Astérisque

MARIE DUFLO

Formule de Chernoff pour des chaînes de Markov (d'après Donsker et Varadhan)

Astérisque, tome 68 (1979), p. 99-124

<http://www.numdam.org/item?id=AST_1979__68__99_0>

Formule de Chernoff pour des chaînes de Markov

(d'après Donsker et Varadhan)

Marie DUFLO.

A - PRÉLIMINAIRES

Notations

$(E , \mathcal{E})$ espace mesurable .

$\mathcal{P}$ ensemble des probabilités sur $(E , \mathcal{E})$.

$\mathcal{B}$ ensemble des v.a. bornées sur $(E , \mathcal{E})$.

Pour $u \in \mathcal{B}$, on note $||u|| = \sup_{x \in E} |u(x)|$.

Pour $\alpha \in \mathcal{P}$, on note $||\alpha|| = \sup_{\substack{u \in \mathcal{B} \\ ||u|| \leqslant 1}} \left|\int u \, d\alpha\right|$.

$\hat{\mathcal{U}}$ est l'ensemble des v.a. positives u sur E , telles que u et $\frac{1}{u}$ soient bornées.

Dans presque tout l'exposé E sera supposé <u>polonais</u> ; (métrique, complet, séparable) ; alors $\mathcal{E}$ sera sa tribu borélienne , $\mathcal{P}$ sera muni de la topologie de la convergence étroite et $\underline{\underline{\mathcal{P}}}$ sera sa tribu borélienne pour cette topologie. La distance de E sera notée d . Enfin on notera $\mathcal{C}$ l'ensemble des fonctions continues bornées de E dans $\mathbb{R}$ et $\mathcal{U}$ l'ensemble $\hat{\mathcal{U}} \cap \mathcal{C}$.

I - INFORMATION DE KULLBACK

1 - Définition

$(E, \mathcal{E})$ étant un espace mesurable quelconque, et α et β deux probabilités sur $(E, \mathcal{E})$, l'information de Kullback de α par rapport à β est :

$$I_\beta(\alpha) = \sup_{u \in \mathcal{U}} \left[\int \text{Log}\, u \, d\alpha - \text{Log} \int u \, d\beta\right]$$

On retrouve la définition classique de l'information par le :

Théorème

$I_\beta(\alpha)$ est finie si et seulement si α est absolument continue par rapport à β $(\alpha < < \beta)$ et $\text{Log}\,\frac{d\alpha}{d\beta}$ est α-intégrable .

Alors $I_\beta(\alpha) = \int (\text{Log}\,\frac{d\alpha}{d\beta})\, d\alpha$.

Démonstration : voir Chapitre III).

2 - Proposition

Si E est polonais ,

$$I_\beta(\alpha) = \sup_{u\in\mathcal{U}} \left[\int \mathrm{Log}\, u \, d\alpha - \mathrm{Log} \int u \, d\beta\right] .$$

Démonstration

Soit $I \in \mathbb{R}$ et

$$\mathcal{A} = \{u \ , \ u \in \hat{\mathcal{U}}, \ \int \mathrm{Log}\, u \, d\alpha - \mathrm{Log} \int u \, d\beta \leq I\} .$$

On suppose que $\mathcal{A}$ contient $\mathcal{U}$; montrons qu'elle contient $\hat{\mathcal{U}}$. Si une suite (f_n) contenue dans $\mathcal{A}$ croît, ou décroît vers une fonction f de $\hat{\mathcal{U}}$, alors f est dans $\mathcal{A}$.

Si $(G_n)_{n=1\dots N}$ est un recouvrement ouvert de E, à tout $p \geq 1$, on peut associer $f_n^p = p\left[d(x,G_n^c) \wedge \frac{1}{p}\right]$, $n = 1 \dots N$. Soit $(a_n)_{n=1\dots N}$ une suite de nombres strictement positifs. La suite $\sum_{n=1}^{N} a_n f_n^p$ croît si p tend vers ∞ vers $\sum_{n=1}^{N} a_n 1_{G_n}$:

Donc $\sum_{n=1}^{N} a_n 1_{G_n}$ est dans $\mathcal{A}$.

Soit $(A_n)_{n=1\dots N}$ un recouvrement mesurable de E .

Pour tout entier p on peut trouver un ouvert G_n^p contenant A_n tel que $\beta(G_n^p) - \beta(A_n) \leq \frac{1}{p}$ $(n = 1 \dots N)$.

$$\int \mathrm{Log}\left(\sum_{n=1}^{N} a_n 1_{A_n}\right) d\beta - \mathrm{Log} \sum_{n=1}^{N} a_n \beta(A_n)$$

$$\leq \int \mathrm{Log}\left(\sum_{n=1}^{N} a_n 1_{G_n^p}\right) d\beta - \mathrm{Log} \sum_{n=1}^{N} a_n \beta(G_n^p) + \mathrm{Log} \frac{\sum_{n=1}^{N} a_n \beta(G_n^p)}{\sum_{n=1}^{N} a_n \beta(A_n)}$$

$$\leq I + \mathrm{Log}\, \{1 + \frac{1}{p \sum_{n=1}^{N} a_n \beta(A_n)}\}$$

Ceci est vrai pour tout $p \in \mathbb{N}$, donc $\sum_{n=1}^{N} a_n 1_{A_n}$ est dans $\mathcal{A}$. Toute fonction de $\hat{\mathcal{U}}$ est limite croissante de telles fonctions donc $\hat{\mathcal{U}} = \mathcal{A}$ ■

<u>Corollaire</u>

Les fonctions $\alpha \longmapsto I_\beta(\alpha)$ et $\beta \longmapsto I_\beta(\alpha)$ sont semi continues inférieurement (s.c.i.) de $\mathcal{P}$ dans $[0,\infty]$.

En effet pour tout $u \in \mathcal{U}$, les fonctions

$$\alpha \longmapsto \int \mathrm{Log}\, u\, d\alpha - \mathrm{Log} \int u\, d\beta \quad \text{et} \quad \beta \longrightarrow \int \mathrm{Log}\, u\, d\alpha - \mathrm{Log} \int u\, d\beta$$

sont continues ■

3 - Il existe une fonction Φ à valeurs dans $[0,2]$ telle que $\lim_{\ell \to 0} \Phi(\ell) = 0$ et telle que :

$$||\alpha - \beta|| \leqslant \phi[I_\beta(\alpha)]$$

(D'où le résultat classique : $I_\beta(\alpha) = 0$, si et seulement si $\alpha = \beta$).

<u>Démonstration</u>

Soit $u = (b+1)1_A + 1_{A^c}$, pour $A \in \mathcal{E}$

$$\alpha(A)\,\mathrm{Log}(b+1) - \mathrm{Log}(1 + b\,\beta(A)) \leqslant I_\beta(\alpha) = I$$

$$\alpha(A)\,\mathrm{Log}(b+1) - b\,\beta(A) \leqslant I$$

$$b[\alpha(A) - \beta(A)] \leqslant I + \alpha(A)[b - \mathrm{Log}(b+1)]$$

$$\alpha(A) - \beta(A) \leqslant \inf_{b>0} \frac{I + (b - \mathrm{Log}(b+1))}{b} = 2\hat{\phi}(I)$$

$$||\alpha - \beta|| = 2 \sup_A (\alpha(A) - \beta(A)) \leqslant \phi(I) = 2\{\hat{\phi}(I) \wedge 2\}$$

$$\phi(0) = 0 .$$

Soit $\Psi(b) = \frac{I}{b} + 1 - \frac{\mathrm{Log}(b+1)}{b}$

$$\Psi'(b) = -\frac{I}{b^2} + \frac{\mathrm{Log}(b+1)}{b^2} - \frac{1}{b(b+1)} \quad .$$

Si $I > 0$, le minimum est atteint pour la valeur $\hat{b} > 0$ telle que :

$$\mathrm{Log}(\hat{b} + 1) = I + \frac{\hat{b}}{\hat{b}+1}$$

et
$$\Psi(\hat{b}) = 1 - \frac{1}{\hat{b}+1} = \frac{\hat{b}}{\hat{b}+1} \quad .$$

Il est facile de voir que si I tend vers 0 , $\hat{b}$ tend vers 0 donc $\phi(I)$ tend vers 0 ■

B - MAJORATION DE CHERNOFF POUR DES CHAÎNES DE MARKOV CONTROLÉES

On reproduit ici des résultats de N. MAIGRET [4].

On considère une chaîne de Markov contrôlée. Les espaces des états et des actions sont des espaces polonais munis de leurs tribus boréliennes $(E , \mathcal{E})$ et $(A , \mathcal{A})$ respectivement. On donne une probabilité de transition π de $(E \times A , \mathcal{E} \otimes \mathcal{A})$ dans $(E , \mathcal{E})$. L'espace des trajectoires est $(\Omega,\mathfrak{A}) = (E \times A , \mathcal{E} \otimes \mathcal{A})^{\mathbb{N}}$: si $\omega = (x_n, a_n)_{n \in \mathbb{N}}$, on note $X_n(\omega) = x_n$ et $A_n(\omega) = a_n$. Soit $\mathcal{B}_n = \sigma(X_o, \ldots, X_n, A_o, \ldots, A_{n-1})$. Une stratégie δ est une suite $(\delta_n)_{n \in \mathbb{N}}$, où δ_n est une probabilité de transition de $(E , \mathcal{E})^n$ dans $(A , \mathcal{A})$: $\delta_n(x_o, \ldots, x_{n-1}; \cdot)$ est la loi de probabilité de la $n^{\text{ième}}$ action choisie, si les observations antérieures sont $(x_o, \ldots, x_{n-1})$. On peut alors pour tout état initial x et toute stratégie δ construire une probabilité P_x^δ sur $(\Omega , \mathfrak{A})$ telle que :

$$P_x^\delta(X_o = x) = 1$$

et pour tout $B \in \mathcal{E}$ et $C \in \mathcal{A}$

$$P_x^\delta[X_{n+1} \in B, A_n \in C \,|\, \mathcal{B}_n] = \int \pi(X_n,a,dy)\, 1_B(y)\, 1_C(a)\, \delta_n(X_o,\ldots,X_{n-1};da) \quad .$$

On notera $\mathscr{C}$, $\mathcal{U}$, $\mathcal{P}$ les ensembles des v.a. continues bornées, des v.a. continues bornées ainsi que leurs inverses, des probabilités sur $(E \times A \times E, \mathcal{E} \otimes \mathcal{A} \otimes \mathcal{E})$; $\mathcal{P}$ est muni de la topologie de la convergence étroite. Les mêmes ensembles avec les indices 1 ou 2 sont définis sur E et $E \times A$ respectivement. On note enfin

$$\pi f(x,a) = \int \pi(x,a,dy)\, f(y) \qquad \text{pour} \quad f \in \mathscr{C}_1$$

$$= \int \pi(x,a,dy)\, f(x,a,y) \qquad \text{pour} \quad f \in \mathscr{C} \quad .$$

On suppose π fellérienne, c'est-à-dire que $\pi\,\mathscr{C}$ est continue dans $\mathscr{C}$.

Soit $\alpha \in \mathcal{P}_1$ et s une transition de $(E, \mathcal{E})$ dans $(A, \mathcal{A})$.

Soient $\alpha \otimes s$ et $\alpha \otimes s \otimes \pi$ les probabilités définies pour $B \in \mathcal{E} \otimes \mathcal{A}$ et $C \in \mathcal{E} \otimes \mathcal{A} \otimes \mathcal{E}$ par :

$$\alpha \otimes s(B) = \int \alpha(dx)\ s(x,da)\ 1_B(x,a)$$

$$\alpha \otimes s \otimes \pi(C) = \int \alpha(dx)\ s(x,da)\ \pi(x,a,dy)\ 1_C(x,a,y)\ .$$

Si (X_n, A_n) a la loi $\alpha \otimes s$, alors (X_n, A_n, X_{n+1}) a la loi $\alpha \otimes s \otimes \pi$.

Pour décrire la chaîne controlée, il s'agit de décrire π ; c'est pourquoi on considère les fonctions de répartition empiriques L^n définies en posant pour $\omega \in \Omega$ et $B \in \mathcal{E} \otimes \mathcal{A} \otimes \mathcal{E}$

$$L^n(\omega,B) = \frac{1}{n} \sum_{k=o}^{n-1} 1_B(X_k, A_k, X_{k+1})(\omega)\ .$$

La fonction L^n est mesurable de $(\Omega, \mathcal{a})$ dans $\mathcal{P}$ muni de sa tribu borélienne.

Pour toute probabilité $\lambda \in \mathcal{P}$, on note λ_i sa i^e marginale $(i = 1, 2, 3)$. Il existe une transition s de $(E, \mathcal{E})$ dans $(A, \mathcal{A})$ telle que $\lambda_1 \otimes s$ soit la marginale de λ sur $(E \times A, \mathcal{E} \otimes \mathcal{A})$. Intuitivement, si $\mathcal{L}^n$ est proche de λ, elle a tendance à être proche de $\lambda_1 \otimes s \otimes \pi$, qui est l''Écoulement naturel de marginale $\lambda_1 \otimes s$". L'information apportée par λ sur π est d'autant plus grande que λ diffère de $\lambda_1 \otimes s \otimes \pi$, donc que $I_{\lambda_1 \otimes s \otimes \pi}(\lambda)$, noté $I(\lambda)$, est grand.

<u>Etude de $I(\lambda) = I_{\lambda_1 \otimes s \otimes \pi}(\lambda)$</u>

a) $$I(\lambda) = \sup_{u \in \mathcal{U}} \left[\int \text{Log}\ u\ d\lambda - \text{Log} \int u\ d\lambda_1 \otimes s \otimes \pi\right]$$

$$= \sup_{u \in \mathcal{U}} \left[\int \text{Log}\ \frac{u(x,a,y)}{\pi\ u(x,a)}\ d\lambda\ (x,a,y)\right]$$

Démonstration

La première expression est la définition de $I(\lambda)$.

La seconde notée J majore $I(\lambda)$ d'après l'inégalité de Jensen . Soit :

$$\mathcal{V} = \{v \; ; \; v(x,a,y) = \operatorname{Log} \frac{u(x,a,y)}{\pi\, u(x,a)} \quad \text{pour} \quad u \in \mathcal{U}\} .$$

Pour $v \in \mathcal{V}$ la première expression vaut :

$$\int \operatorname{Log} v \, d\lambda = \int \operatorname{Log} \frac{u(x,a,y)}{\pi\, u(x,a)} \, d\lambda(x,a,y) \quad .$$

Comme $\mathcal{V}$ est inclus dans $\mathcal{U}$, J majore $I(\lambda)$.

b) La fonction $I(.)$, borne supérieure de fonctions continues sur $\mathcal{P}$ est semi continue inférieurement.

Si Γ est un borélien de $\mathcal{P}$, on posera :

$$I(\Gamma) = \inf_{\lambda \varepsilon \Gamma} I(\lambda) \quad .$$

(Si Γ est vide , on pose $I(\Gamma) = +\infty$) .

Si K est un compact de $\mathcal{P}$, il existe un $\lambda \in K$ telle que $I(\lambda) = I(K)$.

c) Pour toute $v \in \mathcal{V}$, tout $x \in E$ et toute stratégie δ

$$E_{x}^{\delta}[\exp\,(\sum_{K=o}^{n-1} v(X_K, A_K, X_{K+1}))] = E_{x}^{\delta}[\exp(n \int v \, dL^n)]$$

$$= E_{x}^{\delta}\{\prod_{K=o}^{n-2} \frac{u(X_K, A_K, X_{K+1}}{\pi\, u(X_K, A_K)} \; \frac{E_X^{\delta}[u(X_{n-1}, A_{n-1}, X_n) \mid \mathcal{B}_{n-1}]}{\pi\, u(X_{n-1}, A_{n-1})}\}$$

$$= E_{x}^{\delta}[\prod_{K=o}^{n-2} \frac{u(X_K, A_K, X_{K+1})}{\pi\, u(X_K, A_K)}] = \ldots = 1 \quad .$$

Soit Γ un borélien de $\mathcal{P}$. Par l'inégalité de Tchebichev, on obtient :

$$\boxed{P_{x}^{\delta}[L^n \in \Gamma] \leqslant \exp[-n \sup_{v \varepsilon \mathcal{V}} \inf_{\lambda \varepsilon \Gamma} \int v \, d\lambda]}$$

Reste à obtenir une propriété de minimax pour trouver une majoration du type $\exp[-n\ I(\Gamma)]$. C'est possible si Γ est un compact K de $\mathcal{P}$. Soit $i < I(K)$. La famille d'ouverts $\{\lambda; \int v\, d\lambda > i\}_{v\in\mathcal{V}}$ recouvre K ; il existe une partie finie $\{v_1, v_2, \ldots, v_N\}$ de $\mathcal{V}$ telle que

$$K = \bigcup_{j=1}^{N} K_j \quad , \quad \text{où } K_j = K \cap \{\lambda \ ; \int v_j\, d\lambda > i\} \quad .$$

$$P_x^\delta(L^n \in K_j) \leqslant e^{-ni}$$

$$P_x^\delta(L^n \in K) \leqslant N \sup_j P_x(L^n \in K_j) \leqslant N\, e^{-ni}$$

$$\boxed{\overline{\lim_{n\to\infty}}\ \frac{1}{n} \log \sup_{x,\delta} P_x^\delta\ (L^n \in K) \leqslant -I(K)} \qquad \blacksquare \ .$$

d) Etude des marginales de L^n , L_1^n et L_3^n .

Soit $\|\ \|$ la norme uniforme sur $\mathcal{P}$.

Soit $\Delta = \{\lambda \ ; \lambda_1 = \lambda_3\}$ et $\Delta_n = \{\lambda \ ; \|\lambda_1-\lambda_3\| \leqslant \frac{2}{n}\}$.

La fonction $\lambda \longmapsto \|\lambda_1-\lambda_3\|$ de $\mathcal{P}$ dans $\mathbb{R}$ est s.c.i. ; donc Δ_n et Δ sont fermés dans $\mathcal{P}$. Pour tout (x,δ) , $P_x^\delta(L^n \in \Delta_n)$ est égale à 1 .

Soit K un compact de $\mathcal{P}$. Si $K \cap \Delta = \bigcap_n (K \cap \Delta_n)$ est vide , il existe un n_o tel que $K \cap \Delta_n$ soit vide, donc $P_x\ (L^n \in K) = 0$, dès que n est supérieur à n_o . Sinon $I(K \cap \Delta) = \lim_n \uparrow I(K \cup \Delta_n)$. En effet pour chaque n , il existe $\lambda_n \in K \cap \Delta_n$ telle que $I(\lambda_n) = I(K \cap \Delta_n)$. Soit $\lambda \in K \cap \Delta$ une valeur d'adhérence de (λ_n) . La fonction I étant s.c.i.

$$I(K \cap \Delta) \leqslant I(\lambda) \leqslant \underline{\lim}_{n\to\infty} I(\lambda_n) = \lim_{n\to\infty} \uparrow I(K \cap \Delta_n)$$

et
$$I(K \cap \Delta) \geqslant \lim_{n\to\infty} \uparrow I(K \cap \Delta_n) \qquad \text{est évident .}$$

De plus, pour tout $m \in \mathbb{N}$

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \operatorname{Log} \sup_{x,\delta} P_x^\delta(L^n \in K) = \overline{\lim_{n\to\infty}} \frac{1}{n} \operatorname{Log} \sup_{x,\delta} P_x(L^n \in K \cap \Delta_m) \leqslant -I(K \cap \Delta_m)$$

<u>Théorème</u>

Si E et A sont polonais et π fellerienne, pour tout compact K de $\wp$:

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \operatorname{Log} \sup_{x,\delta} P_x^\delta(L^n \in K) \leqslant -I(K \cap \Delta) \quad .$$

Cette majoration exponentielle n'a bien sûr d'intérêt que lorsque $I\,(K \cap \Delta)$ n'est pas nul. Il est facile de voir ([4]) que $I\,(K \cap \Delta)$ est nul si et seulement si il existe une stratégie markovienne stationnaire s et une probabilité μ telles que $\mu \otimes s \otimes \pi$ soit dans K. On trouvera également dans [4] une application au problème du contrôle des chaînes de Markov sous contrainte, ainsi que l'extension aux fermés de $\wp$ dans le cas d'expériences indépendantes contrôlées.

C - MAJORATIONS DE CHERNOFF POUR DES CHAÎNES DE MARKOV NON CONTROLÉES

On supprime désormais actions, et stratégies (A est réduit à un point). Les notations restent les mêmes, mais $\mathcal{E}$, $\mathcal{U}$ et $\mathcal{P}$ sont définis sur $E^2 \otimes \mathcal{E}^{\otimes 2}$; à $\lambda \in \mathcal{P}$, on associera ses deux marginales λ_1 et λ_2 ; $(\Omega,\mathfrak{a}) = (E, \mathcal{E})^{\mathbb{N}}$; $\{\Omega, a, (P_x)_{x \in E}, (X_n)_{n \in \mathbb{N}}\}$ est une chaîne de Markov à valeurs dans $(E, \mathcal{E})$, de transition π.

Puisque E est polonais, on peut le compactifier ; soit $\hat{E}$ son compactifié et $\hat{\mathcal{E}}_1$ l'ensemble des v.a. continues sur E qui ont un prolongement continu à $\hat{E}$. Si E est localement compact à base dénombrable on le compactifie en ajoutant un point à l'infini ; si $\overset{\circ}{\mathcal{E}}_1$ est l'ensemble des fonctions continues sur E qui tendent vers 0 à l'infini, une fonction de $\hat{\mathcal{E}}_1$ est somme d'une fonction de $\overset{\circ}{\mathcal{E}}_1$ et d'une constante. On suppose désormais $\pi\hat{\mathcal{E}}_1$ inclus dans $\hat{\mathcal{E}}_1$. Alors π peut être prolongée à $\hat{E}$ par continuité pour la topologie étroite; π est alors fellérienne. L'information $I(\lambda) = I_{\lambda_1 \otimes \pi}(\lambda)$ d'une probabilité $\lambda \in \mathcal{P}$, ne change pas si elle est considérée comme une probabilité λ de $\hat{E}^2$ concentrée sur E^2 . Le théorème obtenu en B est donc encore valable.

Projection de l'information sur E

Soit α une probabilité sur $(E , \mathcal{E})$. On note $\mathcal{P}_\alpha^2$ l'ensemble des probabilités sur $(E^2, \mathcal{E}^{\otimes 2})$ dont les deux marginales valent α et

$$I_1(\alpha) = \inf\{I(\lambda) \ ; \ \lambda \in \mathcal{P}_\alpha^2\} .$$

Si Γ est un borélien de $\mathcal{P}_1$, on note $I_1(\Gamma) = \inf\{I_1(\alpha) \ ; \ \alpha \in \mathcal{P}_1\}$

Théorème

Si E est polonais et si $\pi\hat{\mathcal{E}}_1$ est inclus dans $\hat{\mathcal{E}}_1$ (en particulier si E est l.c.d. et si $\pi\overset{\circ}{\mathcal{E}}_1$ est inclu dans $\overset{\circ}{\mathcal{E}}_1$) , alors pour toute $\alpha \in \mathcal{P}_1$

$$I_1(\alpha) = \sup_{u \in \mathcal{U}_1} \int \operatorname{Log} \frac{u}{\pi u} \, d\alpha$$

Ce résultat est l'un des plus délicats établit dans [1] .

Démonstration

a) Soit J le terme de droite. Pour $u \in \mathcal{U}_1$,

$$I_1(\alpha) \geq \int \operatorname{Log} u(y)\, d\alpha(y) - \int \operatorname{Log} \pi u(x)\, d\alpha(x) \geq J \ .$$

b) Il suffit de montrer le résultat lorsque E est compact. En effet on obtiendra alors le résultat sur le compactifié $\hat{E}$ et

$$I_1(\alpha) = \sup_{u \in \mathcal{U}_1 \cap \hat{\mathcal{C}}} \int \operatorname{Log} \frac{u}{\pi u}\, d\alpha \leq J \ .$$

c) Supposons E métrique compact.

Soit $A = \{\lambda \ ; \ I_{\alpha \otimes \pi}(\lambda) \leq J\}$. Supposons $I_1(\alpha) > J$. Alors A et $\mathcal{P}_\alpha^2$ sont des compacts convexes disjoints de $\mathcal{P}$. D'après le théorème de séparation , il existe une $v \in \mathcal{C}$ telle que :

$$\begin{cases} \int v\, d\lambda \leq 0 & \text{si} \quad \lambda \in \mathcal{P}_\alpha^2 \\ \int v\, d\lambda \geq 1 & \qquad \lambda \in A \end{cases}$$

Montrons que pour tout $\varepsilon > 0$, il existe deux fonctions f et g de $\mathcal{C}_1$ telles que :

$$\begin{cases} v(x,y) \leq f(x) - g(y) + \varepsilon \\ \int f\, d\alpha = \int g\, d\alpha \ . \end{cases}$$

Soit en effet H la fonction de $\mathcal{P} \times \mathcal{C}_1 \times \mathcal{C}_1$ dans $\mathbb{R}$ définie par :

$$H(\lambda,f,g) = \int v\, d\lambda - \left(\int f\, d\lambda_1 - \int f\, d\alpha\right) + \left(\int g\, d\lambda_2 - \int g\, d\alpha\right)$$

Pour $\lambda \in \mathcal{P}_\alpha^2$, $H(\lambda,f,g) \leq 0$.

Sinon, par exemple $\lambda_1 \neq \alpha$ et il existe une $f \in \mathcal{C}$ telle que

$\int f\, d\lambda_1 - \int f\, d\alpha > 0$.

$$\inf_{f,g} H(\lambda,f,g) \leqslant \int v\, d\lambda - \sup_n \left[\left(\int f\, d\lambda_1 - \int f\, d\alpha\right) n\right] = -\infty \ .$$

Donc $\sup_{\lambda\in\mathcal{P}} \inf_{f,g} H(\lambda,f,g) \leqslant 0$.

Pour (f,g) fixé H est affine en λ et pour λ fixé H est affine en (f,g). Donc d'après le théorème du minimax de Sion ([7],[5])

$$\inf_{f,g} \sup_{\lambda} H[\lambda,f,g] = \sup_{\lambda} \inf_{f,g} H(\lambda,f,g) \leqslant 0 \ .$$

$$\sup_{\lambda} H(\lambda,f,g) = \sup_{x,y} \left[v(x,y) - f(x) + g(y) + \int f\, d\alpha - \int g\, d\alpha\right] \ .$$

Pour $\varepsilon > 0$, il existe $(\hat{f},\hat{g}) \in \mathcal{C}_1^2$ tel que :

$$\sup_{x,y} \left[v(x,y) - \hat{f}(x) + \hat{g}(y)\right] \leqslant \int \hat{g}\, d\alpha - \int \hat{f}\, d\alpha + \varepsilon \ .$$

Soit $\begin{cases} f = \hat{f} - \int \hat{f}\, d\alpha + \int \hat{g}\, d\alpha \\ g = \hat{g} \end{cases}$: $\begin{cases} \int f\, d\alpha = \int g\, d\alpha \\ v(x,y) \leqslant f(x) - g(y) + \varepsilon . \end{cases}$

Prenons $\varepsilon < 1$; soit $\lambda \in \mathcal{P}_\alpha^2$ tel que $\int [f(x) - g(y)]\lambda(dx,dy)$ soit nulle.

Alors λ n'est pas dans A et $I_{\alpha\otimes\pi}(\lambda) > J$.

Soit $w(x,y) = f(x) - g(y)$. La transformée de Cramer de la probabilité $w(\alpha\otimes\pi)$ est :

$$\lambda_w^{\alpha\times\pi}(a) = -\inf_u \mathrm{Log} \int e^{u(w-a)}\, d\alpha\otimes\pi$$

$$= \inf_{\{\lambda\, ;\, \int w\, d\lambda = a\}} I_{\alpha\otimes\pi}(\lambda)$$

Donc $\lambda_W^{\alpha\otimes\pi}(0) = -\inf_u \mathrm{Log} \int e^{-u[f(x)-g(y)]}\ \alpha(dx)\pi(x,dy) \geqslant J$.

Pour tout $\eta > 0$, il existe un u tel que :

$$\int e^{u[f(x)-g(y)]}\ \alpha(dx)\ \pi(x,dy) \leqslant e^{-J+\eta}$$

Posons $W(y) = e^{-ug(y)}$ et $V(x) = e^{-uf(x)}$:

$$\int \frac{\pi W}{V}\, d\alpha \leqslant e^{-J+\eta}$$

$$\int \mathrm{Log}\, \frac{\pi W}{V}\, d\alpha = \int \mathrm{Log}\, \pi W\, d\alpha - \int \mathrm{Log}\, V\, d\alpha \leqslant \mathrm{Log} \int \frac{\pi W}{V} \leqslant -J+\eta$$

Mais $\int \mathrm{Log}\, V\, d\alpha = -\int u\, f\, d\alpha = -\int u\, g\, d\alpha = \int \mathrm{Log}\, W\, d\alpha$

$$-\, I_1(\alpha) \leqslant \int \mathrm{Log}\, \pi W\, d\alpha - \int \mathrm{Log}\, W\, d\alpha \leqslant -J+\eta$$

et $\qquad I \leqslant I_1(\alpha)$. ■

<u>Projection de la majoration sur E</u>

<u>Théorème</u>

Si E est polonais et si $\pi\hat{\mathcal{E}}_1 \subset \hat{\mathcal{E}}_1$, pour tout compact K de $\mathcal{P}_1$

$$\overline{\lim_{n\to\infty}}\ \frac{1}{n}\, \mathrm{Log}\ \sup_x P_x[L_1^n \in K] \leqslant -\, I_1(K)$$

Démonstration :

Posons $\mathcal{V}_1 = \{\mathrm{Log}\, \frac{u}{\pi u}\ ;\ u \in \mathcal{U}\}$. Grâce au théorème précédent, on établit comme sur E^2 que :

- pour $v \in \mathcal{V}_1$, $E_x[\exp(n \int v\ d\, L_n^1)] = 1$

- pour Γ borélien de $\mathcal{P}_1$, $P_x[L_1^n \in \Gamma] \leqslant \exp(-n \sup_{v\in\mathcal{V}} \inf_{\lambda\in\Gamma} \int v\ d\lambda)$

La démonstration s'achève comme sur E^2 .

Remarque

L'ensemble $\mathcal{P}^2_\alpha$ est un compact $\mathcal{P}$: il est fermé et c'est une partie tendue de $\mathcal{P}$. En effet, pour tout $\varepsilon > 0$, il existe K compact de E tel que $\alpha(K^c) \leqslant \varepsilon$; pour $\lambda \in \mathcal{P}^2_\alpha$, $\lambda((K \times K)^c) \leqslant \lambda(K^c \times E) + \lambda(E \times K^c) \leqslant 2\varepsilon$. Donc il existe une $\lambda \in \mathcal{P}^2_\alpha$ telle que $I_1(\lambda) = I(\lambda)$.

D'autre part, $\alpha \longrightarrow I_1(\alpha)$ est s.c.i. et pour tout compact K de $\mathcal{P}$, il existe $\alpha \in K$ et $\lambda \in \mathcal{P}^2_\alpha$ tels que $I_1(K) = I_1(\alpha) = I(\lambda)$.

Application au cas où E est un espace métrique compact

Dans ce cas, soit $f \in \mathcal{C}$. La fonction $\lambda \longrightarrow \int f\, d\lambda$ est continue de $\mathcal{P}$ dans $\mathbb{R}$. Soit F un fermé de $\mathbb{R}$

$$\overline{\lim}\ \frac{1}{n} \operatorname{Log} \sup_x P_x \left[\frac{1}{n} \sum_{k=o}^{n-1} f(X_k, X_{k+1}) \in F\right] \leqslant -H_f(F)$$

où $H_f(F) = \inf\limits_{\{\lambda ; \int f\, d\lambda \in F,\ \lambda_1 = \lambda_2\}} I(\lambda)$.

De même si f est s.c.i. de E dans $\mathbb{R}$, $\lambda \longmapsto \int f\, d\lambda$ est s.c.i. de $\mathcal{P}$ dans $\mathbb{R}$ et pour $a \in \mathbb{R}$

$$\overline{\lim_x}\ \frac{1}{n} \operatorname{Log} \sup_x P_x \left[\frac{1}{n} \sum_{k=o}^{n-1} f(X_k, X_{k+1}) \leqslant a\right] \leqslant -H_f((-\infty a]) .$$

De même, soit $g \in \mathcal{C}_1$, et F un fermé de $\mathbb{R}$

$$\overline{\lim}\ \frac{1}{n} \operatorname{Log} \sup_x P_x \left[\frac{1}{n} \sum_{k=o}^{n-1} g(X_k) \in F\right] \leqslant -h_g(F)$$

où $h_g(F) = \inf\limits_{\{\alpha ; \int F\, d\alpha \in \Gamma\}} I_1(\alpha)$.

La fonction g peut être prise s.c.i., si $F = (-\infty, a]$. ∎

Extensions : Pour obtenir une majoration du type précédent pour les ensembles fermés de $\mathcal{P}$, Donsker et Varadhan font une hypothèse qui revient à peu près à $\{\pi(x,\cdot)\}_{x \in E}$ est une famille tendue ... ce qui en dehors du cas indépendant, et du cas où E est compact semble peu fréquent.

Conditions pour que les bornes supérieures obtenues soient strictement négatives

Proposition

$I(\lambda) = 0$ si et seulement si $\lambda = \lambda_1 \otimes \pi$

$I_1(\lambda) = 0$ si et seulement si $\alpha = \alpha\,\pi$ (α est une probabilité invariante de la chaîne)

Démonstration :

Pour $\alpha \in \mathcal{P}_1$, $\alpha\pi$ est la probabilité qui à $A \in \mathcal{E}$ associe

$$\alpha\,\pi\,(A) = \int \alpha(dx)\ \pi(x,A)\ .$$

La seconde marginale de $\alpha \otimes \pi$ est $\alpha\pi$.

$I(\lambda) = I_{\lambda_1 \otimes \pi}(\lambda)$ est nulle si et seulement si $\lambda = \lambda_1 \otimes \pi$.

Il existe une $\lambda \in \mathcal{P}_\alpha$ telle que $I_1(\lambda) = I^2(\lambda)$.

Donc si $I_1(\lambda)$ est nulle, $\lambda = \alpha \otimes \pi$; mais la seconde marginale de λ est α, et $\alpha\pi = \alpha$. ∎

Conséquence

Soit K un compact de $\mathcal{P}_1$; si K ne contient aucune probabilité invariante, $I_1(K)$ est strictement positif et pour tout $\varepsilon > 0$, il existe n_o tel que pour $n \geqslant n_o$ et pour tout $x \in E$

$$P_x[L_1^n \in K] \leqslant e^{-n[I_1(K)-\varepsilon]}\ .$$

Cas des chaînes de markov irréductibles

Soit π_n la n^e itérée de π : $P_x(X_n \in \cdot) = \pi_n(x,\cdot)$.

Pour $z \in]0,1[$, on considère la transition :

$$\pi^z(x,\cdot) = (1-z) \sum_{n \geqslant 1} z^{n-1}\ \pi_n(x,\cdot)\ .$$

La chaîne est dite irréductible si il existe une mesure ϕ telle que $\phi << \pi^z(x,\cdot)$ pour tout $x \in E$. Pour une telle chaîne ou bien il n'existe pas de probabilité invariante (cas transient ou récurrent nul) ou bien il existe une probabilité invariante μ unique telle que pour toute v.a. f positive sur E^2 et tout $x \in E$ on ait :

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=o}^{n-1} f(X_k, X_{k+1}) = \int f \, d\mu \otimes \pi \qquad P_x \text{ p.s.}$$

Si E est compact, le seul cas possible est le cas récurrent positif [8]. Donc pour f s.c.i. de E^2 dans $\mathbb{R}$ et pour $a < \int fd\, \mu \otimes \pi$, il existe un $b > 0$ et un n_o , tels que $n \geqslant n_o$ implique :

$$\sup_x P_x \left[\frac{1}{n} \sum_{k=o}^{n-1} f(X_k, X_{k+1}) \leqslant a\right] \leqslant e^{-nb} \quad .$$

Un théorème analogue a été établi par N. Maigret (exposé suivant) sans hypothèse topologique pour les chaînes récurrentes au sens de Doeblin. Mais la chaîne précédente est récurrente au sens de Doeblin dès que sa mesure invariante a un support non vide ([2]) .

On est donc dans un cadre très voisin de celui de N. Maigret.

D - MINORATIONS DE CHERNOFF POUR UNE CHAÎNE DE MARKOV IRRÉDUCTIBLE

I - TRANSFORMATION DE LA CHAÎNE DE MARKOV

Soit $\alpha \in \mathcal{P}_1$ telle que $I_1(\alpha)$ soit finie. On a vu qu'il existe une $\lambda \in \mathcal{P}^2_\alpha$ telle que $I_1(\alpha) = I_{\alpha\otimes\pi}(\lambda)$. Puisque $\lambda_1 = \alpha$ et E est polonais, il existe une transition $\bar{\pi}$ de E dans E telle que $\lambda = \alpha\otimes\bar{\pi}$ (et $\alpha\otimes\bar{\pi} << \alpha\otimes\pi$) . En outre $\alpha\bar{\pi} = \alpha$: α est invariante pour $\bar{\pi}$. On cherche à montrer qu'on peut choisir $\bar{\pi}$ telle que la chaîne associée à $\bar{\pi}$ soit récurrente Harris positive. La transformation ainsi obtenue permettra d' obtenir des minorations comme la transformation des lois dans le cas indépendant. On suppose toujours E polonais et $\pi\hat{\mathcal{C}}_1 \subset \hat{\mathcal{C}}_1$.

a) Considérons tout d'abord une suite $(u_n)_{n\in\mathbb{N}} \subset \mathcal{U}$ telle que :

$$I(\lambda) = I_1(\alpha) = \lim_{n\to\infty} \int \frac{u_n}{\pi u_n}\, d\alpha \quad .$$

Soit $\lambda_n(dx,dy) = \alpha(dx)\, \dfrac{u_n(y)}{\pi u_n(x)}\, \pi(x,dy)$.

La probabilité $\alpha\otimes\bar{\pi}$ est absolument continue par rapport à $\alpha\otimes\pi$ donc à λ_n et

$$\begin{aligned} I_{\lambda_n}(\lambda) &= \int \mathrm{Log}\, \frac{d\lambda}{d\lambda_n}\, d\lambda \\ &= \int \mathrm{Log}\, \frac{d\lambda}{d\,\alpha\otimes\pi}\, d\lambda + \int \mathrm{Log}\, \frac{d\,\alpha\otimes\pi}{d\lambda_n}\, d\lambda \\ &= I_1(\alpha) - \int \mathrm{Log}\, \frac{u_n(y)}{\pi u_n(x)}\, d\lambda(x,y) = I_1(\alpha) - \int \mathrm{Log}\, \frac{u_n}{\pi u_n}\, d\alpha \end{aligned}$$

$$\lim_{n\to\infty} I_{\lambda_n}(\lambda) = 0 \ .$$

D'après A.3, $\lim_{n\to\infty} ||\lambda_n - \lambda|| = 0$ et $\dfrac{u_n(y)}{\pi u_n(x)}$ converge dans $L^1(\alpha\otimes\pi)$ vers $\dfrac{d\lambda}{d\,\alpha\otimes\pi}(x,y)$.

Suivant [10] , considérons :

$$\Delta = \{\theta \ ; \ \theta \in L^1(\alpha \otimes \pi), \ \theta(x,y)\theta(x',y') = \theta(x,y')\theta(x',y)$$
$$\text{pour } (\alpha \otimes \pi)^2 \text{ presque tout } (x,y,x',y')\}$$

Δ est fermé dans $L^1(\alpha \otimes \pi)$; si $\hat{\theta}$ est une version de $\theta \in \Delta$ non nulle , il existe $(x' \ y')$ où $\hat{\theta}(x' \ y') \neq 0$: $\hat{\theta}(x,y) = \dfrac{\hat{\theta}(x'y) \ \hat{\theta}(x y')}{\hat{\theta}(x'y')} = \dfrac{a(x)}{b(y)}$.

Donc $\dfrac{d\lambda}{d\alpha \otimes \pi}(x,y)$ a une version égale à $\dfrac{a(x)}{b(y)}$ où a est v.a. finie et b une v.a. strictement positive.

La suite $\displaystyle\int \frac{u_n(y)}{\pi \ u_n(x)} \pi(x,dy) = 1$ converge vers $\displaystyle\int \frac{a(x)}{b(y)} \pi(x,dy) = 1$

pour tout x , sauf sur un ensemble N négligeable pour α , donc pour tout x en modifiant a sur l'ensemble N.

Posons : $\begin{cases} \overline{\pi}(x,\cdot) = \dfrac{a(x)}{b(y)} \pi(x,\cdot) \\ \\ h(x,y) = \dfrac{a(x)}{b(y)} \end{cases}$

Alors $\overline{\pi}$ est une probabilité de transition et α est invariante pour $\overline{\pi}$.

- <u>Supposons qu'il existe une $\beta \in \mathcal{P}$ telle que pour tout $x \in E$, $\pi(x,\cdot)$ et β soient équivalentes</u> .

On pose $\pi(x,dy) = \pi(x,y) \beta(dy)$. Alors la chaîne associée à $\overline{\pi}$ est α irréductible.

En effet $\alpha = \alpha \overline{\pi} \ll \beta$. Soit $p = \dfrac{d\alpha}{d\beta}$. Pour β presque tout y

$$p(y) = \frac{1}{b(y)} \int \alpha(dx) \ a(x) \ \pi(x, y)$$

donc $\alpha(p > 0) = \alpha(b < \infty) = 1$.

Soit $A \in \mathcal{E}$ tel que $\alpha(A) > 0$.

$$\overline{\pi}(x, A) = a(x) \int \frac{\pi(xy)}{b(y)} 1_A(y) \beta(dy) > 0 .$$

La chaîne de Markov associée à $\overline{\pi}$ est donc récurrente positive de probabilité invariante α .

2 - MINORATION DE CHERNOFF LORSQUE LES PROBABILITÉS $(\pi(x,\cdot))_{x\varepsilon E}$ SONT TOUTES ÉQUIVALENTES

Soit $\alpha \varepsilon \mathcal{P}$ tel que $I_1(\alpha) < \infty$. On lui associe d'après ce qui précède une transition $\overline{\pi}$, d'une chaîne de Markov récurrente positive de probabilité invariante α , et une v.a. h sur $E \times E$ telle que pour tout x , $\frac{d\,\overline{\pi}(x,\cdot)}{d\,\pi(x,\cdot)} = h(x,\cdot)$.

$$I_1(\alpha) = I_{\alpha\otimes\pi}(\alpha\otimes\overline{\pi}) = \int \operatorname{Log} h \; d\,\alpha\otimes\pi \quad .$$

Pour tout x , il existe un ensemble N , $\overline{\pi}(x,\cdot)$ négligeable tel que pour tout $A \varepsilon \mathcal{E}$:

$$\pi(x,A) = \int 1_A(y) \; \frac{1}{h(x,y)} \; \overline{\pi}(x,dy) + \overline{\pi}(x, A \cap N)$$

$$\geqslant \int 1_A(y) \; \frac{1}{h(x,y)} \; \overline{\pi}(x,dy)$$

Considérons sur l'espace canonique $(\Omega , \mathcal{A})$, la famille de probabilités $(\overline{P}_x)_{x\varepsilon E}$ telles que :

$$\begin{cases} \overline{P}_x(X_o = x) = 1 \\ \overline{P}_x[X_{n+1} \varepsilon A \mid \mathcal{B}_n] = \overline{\pi}(X_n, A) \end{cases}$$

Soit $B \varepsilon \sigma(X_1 \ldots X_n)$. On peut lui associer $\Gamma \varepsilon \mathcal{E}^n$ tel que $B = \{(X_1 \ldots X_n) \varepsilon \Gamma\}$.

$$P_x(B) = \int 1_\Gamma(x_1,\dots,x_n)\pi(x,dx_1)\pi(x_1,dx_2) \dots \pi(x_{n-1},dx_n)$$

$$\geqslant \int 1_\Gamma(x_1,\dots,x_n) \frac{1}{h(x,x_1)\dots h(x_{n-1},x_n)} \overline{\pi}(x,dx_1)\overline{\pi}(x_1,dx_2)\dots\overline{\pi}(x_{n-1},dx_n)$$

$$\geqslant \overline{E}_x\left[1_B \frac{1}{h(X_oX_1)h(X_1X_2) \dots h(X_{n-1}X_n)}\right]$$

Soit V un voisinage de α . Prenons $B = \{L_2^n \varepsilon V\}$.

La chaîne $(\Omega , \mathcal{A}, \overline{P}_x , (X_n)_{n\varepsilon \mathbb{N}})$ étant récurrente :

a) $\overline{P}_x$ p.s. (L_2^n) converge en loi vers α donc $\lim\limits_{n\to\infty} \overline{P}_x[L_2^n \varepsilon V] = 1$.

b) $\overline{P}_x$ p.s. $\frac{1}{n} \sum\limits_{k=o}^{n-1} \text{Log } h(X_k,X_{k+1})$ tend vers $I_1(\alpha) = \int \text{Log } h \, d\alpha \otimes \pi$

$$\lim_{n\to\infty} \overline{P}_x \left[\frac{1}{n} \sum_{k=o}^{n-1} \text{Log } h(X_k, X_{k+1}) \leqslant I_1(\alpha) + \varepsilon\right] = 0$$

$$P_x(B) \geqslant e^{-n(I+\varepsilon)}\{1 - \overline{P}_x(L_2^n \varepsilon V^c) - \overline{P}_x\left[\sum_{k=o}^{n-1} \text{Log } h(X_k,X_{k+1}) \leqslant I_1(\alpha) + \varepsilon\right]\}$$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \text{Log } P_x(L_2^n \varepsilon V) \geqslant - I_1(\alpha) .$$

Supposons V ouvert , soit W un voisinage de α fermé inclus dans V . Il existe un n tel que $\{\beta \; ; \sup\limits_{\gamma\varepsilon W} ||\beta-\gamma|| \leqslant \frac{2}{n}\}$ soit inclus dans V .

Alors $\{L_2^n \varepsilon W\} \subset \{L_1^n \varepsilon V\}$ et

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \text{Log } P_x(L_1^n \varepsilon V) \geqslant -I_1(\alpha) .$$

Théorème

Supposons E polonais et $\pi\hat{\mathcal{C}} \subset \hat{\mathcal{C}}$.

Si les transitions $\{\pi(x,\cdot)\}_{x\in E}$ sont toutes équivalentes, alors pour tout ouvert G de $\mathcal{P}$ et pour tout x :

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \text{ Log } P_x(L_1^n \in G) \geqslant -I_1(G)$$

En effet la relation que l'on vient d'obtenir est évidente, si $I_1(\alpha) = \infty$ (ce qui est en particulier le cas lorsque α n'est pas absolument continue par rapport à β).

3 - EXTENSION

Soit $z \in]0,1[$. Considérons la transition $\pi^z = (1-z) \sum_{n\geqslant 1} z^{n-1} \pi_n$. On peut construire une chaîne de Markov associée à π^z de la manière suivante . Soit sur $(\Omega', \mathcal{A}', P')$ une suite $(\tau_n)_{n\geqslant 1}$ de v.a. indépendantes telles que :

$$P'(\tau_n = i) = (1 - z)\, z^{i-1} \qquad (i \geqslant 1) \quad .$$

Soit $\begin{cases} T_o = 0 \\ T_n = \tau_1 + \ldots \tau_n \quad \text{si} \quad n \geqslant 1 . \end{cases}$

Alors $\{\Omega \times \Omega' , \mathcal{A} \otimes \mathcal{A}' , P_x \otimes P' , (Y_n)_{n\in\mathbb{N}}\}$ est une chaine de markov associée à π^z si on pose :

$$Y_n(\omega,\omega') = X_{T_n(\omega')}(\omega)$$

Montrons que le théorème précédent est encore valable si les probabilités $\{\pi^z(x,\cdot)\}_{x\varepsilon E}$ sont équivalentes à une mesure β . Ce sera en particulier le cas si la chaîne est une chaîne récurrente (positive ou nulle) de mesure invariante μ et si pour tout x , $\pi(x,\cdot) << \mu$ (car, sous les hypoyhèses de récurrence, $\pi^z(x,\cdot) >> \mu$).

Soit $\alpha \varepsilon \mathcal{P}$, telle que $\alpha << \beta$. Considérons V un voisinage ouvert de α et W un autre voisinage de α , tel que pour $\varepsilon > 0$,
$\{\nu \; ; \sup_{\gamma\varepsilon W} ||\gamma - \nu|| \leqslant 2\varepsilon\} \subset V$.

Soit $A \varepsilon \mathcal{E}$. Posons :

$$\overline{L}_1^n(\omega,\omega',A) = \frac{1}{n} \sum_{k=o}^{n-1} 1_A[Y_k](\omega,\omega')$$

Supposons : $\dfrac{T_{n-1} - (n-1)}{n} \leqslant \varepsilon$. Pour tout A

$$\overline{L}_1^n(A) \leqslant L_1^{T_{n-1}}(A) \leqslant L_1^n(A) + \frac{\varepsilon}{2} \quad .$$

Mais parmi les T_{n-1} termes qui figurent dans $L_1^{T_{n-1}}(A)$ au plus $n\varepsilon$ ne figurent pas dans L_n^1 , donc :

$$L_1^{T_{n-1}}(A) \leqslant \overline{L}_1^n(A) + \varepsilon \quad .$$

Donc $$||L_1^n - \overline{L}_1^n|| \leqslant 2\varepsilon$$

$$P_x(L_1^n \varepsilon V) \geqslant P_x(\overline{L}_1^n \varepsilon W) - P'\left[\frac{T_{n-1}-(n-1)}{n} \geqslant \frac{\varepsilon}{2}\right]$$

$$\geqslant - I_1^z(\alpha) - P'(\frac{T_n}{n} \geqslant 1 + \frac{\varepsilon}{2})$$

où I_1^z est l'information calculée à partir de π^z .

On cherche à passer à la limite si z tend vers 0

$$E'(e^{\sigma T_n}) = (1-z) \sum_{i=1}^{n} z^{i-1} e^{\sigma_i} = \frac{(1-z)e^\sigma}{1-z\, e^\sigma} \quad \text{si } z\, e^\sigma < 1$$

$$P'(T_n \geqslant n(1+\varepsilon)) \leqslant e^{-n(1+\varepsilon)\sigma} E'(e^{\sigma T_n}) \leqslant e^{-n(1+\varepsilon)} \frac{(1-z)^n e^{n\sigma}}{(1-z\, e^\sigma)^n}$$

$$\frac{1}{n} \operatorname{Log} P'\ (T_n \geqslant n(1+\varepsilon)) \leqslant -\varepsilon\sigma + \operatorname{Log}(1-z) - \operatorname{Log}(1-z\ e^{\sigma})$$

$$\lim_{z\to 0} \frac{1}{n} \operatorname{Log} P'(T_n \geqslant n(1+\varepsilon)) = 0$$

et cette limite est atteinte uniformément en n .

D'autre part $\lim\limits_{z\to 0} I_1^z(\alpha) = I_1(\alpha)$. En effet :

$$I_1^z(\varepsilon) = \sup_{u\varepsilon} \int \operatorname{Log} \frac{u}{\pi^z u}\, d\alpha \leqslant \sup_{u\varepsilon} \int \operatorname{Log} \frac{u}{(1-z)\pi u}\, d\alpha$$

$$\leqslant \frac{1}{1-z} I_1(\alpha)$$

$$\overline{\lim_{z\to 0}}\ I_1^z(\alpha) \leqslant I_1(\alpha)$$

$$I_1(\alpha) = \sup_{u\varepsilon\,\mathcal{U}_1} \int \operatorname{Log} \frac{u}{\pi u}\, d\alpha \leqslant \sup_{u\varepsilon\mathcal{U}_1} \underline{\lim_{z\to 0}} \int \operatorname{Log} \frac{u}{\pi^z u}\, du$$

$$\leqslant \underline{\lim_{z\to 0}} \sup_{u\varepsilon\,\mathcal{U}_1} \int \operatorname{Log} \frac{u}{\pi^z u}\, du \leqslant \underline{\lim_{z\to 0}}\ I_1^z(\alpha) .$$

Donc $\underline{\lim\limits_{n\to\infty}}\ P_x(L_1^n \varepsilon V) \geqslant -I_1(\alpha)$.

<u>Théorème</u>

Supposons E polonais , $\pi\,\hat{\mathcal{C}}_1 \subset \hat{\mathcal{C}}_1$ et la chaîne associée à π récurrente Harris positive de mesure invariante μ. Si $\pi(x,\cdot)$ est absolument continue par rapport à μ pour tout x de E , alors pour tout ouvert G de $\mathcal{P}$ et tout $x \varepsilon E$,

$$\underline{\lim_{n\to\infty}}\ P_x(L_1^n \varepsilon G) \geqslant -I_1(G) \quad .$$

<u>Remarque</u> : Le théorème reste vrai si les probabilités $(\pi^z(x,\cdot))_{x\varepsilon E}$ sont toutes équivalentes, sans hypothèse de récurrence .

<u>Corollaire 1</u>

Sous les mêmes hypothèses que le théorème, soit f continue de E dans $\mathbb{R}$ et O ouvert de $\mathbb{R}$. Pour tout x :

$$\underline{\lim}\ \frac{1}{n}\ \mathrm{Log}\ P_x\left[\frac{1}{n}\sum_{k=o}^{n-1} f(X_k) \in O\right] \geqslant - \inf_{\{\alpha;\int d\alpha \in O\}} I(\alpha)$$

Si f est s.c.i. et O une demi droite $]-\infty, a[$ le théorème reste vrai.

Corollaire 2

Pour une chaîne de Markov fellerienne, définie sur un espace compact métrique et récurrente de probabilité invariante μ telle que $\pi(x,\cdot)$ soit absolument continue par rapport à μ pour tout x, on a pour toute f continue de E dans $\mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{n}\ \mathrm{Log}\ E_x\left[\exp \sum_{i=o}^{n-1} f(X_i)\right] = \sup_{\alpha\in \mathcal{P}_1} \left(\int f\ d\alpha - I_1(\alpha)\right) .$$

Cela résulte des majorations et minorations obtenues dans ce cas par une démonstration de Varadhan [9]. Inversement Gärtner établit dans [3] des théorèmes de grandes déviations en supposant que pour toute v.a. bornée la limite de la suite $\frac{1}{n}\ \mathrm{Log}\ E_x\left[\exp\left(\sum_{i=o}^{n-1} f(X_i)\right)\right]$ existe et est finie.

BIBLIOGRAPHIE

[1] DONSKER - VARADHAN : Asymptotic evaluation of certain Markov process expectations for large time. Com. on pure and appl. maths.
I Vol. 28 p. 1-47 (1975)
II Vol. 29 p. 279-301 (1976)
III Vol. 29 p. 389-461 (1976)

[2] COGBURN : A uniform theory for sums of Markov chain transition probabilities Ann. prob. Vol. 3 , 2, p. 191-214 (1975)

[3] GÄRTNER : On large deviations for the invariant measure. Theory of prob. and appl., Vol. 22, p. 24-39 (1977)

[4] MAIGRET : Inégalités de Chernoff pour des chaines de Markov controlées. Z. Wahrscheinlichkeit (1979).

[5] PARTHASARATHY - RAGHARAN : Some topics in two person games (Elsevier 1971)

[6] REVUZ : Markov chains. Holden Day

[7] SION : On general minimax theorem Pacific J. Math. 8 p. 171-176

[8] TUOMINEN.TWEEDIE : Markov chains with continuous components - A paraître .

[9] VARADHAN : Asymptotic probabilities and differential equations - Com. on pure and appl. Maths. Vol. 19, p. 261-285 (1966)

[10] JEAN SAINT RAYMOND : Quelques remarques sur un article de Donsker et Varadhan - Séminaire de probabilités XII Strasbourg (Springer-Verlag)

Marie DUFLO

Université Paris-Nord
Département Mathématiques
E.R.A. 532 "Statistique Appliquée"
Avenue Jean-Baptiste Clément
93430 VILLETANEUSE.