soit* $\mathcal{C}_b^*(E)$ *(resp.* $\mathcal{C}_b^*(E^2)$*) l'ensemble des fonctions de* E *dans* $\mathbb{R}$ *(resp. de* E^2 *dans* $\mathbb{R}$*) bornées et continues hors de* $D(p)$ *(resp. hors de* $D^2(p)$*). Si* p *est une probabilité de transition presque fellérienne, on a* $p\mathcal{C}_b^*(E) \subset \mathcal{C}_b^*(E)$ *(resp.* $p\mathcal{C}_b^*(E^2) \subset \mathcal{C}_b^*(E^2)$*).*

Les exemples suivants sont dus à Julien Worms et étudiés par lui dans [Wor00] et [Wor99].

Exemple 2.1 (Transition dominée par une même mesure) *Prenons $E = \mathbb{R}^d$ et γ une mesure σ-finie sur E. On suppose que, pour tout x,*

$$p(x, dy) = \rho(x, y)\gamma(dy),$$

où ρ est une fonction borélienne et majorée par une constante R. On suppose enfin que $\gamma(D(p)) = 0$. Alors p est presque fellérienne. Il existe en effet, pour tout $\delta > 0$ donné, un ouvert G_δ contenant $D(p)$ et tel que $\gamma(G_\delta) < \frac{\delta}{R}$. Ceci entraine que $p(x, G_\delta) < \delta$.

Exemple 2.2 (Modèle ARF(1)) *Soit un modèle ARF(1) à valeurs dans $\mathbb{R}^d$, défini pour $n \geq 1$ par*

$$X_n = f(X_{n-1}) + \sigma(X_{n-1})\varepsilon_n,$$

où $(\varepsilon_n)_n$ est une suite i.i.d. de loi ξ et indépendante de X_0. f est une fonction borélienne, d'ensemble de discontinuité noté $\mathcal{D}_f$ et $\sigma : \mathbb{R}^d \to [a, \infty[$ $(a > 0)$ est borélienne, d'ensemble de discontinuité $\mathcal{D}_\sigma$. Alors, la probabilité de transition p de la chaîne de Markov $(X_n)_n$ vérifie

$$p(x, dy) = \xi\left(\frac{1}{\sigma(x)}(dy - f(x))\right); \;\; \mathcal{D}(p) \subset \mathcal{D}_f \cup \mathcal{D}_\sigma.$$

Supposons que ξ ait une densité bornée $h(\cdot)$ par rapport à la mesure de Lebesgue. Alors

$$p(x, dy) = \frac{1}{\sigma(x)} h\left(\frac{1}{\sigma(x)}(y - f(x))\right) dy,$$

et le modèle est presque fellérien dès que $\mathcal{D}_f \cup \mathcal{D}_\sigma$ est négligeable par rapport à la mesure de Lebesgue.

Théorème 2.2 *Pour une probabilité de transition presque fellérienne, on a les propriétés suivantes.*

a) Si $(\zeta_n)_n$ est une suite de $\mathcal{P}(E^2)$ convergeant vers une probabilité ζ dont les deux marginales sont égales à μ, alors on a

$$\mu(D(p)) > 0 \Rightarrow \; K\left(\zeta_n \mid \zeta_n^1 \oslash p\right) \to \infty \;\; \text{et } I(\zeta) = I^1(\mu) = \infty.$$

b) Les fonctions I et I^1 sont convexes et s.c.i. (ce sont des taux vagues convexes).

c) $I(\zeta) = 0$ si, et seulement si, $\zeta^1 = \zeta^2 = \mu$ et $\mu p = \mu$; $I^1(\mu) = 0$ si, et seulement si, $\mu p = \mu$.

2.2.5 Taux spectral et taux de Donsker-Varadhan

Soient f ou F des fonctions réelles boréliennes majorées par une constante, définies sur un espace mesurable $(E, \mathcal{E})$ ou sur $(E, \mathcal{E})^2$. Pour une probabilité de transition arbitraire p, on considère la version canonique d'une chaîne de Markov de transition p, $(\Omega, \mathcal{A}, (\mathbb{P}_x)_{x \in E})$. Pour $\phi \in \mathcal{B}_b(E)$ et pour $A \in \mathcal{E}$, on pose

$$T_F\phi(x) = \int e^{F(x,y)}\phi(y)p(x, dy) = \mathbb{E}_x\left(e^{F(x,X_1)}\phi(X_1)\right),$$

$$T_F(x,A)=\mathbb{E}_x\left(e^{F(x,X_1)}\mathbb{I}_A(X_1)\right),$$

et les formules correspondantes pour f, avec $F(x,y)=f(y)$. On a, par récurrence,

$$T_F^{(n)}(x,A)=\mathbb{E}_x\left(e^{F(x,X_1)+\ldots+F(X_{n-1},X_n)}\mathbb{I}_A(X_n)\right),$$

$$\int e^{F(x,y)}T_F^{(n-1)}(y,A)\nu(dx)p(x,dy)=\mathbb{E}_\nu\left(e^{F(X_0,X_1)+\ldots+F(X_{n-1},X_n)}\mathbb{I}_A(X_n)\right).$$

On a défini un opérateur borné sur $\mathcal{B}_b(E)$, de norme $\|T_F\|=\sup_{x\in E}T_F(x,E)$. L'inverse du rayon de convergence de la série entière $\sum\|T_F^{(n)}\|z^n$ est le rayon spectral $r(T_F)=\lim_n\|T_F^{(n)}\|^{1/n}$. On a

$$\lim_n\frac{1}{n}\log\|T_F^{(n)}\|=\log r(T_F)\stackrel{\triangle}{=}\Lambda(F),$$

$$\lim_n\frac{1}{n}\log\|T_f^{(n)}\|=\log r(T_f)\stackrel{\triangle}{=}\Lambda(f);$$

$$\mathbb{E}_\nu\left(e^{F(X_0,X_1)+\ldots+F(X_{n-1},X_n)}\right)\leq\|T_F^{(n-1)}\|\int e^{F(x,y)}\nu(dx)p(x,dy);$$

$$\limsup_n\frac{1}{n}\log\mathbb{E}_\nu\left(e^{F(X_0,X_1)+\ldots+F(X_{n-1},X_n)}\right)\leq\limsup_n\frac{1}{n}\log\int e^{F(x,y)}T_F^{(n-1)}(y,E)\nu(dx)p(x,dy)\leq\Lambda(F).$$

Proposition 2.5 (Conjugaison entre le rayon spectral et le taux de Donsker-Varadhan) *Pour $\chi\subset\mathcal{B}_{maj}(E^2)$ ou $\chi'\subset\mathcal{B}_{maj}(E)$, et pour $\zeta\in\mathcal{P}(E^2)$ ou $\zeta\in\mathcal{P}(E^2)$, notons*

$$J_\chi(\zeta)=\sup_{G\in\chi}\left\{\int Gd\zeta-\Lambda(G)\right\},\ J^1_{\chi'}(\nu)=\sup_{g\in\chi'}\left\{\int gd\nu-\Lambda(g)\right\}.$$

1) *Sur un espace mesurable quelconque, on a*

$$I(\zeta)=J_{\mathcal{B}_{maj}(E^2)}(\zeta)=J_{\mathcal{B}_b(E^2)}(\zeta),\ I^1(\nu)=J^1_{\mathcal{B}_{maj}(E)}(\nu)=J^1_{\mathcal{B}_b(E)}(\nu).$$

2) *Si E est polonais et p fellérienne,*

$$I(\zeta)=J_{\mathcal{C}_{maj}(E^2)}(\zeta)=J_{\mathcal{C}_b(E^2)}(\zeta),\ I^1(\nu)=J^1_{\mathcal{C}_{maj}(E)}(\nu)=J^1_{\mathcal{C}_b(E)}(\nu).$$

Si p est presque fellérienne, on a

$$I(\zeta)=J_{\mathcal{C}^*_{maj}(E^2)}(\zeta)=J_{\mathcal{C}^*_b(E^2)}(\zeta),\ I^1(\nu)=J^1_{\mathcal{C}^*_{maj}(E)}(\nu)=J^1_{\mathcal{C}^*_b(E)}(\nu).$$

2.2.6 Références

La plupart des principes variationnels énoncés dans les sections 2.2.1 et 2.2.2. ont été prouvés initialement par Donsker-Varadhan dans les articles [DonVar75a] et [DonVar76]. Dupuis-Ellis en font un bilan très clair (voir le chapitre 1.4 de [DupEll97], p. 32-40 pour les propositions 2.1, 2.2. et 2.3. et la proposition 4.5.1, p. 108 pour la proposition 2.4.). On pourra voir les chapitres 8 et 9 de [DupEll97] pour le lemme 2.3. et l'appendice C du même ouvrage pour le lemme 2.2.
Le théorème 2.1 de projection a été prouvé initialement par Donsker et Varadhan (voir [DonVar76], théorème 2.1) sur un compact, puis sur un polonais par compactification. Cependant la preuve donnée ici tire parti d'un théorème dû à Strassen (voir [Stra85] ou Lin[99]) qui la rend plus simple (voir la preuve en 2.2.7). En effet, le résultat est prouvé directement sur un espace polonais, sans compactification.
Pour ce qui est de la comparaison entre le taux spectral et le taux de Donsker-Varadhan, le cas fellérien, c'est à dire la relation $\sup_{g\in\mathcal{C}_b(E)}\left\{\int g d\nu - \log r_{T_g}(\nu)\right\} = I^1(\nu)$ a été prouvé par De Acosta [Aco85a] Nous reviendrons en 2.4.1 et surtout en 3.5 sur les travaux utilisant le rayon spectral pour les grandes déviations et sa comparaison avec le taux de Donsker-Varadhan.

2.2.7 Preuves des résultats de 2.2.2 et 2.2.5

♡ **Preuve du lemme 2.1**
Par la définition de $\mathcal{D}(p)$, pour $g\in\mathcal{C}_b(E_2)$, pg est continue en dehors de $\mathcal{D}(p)$. Soit $(\alpha_n)_n$ une suite de $\mathcal{P}(E_1)$; pour prouver la convergence étroite de $\alpha_n\oslash p$ vers $\alpha\oslash p$, il suffit de prouver la convergence des intégrales $\int g_1\otimes g_2 d\alpha_n\oslash p = \int g_1 p(g_2)d\alpha_n$ vers $\int g_1 p(g_2)d\alpha$, quels que soient $g_1\in\mathcal{C}_b(E_1)$, $g_2\in\mathcal{C}_b(E_2)$ (en notant $g_1\otimes g_2(x_1,x_2) = g_1(x_1)g_2(x_2)$. Pour une preuve de ce résultat, on pourra voir par exemple Ethier-Kurtz [EthKur86] p. 115). Si $(\alpha_n)_n$ converge étroitement vers α, $\lim_n \int g_1p(g_2)d\alpha_n = \int g_1 p(g_2)d\alpha$ si $g_1p(g_2)$ est α-continue donc si $\alpha(\mathcal{D}(p) = 0$. En particulier, si $(x_n)_n$ tend vers x, où x est dans le complémentaire de $\mathcal{D}(p)$, $\delta_{x_n}\oslash p$ converge étroitement vers $\delta_x\oslash p$ et, pour $G\in\mathcal{C}_b(E_1\times E_2)$, $pG(x_n)$ tend vers $pG(x)$; pG est continue en dehors de $\mathcal{D}(p)$.

♣ **Preuve du lemme 2.2.**

a) Selon la remarque 2.2. qui suit la proposition 2.3, il existe une partie $\mathcal{H}$ dénombrable de $\mathcal{C}_b(E_2)$ telle que, pour tout $x\in E_1$,

$$K\left(q(x,\cdot)\mid p(x,\cdot)\right) = \sup_{g\in\mathcal{H}}\left\{\int g(y)q(x,dy) - \log\int e^{g(y)}p(x,dy)\right\}.$$

On en déduit la mesurabilité de $x\mapsto K\left(q(x,\cdot)\mid p(x,\cdot)\right)$.

b) Montrons l'égalité (**??**). On suppose en premier lieu que le membre de droite est fini, ce qui implique $\nu << \mu$ et $p(x,\cdot) << q(x,\cdot)$ pour tout x hors d'un ensemble ν-négligeable de E_1. Moyennant une modification de $p(x,\cdot)$ sur cet ensemble (on peut le remplacer par $q(x,\cdot)$), on a $p(x,\cdot) << q(x,\cdot)$ pour tout x. On pose

$$\psi(x)\triangleq\frac{d\nu}{d\mu}(x),\quad \Phi(x,y)\triangleq\frac{dq(x,\cdot)}{dp(x,\cdot)}(y).$$

Alors, $\nu \oslash q << \mu \oslash p$ avec, pour tout $(x,y) \in E_1 \times E_2$, $\frac{d\nu\oslash q}{d\mu\oslash p}(x,y) = \psi(x)\Phi(x,y)$ et

$$\begin{aligned} K\left(\nu \oslash q \mid \mu \oslash p\right) &= \int \log\left(\psi(x)\Phi(x,y)\right)\nu(dx)p(x,dy) \\ &= \int \log\left(\psi(x)\right)\nu(dx) + \int \log\left(\Phi(x,y)\right)\nu(dx)q(x,dy) \\ &= K\left(\nu \mid \mu\right) + \int K\left(q(x,\cdot) \mid p(x,\cdot)\right)\nu(dx). \end{aligned}$$

On prouve maintenant cette égalité en supposant le membre de gauche fini. On a dans ce cas $\zeta = \nu \oslash q << \mu \oslash p = \xi$ et on peut poser :

$$\phi(x,y) \stackrel{\triangle}{=} \frac{d\zeta}{d\xi}(x,y), \quad \psi(x) \stackrel{\triangle}{=} \int_{E_2} \phi(x,y)q(x,dy), \quad \begin{aligned} q^*(x,dy) &= \frac{\phi(x,y)}{\psi(x)}p(x,dy), \text{ si } \psi(x) > 0 \\ &= 1 \text{ sinon.} \end{aligned}$$

Ainsi, $\nu = \psi\mu$. $\nu(\{\psi = 0\}) = 0$ donc $\zeta = \nu \oslash q = \nu \oslash q^*$. Par conséquent,

$$\begin{aligned} K\left(\zeta \mid \xi\right) &= \int \log(\phi(x,y))\nu \oslash q^*(dx,dy) \\ &= \int \log \psi(x)\nu(dx) + \int \log\left(\frac{\phi(x,y)}{\psi(x)}\right)\nu \oslash q^*(dx,dy) \\ &= K\left(\nu \mid \mu\right) + \int_{E_1} K\left(q^*(x,\cdot) \mid p(x,\cdot)\right)\nu(dx). \ \diamondsuit \end{aligned}$$

♠ **Preuve des lemmes 2.3. et 2.4.**

- **Lemme 2.3.**
  La fonction $(\zeta,\mu) \to K\left(\zeta \mid \mu \oslash p\right)$ est une borne supérieure pour $G \in \mathcal{C}_b(E)$ des fonctions $(\zeta,\mu) \to \int G d\zeta - \log \int p(e^G)d\mu$ qui sont convexes et continues en tout (ζ,μ) tel que $\mu(\mathcal{D}(p)) = 0$. D'où le résultat.
- **Lemme 2.4.**
  La fonction log étant concave, on a, pour tout $G \in \chi$, $\log \int p(e^G)d\mu \geq \int \log p(e^G)d\mu$ et$K\left(\zeta \mid \mu \oslash p\right) \leq \sup_{G\in\chi} \int \left(G - \log p(e^G)\right)d\zeta$.

  Montrons l'inégalité inverse. Soit $G \in \chi$. On pose $F(x,y) \stackrel{\triangle}{=} G(x,y) - \log(p(e^G)(x))$; $F \in \mathcal{B}_{maj}(E_1 \times E_2)$ et

$$\int e^{F(x,y)}\mu(dx)p(x,dy) = 1;$$

$$K\left(\zeta \mid \mu \oslash p\right) \geq \int F d\zeta - \log \int e^F d(\mu \oslash p) = \int \left(G(x,y) - \log(p(e^G)(x))\right)\zeta(dx,dy). \ \diamondsuit$$

♢ **Preuve du théorème 2.1.**

∗ ***Préliminaires***

Nous ferons appel au théorème de Strassen suivant.

Théorème (Strassen) *Soient (E_1,d_1) et (E_2,d_2), deux espaces polonais munis de leurs tribus boréliennes, A et B deux convexes de $\mathcal{P}(E_1 \times E_2)$, A supposé compact et B supposé fermé. Les deux assertions suivantes sont équivalentes.*

i) Pour tous $\zeta \in A$, $\xi \in B$, $(\zeta^1, \zeta^2) \neq (\xi^1, \xi^2)$.
ii) Il existe un couple (f, g) *de* $\mathcal{C}_b(E_1) \times \mathcal{C}_b(E_2)$ *tel que*

$$\sup_{\zeta \in A} \int (f(x) + g(y))\zeta(dx, dy) < \inf_{\xi \in B} \int (f(x) + g(y))\xi(dx, dy).$$

Prenant $A = \mathcal{C}(\mu, \nu)$ et remplaçant (f, g) par $(-f, -g)$ dans le théorème, on obtient le corollaire suivant

Corollaire *Soient* $\mu \in \mathcal{P}(E_1)$, $\nu \in \mathcal{P}(E_2)$ *et* Λ *un convexe de* $\mathcal{P}(E_1 \times E_2)$, *fermé pour la convergence étroite. Les deux propriétés suivantes sont équivalentes.*
i) $\Lambda \cap \mathcal{C}(\mu, \nu) = \emptyset$.
ii) Il existe un couple (f, g) *de fonctions de* $\mathcal{C}_b(E_1) \times \mathcal{C}_b(E_2)$, *tel que*

$$\int f d\mu + \int g d\nu > \sup_{\zeta \in \Lambda} \int (f(x) + g(y))\zeta(dx, dy).$$

$*$ ***Preuve du théorème***

Notons $J_{maj}(\mu, \nu) = \sup_{g \in \mathcal{B}_{maj}(E_2)} \{\int g d\nu - \int \log p(e^g) d\mu\}$ et $J_b(\mu, \nu) = \sup_{g \in \mathcal{C}_b(E_2)} \{\int g d\nu - \int \log p(e^g) d\mu\}$.

a) L'inégalité $\overline{I}(\mu, \nu) \geq J_{maj}(\mu, \nu)$ est la plus simple à démontrer. Soient une fonction $g \in \mathcal{B}_{maj}(E_2)$ et $\zeta \in \mathcal{C}(\mu, \nu)$.
Si on pose $G(x, y) = g(y)$, la fonction G appartient à $\mathcal{B}_{maj}(E_1 \times E_2)$ et on a

$$\int \log \frac{e^{G(x,y)}}{p(e^G)(x)} \zeta(dx, dy) = \int \log e^g(y)\nu(dy) - \int \log p(e^g)(x)\mu(dx).$$

Par conséquent, $\inf_{\zeta \in \mathcal{C}(\mu,\nu)} K\left(\zeta \mid \zeta^1 \oslash p\right) \geq \sup_{g \in \mathcal{B}_{maj}(E_2)} \{\int g d\nu - \int \log p(e^g)(x) d\mu\}$ ce qui donne le résultat recherché.

b) Montrons que $\overline{I}(\mu, \nu) \leq J_b(\mu, \nu)$. lorsque $\mu(D(p)) = 0$. Soit $\gamma < \overline{I}(\mu, \nu)$. Prouvons que $J_b(\mu, \nu) \geq \gamma$. Pour cela, on introduit l'ensemble $\Lambda \subset \mathcal{P}(E_1 \times E_2)$, défini par

$$\Lambda = \{\zeta \in \mathcal{P}(E_1 \times E_2) \mid K\left(\zeta \mid \mu \oslash p\right) \leq \gamma\}.$$

Par le lemme 2.3, c'est une partie convexe fermée de $\mathcal{P}(E_1 \times E_2)$, disjointe de $\mathcal{C}(\mu, \nu)$. Selon le théorème de Strassen, on peut donc affirmer qu'il existe deux fonctions $h \in \mathcal{C}_b(E_1)$ et $k \in \mathcal{C}_b(E_2)$ telles que

$$\int h d\mu + \int k d\nu > \sup_{\zeta \in \Lambda} \int (h(x) + k(y))\zeta(dx, dy). \tag{2.2}$$

En particulier, on peut supposer que $\int h d\mu = \int k d\nu = 0$, quite à prendre $h - \int h d\mu$ et $k - \int k d\nu$. On introduit la fonction w, définie sur $E_1 \times E_2$ par

$$w(x, y) = h \oplus k(x, y) \stackrel{\triangle}{=} h(x) + k(y).$$

C'est une fonction continue et bornée sur $E_1 \times E_2$. La transformée de Cramér de $w(\mu \otimes p)$ est donnée par

$$\Lambda^*_{w(\mu\otimes p)}(a) = \sup_{z\in\mathbb{R}} \left\{ za - \log \int e^{zw(x,y)} \mu(dx)p(x,dy) \right\} = -\inf_{z\in\mathbb{R}} \left\{ \log \int e^{z(w(x,y)-a)} \mu(dx)p(x,dy) \right\}.$$

Par la relation de dualité entre la transformée de Cramér et l'information de Kullback sur laquelle nous reviendrons dans le chapitre 3 (théorème 3.4), on a

$$\Lambda^*_{w(\mu\otimes p)}(a) = \inf \left\{ K(\zeta \mid \mu \otimes p) \mid \int w(x,y)\zeta(dx,dy) = a \right\}.$$

Or, si $\int w(x,y)\zeta(dx,dy) = 0$, l'inégalité (**??**) implique que ζ n'est pas dans Λ et donc que $K(\zeta \mid \mu \otimes p) > \gamma$. On obtient par conséquent

$$-\inf_{z\in\mathbb{R}} \left\{ \log \int e^{zw(x,y)} \mu(dx)p(x,dy) \right\} = \Lambda^*_{w(\mu\otimes p)}(0) \geq \gamma.$$

Pour tout $\eta > 0$, il existe $z_\eta \in \mathbb{R}$ tel que

$$\log \int e^{z_\eta h\oplus k(x,y)} \mu(dx)p(x,dy) = \log \int e^{z_\eta h(x)} p(e^{z_\eta k})(x)\mu(dx) \leq -\gamma + \eta.$$

On pose $g(y) = z_\eta k(y)$. C'est une fonction de $\mathcal{C}_b(E_2)$. L'inégalité de Jensen entraine

$$\int \log p(e^g)d\mu + z_\eta \int hd\mu \leq \log \int e^{z_\eta h} p(e^g)d\mu \leq -\gamma + \eta.$$

On obtient finalement, puisque $\int gd\nu = \int hd\mu = 0$,

$$\int gd\nu - \int \log p(e^g)d\mu \geq \gamma - \eta;$$

$$J_b(\mu,\nu) \geq \int gd\nu - \int \log p(e^g)d\mu \geq \gamma - \eta.$$

Cette inégalité étant vraie pour tout $\eta > 0$ et tout $\gamma < \overline{I}(\mu,\nu)$,

$$J_b(\mu,\nu) \geq \overline{I}(\mu,\nu).$$

On a donc $J_b(\mu,\nu) = \overline{I}(\mu,\nu) = J_{maj}(\mu,\nu)$. $\diamondsuit$

$\triangle$ **Preuve partielle du théorème 2.2.**

a) On admet le résultat de la proposition 9.2.5 de [DupEll97] qui établit la partie a). La preuve n'est pas difficile, mais un peu technique.

b) Il reste à montrer la semi-continuité inférieure de I et de I^1. Pour I, cela découle du a). Par le corollaire 2.1, en tout ζ tel que $\zeta^1(D(p)) = 0$, I est s.c.i. Si $\zeta^1(D(p)) > 0$ et si $(\zeta_n)_n$ converge étroitement vers ζ, on a $I(\zeta) = \lim_n I(\zeta_n) = \infty$. I est donc s.c.i.

Pour ce qui est de I^1, soit μ une probabilité sur E et $(\mu_n)_n$ une suite qui converge étroitement vers μ. Si $\mu(D(p)) = 0$, I^1 est s.c.i. en μ par le corollaire 2.1.
Dans le cas contraire, on va montrer que $\lim_n I(\mu_n) = \infty$. Supposons en effet cette propriété fausse ; alors on note $(\mu_{a(n)})_n$ une sous-suite telle que $\limsup_n I^1(\mu_{a(n)}) =$

$L < \infty$. Cela signifie que pour n assez grand, on a $\mu_{a(n)}(D(p)) = 0$ et il existe une suite $(\zeta_{a(n)})_n \in \mathcal{C}(\mu_{a(n)}, \mu_{a(n)})$ telle que

$$I^1(\mu_{a(n)}) = K\left(\zeta_{(a(n)} \mid \mu_{a(n)} \oslash p\right).$$

La suite $(\zeta_{a(n)})_n$ est tendue et, il existe une suite extraite $(\zeta_{b(n)})_n$ qui converge vers $\zeta \in \mathcal{C}(\mu, \mu)$. $I(\zeta) \leq L < \infty$ donc, en appliquant la partie a), on obtient une contradiction. $\diamondsuit$

○ **Preuve de la proposition 2.5.**

1) On prouve la première partie de la proposition : en 1a), $I(\zeta) \geq J_{\mathcal{B}_{maj}(E^2)}$ et en 1b) $I(\zeta) \leq J_{\mathcal{B}_b(E^2)}$.

1a) Soit $G \in \mathcal{B}_{maj}(E^2)$ et a un réel tel que $e^a \geq 1/r(T_G)$. En posant $T_G^0 = 1$, la fonction u définie par $e^u(x) = \sum_{n=0}^{\infty} e^{an} T_G^n(1)(x)$ est positive et bornée. Donc :

$$\int e^{G(x,y)+u(y)} p(x, dy) = T_G(e^u)(x) = e^{-a}(e^u(x) - 1),$$

$$G(x,y) + u(y) - \log \int e^{G(x,y)+u(y)} p(x, dy) = G(x,y) + a - \log(e^{u(x)} - 1) + \log(e^{u(y)}).$$

Si $F(x,y) = G(x,y) + u(y)$ et $\zeta^1 = \zeta^2 = \nu$, quel que soit $a < -\Lambda(G)$,

$$\int \left(F - \log p(e^F)\right) d\zeta \geq \int G d\zeta + a + \int (\log e^u - \log(e^u - 1))\, d\nu \geq \int G d\zeta + a,$$

$$I(\zeta) \geq \int \left(F - \log p(e^F)\right) d\zeta \geq \int G d\zeta - \Lambda(G).$$

Ceci étant vrai pour toute fonction $G \in \mathcal{B}_{maj}(E^2)$, on a, quelle que soit $\zeta \in \mathcal{P}(E^2)$,

$$I(\zeta) \geq \sup_{G \in \mathcal{B}_{maj}(E^2)} \left\{ \int G d\zeta - \Lambda(G) \right\}.$$

Pour $g \in \mathcal{B}_{maj}(E)$, on pose $f = g + u$ et l'on conclut plus facilement car on a

$$f - \log p(e^f) = f + a + \log \frac{e^u - 1}{e^u} \geq f + a.$$

1b) Soit $G \in \mathcal{B}_b(E^2)$ et $F(x,y) = G(x,y) - \log p(e^G)(y)$. Pour $u = p(e^G)$, on a

$$T_F(u)(x) = \int \frac{e^{G(x,y)}}{p(e^G)(y)} p(e^G)(y) p(x, dy) = u(x).$$

Donc u est un vecteur propre de T_F associé à la valeur propre 1 ; $\sum T_F^n(u)(x) = \infty$ et $1/r(T_F) \leq 1$, c'est à dire que $r(T_F) \geq 1$.

Or, e^G est minorée par une constante $c > 0$ et $u = T_F^n(u) \geq c T_F^n(1)$,

$$\|u\|_\infty \geq c \|T_F^n(1)\|_\infty = c\|T_F^n\|,$$

$$1 = \lim_n \left(\frac{\|u\|_\infty}{c}\right)^{1/n} \geq r(T_F).$$

Finalement, $r(T_F) = 1$, $\Lambda(F) = 0$ et, si $\zeta^1 = \zeta^2$,

$$\begin{aligned}\int \left(G(x,y) - \log p(e^G)(y)\right) \zeta(dx,dy) &= \int \left(G - \log p(e^G)\right) d\zeta \\ &= \int F d\zeta - \Lambda(F) \\ &\leq \sup_{H\in\mathcal{B}_b(E^2)} \left\{\int H d\zeta - \Lambda(H)\right\} = J_{\mathcal{B}_b(E^2)}(\zeta).\end{aligned}$$

De même, si $g \in \mathcal{B}_b(E)$, on obtient le résultat recherché en posant $G(x,y) = g(y)$.

2) Pour prouver la partie 2), il suffit de montrer que $I(\zeta) \leq J_{\mathcal{C}_b(E^2)}(\zeta)$ pour le cas fellérien (resp. $I(\zeta) \leq J_{\mathcal{C}_b^*(E^2)}(\zeta)$ pour le cas presque-fellérien).
On reprend la partie 1b) de la preuve pour $G \in \mathcal{C}_b(E^2)$ et p fellérienne. On a alors $F \in \mathcal{C}_b(E^2)$ et

$$I(\zeta) = \int \left(G(x,y) - \log p(e^G)(y)\right) \zeta(dx,dy) \leq \sup_{H\in\mathcal{B}_b(E^2)} \left\{\int H d\zeta - \Lambda(H)\right\} = J_{\mathcal{C}_b(E^2)}(\zeta).$$

De même, si p est presque fellérienne et si $G \in \mathcal{C}_b^*(E^2)$, on a $F \in \mathcal{C}_b^*(E^2)$ et l'on conclut de manière identique. $\diamondsuit$

2.3 PGD markoviens supérieurs pour les lois empiriques

Dans cette section, (E,d) est un espace polonais et $p : E \hookrightarrow E$ une probabilité de transition. E est l'espace des états de la chaîne de Markov étudiée. Sauf avis contraire, on utilise la version canonique de la chaîne de Markov de transition p ; $(\Omega, \mathcal{A}, (\mathbb{P}_x)_{x\in E}, (X_n))$, où $(X_n)_n$ est, pour $\mathbb{P}_x$, une chaîne de Markov de transition p et d'état initial x ; sous $\mathbb{P}_\nu = \int \nu(dx)\mathbb{P}_x$, (X_n) est une chaîne de loi initiale ν.
La suite des **lois empiriques bi-dimensionnelles** associée à la chaîne $(X_n)_n$ est définie par $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_{i-1},X_i}$; L_n^1 et L_n^2 sont les premières et secondes marginales de L_n. On note I et I^1 les taux markoviens vagues étudiés en 2.2.3. et 2.2.4.

2.3.1 PGD vague supérieur uniforme

Pour obtenir un PGD vague supérieur, seule la régularité de la probabilité de transition est nécessaire. Le théorème suivant donne un PGD vague supérieur uniforme pour toutes les lois initiales.

Théorème 2.3 (Donsker-Varadhan vague uniforme) *Soit $(X_n)_n$ une chaîne de Markov sur E, de probabilité de transition p fellérienne.*

1) *La suite $(L_n)_n$ satisfait un principe de grandes déviations vague supérieur uniforme, de vitesse n et de taux I : pour toute partie compacte $\Gamma \in \mathcal{P}(E^2)$,*

$$\limsup_n \frac{1}{n} \sup_{\alpha\in\mathcal{P}(E)} \log \mathbb{P}_\alpha (L_n \in \Gamma) \leq -I(\Gamma).$$

2) *Les suites $(L_n^1)_n$ et $(L_n^2)_n$ satisfont un principe de grandes déviations vague supérieur uniforme, de vitesse n et de taux I^1.*

2.3.2 PGD étroit supérieur

♣ Fonction de Lyapounov et tension sur $\mathcal{P}(E)$

On rappelle qu'une fonction de Lyapounov V est une fonction borélienne positive dont les épigraphes sont relativement compacts.

Pour $M > 0$, on note $\Gamma_M \stackrel{\triangle}{=} \{\nu \in \mathcal{P}(E) \mid \int V d\nu \leq M\}$. C'est une partie tendue de $\mathcal{P}(E)$. En effet, pour $A > 0$ arbitraire, $K_A \stackrel{\triangle}{=} \overline{\{x \in E \mid V(x) \leq A\}}$ est une partie compacte de E. Pour tout $\nu \in \Gamma_M$ on a

$$\nu(K_A^c) \leq \frac{\int V d\nu}{A} \leq \frac{M}{A}.$$

On en déduit une conséquence du théorème de Donsker-Varadhan vague.

Corollaire 2.2 *Dans le cadre du théorème 2.3, on suppose qu'il existe une fonction de Lyapounov V et une partie $\mathcal{H}$ de $\mathcal{P}(E)$ telles que*

$$\limsup_n \frac{1}{n} \sup_{\nu \in \mathcal{H}} \log \mathbb{E}_\nu \left(e^{n \int V dL_n^1} \right) < \infty. \ \textit{et} \ \limsup_n \frac{1}{n} \sup_{\nu \in \mathcal{H}} \mathbb{E}_\nu \int e^{nV dL_n^2} < \infty.$$

Alors, les PGD vagues du théorème 2.3 deviennent étroits et uniformes pour $\nu \in \mathcal{H}$.

♠ Donsker-Varadhan étroit

Divers principes de grandes déviations étroits supérieurs pour la suite des lois empiriques apparaissent dans la littérature. Le théorème donné ici est approximativement celui de Donsker-Varadhan dans [DonVar76] (voir la section 2.3.3. pour plus de précisions). Cette version est uniforme et porte sur une chaîne presque-fellérienne.

Définition 2.4 (Condition de rappel exponentiel associée à (U, V)) *p vérifie la condition de rappel associée à (U, V) s'il existe une fonction mesurable $U : E \to \mathbb{R}_+$, une fonction V et une constante C, positive ou nulle, telles que $V + C$ soit une fonction de Lyapounov et que l'on ait :*

$$V \leq U - \log p(e^U).$$

Théorème 2.4 (PGD étroit supérieur) *On suppose que la probabilité de transition p de la chaîne est presque fellérienne et vérifie la condition de rappel exponentiel associée à (U, V). Alors*

(i) I et I^1 sont des taux étroits.

(ii) La suite $(L_n)_n$ satisfait un principe de grandes déviations uniforme étroit supérieur de vitesse n et de taux étroit I : pour toute partie fermée $F \in \mathcal{P}(E^2)$ et toute partie compacte $\mathcal{H}$ de $\mathcal{H}_M = \{\nu \in \mathcal{P}(E) \mid \int U d\nu \leq M\}$ $(M > 0)$,

$$\limsup_n \sup_{\nu \in \mathcal{H}} \frac{1}{n} \log \mathbb{P}_\nu (L_n \in F) \leq -I(F).$$

Les suites $(L_n^1)_n$ et $(L_n^2)_n$ satisfont à un principe de grandes déviations uniforme sur $\mathcal{H}_M$ étroit supérieur, de vitesse n et de taux étroit I^1.

2.3.3 Références

Le PGD vague énoncé dans le théorème 2.3 (partie 2)) a été prouvé par Donsker-Varadhan dans [Donvar75a] pour un espace E compact, puis dans [DonVar76] pour l'extension aux espaces polonais. La partie 1) est une version bi-dimensionnelle qui se démontre approximativement de la même manière. Pour passer au PGD étroit, Donsker et Varadhan se servent d'une condition de rappel proche de la définition 2.4. donnée ici et empruntée à Dupuis-Ellis [DupEll97]. L'extension presque fellérienne provient, elle aussi de Dupuis-Ellis (voir le chapitre 9 de [DupEll97]).
Des PGD markoviens pour les lois bidimensionnelles $(L_n)_n$ sont prouvés par Ellis [Ell88] et Lipster [Lip96]. La démonstration du PGD étroit donnée ici est inspirée de celle de Dupuis et Ellis donnée dans les chapitres 8 et 9 de [DupEll97]. Elle est adaptée à notre contexte, un peu différent. Outre le fait que nous travaillons ici sur des lois empiriques de type $\frac{1}{n}\sum_{i=1}^n \delta_{X_{i-1},X_i}$, nous prouvons l'uniformité pour $(\mathcal{P}_\nu)_{\nu\in\mathcal{H}}$, $\mathcal{H}$ compact de $H_M = \{\nu \in \mathcal{P}(E);\ \int U d\nu \leq M\}$, alors que De Acosta [Aco90] la prouvent pour $(\mathcal{P}_x)_{x\in K}$, K compact de E lorsque U est bornée sur les compacts.

2.3.4 Preuves

♣ Préliminaire

Soient G et g boréliennes positives, respectivement définies sur E^2 (resp. E). On a pour tout $\alpha \in \mathcal{P}(E)$ et tout n (par un raisonnement par récurrence sur les espérances conditionnelles),

$$\mathbb{E}_\alpha\left(e^{n\int(G-\log p(e^G))dL_n}\right) = \mathbb{E}_\alpha\left(e^{\sum_{i=1}^n G(X_{i-1},X_i)-\log p(e^G)(X_{i-1})}\right) = 1.$$

Par le même argument,

$$\mathbb{E}_\alpha\left(e^{\sum_{i=1}^n g(X_{i-1},X_i)-\log p(e^g)(X_{i-1})}\right) = 1,$$

$$\mathbb{E}_\alpha\left(e^{n\int(g-\log p(e^g))dL_n^1}\right) \leq \mathbb{E}_\alpha\left(e^{g(X_0)+\sum_{i=1}^{n-1} g(X_i)-\log p(e^g)(X_{i-1})}\right) = \int e^g d\alpha,$$

$$\mathbb{E}_\alpha\left(e^{n\int(g-\log p(e^g))dL_n^2}\right) \leq \mathbb{E}_\alpha\left(e^{\log p(e^g)(X_0)+\sum_{i=1}^{n} g(X_i)-\log p(e^g)(X_{i-1})}\right) = \int p(e^g)d\alpha.$$

Pour Γ borélien de $\mathcal{P}(E^2)$ ou de $\mathcal{P}(E)$,

$$\frac{1}{n}\log \mathbb{P}_\alpha(L_n \in \Gamma) \leq -\inf_{\zeta\in\Gamma}\left\{\int (G-\log p(e^G))d\zeta\right\},$$

$$\frac{1}{n}\log \mathbb{P}_\alpha(L_n^1 \in \Gamma) \leq \frac{1}{n}\log\int e^g d\alpha - \inf_{\zeta\in\Gamma}\left\{\int (g-\log p(e^g))d\zeta\right\},$$

$$\frac{1}{n}\log \mathbb{P}_\alpha(L_n^2 \in \Gamma) \leq \frac{1}{n}\log\int p(e^g) d\alpha - \inf_{\zeta\in\Gamma}\left\{\int (g-\log p(e^g))d\zeta\right\}.$$

♠ **Preuve du PGD vague lorsque p est fellérienne**

Soit Γ un compact de $\mathcal{P}(E^2)$ (resp. $\mathcal{P}(E)$) et $\lambda < I(\Gamma)$ (resp. $\lambda < I^1(\Gamma)$). Pour $\zeta \in \Gamma$, il existe un élément $G \in \mathcal{C}_b(E^2)$ (resp. $G \in \mathcal{C}_b(E)$) tel que $\zeta \in \left\{\xi \mid \int (G - \log p(e^G))d\xi > \lambda\right\}$. Le compact Γ est réunion d'une sous-famille finie de ces ouverts :

$$\Gamma \subset \cup_{j=1..N} \left\{\xi \mid \int (G_j - \log p(e^{G_j}))d\xi > \lambda\right\} \stackrel{\triangle}{=} \cup_{i=j..N} \Gamma_j.$$

On a

$$\frac{1}{n} \log \mathbb{P}_\alpha(L_n \in \Gamma_j) \leq -\lambda,$$

$$\text{resp. } \frac{1}{n} \log \mathbb{P}_\alpha(L_n^i \in \Gamma_j) \leq \frac{M}{n} - \lambda \text{ pour } i = 1, 2,$$

M désignant $\sup_j \left\{\int e^{f_j} d\alpha \vee \int p(e^{f_j})d\alpha\right\}$. On en déduit le PGD vague annoncé. ♢

Preuve du corollaire 2.2 Il existe $M > 0$ et $N \in \mathbb{N}$ tels que, pour tout $n > N$,

$$\sup_{\nu \in \mathcal{H}} \mathbb{E}_\nu \left(e^{n \int V dL_n^1}\right) \leq e^{nM},$$

$$\sup_{\nu \in \mathcal{H}} \mathbb{E}_\nu \left(e^{n \int V dL_n^2}\right) \leq e^{nM}.$$

Donc, pour $i = 1, 2$ et $V \oplus V(x, y) = V(x) + V(y)$

$$\sup_{\nu \in \mathcal{H}} \mathbb{P}_\nu \left(\int V dL_n^i \geq R\right) \leq e^{-n(R-M)},$$

$$\sup_{\nu \in \mathcal{H}} \mathbb{P}_\nu \left(\int V \oplus V dL_n \geq 2R\right) \leq 2e^{-n(R-M)}.$$

On en déduit la tension exponentielle des suites $(L_n)_n$ et $(L_n^i)_n$, ce qui donne le résultat. ♢

Sous la condition de rappel exponentiel, on obtient sans peine, lorsque p est fellérienne, un PGD étroit. En effet, d'après le préliminaire, le corollaire 2.2 s'applique alors à $\mathcal{H} = \left\{\nu \in \mathcal{P}(E) \mid \int e^U + p(e^U)d\nu \leq M\right\}$, ce qui conduit immédiatement à un PGD étroit uniforme sur $\mathcal{H}$. La preuve qui suit améliore un peu l'uniformité et est valable dans le cas presque-fellérien. En contrepartie, elle suit la méthode plus délicate de Dupuis-Ellis.

♢ **Preuve du PGD étroit dans le cas presque fellérien**

(i) **Les taux sont étroits**

Il suffit de montrer que les épigraphes de I et I^1 sont relativement compacts.

Pour I^1, on a (par le théorème de Beppo-Lévi avec les fonctions $U \wedge k$ pour $k \in \mathbb{N}$),

$$I^1(\nu) \geq \int (U - \log p(e^U))d\nu \geq \int V d\nu.$$

Le fermé $\{\nu \mid I^1(\nu) \leq M\}$ est contenu dans $\{\nu \mid \int V d\nu \leq M\}$ qui est relativement compact. I^1 est un taux étroit.

La fonction $V \oplus V(x,y) = V(x) + V(y)$ est aussi à épigraphes compacts sur E^2

$$\left\{\zeta \in \mathcal{P}(E^2);\ I(\zeta) \leq M\right\} \subset \left\{\zeta \in \mathcal{P}(E^2);\ \int (V \oplus V) d\zeta \leq 2M\right\} \text{ qui est relativement compact ;}$$

I est donc un taux étroit. $\diamondsuit$

(ii) Démonstration du PGD étroit

Soit $\mathcal{H}$ une partie compacte de $\mathcal{H}_M$. Selon le théorème 1.3 et la remarque 1.1, prouver le PGD supérieur uniforme pour $\nu \in \mathcal{H}$ de taux I pour la suite $(L_n)_n$ revient à prouver que pour une suite de probabilités $(\nu_n)_n$ telle que $(\int U d\nu_n)$ soit bornée et pour $\Phi : \mathcal{P}(E^2) \to \mathbb{R}$ lipschitzienne et bornée, on a

$$\limsup_n \frac{1}{n} \log \mathbb{E}_{\nu_n}\left(e^{-n\Phi(L_n)}\right) \leq -\inf_{\zeta \in \mathcal{P}(E^2)} \{\Phi(\zeta) + I(\zeta)\}$$

On va donc prouver un principe de Laplace supérieur pour la suite $\Lambda_n(\Phi) = \frac{1}{n} \log \mathbb{E}_{\nu_n}\left(e^{-n\Phi(L_n)}\right)$.

• **Représentation variationnelle à l'instant n pour le calcul de $\Lambda_n(\Phi)$**

Il s'agit de majorer la quantité $\Lambda_n(\Phi)$. Pour la loi initiale ν_n et $j \leq n$, on pose

$$L_{j,n} \triangleq \frac{1}{n} \sum_{i=1}^{j} \delta_{X_{i-1},X_i} \quad \text{et } L_{n,n} = L_n.$$

La suite $(X_j, L_{j,n})_j$ est une chaîne de Markov de loi initiale $\nu_n \otimes \delta_0$ et de transition

$$q_{j,n}\left((y,\rho); \bullet\right) = \text{ loi de } \left(X, \rho + \frac{1}{n}\delta_y \otimes \delta_X\right) \quad \text{avec } X \text{ de loi } p(y, \bullet).$$

On désigne par $\mathcal{F}_{j,n} = \sigma(X_i, i \leq j)$. Nous définissons $(H_{j,n})_j$ par :

$$\begin{aligned} H_{j,n}(X_j, L_{j,n}) &\triangleq -\frac{1}{n} \log \mathbb{E}_{\nu_n}\left(e^{-n\Phi(L_n)} \mid \mathcal{F}_{j,n}\right), \\ H_{0,n}(X_0, 0) &= -\Lambda_n(\Phi), \\ H_n(X_n, L_n) &= \Phi(L_n). \end{aligned}$$

Et, par le principe variationnel (proposition 2.1),

$$\begin{aligned} H_{j,n}(y,\rho) &= -\frac{1}{n} \log \int e^{-nH_{j+1,n}(z,\rho+\frac{1}{n}\delta_u \otimes \delta_z)} \\ &= \inf_{\zeta \in \mathcal{P}(E^2)} \left\{\frac{1}{n} K\left(\zeta \mid \delta_y \oslash p\right) + \int H_{j+1,n}\left(z, \rho + \frac{1}{n}\delta_u \times \delta_z\right) \zeta(du, dz)\right\}. \end{aligned}$$

Si on note $\mathcal{M}_C(E^2)$ l'espace des mesures ζ sur E^2 telles que $\zeta(E^2) \leq C$, cela correspond à un système d'équations de la programmation dynamique d'une chaîne contrôlée à valeurs dans $E \times \mathcal{M}_{j/n}(E^2)$ à l'instant j, avec la loi initiale $\nu_n \oslash \delta_0$ et l'espace des contrôles $\mathcal{P}(E^2)$, la transition à l'instant j étant

$$Q_{j,n}(y, \rho, \zeta; \bullet) = \text{ loi de } (Z^2, \rho + \frac{1}{n}\delta_{Z^1,Z^2}) \text{ où } Z = (Z^1, Z^2) \text{ a la loi } \zeta.$$

Le coût final est $(y, \rho) \to \Phi(\rho)$ et le coût courant est $c_j(y, \rho, \zeta) = \frac{1}{n}K\left(\zeta \mid \delta_y \oslash p\right)$. La stratégie optimale est une probabilité de transition $s^{opt}_{j,n}(y, \rho; \cdot)$ de $E \times \mathcal{M}_{j/n}(E^2)$ dans $\mathcal{P}(E^2)$ donnée par la loi de Gibbs :

$$s^{opt}_{j,n}(y, \rho; du, dz) \overset{\triangle}{=} \frac{e^{-nH_{j+1,n}\left(z,\rho+\frac{1}{n}\delta_u\otimes\delta_z\right)}\delta_y \oslash p(du, dz)}{\int_{E^2} e^{-nH_{j+1,n}\left(z,\rho+\frac{1}{n}\delta_u\otimes\delta_z\right)}\delta_y \oslash p(du, dz)}.$$

Elle génère une suite markovienne non homogène $(X^{opt}_j, L^{opt}_{j,n})_{0\leq j\leq n}$ de loi initiale $\nu_n \otimes \delta_0$ adaptée à la filtration $\left(\mathcal{F}^{opt}_{j,n}\right)$ avec

$$\mathbb{P}\left(X^{opt}_{j+1,n} \in dz / X^{opt}_{j,n}, L^{opt}_{j,n}\right) = s^{opt}_{j,n}(X^{opt}_{j,n}, L^{opt}_{j,n}; E \times dz) \overset{\triangle}{=} \nu^{opt}_{j,n}(X^{opt}_{j,n}, L^{opt}_{j,n}; dz).$$

Les lois marginales de $s^{opt}_{j,n}$ sont

$$s^{opt,1}_{j,n}(X^{opt}_{j,n}, L^{opt}_{j,n}; \cdot) = \delta_{X^{opt}_{j,n}}, \quad s^{opt,2}_{j,n}(X^{opt}_{j,n}, L^{opt}_{j,n}; \cdot) = \nu^{opt}_{j,n}(X^{opt}_{j,n}, L^{opt}_{j,n}; \cdot).$$

Ainsi, le coût de cette stratégie optimale à l'étape n est

$$-\Lambda_n(\Phi) = \mathbb{E}\left(\frac{1}{n}\sum_{j=0}^{n-1} K\left(s^{opt}_{j,n}(X^{opt}_{j,n}, L^{opt}_{j,n}; \cdot)/\delta_{X^{opt}_{j,n}} \oslash p(\cdot)\right) + \Phi(L^{opt}_n)\right).$$

- **Etude asymptotique, preuve du principe de Laplace**

Par la convexité de l'information de Kullback,

$$-\Lambda_n(\Phi) \geq \mathbb{E}\left(K\left(\frac{1}{n}\sum_{j=0}^{n-1} s^{opt}_{j,n}(X^{opt}_{j,n}, L^{opt}_{j,n}; \cdot)/\frac{1}{n}\sum_{j=0}^{n-1}\delta_{X^{opt}_{j,n}} \oslash p(\cdot)\right) + \Phi(L^{opt}_n)\right).$$

On pose $A_{j,n} = \frac{1}{n}\sum_{i=0}^{j-1} s^{opt}_{i,n}(X^{opt}_{j,n}, L^{opt}_{j,n}; \cdot)$ et $A_n = A_{n,n}$

$$A^1_n = L^{opt,1}_n \text{ et } A^2_n = \frac{1}{n}\sum_{j=0}^{n-1}\nu^{opt}_{j,n}(X^{opt}_{j,n}, L^{opt}_{j,n}; \bullet)$$

On prouvera à la fin de la démonstration les majorations suivantes, en notant $-C$ un minorant de la fonction V :

$$\mathbb{E}\left(\int V dL^{opt,2}_n\right) \leq 2\|\Phi\|_\infty + \frac{M + C}{n},$$

$$\mathbb{E}\left(\int V dL^{opt,1}_n\right) \leq 2\|\Phi\|_\infty + \frac{M}{n},$$

et

$$\mathbb{E}\left(\int V \oplus V dL_n^{opt}\right) \leq 4\|\Phi\|_\infty + \frac{2M+C}{n} = K.$$

Ainsi :

$$\mathbb{E}\left(\int V dA_n^2\right) = \mathbb{E}\left(\int V dL_n^{opt,2}\right) \leq K.$$

Par conséquent, la suite de variables aléatoires (A_n, L_n^{opt}) à valeurs dans $\mathcal{P}(E) \times \mathcal{P}(E)$ est tendue (puisque V est une fonction de Lyapounov). De toute suite extraite, on peut extraire une sous-suite $(A_{u(n)}, L_{u(n)}^{opt})$ qui converge en loi vers (A, L) et, par la représentation de Skorokhod, on peut supposer que cette convergence est presque sûre. Par conséquent, presque sûrement, $A^1 = A^2 = L^1 = L^2$.
Si $L^1(D(p)) > 0$ sur un ensemble de probabilité strictement positive, alors on a aussi

$$\lim_n K\left(A_{u(n)} \mid L_{u(n)}^{opt,1} \oslash p\right) = \infty$$

sur un ensemble de probabilité strictement positive et

$$\lim_n \mathbb{E}\left(K\left(A_{u(n)} \mid L_{u(n)}^{opt,1} \oslash p\right)\right) = \infty.$$

Ceci est impossible puisque l'on a

$$\sup_n \mathbb{E}\left(K\left(A_{u(n)} \mid L_{u(n)}^{opt,1} \oslash p\right)\right) \leq 2\|\Phi\|_\infty < \infty.$$

Finalement, on a presque sûrement $L^1(D(p)) = 0$ et $L_{u(n)}^{opt,1} \oslash p$ converge étroitement vers $L \oslash p$, presque sûrement. Par la semi-continuité inférieure de K et la continuité de Φ,

$$\liminf_n \left(K\left(A_{u(n)}/L_{u(n)}^{opt,1} \oslash p\right) + \Phi(L_{u(n)}^{opt})\right) \geq K\left(L/L^1 \oslash p\right) + \Phi(L).$$

Par le lemme de Fatou,

$$\begin{aligned} \limsup_n \Lambda_n(\Phi) &\leq -\liminf_n \mathbb{E}\left(K\left(A_{u(n)}/L_{u(n)}^{opt,1}\right) + \Phi(L_{u(n)}^{opt})\right) \\ &\leq -\inf_{\zeta \in \mathcal{P}(E^2);\, \zeta^1 = \zeta^2} \left\{K\left(\zeta/\zeta^1 \oslash p\right) + \Phi(\zeta)\right\}. \\ &\leq -\inf_{\zeta \in \mathcal{P}(E^2)} \left\{I(\zeta) + \Phi(\zeta)\right\}. \end{aligned}$$

Ainsi le principe de Laplace supérieur est vérifié puisque ceci est valable pour toute suite $(\nu_n)_n$ de $\mathcal{H}_M$. Le PGD supérieur uniforme sur $\mathcal{H}_M$ est satisfait.

On termine la preuve en démontrant les majorations données plus haut. La fonction U étant bornée inférieurement, on a par la remarque 2.1 de 2.2.1.

$$\int U(y)\nu(dy) - \log \int e^U(y)p(x,dy) \leq K\left(\nu(\cdot) \mid p(x,\cdot)\right);$$

$$\begin{aligned} \mathbb{E}\left(U\left(X_{j+1,n}^{opt}\right) - U\left(X_{j,n}^{opt}\right)\right) &= \mathbb{E}\left(\mathbb{E}\left(U\left(X_{j+1,n}^{opt}\right) - U\left(X_{j,n}^{opt}\right) \mid \mathcal{F}_{j,n}^{opt}\right)\right) \\ &= \mathbb{E}\left(\int U(y)\nu_{j,n}^{opt}\left(X_{j,n}^{opt}, L_{j,n}^{opt}; dy\right) - U(X_{j,n}^{opt})\right) \end{aligned}$$

$$\begin{aligned} &= \mathbb{E}\left(\int U(y)\nu^{opt}_{j,n}\left(X^{opt}_{j,n}, L^{opt}_{j,n}; dy\right) - \log\int e^U(y)p(X^{opt}_{j,n}, dy)\right) \\ &\quad -\mathbb{E}\left(V(X^{opt}_{j,n})\right) \\ &\leq \mathbb{E}\left(K\left(\nu^{opt}_{j,n}\left(X^{opt}_{j,n}, L^{opt}_{j,n}; \cdot\right) \mid p(X^{opt}_{j,n}, \cdot)\right)\right) - \mathbb{E}\left(V\left(X^{opt}_{j,n}\right)\right). \end{aligned}$$

En sommant de $j = 0$ à $j = n-1$, on obtient

$$-\mathbb{E}\left(U\left(X^{opt}_{0,n}\right)\right) \leq \mathbb{E}\left(\sum_{j=0}^{n-1} K\left(\nu^{opt}_{j,n}\left(X^{opt}_{j,n}, L^{opt}_{j,n}; \cdot\right) \mid p(X^{opt}_{j,n}, \cdot)\right)\right) - \mathbb{E}\left(\sum_{j=0}^{n-1} V\left(X^{opt}_{j,n}\right)\right),$$

et donc,

$$\mathbb{E}\left(\int V dL^{opt,1}_n\right) \leq 2\|\Phi\|_\infty + \frac{1}{n}\int U d\nu_n \leq 2\|\Phi\|_\infty + \frac{M}{n}.$$

$$\mathbb{E}\left(\int V dL^{opt,2}_n\right) \leq \Delta + \frac{1}{n}\left(\int U d\nu_n - \int V d\nu_n\right) \leq 2\|\Phi\|_\infty + \frac{M+C}{n}. \diamondsuit$$

2.4 Commentaires, exemples et contre-exemples

2.4.1 Taux spectral et oubli de la loi initiale

• **PGD supérieur vague spectral** Dans le cas fellérien, voici une autre preuve du PGD vague supérieur. Pour toute fonction $F \in \mathcal{C}_b(E^2)$, on a prouvé en 2.2.5. la relation

$$\limsup_n \frac{1}{n}\log\mathbb{E}_\nu\left(e^{F(X_0,X_1)+\ldots+F(X_{n-1},X_n)}\right) \leq \limsup_n \frac{1}{n}\log\int e^{F(x,y)}T_F^{(n-1)}(y,E)\nu(dx)p(x,dy) \leq \Lambda(F),$$

avec $\Lambda(F) = \log r(T_F)$, $r(T_F) = \lim_n \|T_F^n\|^{1/n}$, rayon spectral de l'opérateur T_F. Par la proposition 1.6 de De Acosta, cela conduit à un PGD vague supérieur de taux conjugué de Λ, c'est à dire selon la proposition 2.5. de taux égal au taux de Donsker-Varadhan lorsqu'il est fini.
Cette preuve est celle de De Acosta ([Aco85a] et [Aco90]) qui obtient par exemple ainsi le corollaire 2.1. L'un des interêts de sa méthode est son extension à des espaces d'états $(E,\mathcal{E})$ généraux, munis de la *τ-topologie*, topologie sur $\mathcal{P}(E)$ rendant continues les applications $\mu \mapsto \int f d\mu$, pour toute fonction $f \in \mathcal{B}_b(E)$.

• **Optimalité du PGD vague supérieur ?**
Si l'on n'avait pas majoré brutalement l'expression $\limsup_n \frac{1}{n}\log\int e^{F(x,y)}T_F^{(n-1)}(y,E)\nu(dx)p(x,dy)$, mais considéré $R_\nu(F)$ inverse du rayon de convergence de la série

$$\sum\left(\int e^{F(x,y)}T_F^{n-1}(y,E)\nu(dx)p(x,dy)\right)z^n,$$

et $\lambda_\nu(F) = \log R_\nu(F)$, on aurait

$$\limsup_n \frac{1}{n}\log\mathbb{E}_\nu\left(e^{F(X_0,X_1)+\ldots+F(X_{n-1},X_n)}\right) = \lambda_\nu(F) \leq \Lambda(F),$$

et donc un PGD supérieur de taux convexe λ_ν^* défini par $\lambda_\nu^* = \sup_{F\in\mathcal{C}_b(E^2)} \{\int F d\zeta - \lambda_\nu(F)\} \geq I(\zeta)$. Ainsi, en général, on peut obtenir un PGD supérieur dépendant de la loi initiale. L'article de Bahadur et Zabell [BahZab79] éclaire bien le rôle de la loi initiale et est illustré par l'exemple 2.3 suivant.
D'autre part, les exemples suivants prouvent que l'on peut obtenir des PGD de taux parfois non convexes.
Finalement, dans ces cas, les PGD supérieurs de Donsker-Varadhan ne sont pas optimaux, mais ils ont le mérite d'être simples. On n'est assuré de l'optimalité que lorsqu'un PGD inférieur est obtenu avec le même taux. On reviendra sur ce point en 3.5.
L'exemple suivant provient de [Din93].

Exemple 2.3 (Une chaîne de Markov pour laquelle le bon PGD dépend de la loi initiale) *On considère la chaîne de Markov à valeurs dans l'espace discret $\{1, 2, 3, 4\}$ de transition p représentée par la matrice suivante :*

$$\begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Il y a deux classes récurrentes. Sur ces classes récurrentes, on a des PGD étroits de taux respectivement $J^1(1, .)$ et $J^1(2, .)$, les taux de Donsker-Varadhan de la chaîne restreinte à $\{1, 2\}$ et $\{3, 4\}$:

$$\begin{aligned} J^1(1,\nu) &= \nu(1)\log 2 \ \ \textit{si}\ \nu(3) = \nu(4) = 0, \\ &= \infty \ \ \textit{sinon.} \end{aligned}$$

$$\begin{aligned} J^1(2,\nu) &= \nu(3)\log 2 \ \ \textit{si}\ \nu(1) = \nu(2) = 0, \\ &= \infty \ \ \textit{sinon.} \end{aligned}$$

Par les théorèmes 2.1 et 2.2 de [DinZab92], pour une loi initiale qui charge chacune des deux classes récurrentes, on a un PGD de taux $J^1(.) = inf(J^1(1, .), J^1(2, .))$:

$$\begin{aligned} J^1(\nu) &= \nu(1)\log 2 \ \ \textit{si}\ \nu(3) = \nu(4) = 0, \\ &= \nu(3)\log 2 \ \ \textit{si}\ \nu(1) = \nu(2) = 0, \\ &= \infty \ \textit{sinon.} \end{aligned}$$

Soit ν une probabilité et I_d la matrice identité. Les seules probabilités de transition absolument continues par rapport à p sont $q_{t_1,t_2} \triangleq \begin{pmatrix} q_{t_1} & 0 \\ 0 & q_{t_2} \end{pmatrix}$ où $q_t = \begin{pmatrix} t & 1-t \\ 0 & 1 \end{pmatrix}$ pour des $t_1, t_2 \in [0, 1]^2$.
A partir de là, nous pouvons calculer le taux de Donsker-Varadhan I^1 :

* *Si* $\nu(1) > 0$ *et* $\nu(3) > 0$ *alors* $I = q_{1,1}$ *est la seule probabilité de transition pour laquelle* ν *est invariante, ainsi*

$$I^1(\nu) = I(\nu \otimes I_d) = \nu(1)\log 2 + \nu(3)\log 2.$$

* *si* $\nu(1) = 0$ *alors, pour tout* $t \in [0,1]$*,* ν *est une probabilité invariante pour* $q_{1,t}$ *et*

$$I^1(\nu) = \inf_{t\in[0,1]} I(\nu \oslash q_{t,1}) = \nu(3)\log 2.$$

* *Si* $\nu(3) = 0$ *alors, pour tout* $t \in [0,1]$*,* ν *est invariante pour* $q_{t,1}$ *et*

$$I^1(\nu) = \inf_{t\in[0,1]} I(\nu \oslash q_{1,t}) = \nu(1)\log 2.$$

Finalement, $I^1(\nu) = \nu(1)\log 2 + \nu(3)\log 2 \le J^1(\nu)$*. On a un PGD supérieur avec le taux de Donsker-Varadhan mais il n'est pas otpimal.*

2.4.2 Sur la condition de rappel exponentiel et la stabilité

On rappelle qu'un modèle markovien est stable sur (E,d) s'il existe une probabilité invariante unique μ telle que, presque-sûrement, $(L_n)_n$ converge étroitement vers μ. Contrairement à la récurrence, cette propriété suppose une structure topologique sur l'espace d'états. Elle n'implique pas la récurrence comme le montre l'exemple $X_{n+1} = \frac{1}{2}X_n + \varepsilon_{n+1}$ avec $\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_n = -1) = 1/2$ qui est stable de loi stationnaire Lebesgue sur $[-2,2]$ mais non récurrente (partant de l'état initial 0, $\L_n^1(\mathbb{Q}) = 1$).
Pour un modèle presque fellérien, la condition de rappel exponentiel implique l'existence d'une probabilité invariante μ (voir [DupEll97], proposition 9.2.6.). Lorsque cette probabilité est unique (par exemple, quand le modèle est irréductible), le taux I^1 ne s'annule qu'en μ. On en déduit que

$$\forall r > 0,\ \limsup \frac{1}{n}\log \mathbb{P}\left(d_{Dudley}(L_n,\mu) \ge r\right) < 0.$$

Ici, la convergence de L_n vers μ a une vitesse exponentielle. On parlera de **stabilité exponentielle**.

Exemple 2.4 (La condition de rappel exponentiel pour un modèle autorégressif)

Considérons le modèle suivant défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$*, à valeurs dans* $\mathbb{R}^d$ *:*

$$X_{n+1} = f(X_n) + \sigma(X_n)\varepsilon_{n+1}.$$

* $(\varepsilon_n)_n$ *est une suite de variables aléatoires i.i.d, indépendante de* X_0*.*
* $f : \mathbb{R}^d \to \mathbb{R}^d$ *est une application mesurable, bornée sur toute partie compacte de* $\mathbb{R}^d$*.*
* $\sigma : \mathbb{R}^d \to \left[\frac{1}{r}, r\right]$ $(r \ge 1)$ *est une application borélienne.*

La transition de cette chaîne de Markov est fellérienne si f et σ sont continues ; elle est presque fellérienne si f et σ sont continues Lebesgue presque partout et si ε_1 a une densité bornée par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Ce cadre permet d'étudier d'important modèles en économétrie, tels que les modèles autorégressifs à seuil pour lesquels la fonction f est de la forme :

$$f = \sum_{i=1}^{K} f_i \, \mathbb{I}_{\Gamma_i}$$

avec f_i continues, Γ_i disjoints, de frontières Lebesgue-négligeables et tels que $\cup\Gamma_i = \mathbb{R}^d$. Il est à noter que, pour ce type de modèle, on utilise la version définie ci-dessus différente de la version canonique adoptée en général (voir le début de la section 2.3) ; tous les résultats ne portant que sur les lois, le passage d'une version à l'autre ne pose pas de problème.
Observons les contraintes imposées par la condition de rappel exponentiel dans ce cas :
La condition de rappel exponentiel associée à (U,V) est vérifiée pour le modèle s'il existe une fonction borélienne positive U, telle que

$$\lim_{\|x\|\to\infty} \left(U(x) - \log \mathbb{E}\left(\exp\left(U(f(x)+\sigma(x)\varepsilon_1)\right)\right)\right) = +\infty.$$

Il suffit en effet de prendre $V = U - \log p(e^U)$. Alors, pour $\rho > 0$,

$$\log \mathbb{E}\left(\exp\left(\rho\|f(x)+\sigma(x)\varepsilon_1\|\right)\right) \leq \rho\|f(x)\| + \log \mathbb{E}\left(\exp\left(\rho r\|\varepsilon_1\|\right)\right).$$

Si $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ est une fonction croissante, pour tout $s \in]0,1[$, on a

$$\phi\left(\|f\| + r\|\varepsilon_1\|\right) \leq \max\left\{\phi(s^{-1}\|f\|), \phi((1-s)^{-1}r\|\varepsilon_1\|)\right\} \leq \phi(s^{-1}\|f\|)+\phi((1-s)^{-1}r\|\varepsilon_1\|).$$

Par conséquent,

$$\log \mathbb{E}\left(\exp\left(\phi(\|f(x)+\sigma(x)\varepsilon_1\|)\right)\right) \leq \phi(s^{-1}\|f(x)\|) + \log \mathbb{E}\left(\exp\left(\phi((1-s)^{-1}r\|\varepsilon_1\|)\right)\right).$$

On peut donc choisir $U(x) = \phi(\|x\|)$ sous diverses hypothèses, une condition plus faible sur le bruit menant à une contrainte plus forte sur la fonction f.

a) *Sous la condition de Cramér, $\mathbb{E}\left(\exp(\tau\|\varepsilon_1\|)\right) < \infty, (\tau > 0)$, on peut prendre $\phi(t) = \frac{\tau(1-s)t}{r}$ et la condition sur f est*

$$\lim_{\|x\|\to\infty} \left(\|x\| - s^{-1}\|f(x)\|\right) = +\infty.$$

b) *Sous la condition $\mathbb{E}\left(\exp(\tau\|\varepsilon_1\|^\beta)\right) < \infty, (\tau > 0)$, on peut prendre, pour tout $s \in]0,1[$, $\phi(t) = \frac{\tau(1-s)^\beta}{r^\beta} t^\beta$ et la condition sur f se traduit par*

$$\lim_{\|x\|\to\infty} \left(\|x\|^\beta - s^{-\beta}\|f(x)\|^\beta\right) = +\infty.$$

c) *Si on suppose uniquement que le bruit est intégrable* $\mathbb{E}(\|\varepsilon_1\|) < \infty$, *alors* $\phi(t) = \log(t)$ *et la condition sur* f *est*

$$\lim_{\|x\|\to\infty} \frac{\|x\|}{\|f(x)\|} = +\infty.$$

L'exemple suivant provient de [BaxJaiVar91].

Exemple 2.5 (Chaîne de Markov récurrente ne satisfaisant pas le PGD étroit) *On se place sur l'espace d'états* $E = \mathbb{N}$ *et l'on considère, pour tout* $u \in]0,1[$, *la chaîne de Markov* $(X_n)_n$, *dont la suite des lois empiriques bi-dimensionnelles est notée* $(L_n)_n$ *de probabilité de transition* p_u, *définie par*

$$p_u(x, x+1) = u = 1 - p_u(x, 0).$$

Cette chaîne est récurrente, de loi stationnaire μ_u *donnée par*

$$\mu_u(x) = u^x(1-u).$$

Si t *est une application de* $\mathbb{N}$ *dans* $[0,1]$ *et si* p_t *est définie par*

$$p_t(x, x+1) = t(x) = 1 - p_t(x, 0),$$

une loi stationnaire satisfait impérativement $\mu_t(x+1) = t(x)\mu_t(x)$.
Si on note $r(x) = t(0)t(1)...t(x-1)$, *on a* $\mu_t(x) = r(x)\mu_t(0)$.
Par conséquent, il existe une telle probabilité invariante si et seulement si

$$1/R \stackrel{\triangle}{=} \sum_x r(x) < \infty,$$

et, dans ce cas, on a $\mu_t(x) = Rr(x)$. *De plus, en posant* $r(0) = 1$,

$$I(\mu_t \oslash p_t) = R \sum_x r(x) \left(t(x) \log\left(\frac{t(x)}{u}\right) + (1 - t(x)) \log\left(\frac{1 - t(x)}{1-u}\right)\right).$$

Les mesures de la forme $\mu_t \oslash p_t$ *sont les seules de taux fini et, par conséquent, les* μ_t *sont les seules mesures telles que* $I^1(\mu_t) < \infty$.
S'il existe un x *tel que* $t(x) = 0$, *on note* m *le plus petit d'entre eux.* μ_t *est concentrée sur* $[0, m]$. *Si on prend par exemple la fonction* t_m *égale à* 1 *pour* $x < m$, μ_{t_m} *est la loi uniforme sur* $[0, m]$ *et*

$$I(\mu_{t_m} \oslash p_{t_m}) = \frac{m}{m+1} \log(1/u) + \frac{1}{m+1} \log\left(\frac{1}{1-u}\right) \le \log\left(\max\left(\frac{1}{u}, \frac{u}{1-u}\right)\right) = a.$$

L'épigraphe $\{\zeta \mid I(\zeta) \le a\}$ *contient la suite* $\zeta_m = \mu_{t_m} \oslash p_{t_m}$ *et on a*

$$\zeta_m(x, x+1) = \frac{1}{m+1} \quad \text{si } x \le m-1,$$

$$\zeta_m(m, 0) = \frac{1}{m+1}.$$

Cette suite converge vaguement vers 0*, n'a pas de valeur d'adhérence pour la convergence étroite. Le taux* I *n'est pas étroit. D'ailleurs, on n'a pas, pour tout fermé* F *de* $\mathcal{P}(\mathbb{N})$*, la borne supérieure*

$$\limsup_n \frac{1}{n} \log \mathbb{P}\left(L_n^1 \in F\right) \leq -I^1(F).$$

$$I^1(\nu) = \sup_{g \in \mathcal{C}_b(\mathbb{N})} \sum \log \left(\frac{e^g(x)}{(1-u)e^g(0) + ue^g(x+1)} \right) \nu(x).$$

Si on considère ν *la loi uniforme sur* $\{l+1, ..., l+m\}$ *et* $g_k(l+1) = \log(k)$*,* $g_k(x) = 0$ *sinon ;*

$$I^1(\nu) \geq \sup_k \frac{1}{m} \log k = \infty.$$

Ceci était prévisible car ν *n'est pas de la forme* μ_t *définie plus haut. L'ensemble* F *donné par*

$$F = \{\nu \mid \exists m \geq 1, \exists l \geq 1, \nu \text{ uniforme sur } \{l+1, ..., l+m\}\}$$

est fermé de taux infini. Or, pour tout $x \geq 1$*,*

$$\mathbb{P}_x\left(L_n^1 \in F\right) \geq \mathbb{P}_x\left(X_0 = x, X_1 = x+1, ...X_n - 1 = x+n-1\right) = u^{n-1}.$$

et

$$\liminf_n \frac{1}{n} \log \mathbb{P}_x\left(L_n^1 \in F\right) \geq \log u.$$

La borne supérieure n'est pas vérifiée. $\diamondsuit$

Chapitre 3

Transports de PGD : des lois empiriques vers les moyennes empiriques

3.1 Introduction

3.1.1 Position du problème

Soit $(L_n)_n$ une suite de probabilités aléatoires sur (E, d) polonais et f une fonction continue de E dans un espace de Banach B. Si J est un taux (vague ou étroit) sur $\mathcal{P}(E)$, son image par f est le **taux image** dont la définition est la suivante.

Définition 3.1 *Pour un espace polonais (E, d), on considère un taux J (vague ou étroit) défini sur $\mathcal{P}(E)$ et une fonction f continue de E dans un espace de Banach séparable B. Le **taux image** de J par f est défini par*

$$h_f(x) = \inf \left\{ J(\nu) \mid \ \nu \in \mathcal{P}(E), \ \int \|f\| \, d\nu < \infty \ et \int f d\nu = x \right\}.$$

Les questions posées sont les suivantes :

1) (L_n) satisfaisant un PGD supérieur (vague ou é troit) de taux J, peut on obtenir un PGD supérieur (vague ou é troit) pour $(\int f dL_n)$ dont le taux serait h_f ou, si h_f n'est pas s.c.i, sa régularisée s.c.i. La réponse é tant en général négative, peut on donner une réponse partielle à ce problème ?
2) (L_n) satisfaisant un PGD inférieur de taux J, peut on obtenir un PGD inférieur pour $(\int f dL_n)$ dont le taux serait h_f ou sa régularisée s.c.i. ? Peut on donner une réponse partielle à ce problème ?

Bien sûr, par le principe de contraction, si la suite $(L_n)_n$ satisfait un PGD étroit de taux étroit J, la réponse à ces deux questions est positive pour toute fonction $f \in \mathcal{C}_b(E)$.

3.1.2 Plan de ce chapitre

La section 3.2. donne des réponses positives à la question 1), pour des couples équilibrés de fonctions.

La section 3.3. donne des réponses aux deux questions, pour des fonctions convenablement approchées par des fonctions continues bornées ou par des couples "équilibrés". Dans la section 3.4, on applique ce qui précède au cas i.i.d et à des exemples markoviens. Enfin, la section 3.5 donne un résultat de grandes déviations inférieur pour des fonctionnelles additives, sous certaines conditions d'irréductibilité.

3.1.3 Rappels et notations

a) **Rappels (cf avant-propos et 2.3.2)**. Une fonction de Lyapounov V est une fonction borélienne positive, à épigraphes relativement compacts (donc compacts dès que V est s.c.i.).
On note $B(x,r)$ et $\overline{B}(x,r)$ les boules de centre x et de rayon r, respectivement ouverte et fermée. Soit I une fonction de E dans $[0,\infty]$. Sa régularisée s.c.i. est

$$\lim_{r\to 0}\uparrow I(B(x,r)) = \lim_{r\to 0}\uparrow I\left(\overline{B}(x,r)\right) \stackrel{\triangle}{=} I^{sci}(x).$$

En général, $I^{sci} \leq I$ avec égalité si I est s.c.i.

b) Soient (E,d) polonais, B un espace de Banach séparable et $\mu \in \mathcal{P}(E)$. Soit $f: E \to B$ une fonction continue. Si $\int \|f\| d\mu < \infty$, alors il existe $\int f d\mu$ unique tel que, pour tous $\theta \in B'$, on ait $\int <\theta, f> d\mu = <\theta, \int f d\mu>$ et $\|\int f d\mu\| \leq \int \|f\| d\mu$. Ce résultat est prouvé par exemple dans Rudin ([Rud73], théorèmes 3.27 et 3.29) pour K compact séparable. On peut l'étendre à B car, par la tension de μ, il existe une suite $(K_N)_N$ de compacts tels que $\lim_N \mu(K_N) = 1$. La suite $(\int_{K_N} f d\mu)_N$ est une suite de Cauchy dont la limite est l'intégrale $\int f d\mu$ cherchée.

c) **Troncature et uniforme intégrabilité**
Soit $a > 0$. Selon l'usage, si f est une fonction borélienne dé finie sur E à valeurs dans un espace de Banach B, on lui associe les fonctions bornées :

$$\begin{aligned}
\tau_a(f)(x) &= f(x)\min(1, a/\|f(x)\|) \\
&= f(x) \text{ si } \|f(x)\| \leq a \text{ et } = af(x)/\|f(x)\| \text{ si } \|f(x)\| \geq a, \\
\|\tau_a(f)\| &\leq \|f\| \text{ et, pour tout } x,\ \lim_{a\to\infty} \tau_a(f)(x) = f(x).
\end{aligned}$$

Pour $\mathcal{H} \subset \mathcal{P}(E)$, f est **uniformément intégrable** relativement à $\mathcal{H}$ si f est intégrable par rapport à chaque $\nu \in \mathcal{H}$ et si $\nu \mapsto \int f d\nu$ est continue sur $\mathcal{H}$. Voici deux conditions suffisantes d'uniforme intégrabilité relative à $\mathcal{H}$, valables lorsque $f: E \to B$ est continue et intégrable pour tout élément ν de $\mathcal{H}$.

$*$ **Condition suffisante de troncature par la norme**
La propriété suivante est une condition suffisante à l'uniforme intégrabilité de f relativement à $\mathcal{H}$.

$$\lim_{a\to\infty} \sup_{\nu\in\mathcal{H}} \int_{\|f\|>a} \|f\| d\nu = 0 \text{ ce qui équivaut à } \lim_{a\to\infty} \sup_{\nu\in\mathcal{H}} \int \|f - \tau_a(f)\| d\nu = 0.$$

Cette condition est aussi nécessaire si $B = \mathbb{R}^d$.

On rappelle qu'une fonction ρ borélienne de $\mathbb{R}_+$ dans $\mathbb{R}_+ \cup \{+\infty\}$ est surlinéaire si

$$\lim_{t\to\infty} \frac{\rho(t)}{t} = +\infty.$$

Une condition suffisante (qui implique la précédente) à l'uniforme intégrabilité de f relative à $\mathcal{H}$ est l'existence d'une fonction surlinéaire ρ telle que :

$$\lim_{a\to\infty} \sup_{\nu\in\mathcal{H}} \int_{\|f\|\geq a} \rho(\|f\|)\, d\nu < \infty.$$

$*$ **Condition suffisante de troncature par une fonction de Lyapounov convexe sur un espace de Banach**

Soit U une fonction de Lyapounov convexe de B dans $\mathbb{R}_+$, nulle en 0 et strictement positive ailleurs. On suppose que

$$\lim_{a\to\infty} \sup_{\nu\in\mathcal{H}} \int_{U\circ f\geq a} U\circ f d\nu = 0.$$

Alors, f est uniformément intégrable relativement à $\mathcal{H}$.

En effet, soit une suite ν_n de $\mathcal{H}$ convergeant vers $\nu \in \mathcal{H}$. Puisque f est bornée sur $\{x \mid U\circ f(x) \leq a\}$ (U est de Lyapounov), dès que $\nu(\{x \mid U\circ f(x) = a\}) = 0$, on a

$$\lim_n \int f \mathbb{I}_{U\circ f\leq a} d\nu_n = \int f \mathbb{I}_{U\circ f\leq a} d\nu.$$

Mais, comme U est convexe et $U(0) = 0$, on a

$$U\left(\int f\mathbb{I}_{U\circ f\geq a} d\nu\right) \leq \int U\left(f\mathbb{I}_{U\circ f\geq a}\right) d\nu = \int_{U\circ f\geq a} U\circ f d\nu,$$

et donc

$$\lim_{a\to\infty} \sup_{\nu\in\mathcal{H}} U\left(\int f\mathbb{I}_{U\circ f\geq a} d\nu\right) = 0.$$

Comme U ne s'annule qu'en 0, $\lim_{\delta\to 0} \sup\{\|x\| \mid U(x) \leq \delta\} = 0$ et

$$\lim_{a\to\infty} \sup_{\nu\in\mathcal{H}} \left\|\int f\mathbb{I}_{U\circ f\geq a} d\nu\right\| = 0,$$

ce qui permet de conclure ; $\nu \mapsto \int f d\nu$ est continue sur $\mathcal{H}$. $\diamondsuit$

3.2 Transport d'un PGD vague supérieur empirique pour des couples équilibrés

En 3.2 et 3.3, le cadre est le suivant.

Axiome 3.1

- $(L_n)_n$ *est une suite de probabilités aléatoires définies sur un espace polonais* (E, d) *satisfaisant un PGD vague supérieur de taux vague* J *et de vitesse* $v = (v_n)_n$ *;*
- V *est une fonction de Lyapounov s.c.i. définie sur* (E, d) *;*
- f *est une fonction continue de* (E, d) *dans un espace de Banach séparable* B*.*

3.2.1 Les couples équilibrés

Dans la situation considérée, c'est à dire en supposant un PGD vague supérieur pour la suite $(L_n)_n$, le transport n'est même plus évident pour $f \in C_b(E)$ car, pour Γ compact de E, $\{\nu \mid \int f d\nu \in \Gamma\}$ n'est en général pas compact dans $\mathcal{P}(E)$.
V étant une fonction de Lyapounov s.c.i, quel que soit $R > 0$, $\{\nu \mid \int V d\nu \leq R\}$ est compact. Supposons que la fonction f est uniformément intégrable relativement à l'ensemble $\{\nu \mid \int V d\nu \leq R\}$; alors, pour tout Γ fermé de E, $\{\nu \mid \int f d\nu \in \Gamma$ et $\int V d\nu \leq R\}$ est un compact de $\mathcal{P}(E)$. On est naturellement conduit à la définition et au théorème de transport qui suivent.

Définition 3.2 (Couple équilibré) (f, V) *est un couple équilibré si :*
a) $V : E \to \mathbb{R}_+$ *est une fonction de Lyapounov s.c.i,*
b) f *une fonction continue de* E *dans l'espace de Banach séparable* B, *uniformément intégrable sur chacun des épigraphes* $\{\nu \mid \int V d\nu \leq R\}$ $(R > 0)$*.*

Conditions suffisantes. Pour V fonction de Lyapounov s.c.i, les conditions suivantes souvent faciles à vérifier suffisent à assurer que (f, V) est équilibré :
$CS1)$ Il existe une fonction surlinéaire ρ telle que, pour toute probabilité ν,

$$\int \rho(\|f\|)\, d\nu \leq \int V d\nu.$$

$CS2)$ Il existe une fonction surlinéaire ρ telle que $\rho(\|f\|) \leq V$.

Théorème 3.1 *Soit* (f, V) *un couple équilibré. On rappelle que*

$$h_{f,V}(\{x\} \times [0, R]) = \inf\left\{J(\nu) \mid \nu \in \mathcal{P}(E),\ \int f d\nu = x, \int V d\nu \leq R\right\}.$$

On a les propriétés suivantes.
1) *Pour tout* $R > 0$, *la fonction* $x \mapsto h_{f,V}(\{x\} \times [0, R]) \stackrel{\triangle}{=} h_{f,V,R}(x)$ *est s.c.i.*
2) *Pour toute partie* Γ *fermée de* B,

$$\limsup_n \frac{1}{v_n} \log \mathbb{P}\left(\int f dL_n \in \Gamma,\ \int V dL_n \leq R\right) \leq -h_{f,V,R}(\Gamma).$$

3.2.2 Démonstration du théorème 3.1

2) est clair puisque $\{\nu \mid \int f d\nu \in \Gamma,\ \int V d\nu \leq R\}$ est compact.
Pour 1), il s'agit de prouver que $h_{f,V,R} = h^{sci}_{f,V,R}$ donc que :

$$h_{f,V,R}(x) \leq \lim_{r \to 0} \uparrow h_{f,V,R}\left(\overline{B}(x, r)\right) = \lim_{r \to 0} \uparrow J\left(\left\{\nu \mid \int f d\nu \in \overline{B}(x, r) \text{ et } \int V d\nu \leq R\right\}\right).$$

Mais, $\left\{\nu \mid \int f d\nu \in \overline{B}(x, r) \text{ et } \int V d\nu \leq R\right\}$ étant compact et J étant s.c.i, il existe au moins une probabilité ν_r telle que

$$J(\nu_r) = J\left(\left\{\nu \mid \int f d\nu \in \overline{B}(x, r) \text{ et } \int V d\nu \leq R\right\}\right) \text{ et } \int f d\nu_r = y_r \in \overline{B}(x, r).$$

Si ν est une valeur d'adhérence de $(\nu_r)_r$ lorsque $r \to 0$, on a $J(\nu) \leq \liminf J(\nu_r)$, $\int f d\nu = x$ et $\int V d\nu \leq R$. Donc

$$h_{f,V}(\{x\} \times [0,R]) \leq J(\nu) \leq \lim_{r \to 0} \uparrow h_{f,V}\left(\overline{B}(x,r) \times [0,R]\right). \diamondsuit$$

3.3 Approximation exponentielle et transport pour des couples exponentiellement équilibrés

3.3.1 Approximation exponentielle

Définition 3.3 ***(Approximation exponentielle contrôlée par une fonction de Lyapounov)***
Soient f et $(f_a)_{a>0}$ des fonctions boréliennes de E dans B.

1) 1a) $(f_a)_{a>0}$ *est une (v_n)-approximation exponentielle de f si, pour tout $r > 0$,*

$$\lim_{a \to \infty} \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(\left\|\int (f - f_a)\, dL_n\right\| \geq r\right) = -\infty.$$

1b) *C'est une **bonne** (v_n)-approximation exponentielle de f si l'on a de plus, pour tout $\gamma > 0$,*

$$\lim_{a \to \infty} \sup\left\{\| \int (f - f_a) d\nu \| \mid J(\nu) \leq \gamma\right\} = 0.$$

2) 2a) *Soit V une fonction de Lyapounov s.c.i. sur E. (f_a) est une (v_n)-approximation exponentielle de f, contrôlée par V si, pour tout $R > 0$ et tout $r > 0$, on a*

$$\lim_{a \to \infty} \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(\left\|\int (f - f_a)\, dL_n\right\| \geq r \text{ et } \int V dL_n \leq R\right) = -\infty.$$

2b) *C'est une **bonne** (v_n)-approximation exponentielle de f contrôlée par V si l'on a de plus, pour tout $\gamma > 0$ et pour tout $R > 0$,*

$$\lim_{a \to \infty} \sup\left\{\| \int (f - f_a) d\nu \| \mid J(\nu) \leq \gamma,\ \int V d\nu \leq R\right\} = 0.$$

Remarque 3.1 *Comme $\|\int (f - f_a)\, d\nu\| \leq \int \|f - f_a\| d\nu$, on obtient des conditions suffisantes souvent utiles en remplaçant dans les expressions précédentes $\|\int (f - f_a)\, dL_n\|$ par $\int \|f - f_a\| dL_n$.*

Exemple 3.1 (Troncatures) *Considérons les fonctions tronquées $\tau_a(f)$ définies en introduction. On remarque que*

$$\begin{aligned} f - \tau_a(f) &= \mathbb{I}_{\|f\|>a}\left(f - a\left(f/\|f\|\right)\right), \\ \|f - \tau_a(f)\| &\leq 2\|f\|\mathbb{I}_{\|f\|>a}. \end{aligned}$$

Donc :

$$\left\|\int (f - \tau_a(f))\, d\nu\right\| \leq \int \|f - \tau_a(f)\| d\nu \leq 2 \int \|f\|\mathbb{I}_{\|f\|>a} d\nu.$$

*Par conséquent, la condition d'**uniforme intégrabilité exponentielle de vitesse** $(v_n)_n$*

$$\forall r > 0, \lim_{a\to\infty} \limsup \frac{1}{v_n} \log \mathbb{P}\left(\int \|f\| \mathbb{I}_{\|f\|>a} dL_n \geq r\right) = -\infty,$$

*(resp. la condition d'**uniforme intégrabilité exponentielle de vitesse** $(v_n)_n$ **contrôlée par** V*

$$\forall R > 0, \forall r > 0, \quad \lim_{a\to\infty} \limsup_n \frac{1}{n} \log \mathbb{P}\left(\int \|f\| \mathbb{I}_{\|f\|>a} dL_n \geq r \ \ et \ \int V dL_n \leq R\right) = -\infty),$$

entraine la (v_n)-approximation exponentielle (resp. en ajoutant contrôlée par V) de f par ses tronquées $(\tau_a(f))$.

Exemple 3.2 (Suite i.i.d.) *Soit (Y_n) une suite i.i.d. de loi μ à valeurs dans E. Selon le théorème de Sanov (voir 1.3.4), $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ satisfait toujours un PGD é troit de taux $K(\bullet|\mu)$.*

a) Notons $\tau_a(x) = x\min(1, a/\|x\|)$; la suite $(\tau_a(Y_n))$ est i.i.d. et bornée. Supposons la condition de Cramér renforcée vérifiée par μ : pour tout $t > 0$, on a $\int \exp(t\|x\|)\mu(dx) < \infty$. Alors, pour tout $a > 0$, si $a \uparrow \infty$, $\|Y_1 - \tau_a(Y_1)\| \downarrow 0$ donc on a, pour tout $t > 0$,

$$\lim_{a\to\infty} \mathbb{E}(\exp(t\|Y_1 - \tau_a(Y_1)\|)) = 1;$$

par conséquent, pour tout $r > 0$,

$$\sup_u \{ur - \log \mathbb{E}(\exp(u\|Y_1 - \tau_a(Y_1)\|))\} \geq tr - \log \mathbb{E}(\exp(t\|Y_1 - \tau_a(Y_1)\|)).$$

Par le théorème de Cramér appliqué à la suite i.i.d. $(\|Y_n - \tau_a(Y_n)\|)$

$$\begin{aligned}
\limsup_n \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \|Y_i - \tau_a(Y_i)\| \geq r\right) &\leq -[tr - \log \mathbb{E}(\exp(t\|Y_1 - \tau_a(Y_1)\|))], \\
\lim_{a\to\infty} \limsup_n \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \|Y_i - \tau_a(Y_i)\| \geq r\right) &\leq -tr.
\end{aligned}$$

Ceci étant vrai pour tout $t > 0$,

$$\lim_{a\to\infty} \limsup_n \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \|Y_i - \tau_a(Y_i)\| \geq r\right) \leq -\infty.$$

Ainsi, $(\tau_a)_{a>0}$ est une approximation exponentielle de la fonction identité $id(x) = x$.

b) Lorsque la loi μ est centrée, on aura parfois intérêt à utiliser une approximation i.i.d. et centrée. On pose $m_a = \mathbb{E}(\tau_a(Y))$ et $\rho_a(x) = x\min(1, a/\|x\|) - m_a$. Ici, les suites $(\rho_a(Y_n))$ et $(Y_n - \rho_a(Y_n))$ sont i.i.d, centrées et, comme en a),

$$\lim_{a\to\infty} \limsup_n \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \|Y_i - \rho_a(Y_i)\| \geq r\right) = -\infty;$$

$(\rho_a)_{a>0}$ est une approximation exponentielle de la fonction id.

3.3.2 Transport d'un PGD vague supérieur empirique pour un couple exponentiellement équilibré

Définition 3.4

1) 1a) *Une fonction f est (v_n)-**exponentiellement bornée**, s'il existe une famille $(f_a)_{a>0}$ de fonctions continues et bornées qui est une (v_n)-approximation exponentielle de f ;*

   1b) *elle est (v_n)-**exponentiellement bien bornée**, si $(f_a)_{a>0}$ est une bonne approximation exponentielle.*

2) 2a) *Un couple (f, V) est (v_n)-**exponentiellement équilibré,** s'il existe une famille (f_a) qui est une (v_n)-approximation exponentielle de f contrôlée par V et telle que, pour tout a, (f_a, V) est un couple équilibré ;*

   2b) *il est (v_n)-**exponentiellement bien équilibré si $(f_a)_a$ est une bonne approximation exponentielle contrôlée par V.***

.

Théorème 3.2

1) Transport par une fonction exponentiellement bornée.

1a) *On suppose que (L_n) satisfait un PGD étroit supérieur de vitesse $(v_n)_n$ et de taux étroit J et que $(f_a)_{a>0}$, famille de fonctions continues et bornées, est une (v_n)-approximation exponentielle de f. Alors la suite $(\int f dL_n)$ satisfait un PGD vague supérieur de vitesse $(v_n)_n$ de taux vague*

$$h_f^{app}(x) = \lim_{r\to 0} \uparrow \limsup_{a\to\infty} h_{f_a}(B(x,r)) = \lim_{r\to 0} \uparrow \limsup_{a\to\infty} h_{f_a}\left(\overline{B}(x,r)\right).$$

1b) *Si $(f_a)_a$ est une bonne approximation exponentielle de f, alors $h_f^{app} = h_f$.*

2) Transport par un couple exponentiellement équilibré

2a) *On suppose que (L_n) satisfait un PGD vague supérieur de vitesse $(v_n)_n$ et de taux vague J. On suppose que $(f_a)_{a>0}$ est une (v_n)-approximation exponentielle de f contrôlée par V et que, pour tout a, (f_a, V) est un couple équilibré. Alors, pour tout compact Γ de E et tout $R > 0$, on a*

$$\limsup \frac{1}{v_n} \log \mathbb{P}\left(\int f dL_n \in \Gamma \text{ et } \int V dL_n \leq R\right) \leq -h_{f,V,R}^{app}(\Gamma),$$

où $h_{f,V,R}^{app}$ est le taux vague défini par

$$h_{f,V,R}^{app}(x) = \lim_{r\to 0} \uparrow \limsup_{a\to\infty} h_{f_a,V,R}(B(x,r)) = \lim_{r\to 0} \uparrow \limsup_{a\to\infty} h_{f_a,V,R}\left(\overline{B}(x,r)\right).$$

2b) *Si $(f_a)_a$ est une bonne approximation exponentielle de f contrôlée par V, alors $h_{f,V,R}^{app} = h_{f,V,R}$ qui est donc un taux vague.*

3.3.3 Transport d'un PGD inférieur

Dans les exemples markoviens qui nous intéressent, le transport inférieur sera le plus souvent démontré par une preuve spécifique. Nous transposons cependant ici la méthode des approximations exponentielles et nous montrons son intérêt dans le cas i.i.d.

Si f est continue et bornée, l'application $\nu \to \int f d\nu$ étant continue, le transport d'un PGD inférieur de taux J pour (L_n) ne pose pas de problème.

$$\liminf \frac{1}{n} \log \mathbb{P} \left(\int f dL_n \in B(x,r) \right) \geq -h_f(B(x,r));$$

$(\int f dL_n)$ satisfait un PGD inférieur de taux h_f^{sci}. Voici l'analogue du théorème 3.2.

Théorème 3.3

1) Transport d'un PGD inférieur
On suppose que (L_n) satisfait un PGD inférieur de vitesse $(v_n)_n$ et de taux étroit J et que $(f_a)_{a>0}$, famille de fonctions continues et bornées, est une (v_n)-approximation exponentielle de f.

1a) *La suite $(\int f dL_n)$ satisfait un PGD inférieur de vitesse $(v_n)_n$ et de taux vague*

$$h_f^{app-}(x) = \lim_{r\to 0} \uparrow \lim \inf_{a\to\infty} h_{f_a}(B(x,r)) = \lim_{r\to 0} \uparrow \lim \inf_{a\to\infty} h_{f_a}\left(\overline{B}(x,r)\right).$$

1b) *Si (f_a) est une bonne approximation exponentielle de f, alors $h_f^{app-} = h_f$.*

2) Transport d'un PGD étroit

2a) *On suppose que (L_n) satisfait un PGD de vitesse $(v_n)_n$ et de taux étroit J. Alors, la suite $(\int f dL_n)_n$ satisfait un PGD vague de vitesse $(v_n)_n$ et de taux vague $h_f^{app-} = h_f^{app}$.*

2b) *Si (f_a) est une bonne approximation exponentielle de f, alors $h_f^{app-} = h_f^{app} = h_f$.*

3.3.4 Références

La notion d'approximation exponentielle est étudiée par Dembo et Zeitouni ([DemZei98] p.130-137) dans le cas de transports d'un bon PGD étroit (résultats de la partie 2) du théorème 3.3). En 3.3, nous avons essentiellement repris ses techniques en les adaptant au transport d'un PGD supérieur (resp. inférieur) étroit puis en les relativisant à une fonction de Lyapounov, dans le cas du transport d'un PGD supérieur vague sur les lois empiriques. La partie 2b) du théorème 3.3 est prouvée par Deuschel et Stroock ([DeuStr89] lemme 2.1.4).

3.3.5 Démonstrations

♣ Démonstration du théorème 3.2

$*$ Montrons que ces taux approchés sont s.c.i. Cela revient à montrer que leurs épigraphes sont fermés (ou que leurs complémentaires sont ouverts). Notons h^{app} l'un ou l'autre d'entre eux et h_a pour h_{f_a} ou $h_{f_a,V,R}$. Si $h^{app}(x) > \alpha$, il existe un r tel que

$$\limsup_{a\to\infty} h_a(B(x,r)) > \alpha.$$

Pour $y \in B(x,r/2)$, $B(y,r/2) \subset B(x,r)$ et $h_a(B(y,r/2)) \geq h_a(B(x,r))$,

$$\limsup_{a\to\infty} h_a(B(y,r/2)) > \alpha \text{ donc } h^{app}(y) > \alpha.$$

Ainsi, $B(x,r/2) \subset \{y \mid h^{app}(y) > \alpha\}$.

* Montrons les PGD vagues supérieurs. En notant $A_n = E$ dans la partie 1 et $A_n = \{\int V dL_n \leq R\}$ dans la partie 2, pour tout $r > 0$,

$$\begin{aligned}\mathbb{P}\left(\left\{\int f dL_n \in \overline{B}(x,r)\right\} \cap A_n\right) &\leq \mathbb{P}\left(\left\{\int f_a dL_n \in \overline{B}(x,2r)\right\} \cap A_n\right) \\ &+ \mathbb{P}\left(\left\|\int (f - f_a)\, dL_n\right\| > r\right).\end{aligned}$$

Dans les deux cas,

$$\begin{aligned}&\limsup \frac{1}{v_n} \log \mathbb{P}\left(\left\{\int f dL_n \in \overline{B}(x,r)\right\} \cap A_n\right) \\ \leq\ & \sup \begin{pmatrix} \limsup \frac{1}{v_n} \log \mathbb{P}\left(\left\{\int f_a dL_n \in \overline{B}(x,2r)\right\} \cap A_n\right), \\ \limsup \frac{1}{v_n} \log \mathbb{P}\left(\|\int (f - f_a)\, dL_n\| > r\right) \end{pmatrix} \\ \leq\ & \sup\left\{-\inf\left\{h_a\left(\overline{B}(x,2r)\right)\right\}, \limsup \frac{1}{v_n} \log \mathbb{P}\left(\left\|\int (f - f_a)\, dL_n\right\| > r\right)\right\};\end{aligned}$$

Ceci étant vrai pour tout a, on obtient en faisant tendre a vers l'infini

$$\limsup \frac{1}{v_n} \log \mathbb{P}\left(\left\{\int f dL_n \in \overline{B}(x,r)\right\} \cap A_n\right) \leq -\limsup_{a \to 0} h_a\left(\overline{B}(x,2r)\right),$$

d'où

$$\lim_{r \to 0} \downarrow \limsup \frac{1}{v_n} \log \mathbb{P}\left(\left\{\int f dL_n \in \overline{B}(x,r)\right\} \cap A_n\right) \leq -h^{app}(x).$$

Le passage au PGD vague supérieur est classique.

* Dans le cadre de 1b) ou 2b), on introduit la notation $A = E$ dans le cadre de 1) et $A = \{\nu \mid \int V d\nu \leq R\}$ dans le cadre de la partie 2). On a, pour tout $r > 0$,

$$h_a\left(\overline{B}(x,r)\right) = J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,r)\right\} \cap A\right).$$

Remarquons d'abord que, les ensembles $A \cap \{\nu \mid J(\nu) \leq \gamma\}$ sont des compacts sur lesquels la fonction $\nu \to \int f d\nu$ est continue. Cela résulte facilement de l'hypothèse de bonne approximation.

Montrons $h^{app}(x) \geq h(x)$. Il n'y a rien à prouver si $h^{app}(x) = \infty$. Supposons donc $h^{app}(x) < \gamma$. Pour tout $r \leq r_0$, on a $\liminf_a h_a\left(\overline{B}(x,r)\right) \leq \limsup_a h_a\left(\overline{B}(x,r)\right) < \gamma$.
Cela veut dire que, pour tout $r \leq r_0$,

$$\begin{aligned}\gamma &\geq \liminf_a J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,r)\right\} \cap A\right) \\ &= \liminf_a J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,r)\right\} \cap A \cap \{\nu \mid J(\nu) \leq \gamma\}\right)\end{aligned}$$

et, pour $a \geq a_0$,

$$\begin{aligned} h\left(\overline{B}(x,2r)\right) &\leq J\left(\left\{\nu \mid \int f d\nu \in \overline{B}(x,2r)\right\} \cap A \cap \{\nu \mid J(\nu) \leq \gamma\}\right) \\ &\leq J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,r)\right\} \cap A \cap \{\nu \mid J(\nu) \leq \gamma\}\right), \\ h\left(\overline{B}(x,2r)\right) &\leq \liminf_a h_a\left(\overline{B}(x,r)\right) \leq \gamma. \end{aligned}$$

Or, $\left\{\nu \mid \int f d\nu \in \overline{B}(x,2r)\right\} \cap A \cap \{\nu \mid J(\nu) \leq \gamma\}$ est compact, et il existe un ν_r dans cet ensemble tel que $J(\nu_r) = J\left(\left\{\nu \mid \int f d\nu \in \overline{B}(x,2r)\right\} \cap A \cap \{\nu \mid J(\nu) \leq \gamma\}\right)$ et $\int f d\nu_r \in \overline{B}(x,2r)$. Pour $r \leq r_0$, la suite (ν_r) est contenue dans $A \cap \{\nu \mid J(\nu) \leq \gamma\}$; il existe une suite (r_n) majorée par r_0 qui tend vers 0 et telle que (ν_{r_n}) converge étroitement vers ν; $\int f d\nu_r \to \int f d\nu = x$. Donc :

$$h(x) \leq J(\nu) \leq \liminf J(\nu_{r_n}) \leq \limsup_{r\to 0} \liminf_a h_a\left(\overline{B}(x,r)\right) \leq \gamma.$$

Ceci étant vrai dès que $h^{app}(x) > \gamma$, on a bien $h(x) \leq h^{app}(x)$.

Montrons que $h^{app} \leq h^{sci}$. On suppose $h^{sci} < \gamma < \infty$;

$$\begin{aligned} h\left(\overline{B}(x,r)\right) &= J\left(\left\{\nu \mid \int f d\nu \in \overline{B}(x,r)\right\} \cap A\right) \\ &= J\left(\left\{\nu \mid \int f d\nu \in \overline{B}(x,r)\right\} \cap A \cap \{\nu \mid J(\nu) \leq \gamma\}\right) \end{aligned}$$

et, pour $a \geq a_0$,

$$\begin{aligned} h\left(\overline{B}(x,r)\right) &\geq J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,2r)\right\} \cap A \cap \{\nu \mid J(\nu) \leq \gamma\}\right) \\ &\geq h_a\left(\overline{B}(x,2r)\right) \end{aligned}$$

ce qui permet de conclure.

Comme $h \leq h^{app} \leq h^{sci} \leq h$ on a fini. ♢

♠ **Démonstration du théorème 3.3**

1) On prouve la partie 1a). Montrons h^{app-} est s.c.i. Si $h^{app-}(x) > \alpha$, il existe un r tel que

$$\liminf h_a(B(x,r)) > \alpha.$$

Pour $y \in B(x,r/2)$, $B(y,r/2) \subset B(x,r)$ et $h_a(B(y,r/2)) \geq h_a(B(x,r))$,

$$\liminf_{a\to\infty} h_a(B(y,r/2)) > \alpha \text{ donc } h^{app}(y) > \alpha.$$

Ainsi $B(x,r/2) \subset \{y \mid h^{app}(x) > \alpha\}$ et $(h^{app-}(x))^{sci} > \alpha$; $(h^{app-}(x))^{sci} \geq h^{app-}(x)$.

Montrons le PGD inférieur. Pour tout $r > 0$,

$$\begin{aligned} &\mathbb{P}\left(\left\{\int f_a dL_n \in B(x,r)\right\}\right) \\ \leq\ &\mathbb{P}\left(\left\{\int f dL_n \in B(x,2r)\right\}\right) + \mathbb{P}\left(\left\|\int (f-f_a)\, dL_n\right\| > r\right). \end{aligned}$$

Donc,

$$\begin{aligned}
&-h_a\left(B\left(x,r\right)\right)\\
\leq\ &\liminf \frac{1}{v_n}\log\mathbb{P}\left(\left\{\int f_a dL_n\in B\left(x,r\right)\right\}\right)\\
\leq\ &\sup\left(\liminf_n \frac{1}{v_n}\log\mathbb{P}\left(\int f dL_n\in B\left(x,2r\right)\right),\ \limsup \frac{1}{v_n}\log\mathbb{P}\left(\left\|\int\left(f-f_a\right)dL_n\right\|>r\right).\right).
\end{aligned}$$

Ceci étant vrai pour tout a, on obtient en faisant tendre a vers ∞,

$$-\liminf_a\left(-h_a\left(B\left(x,r\right)\right)\right)\leq\liminf\frac{1}{v_n}\log\mathbb{P}\left(\left\{\int f_a dL_n\in B\left(x,2r\right)\right\}\right),$$

d'où

$$\lim_{r\to 0}\downarrow\liminf\frac{1}{v_n}\log\mathbb{P}\left(\left\{\int f dL_n\in\overline{B}\left(x,r\right)\right\}\right)\geq -h^{app-}\left(x\right),$$

ce qui implique le PGD inférieur.

2) Preuve de 1b). Dans tous les cas, on a $h^{app-}\leq h^{app}\leq h$. La preuve donnée ci-dessus de la partie 1b) du théorème 3.2, montrait en fait que $h^{app-}=h$, d'où le résultat annoncé.

3) Lorsque les théorèmes 3.2 et 3.3. s'appliquent, on a le taux du PGD inférieur qui est supérieur ou égal au taux du PGD supérieur ; donc $h^{app}=h^{app-}$. $\diamondsuit$

3.4 Transports i.i.d. et markoviens

3.4.1 Transports i.i.d.

Avant d'aborder le cas markovien, il est bon de revenir au cas i.i.d. Pour une probabilité μ sur un espace de Banach séparable B, on considère la log-Laplace et la transformée de Cramér définies pour $x\in B$ et $\theta\in B'$ par :

$$\Lambda\left(\theta\right)=\log\int e^{<\theta,x>}\mu\left(dx\right)\ \text{ et }\ \Lambda^*\left(x\right)=\sup_{\theta\in B'}\left\{<\theta,x>-\Lambda\left(\theta\right)\right\}.$$

On considère aussi l'image par $h=h_{id}$ de l'information de Kullback relative à μ par la fonction identité, $id\left(x\right)=x$:

$$h\left(x\right)=\inf\left\{K\left(\nu|\mu\right)\mid\int\|y\|d\nu<\infty\text{ et }\int y\nu\left(dy\right)=x\right\}.$$

Soit (X_n) une suite i.i.d. de loi μ à valeurs dans un espace de Banach. Par le théorème de Sanov, la suite des moyennes empiriques satisfait un PGD de taux $\nu\to K\left(\nu|\mu\right)$. Sous la condition de Cramér renforcée, on a vu dans l'exemple 3.2 que, posant $\tau_a\left(x\right)=x\min\left(1,a/\|x\|\right)$, (τ_a) est une bonne approximation exponentielle de id.

Théorème 3.4 *Dans le cadre décrit ci-dessus et sous la condition de Cramér renforcée, on a, pour tout x, l'égalité $\Lambda^*\left(x\right)=h\left(x\right)$.*

La preuve de ce théorème utilisera le lemme suivant que nous énonçons ici car il sera encore utile à d'autres reprises.

Lemme 3.1 (Outil pour l'uniforme intégrabilité)) *Soient ν et μ deux probabilités, ϕ une fonction borélienne positive, $c < \infty$ et $\tau \geq 1$. On a*

$$\int_{\{\phi \geq c\}} \phi d\nu \leq \int_{\{\phi \geq c\}} \exp(\tau\phi)\, d\mu + \frac{1}{\tau} K(\nu|\mu).$$

3.4.2 Transports markoviens

Préliminaire Toutes les propriétés énoncées dans la section précédente sont encore valables (sans aucun changement dans les preuves) si l'on l'on remplace partout $\mathbb{P}$ (ou l'espérance par rapport à $\mathbb{P}$) par une famille $(\mathbb{P}_\alpha)_{\alpha \in H}$ (ou les espérances $(\mathbb{E}_\alpha)_{\alpha \in H}$ correspondantes). A un PGD sur les lois empiriques uniforme (et éventuellement à une propriété d'approximation exponentielle uniforme), correspondent des résultats uniformes.
Cadre et notations On se place dans le cadre markovien étudié en 2.3. avec les mêmes notations. Tout ce qui précède s'applique aux lois empiriques (L_n) de (X_{n-1}, X_n) ou (L_n^1) ou encore (L_n^2) avec les taux I ou I^1.
Soit une fonction de Lyapounov V. Pour g continue $E \to B$ on note h_g^1 le taux image de I^1 par g et $h_{g,V,R}^1$ le taux image relatif à $\{V \leq R\}$; de même pour G continue $E \times E$ dans B, on note h_G et $h_{G,V,R}$. Tous les résultats énoncés plus haut s'appliquent. Bien entendu, plus rien n'est valable si l'on supprime l'hypothèse de continuité de la fonction G. Voici quelques exemples pour illustrer ces propos.

Exemple 3.3
Cet exemple montre l'importance de la régularité de la fonction G pour pouvoir appliquer les théorèmes de transport.

On considère le modèle suivant

$$X_0 = x, \; X_{n+1} = \frac{1}{2} X_n + \varepsilon_{n+1}$$

où $(\varepsilon_i)_i$ est une suite i.i.d. de loi $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = \frac{1}{2}$. Nous prenons $E = [-2, 2]$. La probabilité de transition de cette chaîne de Markov est

$$p(x, .) = \frac{1}{2}\delta_{x/2+1} + \frac{1}{2}\delta_{x/2-1}$$

et sa loi invariante est la mesure de Lebesgue sur $[-2, 2]$. Bien entendu, cette transition est fellérienne et vérifie la condition de rappel exponentielle. Les seules transitions sur $[-2, 2]$ absolument continues par rapport à p sont de la forme :

$$q_{t(.)}(x, .) = t(x)\delta_{x/2+1} + (1 - t(x))\delta_{x/2-1} \text{ (où } t \text{ est une fonction à valeurs dans } [0, 1]).$$

Soit alors ν une probabilité invariante pour $q_{t(.)}$ et $\mathcal{D}$ le sous-ensemble suivant de $[-2, 2]$:

$$\left\{ 0, \sum_{j=0}^{n} \frac{i_j}{2^j}, \; n \in \mathbb{N}, \; \{i_0, i_1, ..., i_n\} \in \{-1, 1\}^{n+1} \right\}.$$

Nous allons maintenant prouver que, pour toute probabilité de transition $q_{t(.)}$ absolument continue par rapport à p, sa probabilité invariante ν est telle que $\nu(\mathcal{D}) = 0$

– *Nous commencons par prouver que* $\nu(0) = \nu(1) = \nu(-1) = 0$
Si $q_t(x,2) > 0$ *alors* $x = 2$ *et on obtient* $\nu q_{t(.)}(2) = t(2)\nu(2) = \nu(2)$. *Cela donne* $\nu(2) > 0 \Rightarrow t(2) = 1$. *De même,* $\nu(-2) > 0 \Rightarrow t(-2) = 0$.
Mais, si $q_{t(.)}(x,0) > 0$, *alors* $x = 2$ *ou bien* $x = -2$. *On a donc* $\nu(0) = \nu q_{t(.)}(0) = (1 - t(2))\nu(2) + t(-2)\nu(-2)$. *Par conséquent* $\nu(0) = 0$.
D'autre part, si $q_{t(.)}(x,1) > 0$, *alors* $x = 0$. *Ainsi,* $\nu(1) = \nu(0)q_{t(.)}(0,1) = 0$ *et* $\nu(1) = \nu(-1) = 0$.
– *On prouve ensuite que, pour tout entier* $n \le N$ *et tout* $\{i_0, i_1, ..., i_n\} \in \{-1,1\}^{n+1}$,

$$\nu\left(\sum_{j=0}^{n} \frac{i_j}{2^j}\right) = 0.$$

Supposons que cette propriété soit vraie pour $n \ge N$. *Soit* $a = \sum_{j=0}^{N+1} \frac{i_j}{2^j}$ *avec* $(i_0, i_1, ..., i_{N+1}) \in \{-1,1\}^{N+2}$. *Si* $q_{t(.)}(x,a) > 0$, *alors* $x = 2(a-1)$ *si* $i_0 = 1$ *et* $x = 2(a+1)$ *si* $i_0 = -1$. *Dans les deux cas,* $x = 2\sum_{j=1}^{N+1} \frac{i_j}{2^j} = \sum_{j=0}^{N} \frac{i_{j+1}}{2^j}$. *Nous avons supposé que* $\nu(x) = 0$ *donc* $\nu(a) = 0$. *Cela prouve que la propriété est juste pour* $N+1$.

Nous avons prouvé par récurrence que $\nu(\mathcal{D}) = 0$. *Par conséquent, le taux* $h_{\mathbb{I}_D}$ *est infini partout sauf en* 0 *où il est nul. Partant de l'état initial* 0, *la chaîne reste dans* $\mathcal{D}$ *et*

$$\frac{1}{n}\log \mathbb{P}_0(\int \mathbb{I}_D dL_n^1 = 1) = 0;$$

pour la loi initiale δ_0, $(\int \mathbb{I}_D dL_n^1)_n$ *ne satisfait pas un PGD supérieur (même vague) de taux* $h_{\mathbb{I}_D}$.

Exemple 3.4 (conséquences d'une condition de rappel exponentiel) *Supposons que la chaîne de Markov est presque fellérienne et satisfait la condition de rappel exponentiel associée à* (U,V) *(voir la définition 2.4 de 2.3.2). Rappelons que, pour tout* M, *les lois empiriques satisfont des PGD supérieurs étroits uniformes sur tout compact de* $\{\nu | \int U d\nu \le M\}$. *En outre, si l'on note* $H_M = \left\{\nu | \int e^U d\nu \le M\right\}$,

$$\begin{aligned} \sup_n \sup_{\nu \in H_M} \mathbb{E}_\nu\left(\exp\left(n\int V dL_n\right)\right) &= A_M < \infty, \\ \forall r > 0,\ \sup_n \sup_{\nu \in H_M} \mathbb{P}_\nu\left(\int V dL_n > r\right) &\le A_M e^{-r}. \end{aligned}$$

Soit g *une fonction continue* $E \to B$ *ou* G *continue* $E^2 \to B$. *On a les propriétés évidentes suivantes :*

a) *La suite* $(\int V dL_n)$ *est exponentiellement tendue.*

b) *Si* ρ *est surlinéaire et* $\rho(\|g\|) \le V$ *ou* $\rho(\|G\|) \le V \oplus V$, *le couple* (g,V) *ou le couple* (G,V) *sont équilibrés. Mais on a ici mieux puisque, pour toute probabilité,*

$$\begin{aligned} \rho(a)\int_{\|g\| \ge a} \|g\| d\nu &\le \int V d\nu, \\ \frac{1}{n}\log \sup_{\nu \in H_M} \mathbb{P}_\nu\left(\int_{\|g\| \ge a} \|g\| dL_n^1 \ge r\right) &\le A_M \exp\left(-r\rho(a)\right), \end{aligned}$$

$$\lim_{a\to\infty}\frac{1}{n}\log\sup_{\nu\in H_M}\mathbb{P}_\nu\left(\int_{\|g\|\geq a}\|g\|dL_n^1\geq r\right) \quad = \quad -\infty.$$

Ainsi, la suite tronquée $(\tau_a(g))$ est une bonne approximation exponentielle de g. De même pour G. On retrouve ainsi la proposition suivante.

Proposition 3.1 *Dans le cadre presque fellérien avec condition de rappel exponentiel associée à (U,V), en notant $H_M=\left\{\nu|\int e^U d\nu\leq M\right\}$ pour $M<\infty$, on a les résultats suivants :*

a) Pour $g:B\to E$, continue et telle que $\rho(\|g\|)\leq V$, $\left(\int g dL_n^1\right)$ satisfait un PGD étroit supérieur de taux étroit h_g^1 uniforme pour les lois initiales $\nu\in H_M$.

b) Pour $G:B\to E\times E$, continue et telle que $\rho(\|G\|)\leq V\oplus V$, on a les mêmes propriétés.

Lorsque la condition de domination de la fonction G par la fonction de Lyapounov associée à la condition de rappel exponentiel n'est pas vérifiée, il est possible de ne pas avoir le PGD étroit supérieur. L'exemple suivant traite d'un tel cas, où la fonctionnelle est de l'ordre de la fonction de Lyapounov. Il prouve que la proposition 3.1 ne peut pas être beaucoup améliorée.

Exemple 3.5

– ***Description de la chaîne de Markov***

$E(.)$ désignant la partie entière, les ensembles d'entiers suivants forment une partition des entiers naturels :

$$\{u_m\overset{\triangle}{=}9m;\ m\in\mathbb{N}\},\ \left\{v_m\overset{\triangle}{=}3\left(m+E\left(\frac{m-1}{2}\right)\right);\ m\in\mathbb{N}\right\}\ \ et\ \ \left\{w_m\overset{\triangle}{=}m+E\left(\frac{m-1}{2}\right);\ m\in\mathbb{N}\right\}.$$

On considère une chaîne de Markov à valeurs dans $\mathbb{N}$ de transition :

$*$ $p(0,u_m)=p_m>0\ \forall m\in\mathbb{N}$,

$*$ $p(u_m,v_m)=p(v_m,w_m)=p(w_m,0)=1$.

Cette chaîne est récurrente et apériodique de loi stationnaire μ telle que :

$*$ $\mu(0)=\frac{1}{4-3p_0}=c$,

$*$ $\mu(u_m)=\mu(v_m)=\mu(w_m)=cp_m$.

Afin de calculer le taux de Donsker-Varadhan I^1, on cherche les probabilités de transition absolument continues par rapport à p. Elles sont de la forme :

$*$ $q(0,u_m)=q_m>0$,

$*$ $q(u_m,v_m)=q(v_m,w_m)=q(w_m,0)=1$.

Les probabilités invariantes pour q sont définies par :

$*$ $\mu_q(0)=\frac{1}{4-3q_0}=c(q)$,

$*$ $\mu_q(u_m)=\mu_q(v_m)=\mu_q(w_m)=c(q)q_m$.

Ainsi, $I(\mu_q/q)=\mu_q(0)K(q(0,.)/p(0,.))=\mu_q(0)\sum_{m=0}^{\infty}q_m\log\frac{q_m}{p_m}$ et, pour toute probabilité sur $\mathbb{N}^2$ qui n'est pas de la forme μ_q/q, le taux I est infini.

Nous avons prouvé que si $I^1(\nu)<\infty$ alors ν est égal à l'une des probabilités μ_q et $I^1(\nu)=c(q)\sum_{m=0}^{\infty}q_m\log\frac{q_m}{p_m}$.

– ***Une fonction g particulière***
Soit g la fonction définie par $g(u_m) = m$, $g(v_m) = -2m$ et $g(w_m) = m$.
Quelle que soit la probabilité ν telle que $I^1(\nu) < \infty$, nous avons $\int g d\nu = 0$ et, par conséquent, $h_g(x) = 0$ si $x = 0$ et $h_g(x) = \infty$ sinon. D'autre part, si on pose $r(t) \stackrel{\triangle}{=} \sum_{j \geq t} p_j$ et $a > 0$, on a pour cette fonction

$$\mathbb{P}_x \left(\int g dL_n^1 \geq a \right) = \mathbb{P}(X_{n-1} = 0) r(an),$$

$$\mathbb{P}_x \left(\int g dL_n^1 \leq -a \right) = \mathbb{P}(X_{n-2} = 0) r(an).$$

De plus, $\mathbb{P}(X_n = 0) \to c = \frac{1}{4 - 3p_0}$.

Par conséquent, si la suite $(p_m)_m$ est telle que $\frac{1}{n} \log r(an) \to R(a) > -\infty$ et si R est continue, strictement croissante vers l'infini alors $(\int g dL_n^1)_n$ vérifie un PGD de taux $J(x) = -R(|x|)$ pour tout état initial. On ne peut donc pas avoir un PGD vague supérieur de taux h_g.

– ***Une condition de rappel exponentiel***
Supposons que la transition p vérifie la condition de rappel exponentiel avec une fonction U positive, croissante vers l'infini et telle que :

$$\begin{aligned} V(0) &= U(0) - \log \textstyle\sum_{m \geq 0} e^{U(9m)} p_m > -\infty, \\ V(u_m) &= U(u_m) - U(v_m) \to \infty, \\ V(v_m) &= U(v_m) - U(w_m) \to \infty, \\ V(w_m) &= U(w_m) - 1 \to \infty. \end{aligned}$$

La condition sur $V(0)$ impose qu'il existe $M > 0$ tel que $\sum_{m \geq 0} e^{U(9m)} p_m \leq M$ donc, $e^{U(9am)} r(am) \leq M$. Ainsi, si $\frac{1}{n} \log r(an) \to R(a) > -\infty$, alors $\limsup_{n \to \infty} \frac{1}{n} U(9an)$ est inférieure à une constante et $U(n) = O(n)$.
Prenons, par exemple, la suite $p_n = \frac{1}{2^{n+1}}$; $\frac{1}{n} \log r(an) \to -[a] \log 2$. Si on considère la fonction $U(m) = \frac{m \log 2}{18}$, alors la condition en 0 est verifiée ainsi, bien sûr, que les trois autres. La condition de rappel exponentiel associée à (U, V) est vérifiée pour cette chaîne de Markov, $\|g\| = O(V)$ sans pour autant avoir le PGD vague supérieur de taux h_g.

3.4.3 Transports autorégressifs

Sur un espace de probabilité $(\Omega, \mathcal{A}, P)$, on considère le modèle markovien autorégressif fonctionnel à valeurs dans $\mathbb{R}^d$ suivant :

$$X_{n+1} = A(X_n) + \sigma(X_n) \varepsilon_{n+1}$$

où A est continue de $\mathbb{R}^d$ dans $\mathbb{R}^d$, σ est une fonction continue de $\mathbb{R}^d$ dans un intervalle $[1/\beta, \beta]$ avec $1 < \beta < \infty$ et $(\varepsilon_n)_{n \geq 1}$ est une suite i.i.d. de vecteurs aléatoires de dimension

d supposée indépendante de X_0; la loi de X_0 est notée α. Si p est la probabilité de transition de cette chaîne de Markov, $p(x,\cdot)$ est la loi de $A(x)+\sigma(x)\varepsilon_1$; c'est une probabilité de transition fellérienne.
Voici un résultat important, sur lequel on reviendra dans le chapitre 4, *qui n'exige aucune condition de rappel.*

Théorème 3.5 (PGD supérieur contrôlé uniforme) *Dans le cadre autorégressif décrit ci-dessus, on considère $\phi : \mathbb{R}^d \to \mathbb{R}^d$ continue et telle que $\|\phi\|^2$ soit une fonction de Lyapounov et la fonction f définie par :*

$$f(x,y) = <\phi(x), (y - A(x))/\sigma(x)> .$$

On considère les variables aléatoires

$$M_n = \sum_{i=1}^{n} <\phi(X_{i-1}), \varepsilon_i> \quad et\ C_n = \frac{1}{2}\left(\|\phi(X_0)\|^2 + \|\phi(X_n)\|^2\right) + \sum_{i=1}^{n-1} \|\phi(X_i)\|^2$$

et, pour $R > 0$, la fonction H_R définie sur $\mathbb{R}$ par :

$$H_R(u) = \inf\left\{ I(\zeta) \mid \int f d\zeta = u \ et \ \int \|\phi\|^2 d\zeta^1 \le R \right\}.$$

Sous la condition de Cramér renforcée sur $\|\varepsilon_1\|^2$ ($\forall t > 0$, $\mathbb{E}\left(\exp\left(t\|\varepsilon_1\|^2\right)\right) < \infty$), la fonction H_R est un taux vague et, quel que soit Γ fermé de $\mathbb{R}$,

$$\limsup_n \frac{1}{n} \log \sup_{\alpha \in \mathcal{P}(\mathbb{R}^d)} \mathbb{P}_\alpha\left(\frac{M_n}{n} \in \Gamma \ et\ \frac{C_n}{n} \le R \right) \le -H_R(\Gamma).$$

3.4.4 Références

Dans le cas i.i.d., l'égalité entre transformée de Cramér et taux image sous l'hypothèse de Cramér renforcée a été bien souvent prouvée dans diverses situations. La première preuve générale est due à Donsker et Varadhan [DonVar76] ; Bretagnolle généralise le théorème 3.4 à un espace lusinien en obtenant, sous la condition de Cramér renforcée, une fonction surlinéaire ρ telle que $\int e^{\rho(\|x\|)}\mu(dx) < \infty$ ce qui permet selon le théorème 3.1 d'affirmer que $\int \rho(\|x\|)\nu(dx)$ est bornée sur $\{\nu \mid K(\nu \mid \mu) \le R\}$. Dupuis et Ellis ([DupEll97]) en donnent une preuve sur $\mathbb{R}^d$ généralisée à un résultat uniforme pour des lois dépendant de paramètres (lemme 6.2.3). La proposition 3.1. vient de [GulLipLot94]. L'exemple 3.4 est inspiré d'un exemple donné par De Acosta et Ney dans [AcoNey98].
Des références relatives aux modèles autorégressifs seront données en 4.4.

3.4.5 Preuves des résultats de 3.3

♣ Démonstration du lemme 3.1 d'uniforme intégrabilité
On remarque d'abord que, pour a et b réels positifs et $\tau \ge 1$, on a par l'inégalité de Young,

$$ab \le \exp(\tau a) + \frac{1}{\tau}(b\log b - b + 1).$$

En effet

$$\sup_a (ab - \exp(\tau a)) = \frac{b}{\tau}\left(\log\frac{b}{\tau} - 1\right) \leq \frac{1}{\tau}(b\log b - b + 1).$$

Si $K(\nu \mid \mu) < \infty$, $d\nu = f d\mu$ et comme $f\log f - f + 1 \geq 0$,

$$\begin{aligned}\int_{\{\phi\geq c\}} \phi d\nu &\leq \int_{\{\phi\geq c\}} \exp(\tau\phi)\, d\mu + \frac{1}{\tau}\int (f\log f - f + 1)\, d\mu \\ &= \int_{\{\phi\geq c\}} \exp(\tau\phi)\, d\mu + \frac{1}{\tau}K(\nu|\mu). \diamondsuit\end{aligned}$$

♠ **Démonstration du théorème 3.4.**

Elle résulte de l'exemple 3.2 et de la partie 2*b*) du théorème 3.3 si l'on vérifie que pour tout $R > 0$,

$$\lim_{a\to\infty} \sup_{\nu |\, K(\nu|\mu)\leq R} \left\| \int (x - \tau_a(x))\, d\nu \right\| = 0;$$

il suffit pour cela de vérifier l'uniforme intégrabilité de id relative à $\{\nu \mid K(\nu|\mu) \leq R\}$ sous la condition de Cramér renforcée.

Or, on a alors pour tout $t > 0$,

$$M(a,t) = \sup_{\nu |\, K(\nu|\mu)\leq R} \int \mathbb{I}_{\|x\|>a} e^{t\|x\|}\mu(dx) \leq e^{-ta} \sup_{\nu |\, K(\nu|\mu)\leq R} \int e^{2t\|x\|}\mu(dx),$$

et, selon le lemme 3.1 donné ci-dessus, si $K(\nu|\mu) \leq R$ et si $\phi(x) = \|x\|$,

$$\begin{aligned}\int \mathbb{I}_{\|x\|>a}\|x\|\nu(dx) &\leq M(a,t) + R/t, \\ \lim_{a\to\infty} \sup_{\nu |\, K(\nu|\mu)\leq R} \int \mathbb{I}_{\|x\|>a}\|x\|\nu(dx) &\leq R/t\end{aligned}$$

et l'on conclut en faisant tendre t vers l'infini. $\diamondsuit$

△ **Démonstration du théorème 3.5**

On suppose que $\|\varepsilon_1\|^2$ vérifie la condition de Cramér renforcée. On a :

$$M_n = \sum_{i=1}^{n} < \phi(X_{i-1}), \varepsilon_i >= \sum_{i=1}^{n} < \phi(X_{i-1}), (X_i - A(X_{i-1}))/\sigma(X_{i-1}) > .$$

Avec les notations de l'exemple 3.2, pour $a > 0$,

$$\begin{aligned} f_a(x,y) &= < \phi(x),\, \tau_a((y - A(x))/\sigma(x)) >, \\ (f - f_a)(x,y) &= < \phi(x), (y - A(x))/\sigma(x) - \tau_a((y - A(x))/\sigma(x)) > .\end{aligned}$$

Montrons que (f_a) est une approximation exponentielle de f contrôlée par la fonction de Lyapounov $V = \frac{1}{2}\|\phi\|^2 \oplus \|\phi\|^2$.

On a :

$$\sum_{i=1}^{n} < \phi(X_{i-1}), \tau_a((X_i - A(X_{i-1}))/\sigma(X_{i-1})) > \;=\; \sum_{i=1}^{n} < \phi(X_{i-1}), \tau_a(\varepsilon_i) >,$$

$$\begin{aligned}
\left(\frac{1}{n}\sum_{i=1}^{n} < \phi\left(X_{i-1}\right), \varepsilon_i - \tau_a\left(\varepsilon_i\right) >\right)^2 &\le \frac{1}{n}\sum_{i=1}^{n}\|\phi\left(X_{i-1}\right)\|^2 \frac{1}{n}\sum_{i=1}^{n}\|\varepsilon_i - \tau_a\left(\varepsilon_i\right)\|^2. \\
\left\{\left\|\int\left(f - f_a\right) dL_n\right\| > r \text{ et } \int V dL_n \le R\right\} &\subset \left\{\left\|\int\left(f - f_a\right) dL_n\right\| > r \text{ et } \int \|\phi\|^2 dL_n^1 \le 2R\right\} \\
&\subset \left\{\sum_{i=1}^{n}\|\varepsilon_i - \tau_a\left(\varepsilon_i\right)\|^2 > nr^2/2R\right\}.
\end{aligned}$$

Comme dans l'exemple 3.2, dès que $\left(\|\varepsilon_n\|^2\right)$ satisfait la condition de Cramér renforcée, la suite des fonctions (f_a) est une approximation exponentielle contrôlée par la fonction de Lyapounov

$$V = \frac{\|\phi\|^2 \oplus \|\phi\|^2}{2}.$$

Montrons maintenant la propriété suivante : pour tous $\gamma > 0$ et $R > 0$,

$$\lim_{a\to\infty} \sup_{\left\{\zeta \mid I(\zeta)\le\gamma \text{ et } \int V d\zeta \le R\right\}} \left\|\int\left(f - f_a\right) d\zeta\right\| = 0.$$

En effet, notant $\zeta = \nu \oslash q$,

$$\left\|\int\left(f - f_a\right) d\zeta\right\|^2 \le \int \|\phi\left(x\right)\|^2 \left(\int \left\|\left(\frac{y - A(x)}{\sigma(x)}\right) - \tau_a\left(\frac{y - A\left(x\right)}{\sigma(x)}\right)\right\|^2 q\left(x, dy\right)\right) \nu\left(dx\right);$$

et, si $\zeta^1 = \zeta^2$,

$$I\left(\zeta\right) = \int K\left(q\left(x, \cdot\right) | p\left(x, \cdot\right)\right) \nu\left(dx\right).$$

Mais ici, pour tout x, $p\left(x, \cdot\right)$ est la loi de $A\left(x\right) + \sigma(x)\varepsilon_1$, donc $(y - A\left(x\right))/\sigma(x) - \tau_a\left(\left(y - A\left(x\right)\right)/\sigma(x)\right)$ a la loi de $\varepsilon_1 - \tau_a(\varepsilon_1)$;

$$\begin{aligned}
&\int \exp\left(t\left\|\left(\frac{y - A(x)}{\sigma(x)}\right) - \tau_a\left(\frac{y - A\left(x\right)}{\sigma(x)}\right)\right\|^2\right) p\left(x, dy\right) \\
\le\ & \mathbb{E}\left(\exp\left(t\|\varepsilon_1 \mathbb{I}_{\|\varepsilon_1\|\ge a}\|^2\right)\right) = M\left(a, t\right)
\end{aligned}$$

et, quel que soit $t, \lim_{a\to\infty} M\left(a, t\right) = 0$. Par la relation variationnelle, on a, pour tout x,

$$\begin{aligned}
t\int \left\|\left(\frac{y - A(x)}{\sigma(x)}\right) - \tau_a\left(\frac{y - A\left(x\right)}{\sigma(x)}\right)\right\|^2 q\left(x, dy\right) &\le M\left(a, t\right) + K\left(q\left(x, \cdot\right) | p\left(x, \cdot\right)\right), \\
\int \left\|\left(\frac{y - A(x)}{\sigma(x)}\right) - \tau_a\left(\frac{y - A\left(x\right)}{\sigma(x)}\right)\right\|^2 \zeta\left(dx, dy\right) &\le \frac{1}{t} M\left(a, t\right) + \frac{1}{t} I\left(\zeta\right);
\end{aligned}$$

pour toute ζ telle que $I\left(\zeta\right) \le \gamma$, quel que soit $\alpha > 0$, on fixe $t > 2\gamma/\alpha$; pour $a \ge a_0$, $M\left(a, t\right) \le \alpha/2$.

Ainsi, pour tout α, il existe un a_0 pour lequel, uniformément sur $\left\{\zeta \mid I\left(\zeta\right) \le \gamma \text{ et } \int \|\phi\|^2 d\zeta^1 \le R\right\}$,

$$\sup_{a \ge a_0} \left\|\int\left(f - f_a\right) d\zeta\right\| \le \sqrt{R\alpha}.$$

Comme le taux est infini si les deux marginales sont différentes, on a :

$$\inf\left(I(\zeta) | \int f d\zeta = u \text{ et } \frac{1}{2}\int \|\phi\| \oplus \|\phi\| d\zeta \leq R\right) = \inf\left(I(\zeta) | \int f d\zeta = u \text{ et } \int \|\phi\| d\zeta^1 \leq R/2\right).$$

Le théorème 3.2 s'applique donc. $\diamondsuit$

3.5 PGD inférieurs markoviens

Pour les chaînes de Markov, nous n'avons pour le moment abordé ni les PGD inférieurs pour les lois empiriques ni leur éventuel transfert vers des fonctionnelles additives. On comble ici partiellement cette lacune. Les méthodes sont spécifiques et les PGD inférieurs sur les fonctionnelles ne se déduiront pas des PGD empiriques par la méthode décrite en 3.3. Selon les commentaires de 2.4, on ne s'étonnera pas de l'introduction d'hypothèses permettant "l'oubli de la loi initiale".
Dans toute cette section, on se place dans la situation décrite en introduction de 2.3, en gardant les notations.

3.5.1 A propos de l'irréductibilité

Définitions et notations
On se place exceptionnellement sur un espace d'états $(E, \mathcal{E})$ mesurable en supposant $\mathcal{E}$ engendrée par une famille dénombrable de parties de E. On considère une probabilité de transition p sur E et on lui associe le noyau et la probabilité de transition définis par :

$$U(x, \bullet) = \sum_{n=1}^{\infty} p^n(x, \bullet) \text{ et } U_a(x, \bullet) = (1-a)\sum_{n=0}^{\infty} a^n p^{n+1}(x, \bullet).$$

Pour une probabilité β sur $(E, \mathcal{E})$, p est **β-irréductible** si, pour tout x, β est absolument continue par rapport à $U(x, \bullet)$; s'il existe une mesure d'irréductibilité, il existe toujours une mesure d'irréductibilité **maximale** α telle que pour toute autre mesure d'irréductibilité β, $\beta << \alpha$. De même, il existe alors toujours un **petit ensemble** $C \in \mathcal{E}$ tel que $\alpha(C) > 0$, un entier $m \geq 1$, une constante $\delta_m > 0$ et une probabilité ξ concentrée sur C pour lesquels :

$$p^m(x, \bullet) \geq \delta_m \mathbb{I}_C(x)\, \xi.$$

Alors, on a aussi $U(x, \bullet) \geq \delta \mathbb{I}_C(x)\, \xi$ pour une constante $\delta > 0$.

Condition de minoration de la chaîne de Markov bivariée
On a vu qu'il est utile d'associer à une chaîne de Markov $(X_n)_{n\geq 0}$ la chaîne $(X_n, X_{n+1})_{n\geq 0}$ dont la probabilité de transition est définie par $q((w,x); \bullet) = \delta_x \oslash p(x, \bullet)$; sous l'hypothèse de minoration $p^m(x, \bullet) \geq \delta_m \mathbb{I}_C(x)\, \xi$, pour $A \in \mathcal{E}^{\otimes 2}$,

$$\begin{aligned} q((w,x); A) &= \mathbb{E}_x(\mathbb{I}_A(x, X_1)), \\ \xi \oslash p^m(E \times C) &= \int p^m(z, C)\, \xi(dz) \geq \delta. \end{aligned}$$

Pour tout x,

$$\begin{aligned}
\mathbb{E}_x\left(\mathbb{I}_A\left(X_{2m-1},X_{2m}\right)\right) &= \int\left(\int\mathbb{I}_A\left(y,z\right)p^m\left(y,dz\right)\right)p^m\left(x,dy\right)\\
&\geq \delta_m\mathbb{I}_C\left(x\right)\int\left(\int\mathbb{I}_A\left(y,z\right)p^m\left(y,dz\right)\right)\xi\left(dy\right),\\
q^{2m}\left(\left(w,x\right);\bullet\right) &\geq \delta_m\mathbb{I}_C\left(x\right)\xi\oslash p^m.
\end{aligned}$$

On est dans la même situation que la chaîne initiale avec le petit ensemble $E\times C$ et la mesure $\xi\oslash p^m$.

Technique classique
Sous la condition de minoration plus forte $p\left(x,\bullet\right)\geq\delta\mathbb{I}_C\left(x\right)\xi$, on étudie une chaîne de Markov de transition p à l'aide d'une méthode de "fission" qui construit une nouvelle chaîne de Markov "atomique" d'espace d'états $E\times\{0,1\}$ dont la première composante est une chaîne de Markov de transition p ; la probabilité de transition de cette nouvelle chaîne est constante sur l'atome $E\times\{1\}$. Cette nouvelle chaîne a une structure de renouvellement lors des passages à l'atome qui permet son étude par des méthodes proches du cas i.i.d. On passe ensuite aux cas $p^m\left(x,\bullet\right)\geq\delta_m\mathbb{I}_C\left(x\right)\xi$ et $U\left(x,\bullet\right)\geq\delta\mathbb{I}_C\left(x\right)\xi$ par des méthodes bien rôdées.

Récurrence et transience
Notant $R_C=\{\omega;\sum_{n=1}^{\infty}\mathbb{I}_C\left(X_n\left(\omega\right)\right)=\infty\}$ on a deux cas.

- Condition de récurrence : $\xi U\left(C\right)=\infty$ et $\mathbb{P}_\xi\left(R_C\right)=1$;
- Condition de transience : $\xi U\left(C\right)<\infty$ et $\mathbb{P}_\xi\left(R_C\right)=0$.

S'il existe une probabilité invariante μ ($\mu p=\mu$), la condition de récurrence est assurée. Alors si f est une fonction $\mu-$intégrable de E dans $\mathbb{R}^d$ et si ν est une loi initiale pour laquelle $\mathbb{P}_\nu\left(R_C\right)=\mathbb{P}_\nu\left(\sum_{n=1}^{\infty}\mathbb{I}_C\left(X_n\left(\omega\right)\right)>0\right)=1$, alors, $\mathbb{P}_\nu-p.s$,

$$\frac{1}{n}\sum_{i=1}^{n}f\left(X_i\right)\rightarrow\int fd\mu.$$

C'est vrai, pour la loi initiale ξ et pour les lois initiales δ_x pour ξ presque tout x.
La chaîne de Markov est dite récurrente de loi invariante μ si cette propriété est vraie pour toute loi initiale. C'est le cas si $\mathbb{P}_x\left(R_C\right)=1$ pour tout x.

3.5.2 Résultats connus sur les PGD inférieurs spectraux

Voici un résumé rapide des résultats de ce type connus.
1) PGD convexes
Le point de départ de ces PGD inférieurs est, sous l'hypothèse de minoration, l'obtention de deux "PGD convexes".

- **(Dinwoodie et Ney ([DinNey95])**
  Pour E polonais et Γ un ouvert convexe de $\mathcal{P}\left(E\right)$,

$$\lim\frac{1}{n}\log\mathbb{P}_\xi\left(\left(L_n^2\in\Gamma\right)\cap\mathbb{I}_C\left(X_n\right)\right)=-J\left(\Gamma\right);$$

– **De Acosta et Ney ([AcoNey98])**
Pour E mesurable arbitraire et f fonction borélienne bornée à valeurs dans un espace de Banach séparable B et si Γ est un ouvert de Γ,
$$\lim \frac{1}{n} \log \mathbb{P}_\xi \left(\left(L_n^2 \in \Gamma \right) \cap \mathbb{I}_C (X_n) \right) = -J_f (\Gamma) .$$
C'est un PGD convexe portant sur des mesures de masse ≤ 1, mais les m éthodes convexes du chapitre 1 s'appliquent et le théorème de Dinwoodie permet d'identifier les taux. On obtient les résultats suivants en reprenant les notations définies en 2.2.5 et utilisées en 2.4.1.
– Dans le cadre empirique, quelle que soit la fonction $f \in C_b (E)$,
$$\begin{aligned} \lim \frac{1}{n} \log \mathbb{E}_\xi \left(\exp \left(n \int f dL_n^2 \right) \cap \mathbb{I}_C (X_n) \right) &= \lim \frac{1}{n} \log \mathbb{E}_{\xi p} \left(\exp \left(\sum_{i=0}^{n-1} f (X_i) \right) \mathbb{I}_C (X_n) \right) \\ &= \lim \frac{1}{n} \log \int T_f^{n-1} (\bullet, C) \, d\xi p = \lambda (f) , \\ J (\nu) &= \sup_{f \in \mathcal{C}_b (E)} \left\{ \int f d\nu - \lambda (f) \right\} . \end{aligned}$$
Une comparaisons avec le taux de Donsker-Varadhan est facile : $\lambda (f) \leq \Lambda (f)$ donc $J (\nu) \geq I^1 (\nu)$; λ ne dépend pas du couple (C, ξ) .
– Dans le cadre des fonctionnelles additives, pour $\theta \in B'$,
$$\begin{aligned} \lim \frac{1}{n} \log \mathbb{E}_\xi \left(\exp \left(n < \theta, \int f dL_n^2 > \right) \cap \mathbb{I}_C (X_n) \right) &= \lim \frac{1}{n} \log \mathbb{E}_{\xi p} \left(\exp \left(\sum_{i=0}^{n-1} < \theta, f (X_i) > \right) \mathbb{I}_C (X_n \right. \\ &= \lim \frac{1}{n} \log \int T_{<\theta,f>}^{n-1} (\bullet, C) \, d\xi p = \lambda_f (\theta) , \\ J_f (x) &= \sup_{\theta \in B'} \{ < \theta, x > - \lambda_f (\theta) \} \\ &\geq \sup_{\theta \in B'} \{ < \theta, x > - \Lambda (< \theta, f >) \} , \end{aligned}$$
le minorant correspondant à la fonctionnelle Λ définie en 2.2.5 et utilisée pour les PGD supérieurs de De Acosta (voir 2.4.1).

2) PGD inférieurs
Les articles cités ci-dessus déduisent (au prix de longs efforts dans le second cas surtout) de la partie 1, que pour toute loi initiale l'hypothèse d'irréductibilité entraine les résultats très généraux suivants.
– Si (E, d) est polonais $\left(L_n^2 \right)$ satisfait un PGD inférieur de taux vague J ;
– Pour E mesurable quelconque et une fonction f borélienne quelconque de E dans un espace de Banach séparable, $\left(\int f dL_n^2 \right)$ satisfait un PGD inférieur de taux vague J_f.

3) Conditions suffisantes à l'obtention d'un PGD empirique
Pour les lois empiriques sur un espace polonais, on a toujours $J \geq I^1$, I^1 étant le taux de Donsker et Varadhan.
L'inégalité peut être stricte même dans le cas dénombrable. On ne connait pas de condition nécessaire et suffisante pour l'égalité $J = I^1$.
Voici, la probabilité de transition étant toujours irréductible, une condition suffisante.

Condition 3.1 (Condition suffisante de De Acosta [?], théorème 6.3 et [?]) *p est fellérienne et, pour tout borélien A absorbant (c'est à dire tel que* $\forall x \in A,\ p\,(x, A) = 1$*) il existe un entier k tel que*

$$\forall x \in A^c,\ p^k\,(x, A^c) = 0.$$

C'est par exemple le cas lorsqu'il existe une mesure d'irréductibilité maximale α *pour laquelle*

$$\forall x,\ p\,(x, \bullet) << \alpha,$$

puisque, dans ce cas, si A est absorbant, $\alpha\,(A^c) = 0$ *et* $\forall x,\ p\,(x, A^c) = 0$.

D'après le théorème 3.2.1, cette condition permet aussi le transport vers des fonctions continues bornées ou pour lesquelles existe une bonne approximation exponentielle par des fonctions continues et bornées.

4) Questions ouvertes sur l'égalité des deux taux spectraux

Il est facile, par la méthode de De Acosta, d'obtenir des PGD supérieurs de taux

$$\overline{J}_f(x) = \sup_\theta \{< \theta, x > -\Lambda(< \theta, f >)\} \leq J_f(x)$$

(voir [Aco85a] par exemple, pour f majorée). Mais, malgré d'évidentes tentatives, la question de bons critères d'égalité des deux taux spectraux reste ouverte, tant sur les classes de probabilités de transition que sur la classe de fonctions F auxquelles permettant d'affirmer l'égalité des taux. C'est ce qui nous a menés à utiliser la méthode de Dupuis et Ellis afin de montrer directement des PGD avec les taux de Donsker et Varadhan.

3.5.3 Vers un principe de Laplace inférieur de fonctionnelle limite conjuguée au taux de Donsker-Varadhan

Nous suivons la méthode utilisée par Dupuis et Ellis ([DupEll97]) pour prouver le PGD inférieur relatif aux lois empiriques.

Selon le théorème 1.4, on obtient, lorsque E est polonais, un PGD inférieur de taux h pour la suite des lois empiriques $(L_n)_n$ si l'on prouve que, pour toute fonction $\Phi : \mathcal{P}(E^2) \to \mathbb{R}$ lipschitzienne et bornée, on a

$$\liminf_n \frac{1}{n} \log \mathbb{E}\left(e^{-n\Phi(L_n)}\right) \geq - \inf_{\zeta \in \mathcal{P}(E^2)} \{\Phi(\zeta) - I(\zeta)\}\,.$$

Lorsque $F : E^2 \to B$ est une fonction borélienne à valeurs dans $\mathbb{R}^d$, on obtient un PGD inférieur de taux h_F^{sci} pour la suite $(\int F dL_n)_n$ si l'on prouve que, pour toute fonction $\Psi : B \to \mathbb{R}$ lipschitzienne et bornée, on a

$$\liminf_n \frac{1}{n} \log \mathbb{E}\left(e^{-n\Psi(\int F dL_n)}\right) \geq - \inf_{x \in B} \{\Psi(x) - h_F(x)\}\,.$$

• **Représentation variationnelle à l'instant n pour les lois empiriques**

Il s'agit de minorer la quantité $\Lambda_n(\Phi) = \frac{1}{n} \log \mathbb{E}\left(e^{-n\Phi(L_n)}\right)$.

On reprend la même méthode que celle utilisée en 2.3.4 pour une loi initiale ν indépendante de n et $j \leq n$, posons :

$$L_{j,n} \stackrel{\triangle}{=} \frac{1}{n} \sum_{i=1}^{j} \delta_{X_{i-1},X_i} \quad \text{et} \quad L_{n,n} = L_n.$$

La suite $(X_j, L_{j,n})_j$ est une chaîne de Markov de loi initiale $\nu \otimes \delta_0$ et de transition

$$q_{j,n}\left((y,\rho);\bullet\right) = \text{ loi de } \left(X, \rho + \frac{1}{n}\delta_y \otimes \delta_X\right) \quad \text{avec } X \text{ de loi } p(y,\bullet).$$

Soit $\mathcal{F}_j = \sigma(X_i, i \leq j)$. Comme en 2.3.4, définissons $(H_{j,n})_j$ par :

$$\begin{array}{rcl} H_{j,n}(X_j, L_{j,n}) & \stackrel{\triangle}{=} & -\frac{1}{n}\log \mathbb{E}\left(e^{-n\Phi(L_n)} \mid \mathcal{F}_j\right), \\ H_{0,n}(X_0, 0) & = & -\Lambda_n(\Phi), \\ H_n(X_n, L_n) & = & \Phi(L_n); \end{array}$$

$$\begin{array}{rcl} H_{j,n}(y,\rho) & = & -\frac{1}{n}\log \int e^{-nH_{j+1,n}(z,\rho+\frac{1}{n}\delta_u\otimes\delta_z)} \\ & = & \inf_{\zeta\in\mathcal{P}(E^2)} \left\{\frac{1}{n}K\left(\zeta \mid \delta_y \oslash p\right) + \int H_{j+1,n}\left(z, \rho + \frac{1}{n}\delta_u \times \delta_z\right)\zeta(du,dz)\right\}. \end{array}$$

Si on note $\mathcal{M}_C(E^2)$ l'espace des mesures ζ sur E^2 telles que $\zeta(E^2) \leq C$, cela correspond à un système d'équations de la programmation dynamique d'une chaîne contrôlée à valeurs dans $E \times \mathcal{M}_{j/n}(E^2)$ à l'instant j, avec la loi initiale $\nu \oslash \delta_0$ et l'espace des contrôles $\mathcal{P}(E^2)$, la transition à l'instant j étant

$$Q_{j,n}(y,\rho,\zeta;\bullet) = \text{ loi de } (Y^2, \rho + \frac{1}{n}\delta_{Y^1,Y^2}) \quad \text{où } Y = (Y^1, Y^2) \text{ a la loi } \zeta.$$

Le coût final est $(y,\rho) \to \Phi(\rho)$ et le coût courant est $c_j(y,\rho,\zeta) = \frac{1}{n}K\left(\zeta \mid \delta_y \oslash p\right)$.

Le coût minimal moyen est majoré par les coûts obtenus en se contentant d'une classe plus restrictive de stratégies que l'on décrit ici.

Définition 3.5 *On considère $\mathcal{R} \subset \mathcal{P}(E^2)$, l'ensemble des probabilités de la forme $\zeta = \alpha \oslash q$ pour lesquelles*
- *$I(\alpha \oslash q) < \infty$ (donc satisfaisant $\alpha q = \alpha$) ;*
- *q est la transition d'une chaîne de Markov récurrente.*

Soit $\zeta = \alpha \oslash q \in \mathcal{R}$. Prenons alors, quel que soit ρ, la stratégie $s(y,\rho;\bullet) = \delta_y \oslash q$. La chaîne contrôlée correspond, étant donnée une chaîne de Markov $(Y_j)_{0\leq j\leq n}$ de loi initiale ν et de transition q, à la suite $(Y_j, \frac{1}{n}\sum_{i=1}^{j}\delta_{Y_{i-1},Y_i})$. On en déduit

$$-\Lambda_n(\Phi) \leq \mathbb{E}\left(\sum_{i=0}^{n-1} K\left(\delta_{Y_i} \oslash q / \delta_{Y_i} \oslash p\right) + \Phi\left(\frac{1}{n}\sum_{i=1}^{n}\delta(Y_{i-1}, Y_i)\right)\right).$$

Sur un espace polonais, la récurrence implique la stabilité c'est à dire que, p.s., $\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{Y_{i-1},Y_i}\right)$ converge étroitement vers ζ. Donc $\Phi\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{Y_{i-1},Y_i}\right)$ tend, p.s. vers $\Phi\left(\zeta\right)$.

En outre, la fonction $x \to K(q(x,) \mid p(x,))$ est s.c.i. donc borélienne et positive. Elle est en outre intégrable par rapport à la probabilité invariante α car

$$\int K(\delta_x \oslash q \mid \delta_x \oslash p)\,\alpha(dx) = \int K(q(x,) \mid p(x,))\,\alpha(dx) = I(\zeta).$$

Donc, grâce à la récurrence, $\left(\frac{1}{n}\sum_{j=1}^{n} K\left(\delta_{Y_j} \oslash q \mid \delta_{Y_j} \oslash p\right)\right)$ tend p.s. vers $I(\zeta)$.
Finalement :

$$\begin{aligned} \liminf \Lambda_n(\Phi) &\geq -I(\zeta) - \Phi(\zeta) \\ \liminf \Lambda_n(\Phi) &= -\sup_{\zeta \in \mathcal{R}} (I(\zeta) + \Phi(\zeta)). \end{aligned}$$

- **Représentation variationnelle à l'instant n pour les fonctionnelles additives**

On pose :

$$\begin{aligned} R_n &= \int F dL_n \\ R_{j,n} &= R_{j-1,n} + \frac{1}{n} F(X_{j-1}, X_j). \end{aligned}$$

Soit Ψ une fonction réelle lipschitziennne et bornée définie sur B,

$$\begin{aligned} -\Lambda_n(\Psi) &= -\frac{1}{n} \log \mathbb{E}(\exp(-n\Psi(R_n))), \\ H_{j,n}(X_j, R_{j,n}) &= -\frac{1}{n} \log \mathbb{E}(\exp(-n\Psi(R_n)) \,|\mathcal{F}_j). \end{aligned}$$

Cette expression se calcule par récurrence en marche arrière par :

$$\begin{aligned} H_{n,n}(X_{n,n}, R_{n,n}) &= \Psi(R_n), \\ H_{j,n}(y,r) &= -\frac{1}{n} \log \int \exp\left(-nH_{j+1,n}\left(z, r + \frac{F(y',z)}{n}\right)\right) \delta_y \otimes p(dy'; dz) \\ &= \inf_{\zeta \in \mathcal{P}(E^2)} \left\{ \frac{1}{n} K(\zeta \mid \delta_y \otimes p) + \int H_{j+1,n}\left(z, r + \frac{F(y',z)}{n}\right) d\zeta(y',z) \right\}. \end{aligned}$$

Ces équations correspondent à la programmation dynamique d'horizon n d'une chaîne de Markov contrôlée à valeurs dans $E \times \mathbb{R}^q$ d'état initial 0, avec l'espace de contrôles $\mathcal{P}(E^2)$ et la transition :

$$Q_{j,n}(y, r, \zeta; \bullet) = \text{loi de } \left(Y^2, r + \frac{1}{n} F(Y)\right) \text{ où } Y = \left(Y^1, Y^2\right) \text{ a la loi } \zeta,$$

le coût final étant $(y, r) \to \Psi(r)$ et le coût courant égal à $c(y,\zeta) = \frac{1}{n} K(\zeta \mid \delta_y \oslash p)$.
On considère maintenant $\zeta \in \mathcal{R}$ et la même stratégie que pour les lois empiriques. La chaîne contrôlée correspond, étant donnée une chaîne de Markov $(Y_j)_{0 \leq j \leq n}$ de loi initiale ν et de transition q, à la suite $\left(Y_j, \frac{1}{n}\sum_{i=1}^{j} F(Y_{i-1}, Y_i)\right)$. On en déduit

$$-\Lambda_n(f) \leq \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n} K\left(\delta_{Y_j} \oslash q \mid \delta_{Y_j} \oslash p\right) + \Psi\left(\frac{1}{n}\sum_{i=1}^{n} F(Y_{i-1}, Y_i)\right)\right)$$

et, par la récurrence, si $\int \|F\| d\zeta < \infty$,

$$\lim_n \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n} K\left(\delta_{Y_j} \oslash q \mid \delta_{Y_j} \oslash p\right) + \Psi\left(\frac{1}{n}\sum_{i=1}^{n} F\left(Y_{i-1}, Y_i\right)\right)\right) = I(\zeta) + \Psi\left(\int F d\zeta\right),$$

et

$$\liminf_n \Lambda_n(\Psi) \geq -\inf\left\{I(\zeta) + \Psi\left(\int F d\zeta\right);\ \zeta \in \mathcal{R},\ \int \|F\| d\zeta < \infty\right\}.$$

3.5.4 Conditions assurant la validité du principe de Laplace inférieur

♣ Méthode générale pour pouvoir conclure

On aura obtenu les PGD inférieurs de taux I et h_F^{sci} si, pour toute $\zeta \in \mathcal{P}\left(E^2\right)$ et tout δ, il existe une probabilité ζ^ρ de $\mathcal{R}$ telle que

- pour les lois empiriques, $\Phi\left(\zeta^\rho\right) + I\left(\zeta^\rho\right) \leq \Phi\left(\zeta\right) + I\left(\zeta\right) + \delta$;
- pour la suite $\left(\int F dL_n\right)$ lorsque $\int \|F\| d\zeta < \infty$,

$$\int \|F\| d\zeta^\rho < \infty \text{ et } \Psi\left(\int F d\zeta^\rho\right) + I\left(\zeta^\rho\right) \leq \Psi\left(\int F d\zeta\right) + I\left(\zeta\right) + \delta.$$

Il suffit bien sûr de faire les preuves pour $I\left(\zeta\right) < \infty$. On rappelle que Φ ou Ψ sont lipschitziennes et bornée. Sur $\mathcal{P}\left(E^2\right)$ on suppose Φ lipschitzienne pour la distance de Dudley.

Considérons une mesure $\rho \in \mathcal{P}\left(E^2\right)$ telle que $I\left(\rho\right) < \infty$ et, pour l'étude de $\left(\int F dL_n\right)$, $\int \|F\| d\rho < \infty$. Posons $\zeta_t = (1-t)\,\zeta + t\rho$ lorsque $0 < t < 0$. On a $d_{Dudley}\left(\zeta_t, \zeta\right) \leq t$ et, par la convexité de I,

$$\begin{aligned} I\left(\zeta_t\right) &\leq (1-t)\, I\left(\zeta\right) + tI\left(\rho\right), \\ \Phi\left(\zeta_t\right) + I\left(\zeta_t\right) &\leq \Phi\left(\zeta\right) + I\left(\zeta\right) + t\|\Phi\|_{\mathcal{LB}} + tI\left(\rho\right); \end{aligned}$$

et, comme $\|\int F d\zeta_t - \int F d\zeta\| \leq t \int \|F\| d\zeta + t \int \|F\| d\zeta_t \leq 2t \int \|F\| d\zeta + t \int \|F\| d\rho$,

$$\Psi\left(\int F d\zeta_t\right) + I\left(\zeta_t\right) \leq \Psi\left(\int F d\zeta\right) + I\left(\zeta\right) + t\|\Psi\|_{\mathcal{LB}}\left(2\int \|F\| d\zeta + \int \|F\| d\rho\right) + tI\left(\rho\right).$$

Pour conclure, il suffit donc de proposer une $\rho \in \mathcal{P}\left(E^2\right)$ telle que $I\left(\rho\right) < \infty$ et, pour l'étude de $\left(\int F dL_n\right)$, $\int \|F\| d\rho < \infty$ ayant la propriété suivante ; pour tout ζ tel que $I\left(\zeta\right) < \infty$,

$$\forall t \in]0,1[,\ \zeta_t = (1-t)\,\zeta + t\rho \in \mathcal{R}.$$

On rappelle que $I\left(\zeta\right) < \infty$ implique $\zeta = \zeta^1 \oslash q$ avec $\zeta^1 q = \zeta^1$ et, pour tout x, $q\left(x, \bullet\right) << p\left(x, \bullet\right)$.

♠ Condition de domination de la probabilité de transition sans hypothèse de récurrence

On considère la condition de domination de la probabilité de transition p suivante, voisine de la condition 3.1 de De Acosta.

Condition 3.2 (Domination de la transition par α) *On suppose qu'il existe une probabilité α et une fonction $h : E^2 \to]0, \infty[$ borélienne telles que*

$$\begin{aligned} \forall x,\ p(x, \bullet) &= h(x, \bullet)\,\alpha, \\ \int (-\log h)\, d(\alpha \otimes \alpha) &< \infty. \end{aligned}$$

On a alors
$$\int K(\alpha | p(x, \bullet))\, \alpha(dx) = \int (-\log h)\, d(\alpha \otimes \alpha) < \infty.$$

Ceci implique que $K(\alpha \otimes \alpha \mid \alpha \oslash p) < \infty$.
Si $I(\zeta) < \infty$, on pose, pour tout $t \in]0, 1[$,

$$\zeta_t = (1 - t)\,\zeta + t\alpha \otimes \alpha.$$

On a
$$I(\zeta_t) \leq (1 - t) I(\zeta) + tK(\alpha \otimes \alpha \mid \alpha \oslash p) < \infty$$

et ζ_t peut s'écrire sous la forme $\zeta^1 \oslash q_t$ où q_t est une transition telle que $q_t(x, \bullet) << p(x, \bullet) << \alpha$.
Comme $I(\zeta_t) < \infty$, ζ_t^1 est une probabilité invariante pour q_t donc $\zeta_t^1 << \alpha$; comme $\zeta_t << \zeta_t^1 \oslash p$, on a aussi $\zeta_t << \alpha \otimes \alpha$.
Soit h_t une densité de ζ_t par rapport à $\alpha \otimes \alpha$; comme $\zeta_t \geq t\alpha \otimes \alpha$, on peut prendre $h_t \geq t$ partout. Pour $h_t^1(x) = \int h_t(x, y)\, \alpha(dy)$, on peut poser :

$$q_t(x, dy) = \frac{h_t(x, y)}{h_t^1(x)} \alpha(dy).$$

Posons $D = \{x; h_t^1(x) < \infty\}$ et $D_N = \{\{x; h_t^1(x) \leq N\}\}$; on a $\zeta_t^1(D) = \lim \zeta_t^1(D_N) = 1$ donc $\alpha(D) = 1$ puisque ζ_t^1 et α sont équivalentes. Pour N assez grand, $\alpha(D_N) > 0$ et en posant $\alpha_N = \frac{\alpha(\bullet \cap D_N)}{\alpha(D_N)}$,

$$q_t(x, \bullet) \geq \frac{t}{N} \mathbb{I}_{D_N}(x)\, \alpha_N(\bullet).$$

On a donc pour q_t une condition de minorisation et une probabilité invariante ; la condition de récurrence est satisfaite. Donc on a la récurrence pour α_N-presque tout état initial ; comme c'est vrai pour tout N assez grand, on a la récurrence pour ζ_t^1-presque tout état initial, c'est à dire pour α-presque tout état initial. Mais la récurrence avec la loi initiale δ_x équivaut à la récurrence pour $q_t(x, \bullet)$ et $q_t(x, \bullet) << \alpha$; on a donc la récurrence pour tout état initial.
On a montré que $\zeta_t \in \mathcal{R}$, pour tout $t \in]0, 1[$.

On aboutit finalement au théorème suivant où aucune hypothèse de continuité ou de rappel n'est faite sur la transition p et sur la fonction F. Ces PGD inférieurs sont valables pour toute loi initiale, mais sans propriété d'uniformité sur des classes de lois initiales. Ceux qui sont énoncés pour L_n^1 ou L_n^2 et pour $f : E \to B$ se déduisent aisément des résultats bivariés.

Théorème 3.6 (PGD inférieur sans condition de stabilité) *On considère une probabilité de transition p sur un espace polonais E satisfaisant la condition 3.2 de domination par α.*
On a alors les résultats suivants en considérant une chaîne de Markov de loi initiale ν dont on note (L_n) les lois empiriques bivariées.

a) Pour les taux de Donsker-Varadhan I et I^1, (L_n) satisfait un PGD inférieur de taux I ; (L_n^1) et (L_n^2) satisfont un PGD inférieur de taux I^1.

b) Soit F (resp. f) une fonction borélienne de E dans $\mathbb{R}^d$ telle que $\int \|F\| d\alpha \otimes \alpha < \infty$ (resp. $\int \|f\| d\alpha < \infty$) ; pour les taux images, $(\int F dL_n)$ satisfait un PGD inférieur de taux h_F^{sci} ; $(\int f dL_n^1)$ et $(\int f dL_n^2)$ satisfont un PGD inférieur de taux h_f^{sci}.

Exemple 3.6 *Soit un modèle autorégressif fonctionnel de dimension d*

$$X_{n+1} = A(X_n) + \sigma(X_n)\varepsilon_{n+1}$$

où le bruit (ε_n) est une suite i.i.d. de vecteurs aléatoires de dimension d ayant une densité g strictement positive par rapport à la mesure de Lebesgue, A et σ sont des fonctions boréliennes définies sur $\mathbb{R}^d$, A à valeurs dans $\mathbb{R}^d$ et σ à valeurs dans un intervalle $[a_1, a_2]$, $0 < a_1 < a_2 < \infty$; X_0 est indépendant du bruit et de loi ν. Alors $p(x, \bullet)$ a, par rapport à la mesure de Lebesgue, la densité $y \to g((y - f(x)/\sigma(x)))$. On peut appliquer le résultat précédent en choisissant une probabilité α ayant une densité r strictement positive par rapport à la mesure de Lebesgue donc en prenant

$$h(x, y) = r(y)\, g((y - A(x)/\sigma(x))).$$

On fera un choix de r imposant

$$\int\int (-\log[g((y - A(x)/\sigma(x)))])\, r(x)\, r(y)\, dxdy + \int (-\log r(y))\, r(y)\, dy < \infty.$$

Pour le PGD inférieur relatif à $(\int F dL_n)$, on impose en plus $\int\int \|F(x,y)\| r(x)\, r(y)\, dxdy < \infty$.
Voici deux exemples.

$*$ Bruit gaussien
Si le bruit est gaussien centré de covariance Γ inversible, $-\log g(x) = \frac{1}{2} x^T \Gamma^{-1} x$ et la condition devient

$$\int\int \frac{1}{2}((y - A(x))/\sigma(x))^T \Gamma^{-1} x ((y - A(x))/\sigma(x))\, r(x)\, r(y)\, dxdy + \int (-\log r(y))\, r(y)\, dy < \infty;$$

c'est à dire $\int (\|x\|^2 + \|A(x)\|^2 + (-\log r(x)))\, r(x)\, dx < \infty$.
Lorsque, pour deux constantes B et C arbitraires,

$$\begin{aligned} \|A(x)\| &\leq C \exp\left(B\|x\|^2\right), \\ \|F(x,y)\| &\leq C \exp\left(B\|x\|^2 + B\|y\|^2\right), \end{aligned}$$

toute loi α gaussienne centrée et de covariance K fois la matrice identité avec $K > B/2$ convient.

$*$ Une condition faible sur la densité
On suppose que pour des constantes B et C arbitraires,

$$\begin{aligned} \|A(x)\| &\leq B\|x\| + C \\ \|F(x,y)\| &\leq C \exp\left(B\|x\|^2 + B\|y\|^2\right). \end{aligned}$$

Ce qui vient d'être écrit s'applique à toute densité g telle que, pour deux constantes $\delta > 0$ et $\gamma > 0$,

$$\log g(x) \geq \delta \exp\left(-\gamma \|x\|^2\right).$$

En effet, on a alors

$$\begin{aligned}
-\log\left(g\left(\frac{y-A(x)}{\sigma(x)}\right)\right) &\leq -\delta\left(\exp\left(-\gamma\left\|\frac{y-A(x)}{\sigma(x)}\right\|^2\right)\right) \\
&\leq -\delta\left(\exp\left(-\gamma a_1^{-2}\|y-A(x)\|^2\right)\right) \\
&\leq -\delta\left(\exp\left(-2\gamma a_1^{-2}\left(\|y\|^2+\|A(x)\|^2\right)\right)\right) \\
&\leq -\delta/2\left(\exp\left(-4\gamma a_1^{-2}\|y\|^2\right)+\exp\left(-4\gamma a_1^{-2}\|A(x)\|^2\right)\right).
\end{aligned}$$

Une loi α gaussienne centrée de covariance K fois la matrice identité conviendra pour K assez grand.
Ces deux exemples montrent que le champs d'application de cette méthode est assez vaste et convient bien à des situations non récurrentes.

♢ Condition d'irréductibilité forte dans le cas récurrent

Sous une hypothèse d'irréductibilité forte, Dupuis et Ellis ont prouvé un PGD inférieur pour la suite des lois empiriques. L'hypothèse est un peu plus faible que celle utilisée initialement par Donsker-Varadhan pour prouver ce même résultat.

Hypothèse 3.1 (Irréductibilité renforcée) *p vérifie les deux conditions suivantes :*

1) Il existe un entier L tel que, pour tout couple $(x, y) \in E^2$, on ait :

$$\sum_{i=L}^{\infty} \frac{1}{2^i} p^i(x, \bullet) << \sum_{i=1}^{\infty} \frac{1}{2^i} p^i(y, \bullet).$$

2) p a une probabilité invariante μ.

Toute probabilité de transition p vérifiant cette hypothèse est associée à une chaîne de Markov récurrente.
A l'aide du lemme suivant, nous allons prouver que la probabilité $\mu \oslash p$ satisfait la propriété permettant d'affirmer que le principe de Laplace inférieur est vérifié.

Lemme 3.2 *Sous l'hypothèse d'irréductibilité renforcée, pour toute probabilité $\zeta \in \mathcal{P}(E^2)$ telle que $I(\zeta) < \infty$, il existe une probabilité de transition q de loi stationnaire ζ^1 telle que $q^n(x, \bullet) << p^n(x, \bullet)$ pour tout $n \in \mathbb{N}$ et tout $x \in E$. De plus $\zeta^1 << \mu$.*

Preuve

Comme $K\left(\zeta/\zeta^1 \otimes p\right) < \infty$, d'après le résultat sur les probabilités conditionnelles régulières rappelé en 2.2.2, il existe une probabilité de transition q de loi stationnaire ζ^1 telle que $\zeta = \zeta^1 \otimes q$. Par suite, $q(x, \bullet) << p(x, \bullet)$ ζ^1 presque-sûrement. On peut modifier q en

posant $q(x,\bullet)=p(x,\bullet)$ aux points x de E pour lesquels ce n'est pas vérifié et ζ^1 reste invariante pour ce q modifié. Pour tout n, $\zeta^1 << q^n$.
Par récurrence, $q^n(x,\bullet) << p^n(x,\bullet)$ pour tout $n\geq 1$. Ainsi, on a

$$\zeta^1 << \sum_{i\geq L} q^i(x,\bullet) << \sum_{i\geq L} p^i(x,\bullet) << \mu, \text{ et donc } \zeta^1 << \mu. \Diamond$$

Pour tout $t\in]0,1[$, on pose $\zeta_t \stackrel{\triangle}{=} (1-t)\zeta + t\mu\oslash p$. Par la convexité du taux I,

$$I(\zeta_t)=K\left(\zeta_t/\zeta_t^1\oslash p\right)\leq (1-t)K\left(\zeta/\zeta\oslash p\right)<\infty \text{et} \int \|G\|d\zeta_t=(1-t)\int\|G\|d\zeta+t\int\|G\|d\mu\otimes p<\infty.$$

Il reste à montrer que ζ_t est un élément de la classe $\mathcal{R}$. Par le lemme 3.2, il existe une probabilité de transition q_t de loi stationnaire ζ_t^1 telle que pour tout n et tout x de E, on ait $q_t^n(x,\bullet) << p_t^n(x,\bullet)$. De plus, $\zeta_t^1 << \mu$. Montrons que q_t vérifie l'hypothèse d'irréductibilité renforcée.

Par définition de ζ_t, $\zeta_t^1(\bullet)\geq t\mu(\bullet)$. Les probabilités ζ_t^1 et μ sont donc équivalentes. On note h la densité de ζ_t^1 par rapport à μ ; cette fonction est supérieure à t. Soient A et B deux boréliens de E ;

$$\int_A q_t(x,B)h(x)\mu(dx)=\zeta_t(A\times B)\geq t\mu\otimes p(A\times B)=t\int_A p(x,B)\mu(dx).$$

Par conséquent , $q_t(x,\bullet)\geq \frac{t}{h(x)}p(x,\bullet)$ pour μ presque tout x. On modifie q_t de la même manière que dans la preuve du lemme 3.2 de façon à ce que cette inégalité soit satisfaite pour tout x de E. Cela ne pose aucun problème car $\zeta_t^1\sim\mu$ donc q_t est modifié sur un ensemble de mesure nulle pour ζ_t^1 qui reste donc invariante pour q_t. Ainsi $q_t^n(x,\bullet)\sim p^n(x,\bullet)$ pour tout n et tout $x\in E$.

Finalement $q_t^n(x,\bullet)\sim p^n(x,\bullet)$ et q_t vérifie, comme p, l'hypothèse d'irréductibilité renforcée donc est la transition d'une chaîne de Markov récurrente.
Ceci conduit au théorème suivant ; là encore, aucunes hypothèses de régularité ne sont faites sur la transition p et la fonctionnelle.

Théorème 3.7 (PGD inférieur pour une chaîne de Markov stable) *On considère une probabilité de transition p sur un espace polonais E satisfaisant l'hypothèse d'irréductibilité renforcée avec la probabilité invariante μ. On a alors les résultats suivant, en considérant une chaîne de Markov de loi intiale ν et de probabilité de transition p.*

a) La suite des lois empiriques $(L_n)_n$ vérifie un PGD inférieur de vitesse n et de taux I ; $(L_n^1)_n$ et $(L_n^2)_n$ satisfont un PGD inférieur de taux I^1.

b) Si $F: E^2\to B$ est une fonction mesurable, intégrable par rapport à $\mu\otimes p$ (resp $f: E\to B$ est une fonction mesurable, intégrable par rapport à μ), la suite des moyennes empiriques $\int FdL_n$ satisfait un PGD inférieur de taux h_F^{sci} ; $(\int fdL_n^1)_n$ et $(\int fdL_n^1)_n$ satisfont un PGD inférieur de taux h_f^{sci}.

3.5.5 Références

La notion d'irréductibilité a été introduite dès 1956 par T. E. Harris. La méthode de fission de la chaîne est due à Cellier [Cel80], Nummelin [Num84] ; Meyn et Tweedie [MeyTwe93] en donnent une étude fouillée, accompagnée d'exemples. Voir aussi Revuz [Rev84] et Duflo [Duf97].

Chapitre 4

PGD autonormalisés

4.1 Introduction

4.1.1 Objectifs de ce chapitre

Comme on l'a vu au cours du premier chapitre, notre méthode pour l'obtention d'un PGD partiel relatif à une certaine classe pour un couple $(\int f dL_n, \int V dL_n)_n$ ($(L_n)_n$ désignant toujours une suite de probabilités aléatoires) passe par deux étapes dont l'approche est différente.

- Obtention d'un PGD vague supérieur pour le couple.
- Preuve de la tension exponentielle partielle relative à la classe considérée.

Les résultats du chapitre 3 portant sur les transports de PGD supérieurs sont traduisibles en termes de PGD supérieurs vagues. On est en mesure de prouver de tels résultats pour des couples équilibrés, mais aussi pour des couples exponentiellement équilibrés. Afin d'établir des PGD autonormalisés, on revient aux classes de Chernov (voir la définition 1.8 au chapitre 1). Ayant obtenu un PGD vague, il reste à prouver la tension exponentielle partielle, ce que permettent certains critères énoncés dans la partie 1.4.4. A partir des PGD partiels relatifs aux classes de Chernov, le "passage au quotient" peut s'avérer délicat pour obtenir des PGD autonormalisés. C'est ce que l'on explique, tout en essayant de dégager des situations pour lesquelles on évite le problème.

4.1.2 Plan du chapitre

La section 4.2 se place dans le cadre général des sections 3.2 et 3.3, où $(L_n)_n$ est une suite de probabilités aléatoires satisfaisant un PGD vague. On traduit en 4.2.1 les résultats du chapitre 3 en un PGD vague supérieur pour $(\int f dL_n, \int V dL_n)_n$ lorsque (f, V) est un couple exponentiellement équilibré à valeurs dans $\mathbb{R}_+ \times B$ (B espace de Banach séparable). Puis, en 4.2.2, on examine les conséquences d'un éventuel PGD partiel pour $(\int f dL_n, \int V dL_n)_n$ relatif à une classe de Chernov $\mathcal{S}(U)$ (définie en 1.4.2), U étant une fonction de Lyapounov définie sur B. On obtient notamment sous certaines conditions un PGD supérieur portant sur les quotients $\left(U\left(\int f dL_n\right) / \int V dL_n\right)_n$ et, dans les cas où $(\int f dL_n, \int V dL_n)_n$ satisfait aussi un PGD inférieur de taux image $h_{f,V}$ (situation étudiée en 3.3 et 3.5 dans le cadre markovien), on obtient un PGD étroit.

En 4.3, on applique cette méthode au cadre général "exponentiellement stable" d'une suite de probabilités aléatoires satisfaisant un PGD étroit. On obtient alors facilement un critère de tension exponentielle partielle de $(\int f dL_n, \int V dL_n)$ relativement à une classe de Chernov $\mathcal{S}(U)$, pour V et U convenables. Lorsque $B = \mathbb{R}^d$, on obtient par exemple des PGD relatifs à $\left(\frac{\|\int f dL_n\|}{\left(\int \|f\|^p dL_n\right)^{1/p}}\right)_n$.

On aborde, dans la section 4.4, des modèles autorégressifs sans hypothèse de stabilité et l'on prouve un PGD autonormalisé général.
En 4.5, on présente le PGD de Heck-Maaouia [HecMaa01] relatifs aux lois empiriques logarithmiques de $(M_n/\sqrt{n})_n$, $(M_n)_n$ étant une fonctionnelle additive martingale d'une chaîne de Markov récurrente. On montre comment les résultats de 4.3 s'appliquent à ce cadre

4.2 Méthode générale pour l'autonormalisation

On se place dans le cadre suivant tout au long de cette section.

Axiome 4.1 *On suppose que :*
- (L_n) *est une suite de probabilités aléatoires définies sur un espace polonais* (E, d) *satisfaisant un PGD vague supérieur de taux vague* J *et de vitesse* $v = (v_n)$;
- V *est une fonction de Lyapounov s.c.i. définie sur* (E, d) ;
- f *est une fonction continue de* (E, d) *dans un espace de Banach séparable* B.

Les résultats des sections 3.2 et 3.3 sont interprétés en termes de PGD vagues en 4.2.1.

4.2.1 PGD vague pour un couple exponentiellement équilibré

On introduit ici un taux noté $h^*_{f,V}$, a priori moins bon que le taux image $h_{f,V}$ (voir la définition 3.1), mais adapté à l'étude des PGD autonormalisés.

Définition 4.1 *On appelle* ***taux image partiel*** *de* (f, V) *la fonction* $h^*_{f,V}$ *définie sur* $B \times \mathbb{R}_+$ *par l'expression :*

$$h^*_{f,V}(x, R) \stackrel{\triangle}{=} h_{f,V}(\{x\} \times [0, R]) = J\left\{\nu \mid \int V d\nu \leq R, \ \int f d\nu = x\right\}.$$

La proposition 4.1 est la traduction du théorème 3.1. en terme de PGD vague. Le taux naturellement obtenu est le taux image partiel défini ci-dessus.

Proposition 4.1 (PGD vague supérieur pour un couple équilibré) *On suppose que le couple* (f, V) *est équilibré.*
1) $h^*_{f,V}$ *est un taux vague ;* $h^*_{f,V}(x, R) = J(\nu_{x,R})$ *avec une probabilité* $\nu_{x,R}$ *telle que* $\int f d\nu_{x,R} = x$ *et* $\int V d\nu_{x,R} \leq R$.
2) *Lorsque* J *est convexe,* $h^*_{f,V}$ *est convexe.*
3) *La suite* $(\int f dL_n, \int V dL_n)$ *satisfait un PGD vague supérieur de vitesse* v *et de taux vague* $h^*_{f,V}$.

Quand (f, V) est exponentiellement équilibré, on obtient le résultat suivant issu du théorème 3.2.

Proposition 4.2 (PGD vague supérieur pour un couple exponentiellement équilibré) *On suppose que (f_a) est une approximation exponentielle de f relative à V et que, pour tout a, le couple (f_a, V) est équilibré. Alors*

1) *La suite $(\int f dL_n, \int V dL_n)$ satisfait un PGD vague supérieur de vitesse v et de taux vague $h^{*app}_{f,V}$ défini par :*

$$h^{*app}_{f,V}(x,R) = \lim_{r\to 0} \uparrow \limsup_{a\to\infty} h_{f_a,V}\left(\overline{B}(x,r)\times[0,R+r]\right).$$

2) *Si le couple (f, V) est (v_n)-exponentiellement bien équilibré, on a les propriétés suivantes.*
2a) $h^{*app}_{f,V} = h^*_{f,V}$ qui est donc un taux vague.
2b) Pour tout $(x,R) \in B\times\mathbb{R}_+$, il existe une probabilité $\nu_{x,R}$ telle que $\int f d\nu_{x,R} = x$ et $\int V d\nu_{x,R} \le R$, $J(\nu_{x,R}) = h^*_{f,V}(x,R)$.
2c) Lorsque J est convexe, $h^*_{f,V}$ est convexe.

4.2.2 PGD autonormalisé et Chernov

En vue d'appliquer le PGD vague de la section précédente pour obtenir un PGD partiel, on se munit des conditions requises regroupées dans l'axiome suivant.

Axiome 4.2 *On suppose que :*

- *(L_n) est une suite de probabilités aléatoires définies sur un espace polonais (E, d) satisfaisant un PGD vague supérieur de taux vague J et de vitesse $v = (v_n)$;*
- *f est une fonction continue de (E, d) dans un espace de Banach séparable B ;*
- *le couple (f, V) est exponentiellement equilibré ;*
- *U est une fonction de Lyapounov convexe à valeurs dans $[0, \infty[$, définie sur B, et le couple $(\int f dL_n, \int V dL_n)_n$ est (v_n)-exponentiellement tendu pour la classe de Chernov*

$$\mathcal{S}(U) \stackrel{\triangle}{=} \left\{A \text{ borélien de } B\times\mathbb{R}_+ \mid \lim_{R\to\infty}\inf\left\{\frac{U(s)}{v} \mid (s,v)\in A,\ v > R\right\} > 0\right\}.$$

Par la proposition 1.1 et la proposition 4.2, on a le résultat suivant.

Corollaire 4.1 *Dans le cadre de l'axiome précédent, on a les résultats suivants.*

1) *$(\int f dL_n, \int V dL_n)$ satisfait un PGD partiel supérieur de vitesse $(v_n)_n$ et de taux $h^{*app}_{f,V}$ relatif à la classe de Chernov $\mathcal{S}(U)$.*
2) *Si (f, V) est (v_n)-exponentiellement bien équilibré,*
2a) $(\int f dL_n, \int V dL_n)$ satisfait un PGD vague supérieur de vitesse v dont le taux $h^*_{f,V}$ est s.c.i. et convexe

$$\begin{aligned} h^*_{f,V}(x,R) &= \inf\left\{J(\nu) \mid \nu\in\mathcal{P}(E), \int f d\nu = x \text{ et } \int V d\nu \le R\right\} \\ &= J(\nu_{x,R}) \text{ pour } \nu_{x,R} \in \left\{\nu\in\mathcal{P}(E) \mid \int f d\nu = x \text{ et } \int V d\nu \le R\right\}; \end{aligned}$$

2b) $(\int f dL_n, \int V dL_n)$ satisfait un PGD partiel supérieur de vitesse $(v_n)_n$ et de taux $h^*_{f,V}$ relatif à la classe de Chernov $\mathcal{S}(U)$.

4.2.3 PGD troué en 0

4.2.3.1 Préliminaire aux PGD troués : le problème "0/0"

De manière générale, un PGD autonormalisé est un PGD portant sur un quotient du type S_n/V_n, où (S_n) est une suite de variables aléatoires à valeurs dans $\mathbb{R}$ et (V_n) est une suite aléatoire à valeurs dans $\mathbb{R}_+$. Même si le couple (S_n, V_n) satisfait un PGD étroit, un principe de contraction menant à un PGD sur S_n/V_n ne s'applique que si V_n est moralement "assez loin de 0", c'est à dire si l'on a par exemple

$$\lim_{r\to 0} \limsup_n \frac{1}{v_n} \log \mathbb{P}(V_n \leq r) = -\infty.$$

Dans un souci d'éclaircir les idées, considérons une suite de variables aléatoires réelles $(X_n)_n$ centrées et intéressons nous à la suite $T_n = \frac{1}{n}\sum_{i=1}^n X_i / \left(\frac{1}{n}\sum_{i=1}^n X_i^p\right)^{1/p}$ $(p > 1)$. L'étude d'un PGD portant sur $(T_n)_n$ est liée aux propriétés asymptotiques de

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \geq r\left(\frac{1}{n}\sum_{i=1}^n X_i^p\right)^{1/p}\right) (r > 0) \text{ et } \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \leq r\left(\frac{1}{n}\sum_{i=1}^n X_i^p\right)^{1/p}\right) (r < 0).$$

L'obtention d'un PGD partiel relatif à la classe de Chernov

$$\left\{A \text{ borélien de } \mathbb{R}\times\mathbb{R}_+ \mid \lim_{R\to\infty} \inf\left\{\frac{s}{v^{1/p}} \mid (s,v) \in A,\ v > R\right\}\right\}$$

permet d'étudier les quantités précédentes et d'obtenir des résultats de grandes déviations pour $r \neq 0$. C'est l'objet de la partie 4.2.3.2 "taux troué en 0".

Pour passer à un vrai PGD pour la suite $(T_n)_n$, il est nécessaire de contourner l'obstacle "0/0", c'est à dire de s'assurer que, asymptotiquement, (S_n, V_n) est "assez loin de $(0,0)$. C'est ce que l'on étudie dans la partie 4.2.3.3 "PGD troué en 0 pour le quotient".

Voici maintenant la technique. On peut dès à présent voir le corollaire 4.2 qui en est un type d'application naturel ; les exemples qui suivent ce corollaire éclairent le problème des trous en 0.

4.2.3.2 Taux troué en 0

♣ Taux troué unilatéral

Pour $\rho > 0$, on note $A_\rho = \{(s,v) \mid U(s) > \rho v\}$. C'est un ouvert de $\mathcal{S}(U)$ dont la fermeture est $\overline{A_\rho} = \{(s,v) \mid\ U(s) \geq \rho v\}$; $(s,v) \in \overline{A_\rho}$ équivaut à $s \times [0,v] \in \overline{A_\rho}$. On a l'égalité suivante :

$$h^*_{f,V}\left(\overline{A_\rho}\right) = h_{f,V}\left(\overline{A_\rho}\right).$$

Soient les fonctions α et $\overline{\alpha}$ définies par

$$\alpha(\rho) \stackrel{\triangle}{=} h_{f,V}(A_\rho) \text{ et } \overline{\alpha}(\rho) \stackrel{\triangle}{=} h_{f,V}\left(\overline{A_\rho}\right).$$

Ce sont des fonctions croissantes et l'on a

$$\alpha(\rho) = \overline{\alpha}(\rho_+).$$

On introduit H un taux vague croissant égal à la régularisée s.c.i. de ces deux fonctions

$$H(\rho) \stackrel{\triangle}{=} \alpha(\rho_-) = \overline{\alpha}(\rho_-).$$

Remarque 4.1 *Supposons que le taux J est étroit et que le couple (f, V) est bien équilibré. Alors, la fonction $\overline{\alpha}$ est continue à gauche et $H(\rho) = h_{f,V}(A_\rho) = h_{f,V}(\overline{A}_\rho)$. En outre, $h_{f,V}(\overline{A}_\rho) = J(\nu_\rho)$ pour une probabilité ν_ρ telle que $\int V d\nu_\rho < \infty$ $U\left(\int f d\nu_\rho\right) \geq \rho \int V d\nu_\rho$.*

Si par ailleurs $\left(\int f dL_n, \int V dL_n\right)_n$ vérifie aussi (par exemple, pour les chaînes de Markov, dans le cadre des théorèmes 3.6 et 3.7) un PGD inférieur de même vitesse et de taux $h^{sci}_{f,V}$,

$$\begin{aligned} -H(\rho_+) \leq -h_{f,V}(A_\rho) = -h^{sci}_{f,V}(A_\rho) &\leq \liminf \frac{1}{v_n} \log \mathbb{P}\left(\left(\int f dL_n, \int V dL_n\right) \in A_\rho\right) \\ &\leq \limsup \frac{1}{v_n} \log \mathbb{P}\left(\left(\int f dL_n, \int V dL_n\right) \in \overline{A_\rho}\right) \\ &\leq -h_{f,V}\left(\overline{A_\rho}\right) = -H(\rho). \end{aligned}$$

L'ensemble $\mathcal{D}_H$ des discontinuités de la fonction H est au plus dénombrable. Selon le lemme 1.4, si $\rho \notin \mathcal{D}_H$, il existe $(x_\rho, v_\rho) \in \overline{A}_\rho$ tel que

$$h^*_{f,V}(x_\rho, v_\rho) = h_{f,V}(\overline{A}_\rho);$$

donc, selon la proposition 4.2, il existe une probabilité ν_ρ telle que $\int V d\nu_\rho < \infty$, $U\left(\int f d\nu_\rho\right) \geq \rho \int V d\nu_\rho$ et $H(\rho) = J(\nu_\rho)$.

♠ **Taux troué en** 0

Supposons que $B = \mathbb{R}$ et que $U(s) = \gamma(|s|)$ où $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ est continue, strictement croissante et convexe. Pour $r > 0$, on définit

$$R_r^+ = \left\{(s, v) \mid \ s > \gamma^{-1}(rv)\right\},$$

$$\alpha^+(r) = h_{f,V}(R_r^+) \text{ et } \overline{\alpha}^+(r) = h_{f,V}\left(\overline{R_r^+}\right) = h^*_{f,V}\left(\overline{R_r^+}\right);$$

et on considère la fonction H^+ croissante s.c.i. sur $]0, \infty[$: $H^+(r) = \alpha^+(r_-) = \overline{\alpha}^+(r_-)$. Pour $r < 0$, on définit

$$R_r^+ = \left\{(s, v) \mid \ s < -\gamma^{-1}(-rv)\right\},$$

$$\alpha^-(r) = h_{f,V}(R_r^-) \text{ et } \overline{\alpha}^-(r) = h_{f,V}\left(\overline{R_r^-}\right) = h^*_{f,V}\left(\overline{R_r^-}\right);$$

et on considère la fonction H^- décroissante s.c.i. sur $]-\infty,0[$: $H^-(r) = \alpha^-(r_+) = \overline{\alpha}^-(r_+)$.

Lorsque (f,V) est un couple équilibré pour lequel on doit utiliser le taux approché $h^{*app}_{f,V}$, on a encore

$$h^{*app}_{f,V}(\overline{A_\rho}) = h^{app}_{f,V}(A_\rho) = \alpha^{app}(\rho) \text{ et } h^{*app}_{f,V}(\overline{R_r^\pm}) = h^{app}_{f,V}(\overline{R_r^\pm}) = \alpha^{app\pm}(r)$$

Les énoncés suivants utilisent alors

$$H(\rho) = \alpha^{app}(\rho_-),\ H^+(r) = \alpha^{app+}(r_-) \text{ et } H^-(r) = \alpha^{app-}(r_+)$$

Proposition 4.3 (Chernov troué en 0**)**

1) *On se place dans le cadre de l'axiome 4.2 et on considère la fonction réelle H définie ci-dessus sur $]0,\infty[$.*

1a) *H est une fonction croissante s.c.i et, pour tout $\rho > 0$,*

$$\limsup_n \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right) \leq -H(\rho) = -H([\rho,\infty[).$$

1b) *Si, en outre, le couple (f,V) est bien équilibré et la suite $(\int f dL_n, \int V dL_n)_n$ vérifie un PGD inférieur de même vitesse et de taux régularisé s.c.i. $h^{sci}_{f,V}$ alors, pour tout $\rho > 0$,*

$$\begin{aligned} -H(]\rho,\infty[) &\leq \liminf_n \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right) \\ &\leq \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right) \leq -H\left([\rho,\infty[\right). \end{aligned}$$

2) *On suppose que $B = \mathbb{R}$ et que γ est une fonction positive, convexe et strictement croissante sur $\mathbb{R}_+$.*

2a) *Pour la fonction H^+ positive, croissante et s.c.i sur $]0,\infty[$ et la fonction H^- positive, décroissante et s.c.i. sur $]-\infty,0[$ définies ci-dessus, on a*

$$\begin{aligned} \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(\int f dL_n \geq \gamma^{-1}\left(r \int V dL_n\right)\right) &\leq -H^+\left([r,\infty[\right) \ \forall r > 0, \\ \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(\int f dL_n \leq -\gamma^{-1}\left(-r \int V dL_n\right)\right) &\leq -H^-\left(]-\infty,r]\right) \ \forall r < 0. \end{aligned}$$

2b) *Si, en outre, le couple est bien équilibré et la suite $(\int f dL_n, \int V dL_n)_n$ vérifie un PGD inférieur de même vitesse et de taux $h^*_{f,V}$ alors, pour tout $r > 0$,*

$$\begin{aligned} -H^+(]r,\infty[) &\leq \liminf_n \frac{1}{v_n} \log \mathbb{P}\left(\int f dL_n \geq \gamma^{-1}\left(r \int V dL_n\right)\right) \\ &\leq \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(\int f dL_n \geq \gamma^{-1}\left(r \int V dL_n\right)\right) \leq -H^+\left([r,\infty[\right); \end{aligned}$$

$$\begin{aligned}-H^-(]-\infty,r[) &\leq \liminf_n \frac{1}{v_n}\log\mathbb{P}\left(\int f dL_n \leq -\gamma^{-1}\left(-r\int V dL_n\right)\right)\\ &\leq \limsup_n \frac{1}{v_n}\log\mathbb{P}\left(\int f dL_n \leq -\gamma^{-1}\left(-r\int V dL_n\right)\right) \leq -H^-\left(]-\infty,r]\right).\end{aligned}$$

3) *On suppose que le couple (f,V) est bien équilibré.*

3a) *Dans le cadre de 1a) lorsque J est étroit ou dans le cadre de 1b) si J est vague, il existe en tout point ρ où H est continue, une probabilité ν_ρ telle que $\int V d\nu_\rho < \infty$, $U\left(\int f d\nu_\rho\right) \geq \rho \int V d\nu_\rho$ et $H(\rho) = J(\nu_\rho)$.*

3b *Dans le cadre de 2a) lorsque J est étroit ou dans le cadre de 2b) si J est vague, en tout point r où $H^\pm$ (suivant que r soit strictement positif ou strictement négatif) est continue, il existe des probabilités $\nu_r^\pm$ telles que $\int V d\nu_r^\pm < \infty$, $\int f d\nu_r^+ \geq \gamma^{-1}\left(r \int V d\nu_r^+\right)$, $\int f d\nu_r^- \leq -\gamma^{-1}\left(-r\int V d\nu_r^-\right)$ et $H^\pm(\pm r) = J\left(\nu_r^\pm\right)$.*

Remarque 4.2 *Dans la partie 3), lorsque J est un taux étroit, les propriétés énoncées sont valables pour tout $\rho > 0$ et tout $r \neq 0$.*

$\diamondsuit$ Où le taux s'annule-t-il ?

On se place dans les cadres décrits dans la partie 3) de la proposition 4.3. La fonction H étant continue à gauche, pour $\rho > 0$, $H(\rho) = 0$ équivaut à $H(\rho_-) = 0$ donc à la propriété

$$\forall a < \rho,\ \exists \nu_a,\ J(\nu_a) = 0,\ \int V d\nu_a < \infty \text{ et } U(\int f d\nu_a) \geq a \int V d\nu_a.$$

De même, $H^+(r) = 0$ pour $r > 0$ (resp. $H^-(r) = 0$ pour $r < 0$) équivaut à

$$\forall a \in]0,r[,\ \exists \nu_a,\ J(\nu_a) = 0,\ \int V d\nu_a < \infty \text{ et } \int f d\nu_a \geq \gamma^{-1}\left(a\int V d\nu_a\right).$$

$$\left(resp. \forall a \in]r,0[,\ \exists \nu_a,\ J(\nu_a) = 0,\ \int V d\nu_a < \infty \text{ et } \int f d\nu_a \leq -\gamma^{-1}\left(-a\int V d\nu_a\right).\right)$$

Nous rencontrerons surtout les deux situations suivantes.

– **J s'annule pour une probabilité μ unique.**
  Alors, si $\int V d\mu < \infty$, H s'annule sur $]0, U(\int f d\mu)/\int V d\mu]$. Si, en outre, $\int f d\mu > 0$, H^+ s'annule sur $]0, \gamma(\int f d\mu)/\int V d\mu]$ et si $\int f d\mu < 0$, H^- s'annule sur $[-\gamma(-\int f d\mu)/\int V d\mu, 0[$. Ailleurs, ces fonctions sont strictement positives.
  Si $\int V d\mu = \infty$, les fonctions H, H^+ et H^- sont toujours strictement positives.

– **J ne s'annule pour aucune probabilité.**
  Alors, les fonctions H, H^+ et H^- sont toujours strictement positives.

4.2.3.3 PGD troué en 0 pour le quotient

Supposons U de Lyapounov, convexe et strictement positive sauf en 0 où elle est nulle (typiquement, pour $B = \mathbb{R}^d$, $U(x) = \|x\|^p$, $(p \geq 1)$) et l'on se place dans le cadre du PGD autonormalisé par U c'est à dire que $V = U \circ f$ ou plus généralement que $U \circ f \leq MV$, M constante strictement positive. Pour toute probabilité ν, $U\left(\int f d\nu\right) \leq \int U \circ f d\nu \leq M \int V d\nu$. Par conséquent, tous les $\overline{A_\rho}$ tels que $\rho > 0$ contiennent $(0,0)$.

Pour l'étude de $h_{f,V}(0,0)$, on remarque que, si ν est une probabilité telle que $U \circ f$ soit intégrable, $\int f d\nu = 0$ et $\int U \circ f d\nu = 0$, la mesure image $f(\nu)$ est égale à la mesure de Dirac δ_0. Donc $h_{f,V}(0,0) = J(\{\nu; f(\nu) = \delta_0\})$.
On va distinguer plusieurs cas de figure, suivant la valeur de $h_{f,V}(0,0)$.

Cas pathologique. Supposons $h_{f,V}(0,0) = 0$; alors on est dans une situation où la proposition 4.3 est totalement sans intérêt car les majorations sont alors triviales. En effet,

$$H(\rho) = H^{\pm}(\pm r) = 0 \text{ pour tous } \rho > 0 \text{ et } r > 0.$$

Cas intermédiaire. Si $\infty > h_{f,V}(0,0) > 0$, cela donne un majorant un peu désagréable, valable pour tous $\rho > 0$ et $r > 0$ de $H(\rho)$ et $H^{\pm}(\pm r)$.
Considérons

$$\begin{aligned} T(s,v) &= \frac{U(s)}{v} \text{ si } (s,v) \neq (0,0) \\ &= 0 \text{ sinon,} \end{aligned}$$

ou, plus simplement $T(s,v) = \frac{U(s)}{v}$ avec la convention $0/0 = 0$.

Pour $\delta > 0$, $\Gamma_\delta = \{(s,v); U(s) \leq M\delta \text{ et } v \leq \delta\}$ est un compact de $B \times \mathbb{R}_+$. Pour tous $\rho > 0$,

$$\mathbb{P}\left(T\left(\int f dL_n, \int V dL_n\right) \geq \rho\right) \leq \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right)$$

Il reste le problème de la majoration. Pour tous $\rho > 0$ et $\delta > 0$,

$$\begin{aligned} &\mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right) \\ \leq\ &\mathbb{P}\left(\left\{U\left(\int f dL_n\right) \geq \rho \int V dL_n\right\} \cap \left\{\left(\int f dL_n, \int V dL_n\right) \notin \Gamma_\delta\right\}\right) \\ &+\mathbb{P}\left(\left(\int f dL_n, \int V dL_n\right) \in \Gamma_\delta\right) \\ \leq\ &\mathbb{P}\left(T\left(\int f dL_n, \int V dL_n\right) \geq \rho\right) + \mathbb{P}\left(\left(\int f dL_n, \int V dL_n\right) \in \Gamma_\delta\right). \end{aligned}$$

Grâce aux PGD vague et partiel supérieurs satisfaits par $(\int f dL_n, \int V dL_n)$,

$$\begin{aligned} &\limsup_n \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right) \\ \leq\ &\sup \left\{ \begin{array}{c} \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(T\left(\int f dL_n, \int V dL_n\right) \geq \rho\right), \\ \limsup_n \frac{1}{v_n} \log \mathbb{P}\left(\left(\int f dL_n, \int V dL_n\right) \in \Gamma_\delta\right) \end{array} \right\}, \end{aligned}$$

et, faisant tendre δ vers 0,

$$\begin{aligned} &\limsup \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right) \\ \leq\ &\sup\left\{\limsup_n \frac{1}{v_n} \log \mathbb{P}\left(T\left(\int f dL_n, \int V dL_n\right) \geq \rho\right), -h_{f,V}(0,0)\right\}. \end{aligned}$$

De plus,

$$\begin{aligned}
& \liminf \frac{1}{v_n} \log \mathbb{P}\left(T\left(\int f dL_n, \int V dL_n\right) \geq \rho\right) \\
\leq\ & \liminf \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right) \\
\leq\ & \sup \left\{ \begin{array}{c} \liminf \frac{1}{v_n} \log \mathbb{P}\left(T\left(\int f dL_n, \int V dL_n\right) \geq \rho\right), \\ \limsup \frac{1}{v_n} \log \mathbb{P}\left(\left(\int f dL_n, \int V dL_n\right) \in \Gamma_\delta\right) \end{array} \right\};
\end{aligned}$$

$$\begin{aligned}
& \liminf \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right) \\
\leq\ & \sup \left(\liminf \frac{1}{v_n} \log \mathbb{P}\left(T\left(\int f dL_n, \int V dL_n\right) \geq \rho\right), -h_{f,V}(0,0)\right).
\end{aligned}$$

Cas favorable. Si $h_{f,V}(0,0) = J(\{\nu; f(\nu) = \delta_0\}) = +\infty$, on obtient pour $\rho > 0$,

$$\begin{aligned}
\limsup \frac{1}{v_n} \log \mathbb{P}\left(\frac{U\left(\int f dL_n\right)}{\int V dL_n} \geq \rho\right) &= \limsup \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right), \\
\liminf \frac{1}{v_n} \log \mathbb{P}\left(U\left(\int f dL_n\right) \geq \rho \int V dL_n\right) &= \liminf \frac{1}{v_n} \log \mathbb{P}\left(\frac{U\left(\int f dL_n\right)}{\int V dL_n} \geq \rho\right).
\end{aligned}$$

De plus, dans ce cas,

$$H(\rho) = \inf\left\{J(\nu); U\left(\int f d\nu\right) \geq \rho \int V d\nu\right\} = \inf\left\{J(\nu) \mid \frac{U\left(\int f d\nu\right)}{\int V d\nu} \geq \rho\right\}$$

et, par la formule de Jensen, pour toute probabilité ν, $U\left(\int f d\nu\right) \leq \int U \circ f d\nu$; ainsi, pour $\rho > M$, $H(\rho) = \infty$.

Dans ce cas, la suite réelle $\left(\frac{U(\int f dL_n)}{\int V dL_n}\right)$ est bornée par M donc exponentiellement tendue et elle satisfait, pour tout $\rho > 0$,

$$\limsup \frac{1}{v_n} \log \mathbb{P}\left(\frac{U\left(\int f dL_n\right)}{\int V dL_n} \geq \rho\right) \leq -H([\rho, \infty[);$$

c'est évidemment encore vrai pour $\rho = 0$ en prenant $H(0) = 0$. Les mêmes raisonnements valent encore pour $H^{\pm}$. D'où la proposition suivante.

Proposition 4.4 (PGD troué en 0) *On se place dans le cadre de l'axiome 4.2, avec une fonction de Lyapounov U convexe, strictement positive partout sauf en 0 et $U \circ f \leq MV$ (pour une constante strictement positive M). On suppose en outre que $J\{\nu; f(\nu) = \delta_0\} = \infty$; cette condition sera désignée par "cas favorable".*

1) 1a) *La suite réelle autonormalisée $\left(\frac{U(\int f dL_n)}{\int V dL_n}\right)$ satisfait un PGD supérieur de taux vague K, où K est définie par*

$$\begin{aligned}
K(\rho) &= H(\rho) \text{ si } \rho > 0, \\
K(\rho) &= 0 \text{ si } \rho = 0. \\
K(\rho) &= +\infty \text{ si } \rho < 0.
\end{aligned}$$

1b) *Si le couple* (f, V) *est exponentiellement bien équilibré et si* $(U(\int f dL_n), \int V dL_n)$ *satisfait aussi un PGD inférieur de taux* $h^{sci}_{f,V}$ *alors* $\left(\frac{U(\int f dL_n)}{\int V dL_n}\right)$ *satisfait un PGD étroit de taux étroit* K.

2) *Si* $B = \mathbb{R}$ *et si* $U(x) = \gamma(|x|)$, *où* γ *est continue, positive, convexe et strictement croissante de* $\mathbb{R}_+$ *dans* $\mathbb{R}_+$, *alors*

2a) *la suite autonormalisée* $\left(\int f dL_n / \gamma^{-1}(\int V dL_n)\right)_n$ *vérifie un PGD supérieur de taux vague* S *défini par*

$$\begin{aligned} S(r) &= H^+(r) \text{ si } r > 0 \\ &= H^-(r) \text{ si } r < 0 \\ &= 0 \text{ si } r = 0. \end{aligned}$$

2b) *Si le couple* (f, V) *est exponentiellement bien équilibré et si* $\left(\int f dL_n /, \gamma^{-1}(\int V dL_n)\right)_n$ *satisfait aussi un PGD inférieur de taux* $h^{sci}_{f,V}$, *alors* $\left(\int f dL_n / \gamma^{-1}(\int V dL_n)\right)_n$ *satisfait un PGD de taux étroit* S.

Remarque 4.3 *Dans le cas particulier sur* $\mathbb{R}$ *où* $U(s) = |s|$, *si le couple est bien équilibré, les taux* H^+ *et* H^- *sont convexes si* J *est convexe.*

4.2.4 Preuves des résultats de 4.2.1

Preuve de la proposition 4.1

Montrons que $h^*_{f,V}$ est égale à sa régularisée s.c.i. c'est à dire que $h^*_{f,V} \leq h^{*sci}_{f,V}$. Ici

$$h^{*sci}_{f,V}(x) = \lim_{r \to 0} \uparrow h_{f,V}\left(\overline{B}(x, r) \times [0, R + r]\right).$$

L'ensemble $\left\{\nu \mid \int V d\nu \leq R + r,\ \int f d\nu \in \overline{B}(x, r)\right\}$ est compact donc il existe une probabilité ν_r telle que

$$\begin{aligned} h_{f,V}\left(\overline{B}(x, r) \times [0, R + r]\right) &= J\left(\left\{\nu \mid \int V d\nu \leq R + r,\ \int f d\nu \in \overline{B}(x, r)\right\}\right) = J(\nu_r) \\ \int f d\nu_r &= y_r \in \overline{B}(x, r), \ \int V d\nu_r \leq R + r; \\ h^{*sci}_{f,V}(x) &= \lim_{r \to 0} J(\nu_r). \end{aligned}$$

Comme V est une fonction de Lyapounov s.c.i. et f est uniformément intégrable sur $\{\nu \mid \int V d\nu \leq R + 1\}$ (qui contient tous les (ν_r) pour $r \leq 1$), il existe une valeur d'adhérence ν de (ν_r) pour laquelle :

$$J(\nu) \leq \liminf_{r \to 0} J(\nu_r) = \lim_{r \to 0} J(\nu_r), \ \int f d\nu = \lim y_r = x, \ \int V d\nu \leq R.$$

Donc $h^*_{f,V}(x) \leq J(\nu) \leq h^{*,sci}_{f,V}(x)$; $h^*_{f,V}$ est un taux vague.

$\{\nu \mid \int f d\nu = x,\ \int V d\nu \leq R\}$ est compact ; donc $h^*_{f,V}(x, R) = J(\nu_{x,R})$ pour une probabilité $\nu_{x,R}$ appartenant à $\{\nu \mid \int f d\nu = x,\ \int V d\nu \leq R\}$.

Lorsque J est convexe, pour deux éléments (x,R) et (y,ρ) de $B\times\mathbb{R}_+$ et $t\in]0,1[$,

$$J\left(t\nu_{x,R}+(1-t)\,\nu_{y,\rho}\right)\leq th^*_{f,V}\left(x,R\right)+(1-t)\,h^*_{f,V}\left(y,\rho\right).$$

Mais $t\nu_{x,R}+(1-t)\,\nu_{y,\rho}\in\left\{\nu\mid\int fd\nu=tx+(1-t)\,y,\ \int Vd\nu\leq tR+(1-t)\,\rho\right\}$;

$$h^*_{f,V}\left(tx+(1-t)\,y,tR+(1-t)\,\rho\right)\leq th^*_{f,V}\left(x,R\right)+(1-t)\,h^*_{f,V}\left(y,\rho\right).$$

et J est convexe.

Montrons le PGD vague supérieur. Soit Γ un compact de $B\times\mathbb{R}_+$; $\{y\mid(x,y)\in\Gamma\}$ est un ensemble borné de $\mathbb{R}_+$ et il existe donc un $R<\infty$ tel que

$$H_\Gamma=\left\{\nu\mid\left(\int fd\nu,\int Vd\nu\right)\in\Gamma\right\}\subset\left\{\nu\mid\int Vd\nu\leq R\right\}.$$

Ainsi H_Γ est une partie relativement compacte de $\mathcal{P}\left(E\right)$, d'où :

$$\limsup\frac{1}{n}\log\mathbb{P}\left(\left(\int fdL_n,\int VdL_n\right)\in\Gamma\right)=\limsup\frac{1}{n}\log\mathbb{P}\left(L_n\in H_\Gamma\right)\leq-J\left(\overline{H_\Gamma}\right)$$

avec $h^*_{f,V}\left(\Gamma\right)=J\left(H_\Gamma\right)\geq J\left(\overline{H_\Gamma}\right)$. On aura achevé la preuve si l'on montre $J\left(H_\Gamma\right)=J\left(\overline{H_\Gamma}\right)$, c'est à dire que, pour tout élément ν de $\overline{H_\Gamma}$, $J\left(\nu\right)\geq h^*_{f,V}\left(\Gamma\right)=J\left(H_\Gamma\right).$ Considérons en effet $\nu\in\overline{H_\Gamma}$ et une suite (ν_n) de H_Γ qui converge vers ν et $(x_n,y_n)=\left(\int fd\nu_n,\int Vd\nu_n\right)\in\Gamma$; pour une certaine suite extraite $\left(x_{a(n)},y_{a(n)}\right)$ tend vers $(x,y)\in\Gamma$ et $\left(\nu_{a(n)}\right)$ tend vers ν. Donc $\int Vd\nu\leq R$ et, le couple étant équilibré, $\int fd\nu=x$. Ainsi $J\left(\nu\right)\geq h^*_{f,V}\left(x,y\right)\geq h^*_{f,V}\left(\Gamma\right).$ ◊

Preuve de la proposition 4.2

1) Preuve de la partie 1) de la proposition.
Montrons que les épigraphes du taux approché sont fermés (ou que leurs complémentaires sont ouverts). Notons h^*_a au lieu de $h^*_{f_a,V}$ et h_a au lieu de $h_{f_a,V}$. Si $h^{*app}_{f,V}\left(x,R\right)>\alpha$, il existe un r tel que

$$\limsup_{a\to\infty}h_a\left(\overline{B}\left(x,r\right)\times[0,R+r]\right)>\alpha.$$

Pour $y\in B\left(x,r/2\right)$, $B\left(y,r/2\right)\subset\overline{B}\left(x,r\right)$ et pour $t\leq R+r/2$,

$$\begin{aligned}\limsup_{a\to\infty}h_a\left(B\left(y,r/2\right)\times[0,t+r/2]\right)&\geq\limsup_{a\to\infty}h_a\left(B\left(x,r\right)\times[0,R+r]\right)>\alpha;\\ h^*_a\left(y,t\right)&>\alpha.\end{aligned}$$

Ainsi $B\left(x,r/2\right)\times[0,R+r/2]\subset\left\{h^{*app}_{f,V}>\alpha\right\}$.

Montrons le PGD vague supérieur, en notant $A_{n,r}=\left\{\int VdL_n\leq R+r\right\}$. Pour tout $r>0$,

$$\begin{aligned}
&\mathbb{P}\left(\left\{\int fdL_n \in \overline{B}(x,r)\right\} \cap A_{n,r}\right) \\
\leq\ &\mathbb{P}\left(\left\{\int f_a dL_n \in \overline{B}(x,2r)\right\} \cap A_{n,r}\right) + \mathbb{P}\left(\| \int (f-f_a)\, dL_n\| > r\right); \\
&\limsup \frac{1}{v_n} \log \mathbb{P}\left(\left\{\int fdL_n \in \overline{B}(x,r)\right\} \cap A_{n,r}\right), \\
\leq\ &\sup \begin{pmatrix} \limsup \frac{1}{v_n} \log \mathbb{P}\left(\left\{\int f_a dL_n \in \overline{B}(x,2r)\right\} \cap A_{n,r}\right), \\ \limsup \frac{1}{v_n} \log \mathbb{P}\left(\| \int (f-f_a)\, dL_n\| > r\right) \end{pmatrix}, \\
\leq\ &\sup \left\{ -\inf \left\{ h_a\left(\overline{B}(x,2r) \times [0,R+2r]\right)\right\},\ \limsup \frac{1}{n} \log \mathbb{P}\left(\| \int (f-f_a)\, dL_n\| > r\right)\right\}.
\end{aligned}$$

Ceci étant vrai pour tout a, on obtient en faisant tendre a vers l'infini,

$$\limsup \frac{1}{v_n} \log \mathbb{P}\left(\int fdL_n \in \overline{B}(x,r),\ \int VdL_n \leq R+r\right) \leq -\lim \sup_{a\to\infty} h_a\left(\overline{B}(x,2r) \times [0,R+2r]\right),$$

d'où

$$\lim_{r\to 0} \downarrow \limsup \frac{1}{v_n} \log \mathbb{P}\left(\left\{\int fdL_n \in \overline{B}(x,r),\int VdL_n \leq R+r\right\}\right) \leq -h^{*app}(x,R).$$

Cela conduit à un PGD vague de taux h^{*app}.

2) Preuve de la partie 2) de la proposition

Soit $A_r = \{\nu \mid \int Vd\nu \leq R+r\}$. On a, pour tout $r>0$,

$$h_a\left(\overline{B}(x,r) \times [0,R+r]\right) = J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,r)\right\} \cap A_r\right).$$

Les ensembles $A_r \cap \{\nu \mid J(\nu) \leq \gamma\}$ sont des compacts sur lesquels la fonction $\nu \to \int fd\nu$ est continue.

Montrons $h^{*app}(x,R) \geq h^*(x,R)$. Il n'y a rien à prouver si $h^{*app}(x) = \infty$. Supposons donc $h^{*app}(x) < \gamma$. Il existe un $r_0 > 0$ tel que, pour tout $r \leq r_0$, on ait $\limsup_a h_a\left(\overline{B}(x,r) \times [0,R+r]\right) < \gamma$.

Cela veut dire que, pour tout $r \leq r_0$,

$$\begin{aligned}
\gamma &> \limsup_a J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,r)\right\} \cap A_r\right) \\
&= \limsup_a J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,r)\right\} \cap A_r \cap \{\nu \mid J(\nu) \leq \gamma\}\right);
\end{aligned}$$

et, pour $a \geq a_0$,

$$\begin{aligned}
h\left(\overline{B}(x,2r) \times [0,R+r]\right) &\leq J\left(\left\{\nu \mid \int fd\nu \in \overline{B}(x,2r)\right\} \cap A_r \cap \{\nu \mid J(\nu) \leq \gamma\}\right) \\
&\leq J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,r)\right\} \cap A_r \cap \{\nu \mid J(\nu) \leq \gamma\}\right), \\
h\left(\overline{B}(x,2r) \times [0,R+r]\right) &\leq \limsup_a h_a\left(\overline{B}(x,r)\right) < \gamma.
\end{aligned}$$

Mais, $\left\{\nu \mid \int f d\nu \in \overline{B}(x,2r)\right\} \cap A_r \cap \{\nu \mid J(\nu) \leq \gamma\}$ est compact, et il existe un ν_r dans cet ensemble tel que $J(\nu_r) = J\left(\left\{\nu \mid \int f d\nu \in \overline{B}(x,2r)\right\} \cap A_r \cap \{\nu \mid J(\nu) \leq \gamma\}\right)$, $\int f d\nu_r \in \overline{B}(x,2r)$ et $\int V d\nu_r \leq R + r$. Pour $r \leq r_0$, la suite (ν_r) est contenue dans $A_{r_0} \cap \{\nu \mid J(\nu) \leq \gamma\}$; il existe une suite (r_n) majorée par r_0 qui tend vers 0 et telle que (ν_{r_n}) converge étroitement vers ν; $\int f d\nu_r \to \int f d\nu = x$ et $\int V d\nu \leq R$. Donc :

$$h^*(x,R) \leq J(\nu) \leq \liminf J(\nu_{r_n}) \leq \limsup_{r\to 0} \liminf_a h_a\left(\overline{B}(x,r) \times [0, R+r]\right) \leq \gamma.$$

Ceci étant vrai dès que $h^{*app}(x,R) > \gamma$, on a bien $h^*(x,R) \leq h^{*app}(x,R)$.

Montrons que $h^{*app} \leq h^{*sci}$. On suppose $h^{*sci} < \gamma < \infty$;

$$\begin{aligned} h\left(\overline{B}(x,r) \times [0,R+r]\right) &= J\left(\left\{\nu \mid \int f d\nu \in \overline{B}(x,r)\right\} \cap A_r\right) \\ &= J\left(\left\{\nu \mid \int f d\nu \in \overline{B}(x,r)\right\} \cap A_r \cap \{\nu \mid J(\nu) \leq \gamma\}\right) \end{aligned}$$

et, pour $a \geq a_0$,

$$\begin{aligned} h\left(\overline{B}(x,r) \times [0,R+r]\right) &\geq J\left(\left\{\nu \mid \int f_a d\nu \in \overline{B}(x,2r)\right\} \cap A \cap \{\nu \mid J(\nu) \leq \gamma\}\right) \\ &\geq h_a\left(\overline{B}(x,2r) \times [0,R+2r]\right), \end{aligned}$$

ce qui permet de conclure.

Comme $h^* \leq h^{*app} \leq h^{*sci} \leq h^*$, on a fini. $\diamondsuit$

Preuve des résultats de la remarque 4.1

Prouvons que $\overline{\alpha}$ est continue à gauche. Soit (f,V) un couple exponentiellement bien équilibré et $(f_a)_{a>0}$ une bonne approximation exponentielle de f, telle que pour tout $a > 0$, (f_a, V) est un couple équilibré. Soit une suite croissante positive $(\rho_n)_n$ convergeant vers $\rho > 0$. On suppose que $\lim_n \overline{\alpha}(\rho_n) = a < \overline{\alpha}(\rho) \leq \infty$. Alors pour tout n, il existe ν_n tel que $U(\int f d\nu_n) \geq \rho_n \int V d\nu_n$ et $\overline{\alpha}(\rho_n) + \frac{1}{n} > J(\nu_n)$. Puisque J est un taux étroit, il existe une sous-suite de $(\nu_n)_n$ (que l'on continue à nommer $(\nu_n)_n$) convergente. On note ν sa limite.

$$\begin{aligned} \left|U\left(\int f d\nu\right) - U\left(\int f d\nu_n\right)\right| &\leq \left|U\left(\int f d\nu\right) - U\left(\int f_a d\nu\right)\right| + \left|U\left(\int f_a d\nu\right) - U\left(\int f_a d\nu_n\right)\right| \\ &\quad + \left|U\left(\int f_a d\nu_n\right) - U\left(\int f d\nu_n\right)\right|. \end{aligned}$$

La fonction U est continue. La bonne approximation

$$\lim_{a\to\infty} \sup\left\{\| \int (f - f_a) d\nu \| \mid \ J(\nu) \leq \gamma\right\} = 0,$$

implique donc $|U(\int f d\nu) - U(\int f d\nu_n)| \to 0$ quand $n \to \infty$. En effet, $(\nu_n)_n \subset \{J(\nu), \int V d\nu \leq R \leq \gamma\}$ pour R et γ bien choisis.

Par conséquent, on a $\lim_n \overline{\alpha}(\rho_n) \geq \liminf_n J(\nu_n) \geq J(\nu) \geq \overline{\alpha}(\rho)$, ce qui apporte une contradiction.

Prouvons que $h^*_{f,V}(\overline{A}_\rho) = J(\nu_\rho)$ pour une probabilité ν_ρ telle que $U\left(\int f d\nu_\rho\right) \geq \rho \int V d\nu_\rho$. Ce résultat est clair si $h^*_{f,V}(\overline{A}_\rho) = +\infty$. Si $h^*_{f,V}(\overline{A}_\rho) \leq a < +\infty$, le même preuve conduit à pouvoir affirmer que $\{\nu \in \mathcal{P}(E) \mid U\left(\int f d\nu\right) \geq \rho \int V d\nu\}$ est une partie fermée, d'où le résultat puisque J atteint son infimum sur tout fermé. $\diamondsuit$

4.3 PGD empirique étroit et PGD autonormalisé

4.3.1 Processus dont la loi empirique satisfait un PGD étroit

On étudie le cas de processus $(X_t)_{t\in T}$ (avec $T = \mathbb{N}$ ou $T = \mathbb{R}_+$) à valeurs dans un espace polonais (E, d) et dont la loi empirique $(L_t)_{t\in T}$ satisfait un PGD étroit supérieur de taux I et de vitesse $v : t \mapsto v_t$ (fonction de T dans $\mathbb{R}_+$ qui croît vers l'infini).
Lorsque le taux de ce PGD ne s'annule que pour une probabilité μ, cela implique la stabilité exponentielle du modèle (voir 2.4.2). Il s'agit donc d'une hypothèse d'ergodicité assez forte mais très générale.

Exemples de processus dont la loi empirique satisfait un PGD étroit. Voici divers cas où la loi empirique vérifie un PGD étroit.

- $E1$. La suite des lois empiriques d'une suite i.i.d. à valeurs dans E vérifie un PGD étroit de taux donné par l'information de Kullback (voir le théorème de Sanov 1.8).
- $E2$. La suite des lois empiriques $(L_n)_n$, $(L^1_n)_n$ ou $(L^2_n)_n$ d'une chaîne de Markov à valeurs dans (E, d) supposée presque fellérienne avec la condition de rappel exponentiel associée à (U, V) satisfait un PGD étroit (c'est le théorème 2.4).
- $E3$. La suite des lois empiriques d'un processus de Markov à temps continu à valeurs dans (E, d) satisfaisant un PGD étroit (voir Donsker et Varadhan [DonVar83], Deuschel et Stroock [DeuStr89]).
- $E4$. Soit $(X_t)_{t\in T}$ un processus gaussien réel stationnaire de densité spectrale continue et tendant vers 0 à l'infini. La suite des lois empiriques d'un tel processus satisfait un PGD étroit de taux convexe (voir [BaxJai93] pour $T = \mathbb{N}$ et [BryDem95] pour $T = \mathbb{R}_+$).

On cherche des critères permettant d'obtenir la tension exponentielle partielle de cette suite. Une clef exhibée en 1.4.4 pour l'obtention de la tension exponentielle est une condition de troncature exponentielle qui sera toujours satisfaite ici.

Troncature exponentielle. Dans le cadre précédent, on a la propriété de troncature très générale suivante.

Lemme 4.1 *Sur un espace polonais, on considère une famille de probabilités aléatoires $(L_t)_{t\in T}$ qui satisfait un PGD étroit supérieur de taux J et de vitesse $(v_t)_t$. Soit ϕ une fonction continue de E dans $[0, \infty[$. Elle satisfait toujours la condition de troncature* *exponentielle*

$$\forall r > 0, \lim_{a\to\infty} \limsup_t \frac{1}{v_t} \log \mathbb{P}\left(L_t(\{x \mid \phi(x) \geq a\}) \geq r\right) = -\infty.$$

Preuve On fixe $r > 0$. Comme $\{x \mid \phi(x) \geq a\}$ est fermé, $\Gamma_a = \{\nu \mid \nu\left(\{x \mid \phi(x) \geq a\}\right) \geq r\}$ est un fermé de $\mathcal{P}(E)$, on a

$$\limsup_t \frac{1}{v_t} \log \mathbb{P}\left(L_t\left(\{x \mid \phi(x) \geq a\} \geq r\right)\right) \leq -J(\Gamma_a).$$

Mais, si a croît vers l'infini, quelle que soit ν, $\lim_{a\to\infty} \nu\left(\{x \mid \phi(x) \geq a\}\right) = 0$; les fermés Γ_a décroissent vers l'ensemble vide et $\lim_a J(\Gamma_a) = \infty$. $\diamondsuit$

4.3.2 Autonormalisation

Le lemme 4.1 mène à une application immédiate des corollaires 1.2 et 1.3 qui permettent de conclure à la tension exponentielle relative à une classe de Chernov convenable. Voici donc un cadre très général, où les résultats "autonormalisés" sont les exactes transpositions de ceux qu'obtenaient Dembo et Shao dans le cas i.i.d.

Axiome 4.3

- *Pour $T = \mathbb{N}$ ou $T = \mathbb{R}_+$, $(L_t)_{t\in T}$ est une famille de probabilités aléatoires définies sur un espace polonais $\mathcal{P}(E)$ satisfaisant un PGD étroit supérieur de taux étroit I et de vitesse $(v_t)_t$.*
- *f est une fonction continue de (E, d) dans un espace de Banach B.*
- *U est une fonction de Lyapounov définie sur B, convexe, continue, nulle en 0 et strictement positive ailleurs ; $h : [0, \infty] \to [0, \infty]$ est une fonction convexe, croissante et finie sur $[0, \infty[$ telle que, si t et z tendent vers l'infini,*

$$\frac{h(tz)}{th(z)} \to \infty.$$

En particulier, sur $\mathbb{R}^d$, on prendra souvent $h(\cdot) = (\cdot)^p$ et $U(\cdot) = \|\cdot\|$ $(p > 1)$.

Lemme 4.2 *Le couple $(f, h \circ U(f))$ est équilibré, dès que $h \circ U(f)$ est une fonction de Lyapounov.*

Preuve On a $\lim_{t\to\infty} h(t) = +\infty$; on va vérifier que $\frac{h(t)}{t} \to \infty$ quand $t \to \infty$. Soit $B > 0$ tel que $h(B+1) - h(B) > 0$. On pose $u = h(B)$, $v = h(B+1)$ et $a = v - u > 0$, $b = u - aB$. Pour tout $t \geq B + 1$ $h(t) \geq at + b$ par la convexité de h. Si on pose $L(t) = \frac{h(t)}{t}$ alors $\frac{L(t)}{L(t^{1/2})} = \frac{h(t)}{t^{1/2}h(t^{1/2})}$. Par conséquent, $L(t)$ tend vers l'infini car, si il existe une suite $(t_n)_n$ croissant vers l'infini et telle que $\limsup L(t_n) < \infty$, alors $\limsup_n \frac{L(t_n)}{L(t_n^{1/2})} \leq \frac{\limsup L(t_n)}{a} < \infty$, ce qui est contradictoire avec l'hypothèse faite sur h.
Donc, h est une fonction surlinéaire et le couple $(U(f), h \circ U(f))$ est équilibré, ce qui implique que f est uniformément intégrable par rapport à $\mathcal{H}_{\mathcal{R}} = \{\nu \mid \int h \circ U(f) d\nu \leq R\}$ car f est intégrable par rapport à tout $\nu \in \mathcal{H}_R$ et elle est uniformément intégrable selon

la condition de troncature par U introduite en 3.1.3. Finalement, $(f, h\circ U(f))$ est un couple équilibré. $\diamondsuit$

On rappelle que la classe de Chernov relative à $(h \circ U)$ est définie par

$$\mathcal{S}(h\circ U) = \left\{ A \text{ borélien de } B\times\mathbb{R}_+ \mid \lim_{R\to\infty} \inf \left\{ \frac{h\circ U(s)}{v} \mid (s,v)\in A,\, v>R \right\} > 0 \right\};$$

la fonction $h^*_{f,V}$ désigne le taux image partiel de (f, V) par I (définition 4.1).

Théorème 4.1 (Transfert d'un PGD empirique étroit en un PGD autonormalisé)

Dans le cadre de l'axiome 4.3, on pose $V = h\circ U(f)$ et l'on suppose que V est de Lyapounov. Alors le couple (f, V) est équilibré et $(\int f dL_t, \int V dL_t)_t$ satisfait un PGD partiel supérieur de vitesse $(v_t)_t$ et de taux vague $h^*_{f,V}$ relativement à la classe de Chernov $\mathcal{S}(h\circ U)$.

En appliquant les propositions 4.3, 4.4 et les remarques 4.2. et 4.3, on obtient le corollaire suivant.

Corollaire 4.2 *Soit f une fonction continue de E dans $\mathbb{R}^d$ telle que $\|f\|$ soit une fonction de Lyapounov sur (E, d). Alors, pour tout $p > 1$,*

1) *La suite $(\int f dL_t, \int \|f\|^p dL_t)_t$ vérifie un PGD partiel supérieur de vitesse $(v_t)_t$ et de taux vague convexe $h^*_{f,\|f\|^p}$ relativement à $\mathcal{S}(\|\cdot\|^p)$.*
2) *En particulier, on a les propriétés suivantes.*

2*a*) *Pour tout $\rho > 0$,*

$$\limsup_t \frac{1}{v_t} \log \mathbb{P}\left(\left\| \int f dL_t \right\|^p \geq \rho \int \|f\|^p dL_t \right) \leq -h_{f,\|f\|^p}\left(\{(s,v) \mid \|s\|^p \geq \rho v\}\right) = -H(\rho),$$

où H est croissante s.c.i. sur $]0,\infty[$. De plus, pour tout ρ sur lequel H est continue, il existe une probabilité ν_ρ telle que $H(\rho) = J(\nu_\rho)$ et $U(\int f d\nu_\rho) \geq \rho \int V d\nu_\rho$.

2*b*) *Lorsque $d = 1$, on a*

$$\begin{aligned}
\forall r>0,\ \limsup_t \frac{1}{v_t}\log\mathbb{P}\left(\int f dL_t \geq r\left(\int |f|^p dL_t\right)^{1/p}\right) &\leq -S(r),\\
\forall r<0,\ \limsup_t \frac{1}{v_t}\log\mathbb{P}\left(\int f dL_t \leq r\left(\int |f|^p dL_t\right)^{1/p}\right) &\leq -S(r).
\end{aligned}$$

avec S défini ainsi

$$\begin{aligned}
S(r) &= H^+(r) = h_{f,V}\left(\left\{(s,v) \mid s \geq r v^{1/p}\right\}\right) \text{ si } r>0,\\
&= H^-(r) = h_{f,V}\left(\left\{(s,v) \mid s \leq r v^{1/p}\right\}\right) \text{ si } r<0,\\
&= 0 \text{ si } r=0,
\end{aligned}$$

où H^+ est convexe positive, croissante et s.c.i. sur $]0,+\infty[$, H^- est convexe positive, décroissante et s.c.i. sur $]-\infty,0[$. De plus, en tout point r où $H^\pm$(suivant que r soit strictement positif ou strictement négatif) est continue, il existe des probabilités $\nu_r^\pm$ telles que $H^\pm(\pm r) = J(\nu_r^\pm)$ avec $\int f d\nu_r^+ \geq \gamma^{-1}(r\int V d\nu_r^+)$ et $\int f d\nu_r^- \leq -\gamma^{-1}(-r\int V d\nu_r^-)$.

3) *Dans le cas favorable où* $I(\{\nu \mid f(\nu) = \delta_0\}) = \infty$, *on a les PGD suivants.*

3*a*) *Si* $d \geq 1$, *la suite* $\left(\frac{\| \int f dL_t \|^p}{\int \|f\|^p dL_t}\right)_t$ *satisfait un PGD supérieur de taux vague* K *défini de la même manière que dans la proposition 4.4.*

3*b*) *Si* $d = 1$, *la suite* $\left(\frac{\int f dL_t}{(\int |f|^p dL_t)^{1/p}}\right)_t$ *satisfait un PGD supérieur de taux vague* S *défini de la même manière que dans la proposition 4.4.*

4) *Dans le cas favorable, lorsque* I *s'annule pour une probabilité* μ *unique, deux cas se présentent.*

4*a*) *Si* $\int \|f\|^p d\mu < \infty$, *alors* $H(\rho) = 0$ *pour tout* $\rho \leq \| \int f d\mu\|^p / \int \|f\|^p d\mu$ *et* $H(\rho) > 0$ *si* $\rho > \| \int f d\mu\|^p / \int \|f\|^p d\mu$.
Si $d = 1$, $S(r) > 0$ *si* r *est positif et* $r > \int f d\mu / \left(\int |f|^p d\mu\right)$ *ou si* r *est négatif et* $r < \int f d\mu / \left(\int |f|^p d\mu\right)$.
En particulier, si $\int f d\mu = 0$, $H(\rho) > 0$ *pour tout* $\rho > 0$ *et* $S(r) > 0$ *pour tout* r *non nul.*

4*b*) *Si* $\int \|f\|^p d\mu = \infty$, $H(\rho) > 0$ *pour tout* $\rho > 0$ *et* $S(r) > 0$ *pour tout* r *non nul.*

Remarque 4.4 *La partie* 4*a*) *ressemble à un théorème de Chernov ;* 4*b*) *met en valeur l'autonormalisation.*

Corollaire 4.3 *Dans le cadre du corollaire 4.2, si l'on fait l'hypothèse que* $(\int f dL_t, \int V dL_t)_t$ *satisfait un PGD inférieur de taux* $h^{sci}_{f,V}$, *alors les PGD supérieurs du point 3) sont des PGD avec les taux étroits* K *ou* S *suivant le cas.*

4.3.3 Quelques exemples

Reprenant le cadre de Shao décrit en 1.4.1, on considère une suite $(X_n)_n$ de variables aléatoires à valeurs dans $\mathbb{R}^d$ et l'on pose, pour $p > 1$,

$$S_n = \sum_{i=1}^n X_i, \quad V_{n,p} = \left(\sum_{i=1}^n \|X_i\|^p\right)^{1/p} \; ; \quad R_{n,p} = \frac{S_n}{n^{1-1/p} V_{n,p}}, \quad T_{n,p} = \frac{\|S_n\|^p}{n^{p-1} V_{n,p}^p}$$

avec la convention $0/0 = 0$.

♣ **Cas i.i.d. de loi** μ**.** Ici, $(X_n)_n$ est une suite i.i.d. de loi μ et les corollaires 4.2 et 4.3 s'appliquent à la fonction identité $f(x) = x$. On a $K(\delta_0 \mid \mu) = -\log \mu(0)$ et $K(\cdot \mid \mu)$ s'annule seulement en μ. On retrouve ainsi le résultat énoncé par Shao dans [Sha97] où la condition $\mu(0) = 0$ avait été, par erreur, oubliée.
Dans le cas favorable où $\mu(0) = 0$, on a les résultats suivants.

a) Si $d \geq 1$, la suite $(T_{n,p})_n$ satisfait un PGD étroit de taux étroit K.

b) Si $d = 1$, $(R_{n,p})_n$ satisfait un PGD étroit de taux étroit S.

Dans le cas intermédiaire où $0 < \mu(0) < 1$, on n'a plus de PGD étroit pour le quotient. Le PGD inférieur est maintenu mais le PGD supérieur est remplacé par un taux tronqué par $K(\delta_0 \mid \mu) = -\log(\mu(0))$.

♠ **Cas markovien de probabilité de transition p presque fellérienne.** Le taux associé aux lois empiriques est ici $I^1(\nu) = \inf \left\{ K\left(\zeta \mid \zeta^1 \oslash p\right) \mid \zeta \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d),\, \zeta^1 = \zeta^2 = \nu \right\}$. La condition $\zeta^1 = \zeta^2 = \delta_0$ signifie $\zeta = \delta_{0,0}$ et

$$K\left(\delta_{0,0} \mid \delta_0 \oslash p\right) = K\left(\delta_0 \mid p(0,\cdot)\right) = -\log p(0,0).$$

Si p est presque fellérienne avec la condition de rappel exponentiel et si $p(0,0) = 0$, on a un PGD supérieur pour la suite $(T_{n,p})_n$ et, si $d = 1$, pour $(R_{n,p})_n$.

Sous la condition d'irréductibilité renforcée (hypothèse 3.1 de la section 3.5.4), le corollaire 4.3 s'applique et les suites autonormalisées $(T_{n,p})_n$ et $(R_{n,p})_n$ satisfont, pour toute loi initiale, un PGD étroit avec les mêmes propriétés que dans le cas i.i.d.
Ce qui précède s'applique aussi à $S_n(f) = \sum_{i=1}^n f(X_i)$ et $V_{n,p}(f) = (\sum_{i=1}^n \|f(X_i)\|^p)^{1/p}$ pour une fonction $f : \mathbb{R}^d \to \mathbb{R}^d$ ne s'annulant qu'en 0, ce qui signifie que $I^1\left(\{\nu \mid\ f(\nu) = \delta_0\}\right) = I^1(\delta_0)$.

♢ **Diffusion du gradient.** Soient V une fonction de Lyapounov de classe $\mathcal{C}^2$ sur $\mathbb{R}^d$ et $(B_t)_{t\geq 0}$ un mouvement brownien sur $\mathbb{R}^d$. On considère la diffusion définie par

$$dX_t = -\nabla V(X_t)dt + dB_t.$$

Sous les conditions suivantes où D est une constante strictement inférieure à 2,

$$\Delta V \leq D(1 + \|\nabla V\|^2), \quad \limsup_{\|x\|\to\infty} \|\nabla V\|/V < \infty,$$

cette diffusion est récurrente et son unique loi stationnaire est la probabilité de Gibbs G_V associée à V dont la densité par rapport à la mesure de Lebesgue est $x \mapsto e^{-2V(x)}/\int e^{-2V(y)}dy$. Alors, pour toute loi initiale, $L_t = \frac{1}{t}\int_0^t \delta_{X_s} ds$ satisfait un PGD de taux étroit J défini par

$$\begin{aligned} J(\nu) &= \frac{1}{2}\int \|\nabla\phi\|^2 dG_V \quad \text{si } \nu = \phi.G_V \text{ et si } \phi \text{ est différentiable dans } \mathcal{L}^2 \\ &= \infty \quad \text{si } \nu \text{ n'est pas absolument continue par rapport à } G_V. \end{aligned}$$

Dès lors, la quantité $J(\delta_0)$ est toujours infinie et, pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}^d$ ne s'annulant qu'en 0, les résultats du corollaire 4.2 s'appliquent à la suite autonormalisée

$$T_{n,p} = \frac{\|\int_0^t f(X_s)ds\|^p}{t^{p-1}\int_0^t \|f(X_s)\|^p ds}$$

ou, si $d = 1$, à

$$R_{n,p} = \frac{\int_0^t f(X_s)ds}{t^{1-1/p}\left(\int_0^t \|f(X_s)\|^p ds\right)^{1/p}}.$$

4.4 PGD autonormalisés pour des modèles autorégressifs

Sans hypothèse de stabilité, la section précédente ne s'applique plus. Néanmoins, il est naturel de tenter d'autonormaliser des martingales. C'est ce que l'on fait ici pour des modèles autorégressifs.

4.4.1 Suite régressive à bruit dont le carré vérifie la condition de cramér renforcée

Sur un espace de probabilité $(\Omega, \mathcal{A}, P)$, on considère le modèle markovien autorégressif fonctionnel à valeurs dans $\mathbb{R}^d$ suivant :

$$X_{n+1} = A(X_n) + \sigma(X_n)\varepsilon_{n+1}$$

où A est continue de $\mathbb{R}^d$ dans $\mathbb{R}^d$, σ est une fonction continue de $\mathbb{R}^d$ dans un intervalle $[1/\beta, \beta]$ avec $1 < \beta < \infty$ et $(\varepsilon_n)_{n\geq 1}$ est une suite i.i.d. de vecteurs aléatoires de dimension d supposée indépendante de X_0; la loi de X_0 est notée α. On rappelle que, si p est la probabilité de transition de cette chaîne de Markov, $p(x, \bullet)$ est la loi de $A(x) + \sigma(x)\varepsilon_1$; c'est une probabilité de transition fellérienne.
Soient $\phi : \mathbb{R}^d \to \mathbb{R}^d$ continue telle que $\|\phi\|^2$ soit une fonction de Lyapounov et la fonction f définie par :

$$f(x,y) =< \phi(x), (y - A(x))/\sigma(x) >,$$
$$\|f(x,y)\| \leq \beta\|\phi(x)\|\|y - A(x)\|.$$

On suppose qu'il existe une constante $0 < M < \infty$ telle que $\|x\| \leq \|\phi(x)\|$, $\|A(x)\| \leq M\|\phi(x)\|$ ce qui implique $\|f(x,y)\| \leq \beta(2+M)\left(\|\phi(x)\| + \|\phi(y)\|\right)^2$.
Comme en 3.4.3, on considère la martingale (M_n) et le processus croissant (C_n) définis par

$$\begin{aligned} M_n &= \sum_{i=1}^{n} < \phi(X_{i-1}), \varepsilon_i > , \\ C_n &= \frac{1}{2}\left(\|\phi(X_0)\|^2 + \|\phi(X_n)\|^2\right) + \sum_{i=1}^{n-1} \|\phi(X_i)\|^2. \end{aligned}$$

Selon le théorème 3.5, on a (sans aucune condition de rappel) les PGD suivants, où I est le taux de Donsker-Varadhan de la chaîne de Markov.

a) Si $\|\varepsilon_1\|^2$ satisfait la condition de Cramér renforcée, $\left(\frac{M_n}{n}, \frac{C_n}{n}\right)$ satisfait un PGD vague supérieur de taux

$$h^*(z,R) = \inf\left\{ I(\zeta) ; \int < \phi(x), (y - A(x)) > \zeta(dx,dy) = z \text{ et } \int \|\phi\|^2 d\zeta^1 \leq R. \right\}.$$

b) Si $\int \|\phi\|^2 d\zeta^1 = 0$ et si $I(\zeta) < \infty$, $\zeta^1 = \zeta^2 = \delta_0$ et $I(\zeta) = K(\delta_{0,0} \mid \delta_0 \oslash p(0,\cdot))$. Par suite, on est dans le cas favorable où $h^*(0,0) = \infty$ si $p(0,0) = 0$
Lorsque $\|\phi\|^2$ ne s'annule qu'en 0, on est dans le cas favorable où $h(0,0) = \infty$ (voir en 4.2.3.) car $\int \|\phi\|^2 d\zeta^1 = 0$ et $I(\zeta) < \infty$ signifient $\zeta^1 = \zeta^2 = \delta_0$ donc $\zeta = \delta_{0,0}$ qui n'est pas absolument continue par rapport à $\delta_0 \oslash p = \delta_0 \oslash p(0, \bullet)$. On a aussi $h(0,0) = \infty$ dans la partie a) si $p(0,0) = 0$.

Pour en déduire des PGD autonormalisés, il reste à obtenir des conditions de tension exponentielle partielle.

On cherche à obtenir la tension exponentielle de la suite $(M_n/n, V_n/n)_n$ relativement à $\mathcal{S}(|\cdot|)$. Si l'on note $K_r = B(0,r) \times [0,r]$, Cela revient, selon le lemme 1.3, à établir que, pour tout $\rho > 0$,

$$\lim_{r\to\infty} \limsup_n \frac{1}{n} \log \sup_\alpha \mathbb{P}_\alpha \left(\left(\frac{M_n}{n}, \frac{C_n}{n} \right) \in A_\rho \cap K_r^c \right) = -\infty$$

avec $A_\rho = \{(m, c) \in \mathbb{R} \times \mathbb{R}_+ \mid \frac{m}{c} > \rho\}$. Or :

$$\left\{ \left(\frac{M_n}{n}, \frac{C_n}{n} \right) \in A_\rho \cap K_r^c \right\} \subset \left\{ \frac{|M_n|}{n} > \rho \frac{C_n}{n} > \rho r \right\} \cup \left\{ \frac{|M_n|}{n} > r,\ \frac{C_n}{n} \leq r \right\};$$

$$\begin{aligned} \left| \frac{M_n}{n} \right| &\leq \left(\frac{1}{n} \sum_{i=1}^n \|\phi(X_{i-1})\|^2 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2 \right)^{1/2} \\ &\leq \left(2\frac{C_n}{n} \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2 \right)^{1/2}. \end{aligned}$$

Par conséquent, si $\left| \frac{M_n}{n} \right| > \rho \frac{C_n}{n} > \rho r$, on a

$$\rho \left(\frac{C_n}{2n} \right)^{1/2} \leq \left(\frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2 \right)^{1/2} \quad \text{et} \quad \frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2 \geq \frac{\rho^2 r}{2};$$

si $\frac{|M_n|}{n} > r,\ \frac{C_n}{n} \leq r$,

$$r < (2r)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2 \right)^{1/2} \quad \text{donc } r/2 < \frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2.$$

Finalement, on obtient

$$\left\{ \left(\frac{M_n}{n}, \frac{C_n}{n} \right) \in A_\rho \cap K_r^c \right\} \subset \left\{ \frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2 > \inf\left(\rho^2/2, 1/2\right) r \right\}.$$

En particulier, si $\|\varepsilon_1\|^2$ vérifie la condition de Cramér, la suite $\left(\frac{M_n}{n}, \frac{C_n}{n} \right)_n$ est exponentiellement tendue relativement à $\mathcal{S}(|\cdot|)$. On a prouvé le lemme suivant.

Lemme 4.3 *Si le carré du bruit vérifie la condition de Cramér,* $\left(\frac{M_n}{n}, \frac{C_n}{n} \right)$ *est exponentiellement tendue par rapport à la classe* $\mathcal{S}(|.|)$.

On se contente de traduire dans le théorème suivant la proposition 4.3. La convexité provient de la remarque 4.2.

Théorème 4.2 (PGD supérieur autonormalisé) *Pour le modèle autorégressif défini ci-dessus, on fait sur le bruit les hypothèses suivantes :*

- $A(0) + \sigma(0)\varepsilon_1 \neq 0$ *presque-sûrement.*
- *Le carré du bruit vérifie la condition de Cramér renforcée.*

On suppose en outre que ϕ *ne s'annule qu'en* 0. *Alors, la suite* $\left(\frac{M_n}{C_n} \right)$ *satisfait un PGD étroit supérieur de vitesse* n *et de taux nul en* 0, *décroissant sur* $]-\infty, 0[$ *et croissant sur* $]0, \infty[$.

Pour obtenir un PGD inférieur avec le même taux, il faut une bonne approximation et le PGD inférieur ce qui conduit au théorème suivant

Théorème 4.3 (PGD autonormalisé) *En addition aux hypothèses du théorème 4.2, on suppose que le bruit a une densité g strictement positive et que l'on se trouve dans un des deux cadres de la section 3.5.4 : récurrence du modèle ou domination de la probabilité de transition (exemple 3.6).*
Alors, pour toute loi initiale, la suite $\left(\frac{M_n}{C_n}\right)$ *satisfait un PGD étroit de vitesse* n *et de taux convexe et décroissant sur* $]-\infty, 0[$*, convexe et croissant sur* $]0, \infty[$.

Remarque 4.5 *Au lieu de* C_n *on aurait pu utiliser* $D_n = \sum_{i=1}^{n} \|\phi(X_i)\|^2$ *qui a les mêmes propriétés que* C_n *si on se limite à des familles de lois initiales concentrées sur une partie bornée de* $\mathbb{R}^d$.

4.4.2 Application aux modèles AR(1) linéaires

♣ Présentation des estimateurs usuels

Dans cette section, nous considérons un modèle autorégressif d'ordre 1 unidimensionnel :

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}$$

où (ε_n) est une suite i.i.d de v.a. centrées non nulles, indépendante de l'état initial X_0.
On considère les estimateurs suivants.

- l'**estimateur des moindres carrés**:

$$\overset{\triangle}{\theta}_n = \frac{\sum_{i=0}^{n-1} X_i X_{i+1}}{\sum_{i=0}^{n-1} X_i^2} = \theta + \frac{\sum_{i=0}^{n-1} X_i \varepsilon_{i+1}}{\sum_{i=1}^{n} X_i^2};$$

- **l'estimateur de Yule-Walker :**

$$\overset{\triangledown}{\theta}_n = \frac{\sum_{i=0}^{n-1} X_i X_{i+1}}{\sum_{i=1}^{n} X_i^2} = \frac{\sum_{i=0}^{n-1} X_i^2}{\sum_{i=1}^{n} X_i^2}\theta + \frac{\sum_{i=0}^{n-1} X_i \varepsilon_{i+1}}{\sum_{i=1}^{n} X_i^2};$$

- **l'estimateur intermédiaire**, qui n'est pas classique, mais qui apparaîtra naturellement dans notre étude :

$$\overset{=}{\theta}_n = \frac{\sum_{i=0}^{n-1} X_i X_{i+1}}{\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=0}^{n-1} X_i^2\right)/2} = \frac{\sum_{i=0}^{n-1} X_i^2}{\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=0}^{n-1} X_i^2\right)/2}\theta + \frac{\sum_{i=0}^{n-1} X_i \varepsilon_{i+1}}{\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=0}^{n-1} X_i^2\right)/2}.$$

Pour l'étude de la consistence, le premier estimateur est le plus simple car, quel que soit θ, par la loi des grands nombres relative aux martingales de carré intégrables, $\overset{\triangle}{\theta}_n - \theta \xrightarrow{p.s.} 0$. L'estimateur des moindres carrés $\overset{\triangle}{\theta}_n$ est donc toujours un estimateur fortement consistant de θ.
Pour les deux autres estimateurs, on a

$$\overset{\triangledown}{\theta}_n - \frac{\sum_{i=0}^{n-1} X_i^2}{\sum_{i=1}^{n} X_i^2}\theta \xrightarrow{p.s.} 0 \text{ et } \overset{=}{\theta}_n - \frac{\sum_{i=0}^{n-1} X_i^2}{\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=0}^{n-1} X_i^2\right)/2}\theta \xrightarrow{p.s.} 0.$$

Dans le cas stable ($|\theta| < 1$), on a $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p.s.} \gamma > 0$.
Dans le cas instable ($\theta = \pm 1$), en supposant que le bruit a un moment d'ordre > 2 fini, on a (cf. par exemple [Duf97], proposition 4.4.24) :

$$\frac{X_n^2}{\sum_{i=1}^n X_i^2} \xrightarrow{p.s.} 0.$$

D'où la consistance forte des trois estimateurs dans les cas stable et instable.
Dans le cas explosif ($|\theta| > 1$), pour une v.a. presque sûrement non nulle Y dont la loi ne charge aucun point (en particulier pas 0),

$$\theta^{-n} X_n \xrightarrow{p.s.} Y = X_0 + \sum_{k=1}^{\infty} \theta^{-k} \varepsilon_k,$$

La convergence est un résultat facile de martingale et la propriété de sa loi est une conséquence d'un théorème de P. Lévy ([Lév31]).
Dans le cas explosif, les deux derniers estimateurs sont des estimateurs fortement consistants d'une fonction simple de θ à partir de laquelle on peut calculer θ; en effet :

$$\theta^{-2n} \sum_{i=1}^n X_i^2 \xrightarrow{p.s.} \frac{\theta^2}{1-\theta^2} Y^2;$$
$$\overset{\triangledown}{\theta}_n \xrightarrow{p.s.} \frac{1}{\theta} \text{ et } \overset{=}{\theta}_n \xrightarrow{p.s.} \frac{2}{\theta + 1/\theta}.$$

Tout cela est fort ancien (Mann-Wald [ManWal43]). Le comportement en loi de ces estimateurs est lui aussi bien connu. Par contre, seul le cas des bruits gaussien avait jusqu'à présent été étudié du point de vue des grandes déviations.

♠ L'état des lieux pour un modèle AR(1) à bruit gaussien

On suppose que le bruit a la loi $\mathcal{N}(0, \sigma^2)$, gaussienne centrée de variance σ^2.

Cas stable

1) *Utilisation de méthodes gaussiennes* En corollaire de méthodes gaussiennes, un PGD est établi par Bryc et Smolenski [BrySmo93] et par Bryc et Dembo [BryDem97] pour les moyennes quadratiques $\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)$.
Parallèlement, Bercu-Gamboa-Rouault [BerGamRou97], eux aussi à partir de résultats gaussiens plus généraux, obtiennent des PGD étroits portant sur le modèle autorégressifs stationnaires - avec la loi initiale $\mathcal{N}\left(0, \sigma^2/\left(1-\theta^2\right)\right)$.

– Pour l'estimateur des moindres carrés avec le taux non convexe :

$$\begin{aligned} h(x) &= \frac{1}{2}\log\frac{1+\theta^2-2\theta x}{1-x^2} \text{ si } \left(\theta - \sqrt{\theta^2+8}\right)/4 \le x \le \left(\theta + \sqrt{\theta^2+8}\right)/4 \\ &= \log|\theta - 2x|. \end{aligned}$$

– Pour l'estimateur de Yule-Walker avec le taux convexe :

$$\begin{aligned} H(x) &= \frac{1}{2}\log\frac{1+\theta^2-2\theta x}{1-x^2} \text{ si } -1 < x < 1, \\ &= \infty \text{ sinon.} \end{aligned}$$

Dans le même cadre, des PGD "précis" sont obtenus par Bercu, Gamboa et Lavielle [BerGamLav00].
Citons aussi des travaux voisins de Florens et Pham portant sur le processus de Ornstein-Uhlenbeck [FloPha96b].
2) *Principes de moyennes déviations (PDM).* En corollaire d'un principe de moyennes déviations relatif aux martingales, Worms ([Wor99a] et [Wor99b]) obtient ici un principe de moyennes déviations pour l'estimateur des moindres carrés. Pour une vitesse (a_n) telle que $a_n = o(n)$, $\left(\sqrt{n/a_n}\left[\overset{\triangle}{\theta}_n - \theta\right]\right)$ satisfait un PGD de vitesse (a_n) et de taux $x \to (x-\theta)^2/2\sigma^2$.

Cas stable, instable et explosif

Deux articles très récents ont présenté des résultats assez surprenants, conduisant à un comportement en grandes déviations des estimateurs de même type dans les trois cas (avec toujours la vitesse n). Ces travaux ont été une motivation des nôtres, visant à obtenir des résultats semblables avec des bruits non gaussiens.
1) *PGD et PDM supérieurs pour* $\left(\overset{\triangle}{\theta}_n - \theta\right)$.
J. Worms [Wor01] utilise une inégalité exponentielle relative aux martingales et des calculs spécifiques au cas gaussien du même type que ceux des auteurs cités ci-dessus et obtient, dans les cas instable et explosif des PGD supérieurs de vitesse n. Il obtient aussi des PDM de vitesse (a_n) avec $a_n = o(n)$, pour $\left(\left|\theta^n\left[\overset{\triangle}{\theta}_n - \theta\right]^{1/a_n}\right|\right)$ dans le cas explosif et pour $\left((n/a_n)\left[\overset{\triangle}{\theta}_n - \theta\right]\right)$ dans le cas instable.
2) *PGD pour l'estimateur de Yule-Walker.*
B. Bercu [Ber01] obtient des PGD (précis) de vitesse n pour l'estimateur de Yule-Walker. Pour le modèle $AR(1)$ avec un bruit de loi $\mathcal{N}(0,1)$ et d'état initial x_0 donné, $\left(\overset{\triangledown}{\theta}_n\right)$ satisfait toujours un PGD de vitesse n, avec les taux suivants :

- dans le cas stable, le taux convexe donné ci-dessus avec la loi stationnaire ;
- dans le cas explosif, un taux convexe sur $]-\infty, 1/\theta[$ et sur $]1/\theta, +\infty[$ mais brutalement discontinu en $1/\theta$:

$$\begin{aligned} H(x) &= \frac{1}{2}\log\left(\frac{1+\theta^2-2\theta x}{1-x^2}\right) \text{ si } x \in]-1,1[\text{ et } x \neq 1/\theta, \\ &= 0 \text{ si } x = 1/\theta, \\ &= +\infty \text{ si } |x| \geq 1; \end{aligned}$$

- dans le cas instable, un taux convexe H^+ si $\theta = 1$ et H^- si $\theta = -1$ avec :

$$\begin{aligned} H^+(x) &= \frac{1}{2}\log\left(\frac{2}{1+x}\right) \text{ si } -1 < x \leq 1, \\ &= +\infty \text{ sinon}; \\ H^-(x) &= \frac{1}{2}\log\left(\frac{2}{1-x}\right) \text{ si } -1 \leq x < 1, \\ &= +\infty \text{ sinon}. \end{aligned}$$

♠ PGD autonormalisé

Voici des résultats qui sont des conséquences des théorèmes 4.2 et 4.3. Dans ce cas, le taux de Donsker-Varadhan $I(\zeta)$ s'annule seulement pour $\zeta = \zeta^1 \oslash p$, avec ζ^1 invariante pour p. Donc jamais si $|\theta| \geq 1$; si $|\theta| < 1$, il s'annule pour une loi gaussienne centrée. Ainsi, pour $R < \infty$,

$$h^*(r, R) = \inf \left\{ I(\zeta) \mid \int x(y - \theta x)\zeta(dx, dy) = r, \int x^2 d\zeta^1 \leq R \right\}$$

ne s'annule que pour $r = 0$. Lorsque le carré du bruit vérifie la condition de Cramér renforcée, cela veut dire que les taux ne s'annulent qu'en 0.

Théorème 4.4 *Soit un modèle autorégressif d'ordre 1 unidimensionnel*

$$X_{n+1} = \theta X_n + \varepsilon_{n+1}$$

ayant un bruit dont le carré satisfait la condition de Cramér renforcée ; en outre $\mathbb{P}(\varepsilon_1 = 0) = 0$*. On ne fait aucune hypothèse sur* θ*.*

– *Alors, la suite* $\left(\overset{=}{\theta}_n - \frac{\sum_{i=0}^{n-1} X_i^2}{\left(\sum_{i=1}^n X_i^2 + \sum_{i=0}^{n-1} X_i^2\right)/2}\theta\right)$ *satisfait un PGD supérieur étroit de vitesse* n *et de taux étroit* $\overset{=}{H}$ *nul en* 0*, convexe et décroissant sur* $]-\infty, 0[$*, convexe et croissant sur* $]0, \infty[$ *; ce PGD est uniforme pour toutes les lois initiales*
– *La suite* $\left(\overset{\triangledown}{\theta}_n - \frac{\sum_{i=0}^{n-1} X_i^2}{\sum_{i=1}^n X_i^2}\theta\right)$ *satisfait un PGD supérieur de taux* $\overset{\triangledown}{H}$ *ayant les mêmes propriétés que* $\overset{=}{H}$*, uniforme sur des classes de lois initiales concentrées sur un compact.*

Les taux $\overset{\triangledown}{H}$ *et* $\overset{=}{H}$ *sont convexes*

Théorème 4.5 *En plus des hypothèses effectuées dans le théorème 4.4, on suppose que le bruit a une densité strictement positive. Alors, pour toute loi initiale, la suite* $\left(\overset{=}{\theta}_n - \frac{\sum_{i=0}^{n-1} X_i^2}{\left(\sum_{i=1}^n X_i^2 + \sum_{i=0}^{n-1} X_i^2\right)/2}\theta\right)$ *satisfait un PGD étroit de vitesse* n *et de taux* $\overset{=}{H}$ *ayant les mêmes propriétés que dans le théorème 4.4.*

Les taux $\overset{\triangledown}{H}$ *et* $\overset{=}{H}$ *sont convexes et strictement positifs pour tout* $r \neq 0$*.*

Pour une loi initiale concentrée sur un compact, la suite $\left(\overset{\triangledown}{\theta}_n - \frac{\sum_{i=0}^{n-1} X_i^2}{\sum_{i=1}^n X_i^2}\theta\right)$ *vérifie un PGD étroit de vitesse* n *et de taux* $\overset{\triangledown}{H}$ *ayant les mêmes propriétés que dans le théorème 4.4.*

♢ Commentaires

$C1$. Nous n'avons pas rejoint tout à fait le cas gaussien pour lequel le carré du bruit ne satisfait qu'une condition de Cramer non renforcée.

$C2$. La méthode de B. Bercu est beaucoup plus précise dans le cas gaussien mais elle n'est pas transposable à d'autres bruits.

$C3$. La formulation donnée ci-dessus (valable pour tout θ) se prête bien, comme c'est souvent le cas pour les PGD autonormalisés, à tout problème de test ou d'intervalle de confiance sur θ; par exemple à un test "$\theta \leq 1$" contre "$\theta > 1$", c'est à dire stable-instable contre explosif.

4.4.3 Fonctionnelles additives martingales markoviennes

♣ Vitesses de convergence d'un théorème de limite centrale

1) *Schéma* Considérons une suite (M_n) de v.a. réelles satisfaisant un théorème de limite centrale : $(M_n/\sqrt{n})$ converge en loi vers une loi normale $\mathcal{N}(0,1)$ - gaussienne centrée de variance 1.
Un moyen d'évaluer la vitesse de cette convergence est donné par des **théorèmes de limite centrale presque sûre** qui donnent une convergence presque sûre du type suivant : sur $\mathcal{P}(\mathbb{R})$,

$$\frac{1}{s_n}\sum_{i=1}^{n}\frac{1}{i}\delta_{[M_i/\sqrt{i}]}\xrightarrow{p.s.}\mathcal{N}(0,1) \text{ avec } s_n=\sum_{i=1}^{n}\frac{1}{i}.$$

Dès lors, on peut tenter d'évaluer la vitesse de convergence de ce nouveau résultat asymptotique (en loi ou en grandes déviations).

Cas du mouvement brownien Pour les grandes déviations, la racine du problème est naturellement un PGD sur les lois empiriques d'un mouvement brownien (B_t). Selon Baxter et Jain [BaxJaiVar91], il n'existe pas de PGD de vitesse t pour les lois empiriques usuelles $\left(\frac{1}{t}\int_0^t \delta_{B_s}ds\right)$. Mais il en existe un pour les moyennes logarithmiques $\left(\frac{1}{\log t}\int_0^t \frac{1}{s}\delta_{[B_s/\sqrt{s}]}ds\right)$. On a alors un PGD sur $\mathcal{P}(\mathbb{R})$ de vitesse $\log t$ et de taux H défini par :

$$\begin{aligned} H(\mu) &= \frac{1}{2}\int \|\nabla\phi(x)\|^2 d\mathcal{N}(0,1) \text{ si } \mu=\phi^2.\mathcal{N}(0,1) \text{ et si } \phi \text{ est différentiable dans } L^2(\mathcal{N}(0,1)),\\ &= \infty \text{ sinon.}\end{aligned}$$

Ce résultat se déduit facilement d'un PGD de taux H et de vitesse t relatif au processus de Ornstein-Uhlenbeck stationnaire $\left(X_t=e^{-t/2}B_{e^t}\right)$. Ce PGD sur le processus de Ornstein-Uhlenbeck résulte des PGD sur les processus gaussiens (Donsker et Varadhan [DonVar85], Bryc et Dembo [BryDem95]) ou des PGD sur les processus hypermélangeants (Chiyonobu et Kusuoka [ChiKus88]). Un exposé complet et clair est celui de Deuschel et Stroock ([DeuStr89]). On y trouve en outre le résultat suivant :

$$K(\mu|\mathcal{N}(0,1))\leq 2H(\mu).$$

Cas i.i.d. et markovien L'objectif naturel était dès lors l'obtention d'un PGD de taux H et de vitesse $\log n$ pour les moyennes logarithmiques $\left(\frac{1}{\log s_n}\sum_{i=1}^n \frac{1}{i}\delta_{[M_i/\sqrt{i}]}\right)$.
La première réponse positive a été donnée par March et Seppäläinen [MarSep97] dans le cas d'une marche aléatoire $(Y_1+...+Y_n)_{n\geq 1}$ où les v.a. Y_i sont i.i.d., centrées, de variance 1 et ayant des moments de tous les ordres finis. On peut regarder aussi [Bro88], [CsoHor92], [ChaMaa00], [Cha01a], [Cha01b] [Maa01a] et [Maa01b].
Ce résultat a été généralisé au cadre markovien par Heck et Maaouia [HecMaa01]. Voici leur résultat.
On considère la version canonique $\left(\Omega,\mathcal{A},(\mathbb{P}_x)_{x\in E},(X_n)_{n\geq 0}\right)$ d'une chaîne de Markov de transition p, d'espace d'états (E,d) et, pour $\mathbb{P}_\nu=\int\nu(dx)\mathbb{P}_x$, il s'agit d'une chaîne de

Markov de loi initiale ν. Ici $(\Omega, \mathcal{A}) = (E, \mathcal{E})^{\mathbb{N}}$ et l'on utilise les opérateurs de translation θ_k définis par $\theta_k\left((y_n)_{n\geq 0}\right) = (y_{n+k})_{n\geq 0}$.
Une **fonctionnelle additive** $(A_n)_{n\geq 0}$ est une suite de v.a. réelles telle que $A_0 = 0$ et, pour tout $n \geq 1$ et toute loi initiale ν, on ait, $\mathbb{P}_\nu - p.s.$,

$$A_n = A_{n-1} + A_1 \circ \theta_{n-1}.$$

C'est une **fonctionnelle additive croissante** si $A_1 \geq 0$.
Une **fonctionnelle additive martingale** $(M_n)_{n\geq 0}$ est une fonctionnelle additive qui est en outre une martingale pour chacune des probabilités $\mathbb{P}_\nu$.

Théorème 4.6 ***(Heck et Maaouia [HecMaa01])***
Hypothèses

a) *La chaîne de Markov est supposée récurrente de probabilité invariante μ et irréductible avec un petit ensemble C (cf. 3.4.1.).*
Si $T_C = \inf\{n \mid X_n \in C$ et $n \geq 1\}$, on suppose enfin que $\sup_{x\in C} \mathbb{E}_x\left(T_C^k\right) < \infty$ pour tout $k \geq 1$.

b) *(M_n) est une fonctionnelle additive martingale telle que, pour tout $k \geq 1$, $\mathbb{E}_\mu\left(|M_n|^k\right) < \infty$ et $\mathbb{E}_\mu\left(M_n^2\right) = \sigma_M^2$.*

Conclusion *Pour toute loi initiale ν telle que, quel que soit $k \geq 1$, $\mathbb{E}_\nu\left(T_C^k\right) < \infty$ et $\mathbb{E}_\nu\left(|M_{T_C}|^k\right) < \infty$, la suite des probabilités aléatoires*

$$\frac{1}{s_n}\sum_{i=1}^{n}\frac{1}{i}\delta_{\left[M_i/\sigma_M\sqrt{i}\right]} \text{ avec } s_n = \sum_{i=1}^{n}\frac{1}{i}$$

satisfait sur $(\Omega, \mathcal{A}, \mathbb{P}_\nu)$ un PGD de taux H.

Commentaires

$C1$. Le taux obtenu est ainsi le même pour toute fonctionnelle additive markovienne normalisée par σ_M. En ce sens, la vitesse de convergence du théorème limite central ne dépend pas de la fonctionnelle additive markovienne.

$C2$. La condition $\sup_{x\in C} \mathbb{E}_x\left(T_C^k\right)$ donnée en a) implique, en notant $\|\bullet\|_{VT}$ la distance en variation totale des probabilités et b une constante :

$$\forall n,\ n^k \sup_{x\in C}\|p^n(x,\bullet) - \mu\|_{VT} < \infty;$$

c'est la **récurrence riemanienne d'ordre** k. Elle implique $\mathbb{E}_\mu\left(T_C^{k-1}\right) < \infty$. Voir Pitman [Pit74] dans le cas atomique, Nummelin et Tuominen [NumTuo82], Duflo ([Duf97] p.292-302).
Donc, dans le théorème de Heck-Maaouia, on peut prendre $\nu = \mu$ ou $\nu = \delta_x$ pour $\mu-$presque tout x.
Comme $\mathbb{E}_\mu\left(|M_n|^k\right) < \infty$ pour tout $k \geq 1$, le théorème s'applique à la loi initiale μ et aux lois initiales δ_x pour $\mu-$presque tout x. Il s'applique aussi à toute mesure ν absolument continue par rapport à μ avec une densité bornée.

$C3$. Des conditions suffisantes peuvent être établies à l'aide du **critère de Hajek** [Haj82]. On suppose qu'il existe une fonction U, deux constantes R et A finies, et un $a \in]0,1[$ tels que

$$pU(x) \leq aU(x)\mathbb{I}_{U(x)>R} + A\mathbb{I}_{U(x)\leq R}.$$

Alors, pour $b \in]1, 1/a[$ et pour une constante γ_b,

$$\mathbb{E}_x\left(b^{T_{\{U\leq R\}}}\right) \leq \gamma_b\left(U(x)+1\right).$$

Ainsi, si $C = \{U \leq R\}$ est un petit ensemble, on a un cadre assurant $\sup_{x\in C}\mathbb{E}_x\left(T_C^k\right)$.

Pour toute loi initiale ν telle que $\int Ud\nu < \infty$, $\mathbb{E}_\nu\left(b^{T_C}\right) < \infty$ donc, pour tout k, $\mathbb{E}_\nu\left(T_C^k\right) < \infty$.

Notons $\Delta M_n = M_n - M_{n-1}$. Par l'inégalité de Burkholder, pour tout $k > 2$, il existe une constante ρ_k pour laquelle

$$\begin{aligned}
\mathbb{E}_\nu\left(\sup_{i\leq n}|M_i|^k\right) &\leq \rho_k\mathbb{E}_\nu\left(\left[\sum_{i=1}^n(\Delta M_i)^2\right]^{k/2}\right)\\
&\leq \rho_k n^{k/2-1}\mathbb{E}_\nu\left(\sum_{k=1}^n|\Delta M_i|^k\right);\\
\mathbb{E}_\nu\left(|M_{T_C}|^k\right) &\leq \gamma_b\left(\int Ud\nu+1\right)\rho_k\sum_n b^{-n}n^{k/2-1}\left(\sum_{k=1}^n\mathbb{E}_\nu\left(|\Delta M_i|^k\right)\right).
\end{aligned}$$

Le théorème s'appliquera à la loi initiale ν si $\int Ud\nu < \infty$ et si, pour tout $k \geq 2$, $\sup_n\mathbb{E}_\nu\left(|\Delta M_n|^k\right) < \infty$.

Exemple 4.1 ***(Modèle autorégressif récurrent)*** *On considère un modèle autorégressif de dimension d*

$$X_{n+1} = A(X_n) + \sigma(X_n)\varepsilon_{n+1}$$

avec les mêmes hypothèses que dans 3.4.3 à l'exception de la continuité de A et σ, inutile ici. On fait les hypothèses suivantes :

$a)$ Conditions sur le bruit. *Le bruit (ε_n) est une suite i.i.d. de v.a. dont la loi est centrée et a une densité par rapport à la mesure de Lebesgue continue, strictement positive ; enfin, pour tout k, $\mathbb{E}\left(\|\varepsilon_1\|^k\right) < \infty$.*

$b)$ Condition de stabilité : $\limsup_{\|x\|\to\infty}\left(\|A(x)\|/\|x\|\right) < 1$.

Le commentaire C3) s'applique à $U(x) = \|x\|^k$ pour tout $k \geq 1$. En outre tout compact de $\mathbb{R}^d$ est un petit ensemble. On prend un petit ensemble $C = \{x; \|x\| \leq R\}$. On obtient, pour un $b > 1$,

$$\mathbb{E}_x\left(b^{T_C}\right) \leq C_b(\|x\|+1) \;\; et \;\; \sup_{x\in C}\mathbb{E}_x\left(b^{T_C}\right) \leq C_b(R+1).$$

En outre, si $\int\|x\|^k\nu(dx) < \infty$, la condition de stabilité implique $\sup_n\mathbb{E}_\nu\left(\|X_n\|^k\right) < \infty$ (voir par exemple [Duf97], lemme 2.3.16).

Bilan. *On considère une loi initiale telle que* $\int \|x\|^k \nu(dx) < \infty$ *pour tout* $k \geq 1$ *et une fonction additive martingale*

$$M_n = \sum_{i=1}^{n} < \phi(X_{i-1}), \varepsilon_i >$$

où ϕ *est une fonction réelle borélienne définie sur* $\mathbb{R}^d$ *avec* $\|\phi(x)\| \leq \alpha\|x\|^m + \beta$ *pour un* $m \geq 1$*, alors*

$$\begin{aligned} \mathbb{E}_\nu\left(|\Delta M_n|^k\right) &= \left(\alpha\mathbb{E}_\nu\left(\|X_{n-1}\|^{mk}\right) + \beta\right)\mathbb{E}\left(\|\varepsilon_1\|^k\right), \\ \sup_n \mathbb{E}_\nu\left(|\Delta M_n|^k\right) &< \infty. \end{aligned}$$

Donc ici, le PGD logarithmique de Heck et Maaouia s'applique pour toute loi initiale satisfaisant $\int \|x\|^k \nu(dx) < \infty$.

$\diamondsuit$ Autonormalisation

Dans le cadre du théorème précédent on a les résultats suivants où l'on note $A_n = \frac{1}{s_n}\sum_{i=1}^n \frac{1}{i}\delta_{\left[M_i/\sigma_M\sqrt{i}\right]}$.

a) (Théorème de limite centrale presque sûr) : p.s., la suite (A_n) converge étroitement vers $\mathcal{N}(0,1)$.

b) (Loi forte quadratique des grands nombres) : p.s., la suite $\int x^2 A_n(dx) = \frac{1}{s_n}\sum_{i=1}^n \frac{M_i^2}{i^2\sigma_M^2}$ tend vers 1.

c) (Loi forte logarithmique des grands nombres) : si f est une fonction borélienne réelle presque partout continue pour la mesure de Lebesgue et telle que $\sup_x |f(x)|/x^2 < \infty$, p.s., la suite $\int f dA_n$ tend vers $\int f d\mathcal{N}(0,1)$.

d) (Théorème de limite centrale avec poids logarithmique et loi du logarithme itéré associée).
Pour $f \in L^2\left(\mathcal{N}\left(0,\sigma_M^2\right)\right)$ d'intégrale nulle et satisfaisant une condition de Lipschitz $|f(x) - f(y)| \leq C|x-y|(1+|x|+|y|)$,

$$\frac{1}{\sqrt{\log n}}\int f dA_n \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\sigma_f^2\right) \text{ et, p.s. } \limsup \frac{1}{\sqrt{2\log n \log\log\log n}}\left|\int f dA_n\right| \leq \sigma_f.$$

Tous ces résultats et d'autres beaucoup plus généraux relatifs aux martingales vectorielles ont été établis par trois mathématiciens tunisiens Touati, Chaabane et Maaouia ([ChaMaaTou98]).
Mais on est dans le cadre d'un PGD empirique étroit. On peut donc ici appliquer le corollaire unidimensionnel, par exemple à la fonction $f(x) = x$; $\int |x|^p d\nu(x) = 0$ signifie $\nu = \delta_0$ donc $H(\nu) = \infty$. On est dans le cas favorable.

Corollaire 4.4 *Dans le cadre du théorème 4.6, pour tout* $p > 1$*, le rapport*

$$\frac{\frac{1}{s_n}\sum_{i=1}^n \frac{1}{i}\frac{M_i}{\sqrt{i}}}{\left(\frac{1}{s_n}\sum_{i=1}^n \frac{1}{i}\left|\frac{M_i}{\sqrt{i}}\right|^p\right)^{1/p}}$$

satisfait un PGD troué avec un taux H_p *qui ne s'annule qu'en* 0*, est croissant convexe sur* $]0,\infty[$*, convexe décroissant sur* $]-\infty,0[$.

Pour $p = 2$, le dénominateur fait apparaître la moyenne logarithmique quadratique. Pour les applications statistiques, il est utile de remarquer que la constante σ_M a disparu.

$\diamondsuit$ **Contiguïté exponentielle**

Soit $(C_n)_{n\geq 0}$ une fonctionnelle additive croissante. On suppose que (C_n/n) converge super-exponentiellement vers 1 pour la vitesse $(\log n)$ et la loi initiale ν, c'est à dire :

$$\forall r > 0,\ \lim_n \frac{1}{\log n} \log \mathbb{P}_\nu \left(\left| \frac{C_n}{n} - 1 \right| \geq r \right) = -\infty.$$

C'est en particulier le cas, si (C_n/n) converge exponentiellement vers γ pour la vitesse n c'est à dire :

$$\forall r > 0,\ \limsup \frac{1}{n} \log \mathbb{P} \left(\left| \frac{C_n}{n} - 1 \right| \geq r \right) < 0.$$

Considérons alors les probabilités aléatoires $A_n = \frac{1}{s_n} \sum_{i=1}^n \frac{1}{i} \delta_{[M_i/\sigma_M \sqrt{i}]}$ et $B_n = \frac{1}{s_n} \sum_{i=1}^n \frac{1}{i} \delta_{[M_i/\sigma_M \sqrt{C_i}]}$. Montrons la contiguïté exponentielle des suites (A_n) et (B_n) c'est à dire :

$$\forall \rho > 0,\ \lim_n \frac{1}{\log n} \log \mathbb{P}_\nu \left(d_{Dudley} \left(A_n, B_n \right) \geq \rho \right) = -\infty.$$

On remarque que, pour des mesures de Dirac, $d_{Dudley}(\delta_a, \delta_b) \leq \inf(2, |a-b|)$. En outre, si l'on a $\left| \left(\sqrt{C_i/i} \right)^{1/2} - 1 \right| \leq r < 1$,

$$\begin{aligned} \left| \frac{M_i}{\sigma_M \sqrt{i}} - \frac{M_i}{\sigma_M \sqrt{C_i}} \right| &= \left| \frac{M_i}{\sigma_M \sqrt{i}} \right| \frac{|\sqrt{C_i/i} - 1|}{\sqrt{C_i/i}} \leq \left| \frac{M_i}{\sigma_M \sqrt{i}} \right| \frac{r}{1-r}, \\ \inf \left(\left| \frac{M_i}{\sigma_M \sqrt{i}} - \frac{M_i}{\sigma_M \sqrt{C_i}} \right|, 2 \right) &\leq 2 \text{ si } |M_i/\sqrt{i}| \geq 2\sigma_M \frac{1-r}{r}, \\ &\leq \frac{r}{1-r} \left| \frac{M_i}{\sigma_M \sqrt{i}} \right| \leq 2 \left(\frac{r}{1-r} \right)^2 \text{ sinon.} \end{aligned}$$

On fixe $\rho > 0$. Pour tous $r > 0$ et $R \in]1, \infty[$, il existe un N tel que, pour $i > N$, on ait

$$\mathbb{P} \left(\left| \sqrt{C_i/i} - 1 \right| \geq \frac{r}{3} \right) \leq i^{-R};$$

Pour $n^{\rho/10} \geq N$,

$$\begin{aligned} d_{Dudley}(A_n, B_n) \leq\ & \frac{2}{s_n} \sum_{i=1}^{n^{\rho/10}} \frac{1}{i} + \frac{2}{s_n} \sum_{i=n^{\rho/10}+1}^{n} \frac{1}{i} \mathbb{I}_{\left| \sqrt{C_i/i} - 1 \right| \geq r} + \frac{2}{s_n} \sum_{i=n^{\rho/10}+1}^{n} \frac{1}{i} \mathbb{I}_{|M_i/\sqrt{i}| \geq 2\sigma_M \frac{1-r}{r}} \\ & + 2 \left(\frac{r}{1-r} \right)^2 \frac{1}{s_n} \sum_{i=n^{\rho/10}+1}^{n} \frac{1}{i}. \end{aligned}$$

L'ensemble $\Gamma_r = \left\{ \nu; \nu \left(|x| > 2\sigma_M \frac{1-r}{r} \right) \geq \rho/4 \right\}$ est un fermé et, si $r \downarrow 0$, ces fermés décroissent vers l'ensemble vide. Par suite $\lim_{r \to 0} H \left(\nu \mid \nu \left(|x| > 2\sigma_M \frac{1-r}{r} \right) \geq \rho/4 \right) = \infty$

et on peut choisir r tel que $H\left(\nu;\nu\left(|x|>\sigma_M\frac{1-r}{r}\right)\geq\rho/4\right)\geq R$. On choisit aussi r assez petit pour que $2\left(\frac{r}{1-r}\right)^2<\rho/4$. Enfin, pour n assez grand $\frac{2}{s_n}\sum_{i=1}^{n^{\rho/10}}\frac{1}{i}<\rho/4$.
Ainsi, pour r choisi assez petit et n assez grand :

$$\begin{aligned}
\mathbb{P}\left(d_{Dudley}\left(A_n,B_n\right)>\rho\right) &\leq \mathbb{P}\left(\frac{2}{s_n}\sum_{i=n^{\rho/10}+1}^{n}\frac{1}{i}\mathbb{I}_{\left|\left(\frac{C_i}{i}\right)^{1/2}-1\right|\geq r}\geq\rho/8\right)+\mathbb{P}_\nu\left(\sum_{i=1}^{n}\frac{1}{i}\mathbb{I}_{|M_i/\sqrt{i}|\geq 2\sigma_M\frac{1-r}{r}}\geq\rho/4\right)\\
&\leq \frac{16\rho}{s_n}\sum_{i=n^{\rho/10}+1}^{n}\frac{1}{i}\mathbb{P}\left(\left|\left(\frac{C_i}{i}\right)^{1/2}-1\right|\geq r\right)+\mathbb{P}_\nu\left(\sum_{i=1}^{n}\frac{1}{i}\mathbb{I}_{|M_i/\sqrt{i}|\geq 2\sigma_M\frac{1-r}{r}}\geq\rho/4\right)
\end{aligned}$$

$$\begin{aligned}
&\limsup\frac{1}{\log n}\log\mathbb{P}\left(d_{Dudley}\left(A_n,B_n\right)>\rho\right)\\
\leq\ & -\inf\left(\limsup\frac{1}{\log n}\log\left(\frac{16\rho}{\log n}\frac{n^{-R\rho/10}}{R}\right),H\left(\nu;\nu\left(|x|\geq\sigma_M\frac{1-r}{r}\right)\geq\rho/4\right)\right)\\
\leq\ & -\inf\left(R,R\rho/10\right).
\end{aligned}$$

Et R étant choisi librement, la preuve est achevée.
On montre de la même manière que (A_n) et (B_n) sont exponentiellement contiguës à (D_n) avec

$$D_n=\left(\frac{1}{s_n}\sum_{i=1}^{n}\frac{1}{i}\delta_{\left[\sqrt{i}M_i/\sigma_M C_i\right]}\right);$$

en écrivant

$$\left|\frac{M_i}{\sigma_M\sqrt{i}}-\frac{M_i}{\sigma_M\sqrt{i}\left(C_i/i\right)}\right|=\left|\frac{M_i}{\sigma_{M\sqrt{i}}}\left(1-\frac{1}{C_i/i}\right)\right|.$$

♡ **Application au modèle AR(1) stable**
On considère le modèle AR(1) de dimension 1 défini avec $|\theta|<1$;

$$X_{n+1}=\theta X_n+\varepsilon_{n+1}.$$

On est dans le cadre du théorème si ε et la loi initiale ont des moments de tous les ordres finis, ε ayant une densité strictement positive et bornée.
On applique ce qui précède à $M_n=\sum_{i=1}^{n}X_{i-1}\varepsilon_i$ et à $C_n=\sum_{i=0}^{n-1}\|X_i\|^2$, donc à l'erreur de l'estimateur des moindres carrés

$$\overset{\wedge}{\theta}_n-\theta=\frac{M_n}{C_n}.$$

Ici, $\sigma_M^2=\gamma^2\sigma^2$ où $\mathbb{E}_\mu\left(X_1^2\right)=\gamma^2$ et $\mathbb{E}\left(\varepsilon_1^2\right)=\sigma^2$.
Dans ce cadre on a (cf. Worms [Wor99])

$$\forall r>0,\ \limsup\frac{1}{n}\log\mathbb{P}_\nu\left(|\frac{1}{n}\sum_{i=0}^{n-1}\|X_i\|^2-\gamma^2|>r\right)<0.$$

On a un PGD étroit pour la suite $\left(\frac{1}{\log n}\sum_{i=1}^{n}\frac{1}{i}\delta_{\left(\frac{\gamma}{\sigma}\sqrt{i}\left(\overset{\wedge}{\theta}_i-\theta\right)\right)}\right)$. Ce qui permet à nouveau d'appliquer le corollaire 4.3 et on a un bon PGD troué pour

$$\frac{1}{\log n}\sum_{i=1}^{n}\frac{1}{\sqrt{i}}\left(\overset{\wedge}{\theta}_i-\theta\right)/\sqrt{\frac{1}{\log n}\sum_{i=1}^{n}\left(\overset{\wedge}{\theta}_i-\theta\right)^2}.$$

Bibliographie

[Aco85a] A. De Acosta. Upper bounds for large deviations of dependent random vectors. *Z. Wahrscheinlichkeitstheorie*, 69 :551–565, 1985.

[Aco85b] A. De Acosta. On large deviations of sums of independent random variables. In *Lecture notes in Math.*, 1153 :1–14. Springer, 1985.

[Aco88] A. De Acosta. Large deviations for vector-valued Functionals of a Markov chain : Lower bounds. *Ann. Prob.*, 16 :925–960, 1988.

[Aco90] A. De Acosta. Large deviations for empirical measures of Markov chains. *J. of Theoretical Prob.*, 3 :395–431, 1990.

[AcoNey98] A. De Acosta and P. Ney. Large deviation lower bounds for arbitrary additive functionals of a Markov chain. *Ann. Prob.*, 26 :1660–1682, 1998.

[Att99] J. G. Attali. *1. Méthodes de stabilité pour des chaînes de Markov non fellériennes, 2. Sur quelques autres problèmes issus des réseaux de neurones, Thèse,* Université Paris 1, 1999

[Aze80] R. Azencott. Grandes déviations et applications. *Ecole de Proba. de Saint-Flour VIII,* number 774 in lectures notes in math., pages 1-176. Springer, 1980.

[BahZab79] R. Bahadur and S. Zabell. Large deviations for the sample mean in general vector spaces. *Ann. Prob.*, 7 :587-621, 1979.

[BalMey00] S. Balaji and S. P Meyn. Multiplicative ergodicity and large deviations for an irreductible Markov chain *Stoch. Proc and their Appl.*, 1 :123–144, 2000.

[Bal88] P. Baldi. Large deviations and stochastic homogenisation. *Ann. Mat. Pura Appl.*, 151 :161–177, 1988.

[BalMey00] S. Balaji and S.P. Meyn. Multiplicative ergodicity and large deviaitoons for an irreductible Markov chain. *Stoch. Proc. and their Appl.*, 1 :123–144, 2000.

[BaxJaiVar91] J.R. Baxter, N.C. Jain and S.R.S. Varadhan. Some familiar examples for which the large deviation principle does not hold. *Commun. Pure Appl. Math.*, 911–923, 1991.

[Ber01] B. Bercu. On large deviations in the gaussian autoregressive process : stable, instable and explosive cases, *Bernoulli*, 7 :299–316, 2001.

[BerGamLav00] B. Bercu, F. Gamboa and M. Lavielle. ESAIM Sharp Large deviations for Gaussian quadratic forms with applications. *ESAIM Prob. Stat.*, 4 :1–24, 2000.

[BerGamRou96] B. Bercu, F. Gamboa and A. Rouault. Large deviations for quadratic forms of stationary gaussian processes. *Stoch. Proc. and their Appl.*, 71 :75–90, 1997.

[BerGasRio02] B. Bercu, E. Gassiat and E. Rio. Concentration inequalities, large and moderate deviaition for self-normalized empirical processes. *To appear in Ann. Prob.*.

[BerRou98] B. Bercu and A. Rouault. Sharp large deviations for the Ornstein-Uhlenback process *Stoch. Proc. and their Appl.*, 71 :75–90, 1997.

[Bre79] J. Bretagnolle. Formules de Chernoff pour les lois empiriques de variables à valeurs dans des espaces généraux. Dans *Séminaire de statistique de l'Université de Paris-Sud. Grandes déviations et applications statistiques*, Astérisque, 60 :33–52, 1979.

[Bro88] G.A. Brosamler. An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.*, 104 :561–574, 1988.

[Bry90] W. Bryc Large deviations by the asymptotic value method. *M. Pinsky editor, Diffusion processes and related problems in analysis*, 1 :447–472, 1990.

[BryDem97] W. Bryc and A. Dembo. Large deviations for quadratic Functionals of gaussian processes. *J. of Theoretical Prob.*, 10 :307–322, 1997.

[Cel80] D. Cellier. *Méthodes de fission pour l'étude de la récurrence des chaines de Markov, Thèse* Université de Rouen, France, 1980.

[Cha01a] F. Chaabane. *Théorèmes limites par moyennisation logarithmique pour les martingales et applications statistiques, thèse* Université de Tunis, Tunisie, 2001.

[Cha01b] F. Chaabane. Invariance principles with logarithmic averaging for martingales. *Studia Math. Sci. Hungar*, 37 :21–52, 2001.

[ChaMaa00] F. Chaabane and F. Maaouia. Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM Probab. Statist.*, 4 :137–190, 2000.

[ChaMaaTou98] F. Chaabane, F. Maaouia and A. Touati. Généralisation du théorème de la limite centrale presque-sûre pour les martingales vectorielles. *C.R. Académie des sciences*, 326(I) :229–232, 1998.

[Che52] H. Chernov. A measure of asymptotics efficiency for tests of a hypothesis based on the sum of observations. *Ann. math. statist.*, 23 :493–507, 1952.

[Chen01] X. Chen. Moderate deviations for Markovian occupation times. *Stoc. Proc. Appl.*, 94 :51–70, 2001.

[ChiKus88] T. Chionobu and S. Kusuoka. The large deviations principle for hypermixing processes. *Probab. Th. Rel. fields*, 78 :627–649, 1988.

[Cra38] H. Cramér. Sur un nouveau théorème-limite de la théorie des probabilités. In *Actualités Scientifiques et Industrielles* number 736 in Colloque consacré à la théorie des probabilités, pages 5–23. Hermann, Paris, 1938.

[CsoHor92] M. Csörgö and L. Horváth. Invariance principles for logarithmic averages. *Math. Proc. Cambridge Philos. Soc.*, 12 :195–205, 1992.

[DemSha98c] A. Dembo and Q. M. Shao. Large and moderate deviations for Hotelling's T^2-statistic. *Preprint.*, 1998.

[DemSha98a] A. Dembo and Q. M. Shao. Self-normalized moderate deviations and LILs. *Stoch. Proc. and their Appl.*, 75 :51–65, 1998.

[DemSha98b] A. Dembo and Q. M. Shao. Self-normalized large deviations in vector spaces. *Progress in probability proceeding Oberwolfach*, 43 :27–32, 1998.

[DemZei98] A. Dembo and O. Zeitouni *Large deviations Techniques and Applications.* Springer, 1998.

[DeuStr89] J. Deuschel and D. W. Stroock. *Large deviations.* Academic press, 1989.

[Din93] I. H. Dinwoodie. Identifying a large deviation rate function. *Ann. Prob.*, 23 : 216–231, 1993.

[DinNey95] I. H. Dinwoodie and P. Ney Occupation measures for Markov chains. *J. Theoretical Prob.*, 8 :679–691, 1995.

[DinZab92] I. H. Dinwoodie and S. L. Zabell. Large deviations for exchangeable random vectors. *Ann. Prob.*, 3 :1147–1166, 1992.

[DonVar75a] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time I. *Commun. Pure Appl. Math.*, 28 :1–47, 1975.

[DonVar76] M.D. Donsker and S.R.S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time III. *Commun. Pure Appl. Math.*, 29 :389–461, 1976.

[DowMeyTwe96] D. Down, S.P. Meyn and R. Tweedie Exponential and uniform ergodicity of Markov processes. *Ann. of Prob.*, 23 :1671–1691, 1996.

[Duf97] M. Duflo. *Random Iterative Models.*, Volume 38 de *Applications of mathematics.* Springer, 1997.

[Dud89] R. Dudley. *Real analysis and probability.* Chapman Hall, 1989.

[DupEll97] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations.* Wiley, 1997.

[Ell84] R. S. Ellis. Large deviations for a general class of random vectors. *Ann. Prob.*, 12 :1–12, 1984.

[Ell85] R. S. Ellis. *Entropy, Large deviations and Statistical Mechanics* Springer, 1985.

[Ell88] R. S. Ellis. Large deviation for the empirical measure of a Markov chain with an application to the multivariate empirical measure *Ann. Prob.*, 16 :1496–1508, 1988.

[EthKur86] S.N. Ethier et T.G. Kurtz. Markov processes : characterization and convergence. Wiley, 1986.

[FloPha97] D. Florens et H. Pham. Large deviations in estimation of an Ornstein-Uhlenbeck model A paraître dans *Journa of applied probability*

[Gar77] J. Găртner. On large deviations from the invariant measure. *Th. Prob Appl.*, 22 :24–39, 1977.

[GulLipLot94] O. V. Gulinskii, R. S. Liptser and S. V. Lototskii. Large deviations for unbounded additive Functionals of a Markov process with discrete time (non compact case). *J. Appl. Math. Stoch. Anal.*, 7(3) :423–436, 1994.

[Haj82] B. Hajek. Hitting-times and occupation time bounds implied by drift analysis and applications. *Adv. applied Prob.*, 14 :502–525, 1982.

[HecMaa01] M. K. Heck and F. Maaouia. The principle of large deviations for martingale additive functionnals of recurrent Markov processes. *Electronic journa of probability*, 6(8) :1–26, 2001.

[Ilt00] M. Iltis. Sharp asymptotics of large deviations for general state space Markov-additive chains in $\mathbb{R}^d$. *Stat. and Prob letters*, 47 :365–380, 2000.

[IscNeyNum85] I. Iscoe, P. Ney and E. Nummelin Large deviations of uniformly recurrent Markov additive process *Advances in applied mathematics*, 6 :373–412, 1985.

[Jai90] N. Jain. Large deviations for additive Functionals of Markov processes : discrete time, non compact case *Ann Prob.*, 18 :1071–1098, 1990.

[KonMey02] I. Kontoyiannis and S.P. Meyn Spectral theory and limit theory for geometrical ergodic Markov processes. A paraître dans *Ann Prob.*, 2002.

[Lez98] P. Lezaud. Chernov-type bound for finite Markov chains. *The annals of appl Prob.*, 8 :849–867, 1998.

[Lez01] P. Lezaud. Chernoff and Berry-Esséen inequalities for Markov processes. *ESAIM Probab. Stat.*, 5 :183–201, 2001.

[Lév31] P. Lévy. Sur les séries dont les termes sont des variables événtuellement indépendantes. Studia math., 3 :119–155, 1931.

[Lin99] T. Lindvall. On strassen's theorem on stochastic domination. *Electronic communications in probability*, 4 :51–59, 1999.

[Lip93] R. Lipster. Large deviations for occupation measures of Markov processe : discrete time, non compact case. *theory Probab. appl.*, vol. 41 No. 1, 1993.

[Maa01a] F. Maaouia. *Théorèmes asymptotiques par moyennisation logarithmique pour les processus de Markov et applications statistiques, thèse* Université de Tunis, Tunisie, 2001.

[Maa01b] F. Maaouia. Principes d'invariance par moyennisation logarithmique pour les processus de Markov *Ann. of Prob.*, 29(4) :1859–1902, 2001.

[Mar96] K. Marton. A measure concentrations inequality for contracting Markov chains. *Geometric and functional analysis*, 6 :556–571, 1996.

[MarSep97] P. March and T. Seppalainen. Large deviations for the almost everywhere central limit theorem. *J. theor. Prob.*, 10 :935–965, 1997.

[Mil61] H. Miller. A convexity property in the theory of random variables defined on a finite Markov chain. *Ann. Math Statist.*, 32 :1260–1270, 1961.

[NeyNum86] P. Ney and E. Nummelin. Markov additive processes (ii) : Eigenvalue properties and limit theorems. *Ann. Prob.*, 2 :561–592, 1986.

[NeyNum87] P. Ney and E. Nummelin. Markov additive processes (ii) : Large deviations. *Ann. Prob.*, 15 :593–609, 1987.

[NumTuo82] E. Nummelin and P. Tuominen Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stoch Proc. and their Appl.*, 12 :187–202, 1982.

[Num84] E. Nummelin *Entropy, General irreductible Markov chains and nonnegative operators* Cambridge university press, 1984.

[OBr96] G. L. O'Brien. Sequencies of capacities, with connections to large deviations theory. *J. theoretical Prob.*, 9 :15–35, 1996.

[OBrSun96] G. L. O'Brien et J. Sun Large deviations on linear spaces. *Prob. Math. Statis.*, 16 :261–273, 1996.

[OBrVer91] G. L. O'Brien et W. Vervaat. *Capacities, large deviations and log-log laws*, pages 43–83. Numéro 25 de *Progr. Probability.* 1991.

[OBrVer95] G. L. O'Brien et W. Vervaat. Compactness in the theory of large deviations *Stoch. Proc. and their appl.*, 57 :1–10, 1995.

[Pit74] J. W. Pitman. Uniform rates of convergence for Markov transition probabilities. *Wahrscheinlichkeitstheorie*, 29 :193–227, 1974.

[Puk91] A. A. Puhalskii. *On functional principle of large deviations*, pages 198–218. VSP/Mokslas, Utrecht-Vilnius, 1991.

[Puk94a] A. A. Puhalskii. On the theory of large deviations. *Theory of Probab. and Applications*, 38 :490–497, 1994.

[Rev84] D. Revuz. *Markov chains* North Holland mathematical library, 1984.

[Rud73] W. Rudin *Functional analysis.* McGraw-Hill, 1973.

[Sam00] P. M.Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Prob*, 28 :416–461, 2000.

[San57] I. N. Sanov On the probability of large deviations of random variables (en russe, 1957). traduction en anglais dans *selected translations in Mathematical Statistics and Probability* , I :213–244, 1961.

[Sha97] Q. M. Shao. Self-normalized large deviations. *Ann. Prob.*, 25 :285–328, 1997.

[Sha97b] Q. M. Shao. A cramer type large deviations result for Student's t-statistics. *J. of theoretical Prob.*, 12 :385–398, 1997.

[Sha98] Q. M. Shao. Recent developments in self-normalized limit theorems. In B. Szyskowicz, editor, *Asymptotic methods in probability and statistics*, 25 :467–480. Elsevier, 1998.

[Stra65] V. Strassen. The existence of probability measures with given marginals. *Annals of Math.*, 36 :423–439, 1965.

[Str93] D. W. Stroock. *Probability theory, an analytic view.* Cambridge university press, 1993.

[Str94] D. W. Stroock. *An introduction to large deviations.* Springer, 1994.

[Var66] S.R.S. Varadhan. Asymptotic probabilities and differential equations. *Commun. Pure Appl. Math.*, 19 :261–286, 1966.

[Var84] S.R.S. Varadhan. *Large deviations and applications.* SIAM, 1984.

[Wor99] J. Worms. Moderate deviations for stable Markov chains and regression models. *Electronic journal of Probability* , 4 :1–28, 1999.

[Wor00] J. Worms. *Principes de déviations modérées pour des martingales et applications statistiques, Thèse* Université de Marne-la-Vallée, France, 2000.

[Wor00b] J. Worms. Moderate deviations for martingales and for some kernel estimators. *Mathematicals methods of statistics*, 2000.

[Wor01] J. Worms. Large and moderate deviations upper bounds for the gaussian autoregressive process. *Statistics and probability letters*, 51 :235–243, 2001.