Z. Wahrscheinlichkeitstheorie verw. Gebiete
38, 1–54 (1977)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Mélanges d'équations différentielles et grands écarts à la loi des grands nombres

R. Azencott et G. Ruget

Université Paris VII, 2 Place Jussieu, U.E.R. de Mathématiques, F-75005 Paris V, France
et IRIA, Domaine de Voluceau, Rocquencourt, BP 5, 78150, Le Chesnay, France

Introduction

On étudie ici le comportement, sur un temps très long, d'un mobile soumis alternativement à plusieurs équations différentielles, les choix successifs étant par exemple espacés d'un laps de temps τ fixe petit, et ayant lieu indépendamment les uns des autres en respectant des statistiques connues. L'origine de ce travail était la recherche de prévisions sur le comportement lent d'un système nerveux à synapses évolutives (cf. les états LSD de [1], interpolés de façon continue comme dans [7]), ou encore l'étude de l'évolution même de l'état des synapses d'un seul neurone à partir de celle des récepteurs de l'acétylcholine. A une toute autre échelle explicative, mais encore dans la théorie de l'apprentissage, on trouve un modèle mathématique voisin dans [8], et avec d'autres motivations, dans le «processus de transport» [9], ou dans des études fines d'algorithmes de gradient stochastique [6]. L'originalité de notre travail est qu'on ne peut s'y contenter de l'approximation par une équation différentielle moyenne, ni d'une approximation par une diffusion, les phénomènes intéressants sur une longue échelle de temps, à savoir, les sauts d'un bassin à un autre du «champ de vecteurs moyen», manifestant de grands écarts par rapport à la loi des grands nombres. Très rapidement, nous avons eu connaissance de [14], qui étudie un phénomène analogue dans le cas des petites diffusions, et dont nous avons respecté la méthode d'attaque. L'autre ingrédient important est la théorie des grands écarts, dont on trouvera établi ce qui nous en est nécessaire au § 1, et qui a dû être redécouverte plusieurs fois, mais dont la première manifestation semble être [3]. Dans des cas simples, on pourrait se contenter des résultats existants, par exemple [10], mais, pour traiter le problème suffisamment significatif indiqué ci-après, nous avons eu besoin de résultats de robustesse qui nous semblent nouveaux.

Pour en revenir à la formalisation du processus, on voit que les données pourraient être une mesure de probabilité μ sur un compact dans l'ensemble des champs de vecteurs C' sur une variété M, et un temps petit τ. En fait, cette donnée ne peut qu'être inaccessible à l'expérience, et on doit pouvoir se contenter de la collection des images μ_m de μ par les évaluations dans les divers espaces tangents $T_m M$. Entre deux choix, on va à peu près tout droit, (voir le § 5 pour une description précise), les prévisions faites n'étant pas sensibles à cet à peu près. On gagne ainsi la possibilité d'envisager des μ_m composées de poids variables sur des vitesses fixes ($M=\mathbb{R}^n$), ce qui serait exclu par le premier formalisme.

Au voisinage de $m \in M$, on peut en première approximation supposer que, dans une carte de M, les μ_m sont constamment égales à μ. D'après Cramer, la probabilité d'aller en un temps t, de m au voisinage d'un m' lui-même proche de m, est à peu près $\exp\left(-\frac{t}{\tau}\lambda_\mu\left(\frac{m\,m'}{t}\right)\right)$, où λ_μ est une indicatrice convexe, nulle au centre de gravité de μ, infinie à l'extérieur de l'enveloppe convexe du support de μ.

Si φ est un chemin $[0, T] \to M$, absolument continu, on peut alors conjecturer que l'intégrale

$$\Lambda(\varphi)=\int_0^T \lambda_{\mu_{\varphi(t)}}(\varphi'(t))\,dt$$

est l'opposé du logarithme de la probabilité qu'une trajectoire du processus issue de $\varphi(0)$ reste, pendant le temps T, uniformément voisine de φ. Une telle évaluation est donnée dans les §§ 6 et 7; elle repose sur les résultats de robustesse du § 2, mais aussi, et l'on peut fabriquer des exemples simples (*) pour se convaincre que cela n'est pas superflu, sur des hypothèses de type lipschitzien concernant la variation des lignes de niveau des λ_{μ_m}; l'objet du § 3 est de donner des conditions naturelles sur les mesures μ_m elles-mêmes, qui assurent la régularité désirée. Au § 4, on étudie la fonctionnelle Λ, et au § 8, le potentiel associé

$$V(K, L)=\inf_{\substack{\varphi(0)\in K\\ \varphi(T)\in L\\ T\in\mathbb{R}^+}} \Lambda(\varphi).$$

La condition $V(x, y)=V(y, x)=0$ définit une relation d'équivalence sur M; nous supposerons qu'il y a un nombre fini de classes d'équivalence K_i, compactes, contenant tous les ensembles ω-limites du champ moyen.

Les méthodes de Ventsell et Freidlin donnent alors, dans notre cas, tous leurs résultats, dont voici un exemple pour la commodité du lecteur: partant de x, nous étudions le point de première sortie d'un domaine U à la frontière duquel le champ de vitesses moyen est rentrant, et qui contient les K_i pour $i\in I$. Posons $V_{ij}=V(K_i, K_j)$ et $V_{i\partial}=V(K_i, \partial U)$, et soit A_i l'ensemble (fermé) des $y\in\partial U$ tels que $V(K_i, y)=V_{i\partial}$.

Alors, pour tout $\varepsilon>0$, avec une probabilité tendant vers 1 lorsque τ tend vers 0, les trajectoires du processus issues de x quittent U en passant à une distance inférieure à ε de $\bigcup_{i\in J} A_i$, où J est défini comme suit:

Soit $K_{i(x)}$ la classe contenant l'attracteur de la trajectoire issue de x du champ moyen. Considérons les graphes G simplement connexes de sommets $I\cup\partial$, tels que, de tout point de I, parte exactement une flèche, et minimisant $\sum_{(i,j)\in \mathrm{arcs}(G)} V_{ij}$. Un indice j appartient à J si et seulement s'il existe un G dans lequel la dernière flèche du chemin menant de $i(x)$ à ∂ est $(j\,\partial)$.

En résumé, pendant de longues périodes de temps, le processus stationne sur un attracteur du champ de vitesses moyen, qu'il décrit d'ailleurs suivant la pondération décrite par Bowen et Ruelle, puis il saute sur un autre attracteur que l'on peut prévoir avec une quasi-certitude si τ est petit. Une application possible est l'étude du comportement d'algorithmes de gradient stochastique dans le cas où la fonction à minimiser possède plusieurs minimums: on trouvera dans [11] une telle situation dans un problème d'identification de système linéaire classique en automatique.

Ayant achevé ce travail, nous nous sommes rendus compte que les grands écarts ont été un centre d'intérêt récurrent de Varadhan: dans [13], il utilise un processus du type étudié ici, mais à μ constant; dans [4], avec Donsker, il explique l'usage possible d'estimations exponentielles comme on en trouve dans cet article, et dans [5], il revient sur le théorème de Cramer. Mais il n'y a pas de recouvrement avec les principales difficultés résolues ici.

(*) Sans hypothèses lipschitziennes sur la variation des μ_m, il peut y avoir une grande instabilité dans le comportement du processus: prenons $M=\mathbb{R}$, et pour tout m, μ_m composée de deux masses 1/2, placées aux vitesses $-1-\sqrt{|m|}$ et $\sqrt{|m|}$. Un premier processus consiste à suivre, entre deux choix, l'un des champs $m\to -1-\sqrt{|m|}$ ou $m\to\sqrt{|m|}$; avec ce processus, il est impossible de passer de $m<0$ à $m>0$. Un autre processus, associ aux mêmes μ_m, consiste à garder, entre deux choix, une vitesse constante; pour celui-ci, la fonctionelle Λ fournit une bonne minoration de la probabilité d'être proche d'un chemin allant de $m<0$ à $m>0$.

1. Transformée de Cramér d'une probabilité

(1.1) *Notations.* Soient $\mathbb{X}$ un espace euclidien, $\mathscr{B}(\mathbb{X})$ la σ-algèbre borélienne des parties de $\mathbb{X}$, μ une probabilité sur $\mathbb{X}$. Soit (X_n), $n\geqq 1$, une suite de variables aléatoires indépendantes, définies sur un espace de probabilité $(\Omega, \mathscr{B}, P)$, à valeurs dans $\mathbb{X}$, et de même loi μ.

On pose $\bar{X}_n=\frac{1}{n}(X_1+X_2+\cdots+X_n)$, et si $A\in\mathscr{B}(\mathbb{X})$

$$-\underline{\sigma}_\mu(A)=\varliminf_{n\to+\infty}\frac{1}{n}\log P(\bar{X}_n\in A),\quad \text{et}\quad -\bar{\sigma}_\mu(A)=\varlimsup_{n\to+\infty}\frac{1}{n}\log P(\bar{X}_n\in A).$$

Il existe une littérature fournie sur la détermination d'une classe large d'ensembles A pour lesquels $\bar{\sigma}_\mu(A)=\underline{\sigma}_\mu(A)$, et sur l'évaluation de

$$\lim_{n\to+\infty}\frac{1}{n}\log P(\bar{X}_n\in A).$$

Le calcul de cette limite fait intervenir une fonction convexe $\lambda_\mu\colon\mathbb{X}\to[0,+\infty]$, attachée à μ, que nous redéfinissons ci-dessous. C'est par souci de complétude que nous reprenons cette question, ainsi que pour introduire les méthodes utiles dans la suite du travail, et pour préciser le comportement de λ_μ au bord de *l'enveloppe convexe fermée du support de μ (notée S_μ dans la suite).* Nous précisons aussi la signification de λ_μ dans les mêmes circonstances, et la dépendance de λ_μ vis-à-vis de (μ, x).

(1.2) *Hypothèse sur* μ. Dans tout ce paragraphe, nous supposons que S_μ engendre $\mathbb{X}$, et qu'il existe $\alpha>1$ tel que $\int_{\mathbb{X}} e^{|x|^\alpha}\, d\mu(x)$ soit finie. Nécessairement *le barycentre* $m=\int_{\mathbb{X}} x\, d\mu(x)$ est alors bien défini.

(1.3) *Définition de* λ_μ. Soit μ une probabilité sur $\mathbb{X}$ vérifiant l'hypothèse 1.2. Nous définissons par récurrence sur la dimension de $\mathbb{X}$ sa *transformée de Cramér* λ_μ, fonction définie sur $\mathbb{X}$ et à valeurs dans $[0, +\infty]$.

(i) *Supposons d'abord que* $\dim \mathbb{X}=1$. L'ensemble des $\xi\in\mathbb{X}$ tels que la transformée de Laplace $\hat{\mu}(\xi)=\int_{\mathbb{X}} e^{\xi x}\, d\mu(x)$ soit finie, a pour intérieur un intervalle ouvert U contenant m.

La fonction $\log \hat{\mu}(\xi)=\varphi(\xi)$ est strictement convexe et analytique sur U, de dérivée

$$\varphi'(\xi)=\frac{\int x\, e^{\xi x}\, d\mu(x)}{\int e^{\xi x}\, d\mu(x)} \tag{1}$$

de sorte que φ' applique de façon croissante U sur l'intérieur de S_μ. On définit une fonction λ_μ strictement convexe et analytique sur $\mathring{S}_\mu$ en posant

$$\lambda_\mu(m)=0 \quad \text{et} \quad \lambda'_\mu(\varphi'(\xi))=\xi. \tag{2}$$

La fonction λ_μ étant convexe, on la prolonge par continuité à S_μ (on est en dimension 1!). Sur $\mathbb{X}-S_\mu$ on pose $\lambda_\mu=+\infty$.

Il est facile de vérifier que

$$e^{-\lambda_\mu(x)}=\inf_{\xi\in\mathbb{X}}[e^{-\xi x}\hat{\mu}(\xi)] \qquad \text{pour } x\in\mathbb{X}. \tag{3}$$

L'infimum est atteint lorsque $x\in\mathring{S}_\mu$, pour un ξ vérifiant $\hat{\mu}'(\xi)=x\,\hat{\mu}(\xi)$; mais il n'est pas nécessairement atteint pour $x\notin\mathring{S}_\mu$. On vérifie aussi facilement que si a est une extrémité de l'intervalle S_μ, $\lambda_\mu(a)$ est fini si et seulement si $\mu\{a\}>0$, et vaut alors $-\log \mu(a)$.

(ii) *Cas où* $\dim\mathbb{X}>1$. Soit $\mathbb{X}^*$ le dual de $\mathbb{X}$. Pour $x\in\mathring{S}_\mu$ nous posons

$$\lambda_\mu(x)=\sup_{\xi\in\mathbb{X}^*}\lambda_{\xi(\mu)}[\xi(x)]. \tag{4}$$

Sur $\mathbb{X}-S_\mu$ nous posons $\lambda_\mu=+\infty$. Il reste à définir λ_μ sur la frontière de S_μ. Ceci va être fait par récurrence. Supposons que nous ayons déjà défini les transformées de Cramer en dimension strictement inférieure à celle de $\mathbb{X}$. Soit F une face de S_μ, et soit W le sous-espace affine de $\mathbb{X}$ engendré par le support de $\mu/_F$. Posons $\mu/_W=\mu(W)\,\nu$ où ν est une probabilité sur W. La fonction $\lambda_\nu\colon W\to[0,+\infty]$ est donc définie d'après l'hypothèse de récurrence. Sur W nous posons $\lambda_\mu=\lambda_\nu-\log\mu(W)$. Il n'est pas difficile de vérifier qu'une telle définition est consistante.

(1.4) **Lemme.** *Sous les hypothèses 1.2,* λ_μ *est, dans l'intérieur de* S_μ, *une fonction convexe et analytique. D'autre part, pour* $\alpha>1$ *et* $C>0$ *fixés, soit* E *l'ensemble des probabilités* ν *sur* $\mathbb{X}$ *telles que*

$$\int e^{|x|^\alpha}\, d\nu(x)\leq C$$

et telles que S_ν *engendre* $\mathbb{X}$.

Alors, E étant muni de la topologie faible, $\lambda_\nu(x)$: $E \times \mathbb{X} \to [0, +\infty]$ est continue en tout point (ν, x) tel que $x \in \mathring{S}_\nu$.

Preuve. L'équation $\hat{\mu}'(\xi) = \hat{\mu}(\xi) x$ possède une solution unique $\xi \in \mathbb{X}^*$ pour tout $x \in \mathring{S}_\mu$. En effet, la fonction $\log \hat{\mu}(\xi)$ est convexe et analytique; il y a donc au plus un ξ où sa différentielle ait une valeur donnée x, et l'ensemble des formes linéaires obtenues comme différentielles de $\log \hat{\mu}(\xi)$ quand ξ varie est convexe. D'après le théorème de Krein-Milman, il ne reste plus qu'à montrer que l'on peut trouver ξ de sorte que $x = \dfrac{\hat{\mu}'(\xi)}{\hat{\mu}(\xi)}$ soit aussi voisin qu'on veut d'un point extrémal x_0 de S_μ.

Soit π un hyperplan d'appui de S_μ en x_0 ne rencontrant S_μ qu'en x_0; puisque $\dfrac{\hat{\mu}'(\xi)}{\hat{\mu}(\xi)}$ est le barycentre de la mesure ν telle que $\dfrac{d\nu}{d\mu}(y) = e^{\langle \xi, y \rangle}$, on voit qu'il suffit de prendre ξ suffisamment grand et orthogonal à π pour rapprocher arbitrairement $\dfrac{\hat{\mu}'(\xi)}{\hat{\mu}(\xi)}$ de x_0.

Il est clair que la solution $\xi(x)$ de $\hat{\mu}'(\xi) = \hat{\mu}(\xi) x$, pour $x \in \mathring{S}_\mu$ est un point où $e^{-\langle \xi, x \rangle} \hat{\mu}(\xi)$ atteint son minimum en $\xi \in \mathbb{X}^*$.

D'autre part d'après (3) et (4), nous avons pour $x \in \mathring{S}_\mu$,

$$e^{-\lambda_\mu(x)} = \inf_{f \in \mathbb{X}^*} \left[\inf_{t \in \mathbb{R}} e^{-t f(x)} f \circ \mu(t) \right].$$

Comme l'infimum entre crochets est atteint, on obtient immédiatement, après changement de notation

$$e^{-\lambda_\mu(x)} = \inf_{\xi \in \mathbb{X}^*} e^{-\langle \xi, x \rangle} \hat{\mu}(\xi), \qquad x \in \mathring{S}_\mu. \tag{5}$$

Par suite, $\xi(x)$ étant comme ci-dessus on *a*, pour $x \in \mathring{S}_\mu$

$$\lambda_\mu(x) = \langle \xi(x), x \rangle - \log \hat{\mu}[\xi(x)]. \tag{6}$$

Ce dernier résultat et des propriétés de régularité évidentes de la solution $\xi(x)$ de l'équation $\dfrac{\hat{\mu}'(\xi)}{\hat{\mu}(\xi)} = x$, $x \in \mathring{S}_\mu$ prouvent immédiatement le lemme. Remarquons en passant que (6) entraîne aussi immédiatement

$$\lambda'_\mu(x) = \xi(x) \qquad \text{pour } x \in \mathring{S}_\mu. \tag{7}$$

(1.5) **Proposition.** *Pour $\alpha > 1$, $C > 0$, soit E l'ensemble de probabilités sur $\mathbb{X}$ défini en 1.4. La fonction $\lambda_\mu(x)$: $E \times \mathbb{X} \to [0, +\infty]$ est semi-continue inférieurement (E est muni de la topologie de la convergence faible). La fonction $\lambda_\mu(\cdot)$ est convexe sur $\mathbb{X}$ entier.*

Ce résultat serait démontré classiquement de façon analytique en prouvant d'abord que, pour tout $x \in \mathbb{X}$,

$$e^{-\lambda_\mu(x)} = \inf_{\xi \in \mathbb{X}^*} [e^{-\langle \xi, x \rangle} \hat{\mu}(\xi)]$$

($\leqq$ est clair; si x est à la frontière de S_μ, on utilise pour montrer $\geqq$ une forme linéaire ξ ayant la restriction adéquate sur la face de S_μ contenant x dans son intérieur, et un fort gradient positif vers l'extérieur de S_μ). Nous en donnerons une démonstration, plus combinatoire et dans l'esprit du § 2, au cours de la preuve de la proposition 1.7.

De la proposition 1.5, résulte aussi la formule, valable pour tout $x \in \mathbb{X}$

$$\lambda_\mu(x) = \sup_{\xi \in \mathbb{X}^*} \lambda_{\xi(\mu)}[\xi(x)].$$

(1.6) *Définition.* Soit μ une probabilité sur $\mathbb{X}$ vérifiant l'hypothèse 1.2, et soit λ_μ sa transformée de Cramer. Nous dirons qu'une partie borélienne A de $\mathbb{X}$ est *μ-régulière* si on a (notations (1.1)

$$\underline{\sigma}_\mu(A) = \bar{\sigma}_\mu(A) = \inf_{x \in A} \lambda_\mu(x)$$

et nous poserons $\sigma_\mu(A) = \underline{\sigma}_\mu(A) = \bar{\sigma}_\mu(A)$. D'après 1.1 on a alors

$$\lim_{n \to +\infty} \frac{1}{n} \log P(\bar{X}_n \in A) = -\sigma_\mu(A).$$

(1.7) **Proposition.** *Soit μ une probabilité sur $\mathbb{X}$ vérifiant 1.2. Alors la famille des ensembles μ-réguliers est stable par union finie et contient les ensembles suivants:*

(i) *les demi-espaces fermés, et plus généralement les convexes fermés A ayant le propriété suivante: soit F la face de S_μ telle que A rencontre F mais aucune autre face de S_μ ayant F dans son adhérence et soit S_F l'enveloppe convexe du support de $\mu|_F$; alors les intérieurs, dans l'espace vectoriel W_F engendré par F, de S_F et $A \cap F$ ont une intersection non vide.*

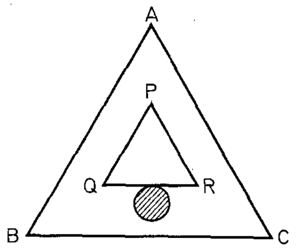
(ii) *les boréliens A de $\mathbb{X}$ ayant la propriété suivante: il existe au moins une suite $x_n \in A$ vérifiant*

$$\inf_{x \in A} \lambda_\mu(x) = \lim_{n \to +\infty} \lambda_\mu(x_n),$$

et telle que, pour tout n, A contienne un voisinage de x_n dans l'espace vectoriel engendré par la face minimale de S_μ passant par x_n.

(1.8) *Exemples.* D'après 1.7 (ii), tout ouvert de $\mathbb{X}$ est μ-régulier; si F est une face de S_μ, tout ouvert de l'espace vectoriel engendré par F est μ-régulier, d'après 1.7 (ii). Tout borélien A vérifiant $\bar{A} = \bar{\mathring{A}}$, et rencontrant $\mathring{S}_\mu$ est μ-régulier (d'après (1.7 (ii) et 1.5). Tout convexe fermé dont l'intérieur rencontre $\mathring{S}_\mu$ est μ-régulier d'après 1.7 (i).

Donnons des exemples d'ensembles qui ne sont pas μ-réguliers. Si μ est absolument continue par rapport à la mesure de Lebesgue, aucune partie Lebesgue-négligeable de $\mathring{S}_\mu$ ne peut être μ-régulière. Voici un exemple de convexe fermé, d'intérieur non vide, et qui n'est pas μ-régulier:



S_μ est un tétraèdre $ABCD$, (D est derrière le plan de figure); μ est la somme de la mesure de Lebesgue de $\mathbb{R}^3$ restreinte au tétraèdre $ABCD$ et de trois masses ponctuelles en P, Q, R.

Le convexe Γ est un cylindre orthogonal au plan de figure, placé devant le plan de figure, sur lequel il s'appuie suivant un disque tangent extérieurement au côté QR du triangle PQR. On peut voir (utiliser les idées développées ci-dessous) que Γ n'est pas μ-régulier.

Le seul exemple de borélien non convexe auquel nous appliquerons effectivement la proposition 1.7 (ii) est le suivant:

(1.9) **Corollaire.** *Pour tout* $a \geqq 0$ *l'ensemble* $A = \{x \in \mathbb{X}: \lambda_\mu(x) \geqq a\}$ *est* μ*-régulier et* $\sigma_\mu(A) = a$. *Remarquons que* $B = \{x \in \mathbb{X}: \lambda_\mu(x) > a\}$ *est aussi* μ*-régulier (car ouvert) mais que* $\sigma_\mu(B)$ *ne vaut* a *que si* $B \neq \mathbb{X} - S_\mu$.

(1.10) *Plan de la preuve des prop. 1.5 et 1.7.* Nous démontrerons 1.7 dans le cas des $\frac{1}{2}$ espaces fermés et des convexes fermés, et la minoration de $P(\bar{X}_n \in A)$ dans le cas non convexe; à l'aide de ces résultats, nous prouverons la semi continuité inférieure de $\lambda_\mu(x)$ pour μ fixe. Nous pourrons alors obtenir la majoration de $P(\bar{X}_n \in A)$ dans le cas où A est non convexe, compact, puis non compact. Nous prouverons alors la s.c. i de $\lambda_\mu(x)$ en (μ, x).

(1.11) *Cas des $\frac{1}{2}$ espaces.* Prouvons d'abord 1.7 lorsque A est un $\frac{1}{2}$ espace fermé, par récurrence sur dim $\mathbb{X}$. Quand dim $\mathbb{X} = 1$, ce résultat est bien connu et s'obtient soit par la méthode du col [3] soit bien plus aisément, à l'aide de (3) et d'estimation à la Cebycheff [2]. Supposons maintenant dim $\mathbb{X} > 1$ et $A \cap \mathring{S}_\mu \neq \emptyset$.

Montrons que le minimum ℓ de λ_μ sur $A \cap \mathring{S}_\mu$ est atteint en un point de $A \cap \mathring{S}_\mu$. Sinon, on construirait une suite $x_n \in A \cap \mathring{S}_\mu$ telle que $\lambda_\mu(x_n)$ décroisse vers ℓ, et telle que x_n tende vers $x \in A \cap \partial S_\mu$; soit δ un bout de demi-droite issu de x et situé dans $A \cap \mathring{S}_\mu$; soit U un voisinage convexe de x et soit Σ le cône des points v de $\mathbb{X}$ vérifiant $\langle \xi, v \rangle \leqq 0$ pour tout $\xi \in \mathbb{X}^*$ tel que le barycentre de la mesure v_ξ définie par $\frac{d v_\xi}{d\mu}(y) = e^{\langle \xi, y \rangle}$ appartienne à U. On vérifie (examiner le graphe de log $\hat{\mu}$) que pour U assez petit, δ est intérieure au cône Σ. Choisissons $y \in \delta \cap U$. Pour n assez grand, tout point z du segment $[y, x_n]$ appartient à Σ. Or pour tout point z de $\mathring{S}_\mu$ on a, d'après (7)

$$\lambda'_\mu(z) = \xi(z)$$

où $\xi(z)$ est tel que z soit le barycentre de la mesure $v_{\xi(z)}$.

La formule des accroissements finis et l'inclusion $z \in \Sigma$ entraînent donc

$$\lambda_\mu(y) - \lambda_\mu(x_n) \leqq 0.$$

On obtient donc $\lambda_\mu(y) \leqq \ell$, ce qui contredit notre hypothèse.

Soit donc $a \in A$ un point de $A \cap \mathring{S}_\mu$ où $\lambda_\mu / A \cap \mathring{S}_\mu$ est minimum. Si a est intérieur au $\frac{1}{2}$ espace A, ce ne peut être que le barycentre m de μ car λ_μ est strictement convexe dans $\mathring{S}_\mu$ avec un unique minimum local atteint en m. Dans ce cas le résultat 1.7 est évident d'après la loi des grands nombres. Si $a \in \partial A$ nécessairement le gradient $\lambda'_\mu(a)$ est orthogonal à ∂A et (cf. preuve de 1.4) $\lambda_\mu(a) = \lambda_{f(\mu)}(1)$ où l'application linéaire $f \in \mathbb{X}^*$ vérifie $f(v)=1$ pour tout $v \in \partial A$. Il suffit alors d'appliquer le résultat en dimension 1 au v.a. $f(X_n)$ pour obtenir

$$-\lambda_\mu(a) = \lim_{n \to +\infty} \frac{1}{n} \log P(\bar{X}_n \in A).$$

Reste à prouver que si $b \in \partial S_\mu \cap A$, on a $\lambda_\mu(b) \geqq \lambda_\mu(a)$. Soit H un hyperplan d'appui de S_μ en b et posons $\mu|_H = \mu(H)\, v$. On a $P(\bar{X}_n \in A) \geqq P(\bar{X}_n \in A \cap H)$. Comme $\bar{X}_n \in H$ est équivalent à $X_i \in H$ pour tout $i \in [1, n]$ on a

$$\begin{aligned} P(\bar{X}_n \in A \cap H) &= P(\bar{X}_n \in A, X_i \in H, 1 \leqq i \leqq n) \\ &= P(\bar{X}_n \in A / X_i \in H, 1 \leqq i \leqq n)\, \mu(H)^n \\ &= P(\bar{X}'_n \in A)\, \mu(H)^n \end{aligned}$$

où les X'_i sont des v.a. indépendantes de même loi v. D'où d'après l'hypothèse de récurrence (c'est-à-dire 1.7 vraie pour A $\frac{1}{2}$ espace fermé, en dimension inférieure à celle de $\mathbb{X}$)

$$\lambda_\mu(a) \leqq -\log \mu(H) + \inf_{b \in A \cap H} \lambda_v(b) = \inf_{b \in A \cap H} \lambda_\mu(b)$$

ce qui achève de montrer que A est μ-régulier. Nous avons utilisé l'hypothèse supplémentaire $A \cap \mathring{S}_\mu \neq \emptyset$ mais si $A \cap \mathring{S}_\mu = \emptyset$, on remplace S_μ par S_F où F est défini comme dans 1.7 (i), et μ par $\mu|_F$, et on est ramené à la situation étudiée. Donc les $\frac{1}{2}$ espaces fermés sont μ-réguliers.

(1.12) *Minoration de $\sigma(A)$ pour A convexe.* Si A est un convexe fermé rencontrant $\mathring{S}_\mu$, on peut reprendre la preuve ci-dessus pour montrer que λ_μ restreinte à $A \cap \mathring{S}_\mu$ y atteint son minimum en un point a. L'hyperplan H tangent à la surface de niveau de λ_μ au point a est nécessairement un hyperplan d'appui de A (sauf dans le cas trivial où a = barycentre de μ et est intérieur à A). Par suite, si H^+ est le $\frac{1}{2}$ espace fermé limité par H et contenant A, on a d'après 1.1 $\bar{\sigma}_\mu(A) \geqq \bar{\sigma}_\mu(H^+)$ et d'après l'étude faite pour les $\frac{1}{2}$ espaces

$$\bar{\sigma}_\mu(H^+) = \inf_{H^+} \lambda_\mu = \lambda_\mu(a). \tag{8}$$

Comme $\inf_A \lambda_\mu \geqq \inf_{H^+} \lambda_\mu$ (à cause de l'inclusion $H^+ \supset A$) et comme $a \in A$, on conclut que

$$\lambda_\mu(a) = \inf_{H^+} \lambda_\mu = \inf_A \lambda_\mu.$$

On obtient ainsi l'inégalité

$$\bar{\sigma}_\mu(A) \geqq \inf_A \lambda_\mu \tag{8}$$

pour A convexe fermé rencontrant $\mathring{S}_\mu$. Si $A \cap \mathring{S}_\mu = \emptyset$, on remplace S_μ par S_F et μ par $\mu|_F$ où F est la face de S_μ définie dans 1.7 (i), ce qui étend la validité de (8) à tous les convexes fermés A vérifiant 1.7 (i).

(1.13) *Remarque.* Soit $\mathscr{C}$ la classe des parties μ-régulières de $\mathbb{X}$. Pour toute paire $A_1, A_2 \in \mathscr{B}(\mathbb{X})$, on a

$$P(\bar{X}_n \in A_1) \vee P(\bar{X}_n \in A_2) \leqq P(\bar{X}_n \in A_1 \cup A_2) \leqq P(\bar{X}_n \in A_1) + P(\bar{X}_n \in A_2)$$

et par suite

$$[-\underline{\sigma}(A_1)] \vee [-\underline{\sigma}(A_2)] \leqq -\underline{\sigma}(A_1 \cup A_2) \leqq -\bar{\sigma}(A_1 \cup A_2) \leqq [-\bar{\sigma}(A_1)] \vee [-\bar{\sigma}(A_2)].$$

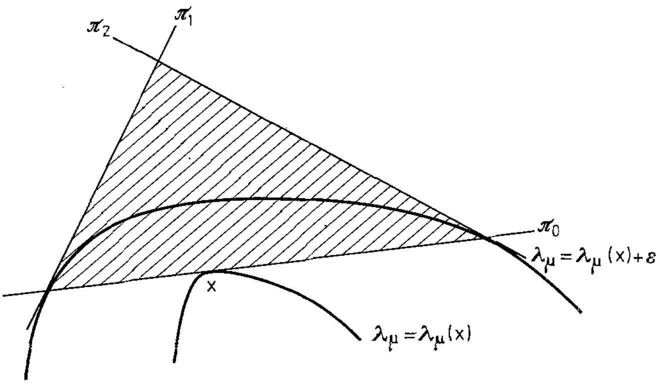
On en déduit immédiatement que $\mathscr{C}$ est stable par réunion finie. Voici maintenant un lemme qui rendra évidente la relation

$$\underline{\sigma}(A) \leqq \inf_{\mathring{A}} \lambda_\mu \qquad \text{pour } A \in \mathscr{B}(\mathbb{X}).$$

(1.14) **Lemme.** *Pour tout $x \in \mathring{S}_\mu$ et pour tout voisinage V de x, on peut trouver un polyèdre T, μ-régulier, contenu dans V, et tel que $\sigma_\mu(T) \leqq \lambda_\mu(x)$.*

Preuve. Un T convenable (hachuré sur la figure) est limité par l'hyperplan π_0 tangent en x à la surface de niveau $\lambda_\mu(\cdot) = \lambda_\mu(x)$, et par un nombre fini d'hyperplans $\pi_1 \dots \pi_r$ tangents à la surface de niveau $\lambda_\mu(\cdot) = \lambda_\mu(x) + \varepsilon$ (ε suffisamment petit) en des points de π_0. Grâce à la stricte convexité de λ_μ en x (hessien défini positif) on peut s'arranger pour que T soit arbitrairement voisin de x (ceci se voit en approchant localement λ_μ par un polynôme de degré 2).

Le fait que T soit μ-régulier et vérifie $\sigma_\mu(T) = \lambda_\mu(x)$ se déduit facilement du fait que la probabilité $P(X_n \in \bigcup_{1 \leqq i \leqq r} \pi_i^+)$ est négligeable devant $P(X_n \in \pi_0^+)$, en utilisant le résultat sur les 1/2 espaces fermés déjà prouvé, la remarque 1.13, et en appelant $\pi_0^+ \dots \pi_r^+$ les $\frac{1}{2}$ espaces fermés convenables limités par $\pi_0 \dots \pi_r$.



(1.15) *Majoration de $\sigma(A)$.* Soit maintenant un convexe fermé A tel que $A \cap \mathring{S}_\mu \neq \emptyset$ et vérifiant 1.7 (i), c'est-à-dire tel que $\mathring{A} \cap \mathring{S}_\mu$ soit non vide. On a vu qu'il existe $\alpha \in A \cap \mathring{S}_\mu$ tel $\lambda_\mu(a) = \inf_A \lambda_\mu$. Puisque λ_μ est continue sur $\mathring{S}_\mu$ et puisque $\mathring{A} \cap \mathring{S}_\mu \neq \emptyset$, il existe une suite x_n tendant vers a, telle que $x_n \in \mathring{A} \cap \mathring{S}_\mu$ et $\lambda_\mu(x_n) \to \lambda_\mu(a)$. D'après le lemme 1.14, il existe des polyèdres μ-réguliers T_n tels que $T_n \subset \mathring{A}$ et tels que $\sigma_\mu(T_n) \leqq \lambda_\mu(x_n)$. La relation 1.1 et l'inclusion $T_n \subset A$ entraînent

$$\underline{\sigma}_\mu(A) < \underline{\sigma}_\mu(T_n) \leqq \lambda_\mu(x_n)$$

d'où quand $n \to +\infty$

$$\underline{\sigma}_\mu(A) < \lambda_\mu(a) = \inf_A \lambda_\mu. \tag{9}$$

Si A vérifie 1.7 (i) et $A \cap S_\mu = \emptyset$ (A convexe fermé), on remplace S_μ par S_F et μ par $\mu(F)^{-1}.\ \mu|_F$ dans le raisonnement précédent, où F est comme dans 1.7 (i). Ainsi (9) est valable pour les A convexes fermés vérifiant 1.7 (i). Comme on a toujours d'après 1.1 $\underline{\sigma}_\mu(A) \geqq \bar{\sigma}_\mu(A)$ les relations (8) et (9) impliquent $\underline{\sigma}_\mu(A) = \bar{\sigma}_\mu(A) = \inf_A \lambda_\mu$ et les ensembles A vérifiant 1.7 (i) sont μ-réguliers.

Si A est maintenant un borélien de $\mathbb{X}$ vérifiant 1.7 (ii), le raisonnement fait ci-dessus pour prouver (9) peut être copié mot pour mot (en utilisant la suite x_n fournie par 1.7 (ii), et prouve que (9) est encore vraie dans ce cas.

En fait l'argument utilisé deux fois pour obtenir (9) prouve encore plus facilement que pour tout borélien A de $\mathbb{X}$, on a

$$\underline{\sigma}_\mu(A) \leqq \inf_{\mathring{A}} \lambda_\mu. \tag{10}$$

(1.16) *Minoration de $\bar{\sigma}_\mu(A)$; cas général.* Admettons provisoirement la s.c. i de $\lambda_\mu(\cdot)$.

Alors si A est un compact de $\mathbb{X}$, il existe pour tout $x \in A$ une boule ouverte B_x de centre x telle que

$$\inf_{\bar{B}_x} \lambda_\mu \geqq (\inf_A \lambda_\mu) - \varepsilon, \qquad \text{où } \varepsilon > 0 \text{ est donné.}$$

On peut évidemment prendre B_x assez petite pour que B_x ne soit tangente à aucune face de S_μ de dimension strictement inférieure à $\dim \mathbb{X}$. D'après 1.15, B_x est alors μ-régulière. Il existe une famille finie $x_1 \dots x_k$ de points de A tels que les $B_{x_1} \dots B_{x_k}$ recouvrent A. Posons $B = B_{x_1} \cup \dots \cup B_{x_k}$.

La classe des ensembles μ-réguliers étant stable par union finie, B est μ-régulier et

$$\sigma_\mu(B) = \inf[\sigma_\mu(B_{x_1}), \dots, \sigma_\mu(B_{x_k})] \geqq (\inf_A \lambda_\mu) - \varepsilon.$$

D'autre part l'inclusion $A \subset B$ et 1.1 entraînent

$$-\bar{\sigma}_\mu(A) \leqq -\bar{\sigma}_\mu(B) = -\sigma_\mu(B) \leqq -(\inf_A \lambda_\mu) + \varepsilon.$$

Ceci étant vrai pour tout $\varepsilon > 0$, on voit que l'inégalité (8) est vraie pour tout A compact de $\mathbb{X}$.

Pour l'étendre au cas non compact, il suffit de faire intervenir un polyèdre K compact, assez grand pour que $(\inf_{\mathbb{X}-K} \lambda_\mu)$ soit supérieur à $(\inf_A \lambda_\mu)$; c'est possible car pour toute forme linéaire ε, les nombres $\inf_{|\varepsilon(x)| \geqq \gamma} \lambda_\mu(x)$ tendent vers $+\infty$ quand $\gamma \to +\infty$.

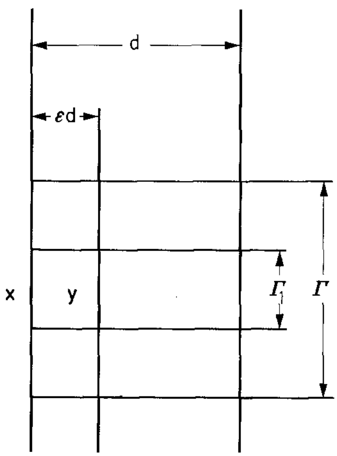
L'ensemble $(\mathbb{X} - K)$, union finie de $\frac{1}{2}$ espaces fermés est μ-régulier et $\sigma_\mu(\mathbb{X} - K)$ est aussi petit qu'on le veut, pour K assez grand.

On a ainsi prouvé l'inégalité (8) pour toute partie borélienne A de $\mathbb{X}$. En particulier (8) et (9) sont vraies pour les A vérifiant 1.7 (ii), ce qui prouve que ces ensembles sont μ-réguliers.

(1.17) *Convexité de* λ_μ. Il nous reste à prouver que λ_μ est convexe et s.c.i., en utilisant seulement 1.7 (i) pour éviter tout cercle vicieux. Pour voir que λ_μ est convexe, il suffit d'étudier λ_μ sur une droite D coupant $\mathring{S}_\mu$, et on sait déjà que λ_μ est convexe sur $D \cap \mathring{S}_\mu$. Soit $x_0 \in D \cap \partial S_\mu$. On a déjà vu, (section 1.11), que $\lambda_\mu(x)$ doit croître lorsque x tend vers x_0 en restant dans $D \cap \mathring{S}_\mu$, et que le minimum de λ_μ sur le segment $[x_0, y]$ avec $y \in D \cap \mathring{S}_\mu$ et voisin de x_0 est atteint en un point de $\mathring{S}_\mu$. On en déduit la convexité de $\lambda_{\mu|D}$ (et, lorsque nous saurons que λ_μ est s.c.i., sa continuité).

(1.18) *Semi-continuité de* $\lambda_\mu(\cdot)$. D'après le lemme 1.4, il suffit de prouver la s.c.i. de $\lambda_\mu(\cdot)$, pour μ fixé, en tout $x \in \partial S_\mu$. Abandonnons provisoirement pour alléger l'écriture, l'indice μ dans λ_μ, σ_μ, S_μ. Nous procédons par récurrence sur la dimension de $\mathbb{X}$, le résultat étant clair en dimension 1. Soit $x \in \partial S$ et choisissons un hyperplan d'appui H de S en x.

L'idée est que $\lambda(x)$ minore à peu près $\sigma(C)$, où C est un voisinage convexe de x dans H, lequel n'est pas beaucoup plus grand que $\sigma(B)$, où B est un convexe intérieur à S mais voisin de x, lequel minore $\lambda(y)$ pour y dans S voisin de x. Donnons seulement, puisqu'il s'agit avant tout d'un exercice de style, la démonstration dans le cas où $\lambda(x)$ est fini, et où μ est à support compact.



Donnons-nous $\Lambda_1 > \Lambda > \Lambda_0 > \exp - \lambda(x)$, arbitrairement voisins de ce dernier; λ étant semi-continue inférieurement sur H, il existe un voisinage convexe fermé γ de x dans H tel que λ reste supérieur à $-\log \Lambda_0$ dans γ; soit Γ un cylindre de directrice γ. Si Γ_1 est un cylindre strictement intérieur à Γ, on peut trouver ε tel que

1°) si, dans une moyenne $\bar{X}_n$, on supprime une proportion de termes inférieure à $3\,\varepsilon$, et si la moyenne initiale appartenait à Γ_1, la moyenne modifiée appartient à Γ.

2°) $2\Lambda\varepsilon$ soit <1.

3°) $\left(\dfrac{e}{2\varepsilon\Lambda}\right)^{2\varepsilon}$ soit voisin de 1.

Une fois ε choisi, on peut trouver m tel que m ainsi que $\left(\dfrac{m\,e}{\varepsilon}\right)^{\varepsilon}$ soient petits devant Λ.

Enfin, choisissons une bande T d'épaisseur d adjacente à H telle que la masse de μ dans $\mathring{T}$ soit inférieure à m.

Soient y et ω un point et un convexe contenus dans Γ_1 et dans la bande adjacente à H d'épaisseur εd. Etant donnée une moyenne $\bar{X}_n \in \omega$, regroupons les X_i en trois paquets: les Y, en nombre p, appartenant à H; les Z, en nombre q, appartenant à $\mathring{T}$, et les autres, W, en nombre r. La probabilité d'avoir $q > n\varepsilon$ est majorée par $\binom{n}{n\varepsilon} m^{n\varepsilon}$, donc par $\left(\left(\frac{m\varepsilon}{\varepsilon}\right)^{\varepsilon}\right)^{n}$ (dans tout ce qui suit, nous utilisons la minoration $p! > (p/e)^p$). Si $q < n\varepsilon$, la moyenne des Y et des W se trouve dans la bande d'épaisseur $2\varepsilon d$ adjacente à H, donc la proportion des W est au plus 2ε; donc, la moyenne des Y appartient à Γ. Ceci permet, pour r assez grand, de majorer $P(\bar{X}_n \in \omega)$ pour des valeurs p, q, r données, $q < n\varepsilon$, $r < 2n\varepsilon$ ipso facto, par

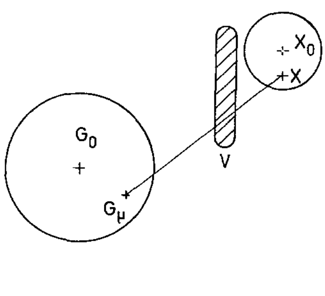
$$\binom{n}{p\ q\ r} \Lambda^p m^q < \Lambda^n \left(\frac{n e m}{q\Lambda}\right)^q \left(\frac{n e}{r\Lambda}\right)^r.$$

Le plus grand $\left(\frac{n e m}{q\Lambda}\right)^q$ possible est obtenu pour $q\Lambda = nm$ et vaut $(e^{m/\Lambda})^n$. Puisque $2\Lambda\varepsilon < 1$, le plus grand $\left(\frac{n e}{r\Lambda}\right)^r$ possible correspond au plus grand r et vaut $\left(\left(\frac{e}{2\varepsilon\Lambda}\right)^{2\varepsilon}\right)^n$. On a donc

$$P(\bar{X}_n \in \omega) < \Lambda^n (n\varepsilon)(2n\varepsilon)\left(e^{m/\Lambda}\left(\frac{e}{2\varepsilon\Lambda}\right)^{2\varepsilon}\right)^n + \left(\left(\frac{m e}{\varepsilon}\right)^{\varepsilon}\right)^n.$$

Grâce aux hypothèses sur m et ε, ceci est inférieur à Λ_1^n, comme ça majore par ailleurs $(\exp - \lambda(y))^n$, on voit que $\lambda(y)$ est presque supérieur à $\lambda(x)$, c.q.f.d.

(1.19) *Semi-continuité de $\lambda_\mu(x)$ vis-à-vis du couple (μ, x).* Soient $\ell < \ell_1 < \lambda_{\mu 0}(x_0)$, arbitrairement proches, G_0 le centre de gravité de μ_0, et G_μ celui de μ, qui reste proche de G_0 d'après la condition sur le moment exponentiel des mesures. On peut trouver un convexe compact V dont l'intérieur coupe le segment $G_0 x_0$, et tel que $\lambda_{\mu_0}(y)$ reste supérieur à ℓ_1 pour $y \in V$. On peut s'arranger pour que V soit intérieur au support de μ_0; il reste alors intérieur au support des μ voisines, et on a $\lambda_\mu(y) > \ell$ pour $y \in V$ et μ voisin de μ_0. Si x est voisin de x_0, et si μ est voisin de μ_0, le segment $G_\mu x$ coupe nécessairement V, et, la fonction λ_μ étant croissante sur ce segment (cf. 1.17), on a $\lambda_\mu(x) > \ell$.



Nous avons donc démontré les propositions 1.5 et 1.7. Remarquons qu'au cours de la preuve, nous avons redémontré le résultat suivant déjà obtenu par Varadhan [4, 13].

(1.20) **Proposition.** *Si A est un borélien de $\mathbb{X}$ et si μ vérifie l'hypothèse 1.2, on a*

$$-\inf_{x\in\mathring{A}}\lambda_\mu(x)<\varliminf\frac{1}{n}\log P(X_n\in A)<\varlimsup\frac{1}{n}\log P(X_n\in A)<-\inf_{x\in A}\lambda_\mu(x).$$

En effet, la première inégalité résulte de (10) et la dernière de (8) qui est valable pour tout borélien A d'après l'étude 1.16. Notons que ces résultats sont moins fins que 1.7 (i), (ii) car par exemple un convexe peut fort bien vérifier 1.7 (i) *sans que* $\inf_{x\in\mathring{A}}\lambda_\mu(x)=\inf_{x\in A}\lambda_\mu(x)$.

Etudions brièvement le comportement de λ_μ lors d'une régularisation de μ par convolution.

(1.21) **Lemme.** *Supposons que μ est à support compact et que S_μ engendre $\mathbb{X}$. Si v_n est une suite de probabilités à support dans un compact fixe et convergeant faiblement vers la masse de Dirac à l'origine, alors pour tout x, $\lambda_\mu(x)$ majore $\lambda_{\mu_* v_n}(x)$ et*

$$\lim_{n\to+\infty}\lambda_{\mu_* v_n}(x)=\lambda_\mu(x).$$

Preuve. D'après la s.c.i. de $\lambda_\mu(x)$ par rapport au couple (μ, x), il suffit de prouver que $\lambda_{\mu_* v_n}(x)$ est majoré par $\lambda_\mu(x)$ pour tout n, et tout x. Commençons par le cas $x\in\mathring{S}_\mu$.

Posons $\log\hat{\mu}=\varphi$, $\log\hat{v}=\psi$; on a $\lambda_\mu(x)=\xi x-\varphi(\xi)$, où ξ est l'unique racine de l'équation $\varphi'(\xi)=x$.

L'équation $\varphi'(\xi+\eta)+\psi'(\xi+\eta)=x$ peut aussi s'écrire

$$\varphi'(\xi+\eta)-\varphi'(\xi)=-\psi'(\xi+\eta),$$

ce qui permettrait de voir que son unique racine η est proche de 0 lorsque v est proche de la masse de Dirac. On a

$$\lambda_{\mu_* v}(x)-\lambda(x)=[\eta\, x-\varphi(\xi+\eta)+\varphi(\xi)]-\psi(\xi+\eta).$$

Le crochet est négatif, puisque φ est convexe (et $\varphi'(\xi)=x$), ainsi que $-\psi$, d'où la propriété désirée.

Si $x\in\partial S_\mu$, soit W un sous-espace de X tel que $\bar{\mu}=\mu|_W$ ait x à l'intérieur de son support, et que $\lambda_\mu(x)=\lambda_{\bar{\mu}}(x)$, lequel vaut par définition $+\infty$ sur $\mathbb{X}-W$, et sur W, $-\log\mu(W)+\lambda_{\bar{\mu}|_{\mu(W)}}$.

Alors, on voit que $\lambda_{\mu_* v}(x)$ est majoré par $\lambda_{\bar{\mu}_* v}(x)$, dont nous allons voir qu'il est inférieur à $\lambda_{\bar{\mu}}(x)$ (insistons sur le fait que nous utilisons passagèrement l'indicatrice λ pour des mesures qui ne sont pas de masse totale 1). On peut reprendre les calculs précédents, mais $\hat{\varphi}$ est maintenant une fonction sur W^*, ξ un vecteur de W^*, et la projection naturelle de $\mathbb{X}^*$ sur W^* est sous-entendue dans les formules. L'expression $\eta\, x-\hat{\varphi}(\xi+\eta)+\hat{\varphi}(\xi)$ ne dépend que de la restriction de η à W, et elle est négative.

2. Robustesse de la loi des grandes deviations

(2.1) *Notations.* $\mathbb{X}$ est un espace euclidien muni de sa norme $|\cdot|$, $\mathscr{B}(\mathbb{X})$ est la σ-algèbre borélienne de $\mathbb{X}$, $M(\mathbb{X})$ est l'ensemble des probabilités sur $\mathbb{X}$. Pour $x\in\mathbb{X}$, $r>0$, on note $B(x,r)$ la boule ouverte de $\mathbb{X}$, de centre x et de rayon r.

Pour $A\subset\mathbb{X}$, $\varepsilon>0$, on pose

$$A^{\varepsilon} = \{x\in X \mid \exists\, y\in A \text{ tel que } |x-y|\leqq\varepsilon\}$$
$$A^{-\varepsilon} = \{x\in A \mid \forall\, y\notin A \text{ on a } |x-y|\geqq\varepsilon\}.$$

La convergence faible sur $M(\mathbb{X})$ peut être définie par la distance de Prokhorov d, qui vérifie «$d(\mu,\nu)\leqq\varepsilon$ avec $\mu,\nu\in M(\mathbb{X})$, $\varepsilon>0$ si et seulement si

$$\mu(A)\leqq\nu(A^{\varepsilon})+\varepsilon, \qquad \nu(A)\leqq\mu(A^{\varepsilon})+\varepsilon \qquad \text{pour tout } A\in\mathscr{B}(X)».$$

Si $\mu\in M(\mathbb{X})$, on note F_μ le support de μ et S_μ l'enveloppe convexe fermée de F_μ.

Dans ce paragraphe on considère un espace measurable $(\Omega,\mathscr{B}(\Omega))$ muni d'une suite croissante $(\mathscr{F}_n)\,n\geqq 0$ de sous σ-algèbres de $\mathscr{B}(\Omega)$, avec $\mathscr{F}_0$ triviale, telle que $\bigcup_{n\geqq 0}\mathscr{F}_n$ engendre $\mathscr{B}(\Omega)$. On se donne une suite $(X_n)\,n\geqq 1$ de variables aléatoires sur Ω à valeurs dans $\mathbb{X}$, telle que pour tout $n\geqq 1$ X_n soit $\mathscr{F}_n$-mesurable. On note $M(\Omega)$ l'espace des probabilités sur $(\Omega,\mathscr{B}(\Omega))$. On suppose que pour tout $n\geqq 1$, pour tout $Q\in M(\Omega)$, il existe un noyau $Q_n(\omega,B)$ où $(\omega,B)\in\Omega\times\mathscr{B}(\Omega)$, qui est $\mathscr{F}_{n-1}$-mesurable en ω pour B fixe, σ-additif en B pour ω fixe, vérifiant pour tout $B\in\mathscr{B}(\Omega)$

$$Q(B|\mathscr{F}_{n-1})=Q_n(\cdot,B), \qquad Q\text{-p.s.}$$

On note $Q_{n,\omega}\in M(\mathbb{X})$ la loi conditionnelle de X_n étant donné $\mathscr{F}_{n-1}$ définie par

$$Q_{n,\omega}(A)=Q_n(\omega, X_n^{-1}(A)) \qquad \text{pour } n\geqq 1,\ A\in\mathscr{B}(\mathbb{X}).$$

On suppose que si $P_{n,\omega}\equiv Q_{n,\omega}$ pour tout $n\in\mathbb{N}$, $\omega\in\Omega$, alors $P=Q$.

On suppose enfin que pour toute probabilité $\mu\in M(\mathbb{X})$ il existe $Q\in M(\Omega)$ telle que $Q_{n,\omega}\equiv\mu$ pour tout $n\geqq 1$ et Q-presque tout $\omega\in\Omega$. Une telle loi Q sera notée P_μ; sous P_μ, les X_n sont indépendantes et de même loi μ.

L'existence de tels triplets $\{\Omega,(\mathscr{F}_n)_{n\geqq 0},(X_n)_{n\geqq 1}\}$ est évidente: il suffit par exemple de munir $\mathbb{X}^{\mathbb{N}}$ de la σ-algèbre produit et de ses projections naturelles sur $\mathbb{X}$, et de définir $\mathscr{F}_n\,n\geqq 1$ comme la σ-algèbre engendrée par les n premières projections.

On pose $\bar{X}_n=\frac{1}{n}(X_1+\cdots+X_n)$. Pour $\mu\in M(\mathbb{X})$ le comportement asymptotique de $\frac{1}{n}\log P_\mu(\bar{X}_n\in A)$ quand $n\to+\infty$ a été étudié au paragraphe précédent. Nous allons préciser le type de petites perturbations de l'indépendance et de l'équidistribution des X_n qui affectent peu ce comportement asymptotique.

(2.2) *Définition.* Si V est une partie de $M(\mathbb{X})$, nous notons $V^{\mathbb{N}}$ l'ensemble des $Q\in M(\Omega)$ tels que pour tout $n\geqq 0$ et pour Q-presque tout $\omega\in\Omega$ on ait $Q_{n,\omega}\in V$.

(2.3) **Théorème.** *Soit $\mu\in M(\mathbb{X})$ une probabilité à support compact, d'enveloppe convexe S_μ. Pour $\delta>0$, notons $V(\delta)$ l'ensemble des $\nu\in M(\mathbb{X})$ tels que $d(\nu,\mu)\leqq\delta$*

et $S_\nu \subset (S_\mu)^\delta$. *Pour tout* $\varepsilon > 0$, *il existe* $\delta > 0$ *et* $N \geqq 1$ *tels que si* $Q \in M(\Omega)$ *est dans* $V(\delta)^{\mathbb{N}}$, *alors elle vérifie*

$$\begin{cases} Q(\bar{X}_n \in A) \leqq e^{\varepsilon n} P_\mu(\bar{X}_n \in A^\varepsilon) \\ P_\mu(\bar{X}_n \in A) \leqq e^{\varepsilon n} Q(\bar{X}_n \in A^\varepsilon) \end{cases} \tag{1}$$

quels que soient $n \geqq N$ *et* $A \in \mathscr{B}(\mathbb{X})$.

(2.4) **Théorème.** *Soient* $\mu \in M(\mathbb{X})$ *et* $\beta > 1$ *tels que* $\int e^{|x|^\beta} d\mu(x) < +\infty$. *Pour* $\delta > 0$, $L > 0$ *notons* $W(\delta, L)$ *l'ensemble des* $\nu \in M(\mathbb{X})$ *tels que* $d(\nu, \mu) \leqq \delta$ *et* $\int e^{|x|^\beta} d\nu(x) \leqq L$.

Pour tout $\varepsilon > 0$, $L > 0$, $I > 0$, *on peut trouver* $\delta > 0$, *et on peut à tout* $A \in \mathscr{B}(\mathbb{X})$ *vérifiant* $\underline{\sigma}_\mu(A^\varepsilon) \leqq I$ *associer un entier* N, *de telle sorte que si* $Q \in M(\Omega)$ *est dans* $W(\delta, L)^{\mathbb{N}}$, *les inégalités* (1) *soient satisfaites pour tout* $n \geqq N$. [*cf.* 1.1 *pour la définition de* $\underline{\sigma}_\mu$].

Les théorèmes 2.3 et 2.4. vont être prouvés simultanément, en plusieurs étapes.

(2.5) *Effet d'une troncature.* Donnons-nous $\beta > 1$, $L > 0$, $\eta_0 > 0$ et des boréliens $K \subset F \subset \mathbb{X}$ avec K borné.

Soit U l'ensemble des $\nu \in M(\mathbb{X})$ tels que

$$\int e^{|x|^\beta} d\nu(x) \leqq L, \qquad \nu(F) \geqq 1 - \eta_0, \qquad \nu(K) \geqq \tfrac{1}{2}. \tag{2}$$

Soit $R > 0$ assez grand pour que $B(0, R) \supset K$. De (2) on tire pour $k \geqq 1$, $\nu \in U$

$$\nu[\mathbb{X} - B(0, kR)] \leqq L e^{-(kR)^\beta} = \eta_k. \tag{3}$$

Soit $\mathscr{L}$ l'ensemble des suites $\Lambda = (\Lambda_j)_{j \geqq -1}$ de parties de $\{1, \dots, n\}$ telles que les Λ_j non-vides forment une partition de $\{1, \dots, n\}$. Définissons une variable aléatoire $J: \Omega \to \mathscr{L}$ par $J = (J_j)_{j \geqq -1}$, en imposant pour $j \geqq -1$ et $k \geqq 1$,

$$\begin{aligned} &j \in J_{-1} \quad \text{si } X_j \in F \cap B(0, R) = F_R \\ &j \in J_0 \quad \text{si } X_j \in B(0, R) - F \\ &j \in J_k \quad \text{si } X_j \in B[0, (k+1)R] - B(0, kR). \end{aligned}$$

Soit $\Lambda \in \mathscr{L}$ et posons $v_i = \mathrm{Card}\,(\Lambda_i)$ de sorte que $\sum_{j \geqq -1} v_j = n$.

Soit $Q \in M(\Omega)$ un élément de $U^{\mathbb{N}}$. De (2), (3) et de la définition 1 on déduit

$$Q(X_j \in A_j \text{ pour } 1 \leqq j \leqq n, J = \Lambda) \leqq \Big(\prod_{k \geqq 0} \eta_k^{v_k}\Big) Q(X_j \in A_j \cap F_R,\ j \in \Lambda_{-1})$$

quels que soient les $A_j \in \mathscr{B}(\mathbb{X})$. De même (2) donne

$$(\tfrac{1}{2})^{n - v_{-1}} Q(X_j \in A_j \cap F_R, j \in \Lambda_{-1}) \leqq Q(X_j \in A_j, j \in \Lambda_{-1};\ X_k \in F_R, 1 \leqq k \leqq n).$$

On rapproche les inégalités obtenues, et à l'aide d'un argument standard (approximation de fonctions par des combinaisons d'indicatrices de «rectangles»), on obtient

$$E_Q[f I_{\{J = \Lambda\}}] \leqq a(\Lambda) E_Q[f \prod_{1 \leqq j \leqq n} I_{F_R}(X_j)] \tag{4}$$

où f est une fonction borélienne positive $f(X_1, \ldots, X_n)$ arbitraire *ne dépendant que* des $(X_j)_{j \in \Lambda_{-1}}$, où $a(\Lambda) = \prod_{k \geqq 0} (2\eta_k)^{v_k}$ et où E_Q désigne l'intégrale par rapport à Q.

Soit $\varepsilon > 0$, et $\mathscr{V}_\varepsilon$ l'ensemble des $\Lambda \in \mathscr{L}$ vérifiant

$$\sum_{k \geqq 0} (1+k) v_k R \leqq \varepsilon n. \tag{5}$$

Si $J = \Lambda \in \mathscr{V}_\varepsilon$, et si $A \in \mathscr{B}(\mathbb{X})$, la relation $(\bar{X}_n \in A)$ implique $\frac{1}{n}\big(\sum_{j \in \Lambda_{-1}} X_j\big) \in A^\varepsilon$, qui, sur l'ensemble $\{|X_j| \leqq R, 1 \leqq j \leqq n\} \subset \Omega$, implique elle-même $\bar{X}_n \in A^{2\varepsilon}$. De (4) et des remarques précédentes on déduit, pour $A \in \mathscr{B}(\mathbb{X})$,

$$\begin{aligned} Q(\bar{X}_n \in A;\, J \in \mathscr{V}_\varepsilon) &\leqq \sum_{\Lambda \in \mathscr{V}_\varepsilon} Q\left(\frac{1}{n} \sum_{j \in \Lambda_{-1}} X_j \in A^\varepsilon;\, J = \Lambda\right) \\ &\leqq \sum_{\Lambda \in \mathscr{V}_\varepsilon} a(\Lambda)\, Q\left\{\frac{1}{n}\Big(\sum_{j \in \Lambda_{-1}} X_j\Big) \in A^\varepsilon;\, X_j \in F_R \text{ pour } 1 \leqq j \leqq n\right\} \\ &\leqq \Big[\sum_{\Lambda \in \mathscr{V}_\varepsilon} a(\Lambda)\Big]\, Q\{\bar{X}_n \in A^{2\varepsilon};\, X_j \in F_R \text{ pour } 1 \leqq j \leqq n\}. \end{aligned} \tag{6}$$

Soit V l'ensemble des suites d'entiers $v = (v_i)_{i \geqq -1}$ telles que $v_i \geqq 0$, $\sum_{i \geqq -1} v_i = n$; soit V_ε l'ensemble des $v \in V$ vérifiant (5). Pour $v \in V$ posons

$$b(v) = \frac{n!}{\prod_{k \geqq -1} (v_k!)} \prod_{k \geqq 0} \eta_k^{v_k}. \tag{7}$$

La majoration brutale

$$Q(\bar{X}_n \in A;\, J = \Lambda) \leqq Q(J = \Lambda) \leqq \prod_{k \geqq 0} \eta_k^{v_k}$$

fournit trivialement l'inégalité

$$Q\bar{X}_n \in A;\, J \notin \mathscr{V}_\varepsilon) \leqq \sum_{v \notin V_\varepsilon} b(v) \tag{8}$$

On a évidemment:

$$\sum_{\Lambda \in \mathscr{V}_\varepsilon} a(\Lambda) \leqq \sum_{\Lambda \in \mathscr{L}} a(\Lambda) = \sum_{v \in V} \left[\frac{n!}{\prod_{k \geqq -1} (v_k!)} \prod_{k \geqq 0} (2\eta_k)^{v_k}\right] = \Big[1 + 2\Big(\sum_{k \geqq 0} \eta_k\Big)\Big]^n. \tag{9}$$

Soit $g(R)$ la fonction $\sup_{k \geqq 1} \frac{1}{(k+1)R} \operatorname{Log} \eta_k$ qui d'après (3) ne dépend que de L et β, et tend vers $-\infty$ quand $R \to +\infty$.

Posons $\gamma(R) = g(R) \vee \frac{1}{R} \operatorname{Log} \eta_0$, et imposons $\gamma(R) < 0$. Si $v \notin V_\varepsilon$ on a alors

$$\prod_{k \geqq 0} \eta_k^{v_k} = \Big(\prod_{k \geqq 0} \eta_k^{\frac{1}{2} v_k}\Big) \prod_{k \geqq 0} \Big(\eta_k^{\frac{1}{(k+1)R}}\Big)^{\frac{1}{2} v_k (k+1) R} \leqq \Big(\prod_{k \geqq 0} \eta_k^{\frac{1}{2} v_k}\Big)\, e^{\frac{1}{2} \gamma(R) \varepsilon n}.$$

Par suite

$$\sum_{v \notin V_\varepsilon} b(v) \leqq e^{\frac{1}{2} \gamma(R) \varepsilon n} \sum_{v \in V} \left[\frac{n!}{\prod_{k \geqq -1} (v_k!)} \prod_{k \geqq 0} \eta_k^{\frac{1}{2} v_k}\right] \leqq e^{\frac{1}{2} \gamma(R) \varepsilon n} \Big(1 + \sum_{k \geqq 0} \eta_k^{\frac{1}{2}}\Big)^n. \tag{10}$$

Etant donnés $\varepsilon>0$ et $I>0$; (9), (10) et (3) permettent de déterminer $a>0$ tel que les conditions $\eta_0 \leqq a$ et $R \geqq \frac{1}{a}$ impliquent

$$\sum_{A \in \mathscr{V}_\varepsilon} a(A) \leqq e^{\varepsilon n} \qquad \sum_{v \notin V_\varepsilon} b(v) \leqq e^{-In}.$$

De (8) et (6) on déduit alors que

$$Q(\bar{X}_n \in A) \leqq e^{\varepsilon n} Q(\bar{X}_n \in A^{2\varepsilon};\ X_j \in F_R,\ 1 \leqq j \leqq n) + e^{-In} \tag{11}$$

dès que $R \geqq 1/a$, $\eta_0 \leqq a$, et $Q \in U^{\mathbb{N}}$.

(2.6) *Effet d'une perturbation de P_μ sur la loi de $\bar{X}_n$ «tronquée».* Soit $\varepsilon>0$ et soit G un compact donné de $\mathbb{X}$.

Soit $(H_n)_{n\in\mathbb{N}}$ une famille d'hyperplans affines de $\mathbb{X}$, tel que chaque H_n soit orthogonal à l'un des «axes de coordonnée» de $\mathbb{X}$. On peut choisir les H_n de telle sorte que la famille $\mathscr{P}$ des composantes connexes de $(\mathbb{X} - \bigcup_n H_n)$ ait les propriétés suivantes: a) chaque pavé $C \in \mathscr{P}$ est de diamètre inférieur à ε, et $C^{-\varepsilon/2} \neq \emptyset$; b) si $C \in \mathscr{P}$ et $\bar{C} \cap G \neq \emptyset$ alors $C \cap G$ est non vide; la propriété b) est réalisée par un nombre fini de *petites* perturbations successives des H_n rencontrant G, qui laissent évidemment «presque invariantes» les familles $\mathscr{P}_1$ (resp. $\mathscr{P}_2$) des $C \in \mathscr{P}$ qui rencontrent G (resp. tels que $\bar{C} \cap G = \emptyset$). Si $\mu \in M(X)$ est donnée, on peut aussi imposer aux H_n d'être μ_k-négligeables pour tout $k \geqq 1$, où μ_k est la loi de $\bar{X}_k$ sous P_μ.

Donnons nous μ, ε et $R>0$; appliquons cette construction avec

$G = \text{Support}(\mu) \cap B(0, R-\varepsilon)$,

et fixons $\mathscr{P}$. L'union F des $C \in \mathscr{P}_1$ est un voisinage de G, contenu dans $B(0,R)$.

Les $\mu(C)$, $C \in \mathscr{P}_1$ forment un ensemble fini fixé de nombres strictement positifs. D'autre part puisque les $C \in \mathscr{P}$ sont de frontières μ-négligeables on a

$$\lim_{\delta \to 0} \mu(C^\delta) = \lim_{\delta \to 0} \mu(C^{-\delta}) = \mu(C).$$

On en déduit l'existence de $\delta_0>0$ tel que si $v \in M(\mathbb{X})$ vérifie $d(v,\mu) \leqq \delta_0$, alors

$$e^{-\varepsilon} \mu(C) \leqq v(C) \leqq e^{\varepsilon} \mu(C) \qquad \text{pour } C \in \mathscr{P}_1. \tag{12}$$

Soit $\mathscr{R}_n$ la σ-algèbre de parties de Ω engendrée par les $X_j^{-1}(\mathscr{P})$, $1 \leqq j \leqq n$. Soit $Q \in M(\Omega)$ un élément de $B_\delta^{\mathbb{N}}$ où $B_\delta = \{v \in M(\mathbb{X}) \colon d(v,\mu) \leqq \delta\}$ avec $\delta \leqq \delta_0$.

Soit $\Omega_F = \{X_j \in F,\ 1 \leqq j \leqq n\}$. De (12) on conclut facilement que

$$e^{-n\varepsilon} \int I_{\Omega_F} f\, dP_\mu \leqq \int I_{\Omega_F} f\, dQ \leqq e^{n\varepsilon} \int I_{\Omega_F} f\, dP_\mu \tag{13}$$

pour toute fonction positive *$\mathscr{R}_n$-mesurable* $f \colon \Omega \to \mathbb{R}^+$. Définissons des variables aléatoires $\mathscr{R}_n$-mesurables $Y_j \colon \Omega \to \mathbb{X}$ $1 \leqq j \leqq n$, en posant, pour $C \in \mathscr{P}$, Y_j = centre du pavé C sur l'ensemble $\{X_j \in C\}$. On pose $\bar{Y}_n = \frac{1}{n}(Y_1 + \cdots + Y_n)$, de sorte que $|\bar{Y}_n - \bar{X}_n| \leqq \varepsilon$. Pour $A \in \mathscr{B}(X)$, on obtient en utilisant (13)

$$\begin{aligned} Q\{(\bar{X}_n \in A) \cap \Omega_F\} &\leqq Q\{(\bar{Y}_n \in A^\varepsilon) \cap \Omega_F\} \leqq e^{n\varepsilon} P_\mu\{(\bar{Y}_n \in A^\varepsilon) \cap \Omega_F\} \\ &\leqq e^{n\varepsilon} P_\mu\{(\bar{X}_n \in A^{2\varepsilon}) \cap \Omega_F\} \end{aligned}$$

et de même

$$P_\mu[(\bar{X}_n \in A)\cap \Omega_F] \leqq e^{n\varepsilon}\, Q[(\bar{X}_n \in A^{2\varepsilon})\cap \Omega_F]. \tag{15}$$

(2.7) *Preuve du Théorème 2.4.* Comme dans l'énoncé du th. 2.4, donnons nous μ, L, I, ε, définissons $W(\delta, L)$ pour un δ à déterminer, et soit $Q \in W(\delta, L)^{\mathbb{N}}$. On détermine $a = a(\mu, L, I, \varepsilon) < \frac{1}{2}$ comme indiqué à la fin de 2.5. On fixe $R \geqq 1/a$ et assez grand pour que $G = S_\mu \cap B(0, R-\varepsilon)$ vérifie $\mu(G) \geqq 1 - a/2$. On fixe les pavés $\mathscr{P}$ comme en 2.6. L'ensemble F de 2.6. est un voisinage de G, donc pour δ petit, et $\nu \in W(\delta, L)$ on a $\nu(F) \geqq 1 - a \geqq 1/2$. Considérons la partie U de $M(\mathbb{X})$ définie par (2) avec $K = F$ et $\eta_0 = a$. Quitte à augmenter L, on peut supposer que $\mu \in U$. Ainsi Q et P_μ sont dans $U^{\mathbb{N}}$, et on peut appliquer (11) à Q et à P_μ, en notant que $F_R = F$. D'après 2.6., pour δ petit, (14) et (15) sont vraies. De (11), (14), (15), on tire, pour tout $n \geqq 1$,

$$\begin{aligned} Q(\bar{X}_n \in A) &\leqq e^{2\varepsilon n}\, P_\mu(\bar{X}_n \in A^{4\varepsilon}) + e^{-In} \\ P_\mu(\bar{X}_n \in A) &\leqq e^{2\varepsilon n}\, Q(X_n \in A^{4\varepsilon}) + e^{-In}. \end{aligned} \tag{16}$$

Le théorème 2.4 se déduit facilement de (16), et de la définition de $\underline{\sigma}_\mu(A)$ donnée en 1.1.

(2.8) **Lemme.** *Soient $n \geqq 1$ et $\pi_n: \Omega \to \mathbb{X}^n$ la projection naturelle $\pi_n = (X_1, \ldots, X_n)$. Pour $\delta > 0$ soit $B_\delta = \{\nu \in M(\mathbb{X}): d(\mu, \nu) \leqq \delta\}$.*

Pour $\varepsilon > 0$, il existe $\delta > 0$ tel que si $Q \in M(\Omega)$ est dans $B_\delta^{\mathbb{N}}$, alors la distance de Prokhorov — dans $M(\mathbb{X}^n)$ — de $\pi_n(Q)$ et $\pi_n(P_\mu)$ est inférieure à ε.

Preuve. Etant donnés n, μ, ε, choisissons R assez grand pour que — notations de 2.6 — l'on ait:

$$P_\mu(\Omega_F) = \mu(F)^n \geqq 1 - \varepsilon/2.$$

Pour δ assez petit on aura $\nu_1(F) \ldots \nu_n(F) \geqq 1 - \varepsilon$ quels que soient $\nu_1 \ldots \nu_n$ dans B_δ, et donc $Q(\Omega_F) \geqq 1 - \varepsilon$ pour $Q \in B_\delta^{\mathbb{N}}$.

Soit $g: \mathbb{X}^n \to [0, 1]$ une fonction continue quelconque fixée. Appliquons la construction de 2.6 en remplaçant ε par α/n avec $\alpha \leqq \varepsilon$. Avec les notations de 2.6, si α est assez petit, il existe $f: \Omega \to [0, 1]$, $\mathscr{R}_n$-mesurable, à support dans Ω_F telle que $|f - I_{\Omega_F} \cdot g \circ \pi_n| \leqq \varepsilon$. Ceci implique que $|\int f\, dP_\mu - \int g \circ \pi_n\, dP_\mu|$ et $|\int f\, dQ - \int g \circ \pi_n\, dQ|$ sont majorés par ε. Pour δ assez petit et Q dans $B_\delta^{\mathbb{N}}$, (13) fournit, puisque $f = f I_{\Omega_F}$,

$$e^{-\alpha} \int f\, dP_\mu \leqq \int f\, dQ \leqq e^{\alpha} \int f\, dP_\mu$$

On en déduit que $|\int g \circ \pi_n\, dP_\mu - \int g \circ \pi_n\, dQ| \leqq 4\varepsilon$.

Sur $M(\mathbb{X}^n)$ on sait que la distance de Prokhorov est équivalente à la distance

$$D(\eta, \eta') = \sum_{k=1}^{+\infty} \frac{1}{2^k}\, [1 \wedge |\int g_k\, d\eta - \int g_k\, d\eta'|]$$

où $\eta, \eta' \in M(\mathbb{X}^n)$, et où g_k est une suite dense dans l'espace des fonctions continues bornées sur $\mathbb{X}^n$, ce qui prouve le lemme.

(2.9) *Effet d'une moyenne partielle.* Fixons un entier $p \geqq 1$ et posons $\Omega' = \Omega$, $\mathscr{B}(\Omega') = \mathscr{B}(\Omega)$, $\mathscr{F}'_n = \mathscr{F}_{np}$ pour $n \geqq 0$ et

$$X'_n = \frac{1}{p}(X_{(n-1)p+1} + X_{(n-1)p+2} + \cdots + X_{np}) \qquad \text{pour } n \geqq 1.$$

Clairement le triplet $\{\Omega', (\mathscr{F}'_n)_{n \geqq 0}, (X'_n)_{n \geqq 1}\}$ possède les propriétés énoncées dans la section 2.1.

Appliquons systématiquement à ce triplet et à $M(\Omega')$ les notations et définitions des sections 2.1 et 2.2. En particulier si j est l'application identique de $M(\Omega)$ dans $M(\Omega')$, on peut écrire pour $\mu \in M(\mathbb{X})$, $j(P_\mu) = P'_{\mu_p}$ où $\mu_p \in M(\mathbb{X})$ est la loi de $\bar{X}_p$ sous P_μ.

Soit $V(\delta) = \{v \in M(\mathbb{X}) \colon d(v, \mu) \leqq \delta$ et $S_v \subset S^\delta_\mu\}$. Soit $Q \in M(\Omega)$ un élément de $V(\delta)^{\mathbb{N}}$ et soit $Q' = j(Q)$. Soit $R = P_\mu \in M(\Omega)$.

Définissons $\pi_n \colon \Omega \to \mathbb{X}^p$ par $\pi_n = (X_{(n-1)p+1}, \ldots, X_{np})$. Soit $q_{n,\omega} \in M(\mathbb{X}^p)$ la loi jointe conditionnelle définie par

$$q_{n,\omega}(A) = Q_{(n-1)p}(\omega, \pi_n^{-1}(A)), \qquad A \in \mathscr{B}(\mathbb{X}^p).$$

Etant donné $\delta' > 0$, on peut (d'après le lemme 2.8) trouver $\delta \leqq \delta'$ assez petit pour que la distance de Prokhorov de $q_{n,\omega}$ et $\mu^{\otimes p}$ soit inférieure à δ', pour tout $n \geqq 1$ et Q-p.tt.$\omega \in \Omega$. Les images de $q_{n,\omega}$ et $\mu^{\otimes p}$ par l'application *contractante* $(x_1 \ldots x_p) \to \frac{1}{p}(x_1 + \cdots + x_p)$ de $\mathbb{X}^p$ dans $\mathbb{X}$ ont encore la même propriété.

Appliquant les définitions 2.1 et 2.2 au triplet $\{\Omega' \ldots\}$, on traduit ce fait par $d(Q'_{n,\omega}, \mu_p) \leqq \delta'$ pour tout $n \geqq 1$ et Q'-p.tt.$\omega \in \Omega'$. On en conclut immédiatement que $Q' \in M(\Omega')$ est un élément de $W(\delta')^{\mathbb{N}}$ où $W(\delta') = \{v \in M(\mathbb{X}) \colon d(v, \mu_p) \leqq \delta'$ et $S_v \subset (S_{\mu_p})^{\delta'}\}$.

La loi de $\bar{X}'_n$ sous Q' est identique à la loi de $\bar{X}_{np}$ sous Q. Si les supports des $Q_{n,\omega}$ sont tous contenus dans un compact fixe on a donc pour p/n assez petit

$$\begin{cases} Q(\bar{X}_n \in A) \leqq Q'(\bar{X}'_{[n/p]} \in A^\varepsilon) \\ Q'(\bar{X}'_{[n/p]} \in A) \leqq Q(\bar{X}_n \in A^\varepsilon) \end{cases} \tag{17}$$

quel que soit $A \in \mathscr{B}(\mathbb{X})$. Donc pour prouver le théorème 2.3 on peut remplacer Q par Q' et μ par μ_p, avec $p \geqq 1$ fixé arbitraire.

Montrons que si μ est à support compact, et si $\alpha > 0$ est donné, on a $(F_{\mu_p})^\alpha \supset S_\mu$ pour p assez grand (F_v est le support de v). Fixons une partition finie de S_μ en boréliens $B_1 \ldots B_n$ de diamètre inférieur à α, et des points $b_i \in B_i$. Si v est une probabilité portée par S_μ, on pose $\tilde{v} = \sum_i v(B_i)\, \delta_{b_i}$; la distance entre les barycentres de v et $\tilde{v}$ est inférieure à α. On en déduit que $S_\mu \subset (S_{\tilde{\mu}})^\alpha$.

On vérifie immédiatement que $(F_{\mu_p})^\alpha \supset F_{(\tilde{\mu})_p}$ pour tout $p \geqq 1$; Si on montre que $[F_{(\tilde{\mu})_p}]^\alpha \supset S_{\tilde{\mu}}$, les deux inclusions obtenues impliquent $(F_{\mu_p})^{3\alpha} \supset [F_{(\tilde{\mu})_p}]^{2\alpha} \supset (S_{\tilde{\mu}})^\alpha \supset S_\mu$. Mais $\tilde{\mu}$ étant à support fini, l'inclusion $[F_{(\tilde{\mu})_p}]^\alpha \supset S_{\tilde{\mu}}$ pour p assez grand est évidente, ce qui prouve notre assertion.

(2.10) *Preuve du théorème 2.3.* Soit $\mu \in M(\mathbb{X})$, à support compact F_μ.

Choisissons le pavage $\mathscr{P}$ comme en 2.6, *mais avec* $G = \mathrm{S}_\mu$. On ne peut plus à priori procéder comme en 2.6 car les pavés $C \in \mathscr{P}_1$ ne vérifiant plus $\mu(C) > 0$.

Comme $\mathscr{P}$ est localement fini, il existe $a>0$ tel que $2a \leqq |x-y|$ pour tout $x \in S_\mu$, $y \in C \subset \mathscr{P}_2$. Choisissons d'après 2.9 p entier assez grand pour que $(F_{\mu_p})^a \supset S_\mu$, et tel que tout $C \in \mathscr{P}_1$ contienne une boule de centre appartenant à S_μ et de rayon $2a$. Choisissons p entier assez grand pour que $(F_{\mu_p})^a \supset S_\mu$.

Tout $C \in \mathscr{P}_1$ rencontre alors F_{μ_p}, et vérifie donc $\mu_p(C)>0$. Avec les notations de 2.9, on peut (p est fixé) appliquer le raisonnement de 2.6 à $\mu'=\mu_p$ et $Q'=j(Q)$ et obtenir (14) (15) pour Q' et $P'_{\mu'}$ dès que $Q' \in M(\Omega')$ est dans $W(\delta')^{\mathbb{N}}$. Mais (notations de 2.6) F est un voisinage de S_μ, (construit à partir de ε et μ), et donc, pour δ' assez petit on a $Q'(\Omega'_F)=P'_{\mu'}(\Omega'_F)=1$, de sorte que (14), (15) deviennent

$$P'_{\mu'}(\bar{X}'_n \in A) \leqq e^{n\varepsilon}\, Q'(\bar{X}'_n \in A^{2\varepsilon}) \quad \text{et} \quad Q'(\bar{X}'_n \in A) \leqq e^{n\varepsilon}\, P'_{\mu'}(\bar{X}'_n \in A^{2\varepsilon}).$$

Il suffit de tenir compte des conclusions de 2.9, en notant que (17) s'applique aussi bien à (Q, Q') qu'à $(P_\mu, P'_{\mu'})$, pour obtenir la preuve du théorème 2.3.

(2.11) *Définition.* Dénotons $h(A,B)$ la *distance de Hausdorff* sur l'ensemble des parties de $\mathbb{X}$, définie par

$$h(A,B)=\inf\{a>0\colon A \subset B^a \text{ et } B \subset A^a\}.$$

Soit $M_c(\mathbb{X})$ l'ensemble des probabilités sur $\mathbb{X}$ qui sont à support compact. Munissons $M_c(\mathbb{X})$ de *la topologie* (T) définie comme suit: $\mu \xrightarrow{(T)} \nu$ si μ tend vers ν faiblement et si $h(S_\mu, S_\nu)$ tend vers 0.

Soit $(\Omega, \mathscr{B}(\Omega))$ un espace mesurable comme en 2.1. Pour $P, Q \in M(\Omega)$ considérons l'ensemble des $\varepsilon>0$ ayant la propriété suivante: il existe des versions $P_{n,\omega}$, $Q_{n,\omega}$, où $n \in \mathbb{N}$, $\omega \in \Omega$, des lois conditionnelles de X_n étant donné $\mathscr{F}_{n-1}$ telles que $d(P_{n,\omega}, Q_{n,\omega}) \leqq \varepsilon$ pour tout $n \in \mathbb{N}$ et *tout* $\omega \in \Omega$.

Soit $\varepsilon(P,Q)$ la borne inférieure de l'ensemble de ces $\varepsilon>0$. Il est clair que $d_{\mathbb{N}}(P,Q)=1 \wedge \varepsilon(P,Q)$ définit une distance sur $M(\Omega)$.

(2.12) **Corollaire.** *Soit $E \subset M_c(\mathbb{X})$ une partie compacte pour la topologie (T). Alors pour tout $\varepsilon>0$, il exists $\delta>0$ et $N \geqq 1$ tels que le théorème 2.3 soit simultanément valable pour tous les $\mu \in E$.*

Preuve. Il s'agit là d'un résultat de robustesse uniforme en $\mu \in E$. Munissons l'ensemble $\mathscr{A}$ des applications de $\mathbb{N} \times \mathscr{B}(\mathbb{X})$ dans $[-\infty, 0]$ d'une distance D en posant $D(f,g) \leqq \varepsilon$ si et seulement si il existe k entier tel que pour tout $n \geqq k$, $A \in \mathscr{B}(\mathbb{X})$, on ait

$$f(n,A) \leqq g(n, A^\varepsilon)+\varepsilon, \qquad g(n,A) \leqq f(n, A^\varepsilon)+\varepsilon.$$

Si $Q \in M(\Omega)$ on lui associe $f_Q \in \mathscr{A}$ en posant $f_Q(n,A)=\dfrac{1}{n}\log Q(\bar{X}_n \in A)$.

Le théorème 2.3 s'énonce aussi bien sous la forme suivante; «si on munit respectivement $M(\Omega)$ et $\mathscr{A}$ des distances $d_{\mathbb{N}}$ et D, et si $\mu \in M_c(\mathbb{X})$, l'application $Q \to f_Q$ de $M(\Omega)$ dans $\mathscr{A}$ est continue au point $P_\mu \in M(\Omega)$».

D'autre part, $M(\Omega)$ et $M_c(\mathbb{X})$ étant respectivement munie de la distance $d_{\mathbb{N}}$ et de la topologie (T), l'application $\mu \to P_\mu$ de $M_c(\mathbb{X})$ dans $M(\Omega)$ est trivialement continue. L'ensemble $F=\{P_\mu\colon \mu \in E\}$ est donc $d_{\mathbb{N}}$-compact dans $M(\Omega)$. Par suite, pour tout $\varepsilon>0$, il existe $\delta>0$ tel que les relations $d_{\mathbb{N}}(Q, Q_0) \leqq \delta$ et $Q_0 \in F$ entraînent $D(f_Q, f_{Q_0}) \leqq \varepsilon$. Q.E.D.

3. Variations lipschitziennes des lignes de niveau de λ_μ

(3.1) *Notations.* Soit $M_c(\mathbb{X})$ l'ensemble des probabilités μ sur l'espace euclidien $\mathbb{X}$, telles que l'enveloppe convexe fermée S_μ du support de μ soit compacte et engendre $\mathbb{X}$. Pour $\mu \in M_c(\mathbb{X})$ et $I \geqq 0$ on pose

$$\Sigma_\mu(I) = \{x \in \mathbb{X} \colon \lambda_\mu(x) \leqq I\} \qquad \text{où } I \geqq 0.$$

Quand $\dim \mathbb{X} = 1$, on note $[g_\mu(I), d_\mu(I)]$ l'intervalle fermé $\Sigma_\mu(I)$.

(3.2) *Définition.* Si $\mu, \nu \in M_c(\mathbb{X})$, on appelle *distance entre les lignes de niveau de λ_μ et λ_ν*, le nombre $D(\lambda_\mu, \lambda_\nu) = \sup\limits_{I \geqq 0} h[\Sigma_\mu(I), \Sigma_\nu(I)]$ où h est la distance de Hausdorff (cf. 2.11).

Quand $x \to \mu_x$ est une application de $U \subset \mathbb{X}$ dans $M_c(\mathbb{X})$, nous disons que «*les lignes de niveau de λ_{μ_x} sont localement lipschitziennes en $x \in U$*» si pour tout compact $K \subset U$, il existe un nombre C_K tel que $D(\lambda_{\mu_x}, \lambda_{\mu_y}) \leqq C_K |x-y|$ quels que soient $x, y \in K$. Notons que ceci implique immédiatement $h(S_{\mu_x}, S_{\mu_y}) \leqq C_K |x-y|$, car on a toujours $h(S_\mu, S_\nu) \leqq D(\lambda_\mu, \lambda_\nu)$.

(3.3) **Lemme.** *Soit π l'ensemble des projections orthogonales de $\mathbb{X}$ sur ses sous-espaces unidimensionnels. Pour $L \in \pi$ et $I > 0$, $\mu \in M_c(\mathbb{X})$, on a:*

$$L(\Sigma_\mu(I)) = \Sigma_{L\mu}(I).$$

En particulier

$$h(\Sigma_\mu(I), \Sigma_\nu(I)) \leqq \sup_{L \in \pi} h(\Sigma_{L\mu}(I), \Sigma_{L\nu}(I)).$$

Preuve. Soient $X_1 \dots X_n$ des v.a. indépendantes équidistribuées de loi μ définies sur $(\Omega, \mathscr{B}, P)$; supposons μ de barycentre 0.

Soient $L \in \pi$, $x \in X$, $a = Lx$ et soit $H = L^{-1}[a, +\infty[$.

De $P(\bar{X}_n \in H) = P(\overline{LX}_n \geqq a)$, on tire:

$$\sigma_\mu(H) = \sigma_{L\mu}([a, +\infty[.$$

Pour $a \geqq 0$ on déduit alors de la convexité de λ_μ et $\lambda_{L\mu}$, de leur semi continuité inférieure, et de la proposition 1.7 l'existence de $v \in \partial H$ tel que

$$\lambda_\mu(v) = \sigma_\mu(H) = \sigma_{L\mu}([a, +\infty[) = \lambda_{L\mu}(a).$$

Par suite $a \in \Sigma_{L\mu}(I)$ entraîne $a = Lv$ avec $v \in \Sigma_\mu(I)$.

L'inclusion inverse est évidente, ce qui prouve $L(\Sigma_\mu(I)) = \Sigma_{L\mu}(I)$.

Par définition de $h(A, B)$, ceci implique la seconde assertion du lemme. Q.E.D.

(3.4) **Lemme.** *Soient $\mu, \nu \in M_c(\mathbb{X})$.*

Supposons que μ, ν soient les lois de deux v.a. réelles Y, Z définies sur $(\Omega, \mathscr{B}, P)$ et telles que $|Y - Z| \leqq \varepsilon$, P.p.s. *Alors $D(\lambda_\mu, \lambda_\nu)$ est majoré par 3ε. Cette situation est en particulier réalisée s'il existe une application borélienne $T \colon \mathbb{X} \to \mathbb{X}$ vérifiant $\| T(t) - t \| \leqq \varepsilon$ pour tout $t \in \mathbb{X}$, et telle que $\nu = T\mu$.*

Preuve. D'après le lemme 3.3, on peut se borner à considérer le cas $\dim \mathbb{X}=1$.

Soient $(Y_1, Z_1)\dots(Y_n, Z_n)$ n copies indépendantes équidistribuées de Y, Z, définies sur $(\Omega_n, \mathscr{B}^{\otimes n}, P^{\otimes n})$.

Soient b_μ, b_ν les barycentres de μ, ν et soit $a \geqq \varepsilon + b_\mu \vee b_\nu$. Puisque les moyennes Y_n, Z_n vérifient $|Y_n - Z_n| \leqq \varepsilon$, $P^{\otimes n}$ p.s., on a:

$$P^{\otimes n}(Y_n \geqq a-\varepsilon) \leqq P^{\otimes n}(Z_n \geqq a) \leqq P^{\otimes n}(Y_n \geqq a+\varepsilon).$$

Ceci devient, quand $n \to +\infty$

$$\lambda_\mu(a-\varepsilon) \leqq \lambda_\nu(a) \leqq \lambda_\mu(a+\varepsilon), \qquad a \geqq \varepsilon + b_\mu \vee b_\nu. \tag{1}$$

Soit $I \geqq 0$ tel que $d_\nu(I) \geqq \varepsilon + b_\mu \vee b_\nu$, et soit $J = \lambda_\nu(d_\nu(I)) \leqq I$.

L'inégalité gauche dans (1) donne

$$d_\nu(I) - \varepsilon \leqq d_\mu(I).$$

Si $J = I$ celle de droite donne

$$d_\mu(I) < d_\nu(I) + \varepsilon.$$

Si $J < I$, $d_\nu(I)$ est l'extrémité droite de S_ν et comme $d_\mu(I) \in S_\mu \subset S_\nu^\varepsilon$, on a encore $d_\mu(I) \leqq d_\nu(I) + \varepsilon$.

Ainsi on obtient $|d_\mu(I) - d_\nu(I)| \leqq \varepsilon$ dès que $d_\nu(I) \geqq \varepsilon + b_\mu \vee b_\nu$, et aussi, par symétrie, dès que $d_\mu(I) \geqq \varepsilon + b_\mu \vee b_\nu$.

Enfin si $d_\mu(I)$ et $d_\nu(I)$ sont inférieurs à $\varepsilon + b_\mu \vee b_\nu$, les inégalités $b_\mu \leqq d_\mu(I)$, $b_\nu \leqq d_\nu(I)$ et $|b_\mu - b_\nu| \leqq \varepsilon$ entraînent

$$|d_\mu(I) - d_\nu(I) \leqq 3\varepsilon.$$

Raisonnement analogue pour g_μ, g_ν.

(3.5) **Lemme.** *Supposons* $\dim \mathbb{X} = 1$ *et soient* $\mu_1, \mu_2 \in M_c(\mathbb{X})$, *de barycentres* b_1, b_2. *Les hypothèses* $b_1 \leqq b_2$, $\hat{\mu}_1(\varepsilon) \leqq \hat{\mu}_2(\varepsilon)$ *pour* $\varepsilon \geqq 0$, *[respectivement,* $\hat{\mu}_1(\varepsilon) \geqq \hat{\mu}_2(\varepsilon)$ *pour* $\varepsilon \leqq 0$*], entraînent* $d_{\mu_1}(I) \leqq d_{\mu_2}(I)$ *[respectivement* $g_{\mu_1}(I) \leqq g_{\mu_2}(I)$*] pour tout* $I \geqq 0$. *En particulier, ces situations sont réalisées si* μ_1, μ_2 *sont les lois de deux v.a. réelles* X_1, X_2 *définies sur* $(\Omega, \mathscr{B}, P)$ *et telles que* $X_1 \leqq X_2$ *P*-p.s.

Preuve. Soit $\mu \in M_c(\mathbb{X})$, de barycentre b.

D'après la section 1, on a pour $x \in S_\mu$, $\lambda_\mu(x) = -\log[\inf_\varepsilon e^{-\varepsilon x}\, \hat{\mu}(\varepsilon)]$.

De plus si $x \in \mathring{S}_\mu$, $\inf_\varepsilon e^{-\varepsilon x}\, \hat{\mu}(\varepsilon)$ est atteint pour $\varepsilon_x = \lambda'_\mu(x)$. Si $x \geqq b$ on a donc $\varepsilon_x = \lambda'_\mu(x) \geqq 0$ d'après la convexité de λ_μ. En particulier on obtient

$$\lambda_\mu(x) = -\log[\inf_{\varepsilon > 0} e^{-\varepsilon x}\, \hat{\mu}(\varepsilon)]$$

pour $x \in \mathring{S}_\mu$ et $x \geqq b$. Cette formule reste clairement valable pour $x \geqq b$ sans autre restriction sur x. Dans la situation qui nous intéresse on obtient donc, pour $x \geqq b_2 \geqq b_1$

$$\lambda_{\mu_1}(x) = -\log[\inf_{\varepsilon > 0} e^{-\varepsilon x} \hat{\mu}_1(\varepsilon)] \geqq -\log[\inf_{\varepsilon > 0} e^{-\varepsilon x}\, \hat{\mu}_2(\varepsilon)] = \lambda_{\mu_2}(x). \tag{2}$$

Supposons $b_2 \leqq d_{\mu_1}(I)$. Alors (2) implique

$$\lambda_{\mu_2}(d_{\mu_1}(I)) \leqq \lambda_{\mu_1}(d_{\mu_1}(I)) \leqq I$$

et par suite

$$d_{\mu_1}(I) \leqq d_{\mu_2}(I).$$

Cette inégalité est encore vraie si $d_{\mu_1}(I) < b_2$, car on a toujours $b_2 \leqq d_{\mu_2}(I)$. L'inégalité concernant $g_{\mu_1}(I)$, $g_{\mu_2}(I)$ se vérifie de façon analogue.

(3.6) *Remarque.* Soient $\mu, \nu \in M(\mathbb{X})$ telles que $d(\mu, \nu) = \varepsilon$.

D'après le théorème de Strassen [12], il existe des v.a. Y, Z définies sur $(\Omega, \mathscr{B}, P)$, à valeurs dans $\mathbb{X}$, de lois μ, ν, et vérifiant $P(|Y-Z| \geqq \varepsilon) \leqq \varepsilon$.

Posons $Z_1 = Z$ sur $\{|Y-Z| < \varepsilon\}$ et $Z_1 = Y$ sur $\{|Y-Z| \geqq \varepsilon\}$.

Soit ν_1 la loi de Z.

On a $\|\nu - \nu_1\| \leqq 2\varepsilon$, et $|Y - Z_1| \leqq \varepsilon$ P-p.s.

Remarquons de plus que si μ et ν sont portées par $[a, b]$, ν_1 est aussi portée par $[a, b]$.

(3.7) **Lemme.** *Soient $\mu, \nu \in M(\mathbb{R})$ telles que $S_\mu = S_\nu = [p, q]$.*

Supposons l'existence de $C > 0$ tel que:

$$|\hat{\mu}(z) - \hat{\nu}(z)| \leqq C|z| [\hat{\mu}(z) \wedge \hat{\nu}(z)] \quad \textit{pour tout } z \in \mathbb{R}. \tag{3}$$

On a alors $D(\lambda_\mu, \lambda_\nu) \leqq C$.

Preuve. De (3) on déduit:

$$\frac{\hat{\nu}}{\hat{\mu}}(z) \leqq 1 + C|z|, \qquad z \in \mathbb{R}$$

et donc en posant $m = \log \hat{\mu}$, $n = \log \hat{\nu}$

$$n(z) - m(z) \leqq C|z| \qquad \text{pour } z \in \mathbb{R}. \tag{4}$$

Pour $y \in \mathbb{R}$ fixé, la fonction $z \to n(z) + (y-z)\, n'(z)$ atteint son maximum $n(y)$ au point $z = y$. D'après (4) ceci entraîne pour tout $y, z \in \mathbb{R}$

$$n(z) + (y-z)\, n'(z) < m(y) + C|y|$$

et donc

$$n(z) - z\,n'(z) \leqq m(y) - y\,m'(y) + y[m'(y) - n'(z)] + C|y| \qquad \text{pour } y, z \in \mathbb{R}. \tag{5}$$

Fixons $I > 0$, et soit $u = d_\mu(I)$, $v = d_\nu(I)$.

Considérons d'abord le cas où $u, v \in]p, q[$.

Alors (cf. section 1), on a $\lambda_\mu(u) = \lambda_\nu(v) = I$, et en posant $y = \lambda'_\mu(u)$, $z = \lambda'_\nu(v)$

$$m'(y) = u \qquad n'(z) = v \tag{6}$$

$$-I = m(y) - y\,m'(y) = n(z) - z\,n'(z). \tag{7}$$

Notons que $y=\lambda'_\mu(u)>0$ car $u=d_\mu(I)$ et rapprochons (5), (6), (7) pour obtenir:

$$0\leqq y[u+C-v]$$

d'où

$$v-u\leqq C.$$

Maintenant, si $u\in]p,q[$ et $v=q$, faisons tendre z vers $+\infty$ dans (5), ce qui donne:

$$-\lambda_\nu(q)\leqq m(y)-y\,m'(y)+y[m'(y)-q+C] \qquad \text{pour } y>0;$$

et en prenant $y=\lambda'_\mu(u)$ on obtient $u-q+C\geqq 0$ et donc $v-u\leqq C$.
Enfin si $u=q$ et $v\in]p,q[$, cette inégalité est trivialement vraie.
Dans tous les cas, on obtient donc

$$v-u=d_\nu(I)-d_\mu(I)\leqq C.$$

Les hypothèses étant symétriques en μ,ν on en conclut que

$$|d_\nu(I)-d_\mu(I)|\leqq C.$$

L'inégalité $|g_\mu(I)-g_\nu(I)|<C$ se prouve de façon analogue.

(3.8) **Lemme.** *Soient $\mu,\nu\in M(\mathbb{R})$ des probabilités à support dans $[-A,+A]$. Supposons que $\mu=\sigma+\alpha$, $\nu=\tau+\beta$ où les mesures positives σ,α,τ,β vérifient:*

(i) $\|\mu-\nu\|\leqq\varepsilon$;

(ii) *il existe un intervalle $[a\,b]$ portant σ,τ, et dont le complémentaire porte α et β;*

(iii) *α et β donnent des masses plus grandes que $m>0$ à $]-\infty,a[$ et $]b,+\infty[$;*

(iv) $|\hat\alpha(z)-\hat\beta(z)|\leqq\varepsilon[\hat\alpha(z)\wedge\hat\beta(z)]$ *pour tout $z\in\mathbb{R}$.*

On a alors $D(\lambda_\mu,\lambda_\nu)\leqq\dfrac{4Ae}{m}\varepsilon$.

Preuve. On peut écrire pour $z\in\mathbb{R}$

$$\hat\sigma(z)-\hat\tau(z)=\int_{[a\,b]} e^{tz}\,d(\sigma-\tau)(t).$$

La valeur absolue de l'intégrale est majorée par εe^{bz} pour $z\geqq 0$ et εe^{az} pour $z\leqq 0$.

D'autre part, on a par définition

$$\hat\mu(z)\geqq\int_{]-\infty,a[} e^{tz}\,d\mu(t)+\int_{]b,+\infty[} e^{tz}\,d\mu(t).$$

D'après (iii), $\hat\mu(z)$ est donc minoré par $m\,e^{bz}$ pour $z\geqq 0$ et par $m\,e^{az}$ pour $z\leqq 0$. En conjonction avec (iv), cette inégalité implique

$$|\hat\mu-\hat\nu|<|\hat\sigma-\hat\tau|+|\hat\alpha-\hat\beta|<\frac{2\varepsilon}{m}\hat\mu. \tag{8}$$

D'autre part pour $|z|<1/2A$ on a, d'après (i)

$$\left|\frac{d}{dz}(\hat{\mu}-\hat{\nu})(z)\right| = \Big|\int_{[-A,+A]} t\, e^{tz}\, d(\mu-\nu)(t)\Big| < 2\varepsilon A\, e^{1/2}$$

$$\hat{\mu}(z) = \int_{[-A,A]} e^{tz}\, d\mu(t) > e^{-1/2}.$$

La formule des acrroissements finis donne alors

$$|\hat{\mu}(z)-\hat{\nu}(z)| \leqq 5\varepsilon A\, e^{1/2}|z| \leqq 2\varepsilon A\, e|z|\, \hat{\mu}(z) \qquad \text{pour } |z| \leqq \frac{1}{2A}. \tag{9}$$

Les inégalités (8) (9) entraînent

$$|\hat{\mu}(z)-\hat{\nu}(z)| \leqq 4\varepsilon \frac{A\, e}{m}|z|\, \hat{\mu}(z) \qquad \text{pour tout } z \in \mathbb{R};$$

les hypothèses étant symétriques en μ et ν, on obtient finalement

$$|\hat{\mu}(z)-\hat{\nu}(z)| \leqq 4\varepsilon\, A\, e/m\, |z|\, [\hat{\mu}(z) \wedge \hat{\nu}(z)] \qquad \text{pour } z \in \mathbb{R}.$$

Le lemme 3.7 permet de conclure. Q.E.D.

(3.9) **Proposition.** *Soit $x \to \mu_x$ une application de $U \subset \mathbb{X}$ dans $M(\mathbb{X})$, ayant les propriétés suivantes:*

(i) *pour tout compact $K \subset U$, les supports des μ_x, $x \in K$, sont contenus dans un compact fixe; de plus le support de μ_x engendre $\mathbb{X}$, pour tout $x \in U$;*

(ii) *l'application $x \to \mu_x$ est localement lipschitzienne, $M(\mathbb{X})$ étant muni de la distance de Prokhorov;*

(iii) *pour tout compact $K \subset U$, il existe $m \in]0, \frac{1}{3}[$ et $C>0$, tels que à chaque projection orthogonale $L: \mathbb{X} \to \mathbb{R}$ de rang 1 et à chaque $x \in K$ on puisse associer des nombres $a_x \leqq b_x$ et une fonction borélienne positive $f_x: \mathbb{R} \to \mathbb{R}$ (la dépendance en L n'est pas indiquée explicitement dans la notation) vérifiant:*

1) *les fonctions $x \to a_x$, $x \to b_x$ et celles des fonctions $\varphi_t: x \to \log f_x(t)$, $t \in \mathbb{R}$ qui ne sont pas identiques à $+\infty$ sont lipschitziennes en $x \in K$ avec la constante de Lipschitz C;*

2) *pour $t \geqq 0$ les fonctions de répartition*

$$A_x(t) = L(\mu_x)(]a_x - t, a_x[) \quad et \quad B_x(t) = L(\mu_x)(]b_x, b_x + t[)$$

sont données par

$$A_x(t) = \int_{-t}^{0} f_x(s)\, ds \quad et \quad B_x(t) = \int_{0}^{t} f_x(s)\, ds;$$

de plus, elles sont majorées par Ct, et $A_x(+\infty)$, $B_x(+\infty)$ sont dans l'intervalle $[m, \frac{1}{3}]$.

Dans cette situation, les lignes de niveau de λ_{μ_x} sont localement lipschitziennes en $x \in U$ (cf. def. 3.2).

Preuve. Montrons que pour tout compact $K \subset U$ on peut trouver $C_K > 0$ telle que

$$D[\lambda_{L(\mu_x)}, \lambda_{L(\mu_y)}] \leqq C_K |x-y|$$

pour tout $x, y \in K$, toute projection orthogonale L de rang 1. D'après le lemme 3.3, ceci établira la proposition. Nous sommes donc ramenés à prouver le résultat quand $\dim \mathbb{X} = 1$, ce que nous supposons désormais.

Soit K compact $\subset U$, et soit $t \in \mathbb{R}$ tel que $f_x(t)$ ne soit pas identique à 0 pour tout $x \in K$. L'hypothèse (iii)-1 donne:

$$|\log f_x(t) - \log f_y(t)| \leqq C|x-y| \quad \text{pour } x, y \in K$$

ce qui entraîne

$$|f_x(t) - f_y(t)| \leqq 6C|x-y| [f_x(t) \wedge f_y(t)] \quad \text{pour } x, y \in K \tag{10}$$

pourvu que $C|x-y| \leqq \operatorname{Log} 2$. Si $f_x(t) = 0$ pour tout $x \in K$ (10) reste trivialement vraie.

Soient $x, y \in K$. Posons $|x-y| = \varepsilon$, $a = a_x \wedge a_y$, $b = b_x \vee b_y$. Définissons T_y: $\mathbb{R} \to \mathbb{R}$ par

$$\begin{aligned} T_y(t) &= t && \text{pour } a_y < t < b_y \\ T_y(t) &= b + t - b_y && \text{pour } b_y \leqq t \\ T_y(t) &= a + t - a_y && \text{pour } t \leqq a_y \end{aligned}$$

et définissons T_x à partir des a_x, b_x de façon analogue.

Pour tout t, $|T_y(t) - t|$ et $|T_x(t) - t|$ sont majorés par $|a_x - a_y| \vee |b_x - b_y|$, donc par $C\varepsilon$.

Posons $\mu = T_x(\mu_x)$, $v = T_y(\mu_y)$.

Le lemme 3.4 implique

$$D(\lambda_{\mu_x}, \lambda_\mu) \leqq 3C\varepsilon \quad \text{et} \quad D(\lambda_{\mu_y}, \lambda_v) \leqq 3C\varepsilon.$$

Soient σ, τ les restrictions de μ, v à $[a\, b]$ et écrivons $\mu = \sigma + \alpha$, $v = \tau + \beta$. Les hypothèses (iii)-1-2, et l'inégalité (10) montrent que $\frac{1}{3}$ minore $\|\sigma\|$, $\|\tau\|$, que m minore $\alpha(]b, +\infty[)$, $\beta(]b, +\infty[)$, $\alpha(]-\infty, a[)$, $\beta(]-\infty, a[)$, et entraînent

$$\|\alpha - \beta\| \leqq 6C\varepsilon, \quad |\hat{\alpha} - \hat{\beta}| \leqq 6C\varepsilon[\hat{\alpha} \wedge \hat{\beta}].$$

Posons $\tau_1 = \dfrac{1 - \|\alpha\|}{1 - \|\beta\|} \tau$ si $\|\beta\| \neq 1$ et $\tau_1 = (1 - \|\alpha\|)\, \delta_b$ si $\|\beta\| = 1$.

On a alors $\|\tau - \tau_1\| = |\|\alpha\| - \|\beta\|| \leqq 6C\varepsilon$, et $v_1 = \tau_1 + \alpha$ est une probabilité.

Le lemme 3.8, entraîne alors $D(\lambda_v, \lambda_{v_1}) \leqq 24C \dfrac{Ae}{m} \varepsilon$, où $[-A, +A]$ est un intervalle fixe contenant les supports de tous les μ_x, $x \in K$.

Par hypothèse, il existe $C_1 > 0$ tel que $d(\mu_x, \mu_y) \leqq C_1 |x-y|$ pour $x, y \in K$. La construction de μ, v, v_1 fournit C_2 (dépendant seulement de C et C_1) tel que $d(\mu, v_1) \leqq C_2 \varepsilon$. Si B est un borélien de $[a, b]$, on a donc

$$\mu(B) = \sigma(B) \leqq \tau_1(B^{C_2 \varepsilon}) + \alpha(B^{C_2 \varepsilon}) + C_2 \varepsilon.$$

Comme $\alpha(B^{C_2\varepsilon}) \leqq \alpha(]a-C_2\varepsilon, a[)+\alpha(]b, b+C_2\varepsilon[)$ est majoré par $2CC_2\varepsilon$-d'après (iii)-2 on obtient

$$\sigma(B) \leqq \tau_1(B^{C_2\varepsilon})+C_2(2C+1)\varepsilon;$$

après échange de σ, τ_1, et puisque $\|\sigma\|=\|\tau_1\| \geqq \frac{1}{3}$, ceci donne:

$$d\left(\frac{\sigma}{\|\sigma\|}, \frac{\tau_1}{\|\tau_1\|}\right) \leqq C_3\varepsilon \quad \text{où } C_3=3C_2(2C+1).$$

Il existe (cf. rem. 3.6) alors une probabilité η portée par $[a\,b]$ telle que $\left\|\frac{\tau_1}{\|\tau_1\|}-\eta\right\| \leqq C_3\varepsilon$, et deux v.a. Y, Z définies sur $(\Omega, \mathscr{B}, P)$, de lois $\sigma/\|\sigma\|$ et η, telles que $|Y-Z| \leqq C_3\varepsilon$, P-p.s.

Posons $\nu_2=\|\tau_1\|+\alpha$.

Le lemme 3.8 montre que

$$D(\lambda_{\nu_1}, \lambda_{\nu_2}) \leqq \frac{4Ae}{m} C_3\varepsilon.$$

D'autre part, à partir de $\Omega, \mathscr{B}, P\;\; Y, Z$ on construit aisément (noter que $\|\sigma\|=\|\tau_1\|$) un espace de probabilité $(\Omega', \mathscr{B}', P')$ et des v.a. Y', Z' définies sur Ω', de lois μ et ν_2, vérifiant $|Y'-Z'| \leqq C_3\varepsilon$, P'-p.s.

Du lemme 3.4, on déduit alors

$$D(\lambda_\mu, \lambda_{\nu_2}) \leqq 3C_3\varepsilon.$$

Finalement, on voit que

$$D(\lambda_{\mu_x}, \lambda_{\mu_y}) \leqq C_4|x-y|, \qquad \text{pour } x, y \in K,$$

où $C_4=3C_3+4\frac{Ae}{m}C_3+24\frac{Ae}{m}C+6C$.

(3.10) *Exemple.* Soit K un compact de $\mathbb{X}$, et μ_x, $x\in\mathbb{X}$, une famille de probabilités de support K telles que μ_x ait une densité $f(x,t)$ $t\in K$ par rapport à la mesure de Lebesgue.

Supposons que pour tout $t\in K$, $x\in\mathbb{X}$ $\frac{\partial f}{\partial x}(x,t)$ existe, que $\frac{\partial f}{\partial x}$ et f soient continues sur $\mathbb{X}\times K$, et que $f(x,t)>0$. Alors les hypothèses de 3.9 sont vérifiées, et les lignes de niveau de λ_{μ_x} sont localement lipschitziennes en x.

(3.11) **Proposition.** *Soit $x\to\mu_x$ une application de $U\subset\mathbb{X}$ dans $M(\mathbb{X})$, vérifiant les hypothèses*

(i) (ii) *de 3.9, ainsi que*

(iii) *l'application $x\to S_{\mu_x}$ est localement lipschitzienne en $x\in U$ (pour la distance de Hausdorff);*

(iv) *pour tout compact $K\subset U$ il existe des constantes $\rho>0$, $\gamma>0$ telles que, pour tout $x\in K$, pour tout point extrémal y de S_{μ_x}, pour toute boule B de centre y et de rayon $r\leqq\rho$ on ait $\mu_x(B)\geqq\gamma r$. Alors les lignes de niveau de λ_{μ_x} sont localement lipschitziennes en $x\in U$.*

Preuve. Les hypothèses ci-dessus étant convenablement préservées par projection orthogonale, le lemme 3.2 permet de se limiter à considérer le cas où $\dim \mathbb{X}=1$. Fixons un compact $K\subset U$ et soient $x, y\in K$ avec $|x-y|=\varepsilon$. Dans la suite, $C, C_1, C_2, \ldots$, sont des nombres positifs ne dépendant que de K. Par hypothèse, on a $d(\mu_x, \mu_y)<C\varepsilon$; d'après la remarque 3.6, il existe une probabilité μ vérifiant $\|\mu-\mu_y\|<2C\varepsilon$ et telle que μ_x, μ soient les lois marginales d'une v.a. à valeurs dans $\mathbb{R}^2$ concentrée sur un voisinage de taille $C\varepsilon$ de la diagonale de $\mathbb{R}^2$. Le lemme 3.4 donne

$$D(\lambda_{\mu_x}, \lambda_\mu)\leqq 3C\varepsilon.$$

D'autre part μ est à support dans $[c,d]$ où $[c,d]$ est n'importe quel intervalle portant μ_x et μ_y. Si on écrit $S_{\mu_y}=[a,b]$, on peut donc d'après (iii) prendre $c=a-C_1\varepsilon$ et $d=b+C_1\varepsilon$.

Ecrivons

$$\mu=\sigma+\alpha, \qquad \mu_\mu=\tau+\beta$$

où

$$]-\infty, b-\rho[\text{ porte } \sigma,\tau \quad \text{et} \quad]b-\rho, +\infty[\text{ porte } \alpha, \beta.$$

Posons $\nu_1=\tau_1+\beta_1$ où $\tau_1=\left(1-\dfrac{\|\alpha\|}{\|\alpha\|\vee\|\beta\|}\right)\delta_{b-\rho}+\tau$

$$\beta_1=\frac{\|\alpha\|}{\|\alpha\|\vee\|\beta\|}\beta.$$

Les inégalités $\|\alpha-\beta\|\leqq 2C\varepsilon$ et $\|\alpha\|\wedge\|\beta\|\geqq\gamma\rho$ entraînent $|\hat\beta_1-\hat\beta|\leqq C_2\varepsilon\hat\beta\wedge\hat\beta_1$, $\|\tau-\tau_1\|<C_2\varepsilon$, et donc d'après le lemme 3.8, $D(\lambda_{\mu_y}, \lambda_{\nu_1})<C_3\varepsilon$. De plus, on a $\|\beta_1\|=\|\alpha\|$ et $\|\mu-\nu_1\|\leqq C_4\varepsilon$.

Posons $\pi=\alpha\wedge\beta_1$.

Par construction $\beta_1-\pi$ est à support dans $[b-\rho, d]$ et de norme $v\leqq\frac12 C_4\varepsilon$; ceci entraîne

$$\hat\beta_1(z)-\hat\pi(z)\leqq v\, e^{dz} \quad \text{pour } z\geqq 0. \tag{11}$$

D'autre part, pour $0\leqq r\leqq\rho$ on a, si $C_4\varepsilon\leqq\frac12\gamma r$:

$$\pi([b-r,b])\geqq\beta([b-r,b])-C_4\varepsilon\geqq\gamma r-C_4\varepsilon\geqq\tfrac12\gamma r.$$

Posons $m=\dfrac{2C_4\varepsilon}{\gamma}$, $p=\frac12\rho$, $q=p+m$, et imposons $\dfrac{4C_4\varepsilon}{\gamma}<\rho$, pour garantir en particulier $m<p<q<\rho$.

On peut écrire d'après le théorème de Fubini

$$\hat\pi(z)=\int_a^b e^{tz}\,d\pi(t)=\int_{-\infty}^{b} z\,e^{sz}\,\pi[s\vee a, b]\,ds$$

$$\hat\pi(z)>\int_{-\infty}^{b-q} z\,e^{sz}\,\pi[b-q,b]\,ds+\int_{b-p}^{b-m} z\,e^{sz}\,\pi[s,b]\,ds.$$

Ceci devient, en minorant $\pi[b-q,b]$ par $\frac{1}{2}\gamma q$ et $\pi[s,b]$ par $\frac{1}{2}(b-s)$,

$$\hat{\pi}(z)>\tfrac{1}{2}\gamma\, e^{bz}\left[q\, e^{-qz}+m\, e^{-mz}-\rho\, e^{-pz}+\frac{1}{z}(e^{-mz}-e^{-pz})\right].$$

La fonction $e^{qz}g(z)$, où $g(z)=q\,e^{-qz}+m\,e^{-mz}-p\,e^{-pz}$ est croissante pour $z\geqq 0$, car $m(q-m)=p(q-p)$ et $q+m-p\geqq 0$. Par suite $g(z)$ est positif pour $z\geqq 0$ et finalement

$$\hat{\pi}(z)\geqq\frac{1}{2}\frac{\gamma}{z}\left[e^{(b-m)z}-e^{(b-p)z}\right]\qquad\text{pour } z\geqq 0. \tag{12}$$

Posons $\nu_2=\tau_1+v\,\delta_{b-p+m}+\delta_t*\pi$, où $t>0$ est à déterminer et $v=\|\beta_1-\pi\|\leqq \frac{1}{2}C_4\,\varepsilon$. D'après (11), pour garantir $\hat{\nu}_2(z)\leqq\hat{\nu}_2(z)$ quel que soit $z>0$, il suffit d'avoir:

$$(e^{tz}-1)\,\hat{\pi}(z)\geqq v\left[e^{dz}-e^{(b-p+m)z}\right]\qquad\text{pour } z\geqq 0.$$

Cette inégalité est sûrement vraie – d'après (12) – dès que

$$(e^{tz}-1)\frac{1}{2}\frac{\gamma}{z}\left[e^{-mz}-e^{-pz}\right]\geqq C_4\,\varepsilon\left[e^{C_1\varepsilon z}-e^{(-p+m)z}\right],\qquad z\geqq 0$$

ce qui équivaut à

$$e^{tz}-1\geqq\frac{2\,C_4\,\varepsilon}{\gamma}\,z\,e^{(C_1\varepsilon+m)z},\qquad\text{pour } z\geqq 0.$$

Le choix

$$t=2\left[\frac{2\,C_4\,\varepsilon}{\gamma}+C_1\,\varepsilon+m\right]=C_5\,\varepsilon$$

garantit clairement cette dernière égalité et donc $\hat{\nu}_1(z)\leqq\hat{\nu}_2(z)$, pour $z\geqq 0$.

Posons $\nu_3=\tau_1+v\,\delta_{b-\rho}+\pi$ et $\nu_4=\tau_1+\alpha$.

La mesure $(\alpha-\pi)$ étant de masse v, et à support dans $[b-\rho,d]$, on a $\hat{\nu}_3(z)\leqq\hat{\nu}_4(z)$ pour $z\geqq 0$. En particulier, pour $I\geqq 0$, le lemme 3.5 entraîne

$$d_{\nu_1}(I)-d_{\nu_4}(I)\leqq d_{\nu_2}(I)-d_{\nu_3}(I).$$

Soit $\nu_5=\tau_1+v\,\delta_{b-\rho}+\delta_t*\pi$.

Le lemme 3.4 et le choix $t=C_5\,\varepsilon$ garantissent $D(\lambda_{\nu_3},\lambda_{\nu_5})\leqq 3\,C_5\,\varepsilon$. Comme $\nu_5-\nu_2$ est à support dans $\left[b-\rho,b-\dfrac{\rho}{4}\right]$ le lemme 3.8 donne

$$D(\lambda_{\nu_5},\lambda_{\nu_2})\leqq C_6\,\varepsilon.$$

Finalement on obtient

$$d_{\nu_1}(I)-d_{\nu_4}(I)\leqq C_7\,\varepsilon\qquad\text{pour } I\geqq 0.$$

On peut clairement échanger les rôles de ν_1 et ν_4 dans la démonstration ci-dessus, pour obtenir

$$|d_{\nu_1}(I)-d_{\nu_4}(I)|\leqq C_7\,\varepsilon\qquad\text{pour } I\geqq 0.$$

Le travail effectué sur μ_y à l'extrémité droite de $[a,b]$ peut être répété sur v_4 $\bar{a}$ l'extrémité gauche de $[a,b]$. Ecrivons

$$\sigma=\alpha'+\sigma' \qquad \tau_1=\alpha_4+\tau_4$$

où α', α_4 sont portées par $]-\infty, a+\rho[$ et σ', τ_4 par $[a+\rho, b-\rho]$.

Comme ci-dessus, on construit τ_6 portée par $[a+\rho, b-\rho]$ vérifiant $\|\tau_6-\tau_4\|<C_8\,\varepsilon$ et telle que la probabilité $v_6=\alpha'+\tau_6+\alpha$ vérifie

$$|d_{v_6}(I)-d_{v_4}(I)|\leqq C_8\,\varepsilon \qquad \text{pour } I\geqq 0.$$

Le lemme 3.8, appliqué à $\mu=\alpha'+\sigma'+\alpha$ et $v_6=\alpha'+\tau_6+\alpha$ (puisque par construction $\|\sigma'-\tau_6\|\leqq C_9\,\varepsilon$ tandis que $\|\alpha'\|\wedge\|\alpha\|\geqq C_{10}\geqq 0$), entraîne $D(\lambda_{v_6},\lambda_\mu)\leqq C_{11}\,\varepsilon$, d'où finalement $|d_{\mu_x}(I)-d_{\mu_y}(I)|\leqq C_{12}\,\varepsilon$ pour $I\geqq 0$.

Les images $\tilde{\mu}_x$ des μ_x par la symétrie de $\mathbb{R}$ autour de 0 vérifient les mêmes hypothèses que les μ_x; comme $d_{\tilde{\mu}_x}(I)=-g_{\mu_x}(I)$, on obtient

$$D(\lambda_{\mu_x},\lambda_{\mu_y})\leqq C_{13}\,\varepsilon=C_{13}\,|x-y|$$

pourvu que $x, y\in K$ et $|x-y|\leqq C_{15}$. Q.E.D.

(3.12) *Exemple.* Soit $a_1 \dots a_n$ une famille finie d'applications lipschitziennes (localement) de $\mathbb{R}^p$ dans $\mathbb{R}^p$. Soit $p_1 \dots p_n$ des applications localement lipschitziennes de $\mathbb{R}^p$ dans $\mathbb{R}$ vérifiant pour $x\in\mathbb{R}^p$,

$$\sum_{i=1}^{n} p_i(x)=1 \quad \text{et} \quad p_i(x)\geqq 0, \qquad 1\leqq i\leqq n.$$

Supposons de plus que l'enveloppe convexe $S(x)$ de la famille des points $a_i(x)$ pour lesquels $p_i(x)\neq 0$ varie de façon localement lipschitzienne en $x\in\mathbb{R}^p$. Alors les probabilités

$$\mu_x=\sum_{i=1}^{n} p_i(x)\,\delta_{a_i(x)}, \qquad x\in\mathbb{R}^p$$

vérifient les hypothèses de la proposition 3.11, et les lignes de niveau de λ_{μ_x} sont localement lipschitziennes en $x\in\mathbb{R}^p$.

Notons d'ailleurs que si les $p_i(x)$ ne s'annulent pas, la condition sur $S(x)$ est automatiquement vérifiée.

4. Fonctionnelle de Cramer associee a un champ de mesures

(4.1) *Hypothèses de base.* Soit M une variété riemannienne. Nous appelons *champ de mesures* sur M toute application associant à chaque $x\in M$ une probabilité μ_x sur l'espace tangent $T_x(M)$. Nous considérons seulement des champs de mesures μ_x, $x\in M$ *faiblement continus*, c'est-à-dire tels que pour toute carte φ: $U\to\mathbb{R}^k$ (U ouvert de M) de M, l'application $x\to d\varphi_x(\mu_x)$ de U dans l'espace de probabilités sur $\mathbb{R}^k$ (muni de la distance de Prokhorov) soit continue. Nous noterons S_x l'enveloppe convexe fermée (dans $T_x(M)$) du support de μ_x. Nous supposerons toujours dans la suite que S_x est compact et engendre $T_x(M)$ pour tout $x \in M$, et que pour toute carte $\varphi: U \to \mathbb{R}^k$, l'application $x \to d\varphi_x(S_x)$ de U dans l'ensemble des compacts de $\mathbb{R}^k$, (muni de la distance de Hausdorff) est continue.

Sur l'espace euclidien $T_x(M)$, la transformée de Cramér de μ_x définit une fonction λ_x à valeurs dans $[0, +\infty]$, nulle au barycentre b_x de μ_x. D'après la proposition 1.5, la fonction

$$\lambda: T(M) \to [0, +\infty]$$

définie par

$$\lambda(x, v) = \lambda_x(v), \qquad \text{où } x \in M, \ v \in T_x(M)$$

est semi continue inférieurement.

Le champ de vecteurs continu $x \to b_x$ sera appelé brièvement «*champ moyen*» ou «*champ de vecteurs moyens*». Pour éviter dans la suite des complications associées à d'éventuels «temps d'explosion» de trajectoires tangentes au champ de convexes S_x, $x \in M$, nous supposerons toujours que le champ de mesures μ_x, $x \in M$ est *borné*, c'est-à-dire que

la distance riemannienne $\rho(x, y)$ sur M tend vers $+\infty$ quand y sort de tout compact de M (1)

pour un $x_0 \in M$, les nombres $C(R) = \sup\{\|v\|_x: \ v \in S_x \text{ et } \rho(x_0, x) < R\}$ vérifient

$$\sum_{n=1}^{+\infty} \frac{1}{C(n)} = +\infty. \tag{2}$$

Dans ces conditions, pour tout compact C de M, pour tout $T > 0$ fini, il existe un compact C_1 de M tel que les relations $\gamma(0) \in C$ et $\gamma'(t) \in S_{\gamma(t)}$ pour preque tout $t \in [0, T]$ impliquent $\gamma[0, T] \subset C_2$ (γ désigne un chemin dans M absolument continu et continu). Un tel chemin sera dit «*tangent au champ de convexes S_x, $x \in M$*».

La distance $\rho_{S,T}(\gamma_1, \gamma_2)$ entre deux chemins γ_i: $[S, T] \to M$, $i = 1, 2$, est définie comme le supremum sur $t \in [S, T]$ de $\rho[\gamma_1(t), \gamma_2(t)]$.

(4.2) *Définition.* Etant donné un champ de mesures μ_x sur la variété riemannienne M, comme ci-dessus, nous poserons, pour tout chemin continu, absolument continu,

$$\gamma: [S, T] \to M$$

$$\Lambda_{S,T}(\gamma) = \int_S^T \lambda_{\gamma(t)}(\gamma'(t))\, dt.$$

Si γ: $[S, T] \to M$ est continu, mais pas absolument continu, nous poserons $\Lambda_{S,T}(\gamma) = +\infty$. On définit ainsi une application $\Lambda_{S,T}$ de l'espace $C_{S,T}(M)$ des chemins continus γ: $[S, T] \to M$, dans $[0, +\infty]$. Remarquons que $\Lambda_{S,T}(\gamma) = 0$ implique $\gamma'(t) = b_{\gamma(t)}$ pour presque tout $t \in [S, T]$ et donc implique que γ est une courbe intégrale du champ moyen b_x, $x \in M$. L'application $\Lambda_{S,T}$ sera appelée *la fonctionnelle de Cramer* associée au champ de mesures μ_x, $x \in M$.

(4.3) **Lemme.** *Soit μ_x, $x \in M$ un champ de mesures sur M vérifiant les hypothèses 4.1. La fonctionnelle de Cramer $\Lambda_{S,T}$ est une fonction semi continue inférieurement, (($C_{S,T}(M)$ étant muni de la topologie de la convergence uniforme sur $[S,T]$) c'est-à-dire que $\Lambda_{S,T}(\varphi) \leqq \varliminf_{\psi \to \varphi} \Lambda_{S,T}(\psi)$.*

De plus, l'ensemble des $\varphi \in C_{S,T}(M)$ telles que $\varphi(0)$ appartienne à un compact de M et telles que $\Lambda_{S,T}(\varphi) < +\infty$ est un compact de $C_{S,T}(M)$.

Preuve. Sans perdre de généralité, prenons $S=0$.

L'hypothèse $\Lambda_{0,T}(\varphi) < +\infty$ entraîne $\lambda_{\varphi(t)}(\varphi'(t)) < +\infty$ pour presque tout $t \in [0,T]$ et donc (cf. section 1), $\varphi'(t) \in S_{\varphi(t)}$ pour presque tout t. Si on impose $\varphi(0) \in K$, K compact dans M, le fait que champ μ_x soit borné (cf. 4.1) entraîne aisément l'existence d'un compact K' ne dépendant que de K et T tel que $\varphi([0,T]) \subset K'$. La relation $\varphi'(t) \in S_{\varphi(t)}$ pour tout $t \in [0,T]$ et $\varphi(t) \in K'$ montre que $\|\varphi'(t)\|_{\varphi(t)}$ est majoré par une constante indépendante de φ. En particulier l'ensemble des φ tels que $\Lambda_{0,T}(\varphi) < +\infty$ et $\varphi(0) \in K$ est équicontinu et donc compact dans $C_{0,T}(M)$.

Pour prouver que $\Lambda_{0,T}$ est s.c.i., considérons une suite $\varphi_n \in C_{0,T}(M)$ convergeant uniformément vers φ avec $\lim_{n \to +\infty} \Lambda_{0,T}(\varphi_n) = I$.

Il suffit bien sûr de considérer le cas $I < +\infty$. D'après ce qui précède, les φ'_n sont uniformément localement bornées (sauf sur une partie Lebesgue-négligeable de $[0,T]$). On peut donc extraire une sous-suite φ_k dont les dérivées φ'_k convergent, pour la topologie faible de $L^2(0,T)$ vers une fonction qui est nécessairement φ', de sorte que φ est nécessairement absolument continue (dans $\mathbb{R}^k$ on écrirait

$$\varphi(t) - \varphi(0) = \lim_{k \to \infty} [\varphi_k(t) - \varphi_k(0)] = \lim_{k \to +\infty} \int_0^t \varphi'_k(t)\, dt = \int_0^t \varphi'(t)\, dt.$$

Pour presque tout $t \in [0,T]$, tout voisinage convexe fermé V de $S_{\varphi(t)}$ contient les $\varphi'_k(s)$ pour k assez grand et presque tout s dans un intervalle ω de centre t. Donc V contient $[\varphi_k(s_2) - \varphi_k(s_1)]/(s_2 - s_1)$ pour k grand et presque tout $s_1 \neq s_2$ dans ω. Par suite V contient $[\varphi(s_2) - \varphi(s_1)]/(s_2 - s_1)$ pour presque toute paire $s_1 \neq s_2$ dans ω, ce qui entraîne $\varphi'(t) \in V$, pour presque tout $t \in [0,T]$.

On va maintenant se ramener au cas (*) où $\varphi'_k(t) \in S^{-\delta}_{\varphi_k(t)}$ pour un $\delta > 0$ fixe, pour tout k assez grand, et presque tout $t \in [0,T]$. Pour cela, sans toucher à la suite de chemins φ_k, on accroit les S_x en régularisant chaque μ_x par convolution. D'après le lemme 1.6, et le theorème de convergence dominée, on peut le faire en diminuant $\Lambda_{0,T}(\varphi)$ d'un $\varepsilon > 0$ arbitraire. Si on a démontré la s.c.i dans le cas (*), pour k grand, $(\Lambda_{0,T}(\varphi) - 2\varepsilon)$ minore les «nouveaux» $\Lambda_{0,T}(\varphi_k)$ et donc à fortiori les $\Lambda_{0,T}(\varphi_k)$ initiaux.

Sous l'hypothèse (*), $\Lambda_{0,T}(\varphi_k)$ est voisin de $\int_0^T \lambda_{\varphi(t)}(\varphi'_k(t))\, dt$ — pour que cette intégrale ait un sens, on se place dans une carte de M ce qui par recollement évident traite le cas global —, en effet, puisque φ_k converge uniformément vers la fonction continue φ, on peut pour tout $\varepsilon > 0$ trouver autour de tout t_0 un intervalle ω et un $N > 0$ tels que

$$|\lambda_{\varphi_k(t)}(v) - \lambda_{\varphi(t)}(v)| \leqq \varepsilon$$

quels que soient $t \in \omega$, $k \geqq N$ et $V \in S^{-S/2}_{\varphi(t)}$.

La convexité des transformées de Cramér $\lambda_x(.)$ permet alors d'écrire

$$\lambda_{\varphi(t)}(v) \geqq \lambda_{\varphi(t)}(\varphi'(t)) + \frac{\partial \lambda_{\varphi(t)}}{\partial v}(\varphi'(t)) \cdot [v - \varphi'(t)].$$

D'où

$$\int_0^T \lambda_{\varphi(t)}(\varphi_k'(t))\,dt \geqq \int_0^T \lambda_{\varphi(t)}(\varphi'(t))\,dt + \int_0^T \frac{\partial \lambda_{\varphi(t)}}{\partial v}(\varphi'(t)) \cdot [\varphi_k'(t) - \varphi'(t)]\,dt.$$

La convergence $\varphi_k' \to \varphi'$ dans $L^2_{[0,T]}$ entraîne alors la s.c.i. désirée.

(4.4) **Lemme** (hypothèses de 4.1). *Soit φ: $[0, T] \to M$ un chemin continu, absolument continu, tel que pour un certain $\delta > 0$ on ait $\varphi'(t) \in S_{\varphi(t)}^{-\delta}$ pour presque tout $t \in [0, T]$. Pour tout $\varepsilon > 0$, pour tout $x \in M$ suffisamment proche de $\varphi(0)$, il existe alors un chemin continu ψ: $[0, T] \to M$, continûment différentiable par morceaux, de dérivée à droite $\psi'(t)$ vérifiant $\psi'(t) \in S_{\psi(t)}^{-\delta/2}$ pour tout $t \in [0, T]$, et tel que $\psi(0) = x$, $\rho_{0,T}(\varphi, \psi) \leqq \varepsilon$, $|\Lambda_{0,T}(\varphi) - \Lambda_{0,T}(\psi)| \leqq \varepsilon$.*

Preuve. Si le lemme est vrai pour chaque intervalle d'une partition de $[0, T]$, (et pour tout $\varepsilon > 0$, $\delta > 0$) il est encore vrai pour $[0, T]$. Il suffit donc de considérer le cas où T est assez petit pour que $[0, T]$ soit contenu dans le domaine d'une carte de M-ce qui nous permet de supposer désormais $M = \mathbb{R}^k$.

L'application $y \to S_y$ est continue par hypothèse. On peut donc trouver un T_0 assez petit, et un voisinage compact V de $[0, T_0]$ tels que les convexes $\Gamma_1 = S_{\varphi(0)}^{-3\delta/5}$ et $\Gamma_2 = S_{\varphi(0)}^{-3\delta/5}$ de $\mathbb{R}^k$ vérifient

$$S_{\varphi(t)}^{-\delta} \subset \Gamma_1 \subset \Gamma_2 \subset S_y^{-\delta/2} \qquad \text{pour } t \in [0, T_0],\ y \in V. \tag{1}$$

D'après notre remarque initiale il suffit d'étudier le cas où $T \leqq T_0$. Clairement, on peut aussi supposer δ assez petit pour que Γ_2 soit un voisinage de Γ_1.

On peut toujours choisir une version de φ' telle que $\varphi'(t) \in \Gamma_1$ pour tout $t \in [0, T]$. Il est possible ensuite d'approcher φ', uniformément sur $[0, T]$, avec une précision arbitraire, par une fonction étagée f: $[0, T] \to \Gamma_2$.

Ecrivons $f = \sum_{i=1}^{n} v_i I_{A_i}$ où $v_i \in \Gamma_2$, $A_i \in \mathscr{B}[0, T]$.

Approchons la «partition» (A_i) de $[0, T]$, par une partition (B_i) de $[0, T]$ où les B_i sont des unions finies d'intervalles de sorte que $[\sum_i \ell(A_i \Delta B_i)]$ soit arbitrairement petit. Ici ℓ désigne la mesure de Lebesgue et Δ la différence symétrique.

La fonction $g = \sum_{i=1}^{n} v_i I_{B_i}$ est à valeurs dans Γ_2, constante par morceaux sur $[0, T]$, et $\ell\{f \neq g\}$ est arbitrairement petit. Par suite pour tout $\alpha > 0$ il existe une fonction constante par morceaux g: $[0, T] \to \Gamma_2$ telle que

$$\ell\{|\varphi' - g| \geqq \alpha\} \leqq \alpha. \tag{2}$$

Posons $\psi(t) = x + \int_0^t g(u)\,du$, où $\rho(x, \varphi(0)) < \alpha$. On a immédiatement

$$|\psi(t) - \varphi(t)| < \alpha[1 + T + C] \qquad \text{pour } t \in [0, T], \tag{3}$$

où C est un majorant de $\|v\|$ pour $v \in \Gamma_2$.

Pour α assez petit on peut donc garantir $\psi[0,T]\subset V$ et $\rho_{0,T}(\varphi,\psi)<\varepsilon$. D'après (1) la dérivée à droite ψ' vérifie $\psi'(t)\in S_{\psi(t)}^{-\delta/2}$. Comme $\lambda_y(v)$ est continu sur $\bigcup_{y\in V} y\times \mathring{S}_y$ et uniformément majoré par C_1 sur $\bigcup_{y\in V} y\times S_y^{-\delta/2}$, les relations (2) et (3) montrent que pour α assez petit, $|\Lambda_{0,T}(\varphi)-\Lambda_{0,T}(\psi)|$ est majoré par ε.

(4.5) *Définition.* Nous appellerons «*chemin à vitesses intérieures*» au champ de convexes S_x, $x\in M$, tout chemin continu φ: $[0,T]\to M$, *continûment différentiable par morceaux*, et tel que pour tout $t\in[0,T[$, on ait $\varphi'(t)\in \mathring{S}_{\varphi(t)}$, où $\varphi'(t)$ désigne la dérivée à droite de φ en t.

Nous appellerons *fonctionnelle de Cramer régularisée* associée au champ μ_x, $x\in M$, l'application $\bar{\Lambda}_{0,T}$: $C_{0,T}(M)\to[0,+\infty]$ définie par

$$\bar{\Lambda}_{0,T}(\varphi)=\varliminf_{\substack{\psi\to\varphi\\ \psi \text{ chemin intérieur}}} \Lambda_{0,T}(\psi),\quad \text{où}\quad \varphi\in C_{0,T}(M)$$

D'après le lemme 4.3 on a toujours $\Lambda_{0,T}(\varphi)\leqq\bar{\Lambda}_{0,T}(\varphi)$ et l'égalité a évidemment lieu si φ est un chemin intérieur. Mais en général $\Lambda_{0,T}\not\equiv\bar{\Lambda}_{0,T}$ et $\bar{\Lambda}_{0,T}$ n'est pas s.c.i.

(4.6) **Proposition.** *Soit μ_x un champ de mesures sur M vérifiant 4.1. Supposons de plus que les lignes de niveau de λ_x varient de façon lipschitzienne (cf. section 3). Alors pour tout $T>0$, $\Lambda_{0,T}$ est identique à $\bar{\Lambda}_{0,T}$.*

Preuve. Nous écrirons la démonstration dans le cas $M=\mathbb{R}^k$, ce qui entraîne le résultat dans le cas général, car on verra que le chemin à vitesses intérieures «approximant» φ peut être choisi partant d'un point arbitraire z pourvu que z soit assez proche de $\varphi(0)$. Pour $x\in M$, $\theta\in[0,T]$, considérons l'application affine $h_{x,\theta}$ de $\mathbb{R}^k$ dans $\mathbb{R}^k$ définie par

$$h_{x,\theta}(v)=\theta b_x+(1-\theta)v,\qquad v\in\mathbb{R}^k. \tag{4}$$

Vérifions que pour tout compact V de M il existe des constantes C et $C_1>0$ telles que pour tout $\theta\in[0,1]$, $a\geqq 0$, la relation $C_1\theta\geqq Ca$ implique

$$h_{x,\theta}(S_x^a)\subset S_x^{-[C_1\theta-Ca]}\qquad x\in V. \tag{5}$$

Il suffit de montrer qu'il existe un angle $\beta>0$ tel que pour tout $x\in V$, pour tout point extrémal v de S_x, S_x contient un cône de révolution de sommet v, d'axe $b_x v$, et ayant un angle au sommet supérieur à β. Or, les lignes de niveau de λ_x étant localement lipschitziennes, il existe $C_2>0$ tel que

$$h(S_x,S_y)<C_2|x-y|\qquad \text{pour } x,y\in V. \tag{6}$$

Comme S_x engendre $\mathbb{R}^k$, on en déduit l'existence de $C_3>0$ tel que S_x contienne la boule B_x de centre b_x et de rayon C_3 pour tout $x\in V$. Les S_x, $x\in V$ étant de diamètres bornés, et convexes, les cônes de base B_x, de sommet $v\in\partial S_x$ ont la propriété annoncée, et (5) est démontrée.

Soit φ un chemin dans M tel que $\Lambda_{0,T}(\varphi)<+\infty$, de sorte que $\varphi'(t)\in S_{\varphi(t)}$ pour presque tout $t\in[0,T]$. Soit V un voisinage compact de $\varphi[0,T]$.

Fixons $\theta\in]0,1[$. Pour presque tout $t\in[0,T]$, posons

$$v(t)=h_{\varphi(t),\theta}(\varphi'(t)).$$

Soit $p_{x,I}$, $x\in M$, $I\geqq 0$ la projection sur le convexe compact $\Sigma(\mu_x, I)$ (notations 3.1), c'est-à-dire que pour $v\in\mathbb{R}^k$, $p_{x,I}(v)$ est le point de $\Sigma(\mu_x, I)$ à distance minimum de v. Posons

$$I(t)=\lambda_{\varphi(t)}[v(t)]$$

et considérons l'équation différentielle

$$\psi'(t)=p_{\psi(t),I(t)}[v(t)] \qquad \text{pour presque tout } t\in[0,T] \tag{7}$$

où $\psi(t)$ est un chemin continu, absolument continu sur $[0,T]$, vérifiant $\psi(0)=\varphi(0)$. Ceci revient à résoudre l'équation

$$\psi(t)=\psi(0)+\int_0^t F(u,\psi(u))\,du, \qquad t\in[0,T] \tag{8}$$

où $F(u,x)=p_{x,I(u)}[v(u)]$ est une fonction borélienne du couple $(u,x)\in[0,T]\times\mathbb{R}^k$ et est localement lipschitzienne en x, uniformément en $u\in[0,T]$. La méthode classique d'itération prouve l'existence d'une solution $\psi(t)$ aux problèmes équivalents (7) et (8). Par construction, pour presque tout $t\in[0,T]$, $|\psi'(t)-v(t)|$ est majoré par la distance de Hausdorff entre $\Sigma(\mu_{\psi(t)}, I(t))$ et $\Sigma(\mu_{\varphi(t)}, I(t))$; l'hypothèse faite sur les lignes de niveau de λ_x entraîne donc

$$|\psi'(t)-v(t)|\leqq C_2|\varphi(t)-\psi(t)|, \qquad \text{pour presque tout } t\in[0,T]. \tag{9}$$

Comme $|h_{x,\theta}(w)-w|$ est majoré par $C_4\theta$ lorsque x reste dans un compact fixe de M (choisi assez grand pour contenir un voisinage de $[0,T]$ et de $\psi[0,T]$), on obtient

$$|\psi'(t)-\varphi'(t)|\leqq C_4\theta+C_2|\psi(t)-\varphi(t)|$$

pour presque tout $t\in[0,T]$. Une technique classique permet d'en déduire (considérer la fonction $u(t)=\int_0^t|\psi(u)-\varphi(u)|\,du$) que, en posant $C_5=C_4\vee C_2$,

$$\rho_{0,T}(\psi,\varphi)\leqq C_5\theta Te^{C_5T}. \tag{10}$$

D'autre part, la définition de $v(t)$ et la relation (5) entraînent $v(t)\in S_{\varphi(t)}^{-C_1\theta}$. En tenant compte de (9), (10) et de (6) on conclut que

$$\psi'(t)\in S_{\psi(t)}^{-\delta} \text{ pour presque tout } t\in[0,T],$$

où

$$\delta=\theta[C_1-2C_2C_5Te^{C_5T}].$$

Par conséquent il existe $C_6>0$ tel que la relation $T\leqq C_6$ entraîne

$$\psi'(t)\in S_{\psi(t)}^{-\frac{1}{2}C_1\theta} \text{ pour presque tout } t\in[0,T]. \tag{11}$$

Enfin, la relation (7) entraîne

$$\lambda_{\psi(t)}[\psi'(t)]\leqq I(t)=\lambda_{\varphi(t)}[v(t)], \qquad \text{pour presque tout } t\in[0,T]$$

tandis que la convexité de λ, et l'égalité $\lambda_x(b_x)=0$ et la définition de $h_{x,\theta}$ impliquent

$$\lambda_x \circ h_{x,\theta} \leqq \lambda_x.$$

On en conclut que

$$\lambda_{\psi(t)}[\psi'(t)] \leqq \lambda_{\varphi(t)}[\varphi'(t)]$$

pour presque tout $t\in[0,T]$, et par suite que $\Lambda_{0,T}(\psi)<\Lambda_{0,T}(\varphi)$.

Donnons nous $\varepsilon>0$, et choisissons θ assez petit pour que $\rho_{0,T}(\psi,\varphi)\leqq\varepsilon$. D'après le lemme 4.4, il existe un chemin à vitesses intérieures (cf. def. 4.5) ψ_1: $[0,T]\to M$ tel que $\rho_{0,T}(\psi_1,\psi)\leqq\varepsilon$, $\psi_1(0)=\psi(0)$, et $|\Lambda_{0,T}(\psi_1)-\Lambda_{0,T}(\psi)<\varepsilon$, ce qui entraîne $\rho_{0,T}(\varphi,\psi_1)\leqq 2\varepsilon$ et $\Lambda_{0,T}(\psi_1)\leqq\Lambda_{0,T}(\varphi)+\varepsilon$. En faisant tendre ε vers 0, on déduit alors de la définition de $\bar{\Lambda}$ que

$$\bar{\Lambda}_{0,T}(\varphi)\leqq\Lambda_{0,T}(\varphi).$$

Puisque l'inégalité opposée est toujours vraie (cf. 4.5), on conclut que $\bar{\Lambda}_{0,T}\equiv\Lambda_{0,T}$.

5. Processus d'apprentissage lent associé à un champ de mesures

(5.1) Soit M une variété Riemannienne connexe, $T(M)$ le fibré tangent de M. Nous appelons *champ de vecteurs borélien* V sur M toute section borélienne $x\to V_x\in T_x(M)$ de $T(M)$. L'ensemble $U(M)$ des champs de vecteurs boréliens est muni de la σ-algèbre engendrée par les applications η_x: $V\to(x,V_x)$ de $U(M)$ dans $T(M)$, où x décrit M. Si $(\Omega,\mathscr{B},P)$ est un espace de probabilité, un champ de vecteurs aléatoire V: $\Omega\to U(M)$ définit naturellement un champ de mesures μ_x, $x\in M$, où la probabilité μ_x sur $T_x(M)$ est la loi de V_x.

Inversement, si (μ_x), $x\in M$ est un champ de mesures *faiblement continu*, on peut toujours construire un espace de probabilité $(\Omega,\mathscr{B},P)$ et un champ de vecteurs aléatoire V: $\Omega\to U(M)$ tel que μ_x soit la loi de V_x pour $x\in M$. Bien entendu, la donnée des μ_x, $x\in M$ ne détermine absolument pas les lois jointes de $(V_{x_1},\ldots,V_{x_p})$, $x_1\ldots x_p\in M$.

(5.2) Soient $(\Omega,\mathscr{B})$ un espace mesurable, X_n: $\Omega\to M$, $n\geqq 0$, une suite de variables aléatoires à valeurs dans M, $\mathscr{F}_n$, $n\geqq 0$ une suite croissante de sous-σ-algèbres de $\mathscr{B}$, et P^τ, $\tau>0$ une famille de probabilités sur $(\Omega,\mathscr{B})$. Les X_n sont supposés $\mathscr{F}_n$-mesurables. Nous dirons que la famille de processus aléatoires $\mathscr{M}^\tau=\{\Omega,\mathscr{B},P^\tau,(X_n)_{n\geqq 0},(\mathscr{F}_n)_{n\geqq 0}\}$ indexée par $\tau>0$ est un *processus d'apprentissage lent associé au champ de mesures* μ_x, $x\in M$, si il existe une suite V^n: $\Omega\to U(M)$, $n\geqq 0$ de champs de vecteurs aléatoires ayant pour tout $\tau>0$ les propriétés suivantes:

(i) V^n est $\mathscr{F}_{n+1}$-mesurable pour tout $n\geqq 0$.

(ii) pour tout $\tau>0$, on a P^τ-p.s. l'égalité

$$P^\tau(V_x^n\in A|\mathscr{F}_n)=\mu_x(A)$$

quels que soient $n\geqq 0$, $x\in M$, $A\in\mathscr{B}[T_x(M)]$.

(iii) pour toute carte locale φ: $U\to\mathbb{R}^k$ (U ouvert de M), et pour tout compact $K\subset U$, il existe des nombres $\alpha>0$, $C>0$, tels que pour $\tau\leqq\alpha$ et $n\geqq 0$, la relation

$X_n \in K$ implique P^τ-presque sûrement la relation $X_{n+1} \in U$ et l'inégalité

$$|\varphi(X_{n+1}) - \varphi(X_n) - \tau\, d\varphi_{X_n}(V^n_{X_n})| \leqq C\tau^2. \tag{1}$$

De façon plus intuitive, ceci revient à dire que sous P^τ, et pour τ petit, X_{n+1} est très proche de la position atteinte à l'instant τ par un point décrivant un chemin tangent à $V^n_{X_n}$, partant de X_n à l'instant 0 et se déplaçant avec la vitesse scalaire constante $|V^n_{X_n}|$.
D'autre part, la loi conditionnelle de $V^n_{X_n}$, étant donné le passé $\mathscr{F}_n$ du processus $\mathscr{M}^\tau$ à l'instant n, ne dépend que de X_n et est égale à μ_{X_n}.

Tous les processus d'apprentissage lent décrits dans Norman [8] rentrent dans la catégorie décrite ci-dessus, avec en fait $M =$ ouvert de $\mathbb{R}^k$ et $C=0$. Notons en passant que les problèmes que nous allons considérer ci-dessous sont non triviaux même quand $M=\mathbb{R}$, $C=0$, et $\mu_x \equiv \mu$ (probabilité fixe sur $\mathbb{R}^k$) pour tout $x \in M$. En ce qui concerne ce cas particulier, nous renvoyons à Varadhan [13].

(5.3) Dans la suite, nous supposerons toujours que le champ de mesures μ_x associé aux $\mathscr{M}^\tau$, $\tau > 0$, vérifie toutes les hypothèses 4.1. Dans ces conditions, on montre facilement que pour toute paire de compacts $K_0 \subset K_1$ de M, il existe T fini et $\alpha > 0$ tels que pour $\tau \leqq \alpha$, la relation $X_0 \in K_0$ impliqué P^τ-p.s. les relations $X_n \in K_1$ pour $0 \leqq n\tau \leqq T$.

(5.4) **Lemme.** *Soit μ_x, $x \in M$, un champ de mesures vérifiant 4.1. Il existe une infinité de processus d'apprentissage lent ($\mathscr{M}^\tau$, $\tau > 0$) associés au champ μ_x, $x \in M$.*

Preuve. Soit W un champ de vecteurs aléatoire défini sur l'espace de probabilité $(\Omega_1, \mathscr{B}_1, P_1)$, tel que pour tout $x \in M$, W_x soit de loi μ_x. Sur $(\Omega_2, \mathscr{B}_2, P_2)$, produit d'une famille dénombrable de copies de $(\Omega_1, \mathscr{B}_1, P_1)$, on obtient en composant W avec les projections naturelles $\Omega_2 \to \Omega_1$, une suite W^n, $n \geqq 0$ de champs de vecteurs aléatoires, indépendants, de même loi que W.

Après un changement (scalaire en chaque point) éventuel de la métrique de M, on peut supposer que les arcs géodésiques ouverts de M de longueur finie sont relativement compacts. Soit $g(x, v, s)$ l'extrémité de l'arc géodésique issu de $x \in M$ tangent en x à $v \in T_x(M)$ et de longueur $s \geqq 0$. Fixons $y \in M$ arbitrairement, et définissons sur Ω_2 des v.a. Y^τ_n, $n \geqq 0$ en posant $Y^\tau_0 = y$ et

$$Y^\tau_{n+1} = g(Y^\tau_n, W^n_{Y^\tau_n}, \tau |W^n_{Y^\tau_n}|).$$

Soit $(\Omega, \mathscr{B})$ l'espace produit $(M \times U(M))^{\mathbb{N}}$, muni de sa tribu produit $\mathscr{B}$ et de la suite d'applications coordonnées naturelles (X_n, v^n) de Ω dans $M \times U(M)$, $n \geqq 0$. Soit Θ_τ l'application mesurable de Ω_2 dans Ω définie par

$$X_n \circ \Theta_\tau = Y^\tau_n \qquad V^n \circ \Theta_\tau = W^n.$$

Soit P^τ l'image de P_2 par Θ_τ; soit $\mathscr{F}_n$ la σ-algèbre de parties de Ω engendrée par $X_0 \dots X_n$, $V^0 \dots V^{n-1}$, $n \geqq 1$ ($\mathscr{F}_0$ engendrée par X_0 seulement). Alors les processus $\mathscr{M}^\tau = \{\Omega, \mathscr{F}_n, X_n, P^\tau\}$ forment pour $\tau > 0$ un processus d'apprentissage lent associé au champ de mesures μ_x, comme on le vérifie facilement – noter que les V^n: $\Omega \to U(M)$ sont bien les champs de vecteurs vérifiant (i), (ii), (iii).

(5.5) *Convergence vers le chemin moyen.* Pour $x \in M$, appelons $\bar{x}(t)$, $t>0$ la courbe intégrale du champ de vecteurs moyen (cf. 4.1) défini par le champ de mesures μ_y, $y \in M$, telle que $\bar{x}(0)=x$. Le résultat suivant indique que, quand $\tau \to 0$, le processus $\mathscr{M}^\tau$ se comporte pour $n\tau \leqq T$ fixé comme le processus «déterministe» qui part de X_0 et suit la trajectoire $\bar{X}_0(t)$ du champ moyen. Nous le donnons sans démonstration pour alléger l'exposé, car ce résultat n'est pas utilisé ci-dessous.

(5.6) **Proposition.** *Soit $\mathscr{M}^\tau$, $\tau>0$ un processus d'apprentissage lent associé à un champ de mesures μ_x, $x \in M$, vérifiant 4.1. Supposons de plus que pour tout compact K il existe une constante a_K telle que (notations de 5.2) la relation $|V_x^n - V_y^n| \leqq a_K |x-y|$ pour tout $x, y \in K$, tout n entier, soit vraie P^τ-p.s., dès que $\tau \in \left]0, \frac{1}{a_K}\right]$.*

Alors, pour tout compact $K \subset M$, pour tout temps S fini, il existe des nombres $C>0$, $\alpha>0$ ayant la propriété suivante: dès que $\tau \leqq \alpha$, la relation $X_0 \in K$ implique P^τ-presque sûrement l'inégalité

$$P^\tau\{\sup_{0 \leqq n\tau \leqq S} \rho[X_n, \bar{X}_0(n\tau)] \geqq C\tau^{1/10} | \mathscr{F}_0\} \leqq C\tau^{11/10}.$$

(5.7) Supposons dans ce paragraphe que P^τ-p.s. on ait $X_0 = x \in M$. Au processus $\mathscr{M}^\tau$ on peut associer une probabilité Q^τ sur l'espace $C(M)$ des fonctions continues sur $[0, T]$ à valeurs dans M. Il suffit par exemple d'associer à P^τ-presque toute trajectoire $\{X_n(\omega), n \geqq 0\}$ — où $\omega \in \Omega$ — de $\mathscr{M}^\tau$ le chemin continu $\varphi_\omega(t)$ défini par $\varphi_\omega(n\tau) = X_n$ et coïncidant avec un arc de géodésique entre X_n et X_{n+1}. La question du choix d'un tel arc (s'il en existe plusieurs) est un peu ennuyeuse car il faut garantir la mesurabilité de φ: $\omega \to \varphi_\omega$ comme application de Ω dans $C(M)$. Ce n'est cependant pas une difficulté sérieuse. On définit alors Q^τ comme l'image de P par φ: $\Omega \to C(M)$. Dès que le champ μ_x, $x \in M$, vérifie 4.1, il est facile de voir directement que, lorsque $\tau \to 0$, Q^τ converge faiblement vers Q, masse de Dirac au point $\bar{x} \in C(M)$, où $\bar{x}$ est la trajectoire (sur $[0 T]$) du champ moyen issue de x.

Le comportement asymptotique des $\mathscr{M}^\tau$, $\tau>0$ quand $\tau \to 0$ est donc très simple si on considère des intervalles de temps fixés ($0 \leqq n\tau \leqq T$). Par contre, la description d'évènements rares associés au processus $\mathscr{M}^\tau$, c'est à dire dont la réalisation exige des intervalles de temps dont la longueur moyenne croît exponentiellement avec τ, est beaucoup plus complexe. C'est cette étude que nous abordons maintenant.

6. Majoration de la probabilité d'un tube de chemins

(6.1) *Notations — Hypothèses.* On considère sur la variété M un champ de mesures μ_x, $x \in M$, vérifiant 4.1. On conserve les notations de 4.1.

La distance $\rho_{S,T}(\gamma_1, \gamma_2)$ entre deux chemins γ_1, γ_2 de M a été définie en 4.1. Si A est un ensemble de chemins de $[S\, T]$ dans M et si γ est un tel chemin, on pose

$$\rho_{S,T}(\gamma, A) = \sup_{\gamma_1 \in A} \rho_{S,T}(\gamma, \gamma_1).$$

Si ν est une probabilité sur $\mathbb{R}^n$, et si r est un entier, on note $\nu^{(r)}$ la loi de $\frac{1}{r}(Z_1+\cdots+Z_r)$ où $Z_1, \ldots, Z_r$ sont des v.a. indépendantes de même loi ν.

(6.2) **Lemme.** *Soit M un ouvert de $\mathbb{R}^n$ et soit μ_x, $x\in M$ un champ de mesures vérifiant 4.1.*

Pour $\theta\in[0,1]$ considérons l'application affine

$$h_{x,\theta}(v)=\theta\, b_x+(1-\theta)v, \quad \text{où } v\in\mathbb{R}^n,\ x\in M.$$

Posons $\nu_{x,\theta}=h_{x,\theta}(\mu_x)$.

A tout compact V de M on peut alors associer des fonctions finies positives $E(\delta)$ où $\delta>0$, et $N(\delta,\varepsilon)$ où $\delta>0$, $\varepsilon>0$ telles que la relation $r\geqq N(\delta,\varepsilon)$ entraîne

$$\nu_{x,\theta}^{(r)}\{v\in\mathbb{R}^n\colon \lambda_x(v)\geqq I \text{ et } v\in S_x^{-\delta}\}\leqq e^{-r[I-\varepsilon E(\delta)]}$$

quels que soient $x\in V$, $\theta\in[0,1]$, $I\geqq 0$, $\delta>0$, $\varepsilon>0$.

Preuve. Posons pour $I\geqq 0$, $\delta>0$, $x\in M$.

$$L(x,I,\delta)=\{v\in S_x^{-\delta}\colon \lambda_x(v)\geqq I\}.$$

La fonction $\left|\frac{\partial}{\partial v}\lambda_x(v)\right|$ est continue sur le compact $\bigcup_{x\in V}(x\times S_x^{-\delta/2})$ de $M\times\mathbb{R}^n$; soit D_δ sa borne supérieure sur ce compact. Alors pour $0<\varepsilon<\delta/2$, $I\geqq 0$, et $x\in V$ on a

$$L(x,I,\delta)^\varepsilon\subset L(x,I-\varepsilon D_\delta,\delta-\varepsilon). \tag{2}$$

Le théorème de Strassen et la formule

$$|h_{y,\theta}(v)-h_{z,\theta}(w)|\leqq|v-w|+|b_y-b_z|, \quad y,z\in M,\ v,w\in\mathbb{R}^n \tag{3}$$

prouvent, grâce aux hypothèses sur les μ_x, que les applications $\varphi_\theta\colon x\to\nu_{x,\theta}$ indexées par $\theta\in[0,1]$ forment une famille équicontinue, $M_c(\mathbb{R}^n)$ étant muni de la topologie (T) de 2.11. D'autre part les $\nu_{x,\theta}$, $x\in V$, $\theta\in[0,1]$ forment évidemment un (T)-compact de $M_c(\mathbb{R}^n)$. Le théorème de robustesse uniforme (cor. 2.12 du théorème 2.3) appliqué à la famille des $\nu_{x,\theta}$, $x\in V$, $\theta\in[0,1]$ fournit donc $a_1=a_1(\delta,\varepsilon,V)>0$ — noter que a_1 est indépendant de $\theta\in[0,1]$ — tel que les relations

$$|x-y|\leqq a_1, \quad x\in V,\ y\in M, \text{ et } r\geqq 1/a_1$$

entraînent

$$\nu_{y,\theta}^{(r)}(A)\leqq e^{r\varepsilon}\,\nu_{x,\theta}(A^\varepsilon) \quad \text{pour tout } A\in\mathscr{B}(\mathbb{R}^n),\ \theta\in[0,1]. \tag{4}$$

D'autre part, $\lambda_y(v)$ est uniformément continue sur le compact $\bigcup_{y\in V} y\times S_y^{-\delta/2}$, et la fonction $y\to S_y$ est uniformément continue sur V. Donc il existe $a_2=a_2(\delta,\varepsilon,V)>0$, avec $a_2\leqq a_1$ tel que les relations $|x-y|\leqq a_2$, $x,y\in V$, $I\geqq 0$ entraînent

$$L\left(y,I,\frac{\delta}{2}\right)\subset L\left(x,I-\varepsilon,\frac{\delta}{4}\right). \tag{5}$$

Les conclusions (2), (4), (5) entraînent que pour $x, y \in V$, $|x-y| \leqq a_2$, $r \geqq 1/a_1$, $\varepsilon \leqq \delta/2$

$$\nu_{y,\theta}^{(r)}[L(y, I, \delta)] \leqq e^{r\varepsilon} \nu_{x,\theta}[L(x, I-\varepsilon(1+D_\delta), \delta/4)]. \tag{6}$$

Soit $J(\delta)$ la borne supérieure (finie) de $\lambda_x(v)$ sur le compact $\bigcup_{x \in V} x \times S_x^{-\delta/4}$. Soit $0 = J_0 < J_1 < \cdots < J_k = J(\delta)$ des nombres tels que

$$J_{p+1} - J_p \leqq \varepsilon \qquad \text{pour } p = 0 \ldots k-1.$$

Soit $\{x_1 \ldots x_\ell\}$ un sous-ensemble fini de V tel que V soit contenu dans l'union des boules de centre x_i, $i \leqq \ell$ et de rayon a_2.

Pour $x \in M$, $J \geqq 0$, $\varepsilon > 0$, il existe (voir section 1.9) un nombre $N(x, J, \varepsilon)$ tel que

$$\mu_x^{(r)}[L(x, J, 0)] \leqq e^{-r(J-\varepsilon)} \qquad \text{pour } r \geqq N(x, J, \varepsilon). \tag{7}$$

Posons

$$N_1 = \frac{1}{a_1} + \sup\{N(x_i, J_p, \varepsilon) \colon 1 \leqq i \leqq \ell,\ 0 \leqq p \leqq k\}. \tag{8}$$

Par construction, N_1 se calcule à partir de δ, ε, V.

Puisque λ_x est positive, convexe, nulle en b_x, on a

$$\lambda_x(h_{x,\theta}(v)) \leqq \lambda_x(v) \qquad \text{pour } x \in M,\ v \in \mathbb{R}^n,\ \theta \in [0, 1].$$

Soient $Z_1 \ldots Z_r$ des v.a. indépendantes de même loi μ_x, définies sur $(\Omega', \mathscr{B}', P')$; notons $\bar{Z}_r$ leur moyenne. Par définition de $\nu_{x,\theta}^{(r)}$, on a

$$\nu_{x,\theta}^{(r)}[L(x, J, \delta/4)] = P'\{\lambda_x[h_{x,\theta}(\bar{Z}_r)] \geqq J,\ h_{x,\theta}(\bar{Z}_r) \in S_x^{-\delta/4}\}$$

et donc

$$\nu_{x,\theta}^{(r)}[L(x, J, \delta/4)] \leqq I_{\{J \leqq J(\delta)\}} P'\{\lambda_x(\bar{Z}_r) \geqq J\}$$

c'est-à-dire

$$\nu_{x,\theta}^{(r)}[L(x, J, \delta/4)] \leqq I_{\{J \leqq J(\delta)\}} \mu_x^{(r)}[L(x, J, 0)]. \tag{9}$$

Pour $J \in [0, J(\delta)]$ il existe un entier $p \in [0, k-1]$ tel que

$$J - \varepsilon \leqq J_p \leqq J \leqq J_{p+1}.$$

Donc si $1 \leqq i \leqq \ell$, $0 \leqq J \leqq J(\delta)$, les inégalités (7), (8), (9) entraînent pour $r \geqq N_1$, $\theta \in [0, 1]$

$$\nu_{x_i,\theta}^{(r)}[L(x_i, J, \delta/4)] \leqq \mu_{x_i}^{(r)}[L(x_i, J_p, 0)] \leqq e^{-r(J_p-\varepsilon)} \leqq e^{-r(J-2\varepsilon)}.$$

L'inégalité ainsi obtenue

$$\nu_{x_i,\theta}^{(r)}[L(x_i, J, \delta/4)] \leqq e^{-r(J-2\varepsilon)} \tag{10}$$

reste strictement vraie pour $J \geqq J(\delta)$, d'après (9).

Combinons (6), (10) et la définition des x_i pour en déduire que pour $y \in V$, $I \geqq 0$, $\theta \in [0,1]$, $0 < \varepsilon \leqq \delta/2$, il existe $N_1(\delta, \varepsilon, V)$ tel que $r \geqq N_1$ entraîne

$$\nu_{y,\theta}^{(r)}[L(y, I, \delta)] \leqq e^{-r(\dot{I} - \varepsilon E(\delta))}$$

où $E(\delta) = 4 + D_\delta$ est déterminé par δ et V. Q.E.D.

(6.3) **Théorème.** *Soit M une variété riemannienne et μ_x, $x \in M$ un champ de mesures vérifiant 4.1. Soit $\mathcal{M}^\tau = \{\Omega, \mathcal{B}(\Omega), P^\tau, X_n\}$ une famille de processus d'apprentissage (indexée par $\tau > 0$) associée au champ μ_x (cf. section 5). Supposons que les transformées de Cramer λ_x du champ μ_x aient des lignes de niveau localement lipschitziennes (cf. Section 3). Notons Λ la fonctionnelle de Cramer (cf. section 4) associée au champ μ_x. Etant donnés $T > 0$, $I \geqq 0$, soit $\Gamma(I)$ l'ensemble des chemins $\gamma\colon [0, T] \to M$ tels que $\Lambda_{0,T}(\gamma) \leqq I$. Soit γ^τ le chemin aléatoire défini par: $\gamma^\tau(t) = X_n$ pour $n\tau \leqq t < (n+1)\tau$.*

Pour tout $\varepsilon > 0$ et pour tout compact $K \subset M$, il existe $\tau_0 > 0$ tel que pour $\tau < \tau_0$, l'inégalité (notations de 6.1)

$$P^\tau\{\rho_{0,T}(\gamma^\tau, \Gamma(I)) \geqq \varepsilon | \mathcal{F}_0\} \leqq e^{-\frac{1}{\tau}(I-\varepsilon)}$$

a lieu P^τ-p.s. sur l'ensemble $\{X_0 \in K\}$.

Preuve. On suppose d'abord que M est un ouvert de $\mathbb{R}^k$. Fixons des compacts K, U, V de M tels que $K \subset \mathring{U}$, $U \subset \mathring{V}$.

Pour $x \in M$, $\theta \in [0,1]$, définissons la transformation affine $h_{x,\theta}$ de $\mathbb{R}^k$ dans $\mathbb{R}^k$ comme en 6.2.

Pour $I > 0$, $v \in \mathbb{R}^k$ et $x \in M$, soit $\tilde{v}_x(I)$ le point du convexe compact $\Sigma(\mu_x, I)$ (cf. 3.1) réalisant le minimum de la distance de v à $\Sigma(\mu_x, I)$. Le champ de vecteur $\tilde{v}(I)\colon x \to \tilde{v}_x(I)$ est localement lipschitzien. Si $\eta(t, y, v, I)$, $t \geqq 0$ est la courbe intégrale de $\tilde{v}(I)$ partant de $y \in M$ à l'instant $t = 0$, η est une fonction borélienne de (t, y, v, I).

Notons $V_n = V_{X_n}^n$ (cf. section 5) les «vitesses» du processus $\mathcal{M}^\tau$, de sorte que P^τ-p.s., $V_n \in S_{X_n}$ et la loi conditionnelle de V_n étant donnée $\mathcal{F}_n$ est μ_{X_n}.

Soit $r \geqq 1$ un entier (déterminé plus bas). Posons pour $j \geqq 0$

$$\bar{V}_{jr} = \frac{1}{r}(V_{jr} + V_{jr+1} + \cdots + V_{jr+r-1})$$

$$W_{jr} = h_{X_{jr},\theta}(\bar{V}_{jr})$$

$$I_j = \lambda_{X_{jr}}(W_{jr})$$

et soit $\tilde{W}_{jr}(I_j)$ le champ de vecteurs (aléatoire) associé comme ci-dessus au vecteur W_{jr} et au nombre I_j.

Soit Ω_0 l'ensemble $\{X_0 \in K\} \subset \Omega$. D'après la section 4.1, il existe $T = T(K, V) > 0$ tel que, pour tout chemin $\gamma\colon [0, T] \to M$, les conditions $\gamma(0) \in K$ et $\gamma'(t) \in S_{\gamma(t)}$, $t \in [0, T]$, impliquent $\gamma(t) \in U$ pour $t \in [0, T]$.

Pour P^τ-p.tt. $\omega \in \Omega_0$, il existe un unique chemin continu différentiable par morceaux, $\eta_\omega\colon [0, T] \to U$ tel que $\eta_\omega(0) = X_0$ et tel que pour tout entier $j \in [0, T/r\tau - 1]$, sur $[jr\tau, (j+1)r\tau]$, η_ω soit une courbe intégrale du champ de vecteurs $\tilde{W}_{jr}(I_j)$ défini plus haut. On note η le chemin aléatoire ainsi défini sur Ω_0.

Soit $t \in [j\,r\,\tau, (j+1)\,r\,\tau[$; par construction, $\eta'(t)$ et W_{jr} appartiennent à $\Sigma(\mu_{\eta(t)}, I_j)$ et $\Sigma(\mu_{X_{jr}}, I_j)$. Les lignes de niveau de λ_x étant localement lipschitziennes, il existe $C_4 > 0$ tel que

$$|\eta'(t) - W_{jr}| \leqq C_4\,|\eta(t) - X_{jr}|.$$

La fonction $\varphi(t) = \eta(t) - (t - j\,r\,\tau)\,W_{jr} - X_{jr}$ vérifie donc

$$|\varphi'(t)| \leqq C_5\,|\varphi(t) + (t - j\,r\,\tau)|$$

puisque P^τ-p.s. sur Ω_0, les $|W_{jr}|$ sont majorés par une constante dépendant de U seulement. Par une manipulation classique utilisant la fonction $\int_{jr\tau}^{t} |\varphi'(s)|\,ds$, on en conclut que

$$|\varphi(t)| \leqq [|\varphi(j\,r\,\tau)| + C_5 (r\,\tau)^2]\,e^{C_5 r \tau} \qquad \text{pour } j\,r\,\tau \leqq t \leqq (j+1)\,r\,\tau. \tag{11}$$

Posons $a_j = \eta(j\,r\,\tau) - X_{jr}$. On a $\varphi(j\,r\,\tau) = a_j$ et

$$a_{j+1} = \varphi[(j+1)\,r\,\tau] + r\,\tau\,W_{jr} + X_{jr} - X_{(j+1)r}. \tag{12}$$

Par définition des processus $\mathscr{M}^\tau$, on a (cf. section 5) P^τ-p.s. sur Ω_0,

$$|X_{(j+1)r} - X_{jr} - r\,\tau\,\bar{V}_{jr}| \leqq C_6 (r\,\tau)^2. \tag{13}$$

Par définition, $h_{x,\theta}$ vérifie $|h_{x,\theta}(v) - v| \leqq C_7\,\theta$ pour $v \in (S_x)^1$, $x \in V$. Par conséquent pour $r\,\tau \leqq C_8$, $(j+1)\,r\,\tau \leqq T$, on aura

$$|W_j - \bar{V}_j| \leqq C_7\,\theta \qquad P^\tau\text{-p.s. sur } \Omega_0. \tag{14}$$

Les inégalités (11), (12), (13), (14) et la condition $r\,\tau \leqq C_8$, impliquent

$$|a_{j+1}| \leqq |a_j|\,e^{C_5 r\tau} + C_9\,r\,\tau\,[r\,\tau + \theta], \quad (j+1)\,r\,\tau \leqq T.$$

Un calcul immédiat donne alors $|a_j| \leqq C_{10}[r\tau + \theta]$ pour $(j+1)\,r\,\tau \leqq T$; puisque, P^τ-p.s. sur Ω_0, $\eta'(t)$ est borné par une constante dépendant de V seulement, on obtient

$$\rho_{0,T}(\eta, \gamma^\tau) \leqq C_{11}[r\,\tau + \theta], \qquad P^\tau\text{-p.s. sur } \Omega_0 \tag{15}$$

Posons $A_j = \int_{jr\tau}^{(j+1)r\tau} \lambda_{\eta(t)}[\eta'(t)]\,dt$.

Par définition du champ $\tilde{W}_{jr}(I_j)$ et de η, $\lambda_{\eta(t)}[\eta'(t)]$ est majoré par $\lambda_{X_{jr}}(W_{jr}) = I_j$, pour $t \in [j\,r\,\tau, (j+1)\,r\,\tau[$ et donc

$$A_j \leqq r\,\tau\,\lambda_{X_{jr}}(W_{jr}), \qquad P^\tau\text{-p.s. sur } \Omega_0. \tag{16}$$

Donnons nous $\delta > 0$, à déterminer plus bas, et imposons $C_1\,\theta \geqq 2\delta$ et $\delta \geqq C C_2\,r\,\tau$. D'après les relations (5) et (6) de la section 4.6, et la définition de W_j, ceci garantit pour $(j+1)\,r\,\tau \leqq T$,

$$W_j \in S_{X_{jr}}^{-\delta}, \qquad P^\tau\text{-p.s. sur } \Omega_0. \tag{16}$$

La définition de $h_{x,\theta}$ et les hypothèses sur les μ_x impliquent l'existence de $C_{12} > 0$ tel que pour tout $\theta \in [0, 1]$ on ait

$$d[h_{x,\theta}(\mu_y), h_{x,\theta}(\mu_z)] \leqq C_{12}\,|y - z| \qquad \text{pour } y, z \in V,\ x \in M. \tag{17}$$

La loi conditionnelle de $h_{X_{jr},\theta}(V_n)$ étant donné $\mathscr{F}_n$ est $h_{X_{jr},\theta}(\mu_{X_n})$. D'autre part, (17) entraîne, P^τ-p.s. sur Ω_0,

$$d[h_{X_{jr},\theta}(\mu_{X_n}), h_{X_{jr},\theta}(\mu_{X_{jr}})] \leqq C_{13}\, r\, \tau \tag{18}$$

pour $0 \leqq j\, r \leqq n \leqq (j+1)\, r \leqq T/\tau$.

Posons $v_{x,\theta} = h_{x,\theta}(\mu_x)$.

L'ensemble E des probabilités de la forme $v_{x,\theta}$, $x \in V$, $\theta \in [0,1]$ est (T)-compact (cf. 2.11). Le théorème de robustesse 2.3 s'applique donc uniformément au voisinage de tout point de E (cf. cor. 2.12).

En tenant compte de (20), de la définition de W_j et du fait que $h_{x,\theta}$ est affine, le théorème 2.3 et le corollaire 2.12 associent à tout $\varepsilon > 0$ un nombre $\alpha = \alpha(\varepsilon, V) > 0$ tel que: la relation $C_{13}\, r\, \tau < \alpha$ entraîne P^τ-p.s. sur Ω_0 la famille d'inégalités.

$$P^\tau(W_j \in B \,|\, \mathscr{F}_{jr}) \leqq e^{r\varepsilon}\, v^{(r)}_{X_{jr},\theta}(B^\varepsilon) \tag{19}$$

où $B \in \mathscr{B}(\mathbb{R}^k)$, $(j+1)\, r\, \tau \leqq T$, et $r \geqq 1/\alpha$.

Soit $z \to B_z$ une application de M dans $\mathscr{B}(\mathbb{R}^k)$ telle que $f(x,z) = I_{B_z}(x)$ soit borélienne sur $\mathbb{R}^k \times M$. Un argument standard montre l'existence de $\psi(z,\omega)$, fonction mesurable de $M \times \Omega$ dans $\mathbb{R}$, telle que $\psi(z,\cdot)$ soit pour tout $z \in M$ une version de $P^\tau(W_j \in B_z \,|\, \mathscr{F}_{jr})$, et telle que $\psi[X_{jr}(\cdot),\cdot]$ soit une version de $P^\tau(W_j \in B_{X_{jr}} \,|\, \mathscr{F}_{jr})$.

Reprenons les notations de 6.1, 6.2. L'argument précédent et (21) entraînent

$$P^\tau\{W_j \in L(X_{jr}, I, \delta) \,|\, \mathscr{F}_{jr}\} \leqq e^{r\varepsilon}\, v^{(r)}_{X_{jr},\theta}[L(X_{jr}, I, \delta)^\varepsilon].$$

Imposons $0 < \varepsilon < \delta/2$ et tenons compte de (2), (16) pour obtenir

$$P^\tau\{\lambda_{X_{jr}}(W_j) > I \,|\, \mathscr{F}_{jr}\} \leqq e^{r\varepsilon}\, v^{(r)}_{X_{jr},\theta}[L(X_{jr}, I - \varepsilon D_\delta, \delta/2]. \tag{20}$$

Utilisons le lemme 6.2 pour déterminer $N = N(\delta, \varepsilon, V)$ et $F = F(\delta.\, V)$ tels que $r > N$, $x \in V$, $\theta \in [0,1]$, $J \geqq 0$ entraînent

$$v^{(r)}_{x,\theta}[L(x, J, \delta/2)] \leqq e^{-r(J - \varepsilon F)}.$$

On en conclut que les relations $r \geqq N$, $C_{13}\, r\, \tau \leqq \alpha$, $r \geqq 1/\alpha$, $(j+1)\, r\, \tau \leqq T$, $C_2\, r\, \tau \leqq \delta/2$, $\delta \leqq C_1\, \theta$, $r\, \tau \leqq C_8$, $I \geqq 0$, entraînent P^τ-p.s. sur Ω_0,

$$P^\tau(\lambda_{X_{jr}}(W_j) \geqq I \,|\, \mathscr{F}_{jr}) \leqq e^{-r(I-E)}$$

où $E = \varepsilon(1+F)$.

Par suite (16) fournit la majoration valable P^τ-p.s. sur Ω_0, pour $(j+1)\, r\, \tau \leqq T$, $s \geqq 0$

$$P^\tau(A_j \geqq s \,|\, \mathscr{F}_{jr}) \leqq e^{-\frac{1}{\tau}(s - E r \tau)}. \tag{21}$$

Posons $\Lambda_j = A_0 + \cdots + A_j$. Supposons que pour un j tel que $(j+2)\, r\, \tau \leqq T$ on ait obtenu, P^τ-p.s. sur Ω_0

$$P^\tau(\Lambda_j \geqq s \,|\, \mathscr{F}_0) \leqq e^{-\frac{1}{\tau}(s - b_j)}, \qquad s \geqq 0. \tag{22}$$

De $\Lambda_{j+1} \doteq \Lambda_j + A_{j+1}$, on déduit, P^τ-p.s. sur Ω_0,

$$\begin{aligned} P^\tau(\Lambda_{j+1} \geqq s \,|\, \mathscr{F}_0) &\leqq P^\tau(\Lambda_j \geqq s - E\, r\, \tau \,|\, \mathscr{F}_0) \\ &\quad + E^\tau\{I_{\{\Lambda_j < s - E r \tau\}}\, P^\tau[A_{j+1} \geqq s - \Lambda_j \,|\, \mathscr{F}_{jr}] \,|\, \mathscr{F}_0\}. \end{aligned} \tag{23}$$

Notons $\omega \to \pi_\omega$ une version de la loi conditionnelle de Λ_j étant donné $\mathscr{F}_0$. Pour P^τ-presque tout $\omega \in \Omega_0$, le 2nd membre de (23) est, d'après (21), (22), majoré par

$$e^{-\frac{1}{\tau}(s-b_j-E r\tau)} + e^{-\frac{1}{\tau}(s-E r\tau)} \int_0^{s-E r\tau} e^{u/\tau}\, d\pi_\omega(u).$$

Ecrivons $e^{u/\tau} = \frac{1}{\tau} \int_{-\infty}^{u} e^{v/\tau}\, dv$ et utilisons le théorème de Fubini pour obtenir

$$\int_0^{s-E r\tau} e^{u/\tau}\, d\pi_\omega(u) = \frac{1}{\tau} \int_{-\infty}^{s-E r\tau} e^{v/\tau}\, \pi_\omega[v \vee 0, s-E\, r\, \tau]\, dv.$$

D'après (21) on majore $\pi_\omega[v \vee 0, s-E\, r\, \tau]$ par $e^{-\frac{1}{\tau}(v-b_j)}$ pour $b_j < v < s-E\, r\, \tau$, et par 1 pour $v \leqq b_j$ ce qui donne pour $s \geqq b_j + E\, r\, \tau$

$$\int_0^{s-E r\tau} e^{u/\tau}\, d\pi_\omega(u) \leqq e^{\frac{1}{\tau} b_j} \left(1 + \frac{s-b_j-E\, r\, \tau}{\tau}\right)$$

et finalement

$$P^\tau(\Lambda_{j+1} > s \mid \mathscr{F}_0) \leqq \left(2 + \frac{S}{\tau}\right) e^{-\frac{1}{\tau}(s-b_j-E r\tau)}$$

inégalité trivialement vraie pour $s < b_j + E\, r\, \tau$. Ainsi il est possible de choisir $b_{j+1} = b_j + 2E\, r\, \tau$ et de prouver (22) par récurrence, avec $b_j = (2j+1)\, E\, r\, \tau$, si on impose

$$\left(2 + \frac{S}{\tau}\right) \exp - \frac{E\, r\, \tau}{\tau} \leqq 1.$$

Soit $T' < T$. Dès que $r\tau < T - T'$, il existe j entier tel que $T' \leqq (j+1)\, r\, \tau \leqq T$. On a alors $b_j \leqq 2ET$ et $\Lambda_j \geqq \Lambda_{0,T'}(\eta)$. Par conséquent, si toutes les restrictions introduites sont satisfaites, on obtient, P^τ-p.s. sur Ω_0

$$P^\tau(\Lambda_{0,T'}(\eta) \geqq s \mid \mathscr{F}_0) \leqq P^\tau(\Lambda_j \geqq s \mid \mathscr{F}_0) \leqq e^{-\frac{1}{\tau}(s-2ET)}.$$

Donnons nous maintenant $\varepsilon_0 > 0$ et $T' < T$. Imposons $C_{11}\, \theta = \frac{1}{2} \varepsilon_0$ et $C_{11}\, r\, \tau \leqq \frac{1}{2} \varepsilon_0$ ce qui garantit

$$\rho_{[0,T]}(\eta, \gamma^\tau) \leqq \varepsilon_0.$$

Imposons $\delta = \frac{1}{2} C_1 \theta$ de sorte que $\delta = \frac{1}{2} C_1 / C_{11}\, \varepsilon_0$. Choisissons ε tel que $2ET = \varepsilon_0$ c'est-à-dire tel que $2T[1 + F(\delta)]\, \varepsilon = \varepsilon_0$. A partir de δ, ε on détermine $N = N(\delta, \varepsilon)$, $\alpha = \alpha(\varepsilon)$, comme indiqué c-dessus.

Fixons r_0 entier, vérifiant $r_0 \geqq N + 1/\alpha$ et $2 \exp(-E r_0) \leqq \frac{1}{2}$. Imposons $0 \leqq s \leqq S$, et à τ les conditions $C_{13}\, r_0\, \tau$, $C.C_2\, r_0\, \tau \leqq \delta$, $r_0\, \tau \leqq C_8$, $r_0\, \tau \leqq T - T'$, $\frac{S}{2} \exp(-E r_0) \leqq \frac{1}{2}$, ce qui revient à prendre $\tau < \tau_0$ où $\tau_0 > 0$. On obtient ainsi, dès que $\tau < \tau_0$ et $s \leqq S$, P^τ-p.s. sur Ω_0,

$$P^\tau[\Lambda_{0,T'}(\eta) \geqq s \mid \mathscr{F}_0] \leqq e^{-\frac{1}{\tau}(s-\varepsilon_0)}$$

et

$$\rho_{0,T}(\eta, \gamma^\tau) \leqq \varepsilon_0.$$

On voit que $\tau_0>0$ est déterminé à partir de $\varepsilon_0>0$, $T'<T$, et des compacts K, U, V.

Passons maintenant à la preuve du théorème 6.3 dans le cas général, où M est une variété quelconque. Donnons nous $T_1>0$ et un compact $K\subset M$. Il existe (cf. 4.1) un compact U vérifiant $\mathring{U}\supset K$, tel que tout chemin différientable par morceaux vérifiant, pour presque tout t, $\varphi'(t)\in(S_{\varphi(t)})^1$ et $\varphi(0)\in K$ reste dans U pour $t\in[0, T_1]$. On peut garantir aussi que $\gamma^\tau[0, T_1]\subset U$, P^τ-p.s. sur $\Omega_0=\{\omega|X_0(\omega)\in K\}$. Soit V un voisinage compact de U; fixons un recouvrement fini $V_1 \dots V_p$ de V par les domaines relativement compacts de cartes $\varphi_1 \dots \varphi_p$.

Construisons par récurrence un chemin (aléatoire) continu η à valeurs dans V continûment différentiable par morceaux tel que

$$\eta(0)=X_0, \qquad \rho_{0,T_1}(\eta, \gamma^\tau)\leqq\varepsilon, \qquad P^\tau\text{-p.s. sur } \Omega_0.$$

Soit k un entier tel que $T-\tau\leqq k\tau\leqq T$. Supposons η construit sur $[0, jk\tau]$, avec j entier, tel que $jk\tau\leqq T_1$, et vérifiant

$$\rho_{0,jk\tau}(\eta,\gamma^\tau)\leqq\varepsilon\frac{jk\tau}{T_1}.$$

Pour ε assez petit $\eta(jk\tau)$ et X_{jk} sont dans le domaine V_i d'une des cartes φ_i. Grâce au choix de T, on peut construire η sur $[jk\tau, (j+1)k\tau]$ dans la carte φ_i par la méthode indiquée dans le cas de $M=\mathbb{R}^n$; étant donné $\varepsilon>0$, et $S>0$, on peut ainsi garantir pour $\tau\leqq\tau_0$ les inégalités suivantes, vraies P^τ-p.s. su Ω_0

$$P\{\Lambda_{[jk\tau,(j+1)k\tau]}(\eta)\geqq s|\mathscr{F}_{jk}\}\leqq e^{-\frac{1}{\tau}(s-\varepsilon)} \tag{24}$$

où $s\leqq S$, et

$$\rho_{jk\tau,(j+1)k\tau}(\eta,\gamma^\tau)\leqq\varepsilon\frac{(j+1)k\tau}{T_1}.$$

Si $T_2<T_1$, on peut trouver j_0 entier tel que $(1+j_0)\leqq\dfrac{T_1}{T-\tau_0}$ et $T_2\leqq(j_0+1)k\tau\leqq T_1$.

Puisque $\Lambda_{0,T_2}(\eta)\leqq\sum\limits_{0\leqq j\leqq j_0}\Lambda_{jk\tau,(j+1)k\tau}(\eta)$ (24) implique trivialement, P^τ-p.s. sur Ω_0

$$P^\tau(\Lambda_{0,T_2}(\eta)\geqq s|\mathscr{F}_0)\leqq\frac{T_1}{T-\tau_0}e^{-\frac{1}{\tau}(s-\varepsilon)} \qquad \text{pour } s\leqq S.$$

Prenant $\tau<\tau_0$ assez petit pour que $\dfrac{T_1}{T-\tau_0}e^{-\frac{\varepsilon}{\tau}}\leqq 1$ on conclut que P^τ-p.s. sur Ω_0,

$$P^\tau(\Lambda_{0,T_2}(\eta)\geqq s|\mathscr{F}_0)\leqq e^{-\frac{1}{\tau}(s-2\varepsilon)}, \qquad s\leqq S.$$

D'autre part, on a simultanément, P^τ-p.s. sur Ω_0, $\rho_{0,T_2}(\eta,\gamma^\tau)\leqq\varepsilon$.

Donnons nous $I\in[0,S]$; et soit $\Gamma(I, T_2)$ l'ensemble des chemins continus ψ tels que $\Lambda_{0,T_2}(\psi)\leqq I$. Pour τ comme ci-dessus, on peut écrire P^τ-p.s. sur Ω_0,

$$\begin{aligned}P^\tau\{\rho_{0,T_2}[\gamma^\tau,\Gamma(I,T_2)]>\varepsilon|\mathscr{F}_0\}&\leqq P^\tau\{\eta\notin\Gamma(I,T_2)|\mathscr{F}_0\}\\&=P^\tau\{\Lambda_{0,T_2}(\eta)\geqq I|\mathscr{F}_0\}\leqq e^{-\frac{1}{\tau}(I-2\varepsilon)}\end{aligned}$$

ce qui prouve le théorème 6.3.

7. Minoration de la probabilité d'un tube de chemins

(7.1) *Notations et hypothèses.* On garde les mêmes notations et hypothèses qu'en 4.1 et 6.1. On note $B(v, a)$ la boule ouverte de centre v et de rayon a dans $\mathbb{R}^k$.

(7.2) **Lemme.** *(Hypothèses de 7.1). Supposons que M soit un ouvert de $\mathbb{R}^k$. Pour tout compact $K \subset M$, pour tout $\varepsilon>0$, $\delta>0$, $a>0$ il existe $N>0$ tel que (notations de 6.1),*

$$\mu_x^{(r)}[B(v, a)] \geqq e^{-r[\varepsilon+\lambda_x(v)]}$$

quels que soient $x \in K$, $v \in S_x^{-\delta}$.

Preuve. Donnons nous ε, $\delta>0$ et $K \subset M$. Il existe $\alpha>0$ tel que les relations $x, y \in K$, $|x-y| \leqq \alpha$ entraînent

$$\begin{cases} S_x^{-2\delta} \subset S_y^{-\delta} \\ |\lambda_x(v)-\lambda_y(\omega)| \leqq \varepsilon \qquad \text{pour } v, \omega \in \mathbb{R}^k,\ |v-\omega| \leqq \alpha. \end{cases} \tag{1}$$

Donnons nous $a \in]0, \alpha]$, et posons $\varepsilon_1 = \varepsilon \wedge a/3$. Les théorème 2.3 et corollaire 2.12 prouvent l'existence de N et de $\alpha_1>0$ tels que les relations $|x-y| \leqq \alpha_1$, $x, y \in K$, $r>N$ entraînent pour tout $B \in \mathscr{B}(\mathbb{R}^k)$

$$\mu_x^{(r)}(B) \geqq e^{-r\varepsilon_1} \mu_y^{(r)}(B^{-\varepsilon_1}). \tag{2}$$

Fixons un recouvrement fini de K par des boules de centres $y_1 \dots y_p \in K$, de rayon $\alpha \wedge \alpha_1$. Pour chaque y_i choisissons une famille finie de boules de $B_{i,1} \dots B_{i,p}$ de rayon $a/3$, de centres appartenant à $S_{y_i}^{-\delta}$, telle que toute boule de centre appartenant à $S_{y_i}^{-\delta}$ et de rayon $2a/3$ contienne au moins une boule B_{ij}. Il existe N_0 tel que $r \geqq N_0$ entraîne

$$\mu_{y_i}^{(r)}(B_{ij}) \geqq e^{-r(\varepsilon_1+\sigma_{ij})} \qquad i=1 \dots p,\ j=1 \dots q \tag{3}$$

où $\sigma_{ij} = \inf\limits_{v \in B_{ij}} \lambda_{y_i}(v)$.

Soit $x \in K$, et $v \in S_x^{-2\delta}$. Choisissons i tel que $|y_i - x| \leqq \alpha \wedge \alpha_1$ et B_{ij} telle que $B_{ij} \subset B(v, 2a/3)$. D'après (2), (3), pour $r \geqq N \vee N_0$ on a

$$\mu_x^{(r)}(B(v, a)) \geqq e^{-r\varepsilon_1} \mu_{y_i}^{(r)}(B(v, 2a/3)) \geqq e^{-r(2\varepsilon_1+\sigma_{ij})}. \tag{4}$$

D'autre part, d'après (1), σ_{ij} est majoré par $(\lambda_x(v)+\varepsilon)$; (4) devient donc, puisque $\varepsilon_1 \leqq \varepsilon$

$$\mu_x^{(r)}(B(v, a)) \geqq e^{-r(3\varepsilon+\lambda_x(v))}$$

inégalité vraie dès que $r \geqq N \vee N_0$, $x \in K$, $v \in S_x^{-2\delta}$.

(7.3) **Théorème.** *Les hypothèses sont celles de 7.1. Soit $\mathscr{M}^\tau = \{\Omega, \mathscr{B}(\Omega), P^\tau, X_n\}$ une famille de processus d'apprentissage (indexée par $\tau>0$) associée au champ μ_x. Soit $\bar{\Lambda}$ la fonctionnelle de Cramer régularisée associée au champ μ_x. Etant donnés $T>0$, un chemin continu, absolument continu $\varphi\colon [0, T] \to M$, et $\varepsilon>0$, il existe $\tau_0>0$ et $\alpha>0$ tels que $\tau<\tau_0$ garantisse l'inégalité (notations de 4.1)*

$$P^\tau\{\rho_{0,T}(\gamma^\tau, \varphi) \leqq \varepsilon | \mathscr{F}_0\} \geqq e^{-\frac{1}{\tau}(\bar{\Lambda}_{0,T}(\varphi)+\varepsilon)}$$

P^τ-p.s. *sur l'ensemble $\{\rho(X_0, \varphi(0)) \leqq \alpha\}$, où le chemin aléatoire γ^τ est comme en 6.3.*

Preuve. Supposons d'abord que M soit un ouvert de $\mathbb{R}^k$. Soit $\psi: [0, T] \to M$ un chemin continu, continûment différentiable par morceaux. On notera $\psi'(t)$ la dérivée à droite de ψ. La distance de $\psi'(t)$ à $(\mathbb{R}^k - S_{\psi(t)})$ est une fonction de t continue par morceaux, strictement positive si on suppose que $\psi'(t) \in \mathring{S}_{\psi(t)}$ pour tout $t \in [0, T]$. Cette hypothèse sur ψ étant faite, il existe $\delta > 0$ tel que

$$\psi'(t) \in S_{\psi(t)}^{-\delta} \quad \text{pour tout } t \in [0, T]. \tag{5}$$

Supposons que M contient $B(\psi(0), 1)$, ce qui n'est pas une restriction, évidemment. Fixons un compact K de M assez grand, contenant $\psi([0, T])$ et tel que P^τ-p.s. on ait $\gamma^\tau([0, T]) \subset K$, dès que $|X_0 - \psi(0)| \leqq 1$.

Donnons nous $\varepsilon > 0$. Puisque $\lambda_x(v)$ est uniformément continue sur $\bigcup_{x \in K} x \times S_x^{-\delta}$, il existe $\varepsilon_1 \leqq 1 \wedge \varepsilon/6$ tel que

$$|\lambda_x(v) - \lambda_y(v)| \leqq \varepsilon \quad \text{pour } x, y \in K, \ v \in S_x^{-\delta}, \ |x - y| \leqq 7\varepsilon_1. \tag{6}$$

Posons $\Omega_0 = \{X_0 - \psi(0) < \varepsilon_1\}$.

Donnons nous un entier $r \geqq 1$. Notons $V_n = V_{X_n}^n$ les «vecteurs vitesses» du processus $\mathscr{M}^\tau$ (cf. section 5). Posons pour j entier

$$\bar{V}_{jr} = \frac{1}{r}[V_{jr} + V_{jr+1} + \cdots + V_{jr+r-1}] \qquad j \geqq 0$$

$$Y_{jr} = Y_{(j-1)r} + r\,\tau\, V_{(j-1)r} \qquad j \geqq 1$$

où Y_0 est une v.a. $\mathscr{F}_0$-mesurable *quelconque* telle que $|X_0 - Y_0| \leqq \varepsilon_1$. Le champ μ_x étant borné, on vérifie aisément l'existence de $\alpha > 0$ tel que l'inégalité $r\tau \leqq \alpha$ entraîne, P^τ-p.s. sur Ω_0,

$$\begin{cases} |Y_{jr} - X_{jr}| \leqq 2\varepsilon_1 & \text{pour } (j+1)\, r\,\tau \leqq T \\ |\psi(t) - \psi(j\,r\,\tau)| \leqq \varepsilon_1 & \text{pour } j\,r\,\tau \leqq t \leqq (j+1)\, r\,\tau \leqq T. \end{cases} \tag{7}$$

Posons $b = \dfrac{1}{2T}\varepsilon_1$ et soit $\beta > 0$ tel que $r\tau \leqq \beta$ implique $|\psi'(t) - \psi'(j\,r\,\tau)| \leqq b$ pour $j\,r\,\tau \leqq t < (j+1)\, r\,\tau \leqq T$. Définissons des parties Ω_j, $j \geqq 0$ de Ω par

$$\Omega_j = \Omega_0 \cap \{\bar{V}_{nr} \in B[\psi'(n\,r\,\tau), b], n = 0, \ldots, (j-1)\}, \qquad j \geqq 1.$$

Il est évident que P^τ-p.s. sur Ω_j, et si $r\tau \leqq \beta$, on a

$$|\psi(n\,r\,\tau) - Y_{nr}| \leqq 2\varepsilon_1 + (n\,r\,\tau)\, 2b\,\varepsilon_1 \leqq 3\varepsilon_1 \qquad \text{pour } 0 \leqq n \leqq j \text{ et } (j+1)\, r\,\tau \leqq T.$$

En tenant compte de (7), on voit que P^τ-p.s., l'événement Ω_j implique

$$\rho_{0, jr\tau}(\psi, \gamma^\tau) \leqq 6\varepsilon_1, \qquad \text{pour } (j+1)\, r\,\tau \leqq T.$$

Prenons $0 < \varepsilon_2 \leqq \varepsilon$ et $\varepsilon_2 \leqq b/2$. L'ensemble des μ_x, $x \in K$ est (d'après l'hypothèse 4.1) (T)-compact (cf. 2.11). La loi conditionnelle de $V_n = V_{X_n}^n$ étant donné $\mathscr{F}_n$ est μ_{X_n}. Les théorème 2.3 et le corollaire 2.12 fournissent donc $\alpha_1 > 0$ tel que les inégalités $r\tau \leqq \alpha_1$, $r \geqq 1/\alpha_1$, et $(j+1)\, r\,\tau \leqq T$ entraînent P^τ-p.s. sur Ω_0

$$P^\tau(\bar{V}_{jr} \in B \,|\, \mathscr{F}_{jr}) \geqq e^{-r\varepsilon_2}\, \mu_{X_{jr}}^{(r)}(B^{-\varepsilon_2}) \tag{8}$$

pour tout $B \in \mathscr{B}(\mathbb{R}^k)$, avec les notations de 6.1.

Appliquons (11) à la boule $B(\psi'(j\,r\,\tau), b)$, et utilisons (5) et le lemme 7.2 pour obtenir l'existence de $N>0$ tel que $r \geqq N$ entraîne P^τ-p.s. sur Ω_0 les inégalités

$$P^\tau\{\bar{V}_{jr} \in B[\psi'(j\,r\,\tau), b] \,|\, \mathscr{F}_{jr}\} \geqq e^{-r\varepsilon_2}\, \mu^{(r)}_{X_{jr}}\{B[\psi'(j\,r\,\tau), b/2]\} \geqq e^{-r[2\varepsilon_2 + \lambda_{X_{jr}}(\psi'(jr\tau))]}. \tag{9}$$

Sur Ω_j, on a P^τ-p.s. $|X_{jr} - \psi(j\,r\,\tau)| \leqq 6\varepsilon_1$, et donc d'après (6), (9) on peut écrire

$$I_{\Omega_j} P^\tau\{\bar{V}_{jr} \in B[\psi'(j\,r\,\tau, b)] \,|\, \mathscr{F}_{jr}\} \geqq e^{-r[2\varepsilon_2 + \varepsilon + \lambda_{\psi(jr\tau)}(\psi'(jr\tau))]}. \tag{10}$$

Le premier membre de (10) est égal à $P^\tau\{\Omega_{j+1} | \mathscr{F}_{jr}\}$ donc si on pose $\pi_j = P^\tau(\Omega_j | \mathscr{F}_0)$, (10) donne

$$\pi_{j+1} \geqq e^{-r[3\varepsilon + \lambda_{\psi(jr\tau)}(\psi'(jr\tau))]'} \pi_j \qquad P^\tau\text{-p.s.}$$

et par récurrence on obtient pour $(j+1)\,r\,\tau \leqq T$, P^τ-p.s.

$$\pi_{j+1} \geqq e^{-r[3\varepsilon(j+1) + \Sigma_{j+1}]} I_{\Omega_0}$$

où $\Sigma_{(+1} = \lambda_{\psi(0)}[\psi'(0)] + \cdots + \lambda_{\psi(jr\tau)}[\psi'(j\,r\,\tau)]$.

Donnons nous $S<T$; prenons $r\tau < T-S$ et $(j+1)r\tau \in [S, T]$. Alors Ω_j est P^τ-p.s. inclus dans $\{\rho_{0,S}(\psi, \gamma^\tau) \leqq 6\varepsilon_1 \leqq \varepsilon\}$. Ceci entraîne donc

$$P^\tau\{\rho_{0,S}(\psi, \gamma^\tau) \leqq \varepsilon | \mathscr{F}_0\}\, I_{\Omega_0} \geqq \pi_{j+1} \geqq e^{-\frac{1}{\tau}[3T\varepsilon + r\tau\Sigma_{j+1}]}. \tag{11}$$

Puisque ψ est continûment différentiable par morceaux, et puisque (ψ, ψ') est dans le domaine où $\lambda.(.)$ est continue, il existe $\alpha_2>0$ tel que $r\tau \leqq \alpha_2$ implique

$$|r\,\tau\,\Sigma_{j+1} - \Lambda_{0,S}(\psi)| \leqq \varepsilon \qquad \text{pour un } j \text{ convenable.} \tag{12}$$

De (11), (12) on déduit que P^τ-p.s. sur Ω_0, on a

$$P^\tau\{\rho_{0,S}(\psi, \gamma^\tau) \leqq \varepsilon | \mathscr{F}_0\} \geqq e^{-\frac{1}{\tau}[\Lambda_{0,S}(\psi) + \varepsilon(3T+1)]}. \tag{13}$$

L'inégalité (13) est valable dès que $r > \dfrac{1}{\alpha_1} \vee N$ et $r\tau < \alpha \wedge \beta \wedge \alpha_1\ (T-S)$. On peut donc choisir et fixer $r \geqq \dfrac{1}{\alpha_1} \vee N$, puis prendre $\tau \leqq \tau_0$ assez petit.

Donnons nous maintenant $S>0$ quelconque et un chemin continu absolument continu $\varphi\colon [0,S] \to M$. Soit $\varepsilon>0$ donné. Choisissons $T>S$ et un chemin continu continûment différentiable par morceaux $\psi\colon [0,T] \to M$ tel que

$$\rho_{0,S}(\varphi, \psi) < \varepsilon/2, \qquad |\Lambda_{0,S}(\varphi) - \Lambda_{0,S}(\psi)| \leqq \varepsilon/2.$$

Appliquons (13) à ψ en y remplaçant ε par $\varepsilon' = \dfrac{1}{3T+1}\dfrac{\varepsilon}{2}$ pour obtenir τ_0 et η tels que $\tau \leqq \tau_0$ entraîne P^τ-p.s. sur Ω_0, l'inégalité

$$P^\tau\{\rho_{0,S}(\varphi, \gamma^\tau) \leqq \varepsilon | \mathscr{F}_0\} \geqq e^{-\frac{1}{\tau}[\Lambda_{0,S}(\varphi) + \varepsilon]}$$

où $\Omega_0 = \{|X_0 - \varphi(0)| \leqq \eta\}$.

L'extension au cas général, où M est une variété différentiable quelconque se fait très facilement à peu près comme dans le théorème 6.3.

(7.4) Considérons désormais un champ de mesures μ_x, $x \in M$ sur une variété riemannienne M vérifiant 7.1, mais aussi tel que les lignes de niveau de λ_x soient localement lipschitziennes (cf. section 3). Soit $\mathscr{M}^\tau$, $\tau > 0$ une famille de processus d'apprentissage associée aux μ_x. Nous supposerons pour simplifier les énoncés ci-dessous l'existence d'un point $m \in M$ tel que $P^\tau(X_0 = m) = 1$.

(7.5) *Définition.* Soit $C_{0,T}(M)$ l'ensemble des chemins continus γ: $[0, T] \to M$ tels que $\gamma(0) = m$. On munit $C_{0,T}(M)$ de la topologie de la convergence uniforme sur $[0, T]$. Pour toute partie G de $C_{0,T}(M)$ posons

$$\Lambda(G) = \inf_{\varphi \in G} \Lambda_{0,T}(\varphi).$$

Pour $\tau > 0$ assez petit (T étant donné, ainsi que m), il existe un unique chemin aléatoire continu φ^τ tel que $\varphi^\tau(n\tau) = X_{n\tau}$ pour $0 \leqq n\tau \leqq T + \tau$ et tel que sur $]n\tau, (n+1)\tau[$, φ^τ soit une géodésique de M.

(7.6) **Proposition.** *Les hypothèses et notations sont celles de 7.4 et 7.5. Pour toute partie borélienne G de $C_{0,T}(M)$, on a*

$$-\Lambda(\mathring{G}) \leqq \varliminf_{\tau \to 0} \frac{1}{\tau} \log P^\tau\{\varphi^\tau[0, T] \in G\} \leqq \varlimsup_{\tau \to 0} \frac{1}{\tau} \log P^\tau\{\varphi^\tau[0, T] \in G\} \leqq -\Lambda(\bar{G}).$$

Preuve. Prenons $\varphi \in \mathring{G}$ tel que $|\Lambda_{0,T}(\varphi) - \Lambda(\mathring{G})| \leqq \varepsilon$.

Soit V un «tube de chemins d'axe φ» tel que $V \subset \mathring{G}$. D'après le th. 7.3, et puisque $\bar{\Lambda}_{0,T}(\varphi) = \Lambda_{0,T}(\varphi)$ (cf. section 4), on a

$$-\Lambda(\mathring{G}) - \varepsilon \leqq -\Lambda_{0,T}(\varphi) \leqq \varliminf \frac{1}{\tau} \log P^\tau\{\varphi^\tau[0, T] \in V\}.$$

Puisque $\{\varphi^\tau[0, T] \in V\}$ implique $\{\varphi^\tau[0, T] \in G\}$, et puisque $\varepsilon > 0$ est quelconque, nous obtenons ainsi la première moitié de l'inégalité 7.6.

Soit K_A l'ensemble des $\varphi \in C_{0,T}(M)$ tel que

$$\Lambda_{0,T}(\varphi) \leqq A \quad \text{où} \quad A < \Lambda(\bar{G}).$$

D'après le lemme 4.3, K_A est compact. Par construction K_A et $\bar{G}$ sont disjoints. Il existe donc $\varepsilon > 0$ tel que $\rho_{0,T}(\varphi, K_A) \geqq \varepsilon$ pour tout $\varphi \in \bar{G}$. D'après le théorème 6.3, on a

$$\varlimsup_{\tau \to 0} \frac{1}{\tau} \log P^\tau\{\rho_{0,T}(\varphi^\tau, K_A) \geqq \varepsilon\} \leqq -A.$$

Ceci implique, puisque $\{\varphi^\tau \in G\}$ est inclus dans $\{\rho_{0,T}(\varphi^\tau, K^A) \geqq \varepsilon\}$,

$$\varlimsup_{\tau \to 0} \frac{1}{\tau} \log P^\tau\{\varphi^\tau \in G\} \leqq -A.$$

Quand $A \nearrow \Lambda(\bar{G})$, on obtient la seconde moitié de l'inégalité 7.6.

8. Le potentiel associé à l'étude asymptotique des processus d'apprentissage lent

(8.0) Nous montrons ici comment on peut atteindre l'énoncé donné en fin de l'introduction, ou plus simplement étudier la sortie d'un domaine au bord duquel le champ moyen est rentrant, et contenant un seul attracteur de ce champ. Il nous faut introduire les potentiels

$$V(x,y)=\inf_{\substack{\varphi_0=x\\ \varphi_T=y}} \Lambda_{0,T}(\varphi)$$

$$\bar{V}(x,y)=\inf_{\substack{\varphi_0=x\\ \varphi_T=y\\ \varphi \text{ à vitesses intérieures}}} \Lambda_{0,T}(\varphi)$$

et en étudier les propriétés de continuité, étude qui est spécifique à notre problème. Ensuite, l'obtention du résultat final ne présente aucune originalité par rapport à [14], et sera très allusive.

Précisons que, *dans les infimums ci-dessus, le laps de temps T n'est pas fixé:* par exemple, pour échapper à un attracteur standard en minimisant Λ, on est obligé de recourir à des T tendant vers l'infini: signalons enfin que l'hypothèse de variation lipschitzienne (et non seulement continue) du champ de mesures μ_m n'intervient qu'à partir du théorème 6. Le lemme (trivial, mais fondamental) 5.1 de [14], qui affirme le caractère localement lipschitzien de V, est évidemment faux dans notre situation: le mouvement d'un point à un autre très voisin de M peut parfaitement être impossible.

Introduisons les relations d'équivalences $\mathscr{R}$ (resp. $\bar{\mathscr{R}}$) définies pas $x\mathscr{R}y$ si et seulement si $V(x,y)=V(y,x)=0$ (resp. ...): nous allons montrer que ces relations coïncident, et que les classes d'équivalence qu'elles définissent sont fermées.

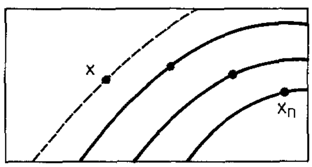
(8.1) **Lemme.** *Si une $\bar{\mathscr{R}}$-classe d'équivalence contient deux points distincts x et y, elle contient sûrement une portion entourant x de trajectoire du champ moyen.*

Preuve. Puisque $\bar{V}(x,y)=0$, il y a une suite φ_n de chemins «à vitesses intérieures» vérifiant $\varphi_n(0)=x$, $\varphi_n(T)=y$, $\Lambda_{0,T_n}(\varphi_n)\to 0$. Les T_n sont nécessairement minorés par un T (même si les μ_m ne sont pas à support compact) et a fortiori $\Lambda_{0,T}(\varphi_n)\to 0$. D'après le lemme 4.3, une suite extraite de $\varphi_n|_{[0,T]}$ converge uniformément vers φ telle que $\varphi(0)=x$ et $\Lambda_{0,T}(\varphi)=0$, ce qui impose que φ soit une trajectoire du champ moyen (en aval de x). Tous les $\varphi(t)$, $0\leqq t\leqq T/2$, sont équivalents à x: on peut aller gratis de x à $\varphi(t)$, puis à $\varphi(T/2)$, et aussi de y à x; il ne coûte presque rien d'aller de $\varphi_n(T)$ à y, et les $\varphi_n(T)$ tendent vers $\varphi(T)$; φ étant une trajectoire du champ moyen, l'hypothèse de non-dégénérescence «Supp μ_m engendre T_mM» entraîne qu'il ne coûte presque rien (au sens de $\bar{V}$) d'aller de $\varphi(T/2)$ à $\varphi_n(T)$. Enfin, pour obtenir une trajectoire du champ moyen en amont de x, et formée de points équivalents à x, on utilise $\bar{V}(y,x)=0$ (il existe ψ_n avec $\psi_n(0)=y$, $\psi_n(T_n)=x$, $\Lambda_{0,T_n}(\psi_n)\to 0$, $T_n\geqq T$) via la limite des $\psi_n|_{[T_n-T,T_n]}$.

(8.2) **Proposition.** *Toute $\bar{\mathscr{R}}$-classe d'équivalence est fermée.*

Preuve. Il est clair que l'ensemble des points de la trajectoire (du champ moyen) passant par x, et équivalents à x, est un intervalle. Le lemme 8.1 montre que, si la classe de x n'est pas réduite à x, cet intervalle I est nécessairement ouvert. Montrons qu'en fait toute la trajectoire de x est alors équivalente à x, ainsi que ses ensembles α et ω-limites: soient $x_n \in I$, x_{n+1} postérieur à x_n sur la trajectoire, et $x_n \xrightarrow[n\to\infty]{} y$. Grâce à une limite de chemins de coûts $\varepsilon_n \to 0$ joignant x_n à x, on obtient un bout de trajectoire issu de y en aval de y, et composé de points équivalents à x, ce qui prouve que y est équivalent à x, donc, dans un premier temps, que l'intervalle I contient toute la trajectoire en aval de x, et, dans un deuxième temps, que l'ensemble ω-limite est équivalent à x (si z est dans l'ensemble ω-limite, on voit facilement que $V(x, z)=0$). On procède de même pour la trajectoire en amont de x et l'ensemble α-limite.

Si maintenant des x_n tous équivalents tendent vers x, et si on veut montrer que x est équivalent aux x_n, le seul cas délicat est celui où, dans un voisinage de x, les trajectoires des x_n sont toutes distinctes, et où le champ moyen en x est non-nul. Dans ce cas, l'hypothèse de non-dégénérescence jointe au fait que les trajectoires entières de tous les x_n sont équivalents, permet clairement de conclure.



(8.3) **Proposition.** *(Lemme 5.2 de [14], adapté et précisé). Soit K une $\bar{\mathscr{R}}$-classe d'équivalence compacte. Pour tout $h>0$, $\delta>0$, il existe $T_0>0$, U ouvert contenant K et contenu dans K^δ tels que, pour tout $x, y \in U$, on puisse aller de x à y en un temps inférieur à T_0, pour un prix (au sens de $\bar{V}$) inférieur à h, et en restant dans K^δ.*

Preuve. Soit U_x l'ensemble des points de $\mathring{K}^\delta$ que l'on peut atteindre à partir de $x \in K$ en un temps $\leqq 1$ et pour moins cher que $h/3$; U_x n'est pas un voisinage de x, mais, grâce à l'hypothèse de non-dégénérescence, c'est un voisinage de la trajectoire moyenne en aval de x sur le temps 1; le lemme 2 montre qu'alors $\bigcup_{x\in K} U_x$ est un voisinage de K.

Soit V_y l'ensemble des points de $\mathring{K}^\delta$ d'où l'on peut rejoindre y en un temps $\leqq 1$ pour moins cher que $h/3$. Comme ci-dessus, on voit que $\bigcup_{x\in K} V_y$ est un voisinage de K.

Soient alors des x_i en nombre fini tels que les U_{x_i} recouvrent K. La proposition est évidemment vérifiée si on choisit pour U un ouvert contenant K et contenu à la fois dans $\bigcup U_{x_i}$ et $\bigcup V_{y_j}$, et $T_0 = 2 + \max_{i,j} \tau_{ij}$, où τ_{ij} est un laps de temps suffisant pour joindre x_i à y_j, pour moins cher que $h/3$, en restant dans $\mathring{K}^\delta$ (on peut rester dans $\mathring{K}^\delta$ parce que K est une *classe* d'équivalence).

(8.4) **Proposition.** *Si M est compacte, et si on peut trouver une réunion finie d'ensembles ω-limites du champ moyen dont tout voisinage contienne tous les autres ensembles ω-limites sauf un nombre fini, alors $V(x, y)=0$ entraîne $\bar{V}(x, y)=0$.*

Preuve. On peut reprendre la preuve de la proposition 3 en remplaçant «K classe d'équivalence compacte» par «K ensemble ω-limite». Si $V(x, y)=0$, pour $\varepsilon>0$

donné, on choisit un nombre fini r d'ouverts U_i contenant tous les ω-limites et tels que, pour toute collection a_i, $b_i \in U_i$, on puisse aller de a_1 à $b_1, \ldots$, et de a_r à b_r pour au total moins que ε. Par plusieurs extractions de sous-suites à partir d'une suite φ_n minimisant $V(x, y)$, on fabrique une suite (sans cycles) de U_i, soit $U_{i_1}, \ldots, U_{i_q}$ telle qu'il existe une trajectoire du champ moyen joignant U_{i_j} à $U_{i_{j+1}}$, et éventuellement x à U_{i_1}, et U_{i_q} à y. On voit maintenant comment, grâce à la proposition 8.3 modifiée, aller de x à y, avec des vitesses intérieures, pour moins que ε.

On peut maintenant recopier les lemmes 5.3 et 5.4 de [14]. Le premier donne une majoration du temps de sortie moyen d'un petit voisinage d'une classe d'équivalence compacte. Le second étudie les voyages dans un compact de M ne rencontrant aucun ensemble ω-limite: si l'on voyage dans K^δ pendant un temps T, cela coûte au moins $a(T-T_0)$, pour a et T_0 convenablement choisis.

On en déduit que la probabilité $P_x^\tau(\tau_K > T)$ que l'instant de première sortie de K soit postérieur à T est inférieure à $\exp\left(-\frac{c(T-T_0)}{\tau}\right)$, pour un c convenablement choisi.

(8.5) **Proposition.** *S'il y a dans M un nombre fini de classes d'équivalence, contenant les ensembles ω-limites, toutes compactes, la fonction $V(x, y)$ est semi-continue inférieurement vis à vis de y.*

Preuve. Soient x fixé, et $y_n \to y$, avec $V(x, y_n) \to I$. Choicissons des φ^n, avec $\varphi^n(0)=x$, $\varphi^n(T_n)=y_n$, $\Lambda_{0,T_n}(\varphi^n) \to I$. Pour $h>0$ arbitraire, on veut trouver φ tel que $\varphi(0)=x$, $\varphi(T)=y$, $\Lambda_{0,T}(\varphi) \leqq I+h$. Soient K_i, $i=1, \ldots, k$ les classes contenant les ensembles ω-limites, δ tel que les K_i^δ soient disjoints, et π le coût minimum pour aller d'un K_i^δ à un autre. Appliquons la proposition 8.3 à chacun des K_i, pour le δ ci-dessus et le coût $h\pi/I$; soit Ω_i l'intersection des U et V obtenus. Une trajectoire φ_n ne peut pas osciller plus de I/π fois entre les Ω_i; lorsqu'elle traverse un Ω_i (y compris lorsqu'elle en sort pour y revenir), on la modifie de sorte à en augmenter peu le prix (de moins que $h\pi/I$ grâce à 8.3), mais à en limiter la durée. La durée de la portion de trajectoire non modifiée, située dans $M - \bigcup \Omega_i$ et de coût moindre que I, est majorée grâce au lemme 5.4 de [14]. On dispose donc désormais de ψ^n, avec $\psi^n(0)=x$, $\psi^n(T_n)=y_n$, $\Lambda_{0,T_n}(\psi^n) \leqq I+h$, et T_n majoré. On extrait une sous-suite ψ^k telle que T_n converge vers T, et on n'a aucun mal, grâce au lemme 4.3, à obtenir le φ désiré, sur $[0, T/2]$ comme limite d'une suite extraite de $\psi^k_{[0,T/2]}$, sur $[T/2, T]$ comme limite d'une suite réextraite de $\psi^k_{[T_n - T/2, T_n]}$.

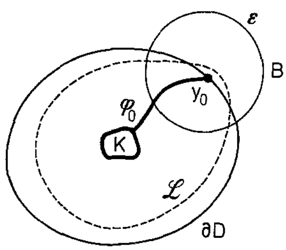
La proposition 8.5 permet de recopier le lemme 5.5 de [14]: si K est une classe d'équivalence stable (pour tous $x \in K$, $y \notin K$, $V(x, y) \neq 0$), si D est un voisinage de K et si $K-D$ ne rencontre pas d'ensemble ω-limite du champ moyen, alors, pour tout x dans un sous-voisinage de K compact dans D, la moyenne $E_x^\tau \theta_D$ de l'instant θ_D de première sortie de D est supérieure à $\exp(a/\tau)$, pour un a bien choisi.

Nous allons enfin faire entrer en scène *toutes les hypothèses sur le champ de mesures* μ_m pour obtenir le

(8.6) **Théorème.** *Posons $V(K, y) = \inf_{x \in K} V(x, y)$, et définissons de même $\bar{V}(K, y)$. Si K est une classe d'équivalence stable, $V(K, y)$ coïncide avec $\bar{V}(K, y)$, et c'est une fonction continue de y.*

La semi-continuité inférieure de $V(K, y)$ résulte de la proposition 8.5, et la semi-continuité supérieure de $\bar{V}(K, y)$ est claire (pour atteindre un voisinage de y à peu près au même prix que y, on modifie un peu la vitesse au bout d'un chemin économique à vitesses intérieures). L'égalité de $V(K, y)$ et $\bar{V}(K, y)$ résulte de la proposition 4.6: si φ, $\varphi(0)=x\in K$, $\varphi(T)=y$, vérifie $\Lambda_{0,T}(\varphi)\leqq V(K, y)+\varepsilon$, la preuve de 4.6 montre qu'il existe ψ, $d(\varphi(0), \psi(0))\leqq\delta$, $\psi(T)=y$, à vitesses intérieures, et telle que $\Lambda_{0,T}(\psi)\leqq\Lambda_{0,T}(\varphi)+\varepsilon$. La proposition 8.3 entraînant, si δ est assez petit, $\bar{V}(K, \psi(0))\leqq\varepsilon$, il en résulte que $\bar{V}(K, y)$ est inférieur à $V(K, y)+3\varepsilon$.

(8.7) Tous les outils sont maintenant disponibles pour étudier par exemple la sortie d'un domaine D contenant un seul attracteur K. Supposons que $V(K, y)$ atteigne son minimum V_0 sur ∂D en un unique y_0. Sur le dessin, est représentée la boule B de centre y_0 et de rayon ε, ainsi qu'une ligne de niveau $\mathscr{L}=\{y\colon V(K, y)= V(K, y)+\omega\}$ (en tirets), ω étant choisi pour que la portion de $\mathscr{L}$ extérieure à B soit à une distance $\delta>0$ de ∂D. Enfin, φ_0 est un chemin de durée T joignant K à y_0 avec un coût excédant V_0 de moins que $\omega/4$. On fait trois paquets des trajectoires du processus issues de x (qu'on peut sans perte de généralité supposer égal à $\varphi_0(0)$): les trajectoires qui ne quittent pour une première fois D qu'après un temps $\theta\geqq T$ et



assez grand, de façon à majorer la probabilité de leur ensemble par $\exp\left(-\frac{V_0+\omega}{\tau}\right)$, en utilisant les remarques suivant la proposition 8.4, appliquées à un petit voisinage de K et à son complément dans D; les trajectoires qui, avant θ, ont quitté D à une distance de y_0 supérieure à ε: ces trajectoires sont à une distance $\geqq\delta$ des trajectoires de coût $\leqq V_0+\omega$, donc leur ensemble a une probabilité $\leqq\exp\left(-\frac{V_0+3\omega/4}{\tau}\right)$, pour τ assez petit; enfin les autres, incluant en particulier les trajectoires distantes de φ_0 de moins que ε, et donc dans leur ensemble de probabilité asymptotiquement $\geqq\exp\left(-\frac{V_0+\omega/2}{\tau}\right)$.

Il est alors clair que, lorsque $\tau\to 0$, la probabilité que la première sortie de D ait lieu dans la boule $B(y_0, \varepsilon)$ tend vers 1.

Terminons en précisant ce qui se passe lorsqu'on agrandit le domaine D: nous supposons qu'à la frontière de D, le champ moyen est strictement rentrant (∂D est de classe C^1, $\bar{D}$ compact contenant pour simplifier un seul attracteur, ponctuel, que nous entourons d'une sphère σ). Soit d un domaine voisin de $\bar{D}$ et le contenant, $V_d=V(\sigma, \partial d)$, E_d l'ensemble des points de ∂d minimisant $V(\sigma, .)$, et F_d l'ensemble des points de ∂D où les trajectoires minimisant $V(\sigma, \partial d)$ sortent pour la première fois de D (les trajectoires minimisantes existent d'après 4.3 et l'absence d'ensembles ω-limites; si φ est une telle trajectoire, le point de F_d associé est $\varphi(\tau_\varphi)$, pour $\tau_\varphi=$

$\inf\{t|\varphi(t)\notin D\}$). Soit V la borne inférieure des V_d, qui peut bien être strictement supérieure à $V(\sigma, \partial D)$.

(8.8) **Proposition.** *La limite des F_d lorsque d décroît vers D est contenue dans l'adhérence de la limite des E_d (il s'agit de limites «extérieures» au sens de la distance de Hausdorff).*

Preuve. Il suffit de majorer la distance de $\varphi(\tau_\varphi)$ à $\varphi(T_\varphi)$; or, l'intégrale $\Lambda_{0,\tau_\varphi}(\varphi)$ est au moins égale à V, et, si d est assez petit, l'intégrale $\Lambda_{0,T\varphi}(\varphi)$ est au plus égale à $V+\varepsilon$, Si ε est assez petit, on peut trouver un voisinage tubulaire $\partial D\times[-1,+1]\xrightarrow{\theta}M$ tel que θ soit l'identité sur $\partial D\times 0$, $\theta(\partial D\times -1)$ soit une surface Δ_1 entre σ et ∂D telle que $V(\Delta_1, \partial D)>\varepsilon$, $\theta(\partial D\times +1)$ soit une surface Δ_2 extérieure à ∂D telle que $V(\partial D, \Delta_2)>\varepsilon$, et que θ^* appliqué au champ moyen (supposé de classe C^1) soit un champ vertical constant. Nous voyons que $\Lambda_{\tau_\varphi T_\varphi}(\varphi)$ est inférieur à ε, donc que $\varphi|_{[\tau_\varphi, T_\varphi]}$ reste dans l'image de θ. Nous pouvons alors minorer $\Lambda_{\tau_\varphi T_\varphi}(\varphi)$ par une intégrale d'énergie portant sur la première projection de $\theta^{-1}\varphi$, il suffit de penser qu'il y a une forme quadratique minorant tous les contours apparents en projection sur $T_m\partial D\times\mathbb{R}$ des graphes de $\lambda_{m\times s}$, $s\in[-1,+1]$ ($\lambda_{m\times s}$ est relatif à la mesure $\theta^*_{m\times s}\mu_{\theta(m\times s)}$). Les intégrales $\Lambda_{\tau_\varphi T_\varphi}(\varphi)$ tendant vers 0 lorsque d tend vers D, il en résulte que la distance entre $\varphi(\tau_\varphi)$ et $\pi_1\theta^{-1}\varphi(T_\varphi)$ tend vers 0, donc aussi la distance entre $\varphi(\tau_\varphi)$ et $\varphi(T_\varphi)$, et ceci uniformément sur les φ minimisants pour une distance de d à D donnée. C.Q.F.D.

Bibliographie

1. Changeux, Courrege, P., Danchin: Theory of epigenesis of neuronal networks by selective stabilization of synapses. Proc. Nat. Acad. Sci. **70**, 2974–2978 (1973)
2. Chernoff, H.: Asymptotic efficiency for tests based on the sum of the observations. Ann. Math. Statist. **23**, 493–507 (1952)
3. Cramer, H.: Sur un nouveau théorème limite de la théorie des probabilités. Colloquium on the theory of Probability. Paris: Hermann 1937
4. Donsker, M. D., Varadhan, S. R. S.: Some problems of large deviation. Congrès de Rome, 1975
5. Donsker, M.D., Varadhan, S.R.S.: Large deviations of Markov processes and asymptotic evaluations of certain Markov processes expectations for large time. In Probabilistic methods for differential equations. Lecture notes in mathematics, **451**. Berlin-Heidelberg-New York: Springer 1975
6. Ljung, L.: Convergence of recursive stochastic algorithms. IFAC Symp. Stochastic Control, Budapest, 1974, and report n° 7403 dir. of automatic control, Lund Inst. of technology
7. Malsburg, C.: Kybernetic, n° 14, p. 85 (1973)
8. Norman, M.F.: Markovian learning processes. SIAM Rev. **16**, n° 2, 143–162 (1974)
9. Papanicolaou, G.C., Kohlerw: Asymptotic theory of mixing stochastic ordinary differential equations. Comm. Pure Appl. Math. **27**, n° 5, 641–668 (1974)
10. Sethuraman, J.: On the probability of large deviations of families of Sample means. Ann. Math. Statist. **35**, 1304–1316 (1964)
11. Södersröm, T.: Convergence properties of the generalized least squares identification method. Automat. **10**, 617–626 (1974)
12. Strassen, V.: Probability measures with given marginals, Ann. Math. Statist. **36**, 423–439 (1965)
13. Varadhan, S.R.S.: Asymptotic probabilities and differential equations. Comm. Pure Appl. Math. **19**, 261–286 (1966)
14. Ventsel, A.D., Freidlin, M.I.: On small random perturbations of dynamical systems. Russian Math. Survey **25**, n° 1, 1–55 (1970)

Reçu le 8 avril 1976