Z. Wahrscheinlichkeitstheorie verw. Gebiete
66, 41–60 (1984)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Théorèmes de grandes deviations pour des mesures aléatoires

F. Portal et A. Touati

Université de Paris-Nord, Centre Scientifique et Polytechnique, Ave. J. B. Clément,
F-93430 Villetaneuse, France

Resume. Il s'agit de majorer l'écart entre deux mesures aléatoires (la loi empirique d'un n échantillon et la vraie loi, un processus ponctuel et son compensateur prévisible, une mesure de Poisson et son intensité ...), ceci en vue d'applications statistiques. L'outil principal est l'information de Kullback des mesures bornées, qu'on étudie dans la première partie. On donne enfin des applications dans les quatrième et cinquième parties. La cinquième partie étant plus spécialement consacrée aux applications statistiques.

I. Introduction

I.1. Si E est muni d'une tribu $\mathscr{E}$ (on choisira toujours pour $\mathscr{E}$ la tribu borélienne $\mathscr{B}_E$, lorsque E est topologique), on note $\hat{\mathscr{U}}(E)$, $\mathscr{M}(E)$, $\mathscr{P}(E)$ respectivement l'ensemble des variables aléatoires positives bornées et d'inverses bornées, l'ensemble des mesures positives bornées et l'ensemble des probabilités sur E. En particulier si E est polonais (métrique séparable, complet), on munit $\mathscr{M}(E)$ et $\mathscr{P}(E)$ de la topologie de la convergence étroite, on note $\|\cdot\|$ la norme de la variation totale sur $\mathscr{M}(E)$ définie par

$$\|\mu\| = \sup\left\{\frac{|\int f\,d\mu|}{\sup\limits_{x\in E}|f(x)|};\ f\colon E\to\mathbb{R} \text{ continue bornée}\right\}$$

et on désigne par $\mathscr{U}(E)$ le sous-ensemble des fonctions continues de $\hat{\mathscr{U}}(E)$. On omettra la mention de $\mathscr{E}$ quand il n'y aura pas d'ambiguité. Enfin, une mesure aléatoire sur $(E,\mathscr{E})$ sera pour nous une mesure de transition positive et σ-finie de $(\Omega,\mathscr{F})$ dans $(E,\mathscr{E})$.

I.2. Un grand nombre de problèmes statistiques peuvent être décrits comme suit: on observe un processus $X=(\Omega,\mathscr{F},\mathbb{P},(X_t)_{t\in\mathscr{T}})$ à valeurs dans $(E,\mathscr{E})$ adapté à une filtration $F=(\mathscr{F}_t)_{t\in\mathscr{T}}$, $\mathscr{T}=\mathbb{N}$ ou $\mathscr{T}=\mathbb{R}^+$. Dans le cas discret l'information qu'apportent les n premières observations sur le phénomène physique décrit par le processus $X=(X_n)_{n\geqq 0}$ est contenue dans la répartition empirique associée aux fonctions aléatoires $X_1, X_2, \ldots, X_n$

$$\eta_n = \sum_{k=1}^{n} \varepsilon_{X_k}$$

(ε_{X_j} étant la masse de Dirac en X_j).

Une question naturelle se pose en statistique séquentielle où l'on suit le phénomène au cours du temps: que peut-on dire de «mieux» sur la loi de X_n connaissant $\mathscr{F}_{n-1}$? C'est-à-dire sur la famille de variables aléatoires $\{P(X_n \in A \mid \mathscr{F}_{n-1}),\ A \in \mathscr{E}\}$. Supposons qu'il existe des versions régulières $\{\lambda_{n-1}(\omega, \cdot)\}_{n\geqq 1}$ des probabilités conditionnelles $P(X_n \in \cdot \mid \mathscr{F}_{n-1})$ et notons $\tilde{\eta}_n = \sum_{k=1}^{n} \lambda_{k-1}(\omega, \cdot)$. Les exemples les plus simples sont les suivants.

Exemple 1. Répartitions empiriques.
Soit $(X_n)_{n\geqq 1}$ une suite de fonctions aléatoires indépendantes de même loi.

$$\tilde{\eta}_n = n\mu.$$

Exemple 2. Chaînes de Markov.
Soit $X=\{\Omega, \mathscr{F}, \mathbb{P}, (\mathbb{P}_x)_{x\in E}, (X_n)_{n\in\mathbb{N}}\}$ une chaîne de Markov, adaptée à $\mathbb{F}$, dont on note π_n la transition à l'instant n (i.e. $\pi_n(X_n, A) = \mathbb{P}(X_{n+1} \in A \mid \mathscr{F}_n)$ pour $A \in \mathscr{E}$). Si $\eta_n = \sum_{k=0}^{n} \varepsilon_{X_{k+1}}$ alors $\tilde{\eta}_n(\cdot) = \sum_{k=0}^{n-1} \pi_k(X_k, \cdot)$.

Pour ces deux exemples il est d'usage d'utiliser η_n pour faire de l'inférence statistique sur $\tilde{\eta}_n$. Ceci constitue le point de départ de l'idée de «compensation», que Aalen [1] a utilisée pour introduire les martingales en statistique des processus. On remarque que, pour tout $A \in \mathscr{E}$, le processus $(M_n^A)_{n\geqq 0}$ défini par $M_n^A = \eta_n(A) - \tilde{\eta}_n(A)$ si $n \geqq 1$ et par $M_0^A = 0$ est une $(\mathbb{F}, \mathbb{P})$ martingale centrée. Par conséquent $\eta_n(A)$ est un estimateur sans biais de $\tilde{\eta}_n(A)$. On peut remarquer aussi que, pour toute fonction $u \in \hat{\mathscr{U}}(E)$, l'espérance conditionelle

$$\begin{aligned}\mathbb{E}[u(X_{k+1}) \mid \mathscr{F}_k] &= \int \lambda_k(dx)\, u(x) = \exp[\operatorname{Log} \int \lambda_k(dx)\, u(x)] \\ &\leqq \exp(\int \lambda_k(dx)\, u(x) - 1),\end{aligned}$$

car pour tout $x \geqq 0$, $\operatorname{Log} x \leqq x-1$. Il en résulte que le processus $(S_n)_{n\geqq 0}$ défini par

$$S_n = \prod_{i=1}^{n} u(X_i) \exp(\int (1-u)\, d\tilde{\eta}_n) \quad \text{si } n \geqq 1 \quad \text{et } S_0 = 1$$

est une $(\mathbb{F}, \mathbb{P})$ surmartingale positive; de sorte que

$$(*) \qquad \mathbb{E}[\exp(\int \operatorname{Log} u(x)\, d\eta_n(x) + \int (1-u(x))\, d\tilde{\eta}_n(x))] \leqq 1.$$

Cette inégalité apparaît fréquemment comme on peut le voir sur les exemples suivants;

Exemple 3. Processus ponctuels marqués.

On suppose données une suite de $\mathbb{F}$-temps d'arrêt finis $(T_n)_{n\geqq 1}$ vérifiant $T_0=0$ et $\lim\limits_{n\to\infty} T_n=+\infty$ et une suite $(X_n)_{n\geqq 1}$ de fonctions mesurables de $(\Omega, \mathscr{F}_{T_n})$ dans $(E,\mathscr{E})$. On définit pour chaque t, la mesure aléatoire $\eta_t=\sum\limits_{n\geqq 1} \varepsilon_{X_n} \mathbb{1}_{\{T_n\leqq t\}}$, $\mathbb{F}^\eta=(\mathscr{F}_t^\eta)_{t\geqq 0}$ étant la filtration naturelle de $(\eta_t)_{t\geqq 0}$ et $(\tilde{\eta}_t)_{t\geqq 0}$ le compensateur prévisible de $(\eta_t)_{t\geqq 0}$. Le processus

$$M_t=\prod_{T_n\leqq t} u(x_n) \exp \int_E (1-u(x))\,\tilde{\eta}_t(dx)$$

est pour tout $u\in\hat{\mathscr{U}}(E)$ une surmartingale. En tout t les mesures $(\eta_t, \tilde{\eta}_t)$ vérifient donc $(*)$:

$$\mathbb{E}\left[\exp \int_E \operatorname{Log} u(x)\,\eta_t(dx)+\int_E (1-u(x))\,\eta_t(dx)\right]\leqq 1.$$

Exemple 4. Mesures aléatoires de Poisson ou de Cox.

On suppose encore pour cet exemple que $(E,\mathscr{E})=(\mathbb{R}^d, \mathscr{B}_{\mathbb{R}}^d)$. Soit λ une mesure de Radon positive σ-finie sur E et η une mesure aléatoire de Poisson d'intensité λ. Si λ est bornée, on a, pour tout $u\in\hat{\mathscr{U}}(E)$

$$(*) \qquad \mathbb{E}\left[\exp\left(\int_E \operatorname{Log} u\, d\eta+\int_E (1-u)\, d\lambda\right)\right]=1.$$

Si λ n'est pas bornée, soient $\lambda^{(r)}$ et $\eta^{(r)}$ respectivement les traces de λ et η sur la boule B_r de centre 0 et de rayon $r(r>0)$ alors la propriété $(*)$ est encore vérifiée pour les couples de mesures $\eta^{(r)}$ et $\tilde{\eta}^{(r)}=\lambda^{(r)}$. Elle s'étend sans aucune difficulté aux mesures de Cox. Compte tenu de ce qui précède, il est donc raisonnable de poser la définition suivante:

Définition. Soient η et $\tilde{\eta}$ deux mesures aléatoires positives bornées sur $(E,\mathscr{E})$. On dit que η est compensée par $\tilde{\eta}$ si, pour toute fonction $u\in\hat{\mathscr{U}}(E)$, on a la propriété suivante:

$$(*) \qquad \mathbb{E}\left[\exp\left(\int \operatorname{Log} u\, d\eta+\int (1-u)\, d\tilde{\eta}\right)\right]\leqq 1.$$

Le mot définition a été mis entre « » car il s'agit d'une définition formelle, a priori sans grand intérêt ($\tilde{\eta}=\eta$, convient!). Elle donne cependant, comme on vient de le voir, un cadre qui s'adapte à beaucoup d'exemples, où $\tilde{\eta}$ est unique moyennant une hypothèse supplémentaire (le plus souvent compensatrice prévisible ...).

Comme on veut utiliser η pour faire de l'inférence statistique sur $\tilde{\eta}$, il est naturel de rechercher dans un premier temps quel type d'information nous apporte η sur $\tilde{\eta}$. Cela nous conduira à étendre l'information de Kullback pour les probabilités aux mesures positives bornées. Les théorèmes de grandes déviations sur l'écart entre η et $\tilde{\eta}$, traités dans la deuxième partie de l'article, seront basés sur cette information. La troisième partie est une application directe des résultats de la deuxième partie aux exemples ci-dessus.

II. Information de Kullback

II.1. Si $\mathbb{P}$ et $\mathbb{Q}$ sont deux probabilités sur $(E, \mathscr{E})$, l'information de Kullback de $\mathbb{P}$ sur $\mathbb{Q}$ est définie dans [3] par

$$I_{\mathbb{Q}}(\mathbb{P}) = \sup_{u \in \mathscr{U}} \{ \int \operatorname{Log} u \, d\mathbb{P} - \operatorname{Log} \int u \, d\mathbb{Q} \} \qquad \text{et } I_{\mathbb{Q}}(\mathbb{P})$$

est fini si et seulement si $\mathbb{P}$ est absolument continue par rapport à $\mathbb{Q}$ et $\operatorname{Log} \frac{d\mathbb{P}}{d\mathbb{Q}} \in L^1(\mathbb{P})$: on a alors la relation $I_{\mathbb{Q}}(\mathbb{P}) = \int \operatorname{Log} \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{P}$.

On généralise cette information aux mesures positives bornées sur $(E, \mathscr{E})$ en suivant des idées de [3]. Si α et β appartiennent à $\mathscr{M}(E)$ l'information de α sur β est encore notée $I_\beta(\alpha)$ et est ici définie par:

$$I_\beta(\alpha) = \sup_{u \in \mathscr{U}} [\int \operatorname{Log} u \, d\alpha + \int (1-u) \, d\beta].$$

II.1.2. **Lemme 1.** *$I_\beta(\alpha)$ est finie si et seulement si $\alpha \ll \beta$ et $\operatorname{Log} \frac{d\alpha}{d\beta} \in L^1(\alpha)$. On a alors*

$$I_\beta(\alpha) = \beta(E) - \alpha(E) + \int \operatorname{Log} \frac{d\alpha}{d\beta} d\alpha.$$

$$I = \beta(E) - \alpha(E) + \int \operatorname{Log} \frac{d\alpha}{d\beta} d\alpha \qquad \text{et } a = \frac{d\alpha}{d\beta}.$$

Démonstration. a) Si $\alpha \ll \beta$ et $\operatorname{Log} \frac{d\alpha}{d\beta} \in L^1(\alpha)$, alors $I_\beta(\alpha) \leqq I$. En effet:

$$\int \operatorname{Log} u \, d\alpha = \int \operatorname{Log} \frac{u}{a} d\alpha + \int \operatorname{Log} a \, d\alpha \leqq \int \left(\frac{u}{a} - 1 \right) d\alpha + \int \operatorname{Log} a \, d\alpha;$$

$$\int \operatorname{Log} u \, d\alpha + \int (1-u) \, d\beta \leqq \beta(E) - \alpha(E) + \int \operatorname{Log} a \, d\alpha < +\infty.$$

b) Si $I_\beta(\alpha) < \infty$ alors $\alpha \ll \beta$ et $\operatorname{Log} \frac{d\alpha}{d\beta} L^1(\alpha)$.

[1] $\alpha \ll \beta$: considérons en effet $u = (b+1) \mathbb{1}_A + \mathbb{1}_{A^c}$

$$\int \operatorname{Log} u \, d\alpha + \int (1-u) \, d\beta = \alpha(A) \operatorname{Log}(b+1) - b \, \beta(A) \leqq I_\beta(\alpha).$$

D'où, si $\beta(A) = 0$: $\operatorname{Log}(b+1) \alpha(A) \leqq I_\beta(\alpha)$, quel que soit $b > 0$ et $\alpha(A) = 0$.

[2] $\operatorname{Log} a \in L^1(\alpha)$: posons $a_n = \left(a \vee \frac{1}{n} \right) \wedge n$.

La fonction $x \mapsto x(\operatorname{Log} x)^-$ est bornée par $\frac{1}{e}$, d'où:

$$\limsup_{n \to \infty} \int a_n (\operatorname{Log} a_n)^- \, d\beta = \int a (\operatorname{Log} a) \, d\beta = \int (\operatorname{Log} a)^- \, d\alpha < +\infty.$$

D'autre part:

$$\int (\operatorname{Log} a_n)^+ \, d\alpha \leqq I_\beta(\alpha) + \int (a_n - 1) \, d\beta + \int (\operatorname{Log} a_n)^- \, d\alpha.$$

$$\limsup_{n\to\infty} \int (\operatorname{Log} a_n)^+ \, d\alpha \leqq I_\beta(\alpha) + \alpha(E) - \beta(E) + \int (\operatorname{Log} a)^- \, d\alpha.$$

$$(\operatorname{Log} a_n)^+ = \operatorname{Log}(a \wedge n)^+ \quad \text{croît vers } (\operatorname{Log} a)^+.$$

On a:

$$\int (\operatorname{Log} a)^+ \, d\alpha < \infty, \qquad \int (\operatorname{Log} a)^- \, d\alpha < \infty$$

et

$$I = \int (\operatorname{Log} a) \, d\alpha + \beta(E) - \alpha(E) \leqq I_\beta(\alpha).$$

Les parties a) et b) prouvent le lemme 1.

II.1.3. **Lemme 2.** *$I_\beta(\alpha) = 0$ si et seulement si $\alpha = \beta$.*

Démonstration. Si $I_\beta(\alpha) = 0$ et $a = \dfrac{d\alpha}{d\beta}$ alors: $\int (1 - a - a \operatorname{Log} a) \, d\alpha = 0$. D'où $a = 1$, β p.s. et $\alpha = \beta$.

II.2. A partir de maintenant E est polonais et $\mathscr{E}$ est sa tribu Borélienne. On montre alors comme dans [3] que $I_\beta(\alpha) = \sup_{u \in \mathscr{U}} \int \operatorname{Log} u \, d\alpha + \int (1-u) \, d\beta$ ($\mathscr{U}$ est le sous-ensemble des applications continues de $\hat{\mathscr{U}}$). Il résulte alors de la continuité de $(\alpha, \beta) \to \int \operatorname{Log} u \, d\alpha + \int (1-u) \, d\beta$ pour toute u de $\mathscr{U}$ que l'application $(\alpha, \beta) \to I_\beta(\alpha)$ est semi-continue inférieurement.

II.3. *Comparaison de $I_\beta(\alpha)$ et $\|\alpha - \beta\|$*

Soit $\|\cdot\|$ la norme de variation totale sur $\mathscr{M}(E)$. Soit $b > 0$ et $u \in \hat{\mathscr{U}}$ défini par $u = (b+1) \mathbb{1}_A + \mathbb{1}_{A^c}$ pour $A \in \mathscr{E}$. Notons $I = I_\beta(\alpha)$, on a:

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

$$\alpha(A) - \beta(A) \leqq \frac{I + \beta(A)(b - \operatorname{Log}(b+1))}{\operatorname{Log}(b+1)} \leqq \frac{I + \|\beta\| (b - \operatorname{Log}(b+1))}{\operatorname{Log}(b+1)}.$$

Posons

$$\hat{I} = \frac{I}{\|\beta\|} \qquad \text{et } f(b) = \frac{\hat{I} + b - \operatorname{Log}(b+1)}{\operatorname{Log}(b+1)}.$$

Alors:

$$\|\alpha - \beta\| + \|\alpha\| - \|\beta\| \leqq 2 \sup_{A \in \mathscr{E}} (\alpha(A) - \beta(A)) \leqq 2 \|\beta\| \inf_{b > 0} (f(b)).$$

Soit $(b+1)\operatorname{Log}(b+1) - b = g(b)$, $f(b)$ est minimum pour $\hat{I} = g(b)$. Comme la fonction $x \mapsto (x+1)\operatorname{Log}(x+1) - x = g(x)$ croît de zéro à l'infini, il existe $\hat{b} > 0$, tel que $\hat{I} = g(\hat{b})$ d'où:

$$\|\alpha - \beta\| + \|\alpha\| - \|\beta\| \leqq 2 \|\beta\| \, g^{-1}\left(\frac{I}{\|\beta\|}\right),$$

$$I_\beta(\alpha) \geqq \|\beta\| \, g\left[\frac{\|\alpha - \beta\| + \|\alpha\| - \|\beta\|}{2 \|\beta\|}\right].$$

II.4. *Information d'un borélien de $\mathscr{M}(E)$*

II.4.1.: On note $\mathscr{B}(\mathscr{M}(E))$ la tribu borélienne de $\mathscr{M}(E)$. Pour $\Gamma \in \mathscr{B}(\mathscr{M}^2(E))$, l'information de Γ est, par définition, le nombre $I(\Gamma) = \inf\limits_{(\alpha,\beta)\in\Gamma} I_\beta(\alpha)$. Il résulte de la propriété de semi-continuité inférieure de $(\alpha,\beta) \mapsto I_\beta(\alpha)$ et du lemme 2 de II.1.3. que, si Γ est compact, $I(\Gamma)$ est nulle si et seulement si Γ ne rencontre pas la diagonale $\mathscr{D} = \{(\alpha,\alpha) \mid \alpha \in \mathscr{M}(E)\}$. Nous utiliserons aussi les remarques suivantes.

Soit $b \in]0,1[$; alors, si $I(\Gamma) < \infty$, pour tout $u \in \mathscr{U}$, l'ensemble

$$\sigma_u^b = \{(\alpha,\beta) \mid \int \operatorname{Log} u\, d\alpha + \int (1-u)\, d\beta > I(\Gamma)(1-b)\}$$

est ouvert et la famille $(\sigma_u^b)_{u\in\mathscr{U}}$ recouvre Γ. Comme Γ est compact, il existe une sous-famille finie $\{\sigma_{u_1}^b, \ldots, \sigma_{u_p}^b\}$ qui recouvre Γ. Les nombre $I(\Gamma)$ et p ne changent pas par une transformation f bijective bimesurable de $(E,\mathscr{E})$ dans $(G,\mathscr{G})$. En effet, si $f(\alpha)$ et $f(\beta)$ désignent les mesures images de α et β par f on a:

$$I_{f(\beta)}(f(\alpha)) = I_\beta(\alpha),\ I(f(\Gamma)) = I(\Gamma) \quad \text{et } f(\Gamma) \subset \bigcup_{i=1}^{p} \sigma_{u_i \circ f^{-1}}$$

(si f n'est pas continue $\sigma_{u_i\circ f^{-1}}$ n'est pas forcément un ouvert, mais cela sera sans importance).

Soit enfin $t > 0$, alors l'ensemble $t\Gamma = \{(t\alpha, t\beta);\ (\alpha,\beta)\in\Gamma\}$ est tel que $I(t\Gamma) = tI(\Gamma)$ et est recouvert par la famille d'ouvert

$$\{(\alpha,\beta) \mid \int \operatorname{Log} u_i\, d\alpha + \int (1-u_i)\, d\beta > tI(\Gamma)(1-b)\}_{i=1,\ldots,p}.$$

II.4.2. *Information d'un compact convexe.* Si $\mathscr{U}$ est muni de la topologie de la convergence uniforme, alors, pour tout compact convexe K de $\mathscr{U}$ et pour tout compact Γ de $\mathscr{M}^2(E)$ ne rencontrant pas la diagonale on notera:

$$\mathscr{G}_{\alpha,\beta}(u) = \int \operatorname{Log} u\, d\alpha + \int (1-u)\, d\beta$$
$$I(K,\Gamma) = \inf_{(\alpha,\beta)\in\Gamma} \sup_{u\in K} \mathscr{G}_{\alpha,\beta}(u), \qquad J(K,\Gamma) = \sup_{u\in K} \inf_{(\alpha,\beta)\in\Gamma} \mathscr{G}_{\alpha,\beta}(u)$$

Proposition. *Pour tout convexe compact Γ de $\mathscr{M}^2(E)$:*

1) $J(K,\Gamma) = I(K,\Gamma)$.
2) $I(\Gamma) = \sup\limits_{K\subset\mathscr{U}} J(K,\Gamma) = \sup\limits_{K\subset\mathscr{U}} I(K,\Gamma)$.

Démonstration. Rappelons le théorème du minimax. Soit K un convexe compact, $\mathscr{F}$ un ensemble convexe de fonctions de $K \to \mathbb{R}$ tel que toute fonction f de $\mathscr{F}$ prenne une valeur positive et soit semi-continue supérieurement. Alors il existe $x_0 \in K$, tel que pour tout $f \in \mathscr{F}, f(x_0)$ soit positive.

1) Démontrons que l'ensemble $\mathscr{F} = \{\mathscr{G}_{(\alpha,\beta)}\}_{(\alpha,\beta)\in\Gamma}$ satisfait aux conditions du théorème du minimax. Puisque Γ est convexe et $(\alpha,\beta) \mapsto \mathscr{G}_{\alpha,\beta}$ est bilinéaire en (α,β), $\mathscr{F}$ est un ensemble de fonctions convexes. La concavité de la fonction logarithme permet d'affirmer que, pour chaque $(\alpha,\beta)\in\mathscr{M}^2(E)$, $u \mapsto \mathscr{G}_{\alpha,\beta}(u)$ est concave. Si $\mathscr{U}$ est muni de la topologie de la convergence uniforme l'application $u \mapsto \mathscr{G}_{\alpha,\beta}(u)$ est continue. Soit alors $b > 0$; par définition de $I(K,\Gamma)$, pour

chaque $(\alpha,\beta)\in\Gamma$, il existe $u\in K$ tel que $\mathscr{G}_{\alpha,\beta}(u)\geqq I(K,\Gamma)-b$. On peut donc appliquer le théorème du minimax et il existe $u_0\in K$ tel que, pour tout $(\alpha,\beta)\in\Gamma$, $\mathscr{G}_{\alpha,\beta}(u_0)\geqq I(K,\Gamma)-b$: $J(K,\Gamma)\geqq I(K,\Gamma)-b$. L'inégalité inverse étant toujours réalisée, on a $J(K,\Gamma)=I(K,\Gamma)$.

2) Il suffit de démontrer que $\sup\limits_{K\in\mathscr{U}} J(K,\Gamma)=I(\Gamma)$.

D'après II.4.1 pour tout $b>0$, il existe $u_1,\dots,u_p$ dans $\mathscr{U}$ tels que:

$$\sup_{1\leqq i\leqq p}\ \inf_{(\alpha,\beta)\in\Gamma}\mathscr{G}_{\alpha,\beta}(u_i)\geqq I(\Gamma)-b.$$

Si K est l'enveloppe convexe de $\{u_1,\dots,u_p\}$

$$I(\Gamma)\geqq I(K,\Gamma)=\sup_{u\in K}\ \inf_{(\alpha,\beta)\in\Gamma}\mathscr{G}_{\alpha,\beta}(u)\geqq I(\Gamma)-b.$$

II.4.3. *Information des compacts*

Lemme. *Soit $(\Gamma_t)_{t\in\mathbb{R}_+}$ une famille décroissante de compacts. Pour*

$$\Gamma=\bigcap_{t>0}\Gamma_t,\qquad \lim_{t\to\infty} I(\Gamma_t)=I(\Gamma).$$

Démonstration. 1) Si Γ est vide, il existe t_0 tel que pour $t\geqq t_0$, Γ_t est vide.

2) Si Γ n'est pas vide, la fonction $t\to I(\Gamma_t)$ est croissante et $I(\Gamma_t)\leqq I(\Gamma)$.

Soit $(t_n)_{n>0}$ une suite de $\mathbb{R}_+$ croissant vers l'infini, les (Γ_{t_n}) étant compacts et la fonction $(\alpha,\beta)\mapsto I_\beta(\alpha)$ étant semi-continue inférieurement, pour tout $n\in\mathbb{N}$, il existe (α_n,β_n) tel que $I_{\beta_n}(\alpha_n)=I(\Gamma_{t_n})$. Soit (α,β) une valeur d'adhérence de la suite (α_n,β_n), $(\alpha,\beta)\in\Gamma$ et

$$I(\Gamma)\leqq I_\beta(\alpha)\leqq\liminf_{n\to\infty} I_{\beta_n}(\alpha_n)=\liminf_{n\to\infty} I(\Gamma_{t_n})$$

puisque I est semi-continue inférieurement. D'où: $I(\Gamma)=\lim I(\Gamma_t)$.

III. Grandes deviations pour des mesures aleatoires

Soit η et $\tilde{\eta}$ deux mesures aléatoires vérifiant la relation $(*)$ de I.1.

III.1. *Comparaison élémentaire de η et $\tilde{\eta}$*
Soit $g: x\mapsto(x+1)\operatorname{Log}(x+1)-x$.

III.1.1. **Lemme.** *Pour tout $a>0$ et $A\in\mathscr{E}$:*

1) $\mathbb{P}(\eta(A)\geqq a(1+b);\ \tilde{\eta}(A)\leqq a)\leqq e^{-ag(b)}$ pour $b>0$,
2) $\mathbb{P}(\eta(A)\leqq(1-b)a;\ \tilde{\eta}(A)\geqq a)\leqq e^{-ag(-b)}$ pour $b\in]0,1[$.

Démonstration. Soit $A\in\mathscr{E}$. En prenant, dans $(*)$, $u=e^c\mathbb{1}_A+\mathbb{1}_{A^c}$ pour une constante $c>0$ strictement positive, on obtient:

$$\mathbb{E}(\exp c\,\eta(A)+(1-e^c)\,\tilde{\eta}(A))\leqq 1,$$

$$\mathbb{E}\left[\mathbb{1}_{\{\eta(A)\geqq(1+b)a;\,\tilde{\eta}(A)\leqq a\}}\exp(c(1+b)\,a+(1-e^c)\,a)\right]\leqq 1,$$

et

$$\mathbb{P}\{\eta(A)\geqq a(1+b);\tilde{\eta}(A)\leqq a\}\leqq \exp-(a\{c(1+b)+1-e^c\}).$$

Ceci est vrai pour tout $c>0$ et comme $\sup_{c>0}\{c(1+b)+1-e^c\}=g(b)$:

$$\mathbb{P}\{\eta(A)\geqq u(1+b);\tilde{\eta}(A)\leqq a\}\leqq e^{-ag(b)}.$$

Un calcul analogue en prenant $c<0$ permet d'obtenir le résultat (2).

III.1.2. **Proposition.** 1) *Pour tout* $a\in]0,1[$ *et* Γ *compact de* $\mathcal{M}^2$, *il existe un entier* p *ne dépendant que de* a *et* Γ *tel que, pour tout*
$t>0$: $\mathbb{P}[(\eta,\tilde{\eta})\in t\Gamma]\leqq p\,e^{-tI(\Gamma)(1-a)}$.

2) *Si de plus* Γ *est convexe alors, pour tout*
$t>0$: $\mathbb{P}[(\eta,\tilde{\eta})\in t\Gamma]\leqq e^{-tI(\Gamma)}$.

Démonstration. De la relation $(*)$ on tire que pour tout Γ de $\mathcal{M}^2$ et tout $u\in\mathcal{U}$:

$$\mathbb{E}[\mathbb{1}_{\{(\eta,\tilde{\eta})\in t\Gamma\}}\exp(\int \operatorname{Log} u\,d\eta+\int(1-u)\,d\tilde{\eta})]\leqq 1.$$

D'où:

$$\mathbb{P}\{(\eta,\tilde{\eta})\in t\Gamma\}\leqq \exp\{-t\sup_{u\in\mathcal{U}}\inf_{(\alpha,\beta)\in\Gamma}(\int \operatorname{Log} u\,d\alpha+\int(1-u)\,d\beta)\}.$$

Deux cas se présentent alors: si Γ est convexe compact, d'après la proposition II.4.2.:

$\sup_{u\in\mathcal{U}}\inf_{(\alpha,\beta)}(\int \operatorname{Log} u\,d\alpha+\int(1-u)\,d\beta)=I(\Gamma)$ et 2) est démontré. Si Γ n'est pas convexe, soit $a\in]0,1[$ quelconque et soit $\sigma^a_{u_1},\ldots,\sigma^a_{u_p})$ un sous-recouvrement de Γ par des ouverts de la forme

$$\sigma^a_u=\{(\alpha,\beta)\,|\int \operatorname{Log} u\,d\alpha+\int(1-u)\,d\beta>I(\Gamma)(1-a)\}:\ \mathbb{P}[(\eta,\tilde{\eta})\in t\,\sigma^a_{u_i}]\leqq e^{-tI(\Gamma)(1-a)}$$

et donc $\mathbb{P}\{(\eta,\tilde{\eta})\in t\Gamma\}\leqq \sum_{i=1}^{p}\exp\{-t\,I(\Gamma)(1-a)\}\leqq p\,e^{-tI(\Gamma)(1-a)}$. D'où l'assertion (1).

III.2.1. En général l'hypothèse de compacité est trop forte pour obtenir des théorèmes de grandes déviations intéressants. Par exemple, $\{(\alpha,\beta);|\int f\,d\alpha-\int f\,d\beta|\geqq a\}$ est, pour une fonction continue bornée, fermé mais n'est pas compact en général. C'est pourquoi on a besoin d'un théorème portant sur les fermés. Soit $(\eta_t,\tilde{\eta}_t)_{t\in T}$, avec $T=\mathbb{R}_+$ où $T=\mathbb{N}$, une famille de mesures aléatoires, telles que pour tout $t\in T$, $\tilde{\eta}_t$ compense η_t.

Théorème. *Soit* Δ *un compact de* $\mathcal{M}$. *On suppose que, pour tout* $t>0$ *et tout* $\omega\in\Omega$, $\dfrac{\tilde{\eta}_t(\omega)}{t}\in\Delta$. *Alors, pour tout* $a\in\mathbb{R}^+$, *il existe un compact convexe* Δ_a *de* $\mathcal{M}$ *tel que*

$$\limsup_{n\to\infty}\frac{1}{t}\operatorname{Log}\mathbb{P}\left[\frac{\eta_t}{t}\in\Delta_a^c\right]<-a.$$

Démonstration. La fonction $\mu \mapsto \|\mu\|$ étant continue sur $\mathcal{M}(E)$ elle est bornée sur le compact Δ; soit m sa borne supérieure. Soit $a>0$ quelconque, $b_i = e^{(a+i)^2}$,

$$c_i = \frac{\sqrt{\log b_i}}{g(b_i)} \text{ et } d \text{ tel que } m g(d) > a.$$

Les suites c_i et $c_i(1+b_i)$ tendent vers zéro lorsque i tend vers l'infini. L'ensemble Δ étant compact et E étant polonais, Δ est tendu. Donc il existe une suite K_i de compacts de E telle que $\tilde{\eta}_t(K_i^c) \leqq c_i t$. L'ensemble $\Delta_a = \{\mu \in \mathcal{M} \mid \mu(E) \leqq m(1+d),\ \mu(K_i^c) \leqq c_i(1+b_i)\}$ est compact convexe et

$$\mathbb{P}\left[\frac{\eta_t}{t} \notin \Delta_a\right] \leqq e^{-mtg(d)} + \sum_{i=1}^{\infty} e^{-c_i t g(b_i)}$$
$$\leqq e^{-at} + \sum_{i=1}^{\infty} e^{-(a+i)t} = \frac{e^{-at}}{1-e^{-at}}.$$

D'où le résultat.

Corollaire. *Sous les hypothèses précédentes, pour tout fermé Φ de $\mathcal{M}^2(E)$ qui ne rencontre pas la diagonale*

$$\limsup_{t \to +\infty} \frac{1}{t} \operatorname{Log} \mathbb{P}\left[\left(\frac{\eta_t}{t}, \frac{\tilde{\eta}_t}{t}\right) \in \Phi\right] \leqq -I(\Phi \cap (\mathcal{M} \times \Delta))$$

avec: $I(\Phi \cap (\mathcal{M} \times \Delta)) > 0$.

Démonstration. Il suffit d'écrire que $\Phi \subset (\Phi \cap \Delta_a \times \Delta) \cup (\Delta_a \times \Delta)^c$ et de remarquer que $\Phi_1^a = \Phi \cap (\Delta_a \times \Delta)$ est compact:

$$\mathbb{P}\left[\left(\frac{\eta_t}{t}, \frac{\tilde{\eta}_t}{t}\right) \in \Phi\right] \leqq \mathbb{P}\left[\frac{\eta_t}{t} \in \Delta_a^c\right] + \mathbb{P}\left[\left(\frac{\eta_t}{t}, \frac{\tilde{\eta}_t}{t}\right) \in \Phi_1^a\right].$$

D'où:

$$\limsup_{t \to +\infty} \frac{1}{t} \operatorname{Log} \mathbb{P}\left[\left(\frac{\eta_t}{t}, \frac{\tilde{\eta}_t}{t}\right) \in \Phi\right] \leqq \sup\{-a, -I(\Phi_1^a)\}.$$

Il suffit alors de montrer que $I(\Phi \cap (\mathcal{M} \times \Delta)) > 0$ car $I(\Phi \cap (\mathcal{M} \times \Delta)) \leqq I(\Phi_1^a)$ implique $\sup(-a, -I(\Phi_1^a)) \leqq \sup(-u, -I(\Phi \cap (\mathcal{M} \times \Delta))$.

Si $I(\Phi \cap (\mathcal{M} \times \Delta)) = \infty$, quel que soit $a>0$, on a:

$$\limsup_{t \to +\infty} \frac{1}{t} \operatorname{Log} \mathbb{P}\left[\left(\frac{\eta_t}{t}, \frac{\tilde{\eta}_t}{t}\right) \in \Phi\right] \leqq -a;$$

d'où le résultat du corollaire.

Si $I(\Phi \cap (\mathcal{M} \times \Delta)) < \infty$, il existe un m tel que l'ensemble $\alpha_m = \{\alpha;\ \exists \beta \in \Delta\ I_\beta(\alpha) \leqq m\}$ ne soit pas vide. Mais alors

$$I(\Phi \cap (\mathcal{M} \times \Delta)) = I(\Phi \cap (\overline{\alpha_{2m}} \times \Delta)) = I(\Phi \cap (\alpha_{2m} \times \Delta));$$

il suffit donc de voir que $I(\Phi \cap (\overline{\alpha_{2m}} \times \Delta)) > 0$. Pour cela montrons que, pour tout m, α_m est relativement compact. Puisque Δ est relativement compact dans

$\mathcal{M}(E)$ avec E polonais, il existe une suite de compact (K_i) de E tel que

$$\sup_{\beta\in\Delta}\beta(K_i^c)\leqq e^{-i}.$$

Soit

$$u_i=e^i\mathbb{1}_{K_i^c}+\mathbb{1}_{K_i};\qquad \int \operatorname{Log} u_i\,d\alpha+\int(1-u_i)\,d\beta=i\alpha(K_i^c)+(1-e^i)\,\beta(K_i^c);$$

pour $\alpha\in\alpha_m$, $\alpha(K_i^c)\leqq m+(e^{-i}-1)\,\beta(K_i^c)$, pour un certain $\beta\in\Delta$.

D'où:

$$\alpha(K_i^c)\leqq\frac{1}{i}[m+(e^i-1)e^i]\xrightarrow[i\to\infty]{}0.$$

Prenant $u=e\mathbb{1}_E$, on obtient de même: $\alpha(E)\leqq m+(e-1)\sup_{\beta\in\Delta}\beta(E)$. Et α_m est relativement compact: le corollaire est démontré.

IV. Quelques applications

IV.1.

IV.1.1. *Répartitions empiriques.*

Théorème. *Soit $X_1,\dots,X_n$, n observations indépendantes d'une fonction aléatoire X de loi μ à valeurs dans un espace polonais $(E,\mathcal{E})$. Soit $\bar\mu_n=\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{X_k}$ la répartition empirique associée à X. Pour tout fermé $\Phi\in\mathcal{P}(E)$ ne contenant pas μ:*

$$\limsup_{n\to\infty}\frac{1}{n}\operatorname{Log}\mathbb{P}[\bar\mu_n\in\Phi]<0.$$

Démonstration. C'est une application directe du paragraphe III.2.1. en remarquant que la mesure $n\bar\mu_n$ est compensée par $n\mu$. Ce résultat figure dans [9] et résulte aussi de [3].

IV.1.2. *Cas indépendant contrôlé.* On donne un ensemble mesurable $(A,\mathcal{A})$ d'actions (ou d'expériences) et une fonction mesurable $a\to\mu_a$ de A dans $\mathcal{P}(E)$. L'espace des épreuves étant $(\Omega,\mathcal{F})$, on définit une suite d'expériences $(A_n)_{n\geqq 0}$, fonctions mesurables de $(\Omega,\mathcal{F})$ dans $(A,\mathcal{A})$ et une suite d'observations $(X_n)_{n\geqq 0}$ fonctions aléatoires de $(\Omega,\mathcal{F})$ dans $(E,\mathcal{E})$. Une stratégie $\delta=(\delta_n)_{n\in\mathbb{N}}$ est définie en donnant $\delta_0\in\mathcal{A}$ constante et, pour tout $n\geqq 1$, une probabilité de transition δ_n de $(E^n,\mathcal{E}^n)$ dans $(A,\mathcal{A})$. On prend $(\Omega,\mathcal{F})=(A\times E,\mathcal{A}\times\mathcal{E})^{\mathbb{N}}$, on note la n^e coordonnée (A_{n-1},X_n). On définit alors pour chaque stratégie δ une probabilité $\mathbb{P}^\delta$ sur $(\Omega,\mathcal{F})$ par

$$\mathbb{P}^\delta[A_n\in da,\ X_{n+1}\in dy\,|\,X_1,\dots,X_n;\ A_0,\dots,A_{n-1}]=\delta_n(X_1,\dots,X_n,da)\,\mu(a,dy).$$

Considérons alors, pour $n\geqq 1$, les mesures aléatoires $\sigma_n=\varepsilon_{A_{n-1}}\times\varepsilon_{X_n}$ et $\tilde\sigma_n=\varepsilon_{A_{n-1}}\times\mu_{A_{n-1}}$. Pour toute stratégie δ et pour toute variable aléatoire $u\in\hat{\mathcal{U}}(A\times E)$ on a:

$$\begin{aligned}&\mathbb{E}^{\delta}[\exp\{\int \operatorname{Log} u(a,x)\,d\sigma_n(a,x)\\&\quad+\int(1-u(a,x))|d\tilde{\sigma}_n(a,x)\}|X_1,\ldots,X_{n-1};A_0,\ldots,A_{n-1}]\\&=\int\delta_{n-1}(X_1,\ldots,X_{n-1},da)\int_E u(a,x)\,\mu(a,dx)\exp\int(1-u(a,x))\,\mu(a,dx)\leqq 1.\end{aligned}$$

Donc $\eta_n=\sum_{k=1}^{n}\sigma_k$ et $\tilde{\eta}=\sum_{k=1}^{n}\tilde{\sigma}_k$ vérifient la relation (*). On suppose la famille $\{\mu_a\}_{a\in\mathscr{A}}$ tendue. Si $\mathscr{A}$ est métrique compact, $(\varepsilon_a\times\mu_a)_{a\in\mathscr{A}}$ est alors une famille tendue. Sinon on utilise les mesures aléatoires

$$\eta_n^0=\sum_{k=1}^{n}\varepsilon_{X_n}\quad\text{et}\quad\tilde{\eta}_n^0=\sum_{k=1}^{n}\mu_{A_{k-1}}.$$

Théorème. *Dans le cadre ci-dessus d'observations indépendantes* $(X_k)_{k\geq 1}$ *conditionnellement à un choix d'expériences décrit par la suite* $\{A_n\}_{n\geq 0}$ *à valeurs dans l'espace* $(A,\mathscr{A})$; *supposons que la famille* $\{\mu_a\}_{a\in A}$ *des lois des observations est tendue:*

a) *Si A est métrique compact, pour tout fermé Φ de $\mathscr{P}^2(A\times E)$ tel qu'il n'existe aucune probabilité δ sur $(A,\mathscr{A})$ telle que $(\delta\times\mu,\,\delta\times\mu)\in\Phi$:*

$$\limsup_{n\to+\infty}\frac{1}{n}\operatorname{Log}\sup_{\delta}\mathbb{P}^{\delta}\left[\left(\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{A_{k-1}}\times\varepsilon_{X_k},\ \frac{1}{n}\sum_{k=1}^{n}\varepsilon_{A_{k-1}}\times\mu_{A_{k-1}}\right)\in\Phi\right]<0.$$

b) *Si A est un ensemble mesurable quelconque, pour tout fermé Φ ne rencontrant pas la diagonale:*

$$\limsup_{n\to+\infty}\frac{1}{n}\operatorname{Log}\sup_{\delta}\mathbb{P}^{\delta}\left[\left(\frac{1}{n}\sum_{k=1}^{n}\varepsilon_{X_k},\ \frac{1}{n}\sum_{k=1}^{n}\mu_{A_{k-1}}\right)\in\Phi\right]<0.$$

Démonstration. La partie b) est une application directe du paragraphe III. Pour a) remarquons que $\frac{\tilde{\eta}_n}{n}\in\{\alpha\in\mathscr{P}(A\times E);\ \alpha=\alpha^{(1)}\times\mu\}=H$, et que l'ensemble H est fermé. Donc:

$$\limsup_{n\to+\infty}\frac{1}{n}\operatorname{Log}\sup_{\delta}\mathbb{P}^{\delta}\left[\left(\frac{\eta_n}{n},\frac{\tilde{\eta}_n}{n}\right)\in\Phi\right]\leqq -I(\Phi\cap(\mathscr{P}(A\times E)\times H)).$$

Mais $\Phi\cap(\mathscr{P}(A\times E)\times H)$ ne rencontre pas la diagonale de $\mathscr{P}^2(A\times E)$, s'il n'existe aucune probabilité δ sur $(A,\mathscr{A})$ telle que $(\delta\times\mu,\delta\times\mu)\in\Phi$.

Remarque. Considérons la stratégie qui consiste à toujours choisir l'action avec la loi δ (indépendante des expériences antérieures). Alors la suite $(A_{n-1},X_n)_{n>1}$ est indépendante de loi $\delta\times\mu$ et la suite $(A_{n-1},\mu_{A_{n-1}})_{n>1}$ aussi. On pouvait donc déduire du cas indépendant sans contrôle que, pour cette stratégie, le théorème ne peut être vrai, si Φ contient $\delta\times\mu$.

IV.2. Chaînes de Markov

IV.2.1. *Chaînes de Markov non homogènes.* Soit $(\Omega,\mathscr{F}(\mathbb{P}_x)_{x\in E},\ (X_n)_{n\in\mathbb{N}})$ une chaîne de Markov non homogène à valeurs dans $(E,\mathscr{E})$ telle que $\mathbb{P}_x\ (X_0=x)$

$=1$. Soit Π_n la transition à l'instant n. Pour u variable aléatoire bornée, strictement positive sur $(E^2, \mathscr{E}^2)$ et $\mathscr{F}_n = \sigma(X_0, \ldots, X_n)$ on a:

$$\mathbb{E}_x[u(X_n, X_{n+1}) | \mathscr{F}_n] = \int \Pi_n(X_n, dy)\, u(X_n, y) = \Pi_n u(X_n)$$

et pour $u \in \hat{\mathscr{U}}(E \times E)$ on obtient:

$$\begin{aligned}&\mathbb{E}_x[\exp(\operatorname{Log} u(X_n, X_{n+1}) + \int (1 - u(X_n, y))\, \Pi_n(X_n dy)) | \mathscr{F}_n]\\ &\quad = \exp\{1 - \Pi_n u(X_n) + \operatorname{Log} \Pi_n u(X_n)\} \leq 1.\end{aligned}$$

Posons $\delta_n = \varepsilon_{X_n, X_{n+1}}$ (la mesure de Dirac en (X_n, X_{n+1}) et $\tilde{\delta}_n = \varepsilon_{X_n} \times \Pi_n(X_n, \cdot)$. Si $\eta_n = \sum_{k=0}^{n-1} \delta_k$ alors η_n est compensée par $\tilde{\eta}_n = \sum_{k=0}^{n-1} \tilde{\delta}_k$ et on peut appliquer les résultats du paragraphe III.1.2. Si de plus, on suppose que la famille $(\Pi_n(x, \cdot))_{x \in E, n \in \mathbb{N}}$ tendue, on peut appliquer les résultats du paragraphe III.2.1. aux mesures aléatoires

$$\eta_n^0 = \sum_{k=0}^{n-1} \varepsilon_{X_{k+1}} \quad \text{et} \quad \tilde{\eta}_n^0 = \sum_{k=0}^{n-1} \Pi_k(X_k, \cdot).$$

Théorème. *Soit $(\Omega, \mathscr{F}, (\mathbb{P}_x)_{x \in E}, (X_n)_{n \in \mathbb{N}})$ une chaîne de Markov non homogène sur un espace polonais $(E, \mathscr{E})$ dont la transition à l'instant n est notée Π_n.*

a) *Pour tout compact Γ de $\mathscr{P}^2(E \times E)$ ne rencontrant pas la diagonale et pour tout $a \in]0, 1[$ on a:*

$$\sup_{x \in E} \mathbb{P}_x\left[\left(\frac{1}{n}\sum_{k=0}^{n-1} \varepsilon_{X_k, X_{k+1}}, \frac{1}{n}\sum_{k=0}^{n-1} \varepsilon_{X_k} \times \Pi_k(X_k, \cdot)\right) \in \Gamma\right] \leq p(a, \Gamma) \exp\{-n(1-a)\, I(\Gamma)\}$$

où $p(a, \Gamma)$ est un entier indépendant de n. Si de plus, Γ est convexe, on a:

$$\sup_{x \in E} \mathbb{P}_x\left[\left(\frac{1}{n}\sum_{k=0}^{n-1} \varepsilon_{X_k, X_{k-1}}, \frac{1}{n}\sum_{k=0}^{n-1} \varepsilon_{X_k} \times \Pi_k(X_k, \cdot)\right) \in \Gamma\right] \leq e^{-nI(\Gamma)}.$$

b) *Si la famille $(\Pi_n(x, \cdot)_{n \in \mathbb{N}, x \in E})$ est tendue, pour tout fermé Φ de $\mathscr{P}^2(E)$ ne rencontrant pas la diagonale:*

$$\limsup_{n \to \infty} \frac{1}{n} \operatorname{Log} \sup_{x \in E} \mathbb{P}_x\left[\left(\frac{1}{n}\sum_{p=0}^{n-1} \varepsilon_{X_{p+1}}, \frac{1}{n}\sum_{p=0}^{n-1} \Pi_p(X_p, \cdot)\right) \in \Phi\right] < 0.$$

IV.2.2. *Chaînes de Markov homogènes.* Si la chaîne précédente est homogène, on note Π la transition à chaque instant; pour $\alpha \in \mathscr{P}(E)$, on note $\alpha \cdot \Pi$ la probabilité définie par $\alpha \cdot \Pi(A) = \int \alpha(dx)\, \Pi(x, A)$ et $\alpha \times \Pi$ la probabilité sur E^2 définie par $\alpha \times \Pi(A \times A') = \int \alpha(dx)\, \mathbb{1}_A(x)\, \Pi(x, y'))$.

On dispose d'un résultat sur les grandes déviations mieux adapté aux chaînes de Markov que le résultat général III.2.1. On dit que α est une mesure invariante par Π, si $\alpha \cdot \Pi = \alpha$.

Théorème. *Soit $(\Omega, \mathscr{F}(\mathbb{P}_x)_{x \in E}, (X_n)_{n \in \mathbb{N}})$ une chaîne de Markov homogène de transition Π, fellérienne, sur un espace E compact. Si Φ est un fermé de $\mathscr{P}^2(E \times E)$ tel* *qu'il n'existe aucune probabilité μ invariante par Π telle que $(\mu\times\Pi,\ \mu\times\Pi)\in\Phi$ alors:*

$$\limsup_{n\to\infty}\frac{1}{n}\operatorname{Log}\sup_{x\in E}\mathbb{P}_x\left[\left(\frac{\eta_n}{n},\frac{\eta_n}{n}\right)\in\Phi\right]<0,$$

Démonstration. Si α est une probabilité sur $E\times E$, on note $\alpha^{(1)}$ et $\alpha^{(2)}$ respectivement la première et la seconde marginale. Soit

$$\Delta=\{\alpha;\alpha^{(1)}=\alpha^{(2)}\},\qquad \Delta_n=\left\{\alpha;\ \|\alpha^{(1)}-\alpha^{(2)}\|\leqq\frac{2}{n}\right\}$$

et

$$H=\{\alpha\in\mathcal{M}(E\times E);\ \alpha^{(1)}(E)=1,\ \alpha=\alpha^{(1)}\times\Pi\}$$

on a alors $\frac{\eta_n}{n}\in\Delta_n$ pour tout $n>0$ et $\frac{\tilde{\eta}_n}{n}\in H$. D'autre part la fonction $\alpha\to\|\alpha^{(1)}-\alpha^{(2)}\|$ étant semi-continue inférieurement de $\mathcal{P}(E)$ dans $\mathbb{R}$, Δ_n et Δ sont fermés.

Le fait que Π soit fellerienne permet d'affirmer que H est fermé. On a donc pour tout $m\leqq n$:

$$\mathbb{P}_x\left[\left(\frac{\eta_n}{n},\frac{\eta_n}{n}\right)\in\Phi\right]=\mathbb{P}_x\left[\left(\frac{\eta_n}{n},\frac{\eta_n}{n}\right)\in\Phi\cap(\Delta_m\times H)\right]$$

Le théorème IV.2.1. implique alors: $\forall m\geqq 1$

$$\limsup_{n\to+\infty}\frac{1}{n}\operatorname{Log}\sup_{x\in E}\mathbb{P}_x\left[\left(\frac{\eta_n}{n},\frac{\eta_n}{n}\right)\in\Phi\right]\leqq -I(\Phi\cap(\Delta_m\times H))$$

et le lemme II.4.3. permet d'affirmer que:

$$\limsup_{n\to\infty} I(\Phi\cap(\Delta_m\times H))=I(\Phi\cap(\Delta\times H)).$$

Or une condition pour que $I(\Phi\cap(\Delta\times H))$ soit strictement positif est que $\Phi\cap(\Delta\times H)$ ne rencontre pas la diagonale $\mathcal{D}$.

Or

$$\Phi\cap(\Delta\times H)\cap\mathcal{D}=\{(\alpha,\alpha)|\alpha^{(1)}=\alpha^{(2)}\alpha=\alpha^{(1)}\times\Pi,\ \alpha^{(1)}(E)=1\}$$

donc s'il n'existe pas de probabilité invariante par Π tel que $(\mu\times\Pi,\ \mu\times\Pi)\in\Phi$, $I(\Phi\cap(\Delta\times H)>0$ et le théorème est démontré. Ce théorème a été obtenu par N. Maigret [9] et est voisin de celui de Donsker et Varadhan [3].

IV.2.3. *Chaînes de Markov contrôlées.* On donne maintenant un ensemble $(A,\mathcal{A})$ mesurable d'actions (ou d'expériences) et une fonction mesurable $a\mapsto\mu_a$ de $\mathcal{A}$ dans $\mathcal{P}(E)$. L'espace des épreuves étant $(\Omega,\mathcal{F})$, on définit $(A_n)_{n\geqq 0}$ et $\delta=(\delta_n)_{n\in\mathbb{N}}$ comme en IV.1.2. Pour chaque n on donne une transition Π_n de $(E\times A,\ \mathcal{E}\times\mathcal{A})$ dans $(E,\mathcal{E})$ et on suppose que, pour $n>0$:

- $\mathbb{P}_x^\delta(X_0=x)=1$
- $\mathbb{P}_x^\delta[A_n\in da, X_{n+1}\in dy|X_1,\dots,X_n,A_0,\dots,A_{n-1}]$
$$=\delta\{(X_1,\dots,X_n);da\}\,\Pi_n(X_n,a,dy).$$

On s'intéressera plus spécialement aux stratégies aléatoires markoviennes stationnaires. Si s est une transition de E dans $\mathscr{A}$, on peut lui associer une stratégie s' définie par $s'=(s(X_n,\cdot))_{n\in\mathbb{N}}$. Si la transition Π_n introduite plus haut est homogène et désignée par Π, on notera $\Pi(x,s(x),C)=\int \Pi(x,y;C)\,s(x,da)$ et Π^s la transition $(x,C)\mapsto \Pi(x,s(x);C)$. Alors pour la stratégie s', (X_n) est une chaîne de Markov homogène de transition Π^s.

Considérons maintenant la mesure aléatoire $\eta_n=\sum_{k=0}^{n-1}\varepsilon_{X_k,A_k,X_{k+1}}$. Elle est compensée par $\tilde\eta_n=\sum_{k=0}^{n-1}\varepsilon_{X_k,A_k}\times\Pi(X_k,A_k;\cdot)$. De même la mesure $\eta_n^0=\sum_{k=0}^{n-1}\varepsilon_{X_{k+1}}$ est compensée par $\tilde\eta_n^0=\sum_{k=0}^{n-1}\Pi_k(X_k,A_k;\cdot)$.

Théorème. *Soit une chaîne de Markov contrôlée, à valeurs dans un espace métrique compact* $(E,\mathscr{E})$.

a) *Si A est métrique compact, pour tout fermé Π de $\mathscr{P}^2(A\times E\times E)$ qui ne rencontre pas la diagonale:*

$$\limsup_{n\to\infty}\frac{1}{n}\operatorname{Log}\sup_{\delta}\sup_{x\in E}\mathbb{P}_x^\delta\left[\left(\frac{1}{n}\sum_{k=0}^{n-1}\varepsilon_{(X_k,A_k,X_{k+1})}\,\frac{1}{n}\sum_{k=0}^{n-1}\varepsilon_{(X_k,A_k)}\times\Pi_k(X_k,A_k;\cdot)\right)\in\Gamma\right]<0.$$

b) *Si $\mathscr{A}$ est un ensemble mesurable quelconque, pour tout fermé Φ qui ne rencontre pas la diagonale:*

$$\limsup_{n\to\infty}\frac{1}{n}\operatorname{Log}\sup_{\delta}\sup_{x\in E}\mathbb{P}_x^\delta\left[\left(\frac{1}{n}\sum_{k=0}^{n-1}\varepsilon_{X_{k+1}},\ \sum_{k=0}^{n-1}\Pi_k(X_k,A_k;\cdot)\right)\in\Phi\right]<0.$$

IV.2.4. *Chaînes de Markov contrôlées homogènes.* On suppose maintenant la chaîne homogène de transition Π. Si α est une probabilité sur $(E\times A\times E,\ \mathscr{E}\times\mathscr{A}\times\mathscr{E})$ on note $\alpha^{(1)}$ et $\alpha^{(3)}$ les première et troisième marginales $\alpha^{(1,2)}$ la marginale sur $E\times\mathscr{A}$. Puisque E est polonais, on peut trouver une transition s de E dans $\mathscr{A}$ tel que $\alpha^{(1,2)}=\alpha\times s$. Si l'on note encore

$$\eta_n=\sum_{k=0}^{n-1}\varepsilon_{X_k,A_k,X_{k+1}},$$

alors

$$\tilde\eta_n=\sum_{k=0}^{n-1}\varepsilon_{X_k,A_k}\times\Pi(X_k,A_k;\cdot).$$

Soit $\Delta=\{\alpha;\ \alpha^{(1)}=\alpha^{(3)}\}$,

$$\Delta_n=\left\{\alpha\in\mathscr{P}(E\times A\times E);\ \|\alpha^{(1)}-\alpha^{(3)}\|\leqq\frac{2}{n}\right\}$$

et

$$H=\{\alpha\in\mathscr{P}(E\times A\times E);\ \alpha=\alpha^{(1,2)}\times\Pi\}.$$

On remarque alors que $\frac{\eta_n}{n} \in \Delta_n$ et $\frac{\tilde{\eta}_n}{n} \in H$ pour tout $n > 0$. Un raisonnement analogue à celui fait pour les chaînes de Markov homogènes non contrôlées permet d'énoncer la proposition suivante analogue à un résultat de [11].

Proposition. *Soit une chaîne de Markov contrôlée, homogène, dont l'espace d'état et l'espace d'actions sont métriques compacts. On suppose que sa transition Π est fellérienne c'est-à-dire possède la propriété suivante: «pour tout $u \in \mathbb{C}_b(E \times A \times E)$ la fonction v: $(x, y) \mapsto \int u(x, a, y)\, \Pi(x, a; dy)$ est continue». Alors pour tout fermé Φ tel qu'il n'existe pas de transition s de E dans $\mathscr{A}$ et de probabilité μ invariante par Π^s telles que les probabilités $(\mu \times s \times \Pi,\ \mu \times s \times \Pi)$ soient dans Φ:*

$$\limsup_{n \to +\infty} \frac{1}{n} \operatorname{Log} \sup_{\delta} \sup_{x \in E} \mathbb{P}_x^\delta \left[\left(\frac{\eta_n}{n}, \frac{\tilde{\eta}_n}{n} \right) \in \Phi \right] < 0.$$

IV.3. Processus ponctuels marqués

IV.3.1. On considère dans ce paragraphe η un processus ponctuel marqué dont le compensateur prévisible est $\tilde{\eta}$. Les notations et définitions sont celles de l'exemple I.2.3. On note $\eta^{(t)}$ et $\tilde{\eta}^{(t)}$ les restrictions de η et $\tilde{\eta}$ à $E^{(t)} =]0, t[\times E$, h_t la fonction de $E^{(t)}$ dans $E^{(1)}$ qui à $(s, x) \in E^{(t)}$ fait correspondre $\left(\frac{s}{t}, x\right) \in E^{(1)}$. On suppose de plus que l'ensemble $\left\{\frac{\tilde{\eta}^{(t)}}{t}(\omega);\ t > 0,\ \omega \in \Omega\right\}$ est relativement compact dans $\mathscr{M}(E^{(t)})$. Alors l'ensemble $\left\{\frac{h_t(\tilde{\eta}_t)}{t}(\omega);\ (t, \omega) \in \mathbb{R}_+^* \times \Omega\right\}$ est relativement compact dans $\mathscr{M}(E^{(1)})$. On est donc dans le cadre du paragraphe III.2.1. D'où le théorème suivante

Théorème. *Soit η un processus ponctuel marqué dont le compensateur prévisible $\tilde{\eta}$ est continu. On suppose que l'ensemble des mesures $\left\{\frac{\tilde{\eta}^{(t)}}{t}(\omega);\ (t, \omega) \in \mathbb{R}_+ \times \Omega\right\}$ est relativement compact dans $\mathscr{M}(E^{(t)})$, alors pour tout fermé Φ de $\mathscr{M}^2(E^{(1)})$ qui ne rencontre pas la diagonale il existe trois constantes $a < \infty$, $b > 0$ et $t_0 > 0$ telles que, pour tout $t > t_0$:*

$$\mathbb{P}\left[\left(\frac{h_t(\tilde{\eta}_t)}{t}, \frac{h_t(\tilde{\eta}^{(t)})}{t}\right) \in \Phi\right] \leqq a e^{-bt}.$$

En particulier:

i) *si Φ_1 est un fermé de $\mathscr{M}^2(E)$, on a pour tout $t \geqq t_0$*

$$\mathbb{P}\left[\left(\frac{\eta_t}{t}, \frac{\tilde{\eta}_t}{t}\right) \in \Phi_1\right] \leqq a e^{-bt}.$$

ii) *Si f est continue bornée de $E^{(1)}$ dans $\mathbb{R}$, on a, pour tout $t \geqq t_0$:*

$$\mathbb{P}\left[\left|\int_0^t \int_E f\left(\frac{s}{t}, x\right) d\eta(s, x) - \int_0^t \int_E f\left(\frac{s}{t}, x\right) d\tilde{\eta}(s, x)\right| \geqq a t\right] \leqq a e^{-bt}.$$

On donne dans la suite des applications de ce théorème.

IV.3.2. *Processus de sauts.* On donne une transition Π de $(E,\mathscr{E})$ dans $(E,\mathscr{E})$, une fonction q: $(E,\mathscr{E})\mapsto(]0,+\infty[,\mathscr{B}_{]0,+\infty[})$ et une famille de probabilités $(\mathbb{P}_x)_{x\in E}$ sur $(\Omega,\mathscr{F})$. On suppose que, pour tout $x\in E$:

$$\mathbb{P}_x(X_0=x)=1, \qquad \mathbb{P}_x(T_n\to\infty)=1,$$
$$\mathbb{P}_x[T_{n+1}-T_n\geqq t,\ X_{n+1}\in dx\,|\,\mathscr{F}_{T_n}]=e^{-q(X_n)t}\Pi(X_n,dx).$$

Soit $Y_t=\sum_{n\in\mathbb{N}} X_n \mathbb{1}_{(T_n\leqq t<T_{n+1})}$. Le processus $(\Omega,\mathscr{F},(\mathbb{P}_x)_{x\in E},(Y_t)_{t\geqq 0})$ est un processus de sauts. On lui associe la mesure ponctuelle

$\eta=\sum_{n\geqq 1}\varepsilon_{X_{n-1},X_n,T_n}$. Si (cf. exemple 3 de I.1) $\mathscr{F}_t=\mathscr{F}_t^\eta\vee\mathscr{F}_0$ le compensateur prévisible du processus ponctuel marqué η est

$$\tilde{\eta}_t(dx,dy)=\int_0^t q(x)\,\varepsilon_{Y_{s-}}(dx)\,\Pi(x,dy)\,ds.$$

On dispose, comme pour les chaînes de Markov, d'un résultat sur les grandes déviations mieux adapté aux processus de sauts que ceux du paragraphe III. Si α est une mesure bornée sur $E\times E$, on notera $\alpha^{(1)}$ et $\alpha^{(2)}$ respectivement la première et seconde marginales,

$$\Delta=\{\alpha\in\mathscr{M}(E\times E);\ \alpha^{(1)}=\alpha^{(2)}\}, \qquad \Delta_t=\left\{\alpha;\ \|\alpha^{(1)}-\alpha^{(2)}\|\leqq\frac{2}{t}\right\}$$

et $H=\{\alpha\in\mathscr{M}(E\times E);\ \exists\mu$ probabilité sur E, avec $\alpha^{(1)}=q\mu$ et $(q\mu)\times\Pi=\alpha\}$.

On note

$$(\mathbb{P}_t f)(x)=E_x f(x_t), \qquad a=q(\Pi-I)$$

le générateur infinitésimal du processus $u^1=\int_0^\infty e^{-t}P_t\,dt$. Si $(q\mu)$ est invariante par Π, alors $\mu a=0$ et donc $\mu u^1=\mu$. Or $\mu u^1(f)-\mu(f)$ est égale à

$$\int_E \mu(dx)\int_0^\infty e^{-t}\{P_t f(x)-f(x)\}\,dt=\int_0^\infty e^{-t}[(\mu P_t(f)-\mu(f)]=0.$$

Comme $t\mapsto\mu P_t f$ est continue à droite $\mu P_t f=\mu(f)$ pour tout $t\geqq 0$, et μ est invariante par P_t. En remarquant que $\frac{\eta_t}{t}\in\Delta_t$ et $\frac{\tilde{\eta}_t}{t}\in\Delta$, un raisonnement analogue à celui fait pour les chaînes de Markov permet d'énoncer la proposition suivante prouvée dans [12].

Proposition. *On suppose l'ensemble* $\left\{\frac{\eta_t}{t}(\omega)\,(\omega,t)\in\Omega\times\mathbb{R}^+\right\}$ *relativement compact, q continue et* Π *fellérienne. Alors, pour tout ensemble fermé* Φ *de* $\mathscr{M}^2(E\times E)$ *tel qu'il n'existe pas de probabilité* μ*, pour laquelle* $(q\mu)$ *est invariante par* Π *et* $((q\mu)\times\Pi,\ (q\mu)\times\Pi)\in\Phi$ *on a:*

$$\limsup_{t\to+\infty}\frac{1}{t}\operatorname{Log}\sup_{x\in E}\mathbb{P}_x\left[\left(\frac{\eta_t}{t},\frac{\tilde{\eta}_t}{t}\right)\in\Phi\right]<0.$$

V. Applications statistiques

Ce paragraphe donne différents types d'applications statistiques des résultats antérieurs. Les démonstrations figurent dans [12] et [14].

V.1. *Cas des variables aléatoires indépendantes*

1) Soit $X=(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans un espace polonais $(E,\mathscr{E})$, indépendantes et de loi μ. Le problème est d'obtenir des majorations de «distances» entre μ et la loi empirique $\bar{\mu}_n=\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{X_i}$. On a le théorème suivant:

Théorème (*cf.* [14]). *Soit T une fonction de $\mathscr{P}(E)$ dans $\mathbb{R}$ continue en μ, alors, pour tout $a>0$,*

$$\limsup_{n\to\infty}\frac{1}{n}\operatorname{Log}P[|T(\mu)-T(\bar{\mu}_n)|>a]<0.$$

2) *Exemples.* On prend $(E,\mathscr{E})=(\mathbb{R},\mathscr{B}_{\mathbb{R}})$ et on confond une probabilité μ et sa fonction de répartition F. Pour $0<p<1$

$$T_p(F)=\tfrac{1}{2}(\sup\{x;F(x)\leqq p\}+\inf\{x;F(x)\geqq p\})$$

$T_p(F)$ est le quantile d'ordre p de F.

Si $\{x;\mathrm{F(x)=p}\}$ est vide ou réduit à un point, T_p est continue en F et le théorème s'applique.

V.2. *Processus de sauts markoviens*

Le but de ce paragraphe est d'établir un théorème de consistance exponentielle et d'étudier un test séquentiel pour des processus de sauts markoviens.

a) *Cadre statistique.* On donne un espace métrique Θ (espace des paramètres) un espace polonais E muni de sa tribu borélienne $\mathscr{E}$ (espace des états) et une famille de transitions (indexée par Θ), $(\Pi(\theta;\cdot,\cdot))_{\theta\in\Theta}$ de $(E,\mathscr{E})$ dans $(E,\mathscr{E})$. On donne enfin une famille de fonctions mesurables strictement positives de $(E,\mathscr{E})$ dans $(\mathbb{R}_+,\mathscr{B}_{\mathbb{R}_+})$, $(q(\theta,\cdot))_{\theta\in\Theta}$. On suppose qu'il existe une transition Π de $(E,\mathscr{E})$ dans $(E,\mathscr{E})$ telle que, pour tout $(\theta,x)\in\Theta\times E$, la mesure $\Pi(\theta;x,\cdot)$ soit absolument continue par rapport à la mesure $\Pi(x,\cdot)$ avec:

$$\frac{d\Pi(\theta;x,\cdot)}{d\Pi(x,\cdot)}=p(\theta,x,\cdot).$$

Sur l'espace canonique $(\Omega,\mathscr{F})=(\mathbb{R}_+\times E,\mathscr{B}_{\mathbb{R}^+}\times\mathscr{E})^{\mathbb{N}}$, on note (T_n,X_n) la $(n+1)^{\text{ème}}$ coordonnée,

$$\mathscr{G}_n=\sigma((T_p,X_p);\,p\leqq n),\qquad Y_t=\sum_{n\geqq 0}X_n\mathbb{1}_{\{T_n\leqq t<T_{n+1}\}},$$

$$\mathscr{F}_t=\sigma(Y_s;\,s\leqq t)\quad\text{et}\quad\mathbb{F}=(\mathscr{F}_t)_{t\geqq 0}.$$

Pour $x\in E$ et $\theta\in\Theta$ on définit les probabilités $\mathbb{P}_x$ et $\mathbb{P}_{x,\theta}$ sur $(\Omega,\mathscr{F})$ par

$$\mathbb{P}_{x,\theta}(X_0=x,\ T_0=0)=\mathbb{P}_x(X_0=x,\ T_0=0)=1,$$

$$\mathbb{P}_{x,\theta}[T_{n+1}-T_n\geqq t,\ X_{n+1}\in dx\,|\,\mathscr{G}_n]=[\exp(-q(\theta,X_n)t)]\,\Pi(\theta;X_n,dx),$$
$$\mathbb{P}_x[T_{n+1}-T_n\geqq t,\ X_{n+1}\in dx\,|\,\mathscr{G}_n]=[\exp(-t)]\,\Pi(X_n,dx).$$

On suppose que $T_n\to\infty$, $\mathbb{P}_{x,\theta}$ et $\mathbb{P}_x$ p.s. lorsque $n\to\infty$. Sur $(\Omega,\mathscr{F},\mathbb{P}_{x,\theta})$ pourvu de la filtration $\mathbb{F}$ la mesure aléatoire

$$\eta_t=\sum_{n\geqq 1}\varepsilon_{X_{n-1},X_n}\cdot\mathbb{1}_{(T_n\leqq t)}$$

a pour compensatrice

$$\tilde{\eta}^\theta_t(dx,dy)=\int_0^t q(\theta,x)\,\Pi(\theta;x,dy)\,\varepsilon_{Y_s}(dx)\,ds$$

et, sur $(\Omega,\mathscr{F},\mathbb{P}_x)$, la compensatrice de

$$\eta_t\quad\text{est}\quad\tilde{\eta}_t(dx,dy)=\int_0^t \Pi(x,dy)\,\varepsilon_{Y_s}(dx)\,ds.$$

Posant $r(\theta,x,y)=q(\theta,x)\,p(\theta,x,y)$, on suppose que

$$\sup_x\int(1-\sqrt{r(\theta,x,y)})^2\,\Pi(x,dy)\qquad\text{est fini.}$$

Alors ([6], [8]), pour tout $t>0$, la restriction de $\mathbb{P}_{\theta,x}$ à $(\Omega,\mathscr{F}_t)$ est absolument continue par rapport à la restriction de $\mathbb{P}_x$ à $(\Omega,\mathscr{F}_t)$ et

$$\begin{aligned}\operatorname{Log}E\left(\frac{d\mathbb{P}_{\theta,x}}{d\mathbb{P}_x}\middle|\mathscr{F}_t\right)&=V_\theta(t)\\&=\int_{E^2}\operatorname{Log}r(\theta,x,y)\,\eta_t(dx,dy)+\int_{E^2}(1-r(\theta,x,y))\,\tilde{\eta}_t(dt,dy).\end{aligned}$$

L'estimateur du maximum de vraisemblance $\hat{\theta}_t$ de θ à l'instant t satisfait (s'il existe), $V_{\hat{\theta}_t}(t)=\sup_\theta V_\theta(t)$.

b) **Théorème** (cf. [12]). *Consistance exponentielle de l'estimateur du maximum de vraisemblance.*

Hypothèses. 1) *E et Θ sont compacts.*

2) *Les fonctions $(\theta,x,y)\mapsto p(\theta,x,y)$ et $(\theta,x)\to q(\theta,x)$ sont continues.*

3) *Pour tout $\theta\in\Theta$, la transition $\Pi(\theta;\cdot,\cdot)$ est fellérienne et, pour toute mesure de probabilité μ_θ invariante par $\Pi(\theta;\cdot,\cdot)$, on a, pour tout $\Phi\neq\theta$*

$$[(q(\theta,\cdot)\,\mu_\theta)\times\Pi(\theta,\cdot)]\,\{(x,y);r(\theta,x,y)\neq r(\Phi;x,y)\}>0.$$

Alors, pour tout voisinage U_θ de θ:

$$\limsup_{t\to\infty}\frac{1}{t}\operatorname{Log}\sup_{x\in E}\mathbb{P}_{x,\theta}[\hat{\theta}_t\in U^c_\theta]<0.$$

c) *Test séquentiel.* On veut étudier le test séquentiel du maximum de vraisemblance pour tester l'hypothèse H_0: «$\theta=\theta_0$» contre l'hypothèse H_1: «$\theta=\theta_1$». La procédure est la suivante. Soit $b>0$; posons

$$\xi_b=\inf\{t>0;\ V_{\theta_0}(t)-V_{\theta_1}(t)\notin]-b,+b[\}.$$

Le temps d'arrêt ξ_b est presque sûrement fini (sous les deux hypothèses). On arrête l'expérience à l'instant ξ_b et

$$\begin{aligned}&\text{si } V_{\theta_0}(\xi_b)-V_{\theta_1}(\xi_b)\geqq b, &&\text{on décide } \theta=\theta_0\\ &\text{si } V_{\theta_0}(\xi_b)-V_{\theta_1}(\xi_b)\geqq -b, &&\text{on décide } \theta=\theta_1.\end{aligned}$$

Les deux principaux critères de «qualité» d'un tel test sont d'une part un critère «économique» (grandeur de ξ_b), d'autre part le critère classique d'estimation des probabilités d'erreur.

Théorème (cf. [12]). *On fait les hypothèses suivantes:*

1) *E est métrique compact.*

2) *Les fonctions $x\mapsto q(\theta_i,x)$ et $(x,y)\mapsto p(\theta_i;x,y)$ sont continues pour $i=0$ et $i=1$. De plus, pour tout x:*

$$q(\theta_0,x)\,\Pi(\theta_0;x,\cdot)\neq q(\theta_1,x)\,\Pi(\theta_1);x,\cdot).$$

3) *$x\mapsto\Pi(\theta_i;x,\cdot)$ est continue de E dans $\mathscr{P}(E)$ pour $i=0$ et $i=1$. Alors, pour $i=0$ ou $i=1$*

a) $\limsup\limits_{t\to\infty}\dfrac{1}{t}\operatorname{Log}\sup\limits_{x}\mathbb{P}_{\theta_i,x}(\xi_b>t)<0$

(décroissance exponentielle de la fonction de répartition de ξ_b).

b) *On peut trouver deux constantes $A>0$ et $h>0$ telles que, pour $x\in E$,*

$$\mathbb{P}_{\theta_0,x}[V_{\theta_0}(\xi_b)-V_{\theta_1}(\xi_b)\leqq -b]=\mathbb{P}_{\theta_0,x}[\textit{choisir } \theta_1]\leqq A\,e^{-hb};$$
$$\mathbb{P}_{\theta_1,x}[V_{\theta_0}(\xi_b)-V_{\theta_1}(\xi_b)\geqq b]=\mathbb{P}_{\theta_1,x}[\textit{choisir } \theta_0]\leqq A\,e^{-hb}.$$

V.3. *Mesures aléatoires de Poisson*

Cadre statistique. On observe une mèsure aléatoire $(\eta(\cdot,\cdot))$ de Poisson et on veut estimer son intensité. On donne un espace métrique Θ (espace des paramètres) et une famille $\{\lambda_\theta\}_{\theta\in\Theta}$ de mesures de Radon positives sur $\mathbb{R}^d$. On note $\mathscr{M}_p(\mathbb{R}^d)$ l'ensemble des mesures ponctuelles sur $\mathbb{R}^d$ et, pour tout $\theta\in\Theta$, on note $\mathbb{P}_\theta$ la loi sur $(\mathscr{M}_p(\mathbb{R}^d),\mathscr{B}(\mathscr{M}_p(\mathbb{R}^d)))$ de la mesure de Poisson d'intensité λ_θ. On suppose que, pour tout θ, λ_θ est absolument continue par rapport à une mesure de Radon λ positive, avec $\dfrac{d\lambda_\theta}{d\lambda}=f(\cdot,\theta)$. On note alors $\mathbb{P}$ la loi de la mesure de Poisson d'intensité λ, $\mathbb{P}_\theta^{(r)}$ et $\mathbb{P}^{(r)}$ désignant les restrictions de $\mathbb{P}_\theta$ et $\mathbb{P}$ à $\mathscr{M}_p(B_r)$, où $B_r=\{x/\|x\|\leqq r\}$.

Alors

$$\frac{d\mathbb{P}_\theta^{(r)}}{d\mathbb{P}^{(r)}}(\mu)=\exp\Big[\int_{B_r}\operatorname{Log}f(x,\theta)\,\mu(dx)+\int_{B_r}(1-f(x,\theta))\,\lambda(dx)\Big].$$

On note

$$L_r(\theta)=\int_{B_r} \operatorname{Log} f(x,\theta)\,\eta(\omega,dx)+\int_{B_r}(1-f(x,\theta))\,\lambda(dx).$$

Un estimateur du maximum de vraisemblance $\hat{\theta}_r$, basé sur les observations qui tombent dans B_r, satisfait $L_r(\hat{\theta}_r)=\sup\limits_{\theta\in\Theta} L_r(\theta)$.

Théorème (cf. [14]). *Consistance exponentielle de $\hat{\theta}_r$ si $r\to\infty$.*

Hypothèses. 1) *L'application $(y,\varphi)\mapsto f(y,\varphi)$ est continue pour $\mathbb{R}^d\times\Theta$.*

2) *Θ est métrique compact.*

3) *Pour tout $\varphi\in\Theta$, l'application $y\mapsto f(y,\varphi)$ est p-homogène.*

4) *Pour tout $\varphi\neq\theta$, $\lambda[f(\cdot,\theta)\neq f(\cdot,\varphi)]>0$.*

Alors, pour tout voisinage U_θ de θ:

$$\limsup \frac{1}{\lambda(B_r)} \operatorname{Log} \mathbb{P}[\hat{\theta}_r\notin U_\theta]<0.$$

References

1. Aalen, O.: Non parametric inference for a family of counting processes. Ann. Statist. **6**, 701–726 (1978)
2. Chernoff, H.: Sequential design of experiments. Ann. Math. Statist. **30**, 755–770 (1959)
3. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time. Comm. Pure Appl. Math. 28, 1–47 (1975)
4. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time. Comm. Pure Appl. Math. **28**, 279–301 (1975)
5. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time. Comm. Pure Appl. Math. **29**, 389–461 (1976)
6. Duflo, M., Florens, D.: Décision statistique – Pas à Pas. Les cours du CIMPA (1981)
7. Jacod, J.: Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. Z. Wahrscheinlichkeitstheorie verw. Gebiete **31**, 235–253 (1975)
8. Liptser, R.S., Shiryayev, A.N.: Statistics of random processes. New-York-Heidelberg-Berlin: Springer 1977
9. Maigret, N.: Théorème de grandes déviations pour les chaînes de Markov contrôlées. Z. Wahrscheinlichkeitstheorie verw. Gebiete **51**, 133–151 (1980)
10. Maigret, N.: Majoration de Chernoff et statistique séquentielle pour des chaînes de Markov récurrentes au sens de Doeblin. Astérisque **68**, 125–142 (1979)
11. Maigret, N.: Statistiques des chaînes contrôlées fellériennes. Astérisque **68**, 143–169 (1979)
12. Portal, F.: Thèse de 3ème cycle, Université Paris-Sud (1980)
13. Sethuraman, J.: On the probability of lange deviations of families of sample means. Ann. Math. Statist. **35**, 1304–1316 (1964)
14. Touati, A.: Thèse de 3ème cycle, Université Paris-Nord (1980)
15. Tuominen, P., Tweedie, R.L.: Markov chains with continuous components, Proc. London Math. Soc. **38**, 89–114 (1979)

Reçu le 25 Mars 1982; en forme révisée le 29 Septembre 1983