JEAN BRETAGNOLLE

Formule de Chernoff pour les lois empiriques de variables à valeurs dans des espaces généraux

Astérisque, tome 68 (1979), p. 33-52

<http://www.numdam.org/item?id=AST_1979__68__33_0>

Formule de Chernoff pour les lois empiriques de variables à valeurs dans des espaces généraux

Jean BRETAGNOLLE.

Introduction

On se propose dans cet exposé d'établir des évaluations asymptotiques de la probabilité que la mesure empirique associée à un n-échantillon appartienne à un sous-ensemble Λ de l'ensemble des lois de probabilité sur un espace $\{\mathcal{X}, \mathcal{B}\}$. Dans les cas raisonnables, cette évaluation se fait facilement pour un ensemble ouvert, ou fermé pour la topologie de la convergence étroire, c'est-à-dire celle de la dualité avec les fonctions continues bornées ; on pourra jouer dans certaines limites sur la topologie de $\mathcal{X}$ sans changer la validité du résultat (voir Lemme 1.1. et les exemples du § 6), et l'objet du premier chapitre est de décrire pour quels espaces $\{\mathcal{X}, \mathcal{B}\}$ on aura cette possibilité. On aurait pu se contenter du cas Lusinien, mais le cas Polonais est plus clair pour le lecteur non habitué. La démonstration des résultats principaux suit Donsker et Varadhan [1], l'amélioration Lusinienne est dûe à P. Assouad.

§ 1. Rappels de Théorie de la mesure

Dans la suite, $\{\mathcal{X}, \mathcal{B}\}$ désigne un espace mesurable et si on le munit d'une topologie $\mathcal{T}$, celle-ci a $\mathcal{B}$ pour σ-algèbre Borélienne : $\mathcal{B} = \sigma(\mathcal{T})$.

Les deux cas traités sont les suivants :

Le cas Polonais : $\mathcal{X}$ est métrisable, séparable, complet ; il est homéomorphe à un G_δ d'un compact métrisable.

Le cas Lusinien : $\{\mathcal{X}, \mathcal{B}\}$ est isomorphe à un Borélien d'un compact métrisable (muni de la σ-algèbre trace, et la (une) topologie compatible est la topologie trace associée, donc métrisable séparable).

Dans les deux cas, la σ-algèbre de Baire (engendrée par les fonctions continues) coïncide avec la Borélienne.

Les mesures positives bornées sont les fonctions d'ensemble σ-additives, positives et finies, et, dans les deux cas, elles possèdent la propriété de régularité intérieure :

$$\forall \quad B \in \mathcal{B}, \quad m(B) = \operatorname{Sup}\{m(K) \mid K \text{ compact contenu dans } B\} .$$

Ce sont également les formes linéaires I positives sur $\mathcal{C}_b(\mathcal{X})$, espace des fonctions numériques continues bornées, telles que :

$$\text{si} \quad f_n \in \mathcal{C}_b(\mathcal{X}), \quad f_n \downarrow 0 , \quad \text{alors} \quad I(f_n) \longrightarrow 0 .$$

La (une) topologie sur l'espace des mesures est celle de la convergence étroite, c'est-à-dire celle de la convergence simple sur les fonctions continues bornées.

$P(\mathcal{X})$ désigne l'ensemble des probabilités sur $\{\mathcal{X}, \mathcal{B}\}$, dont on note $\lambda, \mu, \nu \dots$ les éléments. C'est un fermé. On a :

(1.1) Pour qu'une partie Λ de $P(\mathcal{X})$ soit relativement compacte, il suffit qu'elle soit équitendue, c'est-à-dire, qu'à tout $\varepsilon > 0$, on puisse associer

un compact K_ε de $\mathcal{X}$ tel que $\forall\ \lambda \in \Lambda$, $\lambda(K_\varepsilon^c) < \varepsilon$.
(Cette condition est également nécessaire dans le cas Polonais).

Soit alors $\emptyset$ une fonction de $\mathbb{R}^+$ dans $\mathbb{R}^+$, telle que $\emptyset(x)/x$ tende vers $+\infty$ quand x tend vers $+\infty$. Soit Λ une famille de probabilités absolument continues par rapport à une même μ . Alors :

(1.2) si $\sup_{\lambda\in\Lambda} E_\mu\left(\emptyset\left(\frac{d\lambda}{d\mu}\right)\right) < \infty$, Λ est relativement compacte .

(On emploiera systématiquement la notation $E_\lambda(f)$ pour $\int f\,d\lambda$ quand $\lambda \in P(\mathcal{X})$).

On utilisera le théorème de Hahn-Banach sous la forme suivante :

Soit $\mathcal{K}$ un compact convexe de $P(\mathcal{X})$, disjoint de $\mathcal{F}$, fermé convexe de $P(\mathcal{X})$.

Il existe alors une f dans $\mathcal{C}_b(\mathcal{X})$, telle que :

$$\sup_{\lambda\in\mathcal{K}} E_\lambda(f) < \inf_{\lambda\in\mathcal{F}} E_\lambda(f) .$$

Notons enfin un résultat :

<u>Lemme 1.1.</u> : Soit $\{\mathcal{X}, \mathcal{T}, \mathcal{B}\}$ un espace Lusinien, f_n une suite de fonctions mesurables bornées de $\{\mathcal{X}, \mathcal{B}\}$ dans $\{\mathbb{R}, \mathcal{B}_{\mathbb{R}}\}$. On peut trouver une topologie $\mathcal{T}'$ sur $\mathcal{X}$ engendrant la <u>même</u> σ-algèbre Borélienne, telle que $\{\mathcal{X}, \mathcal{T}', \mathcal{B}\}$ soit Lusinien, et que les f_n soient continues pour $\mathcal{T}'$.

Ce résultat permet donc de changer de topologie sur $\mathcal{X}$, donc de topologie étroite sur $P(\mathcal{X})$.

<u>Références</u> : On peut se reporter aux livres de Neveu, Meyer et Dellacherie, de Bourbaki (ch. 9 Topologie).

§ 2. Information de Kullback

Soit U l'ensemble des fonctions continues, positives, bornées et d'inverse borné de $\mathcal{X}$ dans $\mathbb{R}$. Si λ et μ sont deux probabilités, l'information de Kullback de λ par rapport à μ est définie par :

$$(2.1) \qquad I(\lambda,\mu) = \operatorname*{Sup}_{u\in U} \{E_\lambda(\operatorname{Log} u) - \operatorname{Log}(E_\mu(u))\} .$$

Elle est <u>positive</u>, finie si et seulement si λ est absolument continue par rapport à μ <u>et</u> $E_\lambda(\frac{d\lambda}{d\mu}))$ est fini, et vaut alors :

$$(2.2) \qquad I(\lambda,\mu) = E_\lambda(\operatorname{Log}(\frac{d\lambda}{d\mu})) = E_\mu(\frac{d\lambda}{d\mu} \cdot \operatorname{Log}(\frac{d\lambda}{d\mu})) .$$

(2.3) Elle est convexe et semi-continue inférieurement de son argument λ.

(2.4) Si g_o est mesurable de $\mathcal{X}$ dans $\mathcal{Y}$; $I(g_o\lambda, g_o\mu) \leqslant I(\lambda,\mu)$

(2.5) Pour tout a (> 0), toute μ de $P(\mathcal{X})$, $\mathcal{K}_a = \{\lambda \mid I(\lambda,\mu) \leqslant a\}$ est un convexe compact de $P(\mathcal{X})$.

<u>Démonstration</u>

Posons, à μ fixé, $V(\lambda,u) = E_\lambda(\operatorname{Log} u) - \operatorname{Log}(E_\mu u)$, I_1 la première forme : $I_1(\lambda) = \operatorname*{Sup}_U V(\lambda,u)$, $I_2(\lambda)$ la seconde, définie le cas échant par (2.2). V étant linéaire et continue de λ pour tout u de U, I_1 satisfait (2.3). Sa positivité provient de ce que $1 \in U$.

Si $\lambda \not\prec \mu$, soit K un compact tel que $\lambda(K) > 0$, $\mu(K) = 0$. On peut construire pour tout M une suite u_n de U, $u_n \downarrow M \cdot 1_K + 1$, et alors $\lim_n V(\lambda,u_n) \geqslant \lambda(K)\operatorname{Log}(M+1))$, de sorte qu'alors $I_1(\lambda)$ est infinie.

Montrons maintenant que $I_1(\lambda) = I_2(\lambda)$ si $\lambda \prec \mu$: On construit pour tout couple $0 < \varepsilon < M$ une suite u_n de U, comprise entre ε et M, telle que u_n converge en μ-Probabilité vers $p_{\varepsilon,M} = \varepsilon \vee \frac{d\lambda}{d\mu} \wedge M$. On a alors

$$I_1(\lambda) \geqslant \lim V(\lambda,u_n) = E_\lambda(\operatorname{Log} p_{\varepsilon,M}) - \operatorname{Log}(E_\mu(p_{\varepsilon,M})) .$$

Passant à la limite en ε et M, il vient $I_1(\lambda) \geq I_2(\lambda)$.

Pour démontrer que $I_2 \geq I_1$, supposons un instant que $\frac{d\lambda}{d\mu} \in U$; si on lie alors u et v par $u = v \cdot \frac{d\lambda}{d\mu}$, u et v appartiennent en même temps à U, donc :

$$V(\lambda,u) = E_\lambda(\mathrm{Log}(\frac{d\lambda}{d\mu})) + E_\lambda(\mathrm{Log}\ v) - \mathrm{Log}(E_\lambda v) \leq I_2(\lambda) \quad ,$$

d'après l'inégalité de Jensen appliqué à la fonction Log. Donc $I_1 \leq I_2$.
Sinon, par une procédure de troncage similaire à la précédente, on construit une suite u_n de U telle que $E_\mu(u_n \mathrm{Log}\ u_n)$ tende vers $I_2(\lambda)$, u_n tende vers $\frac{d\lambda}{d\mu}$ dans $L^1(\mu)$, et $E_\mu(u_n) = 1$, cette dernière condition par multiplication par une constante. Si λ_n est définie par $d\lambda_n = u_n \cdot d\mu$, comme λ_n tend vers λ, d'après (2.3), $I_1(\lambda) \leq \liminf I_1(\lambda_n)$; mais $I_1(\lambda_n) = I_2(\lambda_n)$ qui tend vers $I_2(\lambda)$.

(2.4) provient de la seconde forme (2.2), en remarquant que si $\mathfrak{A}$ est la σ-algèbre de $\mathfrak{X}$ engendrée par g_o, la probabilité de densité $E_\mu^{\mathfrak{A}}(\frac{d\lambda}{d\mu})$ a une information égale à $I(g_o\lambda, g_o\mu)$ (d'après le Théorème de transport) et moindre que $I(\lambda,\mu)$ d'après Jensen appliquée à $x \longrightarrow x\ \mathrm{Log}\ x$.

(2.5) : $\mathcal{H}_a$ est convexe fermé d'après (2.3), relativement compact d'après la remarque (1.2) appliquée à $\emptyset(x) = x\ \mathrm{Log}\ x$.

§ 3. Transformée de Cramer

Dans ce chapitre , $\mathcal{X}$ est un espace de Banach séparable, de dual $\mathcal{X}^*$, la forme bilinéaire est notée (θ,a) $(\theta \in \mathcal{X}^*$, $a \in \mathcal{X}$). On s'intéresse particulièrement aux probabilités μ satisfaisant (X désignant l'élément générique de $\mathcal{X}$)

(3.1) Pour tout réel t, $E_\mu(\exp(t||X||))$ est fini .

La transformée de Cramer de la probabilité μ est alors définie par :

(3.2) $$h_\mu(a) = \operatorname*{Sup}_\theta\{(\theta,a) - \mathrm{Log}\, E_\mu \exp(\theta,X)\} .$$

C'est une fonction positive et convexe , en tant qu'enveloppe supérieure de fonctions affines ; remarquons que $(\theta,a) - \mathrm{Log}\, E \exp(\theta,X)$ est concave de son argument θ . Notons que :

(3.3) $$h_{\mu^{\otimes n}}(na) = n \cdot h_\mu(a)$$ et que

(3.4) $$h_{\mu_{+b}}(a) = h_\mu(a-b) ,$$ si μ_{+b} est la translatée par b de μ.

- Première étude du cas unidimensionnel ($\mathcal{X} = \mathbb{R}$)

Proposition 3.1. (Cramer-Chernoff), (voir aussi chapitre I)

Soit S_n la somme de n v.a.r. indépendantes de même loi μ satisfaisant (3.1.) . Alors :

$$\frac{1}{n} \mathrm{Log}\{P_\mu(S_n/n \geq a)\} \leq - \operatorname*{Inf}_{b\geq a} h_\mu(b) .$$

Démonstration

C'est tout simplement l'optimisation de l'inégalité de Markov sur les moments exponentiels. D'après les remarques précédentes, on peut se ramener au cas $n=1$, $a = 0$. Au lieu de θ , écrivons t (réel) : $h(b) = \operatorname*{Sup}_{t\in\mathbb{R}} bt - \mathrm{Log}\, E\, e^{tX}$.

$bt - \mathrm{Log}\, E\, (e^{tX})$ étant concave de t , de dérivée en 0 : $b - E(X)$, on remarque que $h(E(X)) = 0$, donc que $\operatorname*{Inf}_{b\geq 0} h(b) = 0$ si $0 \leq E(X)$, il n'y a rien à démontrer alors. Sinon, pour tout $b \geq 0$, le sup définissant h vaut le sup pour t dans $\mathbb{R}^+$, et son minimum en b est clairement obtenu pour $b = 0$. L'inégalité classique de Markov est $P(X \geq 0) \leq E(e^{tX})$ $(t \geq 0)$, donc $P(X \geq 0) < \inf_{t \geq 0} E(e^{tX})$, c'est-à-dire le résultat.

Remarque

Notons que dans le cas réel, $\lim_{|a| \to \infty} \frac{h(a)}{|a|} = +\infty$, et que h est croissante à l'infini du côté positif : en effet, h est croissante à droite de $E(X)$, et (pour tout t) $\limsup_a \frac{h(a)}{|a|} \geq |t|$.

Revenons maintenant au cas général Banachique :

Proposition 3.2.

Soit μ satisfaisant (3.1). Il existe alors une fonction :

$$\emptyset \ : \ \mathbb{R}^+ \longrightarrow \mathbb{R}^+ \text{, convexe , avec } \lim_{x \to \infty} \emptyset(x)/x = +\infty \text{ ,}$$

telle que :

(3.5) Pour tout t, $E_\mu(\exp(t\,\emptyset\,(||X||)))$ est fini.

Démonstration

Remplaçons μ par la loi de $Z = ||X||$, ce qui revient à la supposer portée par $\mathbb{R}^+$: Si on pose $H(z) = P(Z \geq z)$, on a : (Prop 1) $h(z) \leq \exp(-g(z))$, où $g = h \vee 0$, h transformée de Cramer de μ. Posons $\psi(z) = \sqrt{z \cdot g(z)}$; $\psi(z)/z \to \infty$ et $\psi(z)/g(z) \longrightarrow 0$ quand $z \longrightarrow \infty$, d'après la remarque plus haut. Soit alors $\emptyset$ la plus grande fonction convexe sous ψ (i.e. le sup des fonctions affines moindres de ψ) ; $\emptyset$ possède encore ces deux propriétés. Maintenant,

$$E_\mu \exp(t\emptyset(||X||)) < t \int \exp(t\emptyset(z) - g(z))\, \emptyset'(z)\, dz \ ,$$

dont l'intégrant est majoré à l'infini par $t \cdot \exp(-\emptyset(z)) \cdot \emptyset'(z)$ (puisque $\emptyset/g$ tend vers 0) et l'intégrale est donc convergente.

Proposition 3.3.

Soit μ satisfaisant (3.1), $\emptyset$ une fonction satisfaisant (3.5),

$\mathcal{K}_a$ le convexe compact $\{\lambda \mid I(\lambda,\mu) \leqslant a\}$ de (2.5).

Alors $\sup\limits_{\lambda \in \mathcal{K}_a} E_\lambda(\emptyset(||X||))$ est fini.

Démonstration

Posons $T = \emptyset(||X||)$, $Y = \frac{d\lambda}{d\mu}$, $E = E_\mu$. Sur les T, Y positives, $E(Y) = 1$, nous avons à majorer $E(TY)$ en fonction de $E(e^T)$ et $a = E(Y \text{ Log } Y)$. Les fonctions $f(x) = e^x - 1 - x$ et $g(y) = (1+y) \text{ Log } (1+y) - y$ sont en dualité de Young, ce qui veut dire que f' et g' sont réciproques, et entraîne que $xy < f(x) + g(y)$, soit $E(TY) \leqslant E(e^T - 1 - T) + E((1+Y)\text{Log}(1+Y) - Y) \leqslant E(\exp T) + 1 + a$.

Corollaire

Soit μ satisfaisant (3.1), a réel, θ dans $\mathcal{X}^*$. L'ensemble

$$\mathcal{K}_{a,\theta} = \{\lambda \mid \lambda \in \mathcal{K}_a \; ; \; E_\lambda(\theta, X) = 0\}$$

est convexe compact. Il en est de même pour $\mathcal{K}_{a,T} = \bigcap\limits_{\theta \in T} \mathcal{K}_{a,\theta}$, pour toute partie T de $\mathcal{X}^*$.

Démonstration

Il suffit de montrer que $\mathcal{K}_{a,\theta}$ est fermé, soit, que pour toute λ de $\mathcal{K}_a \setminus \mathcal{K}_{a,\theta}$, il existe une fonction de $\mathcal{C}_b(\mathcal{X})$ séparant λ de $\mathcal{K}_{a,\theta}$. D'après la Prop. 3, une telle λ intègre $\emptyset(||X||)$, donc $||X||$, donc (θ, X). Si $f_M(X) = \{\text{signe de } (\theta,X)\} \cdot \{|(\theta,X)| \wedge M\}$, f_M est dans $\mathcal{C}_b$, et, sur $\mathcal{K}_a$

$$E_\lambda |f_M(X) - E_\lambda(\theta,X)| \leqslant \frac{M}{||\theta||} \cdot \{\emptyset(\frac{M}{||\theta||})\}^{-1} \cdot \sup_{\lambda \in \mathcal{K}_a} E_\lambda(\emptyset(||X||)) \quad ;$$

Uniformément sur $\mathcal{K}_a$, le membre de gauche tend donc vers 0 quand M tend vers l'infini, puisque $\emptyset(x)/x$ tend vers 0 à l'infini. Pour M assez grand, f_M répond donc à la question.

Si maintenant les X_i sont des v.a. à valeurs dans $\mathcal{X}$, (indépendantes et de même loi), rappelons qu'on appelle <u>mesure empirique associée au n-échantillon</u>

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \quad .$$

<u>Proposition 3.4.</u>

Soit μ satisfaisant (3.1) , $\emptyset$ une fonction satisfaisant (3.5).

Pour tout réel $a > 0$, il existe un compact $\mathcal{K}(a)$ et un réel M tels que :

(3.6) $$P_\mu(\hat{\mu}_n \notin \mathcal{K}(a)) \leq 2\cdot a^n \ , \quad \text{et} \quad \operatorname*{Sup}_{\lambda \varepsilon \mathcal{K}(a)} E_\lambda(\emptyset(||X||)) \leq M \quad .$$

<u>Démonstration</u> Soit $A_M = \{\lambda \mid E_\lambda(\emptyset(||X||)) > M\}$;

$$\{\hat{\mu}_n \ \varepsilon \ A_M\} = \{\sum_{i<n} \emptyset(||X_i||) > nM\} \ ,$$

et, comme la v.a. <u>réelle</u> , $\emptyset(||X||)$ a une transformée de Laplace définie partout, on peut appliquer la majoration de la Prop. 1, soit :

$$P_\mu(\hat{\mu}_n \ \varepsilon \ A_M) \leq a^n \quad , \qquad \text{pour } M \text{ bien choisi} \quad .$$

On va maintenant construire un ensemble Λ_a tel que :

(3.7) $$\Lambda_a \text{ est relativement compact , et } P_\mu(\hat{\mu}_n \notin \Lambda_a) \leq a^n \ .$$

La démonstration sera alors terminée en posant $\mathcal{K}(a) = \overline{A_M^c \cap \Lambda_a}$.

A b, d réels tels que $0 < b < d < 1$, associons un compact K de $\mathcal{X}$ tel que $\mu(K^c) \leq b$. Si on pose $Y_i = 1_{X_i \varepsilon K^c}$, $\hat{\mu}_n(K^c)$ s'interprète comme une variable binomiale (la somme des Y_i) de paramètre p moindre que b. La transformée de Cramer des Y est minorée par : $d \operatorname{Log}(\frac{d}{b}) + (1-d) \operatorname{Log}(\frac{1-d}{1-b})$, soit :

$$P_\mu(\hat{\mu}_n(K^c) \geq d) \leq \alpha^n \quad , \quad \text{où} \quad \alpha = (\frac{b}{d})^d \cdot (\frac{1-b}{1-d})^{1-d} \quad ,$$

d'après la Prop. 1.

Choisissons d_m tendant vers 0, puis b_m tels que $\sum_m \alpha_m < a$, associons leur K_m , et posons $\Lambda_a = \{\lambda \mid \text{Pour tout } m \ , \ \lambda(K_m^c) \leq d_m\}$. Puisque d_m tend vers 0 , Λ_a est relativement compact (d'après (1.2)) et

$$P_\mu(\hat{\mu}_n \notin \Lambda_a) \leq \sum_m (\alpha_m)^n \leq (\sum_m \alpha_m)^n < a^n \quad .$$

§4. Calcul de la transformée de Cramer

Théorème 1

Soit $\mathcal{X}$ un Banach séparable, μ une probabilité sur $\mathcal{X}$ satisfaisant (3.1). Pour tout t, $E_\mu(\exp(t\|X\|))$ est fini. Alors :

$$(4.1) \qquad h_\mu(a) = \operatorname*{Inf}_{(\lambda, \lambda\in P(\mathcal{X}); E_\lambda(X)=a)} I(\lambda,\mu) \quad .$$

Démonstration

En vertu de (3.4), on peut supposer $a = 0$; appelons h le membre de gauche, I le membre de droite de (4.1).

A. Démontrons d'abord que $h \leqslant I$.

1. Si $\mathcal{X} = \mathbb{R}$, notons Y la dérivée de Radon-Nycodim de λ $(\lambda \prec \mu)$ par rapport à μ : $Y \geqslant 0$, $E_\mu(Y) = 1$. X désignent l'application identique $x \to x$, on doit montrer que pour tout t réel, tout Y comme plus haut, vérifiant de plus $E_\mu(XY) = 0$ l'inégalité :

$$- \operatorname{Log} E_\mu(\exp tX) \leqslant I(\lambda,\mu) \qquad \text{or}$$

$$- \operatorname{Log} E_\mu(\exp tX) \leqslant -\operatorname{Log} E_\lambda \exp(tX-\operatorname{Log} Y) \leqslant E_\lambda(\operatorname{Log} Y-tX) = I(\lambda,\mu)-tE_\mu XY = I(\lambda,\mu) \ .$$

(La seconde inégalité provenant de Jensen appliqué à $-$ Log).

2. Dans le cas général, soit $\theta \in \mathcal{X}^*$. Si $E_\lambda X = 0$, $E_\lambda(\theta,X) = 0$. Soit $g : X \longrightarrow (\theta,X)$ de $\mathcal{X}$ dans $\mathbb{R}$; le résultat, vrai pour les mesures images par g d'après A.1, entraîne : $-\operatorname{Log} E_\mu(\theta,X) \leqslant I(g_o\lambda, g_o\mu)$, pour toute θ de $\mathcal{X}^*$, toute λ centrée. On applique alors (2.4) : $I(g_o\lambda, g_o\mu) \leqslant I(\lambda,\mu)$.

B. Montrons maintenant que $h \geqslant I$. On peut tout d'abord supposer h fini.

1. Ramenons le cas général au cas de dimension finie : soit T une partie finie de $\mathcal{X}^*$, $S(T)$ le sous-espace linéaire engendré, $\varnothing_T$ l'application naturelle de $\mathcal{X}$ dans $S(T)^*$, identifié dans la suite à un espace $\mathbb{R}^m$, et posons $h_T = \operatorname*{Sup}_{\theta\in S(T)} - \operatorname{Log} E_\mu(\exp(\theta,X))$ ($\leqslant h$, donc finie).

Si le résultat est vrai en dimension finie, il existe une Probabilité Q^T sur $\mathbb{R}^m$, centrée, d'information h_T par rapport à $\emptyset_T \circ \mu$ (voir remarque finale). Soit λ^T la Probabilité sur $\mathcal{X}$ définie par $\frac{d\lambda^T}{d\mu}(X) = \frac{dQ^T}{d\emptyset_T \circ \mu}(\emptyset_T X)$. Les $\mathcal{K}_{h,T}$ du Corollaire de la proposition 3.3. forment donc une famille filtrante de compacts non vides, et toute mesure de leur intersection est <u>centrée</u> ; et d'information moindre que h.

2. Montrons le résultat dans $\mathbb{R}^m$. Si θ est une forme linéaire, notons H_θ^-, H_θ, H_θ^+ les images réciproques par θ de $\{x < 0\}$, $\{x = 0\}$, $\{x > 0\}$. Trois cas se séparent :

a. Cas <u>intérieur</u> : pour toute $\theta \neq 0$, $\mu(H_\theta^-)\cdot\mu(H_\theta^+) > 0$. Si on normalise par $||\theta|| = 1$, l'infimum en t de $E_\mu(\exp t(\theta,X))$ est atteint à distance finie, puisque, les deux demi-espaces étant chargés par μ, la transformée de Laplace tend vers $+\infty$ quand $t \to \pm\infty$. Le convexe $\{\theta | E_\mu \exp(t(\theta,X)) \leq 1\}$ est donc borné dans toutes les directions, donc compact, autrement dit l'inf est atteint, et il existe une θ_o telle que $h = -\mathrm{Log}\ E_\mu \exp(\theta_o,X)$. La Probabilité de densité $\exp(h + (\theta_o,X))$ (par rapport à μ) a une transformée de Laplace qui atteint son minimum en $\theta = 0$, elle est donc centrée. Son information vaut h et le résultat est donc démontré.

b. Il existe θ telle que $\mu(H_\infty^-) = 1$. Ce cas ne peut se produire si h est finie, car $\inf_\theta E_\mu \exp(\theta,X) \leq \lim_{t\to\infty} E_\mu \exp t(\theta_o,X) = 0$.

c. Il existe $\theta_o \neq 0$ telle que $p = \mu(H_{\theta_o}) > 0$ et $\mu(H_{\theta_o}^+) = 0$. En changeant de base, on peut supposer que $H_{\theta_o} = \{x_m = 0\}$, et on décompose μ en $\mu^- + p\cdot\gamma$, où μ^- est une sous-probabilité portée par $\{x_m < 0\}$, γ une Probabilité portée par $\{x_m = 0\}$. On désigne par X' la projection de X sur $\{x_m = 0\}$, identifié à $\mathbb{R}^{m-1}$, on décompose de même les formes linéaires, soit $(\theta,X) = (\theta',X') + tX_m$. Alors $E_\mu \exp(\theta,X) = E_{\mu^-} \exp((\theta',X')+tX_m)+p\cdot E_\gamma \exp(\theta',X')$.
Prenons d'abord l'infimum en t : il se réalise pour $t = +\infty$, et vaut $p\cdot E_\gamma \exp(\theta',X')$ donc $h = -\mathrm{Log}\ p + h_\gamma(0)$. Par ailleurs, toute mesure λ' portée par $\{x_m = 0\}$

et centrée en X' est centrée en X, et d'information $I(\lambda',\mu) = -\operatorname{Log} p + I(\lambda',\nu)$. On est donc ramené en dimension inférieure, de sorte que le dernier cas à traiter est le cas c. en dimension 1. Mais alors la dernière ν est ∂_o, les dernières λ' coïncident avec ν, et $h_{\partial_o}(0) = 0 = I(\lambda',\nu)$, de sorte que la résultat est démontré.

Remarque

La fonction $I(\lambda,\mu)$ étant s.c.i. de λ atteint son infimum I (sur les mesures centrées) dès qu'il est fini. En effet, l'ensemble des Probabilités centrées et d'information moindre que $2I$ est compact.

La mesure Q^T introduite en B.1. existe donc.

Corollaire de l'étude précédente : (Cramer-Chernoff "ouvert" en dimension finie)

Soit μ une Probabilité sur $\mathbb{R}$ satisfaisant (3.1), G un ouvert de $\mathbb{R}^m$

(4.2) Alors $\liminf_n \frac{1}{n} \operatorname{Log} P_\mu(S_n/n \in G) \geq - \operatorname{Inf}_{a\in G} h_\mu(a)$.

Démonstration

On peut trouver un point b centre d'une boule ouverte contenue dans G, "intérieur" au support de μ au sens B.2.a., et tel que $h_\mu(b)$ soit voisin de $\inf_{a\in G} h_\mu(a)$ si ce dernier est fini.

Il existe alors une forme linéaire θ réalisant le sup, soit :

$$h_\mu(b) = (\theta,b) - \operatorname{Log} E_\mu \exp(\theta,X) .$$

On note λ_b la Probabilité de densité $\exp(h_\mu(b) + (\theta,X-b))$. Pour r réel assez petit : $P_\mu(S_n/n \in G) \geq P_\mu(\|S_n/n - b\| \leq r)$, soit, changeant de loi

$$\geq E_{\lambda_b}(1_{\|S_n/n-b\|\leq r} \exp(-nh_\mu(b) - (\theta,S_n-nb))) .$$

La loi λ_b étant centrée en b, $P_{\lambda_b}(\|S_n/n-b\| \leq r)$ tend vers 1, et le terme exponentiel dans l'intégrale est supérieur à $\exp(-nh_\mu(b) - nr\|\theta\|)$, d'où le résultat.

§5. Les Théorèmes de Cramer-Chernoff

Théorème 2

Soit $\mathcal{X}$ un espace Lusinien ou Polonais, μ une Probabilité sur $\mathcal{X}$, $\mathcal{F}$ un fermé de $P(\mathcal{X})$, $\mathcal{G}$ un ouvert de $P(\mathcal{X})$. On a :

$$(5.1) \quad \limsup_n \frac{1}{n} \operatorname{Log} P_\mu(\hat{\mu}_n \in \mathcal{F}) \leq - \inf_{\lambda \in \mathcal{F}} I(\lambda,\mu) .$$

$$(5.2) \quad \liminf_n \frac{1}{n} \operatorname{Log} P_\mu(\hat{\mu}_n \in \mathcal{G}) \geq - \inf_{\lambda \in \mathcal{G}} I(\lambda,\mu) .$$

Théorème 3

Soit $\mathcal{X}$ un espace de Banach séparable, F un fermé, G un ouvert de $\mathcal{X}$, μ une Probabilité sur $\mathcal{X}$ satisfaisant :

(3.1) Pour tout t réel, $E \exp(t||X||)$ est fini. Alors :

$$(5.3) \quad \limsup_n \frac{1}{n} \operatorname{Log} P_\mu(\frac{S_n}{n} \in F) \leq - \inf_{a \in F} h_\mu(a)$$

$$(5.4) \quad \liminf_n \frac{1}{n} \operatorname{Log} P_\mu(\frac{S_n}{n} \in G) \geq \inf_{a \in G} h_\mu(a) .$$

Démonstration

On notera dans la suite $I(\mathcal{F})$, $I(\mathcal{G})$, $h(F)$, $h(G)$ les opposés des seconds membres de (5.1, 2, 3, 4).

Remarquons tout d'abord que pour une famille d'événements $A_{n,m}$, où $n \in \mathbb{N}$, mais m varie dans un ensemble <u>fini</u>, on a :

$$(5.5) \quad \limsup_n \frac{1}{n} \operatorname{Log} P(\bigcup_m A_{n,m}) \leq \operatorname{Max}_m \limsup_n \frac{1}{n} \operatorname{Log} P(A_{n,m}) .$$

1. Démonstration de (5.1) :

Notons $L(\mathcal{F})$ le membre de gauche. Si on pose $\mathcal{F}_a = \overline{\mathcal{F} \cap \Lambda_a}$, où Λ_a est l'ensemble relativement compact construit au cours de la démonstration de la Prop. 3.4. et décrit en (3.7), on a $I(\mathcal{F}) \leq I(\mathcal{F}_a)$, et d'après (5.5)

$$L(\mathcal{F}) \leq \operatorname{Max}(L(\mathcal{F}_a) , \operatorname{Log} a) ,$$

autrement dit, si le second membre est fini, a étant arbitraire, on peut se ramener au cas $\mathcal{F}$ <u>compact</u>.

$\mathcal{F}$ est donc maintenant un compact, disjoint pour $\alpha < I(\mathcal{F})$ du compact convexe $\mathcal{K}_\alpha$ de la formule (2.5) . On peut donc recouvrir $\mathcal{F}$ par un nombre <u>fini</u> d'ouverts convexes relativement compacts $\mathcal{G}_j$ tels que $\overline{\mathcal{G}_j} \cap \mathcal{K}_\alpha = \emptyset$. En appliquant encore (5.5), on est ramené au cas d'un $\mathcal{F}$ convexe compact d'information arbitrairement voisine. D'après Hahn-Banach, il existe f dans $\mathcal{E}_b(\mathcal{X})$ telle que :

$$\sup_{\lambda \in \mathcal{K}_\alpha} E_\lambda(f) < \inf_{\lambda \in \mathcal{F}} E_\lambda(f) = b .$$

Projetons la situation sur $\mathbb{R}$, dont l'élément générique sera noté x , et les Probabilités Q . On a :

$$\{\hat{\mu}_n \in \mathcal{F}\} \subset \{E_{\hat{\mu}_n} f \geqslant b\} = \{S_n/n \geqslant b\} \quad ,$$

où S_n est la somme de n v.a.r. indépendantes, de loi commune $f \circ \mu$. Cette loi satisfait (3.1), puisque son support est borné. On a donc comme majorant pour (5.1), d'après le calcul effectué Prop.3.1. et celui du §4.,

$$- \inf_{Q \mid E_Q(x) \geqslant b} I(Q, f \circ \mu) .$$

Maintenant, toute Q d'information finie par rapport à $f \circ \mu$ et telle que $E_Q(x) \geqslant b$ peut se relever en une λ de même information sur $\mathcal{X}$, par exemple $d\lambda = \frac{dQ}{df\circ\mu} \circ f \cdot d\mu$. Cette λ ne peut appartenir à $\mathcal{K}_\alpha$, puisque $E_\lambda f = E_Q x \geqslant b$, et donc :

$$\inf_{Q \mid E_Q(x) \geqslant b} I(Q, f \circ \mu) \geqslant \alpha \quad ,$$

ce qui termine la démonstration .

2. Démonstration de (5.2) :

Pour tout $\varepsilon > 0$, l'ouvert étroit $\mathcal{G}$ contient un ouvert $\mathcal{H}$ du type :

$$\mathcal{H} = \{\lambda \;\big|\; |E_\lambda f_j - E_\nu f_j| < r_j \quad ; \quad j = 1, 2, \dots, m\} \quad ,$$

où les f_j sont continues bornées, $r_j > 0$, enfin $I(\nu,\mu) < I(\mathcal{G}) + \varepsilon$.

Soit f l'application de $\mathcal{X}$ dans $\mathbb{R}^m$ définie par $f(X) = (f_j(X))$. Notons $x = (x_j)$ l'élément générique de $\mathbb{R}^m$, G_m l'ouvert $\{x \mid |x_j - E f_j| < r_j$; $j = 1, 2, \ldots, m\}$.

On a alors $\{\hat{\mu}_n \varepsilon \mathcal{G}\} \supset \{E_{f\circ\hat{\mu}_n}(x) \varepsilon G_m\}$; $E_{f\circ\hat{\mu}_n}(x)$ peut s'interpréter comme la moyenne arithmétique $\frac{S_n}{n}$ de n v.a. indépendantes, de loi commune $f\circ\mu$, loi satisfaisant évidemment la condition (3.1) puisque les f_j sont bornées .

Comme $\{Q \mid E_Q(x) \varepsilon G_m\}$ contient $\{f\circ\lambda \mid \lambda \varepsilon \mathcal{X}\}$, que

$$\operatorname*{Inf}_{Q|E_Q(x)\varepsilon G_m} I(Q, f\circ\mu) \leq I(\nu,\mu)$$

d'après (2.4), on conclut en utilisant la formule (4.2) .

3. Démonstration de (5.3) :

Soit $P_1(\mathcal{X})$ l'ensemble des Probabilités sur $\mathcal{X}$ intégrant $||X||$, et T l'application de P_1 dans $\mathcal{X}$ définie par $T(\lambda) = E_\lambda(X)$.

A $a > 0$, associons le compact $\mathcal{K}(a)$ et le nombre M de la Prop. (3.4). Je dis que <u>T est continue de $\mathcal{K}(a)$ dans $\mathcal{X}$</u> (topologie étroite sur $\mathcal{K}(a)$, topologie forte sur $\mathcal{X}$). Soit en effet $S = \{(\lambda_n) ; \lambda\}$ où λ_n tend étroitement vers λ dans $\mathcal{K}(a)$. Comme $P(\mathcal{X})$ est polonais, il suffit de montrer que pour toute telle suite S , $||E_\lambda(X) - E_{\lambda_n}(X)||$ tend vers 0 . Comme $\mathcal{X}$ est polonais, S est équitendue : à tout $\varepsilon > 0$, on peut associer un compact K de $\mathcal{X}$ tel que $\nu(K^c) \leq \varepsilon$ pour toute ν de S . Soit $\nu_K = 1_K\cdot\nu + \nu(K^c) . \partial_o$. Alors :

$$(5.6) \qquad ||E_\nu(X) - E_{\nu_K}(X)|| \leq M \cdot (R / \emptyset(R)) + R \cdot \varepsilon \quad ,$$

pour $\nu \varepsilon S$, R réel positif , $\emptyset$ étant la fonction satisfaisant (3.5), d'après (3.6). Le membre de gauche de (5.6) peut donc être rendu arbitrairement petit , en choisissant bien R et ε , puisque $\emptyset(x)/x$ tend vers l'infini à l'infini.

Soit $S_K = \{\nu_K | \nu \varepsilon S\}$. K étant borné , T est continue de S_K dans (topologie étroite, faible). Mais $T(S_K)$ est contenu dans l'enveloppe convexe de $K \cup (0)$, qui est un <u>compact</u> sur lequel topologie faible et forte coïncident.

Terminons la démonstration de (5.3) :

$$\{\frac{S_n}{n} \in F\} \subset \{\hat{\mu}_n \in T^{-1}(F) \cap \mathcal{K}(a)\} \cup \{\hat{\mu}_n \in \mathcal{K}(a)^c\} \quad .$$

On applique donc (5.5), la majoration (3.6), enfin (5.1) au <u>fermé</u> $\mathcal{F} = T^{-1}(F) \cap \mathcal{K}(a)$. Il vient :

$$\limsup \frac{1}{n} \operatorname{Log} P_\mu\{\frac{S_n}{n} \in F\} \leq \operatorname{Max}(\operatorname{Log} a, -I(\mathcal{F})) \quad .$$

D'après la formule (4.1), $I(\mathcal{F}) \geq h(F)$, on termine en faisant tendre a vers 0.

4. Démonstration de (5.4) :

Si $h(G)$ est fini, on choisit a dans G tel que $h_\mu(a)$ soit proche de $h(G)$, puis, d'après la remarque suivant la démonstration du Théorème 1, λ de moyenne a et d'information $h_\mu(a)$, on l'approche par λ' a support borné, d'information proche, de moyenne proche, donc encore dans G. Si P_R est l'ensemble des Probabilités dont le support est contenu dans $||X|| \leq R$, la démonstration précédente montre que T est continue sur P_R (il suffit de remplacer (5.6) par :

$$||E_\nu(X) - E_{\nu_K}(X)|| < R\cdot\varepsilon) \quad . \quad \mathcal{G} = T^{-1}(G) \cap P_R$$

est donc ouvert dans P_R, d'information moindre que celle de λ', on applique (5.2) en remarquant que

$$\{\hat{\mu}_n \in \mathcal{G}\} \subset \{\frac{S_n}{n} \in G\} \quad .$$

§6. Deux exemples

Dans beaucoup d'applications, le problème est de montrer qu'une partie Ω de $P(\mathcal{X})$ est un bon ensemble pour la loi μ au sens suivant :

$$(6.1) \qquad \lim \frac{1}{n} \operatorname{Log} P_\mu\{\hat{\mu}_n \in \Omega\} = -I_\mu(\Omega) \quad ,$$

où naturellement :

$$I_\mu(\Omega) = \inf\{I(\lambda,\mu) \mid \lambda \in \Omega\} \quad .$$

Le Théorème 2 donnera immédiatement le résultat si l'on trouve une topologie Lusinienne sur $\mathcal{X}$ telle que $\overline{\Omega}$ et $\mathring{\Omega}$ étant la fermeture et l'intérieur de Ω pour la topologie étroite associée, on ait :

$$(6.2) \qquad I(\overline{\Omega}) = I(\Omega) = I(\mathring{\Omega}) \ .$$

Dans ce qui suit, $\mathcal{X} = \mathbb{R}$, mais tout est immédiatement généralisable à $\mathbb{R}^d$. On identifie les mesures à leurs fonctions de répartition (f.r.), notées, F, G, H, avec $F_\mu(x) = \mu\{-\infty,x]\}$; $F^-(x)$ vaut $\mu\{-\infty, x[\}$, et $\hat{F}_n$ est la f.r. associée à la loi empirique $\hat{\mu}_n$. On note D(F) l'ensemble (dénombrable) des points de discontinuité de F.

Soit maintenant $\mathcal{D}$ un ensemble dénombrable dense de $\mathbb{R}$. Dans la suite, $\mathbb{R}$ est muni de la topologie (Lusinienne ! d'après le Lemme 1.1.) $\tau = \tau(\mathcal{D})$ plus fine que l'usuelle et rendant continues les $1_{-\infty,d]}$ et les $1_{-\infty,d[}$ pour tout d de $\mathcal{D}$; la topologie étroite sur $P(\mathbb{R})$ est la topologie associée.

Soit maintenant $\Delta(F,G) = ||F-G||_\infty = ||F^- - G^-||_\infty$. Notons que Δ est atteinte (soit pour le couple F,G, soit pour le couple F^-, G^-). Naturellement, sur $P(\mathbb{R})$, la Δ-topologie est plus fine que la topologie étroire usuelle (c'est-à-dire celle de la convergence en loi) mais :

Lemme 6.1.

Si $D(G) \subset \mathcal{D}$, tout Δ-voisinage de G est un voisinage étroit de G.

(autrement dit, $H \longrightarrow \Delta(G,H)$ est étroitement continue si D(G) est contenu dans $\mathcal{D}$).

Démonstration

Fixons n ; soit $x_{i,n} = \text{Sup}\{x \mid G(x) \leqslant \frac{i}{n}\}$, pour $1 < i < n$;
Si $x_{i,n} \in D(G)$, on pose $y_{i,n} = x_{i,n}$. Sinon $y_{i,n}$ est un point de $\mathfrak{D}$ tel que

$$|G(y_{i,n}) - G(x_{i,n})| \;\; \underline{\text{et}} \;\; |G^{-}(y_{i,n}) - G(x_{i,n})|$$

soient moindres que $\frac{1}{n}$. Soit $\mathcal{G}_n$ l'ouvert étroit $\{H \mid \text{Max}(\text{Max}_i |G-H|(y_{i,n})$, $\text{Max}_i |G^{-}-H^{-}|(y_{i,n})) < \frac{1}{n}\}$ (il est ouvert comme intersection finie d'ouverts élémentaires). La monotonie de G et H assure que si $H \in \mathcal{G}_n$, $\Delta(G,H) \leqslant \frac{3}{n}$, d'où le résultat.

On note dans la suite $\overset{\circ\circ}{\Omega}$ le Δ-intérieur de Ω.

Corollaire

Soit Ω un ensemble Δ-fermé ,

On suppose que $I_F(\overset{\circ\circ}{\Omega}) = I_F(\Omega)$. Alors Ω est un bon ensemble (satisfaisant (6.1)) pour F .

Démonstration

On prend $\mathfrak{D} = \mathbb{Q} \cup D(F)$. Si G adhère étroitement à Ω , et si elle est d'information finie , elle est absolument continue par rapport à F , donc $D(G) \subset D(F) \subset \mathfrak{D}$; et d'après le lemme, elle Δ-adhère à Ω , donc elle est dans Ω. Autrement dit , $I(\overline{\Omega}) = I(\Omega)$. Si $I(\Omega) = \infty$, il n'y a rien de plus à démontrer. Si $I(\Omega) < \infty$, toute G de $\overset{\circ\circ}{\Omega}$ d'information finie est étroitement intérieure à Ω , donc $I(\overset{\circ\circ}{\Omega}) = I(\Omega)$. (6.2) étant vérifiée, on peut appliquer le Théorème 2.

Dans [3] , Stone donne une série de conditions sur Ω qui assurent que Ω est un bon ensemble. Une lecture attentive de ces conditions portant sur les partitions finies de Ω , montre qu'elles assurent qu'en fait, pour une topologie étroite associée à l'infinité dénombrable de partitions effectivement utilisées, son ensemble contient un ouvert d'information arbitrairement voisine, et que $I(\overline{\Omega}) = I(\Omega)$, du moins si $\mathcal{H}$ est Lusinien au départ. Notre propos est plutôt de démontrer directement l'application qu'il en donne, soit son Théorème 2 :

Théorème

Soit sur $\mathbb{R}$, F continue ; T une fonctionnelle sur $P(\mathbb{R})$ uniformément continue pour la distance Δ . On pose $\Omega_r = \{G | T(G) \geq r\}$, $I(r) = I_F(\Omega_r)$. Alors, si $I(0)$ est finie, si I est continue à droite en 0 , Ω_o est un bon ensemble pour F .

Démonstration

Clairement , Ω_o est Δ-fermé , $\overset{\circ\circ}{\Omega}_o$ contient Ω_r pour $r > 0$, donc $I(\overset{\circ\circ}{\Omega}_o) = I(\Omega_o)$. On applique le Corollaire .

Remarque

Il me semble que je démontre le résultat sous des hypothèses plus générales (pas besoin de la continuité de F , de l'uniforme continuité de T , de la finitude de $I(\Omega_o)$) .

Dans [2] , Sethuraman étudie en particulier les grands écarts à Kolmogorov-Smirnov, et son résultat est la suivant :

Théorème

Soit $\Omega_a = \{G \mid \Delta(F,G) \geq a\}$. Alors Ω_a est un bon ensemble pour F . (De plus, il calcule explicitement $I_F(\Omega_a)$) .

Démonstration

Posons $I(a) = I_F(\Omega_a)$. On voudrait se ramener au résultat précédent en posant $T(G) = \Delta(F,G) - a$. Remarquons que la partie majoration est toujours valable, Ω_a étant Δ-fermé. Pour la minoration, effectuons une construction préliminaire. Soit $G \neq F$, $b = \Delta(F,G)$, G absolument continue par rapport à F : comme on l'a remarqué au début de ce chapitre, il existe un x de $\mathbb{R}$, tel que :

$$b = G(x) - F(x) \text{ , ou } G^-(x) - F^-(x), \text{ ou } F(x) - G(x), \text{ ou } F^-(x) - G^-(x) .$$

Plaçons nous dans le premier cas, ou nous poserons $p = F(x)$, $q = 1 - p$, $P = G(x) = p + b$, $Q = 1 - P$ avec $p < P \leq 1$, (et $p \geq 0$ car dG est absolument continue par rapport à dF) . Soit $\mathcal{F}_{P,p} = \{H | H(x) = P\}$.

Alors $\mathcal{F}_{P,p}$ est contenu dans Ω_b et son information $I(P,p)$ vaut

$$P \operatorname{Log} \frac{P}{p} + Q \operatorname{Log} \frac{Q}{p} \quad , \text{ atteinte par } H_{P,p} \text{ définie par}$$

$$dH = \frac{P}{p} dF \text{ ou } \frac{Q}{q} dF \text{ suivant que l'on est sur } -\infty, x] \text{ ou }]x, +\infty \ .$$

(convexité de $x \longrightarrow x \operatorname{Log} x$). $I(P,p)$ est croissante et continue de P sur $[p,1]$. Montrons maintenant que pour toute G de Ω_a , et d'information finie, $\liminf \frac{1}{n} \operatorname{Log} P_F\{\hat{F}_n \in \Omega_a\} \geq -I(G,F)$. Conservant les notations précédentes, supposons encore être dans le premier cas ; posant $P_1 = p+a \leq p+b \leq 1$, la monotonie de $I(P,p)$ entraîne que $I(G,F) \geq I(P_1,p)$. Si maintenant $P_1 < 1$, on pourra trouver P' tel que $P_1 < P' < 1$, avec $I(P',p)$ arbitrairement proche de $I(P_1,p)$. Alors $H_{P',p}$ sera dans l'intérieur (pour Δ) de Ω_a , et le raisonnement du Corollaire montre alors que $\liminf \frac{1}{n} \operatorname{Log} \ldots$ (en passant à la limite en P') est supérieur à $I(G,F)$. Le cas restant est celui où $P_1 = p+b = p+a = 1$. Mais alors $I(G,F) \geq I(1,p) = -\operatorname{Log} p$. L'événement $\{X_i \leq x \text{ pour } 1 \leq i \leq n\}$ entraîne alors $\{\Delta(\hat{F}_n,F) \geq a\}$ et a une F-Probabilité <u>égale</u> à $(F(x))^n$, soit p^n, minorée par $\exp(-n\, I(G,F))$.

La démonstration est terminée en faisant une démonstration similaire dans chacun des trois autres cas. On voit aisément comment <u>calculer</u> $I(a)$, évidemment de distribution indépendante de F si dF est diffuse.

Bibliographie :

[1] DONSKER et VARADHAN : Some problems of large deviations, Communication au Congrès de Rome (1975), à paraître

[2] SETHURAMAN : On probability of large deviations of families of sample means Ann. Math. Stat., 35, 1964, p. 1304-1316

[3] STONE : Large deviations for empirical probability measures Annals of Stat. 1974, vol. 2, p. 362-366 .

Jean BRETAGNOLLE

Université Paris-Nord
Département Mathématiques
Avenue Jean-Baptiste Clément
93430 VILLETANEUSE.