

Grandes Déviations

FORMULE DE CHERNOFF ET PROBLEMES STATISTIQUES

ASSOCIES.

Séminaire de Statistique d'Orsay

1977 - 1978 .

79.T.06

Université de Paris-Sud  
Département de Mathématique

Bât. 425

91405 **ORSAY** France

Grandes déviations

FORMULE DE CHERNOFF ET PROBLEMES STATISTIQUES ASSOCIES

E R R A T A

- La première page "Table des matières" est à supprimer.
- La page n° 3 "Introduction" est à supprimer.
- La page 111 "Bibliographie" se trouve après la page VTT.18

Grandes Déviations

FORMULE DE CHERNOFF ET PROBLEMES STATISTIQUES

ASSOCIES.

Séminaire de Statistique d'Orsay

1977 - 1978 .

79.T.06



32303

Université de Paris-Sud  
Département de Mathématique

Bât. 425

91405 **ORSAY** France

Grandes Déviations.

FORMULE DE CHERNOFF ET PROBLEMES STATISTIQUES ASSOCIES

- I. D. DACUNHA-CASTELLE : Formule de Chernoff pour une suite de variables réelles.
- II. D. DACUNHA-CASTELLE : Formule de Chernoff pour des rapports de vraisemblance.
- III. J. BRETAGNOLLE : Formule de Chernoff pour les lois empiriques de variables à valeurs dans des espaces généraux.
- IV. J. DESHAYES, D. PICARD : Grandes et moyennes déviations pour les marches aléatoires.
- V. J. DESHAYES, D. PICARD : Application aux tests de rupture de régression.
- VI. M. DUFLO : Formule de Chernoff pour des chaînes de Markov.
- VII. N. MAIGRET : Statistique des chaînes contrôlées felleriennes.
- VIII. L. BIRGÉ : Vitesses optimales de convergence des estimateurs.
- IX. G. RUGET : Quelques occurrences des grands écarts dans la littérature "électronique".

FORMULE DE CHERNOFF ET PROBLEMES STATISTIQUES ASSOCIES.

- D. DACUNHA-CASTELLE : Introduction
- I. D. DACUNHA-CASTELLE : Formule de Chernoff pour une suite de variables réelles
- II. D. DACUNHA-CASTELLE : Formule de Chernoff pour des rapports de vraisemblance
- III. J. BRETAGNOLLE : Formule de Chernoff pour les lois empiriques de variables à valeurs dans des espaces généraux
- IV. J. DESHAYES, D. PICARD : Grandes et moyennes déviations pour les marches aléatoires.
- V. J. DESHAYES, D. PICARD : Application aux tests de rupture de régression
- VI. M. DUFLO : Formule de Chernoff pour des chaînes de Markov
- VII. N. MAIGRET : Majorations de Chernoff et statistique séquentielle pour des chaînes de Markov récurrentes au sens de Doeblin
- VIII. N. MAIGRET : Statistique des chaînes contrôlées felleriennes
- IX. L. BIRGE : Vitesses optimales de convergence des estimateurs
- X. G. RUGET : Quelques occurrences des grands écarts dans la littérature "électronique".

INTRODUCTION

D. DACUNHA-CASTELLE.

I. LES FONDEMENTS PROBABILISTES DE LA THEORIE DES GRANDES DEVIATIONS.

La théorie des grandes déviations (grands écarts) concerne la loi faible des grands nombres. Si  $X_1 \dots X_n$  est une suite de variables aléatoires indépendantes, équidistribuées, d'espérance  $EX = 0$ , et si  $S_n = X_1 + \dots + X_n$ , la loi faible des grands nombres assure de la convergence vers 0 de  $\frac{S_n}{n}$  (en probabilité). Pour préciser cette convergence, on dispose d'abord du théorème de limite centrale qui étudie la convergence de  $\frac{S_n}{\sqrt{n}}$ . Soit  $a > 0$ , on obtient :

$$P(S_n > \sqrt{n} \sigma a) \sim 1 - \Phi(a)$$

où  $\Phi$  est la répartition de la loi normale

et  $\sigma^2 = EX^2$ . Au lieu d'étudier le grossissement  $\sqrt{n}$  de  $\frac{S_n}{n}$ , on peut étudier les grandes déviations sans grossissement, à savoir  $P(S_n > na)$ . Ces grandes déviations sont rares et le premier résultat sera que

$$P(S_n > na) \sim e^{-nh(a)} \quad (1)$$

où  $h$  est une fonction convenable.

De manière générale, on peut étudier  $P(S_n > B_n a)$ .

$B_n = O(\sqrt{n})$  donne le théorème de limite centrale et  $B_n = O(n)$  les grandes déviations. Entre les deux  $\sqrt{n} = \sigma(B_n)$  et  $B_n = \sigma(n)$ , on trouve la théorie des moyennes déviations, qui conduit à des probabilités de l'ordre de  $\frac{1}{n^\alpha}$  pour des  $\alpha$  convenables,  $\alpha > 0$ , dans certains cas.

Bien entendu, l'ensemble de ces théorèmes ne valent que sur des hypothèses convenables sur la queue des distributions des  $X$ . Très grossièrement la loi des grands nombres demande  $E |X| < \infty$ , le théorème limite centrale  $E X^2 < \infty$  et les résultats de grandes déviations  $E \exp tX < \infty$  pour des  $t > 0$ . Si les premiers résultats sur les grandes déviations sont dus à KHINCHINE [6], puis avant-guerre par exemple à CRAMER [3] ils ont pris sous l'influence des statisticiens le nom générique de formules de Chernoff [2]. En fait, on trouve d'abord des majorations exponentielles du type  $P(S_n > na) \leq e^{-nh(a)}$  qui dans les cas simples sont des applications directes d'une formule de Markov-Tchebychev. La minoration, même dans le cas élémentaire que nous venons d'introduire, a une démonstration intéressante en soi. Elle est une conséquence directe de la loi faible des grands nombres, appliquée à  $\frac{S_n}{n}$ , non pour la probabilité initiale associée à la loi  $dF$  des  $X$ , mais pour une nouvelle probabilité  $P_a$  obtenue en remplaçant  $dF$  par  $dF_a$ , avec  $\frac{dF_a}{dF}(x) = \frac{e^{tx}}{\phi(t)}$ , où  $t$  est choisi tel que  $\int x dF_a(x) = a$ , soit  $\frac{\phi'(t)}{\phi(t)} = a$  si  $\phi(t) = E \exp tX$ .

Ce recentrage de la probabilité permet d'appliquer un théorème ergodique, ici la loi faible ordinaire. Cette idée sera essentielle.

Le chapitre I, élémentaire, souligne que pour qu'une suite  $Z_n$  de loi  $P_n$  satisfasse à la formule (1), il suffit que la transformée de Laplace  $\hat{\phi}_n$  de  $P_n$  existe sur un intervalle contenant 0 et qu'elle ait un comportement ergodique du type  $\frac{1}{n} \log \hat{\phi}_n(t) \rightarrow L(t)$ . Cette convergence assure la loi faible des grands nombres pour les  $dP_{n,a}$  obtenues par recentrage. On a aussi le théorème limite centrale pour les  $P_{n,a}$  qui donne d'ailleurs un développement plus précis que (1).

La technique utilisée pour obtenir (1) est voisine de celle utilisée par exemple par LINNIK et FELLER [5] pour obtenir des formules de moyennes déviations mais aussi des techniques utilisées pour obtenir des développements affinant le théorème de limite centrale, comme les techniques de point de selle.

Intuitivement, si l'on interprète les  $X_i$  comme des variables à valeur

mesure (masses de Dirac) plutôt que comme des réels, on est amené à étudier, non pas  $\frac{S_n}{n}$  mais  $\hat{F}_n = \frac{1}{n} (\delta_{X_1} + \dots + \delta_{X_n})$ , répartition empirique. On sait que cette répartition converge (loi faible) et qu'elle satisfait au théorème de GLIVENKO-CANTELLI (convergence de  $\sqrt{n} \sup_x |\hat{F}_n(x) - F(x)|$  qui est l'analogie du théorème de limite centrale. Il est donc raisonnable d'essayer d'obtenir des équivalents pour  $P(\hat{F}_n \in B)$  où  $B$  est un ensemble exceptionnel ou très rare pour  $\hat{F}_n$ ,  $B \subset \mathcal{P}(\mathbb{R})$ , ensemble des probabilités sur  $\mathbb{R}$ . La formule (1) équivaut à

$$P(\hat{F}_n \in B) \sim e^{-n h(a)} \quad \text{avec}$$

$$B \{G \in \mathcal{P}(\mathbb{R}), \int x dG > a\}$$

De manière générale, on cherchera des formules du type

$$P(\hat{F}_n(B)) \sim e^{-n I(B)} \quad (2)$$

mais de telles formules ne vont exister que pour des ensembles bien particuliers; par contre, ouverts et fermés conduiront à des inégalités.

Les formules (2) ont un sens pour des variables à valeurs dans des espaces très généraux. Le cadre agréable est celui des polonais, étendu pour les nécessités au cas souslinien.

Les formules (1) ont un sens pour des variables à valeurs dans un espace vectoriel. Le cas étudié sera celui des Banach séparables. Dans tous les cas, ce qui est décisif c'est l'extension à  $\mathbb{R}^k$  des formules (1) et (2). (Pour les polonais on se ramènera à ce cas, en considérant des ensembles convexes, des techniques de séparation et de projection).

Une notion apparaît tout de suite comme centrale dans la théorie, c'est celle d'informations de Kullback, avec ses différentes formulations.

Rappelons que si  $P$  et  $Q$  sont deux probabilités, on définit  $I(P, Q)$

par



$$I(P, Q) = E_P \log \frac{dP}{dQ}$$

$$= \int (\log \frac{dP}{dQ}) dP$$

lorsque cette quantité a un sens, et  $+\infty$  si elle n'en a pas.

$I$  a les propriétés fondamentales bien connues d'une information. En particulier elle diminue par image. De plus elle est convexe et s.c.i. par rapport à l'argument  $P$  (et pour la convergence étroite).

Dans la démonstration élémentaire de (1), on introduit la fonction  $h$  définie par

$$h_F(a) = \sup_t (ta - \log \phi(t)) \quad (3)$$

Cette fonction dite transformée de Cramer de  $F$  (qu'elle permet de retrouver), est donc la duale, au sens de Young-Orlicz de l'analyse convexe, de la fonction  $\log \phi(t)$ . Son lien avec l'information, connu depuis longtemps, se traduit par la formule

$$h_F(a) = \inf_{\{G, \int x dG(x)=a\}} I(G, F) \quad (4)$$

Cette formule est un outil technique décisif. Elle permet d'abord d'utiliser (4) pour définir l'exposant  $h(B)$  qui sera naturel pour les extensions (2) de (1), on posera

$$h_F(B) = \inf_{G \in B} I(G, F) \quad (5)$$

$h_F(B)$  sera donc l'exposant mesurant la rareté de l'ensemble  $B$  par rapport à un échantillonnage  $\hat{F}_n$  de  $F$ . (5) a donc un sens sur un espace abstrait quelconque. La formule d'analyse convexe (3) s'étend elle au cas d'un Banach séparable  $E$ , en supposant que  $\phi_F(t) = E_F \exp t \|X\| < \infty$ . On peut poser alors :

$$h_F(a) = \sup_{\theta \in E^*} (\langle \theta, a \rangle - \log E_F \exp \langle \theta, X \rangle) \quad (6)$$

(6) permet d'étendre (4) au cas d'un Banach.

Le chapitre III fait une étude détaillée des résultats concernant les cas Banach séparable et polonais (les principaux résultats sont dus à DONSKER et VARADHAN). On obtient essentiellement

$$\begin{aligned}
 (7) \quad & \limsup \frac{1}{n} \log P_F \left( \frac{S_n}{n} \in B \right) \cong - \inf h_F(B) && \text{si } B \text{ est fermé} \\
 (8) \quad & \limsup \frac{1}{n} \log P_F \left( \frac{S_n}{n} \in B \right) \cong - \inf h_F(B) && \text{si } B \text{ est ouvert} \\
 & && \text{(cas Banach)} \\
 \text{et (9)} \quad & \limsup \frac{1}{n} \log P_F \left( \hat{F}_n \in B \right) \cong - \inf_{\lambda \in B} I(\lambda, F) && B \text{ fermé de } \mathcal{P}(E) \\
 (10) \quad & \limsup \frac{1}{n} \log P_F \left( \hat{F}_n \in B \right) \cong - \inf_{\lambda \in B} I(\lambda, F) && B \text{ ouvert de } \mathcal{P}(E), \\
 & && \text{(cas polonais)}.
 \end{aligned}$$

Dans la pratique statistique, il est intéressant de savoir si des ensembles  $B$  particuliers sont de bons ensembles, à savoir qu'ils réalisent à la fois (9) et (10). Un cas utile pour l'étude des tests de Kolmogorov-Smirnov est celui où l'on a

$$B = \{G, T(G) > a\}$$

où  $T$  est une fonctionnelle uniformément continue sur  $\mathcal{P}(\mathbb{R})$ , munie de la distance de  $(G, G') = \|G - G'\|_{\infty}$ . Lorsque  $F$  est continue, ce type de résultats a été l'objet d'efforts importants. Le passage (élémentaire et direct) du cas polonais au cas souligné donne ce type de résultat, en conclusion du chapitre III, sans effort notable.

Le cadre probabiliste général est donc fixé par le chapitre III, avec trois notions essentielles : information de Kullback, transformée de Cramer et changement de probabilité privilégié recentrant le problème.

Les développements probabilistes sont extrêmement nombreux et touchent tous les types de processus aléatoires sans exception (processus de Markov ponctuels, gaussiens, etc...). Dans ce séminaire ne sont traités que des problèmes bien particuliers, portant sur des ensembles  $B$  très précis et choisis en vue d'applications statistiques. Cependant, les deux chapitres qui y sont consacrés présentent les techniques essentielles. Pour le reste nous renvoyons à [0]

un cours de R. AZENCOTT sur des aspects probabilistes essentiels.

Le chapitre IV, à partir d'idées de Borovkov, étudie les probabilités de certaines grandes déviations des marches aléatoires. Considérant la marche  $S_n = X_1 + \dots + X_n$  introduite plus haut, on définit une suite de processus à valeurs sur  $C_0(0,1)$  en posant

$$s_n\left(\frac{k}{n}\right) = \frac{S_k}{n}, \quad 1 \leq k \leq n; \quad s_n(0) = 0,$$

et  $s_n(t)$  étant obtenue à partir de  $s_n\left(\frac{k}{n}\right)$  par interpolation linéaire.

Les ensembles  $B$  auxquels on s'intéresse sont du type suivant. Soit  $g(t)$  une fonction sur  $(0,1)$ , on étudie

$$B = \{f \in C_0(0,1), \exists t \text{ avec } f(t) \geq g(t)\}$$

$(S_n \in B)$  est donc l'événement : franchissement par la marche aléatoire  $S_n$  de la barrière  $B$ . Une variante est le non-franchissement d'une barrière, soit :

$$B = \{f \in C_0(0,1), \forall t, f(t) < g(t)\}$$

Bien entendu, il faut que  $g$  satisfasse des conditions particulières pour que  $B$  conduise à un événement exceptionnel. Pour le franchissement, il faut qu'il existe des  $t$  tels que  $\frac{g(t)}{t} > EX$ , ceci puisque  $\frac{s(t)}{t} \neq \frac{S_k}{n} \frac{n}{k}$  et que  $\frac{S_k}{k} \rightarrow EX$ .

Si  $f$  est presque partout dérivable, on définit  $W$  par

$$W(f) = \int_0^1 h(f'(t)) dt \quad (11)$$

(quelquefois liée à la notion d'intégrale d'action).

$W$  joue pour ces problèmes de suite de marches le rôle de la transformée de Cramer (de l'information). Par passage à la limite, à partir de la suite des problèmes discrets associés à chaque  $n$ , on obtient des résultats du type suivant, si  $EX = 0$ ,

$$\underline{\lim} \frac{1}{n} \log P_F (s_n \in G) \cong -W(\overset{\circ}{B}) \quad (12)$$

$$\overline{\lim} \frac{1}{n} \log P_F (s_n \in G) \cong -W(\overline{B}) \quad (13)$$

si  $B$  est un borélien de  $C(0,1)$ , d'intérieur  $\overset{\circ}{G}$  et de fermeture  $\overline{G}$ , et

$$W(B) = \inf_{\{f \in G, f(0)=0\}} W(f)$$

Les démonstrations faites pour obtenir (12) et (13) valent pour les moyennes déviations, à condition de poser

$$s_n \left(\frac{k}{n}\right) = \frac{S_k}{x(n)}, \quad \frac{x(n)}{n} \rightarrow 0, \quad \frac{x(n)}{\sqrt{n}} \rightarrow \infty$$

et de remplacer dans (12) et (13),  $\frac{1}{n}$  par  $\frac{n}{x^2(n)}$  et  $W$  par  $W_0$  associée à la loi normale (c'est-à-dire à  $h(t) = \frac{t^2}{2}$ ).

Cette première partie du chapitre IV introduit donc une technique générale de passage du discret au continu. La deuxième partie vise à calculer explicitement  $W(B)$  lorsque  $B$  correspond à un franchissement (ou à un non franchissement de frontière). On trouve par exemple pour le franchissement : s'il existe  $t_0$  tel que  $\frac{g(t_0)}{t_0} < \text{ess sup } X$

$$W(B) = \inf_u h\left(\frac{g(u)}{u}\right).$$

La remarque suivante semble avoir des incidences, qui ne sont pas encore suffisamment explorées, dans de nombreux problèmes tant mathématiques que directement appliqués : s'il existe un  $u_0$  unique réalisant  $\inf_u h\left(\frac{g(u)}{u}\right)$ , alors on obtient une information décisive sur la façon dont s'effectue le franchissement lorsque celui-ci a lieu. A savoir : conditionnellement au franchissement de la frontière  $g$ , la fonction  $\frac{g(u_0)}{u_0} \inf(t, u_0)$  est la trajectoire la plus probable.

Dans certains problèmes, il y a donc une distribution conditionnelle très "pointue" sur l'ensemble des trajectoires franchissant une frontière (cet ensemble étant de probabilité très petite). Des développements du même ordre peuvent être faits pour le même franchissement d'une frontière, ou le non échappement d'un tube.

Le chapitre VI aborde l'extension des formules (1) et (2) au cas d'une chaîne de Markov, contrôlée ou non. Soit  $X_1 \dots X_n, \dots$  une telle chaîne, d'espace d'états  $(E, \underline{\mathcal{E}})$ , polonais en général, et d'espace d'actions  $(A, \underline{\mathcal{A}})$ . La chaîne est caractérisée par la transition sous contrôle  $\pi(x, a, dy)$  et l'on définit la fonction empirique des observations et des contrôles par

$$\hat{F}_n(B) = \sum_{j=1}^n 1_B(X_j, A_j, X_{j+1})$$

où  $A_j$  est le  $j^{\text{ème}}$  contrôle.

Si  $\lambda$  est une probabilité sur  $E \times A \times E$ , on appellera  $\lambda$  sa première marginale sur  $E$ ,  $\lambda_1 \otimes s$  sa marginale sur  $E \times A$  ( $s$  est donc une transition de  $E$  dans  $A$ ), et l'on s'intéressera à l'information de  $\lambda$  par rapport à  $\lambda_1 \otimes s \otimes \pi$ , qui mesurera bien l'écart en situations initiales  $\lambda_1, s$  de  $\lambda$  à la chaîne.

$$\text{On pose donc } I(\lambda) = I(\lambda, \lambda_1 \otimes s \otimes \pi)$$

$$\text{et } I(B) = \inf_{\lambda \in B} I(\lambda),$$

$B$  étant un ensemble de probabilités sur  $E \times A \times E$ .

Le premier résultat obtenu au chapitre VI concernant les chaînes contrôlées est une majoration. On a

$$\overline{\lim}_n \frac{1}{n} \log \sup_{x, \delta} P_x^\delta(\hat{F}_n \in K) \leq I(K)$$

$K$  compact,  $x$  valeur initiale de la chaîne et  $P^\delta$  probabilité associée à la stratégie de contrôle  $\delta$ . Ce résultat uniforme par rapport à  $\delta$  n'a rien de très surprenant, et vaut dans toutes les situations probabilistes connues. Mais il nécessite un certain travail technique. La situation est affinée dans le cas non contrôlé. On peut alors (plus facilement) obtenir une formule (2), en définissant un changement de probabilité convenable pour rendre la chaîne récurrente (ce qui remplace le simple centrage dans le cas des marches aléatoires, qui est lui aussi l'opération nécessaire pour rendre la marche récurrente). Une fois la chaîne récurrente, on peut appliquer un théorème ergodique et obtenir (2) et des

variantes fortes de (1). La nouvelle probabilité est construite à partir de l'information de la manière qui suit.

Soit  $\alpha \in \mathcal{P}(E)$  ensemble des probabilités sur  $E$ ,  $\lambda \in \mathcal{P}(E \times E)$ . On dira que  $\lambda \in \mathcal{P}^2(\alpha)$  si les deux marginales de  $\lambda$  valent  $\alpha$ . Posons alors

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

avec  $I(\lambda) = F(\lambda, \alpha \otimes \pi)$ .

Si  $\alpha$  est  $\pi$ -invariante,  $\alpha \otimes \pi \in \mathcal{P}_2(\alpha)$  est donc  $I_1(\alpha) = 0$ . Il est donc naturel d'étudier la mesure  $\lambda$  qui réalise  $I_1(\alpha)$ , si elle existe. Moyennant de faibles hypothèses topologiques et sur  $\pi$  (type transition fellerienne), on montre qu'il existe une telle  $\lambda$ , qui s'écrit donc  $\lambda = \alpha \otimes \bar{\pi}$ , où  $\bar{\pi}$  est une transition de  $E$  dans  $E$ , telle que  $\alpha \bar{\pi} = \alpha$  (donc  $\alpha$  est une probabilité invariante pour la chaîne de transition  $\bar{\pi}$ ). Mais surtout on montre que  $\bar{\pi}$  est récurrente (récurrente positive au sens de Harris). Et elle définit le changement de probabilité adéquat. Si les  $\pi(x, \cdot)$  sont toutes équivalentes on obtient les formules (9) et (10) pour une chaîne de Markov. Diverses variantes peuvent être développées, y compris en direction de la transformée de Cramer. L'étude du cas des chaînes de Markov permet de mieux comprendre encore les liens entre les notions fondamentales introduites plus haut.

## II. QUELQUES APPLICATIONS DE LA THEORIE DES GRANDES DEVIATIONS.

### A) Applications à la statistique.

La première application, qui conditionne toutes les autres, concerne les tests de vraisemblance et d'abord le test de Neyman-Pearson de deux hypothèses simples  $P_0$  et  $P_1$ .

Considérons d'abord le cas d'un échantillon ordinaire  $X_1 \dots X_n$  et faisons le test de  $F_0$  contre  $F_1$  ( $P_0 = F_0^{\otimes n}$ ,  $P_1 = F_1^{\otimes n}$ ). Soit  $f_0$  et  $f_1$  les densités de  $F_0$  et  $F_1$  par rapport à une mesure dominante, et soit

$$L(X_1, \dots, X_n) = \sum_{j=1}^n \log \frac{f_1(X_j)}{f_0(X_j)}$$

le logarithme du rapport de vraisemblance  $\frac{dP_1}{dP_0}$ . Le test de Neyman-Pearson a une région de rejet de  $F_0$  du type  $(L_n > na)$ , de niveau  $\alpha_n = P_0(L_n > na)$  et de probabilité de fausse alarme  $\beta_n = P_1(L_n \leq na)$ . Lorsque le seuil  $a$  est fixé, on voit d'après la formule (1), que l'on a sensiblement, pour  $a$  convenable,

$$\alpha_n \sim e^{-nh(a)}$$

$h$  étant la transformée de Cramer duale de  $\log \phi(t)$ , où

$$\begin{aligned} \phi(t) &= E_{F_0} \exp t \log \frac{f_1}{f_0} \\ &= \int f_1^t f_0^{1-t} d\nu \end{aligned}$$

$\phi(t)$  est donc directement liée à la transformée d'Hellinger de  $F_0$  et  $F_1$ . Les dérivées de  $\phi$  en  $t = 0$  et  $t = 1$  valent respectivement  $-I(F_0, F_1)$  et  $I(F_1, F_0)$ . En appliquant toujours (1), pour  $a$  fixé, on obtiendra des probabilités de fausse alarme  $\beta_n \sim e^{-n\beta}$  où  $\beta$  se calcule en fonction de  $h$  et  $a$ . Les liens entre  $\alpha, \beta, a$  sont précisés au chapitre II, application directe du chapitre I.

Le calcul est fait aussi pour le cas du test entre deux processus gaussiens stationnaires (afin de justifier une formule utilisée en théorie du signal de manière approchée). L'extension naturelle de ces résultats concerne les

tests de vraisemblance traités par BROWN, puis BIRGE, résultats exposés au séminaire, mais rédigés par ailleurs et à paraître.

Si  $\Theta_0$  et  $\Theta_1$  sont les hypothèses à tester  $\Theta_0 \cup \Theta_1 = \Theta$ , ensemble des paramètres, les tests de vraisemblance sont fondés sur une région de rejet du type  $(T_n > na)$  où

$$T_n = \frac{\sup_{\theta \in \Theta_1} U_n(X_1 \dots X_n, \theta)}{\sup_{\theta \in \Theta_0} U_n(X_1 \dots X_n, \theta)}$$

où  $U_n(X_1 \dots X_n, \theta)$  est la vraisemblance sous  $P_\theta$  de  $X_1 \dots X_n$ .

L'optimalité, au sens de la théorie classique (par exemple de la théorie de la contiguïté) des tests de vraisemblance est difficile à obtenir (bien que récemment on ait commencé à avoir des indications). Par contre la théorie des grandes déviations permet de démontrer cette optimalité, de manière naturelle, mais pour une suite de niveaux exponentiels (ou suivant le résultat de BAHADUR exposé plus loin). En dehors des études générales, signalées plus haut, de nombreux travaux récents ont été faits, y compris dans le cadre séquentiel, où le point de vue grandes déviations est le seul opérationnel à ce jour. Ces études sont menées aux chapitres VII et VIII dans le cadre des chaînes de Markov. Les formules du type (2) permettent d'obtenir une règle d'arrêt asymptotiquement la plus économique (au sens de la moyenne). Les tests séquentiels utilisés ici sont des tests de vraisemblance et les temps d'arrêt sont construits à partir d'estimateurs du maximum de vraisemblance (ou du maximum de contraste). Les solutions asymptotiques obtenues rejoignent beaucoup de procédures empiriques utilisées dans la pratique. Notons que, des méthodes analogues peuvent être développées pour toute classe de problèmes séquentiels portant sur des processus pour lesquels il existe une théorie de la vraisemblance (ce qui a été fait depuis ce séminaire pour le cas de différents problèmes portant par exemple sur les processus ponctuels).

Revenant aux comparaisons des tests, indiquons que nous n'avons pas développé ici le point de vue et le vocabulaire des travaux de BAHADUR [1], qui



est une forme un peu différente d'utilisation pour ces problèmes des grandes déviations.

Si  $T_n$  est une suite de tests de  $\Theta_0$  contre  $\Theta_1$ , de répartition  $G_\theta$  sous  $P_\theta$ , alors le niveau observé est défini par

$$R(T_n) = \inf_0 \{G_\theta(T), \theta \in \Theta_0\}$$

et le test est d'autant meilleur que  $R(T_n)$  est petit. En particulier on peut étudier les situations où  $\frac{1}{n} \log [1-R(T_n)] \rightarrow -h(\theta)$ ,  $h(\theta)$  est dit pente du test dans cette littérature. Un point de vue voisin également développé par cette même tendance est d'étudier et de comparer les tests pour lesquels il existe une suite  $k_n \rightarrow \infty$  et  $p$ ,  $0 < p < 1$  tel que  $P_\theta(T_n > k_n) \rightarrow p$ , pour  $\theta \in \Theta_1$ , et de montrer qu'alors  $\frac{1}{n} \log \alpha_n \rightarrow -h(\theta)$  (indépendant de  $p$ , fausse alarme fixée). Des résultats en ce sens ont été obtenus pour des classes très larges de tests, y compris dans un cadre séquentiel. Plutôt que de comparer les tests par  $h(\theta)$ , le chapitre V donne un exemple de comparaison suivant les probabilités de fausse alarme obtenues pour des niveaux exponentiels, ou suivant les seuils.

Le problème est d'étudier les ruptures de moyenne (plus généralement des ruptures de régression). Le modèle de base est

$$X_i \sim N(\theta, \sigma^2) \quad 1 \leq i \leq r$$

$$X_i \sim N(\theta + d, \sigma^2) \quad r < i \leq N$$

(où  $\sim$  veut dire suit la loi).  $\theta, d, \sigma^2, r$  sont des paramètres inconnus,  $\Theta_0$  correspond à  $d = 0$ ,  $\Theta_1$  à  $d \neq 0$ .

Il existe des procédures classiques pour ces tests, fondées sur le comportement d'une marche aléatoire  $S_r = \sum_{k=1}^r \hat{\epsilon}_k$ , où les  $\hat{\epsilon}_k$  sont une suite de variables indépendantes, de loi  $N(0, \sigma^2)$ , résidus récursifs construits d'après l'observation. Sous  $\Theta_0$ ,  $S_r$  est approximable par un mouvement brownien, en remplaçant  $S_r$  par des mesures sur  $C_0(0,1)$  comme au chapitre IV (voir l'analyse faite plus haut pour les marches aléatoires), alors que sous  $H_1$ ,  $S_r$  a une dérive. Les tests ont pour région de rejet de  $H_0$  la région de franchissement

d'une frontière (que l'on traverse en cas de dérive). Les calculs classiques sont rendus imprécis du fait de l'approximation marche aléatoire-mouvement brownien, qui est de l'ordre de  $\frac{1}{\sqrt{n}}$ , ordre sur lequel porte précisément les comparaisons usuelles. A ces tests on peut opposer le test de vraisemblance. La comparaison et l'étude des différents tests est aussi rendue compliquée par la taille de l'alternative, le caractère discret du paramètre  $r$ . Il est donc raisonnable de commencer par comparer ces tests au sens des grandes déviations, ce qui évite d'abord les problèmes d'approximation discret-continu, permet de montrer l'optimalité du test de vraisemblance et donne donc une référence. De plus certaines procédures très simples sont introduites, dont les performances au sens des grandes déviations sont assez bonnes. On a donc intérêt à en faire des études empiriques pour voir leurs performances pour des échantillons plus courts.

Une critique méthodologique est souvent faite à ces études de grandes déviations parce qu'elles portent sur des échantillons trop longs, mais surtout sur des probabilités totalement inaccessibles. Cette critique peut être fondée (sauf par exemple dans le domaine des communications, où la valeur de  $\beta$  est décisive). On peut y répondre cependant en doutant que les valeurs de mesure du paramètre dans le cadre classique ait un sens toujours clair et surtout par un argument de mathématique heuristique : dans bien des cas, les classements obtenus par grande déviation et par des méthodes classiques (pour des niveaux constants et éventuellement des hypothèses contigues) sont les mêmes. Les grandes déviations ont l'avantage de répondre aux problèmes de classement des tests et de recherche de solutions optimales, là où (à ce jour) les méthodes classiques sont en difficulté. La défense de ces techniques au niveau de la méthodologie nous semble donc possible.

L'exposé du chapitre IX qui termine la partie statistique est marginal en regard des autres exposés (et beaucoup plus théorique). Il comporte essentiellement l'amélioration d'un résultat connu de LE CAM sur l'estimation d'un paramètre à valeurs dans un espace métrique de dimension finie. Il met en évidence le lien

entre les majorations du type (1) et la théorie du maximum de vraisemblance sur une suite de réseaux.

Tous les résultats statistiques présentés dans le séminaire sont originaux.

B) Les applications à la littérature électronique.

Le dernier exposé est une revue des applications des grandes déviations au domaine du codage, de la transmission avec des développements sur des modèles du type ALOHA de communication par satellite. Cet exposé est un exemple de ce que peuvent apporter des idées mathématiques relativement simples à un domaine appliqué. Outre la construction de bornes et la simplification des méthodes pour les problèmes liés à la théorie de SHANNON, le changement de probabilité des grandes déviations est l'outil de la méthode de simulation dite "importance sampling". Le domaine offre de nombreuses voies de travail.

La littérature électronique n'est bien entendu pas la seule application des grandes déviations. Signalons une démonstration simple et éclairante des lois de Zipf en linguistique [8].

BIBLIOGRAPHIE: (Quelques références générales, voir après chaque exposé une bibliographie spécialisée)

- [ 0 ]      AZENCOTT R.  
Cours de l'Ecole d'Eté de Saint-Flour, 1978  
A paraître dans Lectures notes, Springer-Verlag
- [ 1 ]      BAHADUR (1962)  
Stochastic comparison of tests  
Ann. Math. Stat. 31, p. 276-295
- [ 2 ]      CHERNOFF H. (1952).  
A measure of asymptotic efficiency for tests based on the sum of observations  
Ann. Math. Stat. 23, p. 493-507
- [ 3 ]      CRAMER H. (1938)  
Sur un nouveau théorème limite de la théorie des probabilités  
Act. Sc. Ind. n° 736
- [ 4 ]      DOBRUSHIN (1955)  
A general formulation of the fundamental theorem of Shannon  
Usp. Nat. Nauk, vol. 6, p. 3-104
- [ 5 ]      FELLER W.  
An introduction to probability theory and its applications (Wiley)
- [ 6 ]      KHINCHIN (1929)  
Ober eine neuer grenzwertats des Wahrscheinlichertreihnung  
Math. Ann. vol. 101, p. 745-752
- [ 7 ]      RAGHAVACHARI (1970)  
On a theorem of Bahadur  
Ann. Math. Stat. 41, p. 1695-1699
- [ 8 ]      ROUAULT A.  
Démonstration simple des lois de Zipf (à paraître)

[9]

SANOV (1957)

On the probability of large deviations of random variables

Mat. Sb. n° 5, 42, p. 11-44.

D. DACUNHA-CASTELLE  
Mathématiques - Bât. 425  
ERA CNRS 532 "Statistique Appliquée"  
Université Paris-Sud  
91405 ORSAY

## INTRODUCTION

Les textes rassemblés ici correspondent à des exposés faits au Séminaire sur les Grandes Déviations d'Orsay, en 1977-78. Ils contiennent quelques exposés fondamentaux et pour l'essentiel des applications à la statistique.

Les grandes déviations concernent l'étude de phénomènes probabilistes dans lesquels il existe un comportement moyen (quelquefois appelé déterministe) dont on s'écarte beaucoup avec une probabilité très faible, d'ordre exponentiel par rapport à un certain paramètre (en général le temps). Il s'agit alors de trajectoires dont le comportement qualitatif est différent de la trajectoire moyenne. La théorie des grandes déviations mesure donc ces ensembles exceptionnels de trajectoires. Le cadre élémentaire est celui de variables aléatoires indépendantes et équidistribuées, à valeurs dans  $\mathbb{R}$  ou dans un Banach. Il est développé dans les premier et troisième exposés. Le caractère très général de ce dernier permet d'englober des cas de fonctions empiriques de résultats en général développés dans un autre cadre. Le deuxième exposé concernant l'application statistique la plus immédiate, là aussi traitée dans un cadre différent du cadre usuel, s'applique aux processus gaussiens stationnaires. Deux exposés concernant le cas de chaînes de Markov et des problèmes de contrôle, le second ne comportant que des résultats nouveaux. Les problèmes de marches aléatoires sont là, pour introduire l'étude statistique de la rupture de régression, les résultats de cet exposé étant aussi nouveaux. Enfin, on trouvera un exposé de synthèse sur l'application des grandes déviations dans le domaine de l'information et des communications. De toute manière, les différentes notions d'information apparaissent comme un des éléments unificateurs de la théorie. Un exposé marginal traite de la vitesse d'estimation en statistique théorique. L'aspect probabiliste, marqué par des travaux importants, notamment de Ventsel et Friedlin n'est pas développé ici, quoiqu'il ait été abordé au Séminaire. Nous renvoyons au cours de Robert Azencott (Séminaire St Flour, 1978, à paraître) qui en fait une synthèse.

Nous remercions Madame Baillet pour le remarquable travail de préparation et de frappe qu'elle a fourni.

Orsay, février 1979

Equipe de Recherche Associée au C.N.R.S. 532  
"Statistique Appliquée".



Formule de Chernoff pour une suite de variables réelles

Didier DACUNHA-CASTELLE.

I - Le problème

Classiquement, on considère une suite  $(X_n)_{n \in \mathbb{N}}$  de variables aléatoires indépendantes équidistribuées, telles que :

$$\phi(t) = E \exp t X < \infty \quad \text{pour } t \in ]-M_1, M_2[ \quad , \quad M_1, M_2 > 0 .$$

Ceci implique que  $EX$  existe, et que  $S_n = X_1 + \dots + X_n$  satisfait à

$$\frac{S_n}{n} \xrightarrow{\text{p.s.}} EX .$$

Si l'on étudie les grandes déviations de  $\frac{S_n}{n}$  par rapport à  $EX$ , à savoir la quantité  $P(\frac{S_n}{n} > a)$ ,  $a > EX$ , (le cas  $a < EX$  étant trivial), on a une majoration évidente qui est :

$$P(\frac{S_n}{n} > a) \leq E \exp t (\frac{S_n}{n} - a) , \quad t > 0$$

ou encore 
$$\frac{1}{n} \log P(S_n > na) \leq \inf_{t>0} (\log \phi(t) - at)$$

à condition de convenir que  $\phi(t) = \infty$ , là où  $\phi(t)$  n'est pas définie.

La formule de Chernoff consiste à obtenir une égalité à la place d'une majoration mais elle est simplement asymptotique. La fonction convexe  $h(a) = \sup_{t>0} at - \log \phi(t)$  dont nous étudierons ci-dessous les propriétés, s'appelle la fonctionnelle de Cramer de (la loi de)  $X$ . Moyennant une définition correcte, elle permet de retrouver la loi de  $X$  tout comme  $\phi$ . La passage de  $\log \phi$  à  $h$  est une dualité classique sur les fonctions convexes dont nous rappellerons quelques points.

Enfin, il est essentiel de noter que la formule de Chernoff est conséquence de la seule existence des moments exponentiels et d'une certaine stationnarité. Pour cela, nous nous libérerons de l'hypothèse,  $S_n$  somme de variables indépen-

dantes et équidistribuées pour la remplacer par  $Z_n$  est une suite de variables aléatoires (d'abord réelles), telle que :

$$(1) \quad \phi_n(t) = E \exp t Z_n < \infty \quad \text{pour} \quad t \in ]-M_1, M_2[, \quad M_1, M_2 > 0$$

et telle que :

$$(2) \quad \frac{\log \phi_n(t)}{n} \longrightarrow L(t), \quad (n \rightarrow \infty), \quad t \in ]-M_1, M_2[$$

soit  $\phi_n^{1/n}(t) \rightarrow \exp L(t)$ .

Cette dernière hypothèse impliquant évidemment que :

$$\frac{1}{n} E Z_n \longrightarrow L'(0)$$

(2) traduit la stationnarité minimale dans le problème et on cherchera à montrer que :

$$\lim_n \frac{1}{n} \log P(Z_n > na) = \inf_{t>0} (L(t) - at)$$

formule dite de Chernoff. Les étapes suivantes seront d'abord des études de cas particuliers et le passage à des variables à valeurs dans les espaces euclidiens ou fonctionnels, l'évènement  $(Z_n > na)$  étant généralisé en  $(\frac{Z}{n} \in G)$ ,  $G$  convenable.

## II - Dualité et fonctions convexes

Soit  $f$  une fonction convexe sur  $R$  à valeurs dans  $R \cup \{+\infty\}$ .

La duale  $\hat{f}$  de  $f$  est définie par :

$$\hat{f}(t) = \sup_{x \in R} tx - f(x)$$

$\hat{f}$  est évidemment convexe, à valeurs dans  $R \cup \{+\infty\}$ . Dans le cas où  $f$  est dérivable, alors  $\hat{f}$  est dérivable, et  $\hat{f}'$  est la fonction réciproque  $(f')^{-1}$  de  $f'$ .

Les résultats de convergence suivants sont faciles (pour des fonctions de



variables réelles). On note  $Df = \{x, f(x) < \infty\}$  :

1) Soit  $f_n$  une suite de fonctions convexes à valeurs dans  $R \cup \{\infty\}$ , convergeant vers  $f$  en tout point sauf peut-être les points de  $(Df)^*$ , frontière de  $Df$ . Supposons  $D^\circ f \neq \emptyset$ . ( $D^\circ f$  intérieur de  $Df$ ). Alors la convergence est uniforme sur tout compact de  $D^\circ f$  (et de  $R - \overline{Df}$ ), ( $\overline{Df}$  fermeture de  $Df$ ).

2) De plus  $\hat{f}_n(x) \longrightarrow \hat{f}(x)$  sauf peut-être en  $x \in (D\hat{f})^*$ .

3) Soit alors  $Z$  une variable aléatoire (non dégénérée).

$$\phi(t) = E \exp t Z$$

est convexe à valeurs dans  $R \cup \{\infty\}$  ;  $\log \phi(t)$  est convexe (inégalité de Hölder).

Il est immédiat de vérifier que :

$$\{t, \log \phi(t) < \infty\} = \text{support } Z$$

et de manière plus précise si  $b = \sup\{x, x \in \text{support } Z\}$ , on a :

$$\log \phi(b) < \infty \quad \text{Équivaut à} \quad P(Z = b) = 0.$$

Soit  $h$  la duale de  $\log \phi$ .

4) On a  $h(EZ) = 0$ ,  $h'(EZ) = 0$ ,  $h$  est croissante à droite de  $EZ$ .

5) Enfin, l'équation  $(\log \phi)'(t) = \frac{\phi'(t)}{\phi(t)} = x_0$  a une solution  $x$  si et seulement si  $x_0 \in \text{Support}^\circ Z$  ( $\phi$  étant dérivable sur  $\text{Support}^\circ Z$ ).

### III - Formule de Chernoff

Soit  $Z_n$  une suite de variables aléatoires.

On suppose que  $Z_n$  vérifie les trois hypothèses suivantes :

H 1)  $E \exp t Z_n = \phi_n(t) < \infty$  pour  $t \in ]-M_1, M_2[$

$$0 < M_1, \quad 0 < M_2.$$

$$H 2) \quad \phi_n^{1/n}(t) \longrightarrow \exp L(t), \quad \text{pour tout } t \in ]-M_1, M_2[$$

H 3) La suite  $Z_n$  n'est asymptotiquement dégénérée au sens où si  $C_n = \text{support } Z_n$ ,  $D_n = 1/n C_n$  (homothétique de  $C_n$ ), alors on suppose que

$C_n^\circ \neq \phi$ , et  $D = (\lim_n D_n)^\circ \neq \phi$ . (Dans le cas des variables indépendantes, équidistribuées  $X_i$ , H 1 et  $X_i$  non identiquement nulle implique

H 3).

Soit  $B = \text{int } L^{-1}(]M_1, M_2[)$ .

On désigne par  $P_t^n$  la loi définie par  $dP_t^n = \frac{e^{t Z_n}}{\phi_n(t)} dP$  et si  $f$  est

mesurable, on note  $E_t f(Z_n)$  pour  $E P_t^n f(Z_n)$ . Si  $a \in D$ , on sait d'après le § 2, qu'il existe  $t_a^n$  unique tel que  $\frac{1}{n} E_{t_a^n} Z_n = a$ .

En général,  $t_a^n \notin ]M_1, M_2[$ , cependant si  $a \in B$ , pour  $n$  assez grand  $t_a^n \in ]M_1, M_2[$

et :

$$\frac{1}{n} E_{t_a^n} Z_n = \frac{1}{n} \frac{\phi_n'(t_a^n)}{\phi_n(t_a^n)}$$

H 2) implique que  $\frac{1}{n} \log \phi_n(t) \longrightarrow L(t)$  et que toutes les dérivées de  $\frac{1}{n} \log \phi_n$  converge vers celles de  $L$  uniformément sur tout compact de  $] -M_1, M_2[$ .

Les duales  $h_n$  de  $\frac{1}{n} \log \phi_n$  convergent pour  $a \in D$ , uniformément sur tout compact de  $D$ .

Remarquons enfin que l'on déduit de la croissance stricte de  $\frac{1}{n} \frac{\phi_n'}{\phi_n}$  sur  $] -M_1, M_2[$  et de la convergence de  $\frac{1}{n} \frac{\phi_n'}{\phi_n}$  vers  $L'$  que  $t_a^n \longrightarrow t_a$ ,

$t_a \in ] -M_1, M_2[$ .

Si  $t > t_a$ , pour un  $\eta$  fixé,  $\eta > 0$  assez petit, on a pour  $n$  assez grand,  $\frac{1}{n} \frac{\phi_n'}{\phi_n}(t) > a - \eta$ .

Théorème

Si  $Z_n$  vérifie H 1- 2- 3, on a pour  $a \in B$ ,

$$\lim \frac{1}{n} \log P(Z_n > na) = -h(a), \quad h \text{ duale de } L.$$

Démonstration

$$\begin{aligned} P(Z_n > na) &= \phi_n(t) E_t e^{-tZ_n} 1_{(Z_n > na)} \\ &= \phi_n(t) e^{-nat} E_t e^{-nt\left(\frac{Z_n}{n} - a\right)} \end{aligned}$$

$$\begin{aligned} \frac{1}{n} \log P(Z_n > na) &> -at + \frac{1}{n} \log \phi_n(t) \\ &\quad + \frac{1}{n} \log E_t e^{-nt\left(\frac{Z_n}{n} - a\right)} 1_{\left(\frac{Z_n}{n} - a > \epsilon\right)} \end{aligned}$$

(pour tout  $\epsilon > 0$ )

$$\begin{aligned} &> -at + \frac{1}{n} \log \phi_n(t) \\ &\quad + \epsilon t + \frac{1}{n} \log P_t \left(\frac{Z_n}{n} - a > \epsilon\right) . \end{aligned}$$

Comme  $\lim_{n \rightarrow \infty} \frac{Z_n}{n} > a$   $P_t$  p.s. puisque  $t > t_a$ , on a  $P_t\left(\frac{Z_n}{n} - a > \epsilon\right) > \frac{1}{2}$  pour  $n$  assez grand, d'où le théorème, puisque  $\epsilon$  est arbitraire.

Remarque : Nous avons utilisé la loi des grand nombres, sous les mêmes hypothèses en utilisant le théorème limite centrale, on obtient un peu plus. Reprenons en effet la démonstration en faisant  $t = t_a^n$

$$\begin{aligned} \frac{1}{n} \log P(Z_n > na) &> -a t_a^n + \frac{1}{n} \log \phi_n(t_a^n) \\ &\quad + \frac{1}{n} \log E_{t_a^n} \exp \frac{Z_n - na}{\sqrt{n} \sigma_a} \sqrt{n} t_a^n \sigma_a 1_{\left(0 < \frac{Z_n - na}{\sqrt{n} \sigma_a} < \epsilon\right)} \end{aligned}$$

(ou  $\sigma_a^2 = E_{t_a^n} (Y - a)^2 \neq 0$ ),

$$> -a t_a^n + \frac{1}{n} \log \phi_n(t_a^n) + \frac{\epsilon t_a^n \sigma_a}{\sqrt{n}} + \frac{1}{n} \log H_1 \left(]0, \epsilon[ \right)$$

ou  $H_n ]0, \epsilon[ = P\left(0 < \frac{Z_n - na}{\sigma_a \sqrt{n}} < \epsilon\right)$

$$\begin{aligned} \log E_{t_a^n} \exp t \left( \frac{Z_n - na}{\sqrt{n} \sigma_a} \right) &= \log \phi_n \left( t_a^n + \frac{t}{\sigma_a \sqrt{n}} \right) - \log \phi_n \left( t_a^n \right) - \frac{at}{\sqrt{n} \sigma_a} \\ &= \frac{t^2}{2n \sigma_a^2} \left( \frac{\phi'}{\phi} \right)' \left( t_a^n \right) + \frac{t^3}{6n^{3/2} \sigma_a^3} \left( \frac{\phi'}{\phi} \right)'' \left( t_a^n + \alpha \frac{t}{\sigma_a \sqrt{n}} \right), \end{aligned}$$

$0 < \alpha < 1$ ,

or  $\frac{1}{n} \left( \frac{\phi'}{\phi} \right)' \left( t_a^n \right) \longrightarrow L''(t_a)$ ,  $a \in B$

et  $\left( \frac{\phi'}{\phi} \right)'' \left( t_a^n + \alpha \frac{t}{\sigma_a \sqrt{n}} \right)$  est uniformément bornée (pour  $n$  assez grand)

par  $2 \sup_{t \in K} |L''(t)|$ ,  $K$  compact de  $B$  incluant  $a$ .

Donc  $\frac{Z_n - na}{\sqrt{n} \sigma_a} \longrightarrow$  loi normale sous  $P_{t_a^n}$  et par suite  $H_n \left] 0, \varepsilon \right[ > C(\varepsilon)$

pour  $n$  grand ou  $C(\varepsilon) = \frac{1}{2} P(0 < G < \varepsilon)$ , ou  $G$  est normale, d'où la

proposition. Sous les hypothèses du théorème, on a pour tout  $\varepsilon > 0$ ,

$$\frac{1}{n} \log P(Z_n > na) > -h_n(t_a^n) + \frac{2\varepsilon t_a^n}{\sqrt{n}} \sigma_a^2 + \frac{C(\varepsilon)}{n}$$

en particulier dans le cas des variables indépendantes et équidistribuées

$$\frac{1}{n} \log P(Z_n > na) > -h(t_a) + \frac{\varepsilon t_a}{\sqrt{n}} \sigma_a^2 + \frac{2C(\varepsilon)}{n}$$

pour tout  $n$  grand.

Remarque : Au cours du chapitre suivant, nous aurons besoin du même théorème sous les hypothèses suivantes

$$\begin{aligned} \frac{1}{N} L_N(t) &\rightarrow L(t) \text{ pour } t \in [0, 1] \\ \text{et } \frac{Z_N}{N} &\text{ converge en } P_0 \text{ probabilité.} \end{aligned}$$

Il est immédiat de modifier les démonstrations que nous venons de faire pour les appliquer à ce cas.

FORMULE DE CHERNOFF POUR DES RAPPORTS DE VRAISEMBLANCE

D. DACUNHA-CASTELLE.

I. GENERALITES.

Soient  $P_0, P_1$  deux probabilités de densité  $f_0$  et  $f_1$  par rapport à une mesure  $\nu$ . On suppose  $P_0 \neq P_1$  et  $P_0$  équivalente à  $P_1$ .

Soit  $\Phi(t) = E_{P_0} \exp t \log \frac{f_1}{f_0}$ .

Pour  $t \in [0, 1]$ ,  $\Phi(t)$  est finie,  $0 \leq \Phi(t) \leq 1$  d'après l'inégalité de Holder. Si  $t \geq 0$ ,  $\Phi$  est en fait définie (et finie) sur  $[0, s[$ ,  $s = \text{ess sup} \log \frac{f_1}{f_0}$ , ( $s \geq 1$ ). En  $s$ ,  $\Phi$  est finie si et seulement si

$$P_0 \left( \log \frac{f_1}{f_0} = s \right) = 0.$$

On pose  $h(a) = \sup_{t \geq 0} at - \log \Phi(t)$

(fonction de Cramer du couple  $P_0, P_1$ ), et

$$K(P_0, P_1) = E_{P_0} \log \frac{f_0}{f_1} \cong \infty$$

$$K(P_1, P_0) = - E_{P_1} \log \frac{f_0}{f_1} \cong \infty.$$

$\Phi$  est analytique sur  $]0, 1[$  et

$$\lim_{t \downarrow 0} \Phi'(t) = -K(P_0, P_1)$$

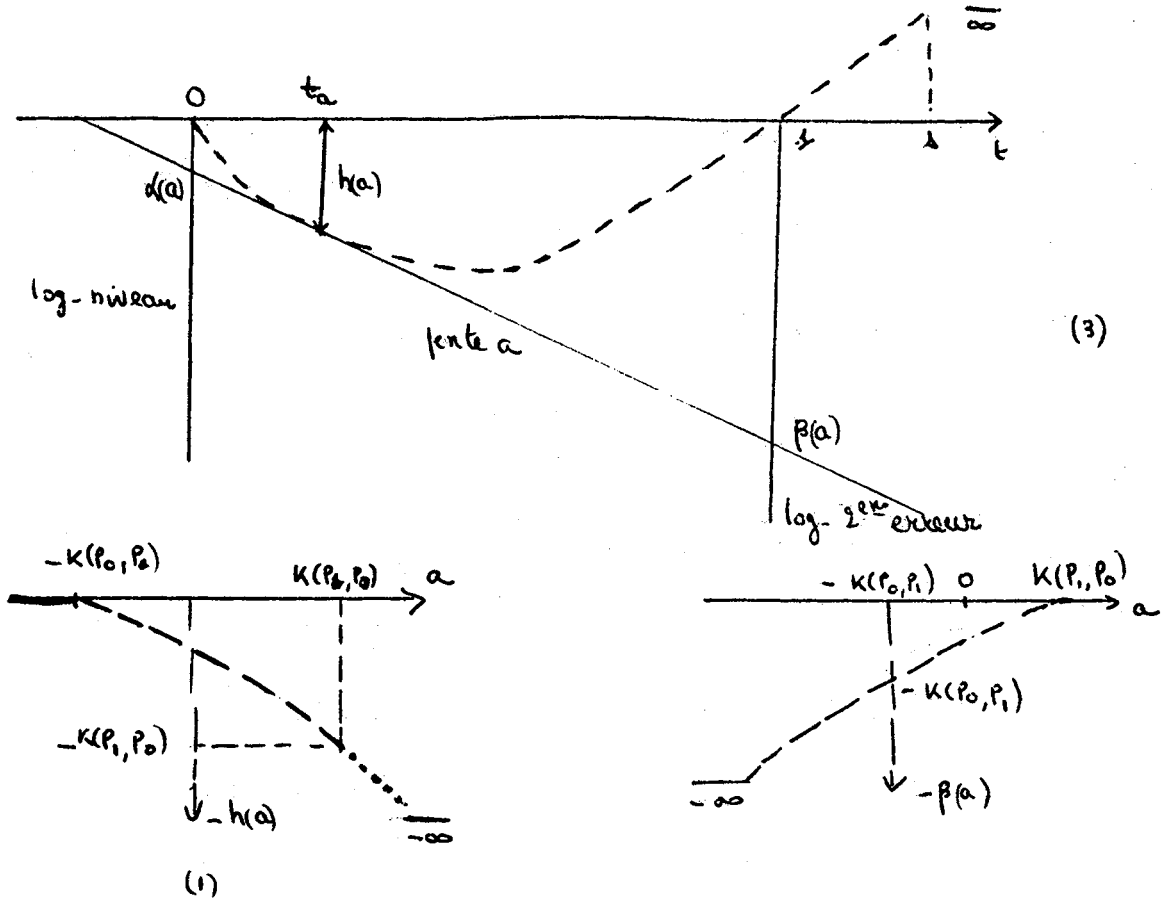
$$\lim_{t \uparrow 1} \Phi'(t) = K(P_1, P_0)$$

On a  $h(a) \rightarrow 0$  pour  $a \rightarrow -K(P_0, P_1)$ , et on définit  $h$  partout dans  $\bar{R}$ , en posant  $h(a) = 0$  pour  $a \leq -K(P_0, P_1)$ .

$h$  ainsi définie est convexe, strictement convexe et continue, sauf en  $s$  dans le cas où  $P_0 \left( \log \frac{f_1}{f_0} = s \right) > 0$ .

Enfin  $E_{P_1} \exp t \log \frac{f_1}{f_0} = \hat{\phi}(1-t)$  et donc si  $a \equiv K(P_1, P_0)$

$$-\sup_{t > 0} (-at - \log \hat{\phi}(1-t)) = a-h(a)$$



Soit  $f$  une fonction convexe sur  $\mathbb{R}$  à valeurs dans  $\mathbb{R} \cup \{+\infty\}$ .

La duale  $\hat{f}$  de  $f$  est définie par

$$\hat{f}(t) = \sup_{x \in \mathbb{R}} tx - f(x)$$

$\hat{f}$  est évidemment convexe, à valeurs dans  $\mathbb{R} \cup \{+\infty\}$ . Dans le cas où  $f$  est dérivable, alors  $\hat{f}$  est dérivable, et  $\hat{f}'$  est la fonction réciproque  $(f')^{-1}$  de  $f'$ .

Les résultats de convergence suivant sont faciles (pour des fonctions de variables réelles). On note  $Df = \{x, f(x) < \infty\}$ .

1) Soit  $f_n$  une suite de fonctions convexes à valeurs dans  $\mathbb{R} \cup \{\infty\}$ ,

convergeant vers  $f$  en tout point, sauf peut-être les points de  $(Df)^*$ , frontière de  $Df$ . Supposons  $\overset{\circ}{Df} \neq \emptyset$  ( $\overset{\circ}{Df}$  intérieur de  $Df$ ). Alors la convergence est uniforme sur tout compact de  $\overset{\circ}{Df}$  (et de  $R - \overline{Df}$ ), ( $\overline{Df}$  fermeture de  $Df$ ).

2) De plus  $\hat{f}_n(x) \rightarrow \hat{f}(x)$  sauf peut-être en  $x \in (\hat{Df})^*$ .

3) Soit alors  $Z$  une variable aléatoire (non dégénérée).  $\phi(t) = E \exp tZ$  est convexe à valeurs dans  $R \cup \{\infty\}$ .  $\log \phi(t)$  est convexe (inégalité de Hölder).

Il est immédiat de vérifier que

$$\overline{\{t, \log \phi(t) < \infty\}} = \text{support } Z$$

et de manière plus précise

si  $b = \sup \{x, x \in \text{support } Z\}$ , on a

$$\log \phi(b) < \infty \text{ équivaut à } P(z = b) = 0$$

Dans la suite interviendra de manière décisive la duale de  $\log \phi$ .

Soit  $(P_{0,n}, P_{1,n})_{n \in \mathbf{N}}$  une suite de couples de probabilités équivalentes sur  $R^n$ ,  $n \in \mathbf{N}$ . On dira que cette suite est de Chernoff si

$$\lim_{n \rightarrow \infty} \phi_n^{1/n}(t) = \exp L(t), \quad t \in [0,1]$$

où  $L$  est une fonction convexe sur  $[0,1]$ .

On a donc  $\frac{1}{n} \log \phi_n(t) = L(t)$ .

Si de plus  $\frac{1}{n} \phi'_n(0)$  converge, on appellera  $K(P_0, P_1)$  sa limite, qui sera définie comme l'information de Kullback de  $P_1$  par rapport à  $P_0$ .

Le lemme de vraisemblance au seuil  $\alpha$  de  $P_0$  contre  $P_1$  a pour région de rejet de  $P_0$ ,  $\left\{ \frac{1}{N} \log \frac{f_{1,N}}{f_{0,N}} > \alpha \right\}$  pour  $N$  fixé.

Le théorème du chapitre précédent s'écrit ici sous la forme suivante :

Théorème : Soit une suite  $(P_{0,n}, P_{1,n})$  de Chernoff, supposons  $K(P_0, P_1)$ ,  $K(P_1, P_0)$  définies et  $\alpha \in (-K(P_0, P_1), K(P_1, P_0))$ . Alors

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log P_{0,N} \left( \frac{1}{N} \log \frac{f_{1,N}}{f_{0,N}} > a \right) = -h(a)$$

où  $h$  est la duale de  $L$ ,  $L(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \log \phi_N(t)$ .

Ce théorème permet de régler asymptotiquement le seuil du test à un niveau  $\exp - N\alpha$ . On devra avoir  $h(a_N) \neq \alpha$ , et l'on aura  $a_N \rightarrow a$  tel que  $h(a) = \alpha$ .

Ce réglage étant fait, on peut calculer

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \log P_{1,N} \left( \frac{1}{N} \log \frac{f_{1,N}}{f_{0,N}} \cong a \right) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \log P_{1,N} \left( \frac{1}{N} \log \frac{f_{0,N}}{f_{1,N}} \cong a \right) \\ &= \beta \end{aligned}$$

tel que  $\alpha + \beta = a$ , d'après une formule vue plus haut. Le test de Neyman-Pearson au niveau  $\exp - n\alpha$  est donc tel que son erreur de 2ème espèce est  $\exp - \beta_n$ , avec  $\lim_{N \rightarrow \infty} \beta_n = a - \alpha$  avec  $\alpha = h(a)$ . (Voir dessin).

Il est immédiat d'appliquer ce résultat à une suite de variables indépendantes équidistribuées. Nous allons l'appliquer à un cas qui n'est pas traité complètement, dans la littérature, celui de deux processus gaussiens stationnaires.

## II. CAS DE DEUX PROCESSUS GAUSSIENS (d'après [9]).

[0] se réfère au chapitre I.

Les notations sont celles de [8] qui contient une bibliographie complète.

$F$  et  $G$  sont deux mesures gaussiennes centrées stationnaires sur  $\mathbb{R}^Z$ , et  $X^N$  désigne l'observation  $(X_1, \dots, X_N)$  du processus des coordonnées. On suppose que pour tout  $N$  assez grand,  $F$  et  $G$  ont des densités  $F_N, G_N$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^N$ . Pour qu'il en soit ainsi, il faut et il suffit que  $F$  et  $G$  aient des densités spectrales  $f, g$  telles que  $\log f, \log g \in L^1(T)$  (s'il n'en est pas ainsi, le problème statistique étudié ici est trivial).



On pose  $L_N(t) = \frac{1}{N} \log \phi_N(t)$

avec  $\phi_N(t) = E_F \exp t \log \frac{dG_N}{dF_N}$ ,  $t \in [0,1]$

On désigne par  $(T_N f)$ ,  $(T_N g)$  la suite des matrices de Toeplitz associées à  $f$  et  $g$ .

On a  $\log \frac{dG_N}{dF_N} = \frac{1}{2} \log \det T_N^{-1} g T_N f + \frac{1}{2} t X^N (T_N^{-1} f - T_N^{-1} g) X^N$

Comme la covariance de  $T_N^{-1/2} f X^N$  est  $I_N$  (identité sur  $\mathbb{R}^N$ ), on a en désignant par  $(\mu_i^N)_{i=1 \dots N}$  les valeurs propres de  $A_N = I_N - T_N^{1/2} f T_N^{-1} g T_N^{1/2} f$ ,

$$N L_N(t) = \frac{t}{2} \log \det T_N^{-1} g T_N f - \frac{1}{2} \sum_{i=1}^N \log (1 - \mu_i^N t).$$

Pour  $|t| < \frac{1}{\sup_{1 \leq i \leq N} |\mu_i^N|}$ , (N fixé), on a

$$\begin{aligned} \sum_{i=1}^N \log (1 - \mu_i^N t) &= - \sum_{p=1}^{\infty} \frac{t^p}{p} \text{Tr } A_N^p \\ &= - \sum_{\ell=1}^{\infty} (-1)^\ell \frac{1}{\ell} \text{Tr} (T_N f T_N^{-1} g)^\ell \left(\frac{t}{1-t}\right)^\ell + N \log (1-t) \end{aligned}$$

Soit  $z \in \mathbb{C}$ . Comme  $T_N f$ ,  $T_N g$  sont régulières on a, dans un voisinage ouvert de 0,  $T_N(zf + g)$  régulière, et donc [1],

$$\frac{d}{dz} \log \det T_N(zf + g) = \text{Tr } T_N f T_N^{-1} (zf + g)$$

$$\frac{d}{dz^\ell} \log \det T_N(zf + g) \Big|_{z=0} = (-1)^{\ell-1} (\ell-1)! \text{Tr} (T_N f T_N^{-1} g)^\ell$$

d'où

$$\sum_{i=1}^N \log (1 - \mu_i^N t) = - \sum_{\ell=1}^{\infty} \frac{1}{\ell!} \log \det T_N(zf + g) \Big|_{z=0}^{(\ell)} \cdot \left(\frac{t}{1-t}\right)^\ell + N \log(1-t)$$

De l'analyticit  au voisinage de 0 de la fonction  $z \rightarrow \log \det T_N(zf + g)$ ,

on tire

$$\sum_{i=1}^N \log (1 - \mu_i^N t) = \log \det T_N \left(\frac{t}{1-t} f + g\right) - \log \det T_N g + N \log (1-t)$$

d'où

$$N L_N(t) = \frac{t}{2} \log \det T_N^{-1} g T_N f - \frac{1}{2} \left[ \log \det T_N \left( \frac{t}{1-t} f + g \right) - \log \det T_N g + N \log(1-t) \right]$$

$$N L_N(t) = \frac{1}{2} \left[ t \log \det T_N f + (1-t) \log \det T_N g - \log \det T_N (tf + (1-t)g) \right] *$$

Cette formule est valable au voisinage de 0.

Mais la fonction  $z \rightarrow \log \det (z T_N f + (1-z) T_N g)$  est analytique dans un  $C$ -voisinage de tout point de  $]0,1[$ , car  $\log(tf + (1-t)g) \in L^1$  (par convexité) et donc  $T_N(tf + (1-t)g)$  est régulière pour  $t \in [0,1]$ .

Par suite, par prolongement analytique, la formule est valable pour  $t \in ]0,1[$  et par continuité pour  $t \in [0,1]$ . Appliquons maintenant le théorème élémentaire de Szegö :

$$\text{si } \log f \in L^1, \frac{1}{N} \log \det T_N f \rightarrow \frac{1}{2\pi} \int_T \log f(\theta) d\theta \text{ quand } N \rightarrow +\infty,$$

et donc :

Théorème : Si  $\log f, \log g \in L^1(T)$  et  $t \in [0,1]$

$$L_N(t) \rightarrow L(t) \quad (N \rightarrow \infty) \quad \text{avec}$$

$$L(t) = \frac{1}{4\pi} \int_{-\pi}^{\pi} [t \log f(\theta) + (1-t) \log g(\theta) - \log(tf(\theta) + (1-t)g(\theta))] d\theta$$

On en déduit, comme dans [0], la formule de Chernoff et les propriétés asymptotiques du lemme de Neyman-Pearson, moyennant la remarque faite en [0] sur le cas où  $M_1=0$  et la deuxième remarque ci-après.

Remarque: 1) Si  $f, g \in B$  (voir [8]), avec  $fg > 0$  on a

$$L_N(t) = L(t) + o\left(\frac{1}{N}\right).$$

Le même type de raisonnement que celui fait en [8] permet de calculer l'approximation quand  $f, g$  sont  $p$ -fois dérivables.

2) Sous ces hypothèses la formule de Chernoff démontrée au chapitre précédent vaut encore H3) remplaçant le fait que  $\phi_N(t)$  est définie sur un intervalle contenant 0 en son intérieur (la modification de démonstration est immédiate). On obtient donc les propriétés asymptotiques du lemme de Neyman-Pearson comme indiqué au § I.

Bibliographie :

- [ 1 ]            DAVIES  
                  Adv. App. Proba. 5, (1973), p. 469-497
  
- [ 2 ]            DZAPARIDZE  
                  Estimateurs asymptotiquement efficaces pour les paramètres du  
                  spectre d'une série chronologique gaussienne  
                  Publication de l'Université de Tbilissi, (1977), en russe.
  
- [ 3 ]            GRENANDER, SZEGÖ  
                  Toeplitz form and their applications  
                  Univ. of California press (1958)
  
- [ 4 ]            GUYON, PRUM  
                  Processus gaussiens sur  $Z^r$   
                  Publication d'Orsay (1977)
  
- [ 5 ]            HIRSCHMANN I.I. Jr.  
                  Recents developments in the theory of finite Toeplitz operators.  
                  Advances in Probability, vol. 1 (1971), p. 103-164
  
- [ 6 ]            PHAN-DINH  
                  Thèse, Grenoble (1975)
  
- [ 7 ]            PISSARENKO  
                  Theory of Probability and its applications, 10, n° 2 (1965),  
                  p. 299-305

- [ 8] D. DACUNHA-CASTELLE  
Remarque sur l'étude asymptotique du rapport de vraisemblance  
de deux processus gaussiens stationnaires  
Note aux C.R. Acad. Sc. Paris, Série A, février 1979 (à paraître)
- [ 9] J. COURSOL, D. DACUNHA-CASTELLE  
Sur la formule de Chernoff pour deux processus gaussiens  
stationnaires,  
Note aux C.R. Acad. Sc. Paris, Série A, mars 1979 (à paraître).

Didier DACUNHA-CASTELLE

Département Mathématique - Bât. 425  
E.R.A. CNRS 532 "Statistique Appliquée"  
Université Paris-Sud (Centre d'Orsay)  
91405 ORSAY.

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  $\Lambda$  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 étroite, 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 Lusinién, 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 due 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  $\sigma$ -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 alors homéomorphe à un  $G_\delta$  d'un compact métrisable.

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

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

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

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

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

$$\text{si } f_n \in \mathcal{E}_b(\mathcal{X}), f_n \geq 0, \text{ alors } I(f_n) \longrightarrow 0.$$

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  $\Lambda$  de  $P(\mathcal{X})$  soit relativement compacte, il suffit qu'elle soit équitendue, c'est-à-dire, qu'à tout  $\epsilon > 0$ , on puisse associer

un compact  $K_\varepsilon$  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  $\phi$  une fonction de  $\mathbb{R}^+$  dans  $\mathbb{R}^+$ , telle que  $\phi(x)/x$  tende vers  $+\infty$  quand  $x$  tend vers  $+\infty$ . Soit  $\Lambda$  une famille de Probabilités absolument continues par rapport à une même  $\mu$ . Alors :

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

(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  $\mathcal{E}_b(\mathcal{X})$ , telle que :

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

Notons enfin un résultat :

Lemme 1.1. : Soit  $\{\mathcal{X}, \tau, \mathcal{B}\}$  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  $\tau'$  sur  $\mathcal{X}$  engendrant la même  $\sigma$ -algèbre Borélienne, telle que  $\{\mathcal{X}, \tau', \mathcal{B}\}$  soit Lusinien, et que les  $f_n$  soient continues pour  $\tau'$ .

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

Références : 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  $\mathfrak{X}$  dans  $\mathbb{R}$ . Si  $\lambda$  et  $\mu$  sont deux Probabilités, l'information de Kullback de  $\lambda$  par rapport à  $\mu$  est définie par :

$$(2.1) \quad I(\lambda, \mu) = \sup_{u \in U} \{E_{\lambda}(\text{Log } u) - \text{Log}(E_{\mu}(u))\} .$$

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

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

(2.3) Elle est convexe et semi-continue inférieurement de son argument  $\lambda$ .

(2.4) Si  $g$  est mesurable de  $\mathfrak{X}$  dans  $\mathcal{Y}$  ;  $I(g_{\circ}\lambda, g_{\circ}\mu) \leq I(\lambda, \mu)$

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

Démonstration

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

Si  $\lambda \not\ll \mu$ , 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  $\lim_n V(\lambda, u_n) \geq \lambda(K) \text{Log}(M+1)$ , de sorte qu'alors  $I_1(\lambda)$  est infinie.

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

$$I_1(\lambda) \geq \lim V(\lambda, u_n) = E_{\lambda}(\text{Log } P_{\varepsilon, M}) - \text{Log}(E_{\mu}(P_{\varepsilon, M})) .$$



Passant à la limite en  $\varepsilon$  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 :

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

d'après l'inégalité de Jensen appliqué à la fonction  $\text{Log}$ . 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  $E_\mu(u_n \text{Log } u_n)$  tende vers  $I_2(\lambda)$ ,  $u_n$  tende vers  $\frac{d\lambda}{d\mu}$  dans  $1L^1(\mu)$ , et  $E_\mu(u_n) = 1$ , cette dernière condition par multiplication par une constante. Si  $\lambda_n$  est définie par  $d\lambda_n = u_n \cdot d\mu$ , comme  $\lambda_n$  tend vers  $\lambda$ , 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  $\mathcal{O}$  est la  $\sigma$ -algèbre de  $\mathfrak{X}$  engendrée par  $g$ , la Probabilité de densité  $E_\mu^{\mathcal{O}} \left( \frac{d\lambda}{d\mu} \right)$  a une information égale à  $I(g_\circ \lambda, g_\circ \mu)$  (d'après le Théorème de transport) et moindre que  $I(\lambda, \mu)$  d'après Jensen appliquée à  $x \longrightarrow x \text{Log } x$ .

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

§ 3. Transformée de Cramer

Dans ce chapitre,  $\mathcal{X}$  est un espace de Banach séparable, de dual  $\mathcal{X}^*$ , la forme bilinéaire est notée  $(\theta, a)$  ( $\theta \in \mathcal{X}^*$ ,  $a \in \mathcal{X}$ ). On s'intéresse particulièrement aux Probabilités  $\mu$  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  $\mu$  est alors définie par :

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

C'est une fonction positive et convexe, en tant qu'enveloppe supérieure de fonctions affines ; remarquons que  $(\theta, a) - \text{Log } E \exp(\theta, X)$  est concave de son argument  $\theta$ . 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  $\mu+b$  est la translatée par  $b$  de  $\mu$ .

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

Proposition 3.1. (Cramer-Chernov)

Soit  $S_n$  la somme de  $n$  v.a.r. indépendantes de même loi  $\mu$  satisfaisant

(3.1.) . Alors :

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

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  $\theta$ , écrivons  $t$  (réel) :  $h(b) = \text{Sup}_{t \in \mathbb{R}} bt - \text{Log } E e^{tX}$  .

$bt - \text{Log } E (e^{tX})$  étant concave de  $t$ , de dérivée en  $0$  :  $b - E(X)$ , on remarque que  $h(E(X)) = 0$ , donc que  $\text{Inf}_{b \geq 0} h(b) = 0$  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,  $\lim_{|a| \rightarrow \infty} \frac{h(a)}{|a|} = +\infty$ , 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  $\mu$  satisfaisant (3.1). Il existe alors une fonction :

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

telle que :

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

Démonstration

Remplaçons  $\mu$  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  $\mu$ . Posons  $\psi(z) = \sqrt{z \cdot g(z)} \cdot \psi(z)/z \rightarrow \infty$  et  $\psi(z)/g(z) \rightarrow 0$  quand  $z \rightarrow \infty$ , d'après la remarque plus haut. Soit alors  $\phi$  la plus grande fonction convexe sous  $\psi$  (i.e. le sup des fonctions affines moindres de  $\psi$ ) ;  $\phi$  possède encore ces deux propriétés. Maintenant,

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

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

Proposition 3.3.

Soit  $\mu$  satisfaisant (3.1),  $\phi$  une fonction satisfaisant (3.5),

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

Alors  $\sup_{\lambda \in \mathcal{K}_a} E_\lambda(\phi(\|X\|))$  est fini.

Démonstration

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

Corollaire

Soit  $\mu$  satisfaisant (3.1),  $a$  réel,  $\theta$  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  $\mathcal{K}_{a,T} = \bigcap_{\theta \in T} \mathcal{K}_{a,\theta}$ , pour toute partie  $T$  de  $\mathcal{X}^*$ .

Démonstration

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

$$E_\lambda |f_M(X) - E_\lambda(\theta, X)| \leq \frac{M}{\|\theta\|} \cdot \left\{ \phi\left(\frac{M}{\|\theta\|}\right) \right\}^{-1} \cdot \sup_{\lambda \in \mathcal{K}_a} E_\lambda(\phi(\|X\|)) ;$$

Uniformément sur  $\mathcal{K}_a$ , le membre de gauche tend donc vers 0 quand  $M$  tend vers l'infini, puisque  $\phi(x)/x$  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 mesure empirique associée au n-échantillon

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

Proposition 3.4.

Soit  $\mu$  satisfaisant (3.1) ,  $\phi$  une fonction satisfaisant (3.5).

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

$$(3.6) \quad P_{\mu}(\hat{\mu}_n \notin \mathcal{K}(a)) \leq 2 \cdot a^n , \quad \text{et} \quad \sup_{\lambda \in (a)} E_{\lambda}(\phi(|X|)) \leq M .$$

Démonstration

Soit  $A_M = \{\lambda \mid E_{\lambda}(\phi(|X|)) > M\}$  ;  $\{\hat{\mu}_n \in A_M\} = \{\sum_{i < n} \phi(|X_i|) > nM\}$  ,

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

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

On va maintenant construire un ensemble  $\Lambda_a$  tel que :

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

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  $Y_i = 1_{X_i \in K^c}$  ,  $\hat{\mu}_n(K^c)$  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 :  $d \log(\frac{d}{b}) + (1-d) \log(\frac{1-d}{1-b})$  , soit :

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

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 ,  $\Lambda_a$  est relativement compact (d'après (1.2)) et

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

§4. Calcul de la transformée de Cramer

Théorème 1

Soit  $\mathfrak{X}$  un Banach séparable,  $\mu$  une Probabilité sur  $\mathfrak{X}$  satisfaisant (3.1).

Pour tout  $t$ ,  $E_{\mu}(\exp(t||X||))$  est fini. Alors :

$$(4.1) \quad h_{\mu}(a) = \inf_{\lambda | \lambda \in P(\mathfrak{X}); E_{\lambda}(X)=a} I(\lambda, \mu) .$$

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  $h \leq I$ .

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

$$- \text{Log } E_{\mu}(\exp tX) \leq I(\lambda, \mu) \quad \text{or}$$

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

(La seconde inégalité provenant de Jensen appliqué à  $-\text{Log}$ ).

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

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

1. Ramenons le cas général au cas de dimension finie : soit  $T$  une partie finie de  $\mathfrak{X}^*$ ,  $S(T)$  le sous-espace linéaire engendré,  $\phi_T$  l'application naturelle de  $\mathfrak{X}$  dans  $S(T)^*$ , identifié dans la suite à un espace  $\mathbb{R}^m$ , et posons

$$h_T = \sup_{\theta \in S(T)} - \text{Log } E_{\mu}(\exp(\theta, X)) \quad (\leq h, \text{ donc finie}).$$

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  $\phi_T \circ \mu$  (voir remarque finale). Soit  $\lambda^T$  la Probabilité sur  $\mathfrak{X}$  définie par  $\frac{d\lambda^T}{d\mu}(X) = \frac{dQ^T}{d\phi_T \circ \mu}(\phi_T X)$ . Les  $\mathcal{K}_{h,t}$  du Corollaire de la Probabilité 3.3. forment donc une famille filtrante de compacts non vides, et toute mesure de leur intersection est centrée; et d'information moindre que  $h$ .

2. Montrons le résultat dans  $\mathbb{R}^m$ . Si  $\theta$  est une forme linéaire, notons  $H_\theta^-, H_\theta, H_\theta^+$  les images réciproques par  $\theta$  de  $\{x < 0\}, \{x = 0\}, \{x > 0\}$ .

Trois cas se séparent :

a. Cas intérieur : 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  $\mu$ , la transformée de Laplace tend vers  $+\infty$  quand  $t \rightarrow \pm\infty$ . Le convexe  $\{\theta \mid E_\mu \exp(t(\theta, X)) \leq 1\}$  est donc borné dans toutes les directions, donc compact, autrement dit l'inf est atteint, et il existe une  $\theta_0$  telle que  $h = -\text{Log } E_\mu \exp(\theta_0, X)$ . La Probabilité de densité  $\exp(h + (\theta_0, X))$  (par rapport à  $\mu$ ) 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  $\theta$  telle que  $\mu(H_\theta^-) = 1$ . Ce cas ne peut se produire si  $h$  est finie, car  $\inf_\theta E_\mu \exp(\theta, X) \leq \lim_{t \rightarrow \infty} E_\mu \exp t(\theta_0, X) = 0$ .

c. Il existe  $\theta_0 \neq 0$  telle que  $p = \mu(H_{\theta_0}) > 0$  et  $\mu(H_{\theta_0}^+) = 0$ . En changeant de base, on peut supposer que  $H_{\theta_0} = \{x_m = 0\}$ , et on décompose  $\mu$  en  $\mu^- + p \cdot \gamma$ , où  $\mu^-$  est une sous-probabilité portée par  $\{x_m < 0\}$ ,  $\gamma$  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  $h = -\text{Log } p + h_\gamma(0)$ . Par ailleurs, toute mesure  $\lambda'$  portée par  $\{x_m = 0\}$

et centrée en  $X'$  est centrée en  $X$ , et d'information  $I(\lambda', \mu) = -\text{Log } p + I(\lambda', \nu)$ .  
 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  $\nu$  est  $\delta_0$ , les dernières  $\lambda'$  coïncident avec  $\nu$ , et  $h_{\delta_0}(0) = 0 = I(\lambda', \nu)$ , de sorte que la résultat est démontré.

Remarque

La fonction  $I(\lambda, \mu)$  étant s.c.i. de  $\lambda$  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-Chernov "ouvert" en dimension finie)

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

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

Démonstration

On peut trouver un point  $b$  centre d'une boule ouverte contenue dans  $G$ , "intérieur" au support de  $\mu$  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  $\theta$  réalisant le sup, soit :

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

On note  $\lambda_b$  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  $\lambda_b$  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-Chernov

Théorème 2

Soit  $\mathcal{X}$  un espace Lusinien ou Polonais,  $\mu$  une Probabilité sur  $\mathcal{X}$ ,

$\mathcal{F}$  un fermé de  $P(\mathcal{X})$ ,  $\mathcal{G}$  un ouvert de  $P(\mathcal{X})$ . On a :

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

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

Théorème 3

Soit  $\mathcal{X}$  un espace de Banach séparable,  $F$  un fermé,  $G$  un ouvert de  $\mathcal{X}$ ,

$\mu$  une Probabilité sur  $\mathcal{X}$  satisfaisant :

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

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

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

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 fini, on a :

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

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  $\Lambda_a$  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)

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

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

$\mathcal{F}$  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 fini 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 :

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

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.,

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

Maintenant, toute  $Q$  d'information finie par rapport à  $f \circ \mu$  et telle que  $E_Q(x) \geq b$  peut se relever en une  $\lambda$  de même information sur  $\mathcal{X}$ , par exemple

$$d\lambda = \frac{dQ}{df \circ \mu} \circ f \circ d\mu. \text{ Cette } \lambda \text{ ne peut appartenir à } \mathcal{K}_\alpha, \text{ puisque}$$

$$E_\lambda f = E_Q x \geq b, \text{ et donc :}$$

$$\inf_{Q | E_Q(x) \geq b} I(Q, f \circ \mu) \geq \alpha,$$

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 :

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

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  $\{x \mid |x_j - E f_j| < r_j ; j = 1, 2, \dots, m\}$ .

On a alors  $\{\hat{\mu}_n \in \mathcal{G}\} \supset \{E_{f \circ \hat{\mu}_n}(x) \in G_m\}$ ;  $E_{f \circ \hat{\mu}_n}(x)$  peut s'interpréter comme la moyenne arithmétique  $\frac{S}{n}$  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  $\{Q \mid E_Q(x) \in G_m\}$  contient  $\{f \circ \lambda \mid \lambda \in \mathcal{K}\}$ , que

$$\inf_{Q \mid E_Q(x) \in G_m} I(Q, f \circ \mu) \leq I(\nu, \mu)$$

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  $T$  est continue de  $\mathcal{K}(a)$  dans  $\mathcal{X}$  (topologie étroite sur  $\mathcal{K}(a)$ , topologie forte sur  $\mathcal{X}$ ). Soit en effet  $S = \{(\lambda_n) ; \lambda\}$  où  $\lambda_n$  tend étroitement vers  $\lambda$  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  $\nu$  de  $S$ . Soit  $\nu_K = 1_K \cdot \nu + \nu(K^c) \cdot \delta_0$ .

Alors :

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

pour  $\nu \in S$ ,  $R$  réel positif,  $\phi$  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  $\varepsilon$ , puisque  $\phi(x)/x$  tend vers l'infini à l'infini.

Soit  $S_K = \{\nu_K \mid \nu \in S\}$ .  $K$  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 compact sur lequel topologie faible et forte coïncident.

Terminons la démonstration de (5.3) :

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

On applique donc (5.5), la majoration (3.6), enfin (5.1) au fermé

$$\mathcal{F} = T^{-1}(F) \cap \mathcal{K}(a) . \text{ Il vient :}$$

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

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,  $\lambda$  de moyenne  $a$  et d'information  $h_{\mu}(a)$ , on l'approche par  $\lambda'$  à 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 :

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

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

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

§6. Deux exemples

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

$$(6.1) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{\mu} \{ \hat{\mu}_n \in \Omega \} = -I_{\mu}(\Omega) ,$$

où naturellement :

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

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

$$(6.2) \quad I(\bar{\Omega}) = I(\Omega) = I(\overset{\circ}{\Omega}) .$$

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 (denombrable) 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  $l_{-\infty, d]}$  et les  $l_{-\infty, d[$  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  $\Delta$  est atteinte (soit par le couple  $F, G$ , soit par le couple  $F^{-}, G^{-}$ ). Naturellement, sur  $P(\mathbb{R})$ , la  $\Delta$ -topologie est plus fine que la topologie étroite usuelle (c'est-à-dire celle de la convergence en loi) mais :

Lemme 6.1.

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

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

Démonstration

Fixons  $n$  ; soit  $x_{i,n} = \text{Sup}\{x \mid G(x) \leq \frac{i}{n}\}$  , 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  $\mathcal{D}$  tel que

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

soient moindres que  $\frac{1}{n}$  . Soit  $g_n$  l'ouvert étroit  $\{H \mid \text{Max}_i (\text{Max}_i |G-H|(y_{i,n}) , \text{Max}_i |G^-H^-(y_{i,n})) < \frac{1}{n}\}$  (il est ouvert comme intersection finie d'ouverts élémentaires).

La monotonie de  $G$  et  $H$  assure que si  $H \in g_n$  ,  $\Delta(G,H) \leq \frac{3}{n}$  , d'où le résultat.

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

Corollaire

Soit  $\Omega$  un ensemble  $\Delta$ -fermé ,

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

Démonstration

On prend  $\mathcal{D} = \mathcal{Q} \cup D(F)$  . Si  $G$  adhère étroitement à  $\Omega$  , et si elle est d'information finie , elle est absolument continue par rapport à  $F$  , donc  $D(G) \subset D(F) \subset \mathcal{D}$  ; et d'après le lemme, elle  $\Delta$ -adhère à  $\Omega$  , donc elle est dans  $\Omega$  . 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  $\Omega$  , 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  $\Omega$  qui assurent que  $\Omega$  est un bon ensemble. Une lecture attentive de ces conditions portant sur les partitions finies de  $\Omega$  , 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  $\mathcal{X}$  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  $\Delta$ . 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  $0$ ,  $\Omega_0$  est un bon ensemble pour  $F$ .

Démonstration

Clairement,  $\Omega_0$  est  $\Delta$ -fermé,  $\overset{\circ\circ}{\Omega}_0$  contient  $\Omega_r$  pour  $r > 0$ , donc  $I(\overset{\circ\circ}{\Omega}_0) = I(\Omega_0)$ . 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  $I(\Omega_0)$ ).

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

Théorème

Soit  $\Omega_a = \{G | \Delta(F,G) \geq a\}$ . Alors  $\Omega_a$  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,  $\Omega_a$  étant  $\Delta$ -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  $\underline{p} \geq 0$  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  $\Omega_b$  et son information  $I(P,p)$  vaut

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

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

(convexité de  $x \rightarrow x \operatorname{Log} x$ ).  $I(P,p)$  est croissante et continue de  $P$  sur  $[p, 1]$  . Montrons maintenant que pour toute  $G$  de  $\Omega_a$  , et d'information finie,  $\liminf \frac{1}{n} \operatorname{Log} P_{\hat{F}_n} \{ \hat{F}_n \in \Omega_a \} \geq -I(G,F)$ . 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  $\Delta$ ) de  $\Omega_a$  , et le raisonnement du Corollaire montre alors que  $\liminf \frac{1}{n} \operatorname{Log} \dots$  (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 égale 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 calculer  $I(a)$  , évidemment distribution free 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.



GRANDES ET MOYENNES DEVIATIONS POUR LES

MARCHES ALEATOIRES.

D. PICARD - J. DESHAYES.

On considère une suite de vecteurs aléatoires indépendants et équidistribués :  $\xi_1, \dots, \xi_n$  de l'espace euclidien  $R^k$  (toutes les normes étant équivalentes, on pourra utiliser la norme :  $\rho(x) = \text{Sup}_{i=1, \dots, k} |x^i|$ )

$$E\xi = 0$$

$$E\xi\xi' = B \text{ avec } \det B \neq 0.$$

On pose  $S_0 = 0$

$$S_j = \xi_1 + \dots + \xi_j, \quad j = 1, 2, \dots, n$$

On définit la ligne polygonale  $s_n(t)$ ,  $t \in [0, 1]$  dans  $R^k$  construite sur les points  $(\frac{j}{n}, \frac{S_j}{x(n)})$  pour  $j = 0, 1, \dots, n$ ; où  $x(n)$  désigne une suite de nombres positifs telle que  $\frac{x(n)}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} \infty$  et  $x(n) = o(n)$ .

On s'intéressera au comportement asymptotique de  $s_n(t)$ ; pour cela on se place dans l'espace  $\mathcal{C}_k[0, 1]$  des fonctions continues de  $[0, 1]$  dans  $R^k$  muni de la norme uniforme notée  $\| \cdot \|$ ; et on étudiera  $P(s_n(\cdot) \in G)$  où  $G$  est un borélien de  $\mathcal{C}_k[0, 1]$ .

Dans la partie I, on introduit la fonctionnelle  $w$  qui joue pour les marches aléatoires le rôle de la transformée de Cramer dans les théorèmes de grandes déviations. En première lecture, on pourra omettre les propriétés de  $w$ . Après avoir énoncé les théorèmes dans la partie II, des exemples d'applications sont mis en évidence dans la partie III. La partie IV est consacrée aux démonstrations et dans la partie V, on envisage le cas où la transformée de Laplace de  $\xi$  existe seulement dans un voisinage de 0.

Cet exposé a été construit d'après des résultats de Borovkov [2] et Mogulskii [3].

I. PRELIMINAIRES ET NOTATIONS.

On désigne par  $\varphi_\xi(\cdot)$  la transformée de Laplace de la variable aléatoire  $\xi$  et par  $h_\xi(\cdot)$  sa transformée de Cramer :

$$\lambda \in \mathbb{R}^k \rightsquigarrow \varphi_\xi(\lambda) = E e^{\langle \lambda, \xi \rangle}$$

$$\alpha \in \mathbb{R}^k \rightsquigarrow h_\xi(\alpha) = \sup_{\lambda \in \mathbb{R}^k} \{ \langle \alpha, \lambda \rangle - \log \varphi_\xi(\lambda) \}$$

On note  $\mathcal{E}_0$  l'ensemble des fonctions de  $\mathcal{C}_k[0,1]$  partant de 0 en  $t=0$  et  $\Gamma$  l'ensemble des lignes polygonales (ayant un nombre fini de sommets) de  $\mathcal{E}_0$ , ces deux ensembles étant munis de la topologie trace.

Les propriétés asymptotiques seront énoncées à l'aide de la fonctionnelle  $w_\xi$  définie par :

$$f \in \Gamma \rightsquigarrow w_\xi(f) = \int_0^1 h_\xi(f'(t)) dt$$

Proposition.  $w_\xi$  est une fonctionnelle semi-continue inférieurement sur  $\Gamma$ .

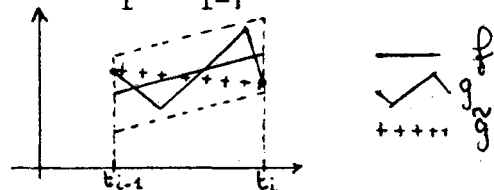
démonstration : soit  $f \in \Gamma$  de sommets  $(t_i, f(t_i))$ ,  $i = 0, 1, \dots, j$ .

Pour un nombre  $\varepsilon$  positif, on notera  $B(f, \varepsilon)$  l'ensemble des fonctions  $g$  de  $\mathcal{C}_k[0,1]$  telles que  $\|g - f\| < \varepsilon$

La fonction  $h_\xi$  étant s.c.i. (cf [1]), on en déduit que  $\forall \eta > 0$ ,  $\exists \varepsilon > 0$  tel que  $g \in B(f, \varepsilon) \cap \Gamma$ ,  $h_\xi\left(\frac{g(t_i) - g(t_{i-1})}{t_i - t_{i-1}}\right) \geq h_\xi\left(\frac{f(t_i) - f(t_{i-1})}{t_i - t_{i-1}}\right) - \eta$

pour  $i = 1, 2, \dots, j$ . Par intégration, on obtient  $w_\xi(\tilde{g}) \geq w_\xi(f) - \eta$ , où  $\tilde{g}$  désigne la ligne polygonale de sommets  $(t_i, g(t_i))$ ,  $i = 0, 1, \dots, j$ .

La fonction  $h_\xi$  étant convexe, on peut appliquer l'inégalité de Jensen :  $\frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} h_\xi(g'(t)) dt \geq h_\xi\left(\frac{g(t_i) - g(t_{i-1})}{t_i - t_{i-1}}\right)$  ce qui donne

$$\int_0^1 h_\xi(g'(t)) dt \geq w_\xi(\tilde{g}) \geq w_\xi(f) - \eta$$


La fonctionnelle  $w_\xi$  étant s.c.i. sur l'ensemble  $\Gamma$  qui est dense dans  $\mathcal{E}_0$ , on la prolonge naturellement en une fonction s.c.i. sur  $\mathcal{E}_0$  en posant :

$$f \in \mathcal{E}_0 \rightsquigarrow w_\xi(f) = \inf_{\{f_n\} \in \Gamma} \lim w_\xi(f_n)$$

$$\|f_n - f\| \rightarrow 0$$

Remarque 1 : Si  $\varphi_\xi$  n'existe pas partout mais seulement dans un voisinage de 0,  $w_\xi(f)$  coïncide avec  $\int_0^1 h_\xi(f'(t)) dt$  pour toutes les fonctions  $f$  de  $\mathcal{C}_0$  telles que  $f'$  soit continue à valeurs à l'intérieur du domaine de définition de  $h_\xi$ .

Remarque 2 : Si  $\varphi_\xi$  existe partout, alors la fonctionnelle  $w_\xi$  coïncide avec  $v_\xi$  définie sur  $\mathcal{C}_0$  par  $v_\xi(f) = \int_0^1 h_\xi(f'(t)) dt$  pour  $f$  absolument continue

$$v_\xi(f) = +\infty \text{ sinon.}$$

Nous allons le montrer à l'aide des deux lemmes suivants :

lemme 1 :  $\forall f \in \mathcal{C}_0, w_\xi(f) \leq v_\xi(f)$

démonstration : si  $v_\xi(f) = +\infty$ , c'est trivial.

si  $f$  est absolument continue, elle admet un module de continuité uniforme  $\delta$ . Désignons par  $\tilde{f}_n$  la ligne polygonale de sommets

$$\left(\frac{j}{n}, f\left(\frac{j}{n}\right)\right)_{j=0, \dots, n}$$

On a :  $v_\xi(f) \geq v_\xi(\tilde{f}_n) = w_\xi(\tilde{f}_n)$  car  $h_\xi$  est convexe.

$$\begin{aligned} \|\tilde{f}_n - f\| &= \sup_{t \in [0,1]} \rho |\tilde{f}_n(t) - f(t)| \\ &\leq \sup_t \{ \rho [\tilde{f}_n(t) - \tilde{f}_n(\frac{[nt]}{n})] + \rho [\tilde{f}_n(\frac{[nt]}{n}) - f(t)] \} \\ &\leq \sup_t \{ \rho [\tilde{f}_n(\frac{[nt]+1}{n}) - \tilde{f}_n(\frac{[nt]}{n})] + \rho [f(\frac{[nt]}{n}) - f(t)] \} \leq 2 \delta(\frac{1}{n}) \end{aligned}$$

$$\|\tilde{f}_n - f\| \rightarrow 0 \Rightarrow v_\xi(f) \geq \underline{\lim} v_\xi(\tilde{f}_n) \geq w_\xi(f).$$

lemme 2 : si  $\varphi_\xi$  existe partout,  $v_\xi$  est s.c.i. Alors on en déduit que pour  $f \in \mathcal{C}_0$

$$v_\xi(f) \leq \inf_{\{f_n\} \in \mathcal{C}_0} \underline{\lim} v_\xi(f_n) \leq \inf_{\{f_n\} \in \Gamma} \underline{\lim} v_\xi(f_n) = w_\xi(f).$$

$$\begin{aligned} \{f_n\} \in \mathcal{C}_0 & \quad \{f_n\} \in \Gamma \\ \|\tilde{f}_n - f\| \rightarrow 0 & \quad \|\tilde{f}_n - f\| \rightarrow 0 \end{aligned}$$

ce qui prouve, avec le lemme 1, l'identité de  $w_\xi$  et  $v_\xi$ .

démonstration : Pour montrer que  $v_\xi$  est s.c.i., on va prouver que pour tout  $K$  positif, les ensembles  $A_K = \{f \in \mathcal{C}_0, v_\xi(f) \leq K\}$  sont fermés. Soit  $f_n$  une suite de  $A_K$  telle  $\|f_n - f\| \xrightarrow{n \rightarrow \infty} 0$  ; montrons que  $f \in A_K$ .

$\int_0^1 h_\xi(f'_n(t)) dt \leq K$  et  $\lim_{\rho(x) \rightarrow \infty} \frac{h_\xi(x)}{\rho(x)} = +\infty$  entraînent que la suite  $f'_n$  est équi-intégrable ; donc elle est relativement compacte pour la topologie  $\sigma(L_1, L_\infty)$ .

Soient  $\gamma_1$  et  $\gamma_2$  deux valeurs d'adhérence de la suite  $f'_n$  pour  $\sigma(L_1, L_\infty)$ , on a

alors :  $f(t) = \int_0^t \gamma_1(u) du = \int_0^t \gamma_2(u) du$ , ce qui implique que  $f$  est absolument continue et  $\gamma_1 = \gamma_2 = f'$ . Donc  $f'_n \rightarrow f'$  pour  $\sigma(L_1, L_\infty)$ . En effet pour toute

fonction  $\eta$ ,  $\int_0^1 h_\xi(\eta(t)) dt \leq K$

$$\int_0^1 \langle g(t), \eta(t) \rangle dt - \int_0^1 \text{Log } \varphi_\xi(g(t)) dt \leq K, \forall g \in L^\infty.$$

Montrons le sens non trivial :

Remarquons d'abord que  $\int_0^1 \langle g, \eta \rangle dt - \int_0^1 \text{Log } \varphi_\xi(g(t)) dt \leq K, \forall g \in L^\infty$

implique :  $\int_0^1 [\langle g(t), \eta(t) \rangle - \text{Log } \varphi_\xi[g(t)]]^+ dt \leq K, \forall g \in L^\infty$ .

et par suite  $\int_0^1 |\langle g(t), \eta(t) \rangle - \text{Log } \varphi_\xi(g(t))|^+ dt \leq K, \forall g$

Prenons  $g(t) = \theta_\varepsilon(\eta(t))$  avec  $\theta_\varepsilon$  défini par :

$$\langle x, \theta_\varepsilon(x) \rangle - \text{Log } \varphi_\xi[\theta_\varepsilon(x)] = (1 - \varepsilon) h_\xi(x)$$

$$\Rightarrow \int_0^1 [(1 - \varepsilon) h_\xi[\eta(t)]]^+ dt \leq K$$

$\varepsilon$  arbitraire  $\Rightarrow \int_0^1 h_\xi[\eta(t)] dt \leq K$ .

Proposition. si  $\varphi_\xi$  existe partout, l'ensemble  $A_K = \{ f \in \mathcal{E}_0, w_\xi(f) \leq K \}$  est de plus compact dans  $\mathcal{E}_0$ .

démonstration : il suffit de trouver un module d'équicontinuité.

Ayant supposé l'existence de  $\varphi_\xi$  partout, nous pouvons considérer sa restriction à toutes les "diagonales" de  $\mathbb{R}^k$  :

Pour  $u$  croissant de 0 à  $+\infty$ ,  $\text{Log } \varphi_\xi(u, \delta)$  est une fonction croissante en  $u$  de 0 à  $+\infty$ , où  $\delta$  désigne l'un des  $2^k$  vecteurs de  $\mathbb{R}^k$  dont les composantes valent  $+1$  ou  $-1$  ; nous pouvons alors considérer la fonction  $\psi_\delta(u)$  inverse de  $\frac{1}{\text{Log } \varphi_\xi(\frac{\delta}{u})}$  : elle est définie pour  $u \in ]0, \infty[$ , croissante et tendant vers 0 en  $0^+.$

Soit  $\psi(u) = \sum_{\delta} \psi_\delta(u)$  définie pour  $u > 0$ , croissante et tendant vers 0 en 0.

Considérons maintenant une fonction  $g$  quelconque de  $A_K$  :

$$0 \leq y \leq x \leq 1 \quad \rho[g(x) - g(y)] \leq \sum_{i=1}^k |g^i(x) - g^i(y)| = \langle \delta, g(x) - g(y) \rangle \text{ pour l'un des } \delta.$$

$$g(x) - g(y) = \int_0^1 g'(t) 1_{[y,x]}(t) dt$$

$$\Rightarrow \frac{\rho[g(x) - g(y)]}{\psi(x-y)} \leq \frac{\langle \delta, g(x) - g(y) \rangle}{\psi_\delta(x-y)} = \int_0^1 \langle g'(t), \frac{\delta \cdot 1_{[y,x]}(t)}{\psi_\delta(x-y)} \rangle dt$$

Comme  $h_\xi$  et  $\text{Log } \varphi_\xi$  sont en dualité de Young

$$\forall \alpha, \forall \lambda, \langle \alpha, \lambda \rangle \leq h_\xi(\alpha) + \text{Log } \varphi_\xi(\lambda)$$

$$\begin{aligned} \text{donc } \int_0^1 \langle g'(t), \delta \frac{1[y,x](t)}{\psi_\delta(x-y)} \rangle dt &\leq \int_0^1 h_\xi(g'(t)) dt + \int_0^1 \text{Log } \varphi_\xi \left[ \frac{1[y,x](t)}{\psi_\delta(x-y)} \right] dt \\ &= w_\xi(g) + \int_y^x \text{Log } \varphi_\xi \left[ \frac{1}{\psi_\delta(x-y)} \right] dt \\ &\leq K + \int_y^x \frac{1}{x-y} dt = K + 1 \end{aligned}$$

Nous obtenons  $\rho[g(x) - g(y)] \leq (K + 1) \psi(x - y)$ ,  $\forall g \in A_K$ , ce qui montre que  $(K + 1) \psi(\cdot)$  est un module d'équicontinuité.

$$\begin{aligned} \text{Pour } G \text{ borélien de } \mathcal{C}_k[0,1], \text{ on pose } W_\xi(G) &= \inf_{g \in G \cap \mathcal{C}_0} w_\xi(g) \text{ si } G \cap \mathcal{C}_0 \neq \emptyset \\ &= +\infty \text{ sinon.} \end{aligned}$$

Des remarques et propositions précédentes, on déduit que :

i) si  $G$  est ouvert,  $W_\xi(G) = W_\xi(G \cap \Gamma)$

ii)  $\forall f \in \mathcal{C}_0, W_\xi[B(f, \varepsilon)] \nearrow w_\xi(f)$  lorsque  $\varepsilon \searrow 0$ .

iii) si  $\varphi$  existe partout,  $W_\xi[B(G, \varepsilon)] \nearrow W_\xi(\bar{G})$  pour tous les boréliens  $G$  vérifiant  $W_\xi(\bar{G}) < \infty$ ,  $\bar{G}$  désignant l'adhérence de  $G$  dans  $\mathcal{C}_k[0,1]$

Lorsque  $\xi$  suit la loi normale :  $\mathcal{N}(0, B)$ , sa transformée de Laplace existe

partout et on obtient les formes particulières suivantes :

$$h_0(\alpha) = \frac{1}{2} \alpha' B^{-1} \alpha$$

$$w_0(f) = \frac{1}{2} \int_0^1 f'(u)' B^{-1} f'(u) du$$

$$W_0(G) = \inf_{g \in G \cap \mathcal{C}_0} w_0(g) \text{ si } G \cap \mathcal{C}_0 \neq \emptyset \text{ et } +\infty \text{ sinon.}$$

## II. ENONCES DES THEOREMES.

### 1. Grandes déviations.

On suppose que les variables  $\xi$  ont une transformée de Laplace sur  $\mathbb{R}^k$ . Soit  $x(n)$  une suite telle que  $\frac{x(n)}{n} \rightarrow 1$ , alors  $\forall G$  borélien de  $\mathcal{C}_k$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P(s_n(\cdot) \in G) \geq -W_\xi^\circ(G), \quad G^\circ \text{ étant l'intérieur de } G.$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P(s_n(\cdot) \in G) \leq -W_\xi(\bar{G}) \quad \text{si } W_\xi(\bar{G}) < \infty.$$

Finalement, pour les boréliens  $G$  vérifiant  $W_\xi(\bar{G}) = W_\xi^\circ(G)$ , on a

$$\frac{1}{n} \text{Log } P(s_n(\cdot) \in G) \xrightarrow{n \rightarrow \infty} -W_\xi(G).$$

## 2. Moyennes déviations

On suppose que les variables  $\xi$  ont une transformée de Laplace dans un voisinage de 0. Soit  $x(n)$  une suite telle que  $\frac{x(n)}{n} \rightarrow 0$  et  $\frac{x(n)}{\sqrt{n}} \rightarrow \infty$ . Alors  $\forall G$  borélien de  $\mathcal{E}_k$ ,

$$\lim_{n \rightarrow \infty} \frac{n}{x(n)^2} \text{Log P}(s_n(\cdot) \in G) \geq -W_0(\overset{\circ}{G})$$

$$\lim_{n \rightarrow \infty} \frac{n}{x(n)^2} \text{Log P}(s_n(\cdot) \in G) \leq -W_0(\bar{G}) \text{ si } W_0(\bar{G}) < \infty$$

Finalement, pour les boréliens  $G$  vérifiant  $W_0(\bar{G}) = W_0(\overset{\circ}{G})$ , on a :

$$\frac{n}{x(n)^2} \text{Log P}(s_n(\cdot) \in G) \rightarrow -W_0(G)$$

Remarques : i) si les variables ne sont pas centrées mais  $E\xi = a$ , il suffit de remplacer  $G$  par  $G_a$  dans les seconds membres des inégalités :

$$G_a = \{g(t) - a.t, g \in G\}$$

ii) si la suite  $x(n)$  est telle que  $\frac{x(n)}{n} \rightarrow \beta \neq 1$ , il suffit de remplacer  $G$  par  $\beta G$  dans les seconds membres.

iii)  $W(\bar{G}) = 0 \Leftrightarrow \bar{G}$  contient la fonction nulle.

Les deux théorèmes seront démontrés parallèlement, il sera donc commode d'utiliser les notations  $h, w, W$ , pour désigner respectivement  $h_\xi, w_\xi, W_\xi$  dans les cas de grandes déviations et  $h_0, w_0, W_0$  dans les cas de moyennes déviations.

## III. EXEMPLES D'APPLICATION.

### 1. Interprétation de $w(f)$

$$\begin{aligned} \text{Si } w(f) < \infty, w(f) &= \lim_{\varepsilon \searrow 0} \lim_{n \rightarrow \infty} \frac{n}{x(n)^2} \text{Log P}(s_n(\cdot) \in B(f, \varepsilon)) \\ &= \lim_{\varepsilon \searrow 0} \overline{\lim}_n \frac{n}{x(n)^2} \text{Log P}(s_n(\cdot) \in B(f, \varepsilon)). \end{aligned}$$

En effet, nous avons vu que  $\forall f \in \mathcal{E}_0, W(B(f, \varepsilon)) \nearrow w(f)$  quand  $\varepsilon \searrow 0$  et d'autre part, le théorème de grandes déviations permet d'écrire :

$$\lim_{n \rightarrow \infty} \frac{n}{x(n)^2} \text{Log P}(s_n(\cdot) \in B(f, \varepsilon)) \geq -W(B(f, \varepsilon))$$

$$\overline{\lim} \frac{n}{x(n)^2} \text{Log P} (s_n(\cdot) \in B(f, \epsilon)) \leq -W(B(f, 2\epsilon))$$

et on fait tendre  $\epsilon$  vers 0

Remarque : Comme  $\varphi_\xi$  existe partout,  $W(B(f, \epsilon))$  est une fonction continue en  $\epsilon$  (suffisamment petit) pour toutes les fonctions  $f$  satisfaisant à la condition :  $f'$  continue à valeurs à l'intérieur du domaine de définition de  $h$ .

$$\text{Cela permet d'en déduire : } W(\overline{B(f, \epsilon)}) = W(B(f, \epsilon))$$

$$\text{et par suite : } \frac{n}{x(n)^2} \text{Log P} (s_n(\cdot) \in B(f, \epsilon)) \rightarrow -W(B(f, \epsilon))$$

et :

$$\lim_{\epsilon \searrow 0} \lim_n \frac{n}{x(n)^2} \text{Log P} (s_n(\cdot) \in B(f, \epsilon)) = -w(f).$$

2. Franchissement de la barrière. ( $k = 1$ )

On prend  $G = \{f \in \mathcal{C}[0, 1] \text{ telles que } \exists t \in [0, 1], f(t) \geq g(t)\}$ , où  $g$  une fonction continue sur  $[0, 1]$  telle que  $g(t) > 0, \forall t$  (le cas contraire serait trivial :  $\frac{1}{n} \text{Log P} (s_n(\cdot) \in G) \rightarrow 0$ ).

Etudions d'abord les grandes déviations.

1er cas : Si  $\frac{g(t)}{t} > \text{ess sup } \xi, \forall t$  ou si  $\frac{g(t)}{t} \geq \text{ess sup } \xi$  dans le cas où  $P(\xi = \text{ess sup } \xi) = 0$ , on a directement :

$$P(s_n(\cdot) \in G) = 0$$

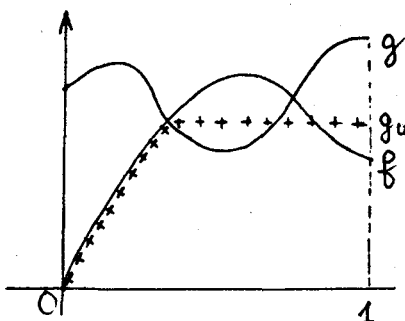
2e cas : si  $\frac{g(t)}{t} \geq \text{ess sup } \xi, \forall t$  dans le cas où  $P(\xi = \text{ess sup } \xi) > 0$

et  $\exists t_0 \in ]0, 1[$  (le plus petit) tel que  $\frac{g(t_0)}{t_0} = \text{ess sup } \xi$ . On obtient directement

$$\begin{aligned} P(s_n(\cdot) \in G) &= P(s_n(t_0) = t_0 \text{ ess sup } \xi) \\ &= [P(\xi = \text{ess sup } \xi)]^{\lfloor nt_0 \rfloor} + 1 \end{aligned}$$

$$\Rightarrow \frac{1}{n} \text{Log P} (s_n(\cdot) \in G) \rightarrow t_0 \cdot \text{Log P}(\xi = \text{ess sup } \xi) = w_\xi(g_{t_0}) = W_\xi(G)$$

Dans ce cas, la minoration donnée par le théorème serait sans intérêt :  $W(G) = \infty$



3e cas : si  $\exists t_0 \in [0,1], \frac{g(t_0)}{t_0} < \text{ess sup } \xi$ .

Pour  $f \in G$ , posons  $u = \inf \{t \in [0,1], f(t) = g(t)\}$ . La fonction  $g_u$  définie par  $g_u(t) = \frac{g(u)}{u} [\inf(t,u)]$  appartient aussi à  $G$  et  $w(f) \geq w(g_u)$  donc  $W(G) = \inf_u w(g_u) = \inf_u h \left[ \frac{g(u)}{u} \right]$

$G$  est évidemment fermé et de plus  $W(G) = W(\overset{\circ}{G})$ . En effet,  $\forall t, \frac{g(t)}{t} < \text{ess sup } \xi$  donc la fonction  $h_{\xi}$  est continue au point  $\frac{g(t)}{t}$ .

$\forall \epsilon > 0, \exists \eta > 0$  tel que  $h_{\xi} \left[ (1 + \eta) \frac{g(t)}{t} \right] \leq h_{\xi} \left[ \frac{g(t)}{t} \right] (1 + \epsilon)$   
 la fonction  $g_t^{\eta} = (1 + \eta) g_t$  vérifie  $w_{\xi}(g_t^{\eta}) \leq w_{\xi}(g_t) [1 + \epsilon]$  et  $g_t^{\eta} \in \overset{\circ}{G}$  donc  $W_{\xi}(\overset{\circ}{G}) \leq W(G) (1 + \eta), \forall \eta > 0$ .

S'il existe  $u_0$  unique réalisant  $\inf_u w(g_u)$ , alors on peut interpréter  $g_{u_0}$  comme la trajectoire la plus probable conditionnellement au fait que l'on franchit la barrière (cf; [ 7 ]).

Pour les moyennes déviations,  $h_0$  est une fonction continue sur  $\mathbb{R}$  pour toute fonction  $g$  continue sur  $[0,1]$ , on aura alors :

$$\frac{n}{x(n)^2} \cdot \text{Log P} \{ \exists t \in [0,1], s_n(t) \geq g(t) \} \rightarrow -W_0(G) = -\inf_t \frac{[g(t)]^2}{2\sigma^2 t}$$

### 3. Non franchissement de barrières : $k = 1$ .

On prend  $G = \{f \in \mathcal{C}[0,1] \text{ telle que } f(t) < g(t), \forall t\}$  où  $g$  de classe  $C^1$  et telle qu'il existe  $t \in ]0,1[$  tel que  $g(t) < 0$  (les autres cas ne sont pas intéressants :  $\frac{1}{n} \text{Log P}(s_n(\cdot) \in G) \rightarrow 0$  ou  $-\infty$ ).

1er cas : Si  $\exists t$  tel que  $\frac{g(t)}{t} < \text{ess inf } \xi$  ou si  $\exists t$  tel que  $\frac{g(t)}{t} = \text{ess inf } \xi$  dans le cas où  $P(\xi = \text{ess inf } \xi) = 0$  on a directement :

$$P(s_n(\cdot) \in G) = 0$$

2e cas :  $\exists t_0 \in ]0,1[$  (le plus grand) tel que  $\frac{g(t_0)}{t_0} = \text{ess inf } \xi$  dans le cas où  $P(\xi = \text{ess inf } \xi) > 0$ . Ce cas sera traité avec le suivant car  $h_{\xi}$  est continue sur  $[\text{ess inf } \xi, 0]$ .

3e cas :  $\frac{g(t)}{t} > \text{ess inf } \xi, \forall t$ .

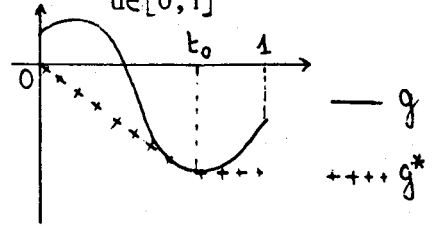
Considérons d'abord la fonction  $\tilde{g}$ , enveloppe convexe inférieure de  $\{(0,0)\} \cup \{(t,y), y > g(t)\}$  ; en vertu des propriétés  $\tilde{g}$  est continue sur



$[0,1]$  et  $g(0) = 0$ .

Considérons maintenant  $t_0 = \inf \{t, g(t) = \inf_{u \in [0,1]} g(u)\}$ , en vertu des propriétés de  $g$ ,  $\tilde{g}(t_0) = g(t_0) < 0$ .

Notons  $g^*$  la fonction :  $g^*(t) = \tilde{g}(t), t \leq t_0$   
 $= g(t_0), t \geq t_0$



Nous allons montrer que  $W(G) = w(g^*)$ . D'abord  $w(g^*) < t_0 \cdot h[\tilde{g}(0)] < \infty$

d'après les propriétés de  $h$  et de  $g$ . La fonctionnelle  $w$  étant s.c.i., elle atteint son minimum sur le compact  $\{f \in G \cap \mathcal{C}_0, w(f) \leq w(g^*)\}$  pour une fonction notée  $\bar{g}$  :

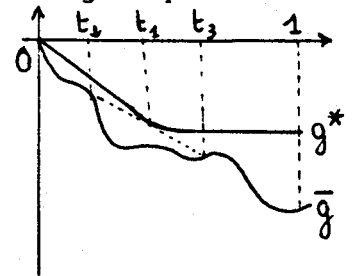
$$W(G) = w(\bar{g}) \leq w(g^*)$$

Supposons que  $w(\bar{g}) < w(g^*)$ .

Si il existe  $t_1$  tel que  $\bar{g}(t_1) < g^*(t_1)$ , considérons une tangente à  $g^*$  au point  $t_1$  ; elle rencontre  $\bar{g}$  en un point  $t_2 < t_1$  car  $\bar{g}(0) = g^*(0) = 0$  et  $g^*$  est convexe. Deux cas se présentent :

1er cas : Cette tangente rencontre  $\bar{g}$  en un point d'abscisse  $t_3 > t_1$ , alors on aurait  $w(\bar{g}) < w(g)$ , ce qui est impossible :

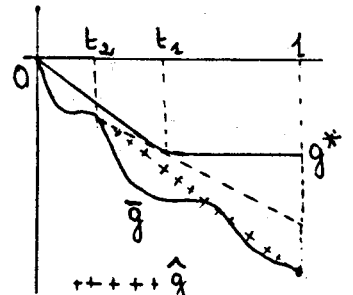
$$\bar{g}(t) = \begin{cases} \bar{g}(t), & t \leq t_2 \text{ et } t \geq t_3 \\ \frac{t_3-t}{t_3-t_2} g(t_2) + \frac{t-t_2}{t_3-t_2} g(t_3), & t_2 \leq t \leq t_3 \end{cases}$$



2e cas : Cette tangente ne rencontre pas  $\bar{g}$  pour  $t > t_1$ , alors on aurait  $w(\hat{g}) < w(\bar{g})$ , ce qui est impossible :

$$\hat{g}(t) = \begin{cases} \bar{g}(t), & t \leq t_2 \\ \frac{1-t}{1-t_2} g(t_2) + \frac{t-t_2}{1-t_2} g(1), & t \geq t_2 \end{cases}$$

ce qui prouve que  $\bar{g} \geq g^*$ .



Considérons maintenant l'ensemble  $\{t \in [0,1], \bar{g}(t) > g^*(t)\}$  :

c'est un ouvert donc réunion au plus dénombrable d'intervalles :

. Pour un intervalle du type  $]t'_n, t''_n[$  avec  $t''_n < 1$ , on a  $g^*(t'_n) = \bar{g}(t'_n)$  et  $g^*(t''_n) = \bar{g}(t''_n)$  ; d'autre part  $g^*$  est linéaire sur  $]t'_n, t''_n[$  car elle est strictement inférieure à  $\bar{g}$  donc à  $g$  sur cet intervalle.

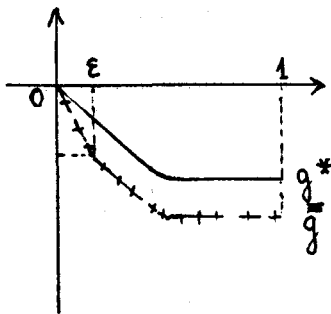
La convexité stricte de  $h$  impliquerait que  $\int_{t'_n}^{t''_n} h(g^{*'}(t)) dt < \int_{t'_n}^{t''_n} h(\bar{g}'(t)) dt$ ,

ce qui contredit l'optimalité de  $\bar{g}$ .

Pour un intervalle du type  $|t', 1|$ , nécessairement  $t' > t_0$  et on aurait  $0 = \int_{t'}^1 h(g^{*'}(t)) dt < \int_{t'}^1 h(\bar{g}'(t)) dt$ . Ce qui prouve que  $\bar{g} = g^*$ .

$G$  est évidemment fermé et de plus  $W(G) = W(\bar{G})$  si  $g(0) > 0$ , et si  $\frac{g(t)}{t} > \text{ess inf } \xi \forall t$ .  
Considérons la fonction de  $G$  définie à partir de  $g^*$  par :

$$\varepsilon = \begin{cases} = [g^{*'}(0) - \varepsilon]t \text{ sur } [0, \varepsilon], \varepsilon \text{ petit de sorte que } g^{*'}(0) - \varepsilon > \text{ess} \\ \text{inf } \xi \\ = g^*(t) + \bar{g}(\varepsilon) - g^*(\varepsilon) \text{ sur } [\varepsilon, 1]. \end{cases}$$



On a :  $0 \leq w(\bar{g}) - w(g^*)$

$$\begin{aligned} &\leq \int_0^\varepsilon |h(g^{*'}(t)) - h(g^{*'}(0) - \varepsilon)| dt \\ &\leq \int_0^\varepsilon |h(g^{*'}(t)) - h(g^{*'}(0))| dt + \varepsilon |h(g^{*'}(0)) - h(g^{*'}(0) - \varepsilon)|. \end{aligned}$$

arbitrairement petit car  $g^{*'}$  continue en 0 et  $h$  est continue en  $g^{*'}(0)$ .

Dans ces exemples,  $g^*$  peut être interprétée comme la trajectoire la plus probable de  $G$ , au sens suivant : asymptotiquement la probabilité pour  $s_n(\cdot)$  d'appartenir à  $G$  est "équivalente" à celle d'être au voisinage de  $g^*$ .

#### IV. DEMONSTRATIONS DES THEOREMES

Comme nous le verrons ces résultats sont obtenus par discrétisation et les outils essentiels sont les théorèmes sur les grandes et moyennes déviations sur  $\mathbb{R}^k$  ; de façon plus précise, nous utiliserons les deux théorèmes suivants :

Théorème 1. Grandes déviations (cf [1]).

Si les variables aléatoires  $\xi_i$  ont une transformée de Laplace définie dans un voisinage de 0, alors pour tout borélien  $Z$  de  $\mathbb{R}^k$  :

$$\liminf_n \frac{1}{n} \text{Log } P \left( \frac{S_n}{n} \in Z \right) \geq - \inf_{z \in Z} h_\xi(z)$$

$$\overline{\lim}_n \frac{1}{n} \text{Log } P \left( \frac{S_n}{n} \in Z \right) \leq - \inf_{z \in B_p(Z, \varepsilon)} h_\xi(z), \forall \varepsilon > 0$$

où  $Z$  désigne l'intérieur de  $Z$  dans  $R^k$  muni de la métrique  $\rho$  et  $B_\rho(Z, \varepsilon)$  désigne  $\{y \in R^k, \exists z \in Z \text{ tel que } \rho(z-y) < \varepsilon\}$ .

Théorème 2. moyennes déviations.

Si les variables aléatoires  $\xi_i$  ont une transformée de Laplace définie dans un voisinage de 0, alors pour tout borélien  $Z$  de  $R^k$  :

$$\begin{aligned} \lim_n \frac{n}{x(n)^2} \text{Log P} \left( \frac{S_n}{x(n)} \in Z \right) &\geq - \inf_{z \in Z} h_0(z) \\ \overline{\lim}_n \frac{n}{x(n)^2} \text{Log P} \left( \frac{S_n}{x(n)} \in Z \right) &\leq - \inf_{z \in B_\rho(Z, \varepsilon)} h_0(z), \quad \forall \varepsilon > 0 \end{aligned}$$

Ce théorème découle directement du théorème de moyennes déviations sur  $R$  (cf [4]) que l'on étend à  $R^k$ , de façon analogue aux grandes déviations.

En suivant la démarche de [3], nous ferons la démonstration en 3 étapes ;

- nous montrons d'abord l'existence d'un  $\varepsilon$ -précompact (voisinage d'ordre  $\varepsilon$  d'un compact) dont la probabilité du complémentaire est négligeable pour le problème .

- nous établissons ensuite les théorèmes dans le cas où  $G$  est une boule centrée sur une ligne polygonale.

- nous étendons enfin les résultats aux boréliens  $G, G \cap \mathcal{E}_0 \neq \emptyset$ .

Lemme 1.

Si les variables aléatoires  $\xi_i$  admettent une transformée de Laplace au voisinage de 0 (resp. partout) dans le cas de moyennes (resp. grandes) déviations,

$\forall C > 0, \forall \varepsilon > 0$ , il existe un nombre fini de lignes polygonales  $f_1, \dots, f_d \in \Gamma$  telles que  $\overline{\lim}_n \frac{n}{x^2(n)} \text{Log P} (s_n(\cdot) \notin \bigcup_{j=1}^d B(f_j, \varepsilon)) \leq -C$

Démonstration

Des nombres  $\varepsilon > 0$  et  $1/N$  étant fixés, considérons l'ensemble

$$K(\varepsilon, 1) = \bigcap_{i=1}^1 \left\{ f \in \mathcal{E}_0, \text{Sup}_{\frac{i-1}{1} \leq t < \frac{i}{1}} \rho \left[ f \left( \frac{i-1}{1} \right) - f(t) \right] < \varepsilon \right\}$$

En prenant les lignes polygonales  $f_j$  à sommets d'abscisses  $0, \frac{1}{1}, \frac{2}{1}, \dots, 1$ , il est clair que  $K(\varepsilon, 1) \subset \bigcup_{j=1}^d B(f_j, \varepsilon)$ ,  $d = 2^{k \cdot 1}$

$$\begin{aligned}
 P(s_n(\cdot) \notin K(\epsilon, 1)) &= P\left(\bigcup_{i=1}^1 \{s_n(\cdot), \sup_{\frac{i-1}{1} \leq t \leq \frac{i}{1}} \rho|s_n(\frac{i-1}{1}) - s_n(t)| > \epsilon\}\right) \\
 &\leq \sum_{i=1}^1 P\left\{\sup_{\frac{i-1}{1} \leq t \leq \frac{i}{1}} \rho[s_n(\frac{i-1}{1}) - s_n(t)] > \epsilon\right\} \\
 &< 1 \cdot P\left(\sup_{0 \leq t \leq \frac{1}{1}} \rho[s_n(t)] > \epsilon\right) \\
 &\leq 1 \cdot P\left(\max_{j \leq \frac{n}{1}} \rho|S_j| > \epsilon \cdot x(n)\right)
 \end{aligned}$$

De l'inégalité bien connue sur  $R$  (par exemple cf [5])

$$P(\max_{j \leq N} |S_j| > \lambda) \leq 2 P(|S_N| > \lambda - \sqrt{2} \sigma \sqrt{N})$$

$$\text{On tire : } P(s_n(\cdot) \notin K(\epsilon, 1)) \leq 1 \sum_{i=1}^k P(\max_{j \leq \frac{n}{1}} [S_j^i] > \epsilon \cdot x(n))$$

$$\leq 2k1 \sup_i P\left(\left[S_{\frac{n}{1}}^i\right] > \epsilon \cdot x(n) - \sqrt{2} \sigma_i \sqrt{\frac{n}{1}}\right)$$

$$\leq 2k1 P\left(\rho\left[S_{\frac{n}{1}}\right] > \epsilon 1 \cdot \frac{x(n)}{1} - \sqrt{2} \sqrt{\frac{n}{1}} \rho(\sigma)\right)$$

En appliquant les théorèmes de moyennes et grandes déviations sur  $R^k$ , où  $\frac{x(n)}{1}$  joue le rôle de  $x\left(\left[\frac{n}{1}\right]\right)$ , on obtient :

$$\begin{aligned}
 \overline{\lim}_n \frac{n}{x^2(n)} \text{Log } P(s_n(\cdot) \notin K(\epsilon, 1)) &\leq -\frac{1}{1} \inf_{z \notin B_\rho(0, (1-1)\epsilon)} h(z)
 \end{aligned}$$

La fonction  $h$  étant la transformée de Cramer d'une variable ayant une

transformée de Laplace définie partout, elle vérifie :  $\frac{h(z)}{\rho(z)} \xrightarrow[\rho(z) \rightarrow \infty]{} \infty$  (cf [1])

On en déduit donc le lemme :

$$\forall \epsilon > 0, \forall C > 0, \exists 1 \text{ tel que } \overline{\lim}_n \frac{n}{x^2(n)} \text{Log } P(s_n(\cdot) \notin K(\epsilon, 1)) \leq -C.$$

Nous allons maintenant montrer que les théorèmes sont vrais

lorsque  $G$  est une boule centrée sur une ligne polygonale  $f$  de  $\Gamma$ . La démonstration

consiste à discrétiser le problème : au lieu de calculer  $P(\rho[s_n(t) - f(t)]$

$< \epsilon, \forall t$ ), on va étudier le comportement asymptotique de l'expression :

$$P\left(\bigcap_{q \in Q} \{\rho[s_n(q) - f(q)] < \epsilon\}\right)$$

où  $Q$  est un ensemble discret ( $r$  étant fixé, on prendra  $Q_r = \{0, \frac{1}{r}, \dots, \frac{r-1}{r}, 1\}$ )

Il y aura donc deux parties :

- écrire un théorème de grandes ou moyennes déviations pour évaluer :

$$P \left( \bigcap_{q \in Q_r} \{ \rho [s_n(q) - f(q)] < \varepsilon \} \right).$$

- approcher  $P (s_n(\cdot) \in B(f, \varepsilon))$  par densité en faisant tendre  $r$  vers l'infini. Il sera commode d'utiliser les notations suivantes :

pour  $g \in \mathcal{C}_0$ , soit  $F_r(g) = \{ g(\frac{1}{r}), \dots, g(\frac{j}{r}) - g(\frac{j-1}{r}), \dots, g(1) - g(\frac{r-1}{r}) \}$

$$F_r^{(n)}(g) = \left\{ g\left(\frac{\lfloor \frac{n}{r} \rfloor}{n}\right), \dots, g\left(\frac{j \lfloor \frac{n}{r} \rfloor}{n}\right) - g\left(\frac{(j-1) \lfloor \frac{n}{r} \rfloor}{n}\right), \dots, g\left(\frac{r \lfloor \frac{n}{r} \rfloor}{n}\right) - g\left(\frac{(r-1) \lfloor \frac{n}{r} \rfloor}{n}\right) \right\}$$

L'application  $F_r$  de  $\mathcal{C}_0$  dans  $\mathbb{R}^{k \cdot r}$  consiste à mettre dans un vecteur de taille  $k \cdot r$  les accroissements de  $g$  sur les segments  $[\frac{j-1}{r}, \frac{j}{r}]$  et l'application  $F_r^{(n)}$  approxime  $F_r$  de sorte que les accroissements soient pris en des points d'abscisse  $\frac{i}{n}$ .

Lemme 2.

Pour tout borélien  $Z$  de  $\mathbb{R}^{k \cdot r}$

$$\liminf_n \frac{n}{x(n)^2} \text{Log} P \{ F_r^{(n)}(s_n) \in Z \} \geq - \inf_{z \in Z} H_r(z)$$

$$\overline{\lim}_n \frac{n}{x(n)^2} \text{Log} P \{ F_r^{(n)}(s_n) \in Z \} \leq - \inf_{z \in B_\rho(Z, \varepsilon)} H_r(z), \forall \varepsilon > 0$$

où  $H_r(z) = \frac{1}{r} \sum_{i=1}^r h(rz_i)$  si  $z = (z_1, \dots, z_r) \in (\mathbb{R}^k)^r$   
 $i = 1$

Démonstration.

C'est une application directe des théorèmes de moyennes et grandes déviations dans  $\mathbb{R}^{k \cdot r}$ , en effet :

$$F_r^{(n)}(s_n) = \left( \frac{\xi_1 + \dots + \xi_{\lfloor \frac{n}{r} \rfloor}}{x(n)}, \dots, \frac{\xi_{(j-1) \lfloor \frac{n}{r} \rfloor + 1} + \dots + \xi_{j \lfloor \frac{n}{r} \rfloor}}{x(n)}, \dots \right)$$

$$= \frac{X_1 + X_2 + \dots + X_{\lfloor \frac{n}{r} \rfloor}}{x(n)}$$

où  $X_j = (\xi_j, \xi_j + \frac{n}{r}, \dots, \xi_j + (r-1) \frac{n}{r})$  est une variable aléatoire de  $R^{kr}$  admettant pour transformée de Cramer au point  $z$  :  $\sum_{i=1}^k h(z_i)$

Montrons d'abord la majoration dans le cas où  $G$  est une boule :

$$\begin{aligned} \overline{\lim}_n \frac{n}{x(n)^2} \text{Log P} (s_n(\cdot) \in B(f, \epsilon)) &\leq \overline{\lim}_n \frac{n}{x(n)^2} \text{Log P} \{F_r^{(n)}(s_n) \in F_r^{(n)} [B(f, \epsilon)]\} \\ &\leq - \inf_{z \in B_\rho [F_r(B(f, \epsilon)), \delta]} H_r(z), \quad \forall \delta > 0. \end{aligned}$$

Mais  $B_\rho [F_r(B(f, \epsilon)), \delta] = F_r [B(f, \epsilon + \delta)]$   
 et  $H_r [F_r(g)] = \frac{1}{r} \sum_{i=1}^r h \left[ \frac{g(\frac{i}{r}) - g(\frac{i-1}{r})}{\frac{1}{r}} \right] = w(g_r)$  où  $g_r$  est la ligne polygonale de sommets  $(\frac{i}{r}, g(\frac{i}{r}))$ ,  $i = 0, 1, 2, \dots, r$ .

Comme  $h$  est convexe, on a  $w(g_r) \leq w(g)$  et comme  $w$  est s.c.i.,  $\lim_{r \rightarrow \infty} w(g_r) = w(g)$   
 et on en déduit :  $\lim_{r \rightarrow \infty} \inf_{g \in B(f, \epsilon + \delta)} w(g_r) \leq \inf_{g \in B(f, \epsilon + \delta)} \lim_{r \rightarrow \infty} w(g_r) = \inf_{g \in B(f, \epsilon + \delta)} w(g)$

d'où la majoration :  $\overline{\lim}_n \frac{n}{x(n)^2} \text{Log P} (s_n(\cdot) \in B(f, \epsilon)) \leq - \inf_{g \in B(f, \epsilon + \delta)} w(g), \quad \forall \delta > 0$

Pour la minoration, on peut écrire :

$$\begin{aligned} P (s_n(\cdot) \in B(f, \epsilon)) &\geq P \{F_r^{(n)}(s_n) \in F_r^{(n)} [B(f, \frac{\epsilon}{3})]\} \\ &\quad - P \{F_r^{(n)}(s_n) \in F_r^{(n)} [B(f, \frac{\epsilon}{3})] \cap s_n(\cdot) \notin B(f, \epsilon)\} \end{aligned}$$

Le deuxième terme est négligeable devant le 1er car dans cet ensemble  $s_n$  est nécessairement "agitée" entre les points  $\frac{j}{r}$ , plus précisément :

$$\{F_r^{(n)}(s_n) \in F_r^{(n)} [B(f, \frac{\epsilon}{3})] \cap \{s_n(\cdot) \notin B(f, \epsilon)\} \cap K(\frac{\epsilon}{3}, r) = \emptyset \text{ pour } r > R,$$

où  $R$  défini à partir du module de continuité de  $f$  :  $\psi_f(\frac{1}{R}) < \frac{\epsilon}{3}$ .

Nous venons de montrer que :

$$\forall \epsilon > 0, \forall C, \exists r \text{ tel que } \overline{\lim}_n \frac{n}{x(n)^2} \text{Log P} \{F_r^{(n)}(s_n) \in F_r^{(n)} [B(f, \frac{\epsilon}{3})] \cap s_n \notin$$

$$B(f, \epsilon)\} \leq -C$$

alors que d'autre part :

$$\begin{aligned} & \lim_{\frac{n}{x(n)^2}} \text{Log P} \{ F_r^{(n)}(s_n) \in F_r^{(n)} [B(f, \frac{\epsilon}{3})] \} \\ &= \lim_{\frac{n}{x(n)^2}} \text{Log P} \{ F_r^{(n)}(s_n) \in B_\rho [F_r^{(n)}(f, \frac{\epsilon}{3})] \} \geq -H_r[F_r(f)] \geq -w(f). \end{aligned}$$

Il ne reste plus qu'à étendre les théorèmes à des boréliens G quelconques. La minoration est immédiate :

$$\begin{aligned} & \lim_{\frac{n}{x(n)^2}} \text{Log P} (s_n(\cdot) \in G) \geq \lim_{\frac{n}{x(n)^2}} \text{Log P} (s_n(\cdot) \in \overset{\circ}{G}) \\ & \geq \lim_{\frac{n}{x(n)^2}} \text{Log P}(s_n(\cdot) \in B(f, \epsilon)), \forall f \in \Gamma \cap \overset{\circ}{G}, \epsilon \text{ étant choisi tel que } B(f, \epsilon) \subset \overset{\circ}{G} \end{aligned}$$

$$\geq -w(f), \forall f \in \Gamma \cap \overset{\circ}{G}$$

$$\text{donc } \lim_{\frac{n}{x(n)^2}} \text{Log P}(s_n(\cdot) \in G) \geq -W(\Gamma \cap \overset{\circ}{G}) = -W(\overset{\circ}{G})$$

Pour la majoration on a :  $\forall \epsilon > 0$ ,

$$P(s_n \in G) \leq P(s_n \in G \cap K(\frac{\epsilon}{3}, 1)) + P(s_n \notin K(\frac{\epsilon}{3}, 1))$$

$\epsilon$  étant fixé, 1 est choisi suffisamment grand pour que le 2e terme soit négligeable devant le 1er.

$$K(\frac{\epsilon}{3}, 1) \subset \bigcup_{i=1}^d B(f_i, \frac{\epsilon}{3}) \Rightarrow \Gamma \cap K(\frac{\epsilon}{3}, 1) \subset \bigcup_i B(f_i, \frac{\epsilon}{3}) \text{ tels que } B(f_i, \frac{\epsilon}{3}) \cap G \neq \emptyset$$

$$\begin{aligned} & \overline{\lim}_{\frac{n}{x(n)^2}} \text{Log P} \{s_n(\cdot) \in G \cap K(\frac{\epsilon}{3}, 1)\} \leq - \inf_i W |B(f_i, \frac{\epsilon}{3} + \frac{\epsilon}{3})|, \text{ indices } i \\ & \text{tels que } B(f_i, \frac{\epsilon}{3}) \cap G \neq \emptyset. \end{aligned}$$

$$\text{soit finalement, } \overline{\lim}_{\frac{n}{x(n)^2}} \text{Log P} (s_n(\cdot) \in G) \leq -W [B(G, \epsilon)], \forall \epsilon > 0.$$

#### V. RESULTATS DANS LE CAS OU $\varphi_\xi$ N'EXISTE PAS PARTOUT

On suppose que  $\varphi_\xi$  existe dans un voisinage de 0. Donc tous les résultats sur les moyennes déviations restent vrais.

En ce qui concerne les grandes déviations, on est confronté avec deux types de problèmes :

- les problèmes liés à la définition et aux propriétés de  $w_\xi$ .

- les problèmes plus spécifiques à la convergence.

1. Propriétés de  $w_\xi$  :

Les principaux ennuis viennent essentiellement du fait que  $\frac{h_\xi(\alpha)}{\rho(\alpha)}$

ne tend plus vers l'infini dans toutes les directions. En particulier, les ensembles  $A_K = \{f \in \mathcal{C}_0, w_\xi(f) \leq K\}$  ne sont plus compacts. De même, la convergence de  $W_\xi(B(G, \epsilon))$  vers  $W_\xi(\bar{G})$ , dans le cas où  $W_\xi(\bar{G}) < \infty$ , n'est plus assurée.

exemple : soit  $\xi$  une variable aléatoire réelle admettant la densité.

$$\frac{C}{1+x^4} \cdot e^{-|x|} \text{ sur } \mathbb{R}.$$

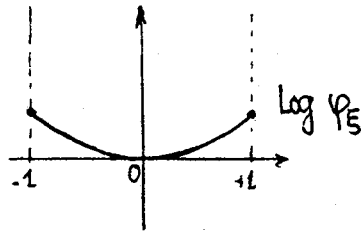
Elle admet clairement une transformée de Laplace sur  $[-1, +1]$

Posons  $\frac{\varphi'_\xi(1)}{\varphi_\xi(1)} = a (< \infty)$

Pour  $\alpha \notin [-a, +a]$ ,  $h_\xi$  est linéaire :

$$h_\xi(\alpha) = |\alpha| - a + h_\xi(a)$$

donc  $\frac{h_\xi(\alpha)}{|\alpha|} \xrightarrow{\alpha \rightarrow \pm \infty} 1.$



L'ensemble  $A_1 = \{f \in \mathcal{C}_0, w_\xi(f) \leq 1\}$  n'est pas compact car il contient la suite

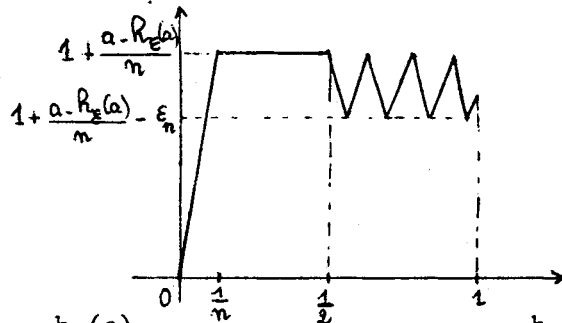
$\{f_n\}_{n \geq h_\xi(a)}$  suivante :

$$f_n(t) = \inf \left[ \frac{n+a-h_\xi(a)}{n}, t(n+a-h_\xi(a)) \right]$$

$$w_\xi(f_n) = \frac{1}{n} h_\xi[n+a-h_\xi(a)] = 1 \text{ si } n \geq h_\xi(a).$$

Considérons l'ensemble  $G = \{g_n\}_{n \geq \sup(2, h_\xi(a))}$

où  $g_n(t) = f_n(t)$  sur  $[0, \frac{1}{2}]$  et est formée d'oscillations polygonales de pente  $\pm a$  d'amplitude  $\epsilon_n \searrow 0$



$$\forall n \geq \sup[2, h_\xi(a)], w_\xi(g_n) = 1 + \frac{h_\xi(a)}{2} \Rightarrow W_\xi(\bar{G}) = W_\xi(G) = 1 + \frac{h_\xi(a)}{2}$$

En revanche,  $\forall \epsilon > 0$  pour  $n$  assez grand,  $\epsilon_n \leq \epsilon \Rightarrow (G, \epsilon) \supset B(G, \epsilon_n) \ni f_n.$

donc  $\forall \epsilon > 0, W_\xi(B(G, \epsilon)) \leq 1.$



2. Résultats de convergence.

Les résultats restent vrais pour les boules centrées sur les lignes polygonales.

Lemme. Soit  $f \in \Gamma$ , telle que  $f'$  prenne ses valeurs dans l'intérieur du domaine de définition de  $h_\xi$ . Pour expliciter cette contrainte, on rappelle que l'intérieur de ce domaine est l'intérieur de l'enveloppe convexe du support de  $\xi$  (voir [6]).

$$\overline{\lim}_n \frac{1}{n} \text{Log } P(s_n(\cdot) \in B(f, \varepsilon)) \leq -W_\xi [B(f, \varepsilon + \delta)], \forall \delta > 0.$$

$$\underline{\lim}_n \frac{1}{n} \text{Log } P(s_n(\cdot) \in B(f, \varepsilon)) \geq -w_\xi(f)$$

Démonstration.

La

majoration utilise seulement le théorème de grandes déviations sur  $R^k$  et donc seulement l'existence de  $\varphi_\xi$  au voisinage de 0.

Pour la minoration nous tronquons les variables aléatoires  $\xi$  de sorte que leur transformée de Laplace existe partout et nous pouvons alors utiliser les résultats précédents.

$$P(s_n(\cdot) \in B(f, \varepsilon)) \geq P(s_n(\cdot) \in B(f, \varepsilon) \mid \rho(\xi_i) \leq M, i) \cdot P(\rho(\xi) \leq M)^n$$

$$\text{donc } \frac{1}{n} \text{Log } P(s_n(\cdot) \in B(f, \varepsilon)) \geq \frac{1}{n} \text{Log } P(s_n(\cdot) \in B(f, \varepsilon) \mid \rho(\xi_i) \leq M, i) +$$

$$\text{Log } P(\rho(\xi) \leq M).$$

Si on note  $w_{\xi, M}$  la fonctionnelle  $w$  correspondant à la variable aléatoire  $\xi$  tronquée par  $\rho(\xi) \leq M$ , les résultats précédents donnent :

$$\forall M, \underline{\lim}_n \frac{1}{n} \text{Log } P(s_n \in B(f, \varepsilon)) \geq -w_{\xi, M}(f) + \text{Log } P[\rho(\xi) \leq M].$$

Lorsque  $M \rightarrow \infty$ ,  $w_{\xi, M}(f) \rightarrow w(f)$  (cf [6]) et  $P(\rho(\xi) \leq M) \rightarrow 1$  ce qui permet d'en déduire immédiatement le :

Théorème.

Si  $\varphi_\xi$  existe dans un voisinage de 0, on a :

$$\forall G \text{ borélien de } \mathcal{C}_k[0, 1], \underline{\lim}_n \frac{1}{n} \text{Log } P(s_n(\cdot) \in G) \geq -W_\xi(G)$$

$$\forall G \text{ } \varepsilon\text{-précompact, } \forall \delta > \varepsilon, \overline{\lim}_n \frac{1}{n} \text{Log } P(s_n(\cdot) \in G) \leq -W_\xi[B(G, \delta)].$$

Démonstration.

$$\begin{aligned} P(s_n(\cdot) \in G) &\geq P(s_n(\cdot) \in \overset{\circ}{G}) \\ &\geq P(s_n(\cdot) \in B(f, \epsilon)), \forall f \in \overset{\circ}{G} \cap \Gamma, \quad \epsilon \text{ déterminé de sorte que } B(f, \epsilon) \subset \overset{\circ}{G}. \end{aligned}$$

. Si  $f$  est tel que  $f'$  prend ses valeurs dans l'intérieur domaine de définition de  $h_\xi$ , alors d'après le lemme précédent  $\underline{\lim} \frac{1}{n} \text{Log } P(s_n(\cdot) \in G) \geq -w_\xi(f)$

. D'autre part si  $f'$  prend des valeurs sur la frontière du domaine de définition de  $h_\xi$ , il suffit de remarquer qu'il existe  $\eta > 0$  et  $f_1$  polygonale telle que  $f'_1$  prend toutes ses valeurs à l'intérieur du domaine de  $h_\xi$  et  $B(f_1, \eta) \subset B(f, \epsilon)$  et  $w_\xi(f_1) \leq w_\xi(f)$ . (Convexité du domaine de définition de  $h_\xi$ ).

Il semble par ailleurs, qu'on ne puisse limiter l'agitation de  $s_n(\cdot)$  comme dans le paragraphe précédent, les queues de  $\xi$  n'étant plus suffisamment petites. (C'est du moins, ce que donne à penser la non-compacité des ensembles  $A_K$ ).

BIBLIOGRAPHIE.

- [1] - J. BRETAGNOLLE : Séminaire 1977-78 d'ORSAY. chap. III
- [2] - A.A. BOROVKOV : "Boundary value problems for random walks and large déviations". Theory of Probability and applications (1967).
- [3] - A.A. MOGULSKII : "Large deviations for trajectories of multi dimensional random walks" Theory of Probability and applications (1976).
- [4] - W. FELLER : "An introduction to probability theory and its applications" Wiley - tome II - chapitre 16.
- [5] - BILLINGSLEY : "Convergence of probability measures" Wiley, page 69.
- [6] - P. BARTFAI : "Connections between the convex analysis and the theory of large deviations". Preprint of the Mathematical Institute of Hungarian academy of sciences. n° 2/ 1977.
- [7] - R. AZENCOTT : "Grandes déviations : théorèmes à la Cramer-Chernoff et petites perturbations de systèmes dynamiques"  
Cours de l'Ecole d'Eté de Saint-Flour, 1978 (à paraître).

Jean DESHAYES

Dominique PICARD

Mathématiques - Bât. 425

ERA CNRS 532 "Statistique Appliquée"

Université Paris-Sud

91405 ORSAY

APPLICATION AUX TESTS DE RUPTURE DE REGRESSION

Jean DESHAYES, Dominique PICARD.

I - VITESSE EXPONENTIELLE DES TESTS utilisés pour détecter une rupture de régression .

Nous reprenons brièvement l'exposé du problème et les tests proposés dans [1].

Soit le modèle :

$$y_k = x'_k \beta_k + u_k \quad k = 1, 2, \dots, n$$

où  $x_k$  désigne le vecteur sur lequel on régresse.

$\beta_k$  est le vecteur des paramètres de régression au temps  $k$  .

$u_k$  est l'erreur : les  $u_k$  sont supposées indépendantes et de loi normale centrée de variance  $\sigma_k^2$  .

L'hypothèse à tester est qu'il n'y a pas de rupture :

$$H_0 = \{ \beta_1 = \beta_2 = \dots = \beta_n \text{ et } \sigma_1 = \sigma_2 = \dots = \sigma_n \} .$$

En fait, nous n'envisagerons ici les propriétés des tests que dans le cas d'une contrehypothèse du type  $\{ \beta_1 = \beta_2 = \dots = \beta_r \neq \beta_{r+1} = \dots = \beta_n \text{ et } \sigma_1 = \sigma_2 = \dots = \sigma_n \}$  et du point de vue asymptotique ; nous supposons observer le processus  $Y_t$  sur  $[0, 1]$  aux instants  $\frac{k}{n}$  ( $k = 1, 2, \dots, n$ ) donc en des instants de plus en plus rapprochés. Sous l'alternative, nous supposons par conséquent qu'il existe  $t_0 \in ]0, 1[$  tel que pour :

$$\begin{aligned} \frac{k}{n} < t_0, & \quad y_{k,n} \text{ suit } \mathcal{N}(x'_{k,n} \beta_0 ; \sigma^2) \\ \frac{k}{n} \geq t_0, & \quad y_{k,n} \text{ suit } \mathcal{N}(x'_{k,n} \beta_1 ; \sigma^2) \end{aligned}$$

Aucun caractère d'optimalité ne se dégage de chacun des tests proposés, ceci étant dû, d'une part à l'immensité de l'alternative, d'autre part au fait que  $t_0$  est mal estimable, même si on restreint l'alternative comme ci-dessous.

Les tests proposés dans [1] sont construits à partir des résidus récursifs (ceux-ci sont obtenus à partir des résidus habituels par orthogonalisation)



$$\omega_k = \frac{y_k - x_k' b_{k-1}}{\sqrt{1 + x_k' (X_{k-1}' X_{k-1})^{-1} x_k}}$$

avec  $b_k = (X_k' X_k)^{-1} X_k' Y_k$  et  $X_k = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_k' \end{pmatrix}$ ,  $Y_k = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}$

Nous allons étudier du point de vue asymptotique, le test basé sur les sommes cumulées de résidus puis celui basé sur les sommes cumulées des carrés de résidus. Pour les deux tests, nous distinguerons les cas  $\sigma$  connu et  $\sigma$  inconnu et nous nous limiterons au cas unidimensionnel (rupture de moyenne).

I - 1) Test construit sur les sommes cumulées de résidus

I - 1. : 1)  $\sigma$  connu

Les variables  $y_{k,n}$  suivent une loi normale  $\mathcal{N}(m_1, \sigma^2)$  pour  $k < nt_0$  et une loi normale  $\mathcal{N}(m_2, \sigma^2)$  pour  $k \geq nt_0$ .

Sous  $H_0$ , on a  $t_0 = 0$  ou  $m_1 = m_2$

sous  $H_1$ , on a  $t_0 \in ]0, 1[$  et  $m_1 - m_2 = d \neq 0$ .

Les résidus récurrents valent :  $\omega_k = \frac{y_k - \bar{y}^{k-1}}{\sqrt{1 + \frac{1}{k-1}}}$ ,  $k = 2, \dots, n$  où

$$\bar{y}^{k-1} = \frac{\sum_{i=1}^{k-1} y_i}{k-1}$$

Le test basé sur les sommes cumulées de résidus consiste à rejeter l'hypothèse  $H_0$  s'il existe  $t$  dans  $]0, 1[$  tel que  $|s_{n-1}(t)| > g(t)$  avec  $g(t) = at + \frac{a}{2}$  pour un  $a > 0$ .

$s_{n-1}(\cdot)$  désigne la ligne polygonale partant de 0 et joignant les points  $(\frac{k-1}{n-1}, \frac{\sum_{i=2}^k \omega_i}{n-1})$ ,  $k = 2, 3, \dots, n$ .

I.1.1. a) Calcul du niveau

Sous  $H_0$ , les  $\omega_k$  sont des variables aléatoires indépendantes  $\mathcal{N}(0, \sigma^2)$ , leur transformée de Cramer est  $h(u) = \frac{u^2}{2\sigma^2}$ .

D'après le chapitre n° IV ([6]), on a :

$$\forall a > 0, \frac{1}{n-1} \text{Log } P_{H_0}(\exists t, |s_{n-1}(t)| > at + \frac{a}{2}) \longrightarrow - \inf_t t \cdot h\left(\frac{at + \frac{a}{2}}{t}\right) = - \frac{a^2}{\sigma^2}$$

ce qui signifie en particulier que ce test a toujours un niveau exponentiel.

I.1.1. b) Calcul de l'erreur de 2ème espèce

Sous  $H_1$ , les  $y_{k,n}$  suivent une loi normale de variance  $\sigma^2$  et de moyenne égale à  $m_1$  pour  $k < nt_0$  et égale à  $m_2$  pour  $k \geq nt_0$ . Alors les résidus récurrents sont indépendants et de loi  $\mathcal{N}(E\omega_k, \sigma^2)$  où :

$$E \omega_k = 0 \quad \text{pour } k < nt_0$$

$$E \omega_k = \frac{\frac{[nt_0]}{k-1} \cdot (m_2 - m_1)}{\sqrt{1 + \frac{1}{k-1}}} = \frac{\frac{[nt_0]}{n} \cdot d}{\sqrt{\frac{k}{n} \left(\frac{k-1}{n}\right)}} \sim \frac{t_0 \cdot d}{\frac{k}{n}} \quad \text{pour } k \geq nt_0.$$

D'après le chapitre n° IV ([6]), l'erreur de 2ème espèce se calcule ainsi :

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1}(|s_{n-1}(t)| < at + \frac{a}{2}, \forall t) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0}(s_{n-1}(t) < at + \frac{a}{2} - t_0 |d| \cdot \mathbb{1}_{t \geq t_0}, \forall t) \\ &= -\omega(g^*) \quad \text{en utilisant les notations du chapitre n° IV ([6], §III.3)}. \end{aligned}$$

Pour avoir une décroissance exponentielle, il faut déjà que :

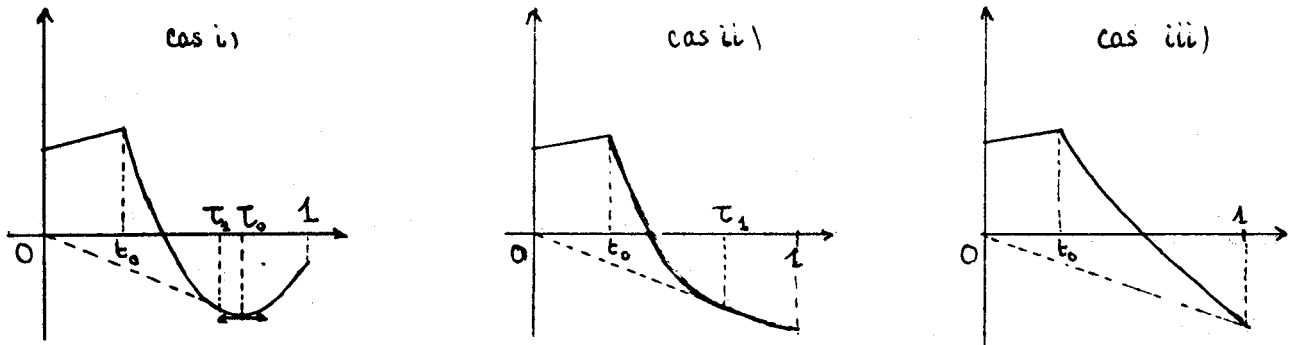
$$\inf_t g(t) \text{ soit négatif. En particulier } g'(t) = a - \frac{t_0 |d|}{t} \cdot \mathbb{1}_{t \geq t_0}$$

doit être négatif en  $t_0$ , ce qui nécessite :  $a < |d|$ .

• Si  $t_0 < \frac{a}{|d|} < 1$ , alors  $g'$  s'annule au point  $\tau_0 = \frac{t_0 |d|}{a} \in [t_0, 1]$ ,

ce qui donne :  $\inf_t g(t) = g(\tau_0) = a \left( t_0 \frac{|d|}{a} + \frac{1}{2} - t_0 \frac{|d|}{a} \text{Log} \frac{|d|}{a} \right)$

• Si  $\frac{a}{|d|} < t_0$ , alors  $\inf_t g(t) = g(1) = a \left( \frac{3}{2} + t_0 \frac{|d|}{a} \text{Log} t_0 \right)$ . Trois cas seront à distinguer pour le calcul explicite :



• Si  $|d| \leq a$ , pas de décroissance exponentielle de l'erreur.

• Si  $t_0 \leq \frac{a}{|d|} < 1$  et  $\frac{a}{|d|} \cdot \frac{1}{2t_0} + 1 + \text{Log} \frac{a}{|d|} \geq 0$ , pas de décroissance exponentielle de l'erreur.

• Cas i)  $t_0 \leq \frac{a}{|d|} < 1$  et  $\frac{a}{|d|} \cdot \frac{1}{2t_0} + 1 + \text{Log} \frac{a}{|d|} < 0$ , la deuxième inégalité peut s'écrire  $\frac{a}{|d|} < \mu(t_0)$  avec  $\mu(t_0)$  solution de  $\frac{\mu}{2t_0} + 1 + \text{Log} \mu = 0$ .

Cette valeur  $\mu(t_0)$  est supérieure à  $t_0$  si et seulement si  $t_0$  est inférieur à  $e^{-3/2}$ . Par conséquent, le cas i) ne se produit pas si  $t_0$  est supérieur à  $e^{-3/2}$ . Notons  $\tau_1$  la racine comprise entre  $t_0$  et  $t_0 \frac{|d|}{a}$  de l'équation :

$g'(t) = \frac{g(t)}{t}$  ;  $\tau_1$  est nécessairement inférieur à  $\tau_0$  dans ce cas puisque la fonction  $g$  est convexe sur  $[t_0, 1]$ . On obtient alors :

$$\begin{aligned} \omega(g^*) &= \int_0^1 h(g^{*'}(t)) dt = \frac{g(\tau_1)^2}{2\sigma^2 \tau_1} + \int_{\tau_1}^1 \frac{g'(t)^2}{2\sigma^2} dt \\ &= t_0 \frac{d^2}{\sigma^2} \left\{ e^{-(1 + \frac{a}{2t_0|d|})} + \frac{a}{|d|} \text{Log} \frac{a}{|d|} + \frac{1}{2t_0} \cdot \frac{a^2}{d^2} \right\} \end{aligned}$$

• Si  $t_0 \geq e^{-3/2}$  et  $t_0 \leq \frac{a}{|d|} \leq \frac{2}{3} t_0 \text{Log} \frac{1}{t_0}$ , alors  $g(1)$  est positif et par conséquent, pas de décroissance exponentielle.

• si  $t_0 < e^{-3/2}$  et  $\frac{a}{|d|} < t_0$  ou bien si  $t_0 \geq e^{-3/2}$  et  $\frac{a}{|d|} < \frac{2}{3} t_0 \text{Log} \frac{1}{t_0}$ , on est amené à distinguer deux cas suivant que  $\tau_1$  est inférieur à 1 ou non ;

le cas ii) correspond à  $\tau_1 \leq 1$ , ce qui se produit si  $t_0 < e^{-1}$  et  $\frac{a}{|d|} < 2t_0(\text{Log } \frac{1}{t_0} - 1)$ , cette dernière borne étant plus petite que  $\frac{2}{3} t_0 \text{Log } \frac{1}{t_0}$  si  $t_0 \geq e^{-3/2}$  et plus grande que  $t_0$  dans le cas contraire. Le cas iii) correspond à  $\tau_1 > 1$ , ce qui se produit pour  $t_0 > e^{-1}$  ou bien  $t_0 < e^{-1}$  et  $\frac{a}{|d|} > 2t_0(\text{Log } \frac{1}{t_0} - 1)$ .

• Cas ii)  $t_0 \leq e^{-3/2}$  et  $\frac{a}{|d|} < t_0$

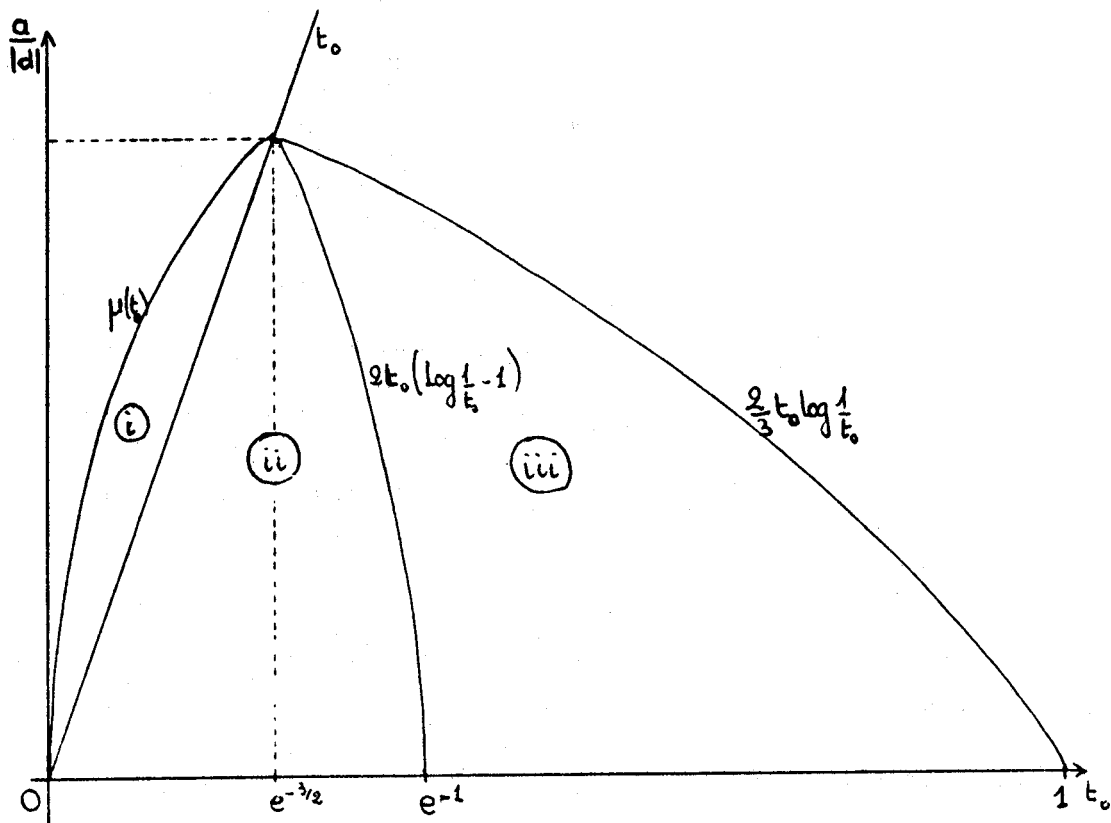
ou  $e^{-3/2} \leq t_0 \leq e^{-1}$  et  $\frac{a}{|d|} \leq 2t_0(\text{Log } \frac{1}{t_0} - 1)$

alors  $\omega(g^*) = \frac{g(\tau_1)^2}{2\sigma^2 \tau_1} + \int_{\tau_1}^1 \frac{g'(t)^2}{2\sigma^2} dt = \frac{t_0 d^2}{\sigma^2} \left\{ \frac{a^2}{d^2} \cdot \frac{1}{t_0} + e^{-(1 + \frac{a}{2t_0|d|})} + \frac{a}{|d|} \cdot \text{Log } t_0 - t_0 \right\}$

• Cas iii)  $t_0 > e^{-1}$  et  $\frac{a}{|d|} < \frac{2}{3} t_0 \text{Log } \frac{1}{t_0}$

ou  $e^{-3/2} \leq t_0 \leq e^{-1}$  et  $2t_0(\text{Log } \frac{1}{t_0} - 1) < \frac{a}{|d|} < \frac{2}{3} t_0 \text{Log } \frac{1}{t_0}$

alors  $\omega(g^*) = \frac{g(1)^2}{2\sigma^2} = \frac{t_0 d^2}{\sigma^2} \left\{ \frac{9}{8t_0} \cdot \frac{a^2}{d^2} + \frac{3a}{2|d|} \text{Log } t_0 + \frac{t_0}{2} \cdot (\text{Log } t_0)^2 \right\}$ .





I.1.2)  $\sigma$  inconnu : niveau

La région de rejet est la même que la précédente si ce n'est que  $\sigma^2$  est ici estimé par l'estimateur classique des moindres carrés

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n (y_k - \bar{y}^n)^2 = \frac{1}{n-1} \sum_{k=2}^n \omega_k^2 \quad (\text{voir [1]}) :$$

$$D = \{ \} t \in [0, 1] \text{ tel que } \frac{|s_{n-1}(t)|}{\hat{\sigma}} > bt + \frac{b}{2} .$$

Du point de vue fonctionnel, D peut s'écrire :

$$D = \{ f \text{ telle que } \} t \in [0, 1] , \frac{|f(t)|}{\left( \int_0^1 f'(t)^2 dt \right)^{1/2}} > bt + \frac{b}{2} .$$

C'est un exemple d'ensemble fonctionnel où la théorie exposée au chapitre n° IV ([6]) ne donne aucune précision : en effet, l'intérieur de D est vide et sa fermeture contient 0 de sorte que  $W(\overset{\circ}{D}) = +\infty$  et  $W(\bar{D}) = 0$ .

• étant donné l'aspect homogène de la contrainte, il est évident que si D n'est pas vide, alors  $\bar{D}$  contient la fonction identiquement nulle.

- si  $b < \frac{\sqrt{2}}{2}$ , alors il existe  $\tau_2$  compris entre 0 et 1 tel que  $\sqrt{\tau_2} > b\tau_2 + \frac{b}{2}$  alors toute fonction f linéaire sur  $[0, \tau_2]$  et constante ensuite appartient à D  $f(t) = C \cdot t \cdot 1_{t < \tau_2} + C \cdot \tau_2 \cdot 1_{t > \tau_2}$  donc D ne sera pas vide .

- si  $b \geq \frac{\sqrt{2}}{2}$ , alors D est vide : il est clair en <sup>effet</sup> qu'il suffit de s'intéresser aux fonctions linéaires sur  $[0, \tau_2]$  car si une fonction f appartient à D, il existe  $\tau$  tel que  $\frac{f(\tau)}{\left( \int_0^1 f'(t)^2 dt \right)^{1/2}} > b\tau + \frac{b}{2}$ , alors la fonction

linéaire :  $\frac{f(\tau)}{\tau} \cdot t \cdot 1_{t \leq \tau} + f(\tau) \cdot 1_{t > \tau}$  est a fortiori dans D. D étant vide,

$\overset{\circ}{D}$  et  $\bar{D}$  le sont aussi.

• Même dans le cas où  $b < \frac{\sqrt{2}}{2}$ , on constate immédiatement que  $\overset{\circ}{D}$  est vide car dans tout  $\varepsilon$ -tube construit autour de f, il existe des fonctions qui "oscillent" trop pour appartenir à D.

Néanmoins,  $P_{H_0}(D)$  décroît exponentiellement comme le montre l'encadrement suivant dans le cas où  $b < \frac{1}{\sqrt{2}}$  :

$$\bullet \forall \varepsilon > 0, P_{H_0}(D) < P_{H_0} \left\{ \exists t, \frac{|S_{n-1}(t)|}{\sigma(1-\varepsilon)} > bt + \frac{b}{2} \right\} + P_{H_0} \{ \hat{\sigma} < \sigma(1-\varepsilon) \}$$

$$\text{or : } \frac{1}{n} \text{Log } P_{H_0} \left\{ \exists t, \frac{|S_{n-1}(t)|}{\sigma} > (1-\varepsilon)bt + \frac{(1-\varepsilon)b}{2} \right\} \longrightarrow -b^2(1-\varepsilon)^2$$

$$\frac{1}{n} \text{Log } P_{H_0} \{ \hat{\sigma} < \sigma(1-\varepsilon) \} \longrightarrow -\frac{(1-\varepsilon)^2 - 1 - \text{Log}(1-\varepsilon)^2}{2}$$

$$\overline{\lim}_n \frac{1}{n} \text{Log } P_{H_0}(D) < -\inf \left\{ b^2(1-\varepsilon)^2, \frac{(1-\varepsilon)^2 - 1 - \text{Log}(1-\varepsilon)^2}{2} \right\}$$

$$\bullet \forall \varepsilon > 0, P_{H_0}(D) > P_{H_0} \left\{ \exists t, \frac{|S_{n-1}(t)|}{\sigma(1+\varepsilon)} > bt + \frac{b}{2} \right\} - P_{H_0} \{ \hat{\sigma} > \sigma(1+\varepsilon) \}$$

$$\frac{1}{n} \text{Log } P_{H_0} \left\{ \exists t, \frac{|S_{n-1}(t)|}{\sigma} > (1+\varepsilon)bt + \frac{(1+\varepsilon)b}{2} \right\} \longrightarrow -b^2(1+\varepsilon)^2$$

$$\frac{1}{n} \text{Log } P_{H_0} \{ \hat{\sigma} > \sigma(1+\varepsilon) \} \longrightarrow -\frac{(1+\varepsilon)^2 - 1 - \text{Log}(1+\varepsilon)^2}{2}$$

$$\sup_{\varepsilon} \frac{(1+\varepsilon)^2 - 1 - \text{Log}(1+\varepsilon)^2}{2(1+\varepsilon)^2} = \frac{1}{2} > b^2$$

$$\text{donc } \exists \varepsilon_0 > 0 \text{ tel que } b^2(1+\varepsilon_0)^2 < \frac{(1+\varepsilon_0)^2 - 1 - \text{Log}(1+\varepsilon_0)^2}{2}$$

$$\text{on en déduit que : } \liminf_n \frac{1}{n} \text{Log } P_{H_0}(D) > -b^2(1+\varepsilon_0)^2.$$

## I.2. Test construit sur les sommes des carrés de résidus - $\sigma$ connu

### I.2.1) Niveau

La région de rejet s'écrit :  $\{ \exists k \in \{2, 3, \dots, n\} \text{ tel que}$

$$\frac{1}{(n-1)\sigma^2} \sum_{i=2}^k \omega_i^2 > \frac{k-1}{n-1} + c \}$$
 soit finalement :  $\{ \exists t \in [0, 1] \text{ tel que } ss_{n-1}(t) > t+c \}$

avec  $c > 0$ , où  $ss_{n-1}(\cdot)$  désigne la ligne polygonale joignant l'origine 0 et

$$\text{les points } \left\{ \frac{k-1}{n-1}; \frac{1}{(n-1)\sigma^2} \sum_{i=2}^k \omega_i^2 \right\} \text{ pour } k = 2, \dots, n.$$

On a vu dans le chapitre n° IV ([6]) que :

$$\frac{1}{n} \text{Log } P_{H_0} (\exists t, ss_{n-1}(t) > t+c) \longrightarrow - \inf_t t \cdot h_0 \left( \frac{t+c}{t} \right)$$

où  $h_0$  est la transformée de Cramer d'une variable suivant une loi  $\chi_1^2$  soit ;

$$h_0(x) = \frac{x-1 - \text{Log } x}{2}, \quad x > 1.$$

Donc :  $\forall c > 0, \frac{1}{n} \text{Log } P_{H_0} (\exists t, ss_{n-1}(t) > t+c) \longrightarrow - \frac{c - \text{Log}(1+c)}{2} = - \mathcal{H}_0(c)$ .

On remarque, c'est la même vitesse que pour le test admettant pour région de rejet seulement  $\left\{ \frac{1}{(n-1)\sigma^2} \sum_{i=2}^n \omega_i^2 > 1+c \right\}$ .

### I.2.2) Erreur de 2ème espèce

Sous  $H_1$ , les variables  $\omega_k$  sont indépendantes et suivant des lois normales de variance  $\sigma^2$  et d'espérance :

$$E \omega_k = \frac{t_0 d}{\sqrt{\frac{k}{n} \left( \frac{k}{n} - 1 \right)}} \cdot 1_{k > n t_0}$$

On a donc à évaluer :  $\frac{1}{n} \text{Log } P_{H_1} (\forall k, \frac{1}{(n-1)\sigma^2} \sum_{i=2}^k \omega_i^2 \leq \frac{k-1}{n-1} + c)$ . Ce

problème n'a pas été traité explicitement dans le chapitre n° IV ([6]), car la suite  $(\omega_k)$ ,  $k = 2, \dots, n$  n'est plus stationnaire. Toutefois, on a de même (cf. [2]) :

$$\frac{1}{n} \text{Log } P_{H_1} (\forall k, \frac{1}{(n-1)\sigma^2} \sum_{i=2}^k \omega_i^2 \leq \frac{k-1}{n-1} + c) \longrightarrow - \inf_{f \in D} \int_0^1 h_t(f'(t)) dt$$

où  $D = \{f \text{ telle que } \forall t \in [0, 1], f(t) \leq t+c \text{ et } f(0) = 0\}$

et  $h_t(\cdot)$  désigne la transformée de Cramer de la variable  $\frac{1}{\sigma^2} \omega_{[n]}^2$ , soit ici :

$$h_t(x) = \sup_u \left( xu + \frac{1}{2} \text{Log}(1-2u) - \frac{t_0^2 d^2}{\sigma^2 t^2} \cdot \frac{u}{1-2u} \cdot 1_{t > t_0} \right)$$

On est amené à résoudre le problème suivant :

Trouver  $\inf_f \int_0^1 h_1(t, f'(t)) dt$  sous la contrainte :

$f(0) = 0$  et  $f(t) \leq C$ ,  $\forall t \in [0, 1]$  si on note  $h_1(t, v) = h_t(v+1)$ . On utilisera le fait que  $h_1$  est décroissante en  $t$  et convexe en  $v$ . Ce problème est équivalent au suivant :

"Trouver  $\inf_g \int_0^1 \int_0^{\gamma(t)} u(t, v) dv$  sous la contrainte :  $\int_0^x g(t) dt \leq c$ ,

$\forall x \in [0, 1]$ , avec :

$$u(t, v) = -\frac{\partial h_1}{\partial v}(t, v)$$

$$\gamma(t) \text{ étant défini par } u(t, \gamma(t)) = 0 \iff \gamma(t) = \frac{t_0 d^2}{\sigma^2 t^2} \cdot 1_{t > t_0}$$

Remarque 1 : Il est évident que la décroissance de l'erreur ne sera pas exponentielle si la fonction  $\gamma(\cdot)$  respecte la contrainte :

$$\int_{t_0}^1 \gamma(t) dt \leq c \text{ soit si } t_0(1-t_0) \frac{d^2}{\sigma^2} < c.$$

Pour  $c \in ]0, t_0(1-t_0) \frac{d^2}{\sigma^2} [$ , on va exhiber la fonction  $g$  réalisant l'infimum.

Remarque 2 : A  $t$  fixé, la convexité de  $h_1$  par rapport à  $v$  implique que  $u(t, v)$  est une fonction décroissante de  $v$  et que les courbes de niveau  $g_\lambda(t)$  définies par :  $v = g_\lambda(t) \iff u(t, v) = \lambda > 0$  sont des fonctions décroissantes de  $t$  sur  $[t_0, 1]$ . On a :

$$g_\lambda(t) = \frac{-2\lambda}{1+2\lambda} + \frac{t_0^2 d^2}{\sigma^2 t^2} \cdot \frac{1}{(1+2\lambda)^2} \cdot 1_{t > t_0}$$

Remarque 3 : Il suffit de considérer les fonctions  $g$  inférieures à  $\gamma$  car pour toute fonction  $g$  réalisant la contrainte, la fonction  $\inf(g, \gamma)$  réalise aussi la contrainte (car  $\gamma \geq 0$ ) et donne une intégrale plus petite.

Nous serons amenés à distinguer deux cas selon que la courbe de niveau  $g_\lambda$

réalisant  $\int_0^1 g_\lambda(t) dt = c$  vérifie :  $g_\lambda(t) > 0$  pour  $t \in [t_0, 1]$  ou non .

Cas i) la courbe de niveau  $g_\lambda(t) = \frac{-2\lambda}{1+2\lambda} + \frac{t_0^2 d^2}{\sigma^2 t^2} \cdot \frac{1}{(1+2\lambda)^2} \cdot 1_{t > t_0}$  telle que  $\int_0^1 g_\lambda(t) dt = c$  , vérifie  $g_\lambda \geq 0$  sur  $[t_0, 1]$ .

Alors montrons que  $g_\lambda$  est meilleure que toute fonction  $g$  inférieure à  $\gamma$  et réalisant  $\int_0^1 g(t) dt \leq c$  .

$$\begin{aligned} \int_0^1 dt \int_{g(t)}^{\gamma(t)} u(t,v) dv &= \int_0^1 dt \int_{g v g_\lambda(t)}^{\gamma(t)} u(t,v) dv + \int_0^1 dt \int_{g(t)}^{g_\lambda(t) v g} u(t,v) dv \\ &\geq \int_0^1 dt \int_{g v g_\lambda(t)}^{\gamma(t)} u(t,v) dv + \lambda \int_0^1 dt \int_{g(t)}^{g_\lambda(t) v g} dv \\ &= \int_0^1 dt \int_{g v g_\lambda(t)}^{\gamma(t)} u(t,v) dv + \lambda \left[ \int_0^1 g v g_\lambda(t) dt - \int_0^1 g(t) dt \right] \\ &\geq \int_0^1 dt \int_{g v g_\lambda(t)}^{\gamma(t)} u(t,v) dv + \lambda \left[ \int_0^1 g_\lambda v g(t) dt - \int_0^1 g_\lambda(t) dt \right] \\ &\geq \int_0^1 dt \int_{g v g_\lambda(t)}^{\gamma(t)} u(t,v) dv + \int_0^1 \int_{g_\lambda(t)}^{g(t) v g} u(t,v) dv \\ &= \int_0^1 dt \int_{g_\lambda(t)}^{\gamma(t)} u(t,v) dv . \end{aligned}$$

On est dans ce cas si  $t_0 \geq \frac{1}{2}$

ou si  $t_0 < \frac{1}{2}$  et 
$$\frac{4t_0(1-2t_0) \frac{d^2}{\sigma^2}}{\left(1 + \sqrt{1 + \frac{4t_0^2 d^2}{\sigma^2}}\right)^2} \leq c < t_0(1-t_0) \frac{d^2}{\sigma^2}$$

on obtient : 
$$\inf_{f \in D} \int_0^1 h_t(f'(t)) dt = \frac{1}{2} \left\{ \text{Log } \Lambda - \frac{\Lambda-1}{\Lambda} + \frac{t_0(1-t_0)d^2}{\sigma^2} \cdot \left(\frac{\Lambda-1}{\Lambda}\right)^2 \right\}$$

avec 
$$\Lambda = 1 + 2\lambda = \frac{1 + \sqrt{1 + 4t_0(1-t_0) \frac{d^2}{\sigma^2}} (1+c)}{2(1+c)}$$

Cas ii) la courbe de niveau  $g_\lambda$  réalisant  $\int_0^1 g_\lambda(t)dt = c$  est telle que  $g_\lambda(1) < 0$ . Nous allons montrer que la courbe réalisant l'infimum est  $\tilde{g}_\lambda$  définie par :

$$\tilde{g}_\lambda(t) = g_\lambda(t)(1_{t < t_0} + 1_{t > t_0} \cap g_\lambda(t) > 0), \lambda \text{ étant tel que } \int_0^1 \tilde{g}_\lambda(t)dt = c$$

Si  $g$  est une fonction réalisant l'infimum, nécessairement  $g$  réalise l'égalité dans la contrainte :  $\int_0^x g(t)dt \leq c$  pour tout  $x$ ; posons alors :

$$\tau_g = \inf \left\{ \tau, \int_0^\tau g(t)dt = c \right\}.$$

Nécessairement  $g$  est nulle pour  $t$  supérieur à  $\tau_g$ . En effet, sur  $[\tau_g, 1]$ , toute "arche de type  $g \geq 0$ " sera précédée d'une ou plusieurs "arches de type  $g \leq 0$ " de surface au moins égale car  $\int_0^\tau g(t)dt = c$  et  $\int_0^x g(t)dt \leq c$  pour tout  $x$ . Ces deux arches sont alors séparées par une courbe de niveau  $g_{\lambda'}, \lambda' > 0$ ; par un raisonnement analogue au cas i), on obtient :

$$\int_{\tau_g}^1 dt \int_{g(t)}^{\gamma(t)} u(t,v)dv \geq \int_{\tau_g}^1 dt \int_0^{\gamma(t)} u(t,v)dv.$$

La fonction  $g$  étant optimale, elle est donc nulle sur  $[\tau_g, 1]$ .

o Un argument analogue permet d'affirmer que  $g$  est nécessairement positive sur  $[t_0, \tau_g]$ .

o Pour achever la démonstration, il suffit de faire la même démonstration que dans le cas i)

$$\begin{aligned} \int_0^1 dt \int_{g(t)}^{\gamma(t)} u(t,v)dv &= \int_0^1 dt \int_{g \vee \tilde{g}_\lambda}^{\gamma(t)} u(t,v)dv + \int_0^1 dt \int_g^{g \vee \tilde{g}_\lambda} u(t,v)dv \\ &\geq \int_0^1 dt \int_{g \vee \tilde{g}_\lambda}^{\gamma(t)} u(t,v)dv + \lambda \left( \int_0^1 g \vee \tilde{g}_\lambda(t)dt - \int_0^1 g(t)dt \right) \\ &= \int_0^1 dt \int_{g \vee \tilde{g}_\lambda}^{\gamma(t)} u(t,v)dv + \lambda \left( \int_0^1 g \vee \tilde{g}_\lambda(t)dt - \int_0^1 \tilde{g}_\lambda(t)dt \right) \end{aligned}$$

D'où

$$\int_0^1 dt \int_{g(t)}^{\gamma(t)} u(t, v) dv \geq \int_0^1 dt \int_{g_\lambda(t)}^{\gamma(t)} u(t, v) dv .$$

On est dans ce cas si  $t_0 < \frac{1}{2}$  et  $0 < c < \frac{4t_0(1-2t_0)\frac{d^2}{\sigma^2}}{(1+\sqrt{1+4t_0^2\frac{d^2}{\sigma^2}})^2}$

on obtient ainsi :  $\inf_{f \in D} \int_0^1 h_t(f'(t)) dt = \frac{1}{2} \text{Log} \frac{1+\sqrt{1+4t_0^2\frac{d^2}{\sigma^2}}}{2} + 1 - \frac{t_0^2 d^2}{2\sigma^2}$

$$+ t_0 \frac{|d|}{\sigma} \cdot \frac{\Lambda^2 - 1}{\sqrt{\Lambda(\Lambda - 1)}} - \sqrt{1 + 4t_0^2 \frac{d^2}{\sigma^2}} + \frac{t_0 d^2}{2\sigma^2} \cdot (\frac{\Lambda - 1}{\Lambda})^2 + \frac{t_0 |d|}{2\sigma} \cdot \text{Log} \frac{2t_0 \frac{|d|}{\sigma} + \sqrt{1 + 4t_0^2 \frac{d^2}{\sigma^2}}}{2\sqrt{\Lambda(\Lambda - 1)} + (2\Lambda - 1)}$$

avec  $\Lambda \neq 1 + 2\lambda$  solution de :  $[\Lambda^2 \frac{c}{t_0} - \frac{d^2}{\sigma^2}]^2 - 4 \frac{d}{\sigma} \Lambda(\Lambda - 1) = 0 .$

II - ETUDE D'UN AUTRE TEST POUR DECELER UNE RUPTURE DANS UN MODELE

Nous avons supposé que les  $y_{k,n}$ ,  $k = 1, \dots, n$  sont indépendantes et suivent une loi  $F_1$  pour  $k < nt_0$  et une loi  $F_2$  pour  $k \geq nt_0$ . Notre problème est toujours de déterminer s'il y a une rupture effective ( $t_0 \neq 0$ , et  $F_1 \neq F_2$ ) ou non. On a vu dans le paragraphe précédent qu'une des difficultés sous-jacentes était la mauvaise estimabilité de  $t_0$  pour de telles observations. On va essayer de contourner cette difficulté en "randomisant" les instants d'observations.

Au lieu d'observer le phénomène à des instants  $\frac{k}{n}$  fixes, on va tirer au sort les instants d'observation, par exemple avec une loi  $\Pi^{\otimes n}$  sur  $[0,1]^n$ , de sorte qu'on obtient un n-échantillon  $(X_1, X_2, \dots, X_n)$  avec  $X_j = Y_{U_j}$  où  $(U_1, U_2, \dots, U_n)$  est un n-échantillon de la loi  $\Pi$  sur  $[0,1]$  et les variables  $Y_{u_j,n}$  suivent la loi  $F_1$  si  $u_j < t_0$  et la loi  $F_2$  si  $u_j \geq t_0$ . Si  $\Pi$  est une loi absolument continue par rapport à la mesure de Lebesgue sur  $[0,1]$  et si on note  $\pi_0 = \Pi(U < t_0)$ , alors  $(X_1, X_2, \dots, X_n)$  est un n-échantillon de la loi mélange :  $\pi_0 F_1 + (1 - \pi_0)F_2$  dont les deux premiers moments sont :  $\pi_0 m_1 + (1 - \pi_0)m_2$  et  $\sigma^2 + \pi_0(1 - \pi_0)d^2$  (on a noté  $m_1$  et  $\sigma^2$  les moments de  $F_1$ ,  $m_2$  et  $\sigma^2$  les moments de  $F_2$  et  $d = m_1 - m_2$ ).

Sous cette forme, l'hypothèse naturelle à tester est :

$$H_0 = \{ \pi_0 (1 - \pi_0) d^2 = 0 \}$$

et la méthode des moments amène à comparer l'estimateur classique de la variance à  $\sigma^2$ .

On va étudier ce test d'un point de vue asymptotique dans le cas très simple où  $F_1 = \mathcal{N}(m_1, \sigma^2)$  et  $F_2 = \mathcal{N}(m_2, \sigma^2)$ . Pour faciliter l'exposition, on conservera la randomisation, sous la forme  $\Pi^{\otimes n}$ , des temps d'observation. En réalité, aussi bien au niveau théorique que pratique, on procédera un peu différemment (l'équivalence des deux procédures étant montrée en appendice) :  $n$  étant fixé, nous supposons disposer de  $N(n)$  (avec  $\frac{N(n)}{n} \xrightarrow{n \rightarrow \infty} \infty$ ) temps possibles d'observations disposés sur  $[0, 1]$  :  $U_1, U_2, \dots, U_N$  et leurs observations associées  $Y_{U_1}, \dots, Y_{U_N}$  avec  $Y_{U_i}$  suit la loi  $F_1$  si  $U_i < t_0$  et la loi  $F_2$  si  $U_i > t_0$ ; et on va effectuer un tirage uniforme et sans remise des  $n$  instants parmi les  $N(n)$ . L'arbitraire dans la loi  $\Pi$  se retrouvant ici dans le choix arbitraire des instants  $U_1, U_2, \dots, U_N$  possibles.

On verra que la loi  $\Pi$  peut être choisie de façon plus ou moins appropriée et qu'on aura intérêt à utiliser l'éventuelle information a priori sur l'instant de rupture  $t_0$ .

## II.1. Définition du test

### II.1.1) $\sigma$ connu

La région de rejet sera du type :  $D = \left\{ \frac{1}{(n-1)\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 > 1+c \right\}$  avec  $c > 0$  déterminé par le niveau  $\alpha$  du test :  $P_{H_0}(D) = \alpha$ .

### II.1.1) $\sigma$ inconnu

Dans ce cas, on conserve le même principe de test mais on remplace  $\sigma^2$  par une estimation faite à l'aide d'un deuxième échantillon de taille  $n$  :  $V_1, V_2, \dots, V_n$  réalisé suivant la même procédure que  $X_1, X_2, \dots, X_n$  mais les instants d'observation étant cette fois tiré selon une loi  $\Pi_n^{\otimes n}$ ,  $\Pi_n$  ayant



pour support  $[0, a_n]$  avec  $a_n$  : suite décroissante vers 0. Ce second échantillon est supposé indépendant du premier et forme un  $n$ -échantillon de la loi  $\pi'_{n,0} F_1 + (1 - \pi'_{n,0}) F_2$  : il est clair que pour  $t_0 > 0$ ,  $\pi'_{n,0}$  vaut 1 pour  $n$  assez grand. D'un point de vue plus pratique, si on présume que la rupture se produit après un instant  $t_1$  ou avant un instant  $t_2$ , on prend alors  $\pi'_n = \Pi'$  loi chargeant uniquement  $[0, t_1]$  ou  $[t_2, 1]$  respectivement.

On pourra donc tester l'hypothèse  $H_0$  à l'aide de la région de rejet

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} > 1 + c$$

avec  $c > 0$  déterminé par le niveau du test.

## II.2. Propriétés asymptotiques

Soit maintenant un  $n$ -échantillon  $(X_1, X_2, \dots, X_n)$  de loi  $F = F_X$ , nous noterons :

$$\phi_n(t) = \int \dots \int e^{-t \sum_{i=1}^n (x_i - \bar{x})^2} dF(x_1) \dots dF(x_n)$$

On suppose que  $\frac{1}{n} \text{Log } \phi_n(t)$  converge vers une fonction notée  $\text{Log } \phi(t)$  en tout point sauf éventuellement à la frontière du domaine de définition de  $\text{Log } \phi$ . On suppose de plus que 0 est un point intérieur à ce domaine. Si on note  $h(\cdot)$  La duale de Young de  $\text{Log } \phi(\cdot)$ , d'après le chapitre n° I ([3]), on a :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)\sigma^2} > 1+c\right) = -h((1+c)\sigma^2), \quad \forall c > 0$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)\sigma^2} < 1+c\right) = -h((1+c)\sigma^2), \quad \forall c < 0$$

### II.2.1) Application à des lois particulières

a) F est la loi  $\mathcal{N}(m, \sigma^2)$

On a dans ce cas  $\phi_n(t) = \left(\frac{1}{1 - 2\sigma^2 t}\right)^{n-1}$ , ce qui donne :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log} P\left(\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)\sigma^2} > 1+c\right) = -H_0(c), \quad c > 0$$

avec 
$$H_0(c) = \frac{c - \text{Log}(1+c)}{2} .$$

b) F est la loi de Bernouilli de paramètre  $\pi_0$

Cette loi F a l'intérêt de montrer que  $\sum_{i=1}^n (X_i - \bar{X})^2$  ne se comporte pas

forcément comme  $\sum_{i=1}^n [X_i - E(X)]^2$  ; elle introduira d'autre part plus facilement

la loi mélange de deux gaussiennes.

$$\phi_n(t) = \sum_{k=0}^n C_n^k \pi_0^k (1 - \pi_0)^{n-k} e^{nt \frac{k}{n} (1 - \frac{k}{n})}$$

Il est facile de montrer que tout  $t$  réel, la limite de  $\frac{1}{n} \text{Log} \phi_{F,n}(t)$  existe et vaut :

$$\text{Log} \phi(t) = \alpha \text{Log} \frac{\pi_0}{\alpha} + (1-\alpha) \text{Log} \frac{1-\pi_0}{1-\alpha} + t \alpha(1-\alpha)$$

avec  $\alpha = \alpha(t)$  égal à l'unique solution comprise entre 0 et 1 de l'équation :

$$\text{Log} \frac{1-\alpha}{\alpha} \cdot \frac{\pi_0}{1-\pi_0} + t(1-2\alpha) = 0 .$$

Toutefois, plutôt que de chercher la duale de Young de  $\text{Log} \phi_F(\cdot)$ , il est plus simple de calculer directement :

$$\begin{aligned} P\left(\sum_{i=1}^n (X_i - \bar{X})^2 < (n-1)\sigma^2(1+c)\right) & \quad (-1 < c < 0) \\ = P\left[\bar{X}(1-\bar{X}) < \frac{n-1}{n} \cdot \pi_0(1-\pi_0)(1+c)\right] \end{aligned}$$

Pour  $0 < \pi_0(1-\pi_0)(1+c) < \frac{1}{4}$ , notons  $u$  la plus petite racine de l'équation :

$$u(1-u) = \pi_0(1-\pi_0)(1+c) .$$

On obtient alors, par application simple des théorèmes de grandes déviations sur les loi de Bernouilli (cf. par exemple [5]), en désignant par

$$\bar{\pi}_0 = \inf(\pi_0, 1 - \pi_0)$$

$$P_{\pi_0}(\bar{X}(1-\bar{X}) < (1-u)) = P_{\pi_0}(\bar{X} < u) + P_{\pi_0}(\bar{X} > 1-u)$$

$$\leq 2 P_{\pi_0}(\bar{X} < u)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{\pi_0}(\bar{X}(1-\bar{X}) < u(1-u)) = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{\pi_0}(\bar{X} < u)$$

$$= u \text{Log } \frac{\pi_0}{u} + (1-u) \text{Log } \frac{1-\pi_0}{1-u}$$

Ce résultat appelle les deux remarques suivantes :

i) pour  $c > 0$ ,  $\sum_{i=1}^n (X_i - \bar{X})^2 > (n-1)\sigma^2(1+c)$  n'est pas une queue exponentielle

ii) pour  $c < 0$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{\pi_0}(\sum_{i=1}^n (X_i - \pi_0)^2 < n \pi_0 (1-\pi_0)(1+c))$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{\pi_0}(\bar{X} < \pi_0 + \frac{c \pi_0 (1-\pi_0)}{1-2\pi_0})$$

$$= (\pi_0 + c \frac{\pi_0 (1-\pi_0)}{1-2\pi_0}) \text{Log } \frac{1}{(1-\pi_0) \frac{1+c}{(1-2\pi_0)}} + (1-\pi_0 - c \frac{\pi_0 (1-\pi_0)}{1-2\pi_0}) \text{Log } \frac{1}{1-c \frac{\pi_0}{1-2\pi_0}}$$

Les deux vitesses exponentielles sont strictement différentes pour

$\sum_{i=1}^n (X_i - \bar{X})^2$  et  $\sum_{i=1}^n (X_i - EX)^2$  dans le cas de variables de Bernoulli :

pour  $c < 0$ ,  $\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P(\sum_{i=1}^n (X_i - \bar{X})^2 < (n-1)\pi_0(1-\pi_0)(1+c)) >$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P(\sum_{i=1}^n (X_i - \pi_0)^2 < (n-1)\pi_0(1-\pi_0)(1+c))$$

c) F est le mélange de deux gaussiennes

$\mathcal{N}(m_1, \sigma^2)$  avec probabilité  $\pi_0$  et  $\mathcal{N}(m_2, \sigma^2)$  avec probabilité  $(1-\pi_0)$ .

On note  $d = m_1 - m_2$  et  $u(t) = \frac{t \cdot d^2}{1 - 2\sigma^2 t}$ ,

$$\phi_n(t) = (1-2\sigma^2 t)^{-\frac{n-1}{2}} \sum_{k=0}^n C_n^k \pi_0^k (1-\pi_0)^{n-k} e^{n \cdot u(t) \cdot \frac{k}{n} (1 - \frac{k}{n})}$$

On déduit du paragraphe précédent que pour tout  $t$  inférieur à  $\frac{1}{2\sigma^2}$ , la fonction  $\frac{1}{n} \text{Log } \phi_n(t)$  converge vers  $\text{Log } \phi(t)$  définie par :

$$\log \phi(t) = -\frac{1}{2} \text{Log}(1-2\sigma^2 t) + \alpha \text{Log} \frac{\pi_0}{\alpha} + (1-\alpha) \text{Log} \frac{1-\pi_0}{1-\alpha} + u(t) \alpha(1-\alpha)$$

avec  $\alpha = \alpha(t)$  solution de :  $\text{Log} \frac{1-\alpha}{\alpha} \cdot \frac{\pi_0}{1-\pi_0} + u(t)(1-2\alpha) = 0$ .

On peut écrire sa duale de Young :

$$h(x) = x \tau + \frac{1}{2} \text{Log}(1-2\sigma^2 \tau) - \alpha \text{Log} \frac{\pi_0}{\alpha} - (1-\alpha) \text{Log} \frac{1-\pi_0}{1-\alpha} - u(\tau) \alpha(1-\alpha)$$

$$\text{avec } \tau = \frac{2x - \sigma^2 - \sqrt{\sigma^4 + 4\alpha(1-\alpha)d^2 x}}{4x \sigma^2}$$

$$\text{et } \alpha(\tau) \text{ solution de : } \text{Log} \frac{\pi_0}{\alpha} \cdot \frac{1-\alpha}{1-\pi_0} + (1-2\alpha) \frac{d^2}{2\sigma^2} \cdot \frac{2x - \sigma^2 - \sqrt{\sigma^4 + 4\alpha(1-\alpha)d^2 x}}{\sigma^2 + \sqrt{\sigma^4 + 4\alpha(1-\alpha)d^2 x}} = 0$$

### II.3. Vitesse exponentielle des erreurs pour ce test

#### II.2.1. $\sigma$ connu

On déduit des paragraphes précédents que pour :  $0 < c < \pi_0(1-\pi_0) \frac{d^2}{\sigma^2}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0} \left( \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)\sigma^2} > 1+c \right) = -H_0(c)$$

$$H_0(c) = \frac{c - \text{Log}(1+c)}{2}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1} \left( \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)\sigma^2} < 1+c \right) = -H_1(c)$$

$$\text{avec } H_1(c) = \frac{c+1}{2} \left[ 1 - \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1)}}{2(c+1)} \right] - \frac{\alpha d^2}{2\sigma^2} \left[ \frac{2(c+1)}{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1)}} - 1 \right] +$$

$$\frac{1}{2} \text{Log} \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1)}}{2(c+1)} - \text{Log} \frac{1-\pi_0}{1-\alpha}$$

où  $\alpha = \alpha(c)$  est la solution de :

$$\text{Log } \frac{\pi_0}{\alpha} \cdot \frac{1-\alpha}{1-\pi_0} + (1-2\alpha) \frac{d^2}{2\sigma^2} \cdot \left[ \frac{2(c+1)}{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2}} (c+1)} - 1 \right] = 0 .$$

On remarque que  $H_1$  dépend de la loi  $\Pi$  choisie pour échantillonner les instants d'observations seulement par l'intermédiaire de  $\pi_0 = \Pi(t < t_0)$  ; et il est facile de montrer que :

$$H_1(c)(\pi_0) \leq H_1(c)(\pi) \leq H_1(c)\left(\frac{1}{2}\right)$$

pour toute valeur de  $\pi$  située entre  $\pi_0$  et  $\frac{1}{2}$  et pour toute valeur de  $c$  positive et inférieure à  $\pi_0(1-\pi_0) \frac{d^2}{\sigma^2}$ . On constate ainsi qu'on aura intérêt à choisir la loi  $\Pi$  la plus simple possible mais en essayant de se rapprocher de la valeur  $\pi_0 = \frac{1}{2}$  : dans la mesure où l'on peut pressentir l'instant de la rupture éventuelle, il faudra faire autant d'observation avant qu'après. Un choix optimal mais théorique de  $\Pi_n$  serait une loi chargeant équitablement les intervalles  $[0, a_n]$  et  $[1-a_n, 1]$  avec  $a_n \xrightarrow{n \rightarrow \infty} 0$ . Plus pratiquement, une rupture présumée dans une plage  $[t_1, t_2]$  conduira à prendre une loi  $\Pi$  chargeant équitablement les intervalles  $[0, t_1]$  et  $[t_2, 1]$ . Avec le meilleur échantillonnage, on obtiendrait :

$$H_1(c)\left(\frac{1}{2}\right) = \frac{c+1}{2} \left( 1 - \frac{1 + \sqrt{1 + \frac{d^2}{\sigma^2}(c+1)}}{2(c+1)} \right) - \frac{d^2}{8\sigma^2} \left( \frac{2(c+1)}{1 + \sqrt{1 + \frac{d^2}{\sigma^2}(c+1)}} - 1 \right) + \frac{1}{2} \text{Log } \frac{1 + \sqrt{1 + \frac{d^2}{\sigma^2}(c+1)}}{2(c+1)}$$

### II.3.2. $\sigma$ inconnu

Par une simple application, dans  $R^2$  cette fois, du chapitre n° I ([3]), on montre que :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0} \left( \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} > c+1 \right) = - \inf_{\substack{(x,y) \in R^2 \\ x \geq (c+1)y}} (\bar{H}_0(x) + \bar{H}_0(y)) \quad \text{si } c > 0$$

$$\text{où } \bar{H}_0(x) = H_0(x-1) = \frac{x-1 - \text{Log } x}{2}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1} \left( \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} < c+1 \right) = - \inf_{\substack{(x,y) \in \mathbb{R}^2 \\ x \leq (c+1)y}} (\bar{H}_0(x) + \bar{H}_1(y)) \quad \text{si } c < \pi_0 (1 - \pi_0) \frac{d^2}{\sigma^2}$$

où  $\bar{H}_1(y) = H_1(y-1) = \frac{y}{2} \left[ 1 - \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} y}}{2y} \right] - \frac{\alpha^2 d^2}{2\sigma^2} \left[ \frac{2y}{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} y}} - 1 \right]$

$$+ \frac{1}{2} \text{Log} \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} y}}{2y} - \text{Log} \frac{1 - \pi_0}{1 - \alpha}$$

avec  $\alpha(y)$  solution de :

$$\text{Log} \frac{\pi_0}{\alpha} \cdot \frac{1 - \alpha}{1 - \pi_0} + (1 - 2\alpha) \frac{d^2}{\sigma^2} \left[ \frac{2y}{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} y}} - 1 \right] = 0 .$$

Les calculs effectués donnent les résultats suivants :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0} \left( \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} > 1+c \right) = - \text{Log} \frac{1 + \frac{c}{2}}{\sqrt{1+c}} , \quad c > 0$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1} \left( \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} < 1+c \right) =$$

$$- \frac{1}{2} \left\{ \text{Log}(1-T) - 2\alpha \text{Log} \frac{\pi_0}{\alpha} - 2(1-\alpha) \text{Log} \frac{1-\pi_0}{1-\alpha} - \alpha(1-\alpha) \frac{\frac{d^2}{\sigma^2} T}{1-T} \right.$$

$$\left. + \text{Log}(1 + (c+1) T) \right\}$$

avec  $T = 2\sigma^2 \tau = \frac{(3c+2) + \alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1) + \sqrt{(c+2)^2 + 6\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1) (c+2) + (\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1))^2}}{4(c+1)}$

et  $\alpha(T)$  solution de :  $\text{Log} \frac{\pi_0}{\alpha} \cdot \frac{1-\alpha}{1-\pi_0} + (1-2\alpha) \frac{d^2}{2\sigma^2} \left[ \frac{T}{1-T} \right] = 0 .$

Remarque 1 : Comme dans le cas où  $\sigma$  est connu, le résultat sera asymptotiquement meilleur si on se rapproche de la proposition  $\pi_0 = \frac{1}{2}$  .

Remarque 2 : Pour un seuil donné  $(1+c)$  , on peut améliorer le niveau asymptotique du test en prenant un nombre différent d'observation au numérateur et au déno-

minateur . En faisant comme ci-dessus  $2n$  observations au total, nous pouvons en faire  $2n\cdot\theta$  pour la statistique du test et  $2n(1-\theta)$  pour estimer la variance  $\sigma^2$ ,  $\theta \in ]0,1[$ . On obtient alors :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0} \left( \frac{\frac{1}{\theta} \sum_{i=1}^{2n\cdot\theta} (X_i - \bar{X})^2}{2n(1-\theta)} > 1+c \right) = - \inf_{x \geq (1+c)y} 2\theta \cdot \bar{H}_0(x) + 2(1-\theta) \bar{H}_0(y)$$

Le niveau le plus petit est atteint avec  $\theta(c) = \frac{1}{\text{Log}(1+c)} - \frac{1}{c}$ , ce qui donne :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0} \left( \frac{\frac{1-\theta(c)}{\theta(c)} \cdot \frac{\sum_{i=1}^{2n\theta(c)} (X_i - \bar{X})^2}{2n(1-\theta(c))}}{\sum_{i=1}^{2n(1-\theta(c))} (V_i - \bar{V})^2} > 1+c \right) = - \left\{ \frac{\text{Log } 1+c}{c} - 1 - \text{Log } \frac{\text{Log } c+1}{c} \right\}$$

Il faut d'autant mieux estimer la variance ( $\theta$  petit) qu'on se fixe un seuil grand ( $\theta(c) \xrightarrow{c \rightarrow \infty} 0$ ).

### III - COMPARAISON DES 2 METHODES DE TESTS

#### III.1. Comparaison du test basé sur les sommes partielles des carrés de résidus avec le test global introduit par échantillonnage du temps ( $\sigma$ connu).

La comparaison de ces deux tests se justifie à de nombreux points de vue :

• d'abord dans leur conception même : le premier consiste en une investigation de toutes les sommes  $\sum_{i=2}^k \omega_i^2 = \sum_{i=1}^k (y_i - \bar{y}^k)^2$ ,  $k = 2, \dots, n$  alors que le second est fondé uniquement sur la dernière somme :  $\sum_{i=1}^n (X_i - \bar{X})^2$ .

• sous  $H_0$ , les variables  $X_i$  et  $Y_i$  ont même loi et nous remarquons, que pour un seuil donné, les deux tests ont même niveau asymptotique.

• sous  $H_1$ , les variables  $X_i$  et  $Y_i$  n'ont plus même loi mais les erreurs de deuxième espèce asymptotiques ont des expression semblables :

- si  $t_0 \geq \frac{1}{2}$  ou si  $t_0 < \frac{1}{2}$  et  $\frac{4t_0(1-2t_0)\frac{d^2}{\sigma^2}}{(1+\sqrt{1+4t_0^2\frac{d^2}{\sigma^2}})^2} \leq c < t_0(1-t_0)\frac{d^2}{\sigma^2}$ , alors

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1} (\forall k, \frac{1}{(n-1)\sigma^2} \sum_{i=1}^k \omega_i^2 \leq \frac{k-1}{n-1} + c) = -h_1(c)$$

$$\text{avec } h_1(c) = \frac{1}{2} \left\{ \text{Log } \Lambda - \frac{\Lambda-1}{\Lambda} + t_0(1-t_0) \frac{d^2}{\sigma^2} \left( \frac{\Lambda-1}{\Lambda} \right)^2 \right\}$$

$$\text{et } \Lambda_{t_0} = \frac{1 + \sqrt{1 + 4t_0(1-t_0) \frac{d^2}{\sigma^2} (1+c)}}{2(1+c)}$$

- pour  $0 < c < \pi_0(1-\pi_0) \frac{d^2}{\sigma^2}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1} \left( \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)\sigma^2} < 1+c \right) = -H_1(c)$$

$$\text{avec } H_1(c) = \frac{1}{2} \left\{ (c+1)(1-\Lambda) - \frac{\alpha^2 d^2}{\sigma^2} \left( \frac{1}{\Lambda} - 1 \right) + \text{Log } \Lambda - 2 \text{Log } \frac{1-\pi_0}{1-\alpha} \right\}$$

$$\text{et } \Lambda_\alpha = \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} (1+c)}}{2(1+c)}, \quad \alpha \text{ étant toujours la solution}$$

$$\text{Log } \frac{\pi_0}{\alpha} \cdot \frac{1-\alpha}{1-\pi_0} + (1-2\alpha) \frac{d^2}{2\sigma^2} \left( \frac{1}{\Lambda} - 1 \right) = 0$$

Pour  $t_0$  (inconnu) et  $c$  fixés, parmi toutes les lois  $\Pi$  possibles, on peut regarder ce qu'on obtient avec celles qui donnent :  $\alpha(c) = t_0$

$$\text{(cela revient à prendre } \pi_{0,c} = \frac{t_0}{(1-2t_0) \left( \frac{2(c+1)}{1 + \sqrt{1 + 4t_0(1-t_0) \frac{d^2}{\sigma^2} (1+c)}} - 1 \right)} ) :$$

$$t_0 + (1-t_0)e$$

$$H_1(c)(\pi_{0,c}) - h_1(c) = \frac{1}{2} \cdot \frac{\Lambda_{t_0} - 1}{\Lambda_{t_0}} \cdot \frac{d^2}{\sigma^2} t_0(2t_0 - 1) - \text{Log } \frac{1-\pi_{0,c}}{1-t_0}$$



Or  $\Lambda_{t_0} > 1 \iff c < t_0(1-t_0)\frac{d^2}{\sigma^2}$

•  $\alpha$  était toujours compris entre  $\pi_0$  et  $\frac{1}{2}$ , donc ici  $t_0$  sera toujours compris entre  $\pi_{0,c}$  et  $\frac{1}{2}$

de sorte que pour  $t_0 > \frac{1}{2}$ ,  $\pi_{0,c} > t_0$  donc  $H_1(c)(\pi_{0,c}) > h_1(c)$

pour  $t_0 < \frac{1}{2}$ ,  $\pi_{0,c} < t_0$  donc  $H_1(c)(\pi_{0,c}) < h_1(c)$ ,

il suffit dans ce dernier cas de prendre  $\Pi$  de sorte que  $\alpha(c) = 1 - t_0$ , cela revient à prendre  $1 - \pi_{0,c}$  au lieu de  $\pi_{0,c}$  :

$$H_1(c)(1-\pi_{0,c}) - h_1(c) = \frac{1}{2} \cdot \frac{\Lambda_{t_0}^{-1}}{\Lambda_{t_0}} \cdot \frac{d^2}{\sigma^2} \cdot (1-t_0)(1-2t_0) - \text{Log} \frac{\pi_{0,c}}{t_0}$$

donc  $H_1(c)(1-\pi_{0,c}) > h_1(c)$ .

On déduit des inégalités précédentes et du fait que  $H_1(c)(\pi) = H_1(c)(1-\pi)$ , que pour toute loi  $\Pi$  telle que  $\pi_0$  soit compris entre  $\pi_{0,c}$  et  $1-\pi_{0,c}$ , le test global est meilleur que celui basé sur la somme des carrés des résidus ; c'est vrai en particulier pour la loi  $\Pi$  uniforme sur  $[0,1]$ . On améliorera encore le résultat pourvu qu'on ait assez d'information a priori sur l'instant  $t_0$  de rupture pour "dilater" ou "contracter" l'intervalle  $[0, t_0]$  pour se rapprocher de  $[0, \frac{1}{2}]$  : situation optimale. Dans les autres cas, le test basé sur la somme des carrés des résidus est meilleur.

Cette comparaison n'est valable que pour les grandes valeurs de  $c$  si  $t_0 < \frac{1}{2}$  : en effet si :

$$0 < c < \frac{4t_0(1-2t_0)}{(1+\sqrt{1+4t_0^2\frac{d^2}{\sigma^2}})^2},$$

les expressions des erreurs de 2ème espèce rendent la comparaison difficilement praticable.

III. 2. Comparaison du test basé sur les sommes partielles de résidus avec le test global introduit par échantillonnage du temps ( $\sigma$  connu)

Le test basé sur les sommes partielles de résidus a pour niveau :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0} \left( \exists t \in [0, 1] , |s_n(t)| > at + \frac{a}{2} \right) = - \frac{a^2}{\sigma^2} , \quad a > 0 .$$

Le test basé sur la variance empirique de l'échantillon a pour niveau :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0} \left( \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)\sigma^2} > 1+c \right) = - \frac{c - \text{Log}(1+c)}{2} , \quad c > 0 .$$

L'ajustement des niveaux se fera donc en prenant  $\frac{a}{\sigma} = \sqrt{\frac{c - \text{Log}(c+1)}{2}}$  .

La comparaison des expressions des erreurs de 2ème espèce asymptotiques s'avérant difficile, nous tirerons notre information de la comparaison des régions où la décroissance des erreurs de 2ème espèce est effectivement exponentielle.

Pour le test basé sur les sommes partielles de résidus, il faut que :

- si  $t_0 \leq e^{-3/2}$ , on ait :  $\frac{a}{|d|} < \mu(t_0)$  avec  $\mu(t_0)$  solution de  $\frac{\mu}{2t_0} + 1 + \text{Log } \mu = 0$
- si  $t_0 \geq e^{-3/2}$ , on ait :  $\frac{a}{|d|} < \frac{2}{3} t_0 \text{Log } \frac{1}{t_0}$

Pour le test basé sur la variance empirique de l'échantillon, il faut que :

$$c < \pi_0(1-\pi_0) \frac{d^2}{\sigma^2} .$$

Pour le premier test, les valeurs de  $a$  donnant une décroissance exponentielle vérifient :

$$\frac{a}{\sigma} \leq \frac{|d|}{\sigma} \times \left\{ \mu(t_0) \cdot 1_{t_0 \leq e^{-3/2}} + \frac{2}{3} t_0 \text{Log } \frac{1}{t_0} \cdot 1_{t_0 \geq e^{-3/2}} \right\} .$$

Pour le second test, les valeurs de  $a$  donnant une décroissance exponentielle vérifient :

- si  $\pi_0(1-\pi_0) \frac{d^2}{\sigma^2}$  est petit ,  $\frac{a}{\sigma} \leq \frac{1}{2} \pi_0(1-\pi_0) \frac{d^2}{\sigma^2}$

• si  $\pi_0(1-\pi_0) \frac{d^2}{\sigma^2}$  est grand ,  $\frac{a}{\sigma} \lesssim \sqrt{\frac{\pi_0(1-\pi_0)}{2} \frac{|d|}{\sigma}}$  .

Afin de faciliter la comparaison, prenons une loi  $\Pi$  telle que  $\pi_0 = t_0$  par exemple la loi uniforme sur  $[0,1]$  . Pour les petites valeurs de  $\frac{|d|}{\sigma}$  , on constate que la région de décroissance exponentielle de l'erreur de 2ème espèce est moins étendue pour le test basé sur la variance asymptotique que pour le test basé sur les sommes de résidus : ce dernier serait donc plus puissant. Pour les grandes valeurs de  $\frac{|d|}{\sigma}$  , on obtient :

$$\begin{aligned} \bullet \mu(t_0) < \frac{t_0(1-t_0)}{2} &\iff \frac{\sqrt{\frac{t_0(1-t_0)}{2}}}{2t_0} + 1 + \text{Log} \sqrt{\frac{t_0(1-t_0)}{2}} > 0 \\ &\iff \sqrt{\frac{1-t_0}{2t_0}} + 2 + \text{Log} \frac{t_0(1-t_0)}{2} > 0 . \end{aligned}$$

Cette dernière inégalité est réalisée pour  $t_0$  inférieur à  $e^{-3/2}$  car la fonction qui intervient est décroissante sur  $[0, \frac{1}{2}]$  et est positive en  $\frac{1}{2}$  .

$$\bullet \frac{2}{3} t_0 \text{Log} \frac{1}{t_0} < \frac{t_0(1-t_0)}{2} \iff \sqrt{\frac{1-t_0}{t_0}} + \frac{2\sqrt{2}}{3} \text{Log} t_0 > 0$$

Cette dernière inégalité est réalisée pour  $t_0$  supérieur à  $e^{-3/2}$  car la fonction qui intervient est décroissante sur  $[0,1]$  et est nulle en 1 .

En résumé pour les grande valeurs de  $\frac{|d|}{\sigma}$  , la région de décroissance exponentielle est plus étendue pour le test basé sur la variance empirique : celui-ci serait donc plus puissant que le test basé sur les sommes de résidus.

APPENDICE

On va montrer qu'asymptotiquement, pour le problème qui nous concerne il est équivalent de faire un échantillonnage de type  $\Pi^{\otimes n}$  ou de faire un tirage sans remise de  $n$  instants parmi  $N(n)$  possibles. La solution de ce problème classique en échantillonnage repose sur la convergence de la loi hypergéométrique vers la loi binomiale.

Notons  $U_1, U_2, \dots, U_N$  les  $N(n)$  instants possibles sur  $[0,1]$  et  $Y_{U_1}, \dots, Y_{U_N}$  les observations associées, les variables  $Y_{U_i}$  étant indépendantes et de loi  $\mathcal{N}(m_1, \sigma^2)$  si  $U_i < t_0$  et  $\mathcal{N}(m_2, \sigma^2)$  si  $U_i \geq t_0$ . On fait un tirage de  $n$  instants parmi les  $N(n)$  de façon uniforme et sans remise : on obtient les observations  $(X_1, \dots, X_n) = (Y_{U_1}, \dots, Y_{U_n})$ . Il nous suffit alors de montrer que  $\frac{1}{n} \text{Log } \phi_n(t)$  converge vers la même fonction limite qu'avec l'échantillonnage de type  $\Pi^{\otimes n}$  :

$$\begin{aligned} \phi_n(t) &= E e^{t \sum_{i=1}^n (X_i - \bar{X})^2} = \sum_{k=0}^n \frac{C_{\pi_0 N}^k C_{(1-\pi_0)N}^{n-k}}{C_N^n} \cdot (1-2\sigma^2 t)^{\frac{n-1}{2}} e^{\frac{nt d^2}{1-2\sigma^2 t} \cdot \frac{k}{n}(1-\frac{k}{n})} \\ &= \sum_{k=0}^n A_k \end{aligned}$$

Pour  $n$  grand,  $\frac{1}{n} \text{Log } \phi_n(t) \sim \frac{1}{n} \text{Log}(\text{Max}_k A_k) \sim \frac{1}{n} \text{Log } A_{\alpha n}$  avec :

$$\alpha = \alpha(t) \text{ solution de : } \text{Log } \frac{\pi_0}{\alpha} \cdot \frac{1-\alpha}{1-\pi_0} + (1-2\alpha) \frac{t d^2}{1-2\sigma^2 t} = 0$$

ce qui donne, en utilisant la formule de Stirling :

$$\frac{1}{n} \text{Log } \phi_n(t) \longrightarrow -\frac{1}{2} \text{Log}(1-2\sigma^2 t) + \alpha \text{Log } \frac{\pi_0}{\alpha} + (1-\alpha) \text{Log } \frac{1-\pi_0}{1-\alpha} + \alpha(1-\alpha) \frac{t d^2}{1-2\sigma^2 t}$$

Bibliographie :

- [ 1 ]        BROWN-DURBIN-EVANS : "Techniques for testing the consistency of regression relationships over time" (1975). JRSS (B), p. 149-199.
  
- [ 2 ]        R. AZENCOTT : "Grandes déviations : théorèmes à la Cramer-Chernoff et petites perturbations de systèmes dynamiques", Cours de l'Ecole d'Eté de Saint-Flour, 1978 (à paraître)
  
- [ 3 ]        D. DACUNHA-CASTELLE : Exposé n° I au Séminaire de Statistique d'Orsay 1977-1978
  
- [ 4 ]        J. BRETAGNOLLE : Exposé n° III au Séminaire de Statistique d'Orsay de 1977-1978
  
- [ 5 ]        P. ERDOS- A.RENYI : "On a new law of large numbers", Journal d'analyse Mathématique n° 22 (1970)
  
- [ 6 ]        J. DESHAYES-D. PICARD : Exposé n° IV au Séminaire de Statistique d'Orsay de 1977-1978.

Jean DESHAYES  
Dominique PICARD  
Mathématique - Bât. 425  
ERA CNRS 532 "Statistique Appliquée"  
Université Paris-Sud  
91405 ORSAY

Formule de Chernoff pour des chaînes de Markov

(d'après Donsker et Varadhan)

Marie DUFLO.

A - PRELIMINAIRES

Notations

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

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

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

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

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

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

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

## I - INFORMATION DE KULLBACK

### 1 - Définition

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

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

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

### Théorème

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

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

### Démonstration

a) Supposons  $\alpha \ll \beta$  et  $a = \text{Log} \frac{d\alpha}{d\beta}$  intégrable par rapport à  $\alpha$  .

Soit  $I = \int \text{Log } a \, d\alpha$  . Soit  $u \in \hat{\mathcal{U}}$  .

$$\int \text{Log } u \, d\alpha = \int \text{Log} \left( \frac{u}{a} \right) \, d\alpha + \int \text{Log } a \, d\alpha$$

$$\leq \text{Log} \int \frac{u}{a} \, d\alpha + I = \text{Log} \int u \, d\beta + I \quad .$$

Donc  $I_\beta(\alpha) \leq I$  .

b) Supposons  $I_\beta(\alpha)$  finie. Pour tout  $b > 0$  , et tout  $A \in \mathcal{E}$  la fonction  $(b+1)I_A + I_{A^c}$  est dans  $\hat{u}$ . Supposons  $\beta(A)$  nul.

$$\alpha(A) \text{Log}(b+1) \leq I_\beta(\alpha) + \text{Log} [1 + b \beta(A)] = I_\beta(\alpha) \quad .$$

Ceci étant vrai pour tout  $b > 0$  ,  $\alpha(A) = 0$ . Donc  $\alpha \ll \beta$  . Soit  $a = \frac{d\alpha}{d\beta}$  .

Si  $\alpha$  est dans  $\hat{u}$ :

$$\int \text{Log } a \, d\alpha - \text{Log} \int a \, d\beta = \int \text{Log } a \, d\alpha = I \leq I_\beta(\alpha) \quad .$$

Sinon on obtient le théorème en approchant  $a$  par  $(a \vee \frac{1}{n}) \wedge n = a_n$  , lorsque  $n$  tend vers l'infini :

$$\int \text{Log } a_n \, d\alpha \leq I_\beta(\alpha) + \text{Log} \int a_n \, d\beta$$

$$\overline{\lim}_{n \rightarrow \infty} \int \text{Log } a_n \, d\alpha \leq I_\beta(\alpha) + \text{Log} \overline{\lim}_{n \rightarrow \infty} \int a_n \, d\beta \leq I_\beta(\alpha) \quad .$$

La fonction  $x \longmapsto x(\text{Log } x)^-$  est bornée, donc si  $n$  tend vers  $\infty$

$$\int (\text{Log } a_n)^- \, a \, d\beta = \int (\text{Log} (a \vee \frac{1}{n}))^- \, a \, d\beta$$

croît vers  $\int (\text{Log } a)^- \, a \, d\alpha$  qui est finie .

Enfin  $(\text{Log } a_n)^+ = (\text{Log} (a \wedge n))^+$  croît vers  $(\text{Log } a)^+$  , si  $n$  tend vers  $\infty$  .

Donc :

$$\int (\text{Log } a)^+ \, d\alpha \leq I_\beta(\alpha) + \int (\text{Log } a)^- \, d\alpha < \infty \quad .$$



Ce qui montre que  $\text{Log } a$  est  $\alpha$  intégrable et que  $I \leq I_\beta(\alpha)$  ■

2 - Proposition

Si  $E$  est polonais ,

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

Démonstration

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

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

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

Si  $(G_n)_{n=1 \dots N}$  est un recouvrement ouvert de  $E$ , à tout  $p \geq 1$ , on peut associer  $f_n^p = p [d(x, G_n^c) \wedge \frac{1}{p}]$ ,  $n = 1 \dots N$ . Soit  $(a_n)_{n=1 \dots N}$  une suite de nombres

strictement positifs. La suite  $\sum_{n=1}^N a_n f_n^p$  croît si  $p$  tend vers  $\infty$  vers

$$\sum_{n=1}^N a_n \mathbb{1}_{G_n} :$$

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

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

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

$$\begin{aligned} & \int \text{Log} \left( \sum_{n=1}^N a_n \mathbb{1}_{A_n} \right) d\alpha - \text{Log} \int \sum_{n=1}^N a_n \beta(A_n) \\ & \leq \int \text{Log} \left( \sum_{n=1}^N a_n \mathbb{1}_{G_n^p} \right) d\alpha - \text{Log} \int \sum_{n=1}^N a_n \beta(G_n^p) + \text{Log} \frac{\sum_{n=1}^N a_n \beta(G_n^p)}{\sum_{n=1}^N a_n \beta(A_n)} \\ & \leq I + \text{Log} \left\{ 1 + \frac{1}{p \sum_{n=1}^N a_n \beta(A_n)} \right\} \end{aligned}$$

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

### Corollaire

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

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

$$\alpha \mapsto \int \text{Log } u \, d\alpha - \text{Log} \int u \, d\beta \quad \text{et} \quad \beta \mapsto \int \text{Log } u \, d\alpha - \text{Log} \int u \, d\beta$$

sont continues

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

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

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

### Démonstration

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

$$\alpha(A) \text{Log}(b+1) - \text{Log}(1 + b \beta(A)) \leq I_\beta(\alpha) = I$$

$$\alpha(A) \text{Log}(b+1) - b \beta(A) \leq I$$

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

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

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

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

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

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

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

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

et

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

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

B - MAJORATION DE CHERNOFF  
POUR DES CHAINES DE MARKOV CONTROLÉES

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

On considère une chaîne de Markov contrôlée. Les espaces des états et des actions sont des espaces polonais munis de leurs tribus boréliennes  $(E, \mathcal{E})$  et  $(A, \mathcal{A})$  respectivement. On donne une probabilité de transition  $\pi$  de  $(E \times A, \mathcal{E} \otimes \mathcal{A})$  dans  $(E, \mathcal{E})$ . L'espace des trajectoires est  $(\Omega, \mathcal{Q}) = (E \times A, \mathcal{E} \otimes \mathcal{A})^{\mathbb{N}}$  : si  $\omega = (x_n, a_n)_{n \in \mathbb{N}}$ , on note  $X_n(\omega) = x_n$  et  $A_n(\omega) = a_n$ . Soit  $\mathcal{B}_n = \sigma(X_0, \dots, X_n, A_0, \dots, A_{n-1})$ . Une stratégie  $\delta$  est une suite  $(\delta_n)_{n \in \mathbb{N}}$ , où  $\delta_n$  est une probabilité de transition de  $(E, \mathcal{E})^n$  dans  $(A, \mathcal{A})$  :  $\delta_n(x_0, \dots, x_{n-1}; \cdot)$  est la loi de probabilité de la  $n^{\text{ième}}$  action choisie, si les observations antérieures sont  $(x_0, \dots, x_{n-1})$ . On peut alors pour tout état initial  $x$  et toute stratégie  $\delta$  construire une probabilité  $P_x^\delta$  sur  $(\Omega, \mathcal{Q})$  telle que :

$$P_x^\delta(X_0 = x) = 1$$

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

$$P_x^\delta[X_{n+1} \in B, A_n \in C | \mathcal{B}_n] = \int \pi(X_n, a, dy) 1_B(y) 1_C(a) \delta_n(X_0, \dots, X_{n-1}; da) .$$

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

$$\begin{aligned} \pi f(x, a) &= \int \pi(x, a, dy) f(y) \quad \text{pour } f \in \mathcal{E}_1 \\ &= \int \pi(x, a, dy) f(x, a, y) \quad \text{pour } f \in \mathcal{E} . \end{aligned}$$

On suppose  $\pi$  fellérienne, c'est-à-dire que  $\pi \mathcal{E}$  est continue dans  $\mathcal{E}$ .

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

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

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

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

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

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

$$\mathcal{L}^n(\omega, B) = \frac{1}{n} \sum_{k=0}^{n-1} 1_B(X_k, A_k, X_{k+1})(\omega).$$

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

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

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

---

$$\begin{aligned} \text{a) } I(\lambda) &= \sup_{u \in \mathcal{U}} \left[ \int \text{Log } u \, d\lambda - \text{Log} \int u \, d\lambda_1 \otimes s \otimes \pi \right] \\ &= \sup_{u \in \mathcal{U}} \left[ \int \text{Log} \frac{u(x, a, y)}{\pi u(x, a)} \, d\lambda(x, a, y) \right] \end{aligned}$$

Démonstration

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned} E_{\star}^{\delta} \left[ \exp \left( \sum_{K=0}^{n-1} v(X_K, A_K, X_{K+1}) \right) \right] &= E_{\star}^{\delta} \left[ \exp \left( n \int v \, dL^n \right) \right] \\ &= E_{\star}^{\delta} \left\{ \prod_{K=0}^{n-2} \frac{u(X_K, A_K, X_{K+1})}{\pi u(X_K, A_K)} \frac{E_X^{\delta} [u(X_{n-1}, A_{n-1}, X_n) | \mathcal{G}_{n-1}]}{\pi u(X_{n-1}, A_{n-1})} \right\} \\ &= E_{\star}^{\delta} \left[ \prod_{K=0}^{n-2} \frac{u(X_K, A_K, X_{K+1})}{\pi u(X_K, A_K)} \right] = \dots = 1. \end{aligned}$$

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

$$P_{\star}^{\delta} [L^n \in \Gamma] \leq \exp \left[ -n \sup_{v \in \mathcal{V}} \inf_{\lambda \in \Gamma} \int v \, d\lambda \right]$$

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

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

$$P_{x, \delta}^{\delta}(L^n \in K_j) \leq e^{-ni}$$

$$P_{x, \delta}^{\delta}(L^n \in K) \leq N \sup_j P_{x, \delta}^{\delta}(L^n \in K_j) \leq N e^{-ni}$$

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \sup_{x, \delta} P_{x, \delta}^{\delta}(L^n \in K) \leq -I(K)$$

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

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

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

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

Soit  $K$  un compact de  $\mathcal{P}$ . Si  $K \cap \Delta = \bigcap_n (K \cap \Delta_n)$  est vide, il existe un  $n_0$  tel que  $K \cap \Delta_n$  soit vide, donc  $P_x(L^n \in K) = 0$ , dès que  $n$  est supérieur à  $n_0$ . Sinon  $I(K \cap \Delta) = \lim_n \uparrow I(K \cap \Delta_n)$ . En effet pour chaque  $n$ , il existe  $\lambda_n \in K \cap \Delta_n$  telle que  $I(\lambda_n) = I(K \cap \Delta_n)$ . Soit  $\lambda \in K \cap \Delta$  une valeur d'adhérence de  $(\lambda_n)$ . La fonction  $I$  étant s.c.i.

$$I(K \cap \Delta) \leq I(\lambda) \leq \lim_{n \rightarrow \infty} I(\lambda_n) = \lim_{n \rightarrow \infty} \uparrow I(K \cap \Delta_n)$$

et  $I(K \cap \Delta) \geq \lim_{n \rightarrow \infty} \uparrow I(K \cap \Delta_n)$  est évident.

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

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \text{Log} \sup_{x, \delta} P_x^\delta(L^n \in K) = \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \text{Log} \sup_{x, \delta} P_x^\delta(L^n \in K \cap \Delta_m) \leq -I(K \cap \Delta_m)$$

Théorème

Si  $E$  et  $A$  sont polonais et  $\pi$  fellerienne, pour tout compact

$K$  de  $\mathcal{P}$  :

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \text{Log} \sup_{x, \delta} P_x^\delta(L^n \in K) \leq -I(K \cap \Delta)$$



C - MAJORATIONS DE CHERNOFF POUR  
DES CHAINES DE MARKOV NON CONTROLEES

On supprime désormais actions, et stratégies (A est réduit à un point).  
Les notations restent les mêmes, mais  $\mathcal{E}$ ,  $\mathcal{U}$  et  $\mathcal{P}$  sont définis sur  $E^2 \otimes \mathcal{Z}^{\otimes 2}$ ; à  $\lambda \in \mathcal{P}$ , on associera ses deux marginales  $\lambda_1$  et  $\lambda_2$ ;  
 $(\Omega, \mathcal{a}) = (E, \mathcal{E})^{\mathbb{N}}$ ;  $\{(\Omega, \mathcal{a}, (P_x)_{x \in E}, (X_n)_{n \in \mathbb{N}})\}$  est une chaîne de Markov à valeurs dans  $(E, \mathcal{E})$ , de transition  $\pi$ .

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

Projection de l'information sur E

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

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

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

Théorème

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

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

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

Démonstration

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

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

b) Il suffit de montrer le résultat lorsque  $E$  est compact.

En effet on obtiendra alors le résultat sur le compactifié  $\hat{E}$  et

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

c) Supposons  $E$  métrique compact.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned} \lambda_{\alpha \otimes \pi}^w(a) &= - \inf_u \text{Log} \int e^{u(w-a)} d\alpha \otimes \pi \\ &= \inf_{\{\lambda ; \int w d\lambda = a\}} I_{\alpha \otimes \pi}(\lambda) \end{aligned}$$

Donc  $\lambda_{\mathcal{W}}^{\alpha \otimes \pi}(0) = - \inf_u \text{Log} \int e^{-u[f(x)-g(y)]} \alpha(dx) \pi(x, dy) \geq J$ .

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

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

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

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

$$\int \text{Log} \frac{\pi W}{V} d\alpha = \int \text{Log} \pi W d\alpha - \int \text{Log} V d\alpha \leq \text{Log} \int \frac{\pi W}{V} d\alpha \leq -J + \eta$$

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

$$- I_1(\alpha) \leq \int \text{Log} \pi W d\alpha - \int \text{Log} W d\alpha \leq -J + \eta$$

et  $I \leq I_1(\alpha)$ .

### Projection de la majoration sur E

#### Théorème

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

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \text{Log} \sup_x P_x [L_1^n \in K] \leq - I_1(K)$$

Démonstration :

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

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

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

La démonstration s'achève comme sur  $E^2$ .

Remarque

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

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

Application au cas où  $E$  est un espace métrique compact

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

$$\overline{\lim} \frac{1}{n} \text{Log} \sup_x P_x \left[ \frac{1}{n} \sum_{k=0}^{n-1} f(X_k, X_{k+1}) \in F \right] \leq -H_f(F)$$

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

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

$$\overline{\lim}_x \frac{1}{n} \text{Log} \sup_x P_x \left[ \frac{1}{n} \sum_{k=0}^{n-1} f(X_k, X_{k+1}) \leq a \right] \leq -H_f((-\infty, a])$$

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

$$\overline{\lim} \frac{1}{n} \text{Log} \sup_x P_x \left[ \frac{1}{n} \sum_{k=0}^{n-1} g(X_k) \in F \right] \leq -h_g(F)$$

où  $h_g(F) = \inf_{\{\alpha; \int f d\alpha \in F\}} I_1(\alpha)$ .

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

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

Conditions pour que les bornes supérieures obtenues soient strictement négatives

Proposition

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

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

Démonstration :

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

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

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

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

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

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

Conséquence

Soit  $K$  un compact de  $\mathcal{P}_1$ ; si  $K$  ne contient aucune probabilité invariante,  $I_1(K)$  est strictement positif et pour tout  $\epsilon > 0$ , il existe  $n_0$  tel que pour  $n \geq n_0$  et pour tout  $x \in E$

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

Cas des chaînes de markov irréductibles

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

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

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

La chaîne est dite irréductible si il existe une mesure  $\phi$  telle que  $\phi \ll \pi^z(x, \cdot)$  pour tout  $x \in E$ . Pour une telle chaîne ou bien il n'existe pas de probabilité invariante (cas transient ou récurrent nul) ou bien il existe une pro-

babilité invariante  $\mu$  unique telle que pour toute v.a.  $f$  positive sur  $E^2$  et tout  $x \in E$  on ait :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k, X_{k+1}) = \int f d\mu \otimes \pi \quad P_x \text{ p.s.}$$

Si  $E$  est compact, le seul cas possible est le cas récurrent positif [8]. Donc pour  $f$  s.c.i. de  $E^2$  dans  $\mathbb{R}$  et pour  $a < \int f d\mu \otimes \pi$ , il existe un  $b > 0$  et un  $n_0$ , tels que  $n \geq n_0$  implique :

$$\sup_x P_x \left[ \frac{1}{n} \sum_{k=0}^{n-1} f(X_k, X_{k+1}) \leq a \right] \leq e^{-nb}.$$

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

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

D - MINORATIONS DE CHERNOFF POUR  
UNE CHAÎNE DE MARKOV IRREDUCTIBLE

I - TRANSFORMATION DE LA CHAÎNE DE MARKOV

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

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

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

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

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

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

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

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

$$\frac{d\lambda}{d\alpha \otimes \pi}(x, y) .$$



Suivant [10], considérons :

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

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

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

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

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

Posons :

$$\begin{cases} \bar{\pi}(x, \cdot) = \frac{a(x)}{b(y)} \pi(x, \cdot) \\ h(x,y) = \frac{a(x)}{b(y)} \end{cases}$$

Alors  $\bar{\pi}$  est une probabilité de transition et  $\alpha$  est invariante pour  $\bar{\pi}$ .

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

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

En effet  $\alpha = \alpha \bar{\pi} \ll \beta$ . Soit  $p = \frac{d\alpha}{d\beta}$ . Pour  $\beta$  presque tout  $y$

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

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

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

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

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

2 - MINORATION DE CHERNOFF LORSQUE LES PROBABILITES  $(\pi(x, \cdot))_{x \in E}$  SONT TOUTES EQUIVALENTES

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

$$\frac{d \bar{\pi}(x, \cdot)}{d \pi(x, \cdot)} = h(x, \cdot) .$$

$$I_1(\alpha) = I_{\alpha \otimes \pi}(\alpha \otimes \bar{\pi}) = \int \text{Log } h \, d \alpha \otimes \pi .$$

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

$$\begin{aligned} \pi(x, A) &= \int 1_A(y) \frac{1}{h(x, y)} \bar{\pi}(x, dy) + \bar{\pi}(x, A \cap N) \\ &\geq \int 1_A(y) \frac{1}{h(x, y)} \bar{\pi}(x, dy) \end{aligned}$$

Considérons sur l'espace canonique  $(\Omega, \mathcal{A})$ , la famille de probabilités

$(\bar{P}_x)_{x \in E}$  telles que :

$$\begin{cases} \bar{P}_x(X_0 = x) = 1 \\ \bar{P}_x[X_{n+1} \in A \mid \mathcal{B}_n] = \bar{\pi}(X_n, A) \end{cases}$$

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

$$B = \{(X_1 \dots X_n) \in \Gamma\} .$$

$$\begin{aligned}
 P_x(B) &= \int 1_{\Gamma}(x_1, \dots, x_n) \pi(x, dx_1) \pi(x_1, dx_2) \dots \pi(x_{n-1}, dx_n) \\
 &\geq \int 1_{\Gamma}(x_1, \dots, x_n) \frac{1}{h(x, x_1) \dots h(x_{n-1}, x_n)} \bar{\pi}(x, dx_1) \bar{\pi}(x_1, dx_2) \dots \bar{\pi}(x_{n-1}, dx_n) \\
 &\geq \bar{E}_x \left[ 1_B \frac{1}{h(X_0, X_1) h(X_1, X_2) \dots h(X_{n-1}, X_n)} \right]
 \end{aligned}$$

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

La chaîne  $(\Omega, \alpha, \bar{P}_x, (X_n)_{n \in \mathbb{N}})$  étant récurrente :

a)  $\bar{P}_x$  p.s.  $(L_2^n)$  converge en loi vers  $\alpha$  donc  $\lim_{n \rightarrow \infty} \bar{P}_x [L_2^n \in V] = 1$ .

b)  $\bar{P}_x$  p.s.  $\frac{1}{n} \sum_{k=0}^{n-1} \text{Log } h(X_k, X_{k+1})$  tend vers  $I_1(\alpha) = \int \text{Log } h \, d\alpha \otimes \pi$

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

$$P_x(B) \geq e^{-n(I-\varepsilon)} \{1 - \bar{P}_x(L_2^n \in V^c) - \bar{P}_x \left[ \sum_{k=0}^{n-1} \text{Log } h(X_k, X_{k+1}) \leq I_1(\alpha) - \varepsilon \right]\}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_x(L_2^n \in V) \geq -I_1(\alpha)$$

Supposons  $V$  ouvert, soit  $W$  un voisinage de  $\alpha$  fermé inclus dans  $V$ . Il existe un  $n$  tel que  $\{\beta; \sup_{\gamma \in W} \|\beta - \gamma\| \leq \frac{2}{n}\}$  soit inclus dans  $V$ .

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_x(L_1^n \in V) \geq -I_1(\alpha)$$

Théorème

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

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_x(L_1^n \in G) \geq -I_1(G)$$

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

3 - EXTENSION

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

$$P'(\tau_n = i) = (1-z) z^{i-1} \quad (i \geq 1)$$

Soit 
$$\begin{cases} T_0 = 0 \\ T_n = \tau_1 + \dots + \tau_n \quad \text{si } n \geq 1. \end{cases}$$

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

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

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

Soit  $\alpha \in \mathcal{G}$ , telle que  $\alpha \ll \beta$ . Considérons  $V$  un voisinage ouvert de  $\alpha$  et  $W$  un autre voisinage de  $\alpha$ , tel que pour  $\varepsilon > 0$ ,

$$\{v; \sup_{\gamma \in W} ||\gamma - v|| \leq 2\varepsilon\} \subset V.$$

Soit  $A \in \mathcal{G}$ . Posons :

$$\bar{L}_1^n(\omega, \omega', A) = \frac{1}{n} \sum_{k=0}^{n-1} I_A[Y_k](\omega, \omega')$$

Supposons :  $\frac{T_{n-1}^{-(n-1)}}{n} \leq \varepsilon$ . Pour tout  $A$

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

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

$$L_1^{T_{n-1}}(A) \leq \bar{L}_1^n(A) + \varepsilon.$$

Donc  $||L_1^n - \bar{L}_1^n|| \leq 2\varepsilon$

$$\begin{aligned} P_x(L_1^n \in V) &\geq P_x(\bar{L}_1^n \in W) - P' \left[ \frac{T_{n-1}^{-(n-1)}}{n} \geq \frac{\varepsilon}{2} \right] \\ &\geq -I_1^z(\alpha) - P' \left( \frac{T_{n-1}}{n} \geq 1 + \frac{\varepsilon}{2} \right) \end{aligned}$$

où  $I_1^z$  est l'information calculée à partir de  $\pi^z$ .

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

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

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

$$\frac{1}{n} \text{Log } P' (T_n \geq n(1+\varepsilon)) \leq -\varepsilon\sigma + \text{Log}(1-z) - \text{Log}(1-z e^\sigma)$$

$$\lim_{z \rightarrow 0} \frac{1}{n} \text{Log } P' (T_n \geq n(1+\varepsilon)) = 0$$

et cette limite est atteinte uniformément en  $n$ .

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

$$\begin{aligned} I_1^z(\varepsilon) &= \sup_{u \in \mathcal{U}} \int \text{Log} \frac{u}{\pi^z u} d\alpha \leq \sup_{u \in \mathcal{U}} \int \text{Log} \frac{u}{(1-z)\pi u} d\alpha \\ &\leq \frac{1}{1-z} I_1(\alpha) \end{aligned}$$

$$\overline{\lim}_{z \rightarrow 0} I_1^z(\alpha) \leq I_1(\alpha)$$

$$\begin{aligned} I_1(\alpha) &= \sup_{u \in \mathcal{U}_1} \int \text{Log} \frac{u}{\pi u} d\alpha \leq \sup_{u \in \mathcal{U}_1} \lim_{z \rightarrow 0} \int \text{Log} \frac{u}{\pi^z u} du \\ &\leq \lim_{z \rightarrow 0} \sup_{u \in \mathcal{U}_1} \int \text{Log} \frac{u}{\pi^z u} du \leq \lim_{z \rightarrow 0} I_1^z(\alpha) . \end{aligned}$$

Donc  $\lim_{n \rightarrow \infty} P_x(L_1^n \in V) \geq -I_1(\alpha)$ .

### Théorème

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

$$\lim_{n \rightarrow \infty} P_x(L_1^n \in G) \geq -I_1(G) .$$

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

### Corollaire 1

Sous les mêmes hypothèses que le théorème, soit  $f$  continue de  $E$  dans  $\mathbb{R}$  et  $0$  ouvert de  $\mathbb{R}$ . Pour tout  $x$  :

$$\underline{\lim} \frac{1}{n} \text{Log } P_x \left[ \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \in 0 \right] \geq - \inf_{\{\alpha; \int f d\alpha \in 0\}} I(\alpha)$$

Si  $f$  est s.c.i. et  $0$  une demi droite  $]-\infty, a[$  le théorème reste vrai.

Corollaire 2

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

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } E_x \left[ \exp \left( \sum_{i=0}^{n-1} f(X_i) \right) \right] = \sup_{\alpha \in \mathcal{P}_1} \left( \int f d\alpha - I_1(\alpha) \right) .$$

Cela résulte des majorations et minorations obtenues dans ce cas par une démonstration de Varadhan [9]. Inversement Gärtner établit dans [3] des théorèmes de grandes déviations en supposant que pour toute v.a. bornée la limite de la suite  $\frac{1}{n} \text{Lo } E_x \left[ \exp \left( \sum_{i=0}^{n-1} f(X_i) \right) \right]$  existe et est finie.

MAJORATIONS DE CHERNOFF ET STATISTIQUE SEQUENTIELLE POUR DES CHAINES DE MARKOV

---

RECURRENTES AU SENS DE DOEBLIN.

---

Nelly MAIGRET.

Résumé :

On établit un théorème de grandes déviations pour une chaîne de Markov récurrente Doeblin, semblable aux théorèmes de DONSKER et VARADHAN [5]. On construit (toujours pour une chaîne récurrente Doeblin) un test analogue à celui de Chernoff dans le cas indépendant [4], [1], [2], le théorème de grandes déviations obtenu permet de montrer que ce test est asymptotiquement le plus économique parmi les tests de force au moins égale.

PREMIERE PARTIE : MAJORATIONS DE CHERNOFF.

$X = (\Omega, \mathcal{A}, (P_x)_{x \in \mathcal{H}}, (X_n)_{n \in \mathbb{N}})$  une chaîne de Markov sur un espace d'état mesurable  $(\mathcal{H}, \underline{\mathcal{H}})$  où la tribu  $\underline{\mathcal{H}}$  est supposée à base dénombrable. On note  $\pi$  le noyau de transition, et  $\pi_n$  la  $n^{\text{ième}}$  itérée de  $\pi_0$ .  $\mathcal{B}_n$  désigne la tribu des événements antérieurs à  $n$  :  $\mathcal{B}_n = \sigma(X_0, \dots, X_n)$ . On note  $\mathcal{P}$  l'ensemble des probabilités sur  $(\mathcal{H}, \underline{\mathcal{H}})$ . Lorsque  $(\mathcal{H}, \underline{\mathcal{H}})$  est métrique,  $\mathcal{P}$  est muni de la topologie de la convergence étroite. On note  $\mathbb{P}_\lambda$  la distribution de la chaîne et  $E_\lambda$  l'espérance lorsque  $\lambda$  de  $\mathcal{P}$  est la loi initiale. Pour simplifier,  $P_x$  et  $E_x$  désigneront respectivement  $P_{\delta_x}$  et  $E_{\delta_x}$ ,  $\delta_x$  étant la mesure de Dirac au point  $x$ . On notera de la même façon, par  $\| \cdot \|$ , la norme uniforme sur  $\mathcal{P}$  et sur l'ensemble des variables aléatoires (v.a.) sur  $\mathcal{H} \times \underline{\mathcal{H}}$ .



La chaîne de Markov  $X$  est supposée récurrente Doeblin, apériodique, de probabilité invariante  $\mu$ , c'est-à-dire : il existe  $c > 0$ , il existe  $\rho$ ,  $0 < \rho < 1$ , tels que pour tout  $n$ ,  $\sup_{x \in \mathcal{X}} \|\pi_n(x, \cdot) - \mu(\cdot)\| \leq c \rho^n$  (cf. [7]).

Sauf précision contraire, les espaces topologiques seront munis de leur tribu borélienne, et les espaces produit de la tribu produit associée.

Enfin, pour tout  $n$  de  $\mathbb{N}$ , on notera  $\mu \otimes \pi_n$  la mesure sur  $\mathcal{X}^{n+1}$  définie par

$$(\mu \otimes \pi_n)(A) = \int I_A(x_0, \dots, x_n) d\mu(x_0) \pi(x_0, dx_1) \dots \pi(x_{n-1}, dx_n)$$

Théorème 1 :

Soit  $g$  une v.a. bornée sur  $\mathcal{X} \times \mathcal{X}$ ,  $m$ ,  $\bar{g}$  et  $\underline{g}$  définis respectivement par  $\int g d\mu \otimes \pi$ ,  $\mu \otimes \pi$ -ess sup  $g$ ,  $\mu \otimes \pi$ -ess inf  $g$ . On note  $h$  la transformée de Cramer de  $g(\mu \otimes \pi)$  définie sur  $[\underline{g}, \bar{g}]$ . Pour tout réel  $a$  strictement supérieur à  $m$ , on a :

$$\overline{\lim}_n \frac{1}{n} \log \sup_{x \in \mathcal{X}} P_x \left( \sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq 0 \right) < 0$$

En particulier si  $a$  est élément de  $]m, \bar{g}[$ , le terme ci-dessus est majoré par le nombre négatif  $\inf_k \left\{ \frac{1}{k} \log (b\rho^k + e^{-h(a)}) \right\}$  où  $b$  est une constante dépendant de  $c$ , de  $g$  et de  $a$ .

Démonstration :

. Supposons que  $a$  est élément de  $]m, \bar{g}[$ .

Alors  $u_0 = h'(a)$  est strictement positif et vérifie

$$e^{-h(a)} = E_{\mu} \left( e^{(g(X_0, X_1) - a)u_0} \right)$$

Pour tout  $x$  de  $\mathcal{X}$ , pour tout  $k$  entier naturel, on a :

$$\begin{aligned}
 |E_x (e^{u_0 (g(X_k, X_{k+1}) - a)} - e^{-h(a)})| &= \int e^{u_0 (g(y, z) - a)} |\Pi(y, dz) (\Pi_k(x, dy) - \mu(dy))| \\
 &\leq \|e^{u_0 (g-a)}\| \|\Pi_k(x, \cdot) - \mu(\cdot)\| \\
 &\leq \|e^{u_0 (g-a)}\| c \rho^k = b \rho^k
 \end{aligned}$$

c'est-à-dire, pour tout  $k$  entier naturel

$$\sup_x E_x (e^{u_0 (g(X_k, X_{k+1}) - a)}) \leq e^{-h(a)} + b \rho^k$$

Soient  $k$  et  $n$  des entiers naturels fixes quelconques. Pour tout  $a$ , on a :

$$\begin{aligned}
 E_x (e^{u_0 \sum_{p=1}^n (g(X_{kp}, X_{kp+1}) - a)}) \\
 &= E_x (e^{u_0 \sum_{p=1}^{n-1} (g(X_{kp}, X_{kp+1}) - a)} E_x (e^{u_0 (g(X_{kn}, X_{kn+1}) - a)}) / \beta^{k(n-1)+1}) \\
 &= E_x (e^{u_0 \sum_{p=1}^{n-1} (g(X_{kp}, X_{kp+1}) - a)} E_{X_{k(n-1)}} (e^{u_0 (g(X_k, X_{k+1}) - a)})
 \end{aligned}$$

En appliquant la relation démontrée ci-dessus, cette expression est majorée par

$$\begin{aligned}
 E_x (e^{u_0 \sum_{p=1}^{n-1} (g(X_{kp}, X_{kp+1}) - a)}) (e^{-h(a)} + b \rho^k) \text{ et en itérant on trouve} \\
 \sup_{x \in \mathcal{X}} E_x (e^{u_0 \sum_{p=1}^n (g(X_{kp}, X_{kp+1}) - a)}) \leq (e^{-h(a)} + b \rho^k)^n
 \end{aligned}$$

Soit  $r$  entier naturel,  $1 \leq r \leq k-1$ . Pour tout  $x$ ,

$$E_x (e^{u_0 \sum_{p=1}^n (g(X_{pk+r}, X_{pk+r+1}) - a)}) = E_x (E_{X_r} (e^{u_0 \sum_{p=1}^n (g(X_{kp}, X_{kp+1}) - a)}))$$

La dernière relation démontrée entraîne :

$$\sup_x E_x (e^{u_0 \{ \sum_{p=1}^n (g(X_{pk+r}, X_{pk+r+1}) - a) \}}) \leq e^{-h(a)} + b \rho^k n$$

Puisque  $a$  est strictement supérieur à  $m$ ,  $u_0$  est strictement positif et on peut écrire :

$$P_x \left( \sum_{p=k}^{nk-1} \{g(X_p, X_{p+1}) - a\} \geq 0 \right) \leq \sum_{r=0}^{k-1} P_x \left( \sum_{p=1}^{n-1} (X_{pk+r}, X_{pk+r+1}) - a \geq 0 \right) \\ \leq \sum_{r=0}^{k-1} E_x (e^{u_0 \sum_{p=1}^{n-1} (g(X_{pk+r}, X_{pk+r+1}) - a)})$$

$$\sup_x P_x \left( \sum_{p=k}^{nk-1} (g(X_p, X_{p+1}) - a) \geq 0 \right) \leq k (e^{-h(a)} + b \rho^k)^{n-1}$$

D'autre part, pour tout  $\epsilon$  strictement positif et pour tout  $a$

$$P_x \left( \sum_{p=0}^n (g(X_p, X_{p+1}) - a) \geq n\epsilon \right) \leq P_x \left( \sum_{p=0}^{k-1} (g(X_p, X_{p+1}) - a) \geq \frac{n\epsilon}{2} \right) \\ + P_x \left( \sum_{p=k}^{\lfloor \frac{n}{k} \rfloor k-1} (g(X_p, X_{p+1}) - a) \geq 0 \right) + P_x \left( \sum_{p=\lfloor \frac{n}{k} \rfloor k}^n (g(X_p, X_{p+1}) - a) \geq \frac{n\epsilon}{2} \right)$$

où pour tout réel  $s$ ,  $\lfloor s \rfloor$  désigne sa partie entière.

$g$  étant une v.a. bornée sur  $(\mathcal{H} \times \mathcal{H})$ , il existe  $n_0(k, \epsilon)$ , tel que

$\forall n \geq n_0$ ,

$$\sup_x P_x \left( \sum_{p=0}^{k-1} (g(X_p, X_{p+1}) - a) \geq \frac{n\epsilon}{2} \right) = 0$$

$$\text{et } \sup_x P_x \left( \sum_{p=\lfloor \frac{n}{k} \rfloor k}^n (g(X_p, X_{p+1}) - a) \geq \frac{n\epsilon}{2} \right) = 0$$

(cette dernière égalité étant vraie puisque  $\sum_{p=\lfloor \frac{n}{k} \rfloor k}^n$  comprend au plus  $k$  termes).

$$\text{De plus, } \sup_x P_x \left( \sum_{p=k}^{\lfloor \frac{n}{k} \rfloor k-1} (g(X_p, X_{p+1}) - a) \geq 0 \right) \leq k (e^{-h(a)} + b \rho^k)^{\lfloor \frac{n}{k} \rfloor - 1}$$

Par conséquent, il existe  $n_0(k, \epsilon)$ , tel que pour tout  $n \geq n_0$ ,

$$\sup_{x \in \mathcal{H}} P_x \left( \sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq n\epsilon \right) \leq k (e^{-h(a)} + b \rho^k)^{\lfloor \frac{n}{k} \rfloor - 1}$$

$$\frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} P_x \left( \sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq n\epsilon \right) \leq \frac{\text{Log } k}{n} + \frac{1}{n} (\lfloor \frac{n}{k} \rfloor - 1) \text{Log} (e^{-h(a)} + b \rho^k)$$

ce qui entraîne

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_x P_x^{n-1} \left( \sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq n\epsilon \right) \leq \frac{1}{k} \text{Log} (e^{-h(a)} + b\rho^k)$$

Cette dernière relation étant vraie pour tout  $k$ , on en déduit

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} P_x^{n-1} \left( \sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq n\epsilon \right) \leq \inf_{k \in \mathbb{N}} \left( \frac{1}{k} \text{Log} (e^{-h(a)} + b\rho^k) \right)$$

Soit maintenant  $\epsilon > 0$  tel que :  $a - \epsilon > m$ . (On a bien sûr  $a - \epsilon < \bar{g}$ ).

Appliquons alors la dernière inégalité démontrée à  $a - \epsilon$ , en notant la constante  $b$  correspondant à un réel  $a$ ,  $b_a$ . On obtient :

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} P_x^{n-1} \left( \sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq 0 \right) \leq \inf_{k \in \mathbb{N}} \left( \frac{1}{k} \text{Log} (e^{-h(a-\epsilon)} + b_{a-\epsilon} \rho^k) \right)$$

Faisons tendre  $\epsilon$  vers 0.

$h$  est une fonction convexe donc continue :  $\lim_{\epsilon \rightarrow 0} h(a - \epsilon) = h(a)$ .

$$\text{On a : } b_{a-\epsilon} = c \parallel e^{h'(a-\epsilon) \cdot (g-a+\epsilon)} \parallel$$

$h'$  est continue donc :  $\lim_{\epsilon \rightarrow 0} b_{a-\epsilon} = b_a$ .

On en déduit que pour tout  $a$ ,  $m < a < \bar{g}$  :

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} P_x^{n-1} \left( \sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq 0 \right) \leq \inf_{k \in \mathbb{N}} \left( \frac{1}{k} \text{Log} (e^{-h(a)} + b_a \rho^k) \right)$$

Cet inf est strictement négatif puisque  $h(a)$  étant strictement positif ( $a \neq m$ ) et  $\rho$  strictement inférieur à 1, il existe  $k_0$  de  $\mathbb{N}$  tel que  $(e^{-h(a)} + b_a \rho^{k_0})$  soit strictement inférieur à 1.

Ceci termine la démonstration de la 4ème partie du théorème. Le résultat de la 1ère se déduit immédiatement puisque la fonction

$$a \rightarrow \overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} P_x^{n-1} \left( \sum_{p=0}^{n-1} (g(X_p, X_{p+1}) - a) \geq 0 \right) \text{ est croissante.}$$

Remarque : Soit  $a \in ]m, \bar{g}[$ . Si  $g$  et  $c$  vérifient :  $\parallel g_{||} \geq a$  et  $c \geq 1$ , la constante  $b_a = c \parallel e^{u_0(g-a)} \parallel$  est supérieure à 1 et  $\inf_{k \in \mathbb{N}} \frac{1}{k} \text{Log} (e^{-h(a)} + b_a \rho^k)$

est minorée par  $\text{Log} \rho$ . Donc en général cet inf ne sera pas  $-\infty$  même si

$\parallel g_{||} \geq a > \bar{g}$ . L'exemple suivant met en évidence cette minoration par  $\text{Log} \rho$ .

Soit la chaîne markovienne à 2 états 0 et 1 de probabilité de transition :  $\pi(0,0) = \rho$ ,  $\pi(1,1) = 1$ . Sa mesure invariante  $\mu$  est  $\delta_1$ . Supposons  $\rho > \frac{1}{2}$ , alors les constantes  $c$  et  $\rho$  sont respectivement 1 et  $\sup(\rho, 1-\rho) = \rho$ .

La v.a.  $g$  définie par :  $g(1) = 0$ ,  $g(0) = 1$  est telle que  $\bar{g} = 0$  et on a :

$$P_0 \left( \sum_{p=0}^{n-1} g(X_p) \geq n \right) = \rho^n .$$

$$\frac{1}{n} \text{Log } P_0 \left( \sum_{p=0}^{n-1} g(X_p) \geq n \right) = \text{Log } \rho = \frac{1}{n} \text{Log } P_0 \left( \sum_{p=0}^{n-1} (g(X_p) - 1) \geq 0 \right) .$$

## DEUXIEME PARTIE : STATISTIQUE.

Maintenant, le noyau  $\pi$  dépend d'un paramètre. L'espace des paramètres  $(\Theta \times \underline{\Theta})$  est un espace topologique muni de sa tribu borélienne. La transition  $\pi$  est une transition de  $\Theta \times \mathcal{H}$  dans  $\mathcal{H}$ . Le processus des observations  $X^\theta = (\Omega, \mathcal{A}, (P_{\theta, x})_{x \in \mathcal{H}}, (X_n)_{n \in \mathbb{N}}$  est une chaîne markovienne récurrente, Doeblin, apériodique, de probabilité invariante  $\mu_\theta$ . On se pose des problèmes d'estimation et de test d'un point de vue séquentiel.

### I. ESTIMATION

#### 1. Définitions. Exemples.

Ce paragraphe reste valable si la chaîne  $X^\theta$  est seulement récurrente Harris positive (cadre 1).

1.1. Fonction de contraste : la définition est celle de Gänsler [6].

Une v.af. sur  $\Theta \times \mathcal{H} \times \mathcal{H}$  est une fonction de contraste si

$\int d\mu_\theta(x) \int \pi(\theta, x, dy) f(\theta, x, y)$  est fini et  $\int d\mu_\theta(x) \int \pi(\theta, x, dy) (f(\varphi, x, y) - f(\theta, x, y)) \geq 0$  l'égalité n'ayant lieu que lorsque  $\varphi = \theta$ .

On notera  $h(\vartheta, \varphi, x, y) = f(\varphi, x, y) - f(\vartheta, x, y)$

et  $K(\vartheta, \varphi) = \int \mu_{\vartheta}(x) \pi(\vartheta, x, dy) h(\vartheta, \varphi, x, y)$

Exemple : On suppose qu'il existe une v.a.  $p$  sur  $\Theta \times \mathcal{H} \times \mathcal{H}$ , strictement positive et une transition  $\lambda$  de  $\mathcal{H}$  dans  $\mathcal{H}$  telles que pour  $x \in \mathcal{H}$  et  $\vartheta \in \underline{\mathcal{H}}$ ,

$$\pi(\vartheta, x; \vartheta) = \int p(\vartheta, x, y) \lambda(x, dy) I_{\vartheta}(y)$$

On suppose aussi que si  $\varphi$  est différent de  $\vartheta$ , l'événement  $\{x; \pi(\varphi, x, \cdot) \neq \pi(\vartheta, x, \cdot)\}$  n'est pas négligeable pour  $\mu_{\vartheta}$ .

Sous ces hypothèses, la v.a.  $f : f(\varphi, x, y) = -\text{Log } p(\varphi, x, y)$  est une fonction de contraste dès que les propriétés d'intégrabilité sont vraies. En outre,  $p$  étant strictement positive, pour tout  $x$  de  $\mathcal{H}$ , l'ensemble  $\{\pi(\varphi, x; \cdot)\}_{\varphi \in \Theta}$  est un ensemble de probabilités équivalentes. Aussi, si  $p$  est bornée,

$$I(\vartheta, \varphi, x) = \int \text{Log} \frac{p(\vartheta, x, y)}{p(\varphi, x, y)} \pi(\vartheta, x; dy) = I_{\pi(\varphi, x; \cdot)}(\pi(\vartheta, x; \cdot)),$$

où  $I_{\mu}(\nu)$  est l'information de Kullback de  $\nu$  par rapport à  $\mu$ . Pour simplifier, on note  $I_{\pi(\varphi, x; \cdot)}(\pi(\vartheta, x; \cdot))$ ,  $I(\vartheta, \varphi, x)$ , et on définit  $I(\vartheta, \varphi)$  par  $\int I(\vartheta, \varphi, x) \mu_{\vartheta}(x)$ . Dès que  $\varphi$  est différent de  $\vartheta$ ,  $I(\vartheta, \varphi)$  est strictement positif.

### 1.2. Estimateur du minimum de contraste :

A l'instant  $n$ , un estimateur de  $\vartheta$  est une fonction mesurable  $\hat{\vartheta}_n$  de  $(\Omega, \mathcal{B}_n)$  dans  $\Theta$ , c'est-à-dire que  $\hat{\vartheta}_n = \tilde{\vartheta}_n(X_1, \dots, X_n)$  pour une fonction mesurable  $\tilde{\vartheta}_n$  de  $\mathcal{H}$  dans  $\Theta$ . On confondra  $\hat{\vartheta}_n$  et  $\tilde{\vartheta}_n$ .

Etant donnée  $f$  une fonction de contraste on pose :

$$L_n(\alpha) = \sum_{k=0}^{n-1} f(\alpha, X_k, X_{k+1})$$

A l'instant  $n$ , un estimateur  $\hat{\vartheta}_n$ ,  $\mathcal{B}_n$  mesurable est un estimateur du minimum de contraste s'il vérifie :

$$L_n(\hat{\vartheta}_n) = \inf_{\alpha \in \Theta} L_n(\alpha)$$

Dans l'exemple précédent cet estimateur est l'estimateur du maximum de vraisemblance. Le théorème de sélection de BROWN [3] assure l'existence d'une suite d'estimateurs du minimum de contraste si  $\mathcal{X}$  est polonais, si  $\Theta$  est un espace métrique compact, et si l'application  $\alpha \rightarrow f(\alpha, x, y)$  est semi continue inférieurement de  $\Theta$  dans  $\mathbb{R}$  pour tout  $x, y$  de  $\mathcal{X}$ .

On pose :

$$S_n(\alpha, \varphi) = \sum_{k=0}^{n-1} (f(\varphi, X_k, X_{k+1}) - f(\alpha, X_k, X_{k+1})) = \sum_{k=0}^{n-1} h(\alpha, \varphi, X_k, X_{k+1})$$

## 2. Consistance exponentielle.

Si  $\lambda_0$  est la loi de probabilité initiale de la chaîne, une suite  $\{\hat{\vartheta}_n\}_{n \in \mathbb{N}}$  d'estimateurs du minimum de contraste est dite consistante si pour tout voisinage  $U(\theta)$  de  $\theta$ ,  $P_{\theta, \lambda_0}$  ps, il existe  $n_0$ , tel que pour tout  $n \geq n_0$ ,  $\hat{\vartheta}_n$  appartient à  $U(\theta)$ . Dans [6] on trouve des hypothèses sous lesquelles une suite d'estimateurs du minimum de contraste est consistante dans le cadre 1.

Le but est d'établir ici, la consistance exponentielle de toute suite  $(\hat{\vartheta}_n)_{n \in \mathbb{N}}$  d'estimateurs du minimum de contraste lorsque  $\Theta$  est fini, et la fonction de contraste bornée.

On définit :  $T_\theta = \inf \{n; \forall p \geq n, \hat{\vartheta}_p = \theta\}$

Théorème 2 : Lorsque  $\Theta$  est fini, et la fonction de contraste bornée, toute suite  $(\hat{\vartheta}_n)_{n \in \mathbb{N}}$  d'estimateurs du minimum de contraste est exponentiellement consistante :

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{X}} P_{\theta, x} (T_\theta > n) < 0.$$

Démonstration :

Pour tout  $x$  de  $\mathcal{X}$ ,  $P_{\theta, x} (T_\theta > n) \leq \sum_{m \geq n} \sum_{\varphi \neq \theta} P_{\theta, x} (S_m(\theta, \varphi) \leq 0)$

Soit  $\varphi$  différent de  $\theta$ .

$$P_{\theta, x} (S_m(\theta, \varphi) \leq 0) = P_{\theta, x} \left( \sum_{k=0}^{m-1} h(\theta, \varphi, X_k, X_{k+1}) \leq 0 \right)$$

Le théorème 1.1. du paragraphe I, appliqué à la fonction bornée  
 $-h(\theta, \varphi, x, y)$  entraîne que pour tout  $\epsilon > 0$ , il existe  $k > 0$  et  $b > 0$ , tels  
 que pour tout  $m$  de  $\mathbb{N}$ ,

$$\sup_x P_{\theta, x} \left( \sum_{k=0}^{m-1} \{h(\theta, \varphi, X_k, X_{k+1}) - (K(\theta, \varphi) - \epsilon)\} \leq 0 \right) \leq k e^{-bm}$$

Pour  $\epsilon = K(\theta, \varphi) > 0$ , il existe  $k > 0$  et  $b > 0$ , tels que pour tout  
 $m$  de  $\mathbb{N}$ ,

$$\sup_x P_{\theta, x} \left( \sum_{k=0}^{m-1} h(\theta, \varphi, X_k, X_{k+1}) \leq 0 \right) \leq k e^{-bm}$$

ce qui entraîne qu'il existe  $k' > 0$  et  $b > 0$  tels que pour tout  $n$  de  $\mathbb{N}$ ,

$$\sup_{x \in \mathcal{H}} P_{\theta, x} (T_{\theta} > n) \leq k' e^{-bn}$$

c'est-à-dire :  $\overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} P_{\theta, x} (T_{\theta} > n) < 0$ .

## II. TEST SEQUENTIEL DU RAPPORT DE VRAISEMBLANCE.

### 0. Introduction :

L'idée de ce paragraphe est inspirée du problème de test résolu par  
 CHERNOFF dans le cas indépendant avec choix d'expériences. Ici, l'espace des  
 paramètres  $\Theta$  est fini (on notera  $q$  sont cardinal). Le processus des obser-  
 vations  $X^{\theta}$  est une chaîne markovienne, récurrente Doeblin, apériodique, de  
 probabilité invariante  $\mu_{\theta}$ , de loi initiale  $\lambda_0$ . On suppose qu'il existe une  
 v.a. sur  $\Theta \times \mathcal{H} \times \mathcal{H}$  strictement positive  $p$ , bornée, et une transition  $\lambda$  de  
 $\mathcal{H}$  dans  $\mathcal{H}$ , telles que pour tout  $x$  de  $\mathcal{H}$ , pour tout  $\Gamma$  de  $\underline{\mathcal{H}}$ ,

$$\pi(\theta, x; \Gamma) = \int p(\theta, x, y) \lambda(x, dy) 1_{\Gamma}(y) .$$

On suppose aussi que dès que  $\varphi$  est différent de  $\theta$ , l'événement  
 $\{x, \pi(\theta, x, \cdot) \neq \pi(\varphi, x, \cdot)\}$  n'est pas négligeable.

C'est le cadre de l'exemple du paragraphe I, dont on reprend les nota-  
 tions : Pour tout  $x$  de  $\mathcal{H}$ , pour tout  $\theta$  et  $\varphi$  de  $\Theta$  :



$$I(\theta, \varphi, x) = \int \text{Log} \frac{p(\theta, x, y)}{p(\varphi, x, y)} \Pi(\theta, x, dy)$$

et  $I(\theta, \varphi) = \int I(\theta, \varphi, x) \mu_{\theta}(dx)$  .

On définira de plus :  $I(\theta) = \inf_{\varphi \in a(\theta)} I(\theta, \varphi)$ .

Sous les hypothèses ci-dessus,  $f(\theta, \dots) = -\text{Log} p(\theta, \dots)$  est une fonction de contraste bornée. On désigne dans la suite par  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  une suite d'estimateurs du minimum de contraste ; d'après le théorème 2.1 précédent, elle est exponentiellement consistante :  $\overline{\lim}_n \frac{1}{n} \text{Log} \sup_x P_{\theta, x} (T_{\theta} > n) < 0$ . On se pose le problème de test suivant :  $(\theta_1, \theta_2)$  étant une partition de  $\Theta$ , on veut savoir à quel élément de la partition appartient la vraie valeur du paramètre.  $c$  est le coût commun à chaque expérience,  $0 < c < 1$ , et  $r(\cdot)$  la perte due à l'erreur de décision.

On note :  $h(\theta) = \theta_i$  si  $\theta$  appartient à  $\theta_i$  ( $i=1$  ou  $2$ ), et  $a(\theta) = \Theta - h(\theta)$ .  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  désigne alors une suite d'estimateurs, telle que pour tout  $n$ ,  $\hat{\theta}_n$  soit  $\mathcal{B}_n$  mesurable et soit défini par :

$$\sum_{k=0}^{n-1} \text{Log} p(\hat{\theta}_n, X_k, X_{k+1}) = \sup_{\alpha \in a(\hat{\theta}_n)} \sum_{k=0}^{n-1} \text{Log} p(\alpha, X_k, X_{k+1}) .$$

Les résultats obtenus ne dépendent pas de la loi initiale  $\lambda_0$  : on notera  $P_{\theta, \lambda_0}$  (respectivement  $E_{\theta, \lambda_0}$ )  $P_{\theta}$  ( $E_{\theta}$ ).

La théorie développée ici est une théorie asymptotique, pour des valeurs de  $c$  petites.

1. Etude d'un temps d'arrêt.

Soit  $b$  positif,  $\theta$  et  $\varphi$  de  $\Theta$ . On définit :

$$\zeta_b = \inf \{n; S_n(\hat{\theta}_n, \hat{\theta}_n) > b\}$$

et  $\zeta_{b, \varphi} = \inf \{n; \forall p \geq n, S_p(\theta, \varphi) > b\}$  .

On a la relation :  $\zeta_b \leq \max(T_{\theta}, \max_{\varphi \in a(\theta)} \zeta_{b, \varphi})$

la suite  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  étant exponentiellement consistante, il existe  $k > 0$  et  $b > 0$ , tels que pour tout  $n$  de  $\mathbb{N}$ , on ait :  $P_\theta(T_\theta > n) \leq ke^{-bn}$ , donc  $P_\theta(T_\theta = \infty) = 0$ .

La chaîne étant récurrente Harris positive et  $\text{Log} \frac{p(\theta, \dots)}{p(\varphi, \dots)} \mu_\theta \otimes \Pi_\theta$  intégrable, un théorème d'ergodicité implique :

$$\frac{1}{n} \sum_{k=0}^{n-1} \text{Log} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})} \xrightarrow{P_\theta \text{ ps}} \int \text{Log} \frac{p(\theta, x, y)}{p(\varphi, x, y)} \Pi(\theta, x, dy) \mu_\theta(dx) = I(\theta, \varphi)$$

donc pour tout  $\varphi$  différent de  $\theta$ , puisque  $I(\theta, \varphi)$  est positif non nul,

$$S_n(\theta, \varphi) \xrightarrow{P_\theta \text{ ps}} \infty$$

ce qui entraîne  $P_\theta(\zeta_{b, \varphi} = \infty) = 0$ .

Finalement :  $P_\theta(\zeta_b = \infty) = 0$ . Le temps d'arrêt  $\zeta_b$  est fini  $P_\theta$  presque sûrement.

## 2. Définition d'un test $\phi_c$ .

On note  $\nu_c$  le temps d'arrêt  $\zeta_{-\text{Log } c}$  défini dans le paragraphe précédent. De la même façon on note  $\nu_{c, \varphi}$  le temps d'arrêt  $\zeta_{-\text{Log } c, \varphi}$ .

On définit le test de Chernoff, noté  $\phi_c$ , ainsi :

sa règle d'arrêt est  $\nu_c$   
 si  $n = \nu_c$ , on arrête et on décide que  $\theta$  est élément de  $h(\hat{\theta}_n)$   
 si  $n < \nu_c$ , on continue, c'est-à-dire on refait une expérience et on regarde à nouveau dans quelle situation on se trouve.

Le test  $\phi_c$  est clos pour toute valeur  $c$  du coût.

$\beta_c(\theta)$  et  $R_c(\theta)$  désigneront respectivement la probabilité d'erreur et le risque associés au test de Chernoff  $\phi_c$ .

3. Propriétés du test.

Le but est de montrer que le test  $\phi_c$  possède une propriété d'optimalité analogue à celle obtenue par Chernoff dans le cas indépendant avec contrôle.

3.a. La probabilité d'erreur

Théorème 3.a : Pour tout  $\theta$  de  $\Theta$ , (card  $\Theta = q$ ), la probabilité d'erreur

$\beta_c(\theta)$  associée au test  $\phi_c$  vérifie :  $\beta_c(\theta) \leq cq$ .

Démonstration : Soit  $\theta$  de  $\Theta$ .

Pour  $\varphi$  élément de  $a(\theta)$ , on définit :  $\Omega_{n,\varphi} = \{\hat{\theta}_n = \varphi, \nu_c = n\}$ .

Alors :  $\beta_c(\theta) = \sum_{n=1}^{\infty} \sum_{\varphi \in a(\theta)} P_{\theta}(\Omega_{n,\varphi})$ , avec :

$$P_{\theta}(\Omega_{n,\varphi}) = P_{\theta} \left( \prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})} \leq c \cap \nu = n \right)$$

Mais pour toute fonction  $g$  mesurable :

$$E_{\theta}(g(X_0, \dots, X_{n-1})) = E_{\varphi}(g(X_0, \dots, X_{n-1})) \prod_{i=0}^{n-1} \frac{p(\theta, X_i, X_{i+1})}{p(\varphi, X_i, X_{i+1})}$$

c'est-à-dire 
$$P_{\theta} = \prod_{i=0}^{n-1} \frac{p(\theta, X_i, X_{i+1})}{p(\varphi, X_i, X_{i+1})} P_{\varphi} \text{ sur } \mathcal{B}_n$$

D'où : 
$$P_{\theta}(\Omega_{n,\varphi}) = E_{\varphi} \left\{ 1_{\left\{ \prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})} \leq c \right\}} 1_{(\nu=n)} \prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})} \right\}$$

$$P_{\theta}(\Omega_{n,\varphi}) \leq c P_{\varphi} \left( \prod_{k=0}^{n-1} \frac{p(\theta, X_k, X_{k+1})}{p(\varphi, X_k, X_{k+1})} \leq c, \cap \nu = n \right)$$

$$\sum_{n=1}^{\infty} P_{\theta}(\Omega_{n,\varphi}) \leq c$$

Par conséquent :  $\beta_c(\theta) \leq cq$

3.b. La durée moyenne du test.

Théorème 3.b. : Pour tout  $\theta$  de  $\Theta$ ,  $E_{\theta}(v_c) \leq \frac{-(1+\frac{\epsilon}{2}) \text{Log } c}{I(\theta)}$

Démonstration : Soit  $\epsilon > 0$  et  $\theta$  un élément de  $\Theta$ .

D'après le théorème 2.1, il existe  $b_1 > 0$  et  $k_1 > 0$  tels que pour tout  $n$  de  $N$ , pour tout  $\varphi$  de  $a(\theta)$ , on ait :

$$P_{\theta}(S_n(\theta, \varphi) \leq \frac{n I(\theta)}{1 + \frac{\epsilon}{2}}) \leq k_1 e^{-b_1 n}$$

De la relation :  $P_{\theta}(v_{c, \varphi} = n) \leq P_{\theta}(S_{n-1}(\theta, \varphi) \leq -\text{Log } c)$ , on déduit qu'il existe  $k_1 > 0$  et  $b_1 > 0$ , tels que pour tout  $\varphi$  de  $a(\theta)$ , pour tout

$n \geq \frac{-(1 + \frac{\epsilon}{2}) \text{Log } c}{I(\theta)}$ , on ait :

$$P_{\theta}(v_{c, \varphi} = n) \leq k_1 e^{-b_1 n}.$$

Ce résultat d'une part, la consistance exponentielle de la suite  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  d'autre part, et la relation :

$$\{v_c = n\} \subset \{T_{\theta} > n\} \cup \left\{ \bigcup_{\varphi \in a(\theta)} v_{c, \varphi} = n \right\}$$

entraîne qu'il existe  $k > 0$  et  $b > 0$  tels que pour tout  $n \geq \frac{-(1 + \frac{\epsilon}{2}) \text{Log } c}{I(\theta)}$

on ait :

$$P_{\theta}(v_c = n) \leq k e^{-bn}.$$

$$\text{Soit alors : } N_c = \frac{-(1 + \frac{\epsilon}{2}) \text{Log } c}{I(\theta)}$$

$$E_{\theta}(v_c) = E_{\theta}(v_c \cdot 1_{(v_c \leq N_c)}) + E_{\theta}(v_c \cdot 1_{(v_c > N_c)})$$

$$E_{\theta}(v_c) \leq N_c + \sum_{n > N_c} n P_{\theta}(v_c = n)$$

$$E_{\theta}(v_c) \leq N_c + k \sum_{n > N_c} n e^{-bn}$$

Comme  $\sum_{n > N_c} n e^{-bn} = \bar{\xi}_c$  tend vers 0 quand  $c$  tend vers 0, il existe

$c_0(\epsilon)$  tel que pour  $c < c_0(\epsilon)$ , on ait :  $\bar{\xi}_c \leq \frac{-\epsilon \text{Log } c}{2k I(\theta)}$ .

Aussi, il existe  $c_0(\epsilon)$ , tel que pour tout  $c < c_0(\epsilon)$ ,

$$E_{\theta}(v_c) \leq \frac{-(1+\epsilon) \text{Log } c}{I(\theta)} .$$

Corollaire : Pour tout  $\theta$  de  $\Theta$ ,  $R_c(\theta) \leq \frac{-(1+\sigma(1)) c \text{Log } c}{I(\theta)}$

Démonstration :

$$R_c(\theta) = c E_{\theta}(v_c) + \beta_c(\theta) r(\theta)$$

Les 2 théorèmes précédents permettent de conclure facilement.

### 3.c. Optimalité asymptotique de la procédure $\phi_c$

On note  $(T, \delta)$  une procédure séquentielle de temps d'arrêt  $T$  et de règle de décision  $\delta$ .  $e(\theta)$  sera la probabilité d'erreur de la procédure  $(T, \delta)$ .

Lemme 3.a : Soit  $b$ ,  $0 < b < 1$  et  $\theta$  de  $\Theta$ . Soit  $(T, \delta)$  une procédure telle que pour tout  $\varphi$ ,  $e(\varphi)$  soit un  $\mathcal{O}(-b \text{Log } b)$ . Pour tout  $\varphi$  de  $a(\theta)$ , pour tout  $\epsilon$ ,  $0 < \epsilon < 1$ , on a :

$$P_{\theta}(S_T(\theta, \varphi) < -(1-\epsilon) \text{Log } b) = \mathcal{O}(-b^{\epsilon} \text{Log } b) .$$

Démonstration : Soit  $\theta$  de  $\Theta$ ,  $\varphi$  appartenant à  $a(\theta)$ , et  $\epsilon$ ,  $0 < \epsilon < 1$ .

On définit :  $\Omega_n = \{T=n, S_n(\theta, \varphi) < -(1-\epsilon) \text{Log } b, \delta \text{ accepte } h(\theta)\}$ .

$$\begin{aligned} \text{Alors : } e(\varphi) &\geq \sum_{n=1}^{\infty} P_{\varphi}(\Omega_n) \\ &\geq \sum_{n=1}^{\infty} P_{\varphi} \left( \prod_{k=0}^{n-1} \frac{p(\varphi, X_k, X_{k+1})}{p(\theta, X_k, X_{k+1})} > b^{1-\epsilon} \right) \cap \{T=n\} \cap \{h(\theta)\} \end{aligned}$$

où " $h(\theta)$ " signifie : la règle de décision décide  $h(\theta)$ .

$$\text{Comme } P_{\varphi} = \prod_{k=0}^{n-1} \frac{p(\varphi, X_k, X_{k+1})}{p(\theta, X_k, X_{k+1})} P_{\theta} \text{ sur } \mathcal{B}_n, \text{ on a :}$$

$$e(\varphi) \geq \sum_{n=1}^{\infty} E_{\theta} \left( 1 - \prod_{k=0}^{n-1} \frac{p(\varphi, X_k, X_{k+1})}{p(\theta, X_k, X_{k+1})} > b^{1-\epsilon} \right) \mathbb{1}_{\{T=n\}} \mathbb{1}_{\{h(\theta) \geq \frac{1}{n} \sum_{k=0}^{n-1} \frac{p(\varphi, X_k, X_{k+1})}{p(\theta, X_k, X_{k+1})}\}}$$

$$e(\varphi) \geq b^{1-\epsilon} \sum_{n=1}^{\infty} P_{\theta}(\Omega_n)$$

$$b^{\epsilon-1} e(\varphi) \geq \sum_{n=1}^{\infty} P_{\theta}(\Omega_n)$$

$$\sum_{n=1}^{\infty} P_{\theta}(\Omega_n) = \mathcal{O}(-b^{\epsilon} \text{Log } b)$$

Comme :

$$\begin{aligned} P_{\theta}(S_T(\theta, \varphi) < -(1-\epsilon)\text{Log } b) &= P_{\theta} \left( \bigcup_{n=1}^{\infty} (T=n, S_n(\theta, \varphi) < -(1-\epsilon)\text{Log } b) \right) \\ &\leq \sum_{n=1}^{\infty} P_{\theta}(\Omega_n) + e(\theta) \end{aligned}$$

On conclut :  $P_{\theta}(S_T(\theta, \varphi) < -(1-\epsilon)\text{Log } b) = \mathcal{O}(-b^{\epsilon} \log b)$

Lemme 3.b. :  $\forall \epsilon > 0, P_{\theta} \left( \max_{0 \leq m \leq n} \min_{\varphi \in a(\theta)} S_m(\theta, \varphi) \geq n(I(\theta) + \epsilon) \right) \xrightarrow[n \rightarrow \infty]{} 0$

Démonstration : Soient  $\epsilon$  et  $\eta$  positifs non nuls.

La chaîne étant récurrente Harris positive, d'après un théorème ergodique

on a :

$$\forall \varphi, \frac{1}{n} S_n(\theta, \varphi) \xrightarrow[n \rightarrow \infty]{} I(\theta, \varphi) \text{ sur } (\Omega, \mathcal{A}, P_{\theta}) \text{ en probabilité.}$$

Il existe donc  $n_0$  tel que pour tout  $n \geq n_0$  :

$$P_{\theta} \left( \max_{n_0 \leq m \leq n} \min_{\varphi \in a(\theta)} S_m(\theta, \varphi) \geq n(I(\theta) + \epsilon) \right) \leq \eta$$

Pour tout  $m$ ,  $S_m(\theta, \varphi)$  est fini. Il existe donc  $n_1$  tel que pour tout

$n \geq n_1$  :

$$P_{\theta} \left( \max_{0 \leq m \leq n_0} \min_{\varphi \in a(\theta)} S_m(\theta, \varphi) \geq n(I(\theta) + \epsilon) \right) = 0$$

D'où , il existe  $n_2$  tel que pour tout  $n$  supérieur à  $n_2$  :

$$P_{\theta} \left( \max_{0 \leq m \leq n} \min_{\varphi \in a(\theta)} S_m(\theta, \varphi) \geq n(I(\theta) + \epsilon) \right) \leq \eta$$

Lemme 3.c. : Soit  $b$ ,  $0 < b < 1$ . Si  $(T, \delta)$  est une procédure telle que pour tout  $\varphi$ , la probabilité d'erreur  $e(\varphi)$  soit un  $\mathcal{O}(-b \text{ Log } b)$ , elle vérifie :

$$E_{\theta}(T) \geq \frac{-(1+\sigma(1)) \text{ Log } b}{I(\theta)} .$$

Démonstration : Soit  $\epsilon > 0$ ,  $0 < b < 1$  et soit  $(T, \delta)$  une procédure vérifiant l'hypothèse du lemme. On définit les constantes  $\epsilon'$ ,  $\epsilon''$  et  $n_b$  par :

$$\epsilon' = \frac{I(\theta)}{1+\epsilon + I(\theta)} , \quad (1-\epsilon'')^2 = (1-\epsilon') \quad \text{et} \quad n_b = \frac{-(1-\epsilon'') \text{ Log } b}{I(\theta) + \epsilon'}$$

On a :

$$P_{\theta}(T \leq n_b) \leq P_{\theta}(T \leq n_b \text{ et pour tout } \varphi \text{ de } a(\theta), S_T(\theta, \varphi) \geq -(1-\epsilon'') \text{ Log } b)$$

$$+ P_{\theta}(\text{il existe } \varphi \text{ de } a(\theta), S_T(\theta, \varphi) < -(1-\epsilon'') \text{ Log } b)$$

Notons  $K_1^b [K_2^b]$  le 1er  $[2^{\text{ème}}]$  terme du membre de gauche.

$$K_1^b \leq P_{\theta} \left( \max_{1 \leq m \leq n_b} \min_{\varphi \in a(\theta)} S_m(\theta, \varphi) \geq n_b (I(\theta) + \epsilon') \right)$$

D'après le lemme 3.b,  $K_1^b$  tend vers 0 quand  $n_b$  tend vers l'infini, c'est-à-dire quand  $b$  tend vers 0.

Quant au lemme 3.a, il permet de conclure que  $K_2^b$  tend vers 0 quand  $b$  tend vers 0.

Par conséquent, il existe  $b_0(\epsilon)$ , tel que pour tout  $b \leq b_0(\epsilon)$ , on ait :

$$P_{\theta}(T \leq n_b) \leq \epsilon'' .$$

$$\text{Comme : } E_{\theta}(T) = \sum_{n=0}^{\infty} P_{\theta}(T_{\theta} > n) \geq n_b P_{\theta}(T > n_b), \quad \text{pour tout } b \leq b_0(\epsilon),$$

on a :

$$\begin{aligned}
E_{\theta}(T) &\geq n_b (1-\epsilon'') \geq \frac{-(1-\epsilon'')^2 \text{Log } b}{I(\theta) + \epsilon'} \\
&\geq \frac{-(1-\epsilon') \text{Log } b}{I(\theta) + \epsilon'} \\
&\geq \frac{-(1+\epsilon) \text{Log } b}{I(\theta)}
\end{aligned}$$

Tous ces résultats permettent de montrer la propriété d'optimalité asymptotique suivante.

Théorème 3.c. : Le test  $\phi_c$  est asymptotiquement le plus économique parmi les tests de force au moins égale.

Démonstration : Soit  $(T, \delta)$  une procédure telle que la probabilité d'erreur  $e(\varphi)$  soit un  $\mathcal{O}(-c \text{Log } c)$  pour tout  $\varphi$ . D'après le lemme 3.3.3, elle vérifie :

$$E_{\theta}(T) \geq \frac{-(1+\sigma(1)) \text{Log } c}{I(\theta)} .$$

Ce résultat, comparé aux résultats des théorèmes 3.a et 3.b précédents permet de conclure que  $\phi_c$  est asymptotiquement le plus économique parmi les tests de force au moins égale.



Bibliographie :

- [ 1] ALBERT A.E. : "The sequential design of experiments for infinitely many states of nature". Ann. Math. Stat., vol. 32 (1961), 774-799
- [ 2] BESSLER S. : "Theory and applications of the sequential design of experiments, k actions and infinitely many experiments". Part I, II. Mech. Report. N° 55, 56. 1960
- [ 3] BROWN : "Measurable selections of extrema"  
Annals of Statistics, vol. 1, n° 5, 902-912 (1973)
- [ 4] CHERNOFF H. : "Sequential design of experiments"  
Ann. Math. Stat., vol. 30 (1959). 755-770.
- [ 5] DONSKER et VARADHAN : "Asymptotic evaluation of certain process expectations for large time" Communication on pure and applied mathematics  
I. vol. 28, 1-47 (1975)
- [ 6] GANSLER : "Note on minimum contrast estimates for Markov processes"  
Metrika 19, 115-130 (1972)
- [ 7] GREY : "Limit theorems for markov chains transitions probabilities"  
Van Nostrand (1971)

Nelly MAIGRET  
Université Paris Nord  
Mathématiques  
ERA CNRS 532 "Statistique Appliquée"  
Avenue J.B. Clément  
93430 VILLETANEUSE.

BIBLIOGRAPHIE

- [1] DONSKER - VARADHAN : Asymptotic evaluation of certain Markov process expectations for large time. Com. on pure and appl. maths.  
I Vol. 28 p. 1-47 (1975)  
II Vol. 29 p. 279-301 (1976)  
III Vol. 29 p. 389-461 (1976)
- [2] COGBURN : A uniform theory for sums of Markov chain transition probabilities  
Ann. prob. Vol. 3 , 2, p. 191-214 (1975)
- [3] GÄRTNER : On large deviations for the invariant measure. Theory of prob. and appl., Vol. 22, p. 24-39 (1977)
- [4] MAIGRET : Inégalités de Chernoff pour des chaînes de Markov contrôlées.  
Z. Wahrscheinlichkeit (1979).
- [5] PARTHASARATHY - RAGHARAN : Some topics in two person games  
(Elsevier 1971)
- [6] REVUZ : Markov chains Holden Day
- [7] SION : On general minimax theorem Pacific J. Math. 8 p. 171-176
- [8] TUOMINEN, TWEEDIE : Markov chains with continuous components -  
A paraître .
- [9] VARADHAN : Asymptotic probabilities and differential equations -  
Com. on pure and appl. Maths. Vol. 19, p. 261-285 (1966)
- [10] JEAN SAINT RAYMOND : Quelques remarques sur un article de Donsker et Varadhan - Séminaire de probabilités XII Strasbourg (Springer-Verlag)

Marie DUFLO

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



STATISTIQUE DES CHAINES  
CONTROLEES FELLERIENNES

Nelly MAIGRET

Résumé :

On se pose ici, pour des chaînes contrôlées felleriennes, des problèmes d'estimation et de tests séquentiels. En s'inspirant du test de Chernoff dans le cas indépendant [1], on construit une procédure. Grâce à un théorème de grandes déviations établi dans [4], on montre que cette procédure est asymptotiquement la plus économique parmi les procédures de force au moins égale, pour le test de deux hypothèses simples.

Axiomatique et hypothèses

L'espace mesurable  $(\Omega, \mathcal{A})$  est l'espace des épreuves, l'espace mesurable  $(\mathcal{H}, \underline{\mathcal{H}})$  celui des observations et  $(\mathcal{A}, \underline{\mathcal{A}})$  est l'espace des actions. Pour toute suite  $(u_n)_{n \geq 0}$ , on note  $u^{(n)} = (u_0, \dots, u_n)$ .

$(A_n)_{n \in \mathbb{N}}$  est le processus des actions (l'évolution du système est contrôlée au temps  $n$ , par une action  $A_n$  mesurable de  $(\Omega, \mathcal{A})$  dans  $(\mathcal{A}, \underline{\mathcal{A}})$ ), et  $(X_n)_{n \in \mathbb{N}}$  est le processus des observations ( $X_n$  est l'état de la chaîne au temps  $n$ ;  $X_n$  est mesurable de  $(\Omega, \mathcal{B}_n)$  dans  $(\mathcal{H}, \underline{\mathcal{H}})$ ). On note  $\mathcal{B}_n$  la tribu des évènements observables jusqu'au temps  $n$ :  $\mathcal{B}_n = \sigma(X^{(n)}, \mathcal{A}^{(n-1)})$ .

A chaque instant  $n$ , on doit choisir l'action  $A_n$  au vu des observations faites jusqu'au temps  $n$ . Une stratégie aléatoire  $\delta = (\delta_n)_{n \in \mathbb{N}}$  est définie de la façon suivante: pour tout  $n$ ,  $\delta_n$  est une probabilité de transition de  $\mathcal{H}^{n+1}$  dans  $\mathcal{A}$ . On note:  $(x^{(n)}; \delta) \longrightarrow \delta_n(x^{(n)}; \delta)$ . Seules seront utilisées les stratégies présentant certaines conditions de cohérence que l'on précisera plus loin. On notera  $\mathcal{D}$  l'ensemble des stratégies cohérentes.  $(\theta, \underline{\theta})$  est l'espace des paramètres. On se donne une probabilité de transition  $\pi$  de  $(\theta \times \mathcal{H} \times \mathcal{A}, \underline{\theta} \times \underline{\mathcal{H}} \times \underline{\mathcal{A}})$  dans  $(\mathcal{H}, \underline{\mathcal{H}})$  définie de la manière suivante: à l'instant  $n$ , si  $X_n = x$ ,  $A_n = a$ , alors  $X_{n+1}$  est dans  $\Gamma$  de  $\underline{\mathcal{H}}$ , avec la probabilité  $\pi(\theta, x, a; \delta)$ . Nous faisons les hypothèses suivantes:

Hypothèse 1

Pour toute stratégie  $\delta = (\delta_n)_{n \in \mathbb{N}}$ , pour tout  $x$  de  $\mathcal{H}$ , pour toute valeur  $\theta$ , il existe une probabilité  $P_{\theta, x}^\delta$  sur  $(\Omega, \mathcal{A})$  unique telle que pour tout  $n$  de  $\mathbb{N}$ , pour tout  $x$  de  $\mathcal{H}$ , pour tout  $\Gamma$  de  $\underline{\mathcal{H}}$ , on ait:

$$P_{\theta, x}^\delta(X_0 = x) = 1$$

et pour  $B \in \underline{\mathcal{H}}$ , pour  $C \in \underline{\mathcal{A}}$ ,

$$P_\theta^\delta(X \in B, A \in C) = \int \pi(\theta, X_n, a, du) 1_B(u) 1_C(a) \delta(X^{(n)}; da)$$

Le processus  $(\Omega, \mathcal{A}, P_x^\delta, (X_n)_{n \in \mathbb{N}})$  ainsi construit est une "chaîne de Markov contrôlée".

Hypothèse 2

A tout  $x$  de  $\mathcal{H}$ , on associe un ensemble  $D(x)$  non vide tel que l'ensemble  $S = \{(x,a) ; a \in D(x)\}$  soit dans  $\underline{\mathcal{H}} \otimes \underline{\mathcal{A}}$ . Une stratégie aléatoire cohérente sera une stratégie aléatoire  $\delta = (\delta_n)_{n \in \mathbb{N}}$  telle que pour tout  $n$ ,  $\delta_n(X^{(n)} ; D(X_n)) = 1$ .

Notations : . Pour toute v.a.  $u$  sur  $(\mathcal{H} \times \mathcal{A} \times \mathcal{H}, \underline{\mathcal{H}} \otimes \underline{\mathcal{A}} \otimes \underline{\mathcal{H}})$ , on définit la v.a.  $\pi u$  sur  $(S, \underline{S})$  par :

$$\pi u(\theta, x, a) = \int u(x, a, y) \pi(\theta, x, a, dy) ,$$

si cette intégrale a un sens .

. Pour toute  $u$  sur  $(\mathcal{H}, \underline{\mathcal{H}})$ , on définit la v.a.  $\pi u$  sur  $(S, \underline{S})$  par :

$$\pi u(\theta, x, a) = \int u(y) \pi(\theta, x, a ; dy) ,$$

toujours si cette intégrale a un sens.

Une partie importante de  $\mathcal{D}$  est le sous ensemble  $\mathcal{A}$  des stratégies aléatoires markoviennes stationnaires cohérentes : si  $s$  est une probabilité de transition de  $\mathcal{H}$  dans  $\mathcal{A}$  telle que pour tout  $x$  de  $\mathcal{H}$ ,  $s(x, D(x)) = 1$ , la stratégie définie par  $\delta_n(x^{(n)}, \cdot) = s(x_n, \cdot)$  est élément de  $\mathcal{A}$ , noté  $s$ .

Notations : Pour tout  $s$  de  $\mathcal{A}$ , pour tout  $x$  de  $\mathcal{H}$ ,  $\pi_\theta^s$  désignera la transition  $(\theta, x, C) \longrightarrow \pi(\theta, x, s(x) ; C)$  et on notera parfois par  $\pi_\theta^s$  la chaîne de Markov contrôlée par  $s$ .

Pour  $C \in \underline{\mathcal{A}}$ , pour  $x \in \mathcal{H}$ ,

$$\pi(\theta, x, s(x) ; C) = \int \pi(\theta, x, a ; C) s(x ; da) .$$

Pour toute  $\alpha$  probabilité sur  $\mathcal{H}$ ,  $\alpha \times \pi^s$  désignera le

probabilité :

$$\int_{\Gamma} l(x,y) \alpha(dx) \pi(\theta,x,s(x) ; dy) = \int_{\Gamma} l_{\Gamma} d(\alpha \times \pi^S) .$$

Hypothèse 3

$\mathcal{H}$ ,  $\mathcal{A}$  et  $\mathcal{O}$  sont métriques compacts et  $S$  est fermé.

Le noyau  $\pi$  est fellerien, c'est à dire que pour toute v.a.  $u$  bornée, continue sur  $(S \times \mathcal{H}, \underline{S} \times \underline{\mathcal{H}})$ ,  $\pi u$  est continue.

Remarque : Sauf précision contraire, les espaces topologiques seront toujours munis de leur tribu borélienne et les espaces produits de la tribu produit.

On note  $\mathcal{P}$  l'ensemble des probabilités sur  $\mathcal{H}$ ,  $\mathcal{P}_3$  celui des probabilités sur  $\mathcal{H} \times \mathcal{A} \times \mathcal{H}$ , concentrées sur  $S$ ; ils sont munis de la topologie de la convergence étroite. Pour  $\lambda$  de  $\mathcal{P}_3$ , on désigne par  $\lambda_1$  et  $\lambda_3$  ses premières et troisièmes marginales. Pour tout  $n$  de  $\mathbb{N}$ ,  $L_n$  est la fonction de répartition empirique : pour tout  $\omega$  de  $\Omega$ , pour tout  $\Gamma$  de  $\underline{S} \times \underline{\mathcal{H}}$ ,

$$L_n(\omega, \delta) = \frac{1}{n} \sum_{k=0}^{n-1} l_{\Gamma}(X_k, A_k, X_{k+1})(\omega) .$$

I - Estimation

1. Définitions - Exemples

1.1. Fonction de contraste

Une v.a.  $f$  sur  $\mathcal{O} \times S \times \mathcal{H}$  est une fonction de contraste si pour tout  $(x,a)$  de  $S$ ,

$$\int f(\theta,x,a,y) \pi(\theta,x,a ; dy)$$

est fini et 
$$K(\theta,\phi,x,a) = \int (f(\phi,x,a,y) - f(\theta,x,a,y)) \pi(\theta,x,a ; dy)$$

est positif et nul si et seulement si  $\phi$  vaut  $\theta$ .

Exemple 1

On suppose qu'il existe une v.a.  $p$  continue sur  $\Theta \times S \times \mathcal{H}$  strictement positive et une transition  $\lambda$  de  $\mathcal{H}$  dans  $\mathcal{H}$ , telles que pour tout  $\Gamma$  de  $\mathcal{H}$ , on ait :

$$\Pi(\theta, x, a ; \Gamma) = \int p(\theta, x, a, y) \lambda(x, dy) 1_{\Gamma}(y) .$$

On suppose aussi que pour tout  $(x, a)$  de  $S$ ,  $\Pi(\phi, x, a ; \cdot)$  est différent de  $\Pi(\theta, x, a ; \cdot)$  dès que  $\phi$  est différent de  $\theta$ .

Sous ces hypothèses, la v.a.  $f : f(\alpha, x, a, y) = - \text{Log } p(\alpha, x, a, y)$  est une fonction de contraste, dès que les conditions d'intégrabilité sont vérifiées. De plus, la v.a.  $p$  étant strictement positive, pour tout  $(x, a)$  de  $S$ , l'ensemble

$$\{\Pi(\phi, x, a ; \cdot)\}_{\phi \in \Theta}$$

est un ensemble de probabilités équivalentes. Aussi, si  $p$  est bornée, on a :

$$\begin{aligned} K(\theta, \phi, x, a) &= \int \text{Log } \frac{p(\theta, x, a, y)}{p(\phi, x, a, y)} \Pi(\theta, x, a ; dy) \\ &= I_{\Pi(\phi, x, a ; \cdot)}(\Pi(\theta, x, a ; \cdot)) \end{aligned}$$

où  $I_{\Pi(\phi, x, a ; \cdot)}(\Pi(\theta, x, a ; \cdot))$  est l'information de Kullback de  $\Pi(\theta, x, a ; \cdot)$  par rapport à  $\Pi(\phi, x, a ; \cdot)$ . On la note  $I(\theta, \phi, x, a)$ . Elle est strictement positive dès que  $\phi$  est différent de  $\theta$ .

Exemple 2

Soit  $g$  une v.a. sur  $\mathcal{H} \times \mathcal{H}$ , bornée. On définit :

$$m(\alpha, x, a) = \int g(x, y) \Pi(\alpha, x, a ; dy)$$

et on suppose que pour tout  $(x, a)$  de  $S$ ,  $m(\alpha, x, a)$  est différent de  $m(\alpha', x, a)$  dès que  $\alpha$  est différent de  $\alpha'$ . La v.a.  $f$  définie par :

$$f(\alpha, x, a, y) = (g(x, y) - m(\alpha, x, a))^2$$

est une fonction de contraste.

## 1.2. Estimateur du minimum de contraste

Etant donnée  $f$  une fonction de contraste, on pose :

$$L_n(\alpha) = \sum_{k=0}^{n-1} f(\alpha, X_k, A_k, X_{k+1}) .$$

A l'instant  $n$ , un estimateur  $\hat{\theta}_n$  est un estimateur du minimum de contraste si :

$$L_n(\hat{\theta}_n) = \inf_{\alpha \in \Theta} L_n(\alpha) .$$

Dans l'exemple 1, cet estimateur est l'estimateur du maximum de vraisemblance.

Le théorème de sélection de Brown assure l'existence d'une suite d'estimateurs du minimum de contraste si  $\Theta$  est un espace métrique compact, et si l'application  $\alpha \longrightarrow f(\alpha, x, a, y)$  est s.c.i de  $\Theta$  dans  $\mathbb{R}$ , pour tout  $(x, a, y)$  de  $S \times \mathcal{H}$ .

On posera :

$$h(\alpha, \phi, x, a, y) = f(\phi, x, a, y) - f(\alpha, x, a, y)$$

et

$$S_n(\alpha, \phi) = \sum_{k=0}^{n-1} h(\alpha, \phi, X_k, A_k, X_{k+1}) = \sum_{k=0}^{n-1} h_k(\alpha, \phi) .$$

## 2. Consistance exponentielle

Si  $\lambda_0$  est la distribution initiale de la chaîne, une suite  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  d'estimateurs de  $\theta$  est fortement consistante si pour tout voisinage  $U(\theta)$  de  $\theta$ , pour toute stratégie cohérente  $\delta$ , la suite  $(\hat{\theta}_n)$  tend vers  $\theta$   $P_{\theta, x}^\delta$  p.s.

Le but est d'établir ici la consistance exponentielle de toute suite  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  d'estimateurs du minimum de contraste lorsque la fonction de contraste  $f$  est continue sur  $\Theta \times S \times \mathcal{H}$ . (Le théorème de Brown assure l'existence d'une suite d'estimateurs du minimum de contraste lorsque  $f$  est continue. On améliore ainsi le résultat de Georgin [2] qui a montré sous les mêmes hypothèses, la consistance forte.

Pour tout voisinage  $U(\theta)$  de  $\theta$ , on définit la v.a. :

$$\Gamma(U(\theta)) = \inf\{n ; p > n, \hat{\theta}_p \in U(\theta)\}$$



Lorsque  $\Theta$  sera fini, on notera  $T_\theta$  la v.a.  $T(\{\theta\})$ .

Théorème 1

Lorsque la fonction de contraste  $f$  est continue sur  $\Theta \times S \times \mathcal{H}$  toute suite  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  d'estimateurs du minimum de contraste est exponentiellement consistante : pour tout  $U(\theta)$  voisinage ouvert de  $\theta$ , on a :

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} \sup_{\delta \in \mathcal{D}} P_{\theta, x}^\delta (P(U(\theta)) > n) < 0 .$$

Démonstration

Soit  $U(\theta)$  un voisinage ouvert de  $\theta$ . L'ensemble  $\{T(U(\theta)) > n\}$  est inclus dans :

$$\bigcup_{m > n} \left\{ \inf_{\phi \in U^c(\theta)} \sum_{k=0}^{m-1} h_k(\theta, \phi) < 0 \right\}$$

et pour tout  $x \in \mathcal{H}$  et  $\delta \in \mathcal{D}$ ,

$$P_{\theta, x}^\delta (T(U(\theta)) > n) \leq \sum_{m > n} P_{\theta, x}^\delta \left( \inf_{\phi \in U^c(\theta)} \sum_{k=0}^{m-1} h_k(\theta, \phi) \leq 0 \right)$$

Soit  $\Gamma = \left\{ \mu \in \mathcal{P}_3, \inf_{\phi \in U^c(\theta)} \int h(\theta, \phi, x, a, y) d\mu(x, a, y) \leq 0 \right\}$ . On a :

$$P_{\theta, x}^\delta (T(U(\theta)) > n) \leq \sum_{m > n} P_{\theta, x}^\delta (L_m \in \Gamma) .$$

Montrons que  $\Gamma$  est un fermé de  $\mathcal{P}_3$ .

Soit  $\mu$  de  $\overline{\Gamma}$ . Il existe alors une suite  $(\mu_n)_{n \in \mathbb{N}}$  de  $\Gamma$  convergeant vers  $\mu$  dans  $\mathcal{P}_3$ . Pour tout  $\lambda$  de  $\mathcal{P}_3$ , l'application :

$$\phi \longrightarrow \int h(\theta, \phi, x, a, y) d\lambda(x, a, y)$$

est continue de  $\Theta$  dans  $\mathbb{R}$ . On définit alors par  $(\phi_n)_{n \in \mathbb{N}}$  une suite de  $U^c(\theta)$  qui vérifie :

$$\mu_n(\phi_n) = \inf_{\phi \in U^c(\theta)} \left( \int h(\theta, \phi, x, a, y) d\mu_n(x, a, y) \right) .$$

Pour tout  $n$ ,  $\mu_n(\phi_n)$  est négatif. On considère maintenant  $(\phi_{n_k})_{k \in \mathbb{N}}$  une sous

suite de  $(\phi_n)_{n \in \mathbb{N}}$ , qui converge vers  $\phi_0$  de  $U^C(\theta)$ . La sous suite  $h(\theta, \phi_{n_k}, x, a, y)$  converge donc vers  $h(\theta, \phi_0, x, a, y)$  uniformément sur  $S \times \mathcal{H}$ . Puisque  $(\mu_{n_k})_{k \in \mathbb{N}}$  tend vers dans  $\mathcal{P}_3$ ,  $(\mu_{n_k}(\phi_0))_{k \in \mathbb{N}}$  tend vers  $\mu(\phi_0)$  et  $(\mu_{n_k}(\phi_{n_k}))_{k \in \mathbb{N}}$  tend aussi vers  $\mu(\phi_0)$ , ce qui permet de conclure que  $\mu(\phi_0)$  est négatif et  $\mu$  appartient à  $\Gamma$  :  $\Gamma$  est un fermé de  $\mathcal{P}_3$ .

Puisque  $f$  est une fonction de contraste,

$$\int h(\theta, \phi, x, a, y) \Pi(\theta, x, a ; dy) d\lambda_1(x) s(x, da)$$

est strictement positif pour toute  $\lambda_1$  de  $\mathcal{P}$  et pour toute  $s$  de  $\mathcal{A}$ , dès que  $\phi$  est différent de  $\theta$ . Comme  $h$  est continue sur le compact  $\Theta \times \Theta \times S \times \mathcal{H}$ , l'application :

$$\phi \longrightarrow \int h(\theta, \phi, x, a, y) \Pi(\theta, x, a ; dy) d\lambda_1(x) s(x, da)$$

est continue de  $\Theta$  dans  $\mathbb{R}$ . Aussi dès que  $\phi$  est différent de  $\theta$ ,

$$\inf_{\phi \in U^C(\theta)} \int h(\theta, \phi, x, a, y) \Pi(\theta, x, a ; dy) d\lambda_1(x) s(x, da)$$

est strictement positif ; c'est à dire que  $\lambda_1 \otimes s \otimes \Pi$  n'appartient pas à  $\Gamma$ , quelles que soient  $\lambda_1$  de  $\mathcal{P}$  et  $s$  de  $\mathcal{A}$ .

Le théorème des grandes déviations [4] entraîne :

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_x \sup_{\delta} P_{\theta, x}^{\delta} (T(U(\theta)) > n) \leq -I(\Gamma \cap \Delta) < 0.$$

## II - $\Gamma$ est séquentiel du rapport de vraisemblance

On se pose ici le même problème de test que dans [3].

L'espace des paramètres est supposé fini (de cardinal  $q$ ),  $(\Theta_1, \Theta_2)$  est une partition de  $\Theta$  et on veut savoir à quel élément de la partition appartient la vraie valeur du paramètre. Dans ce cadre contrôlé, il faudra en plus des règles d'arrêt et de décision, s'intéresser au choix d'un bon contrôle. Chaque expérience a un coût  $c$ ,  $0 < c < 1$ , et on note  $r(c)$  la perte due à l'erreur de décision.

On se place dans le cadre de l'exemple 1, et on notera :

$$J(\theta, \phi) = \inf_{(x, a) \in S} I(\theta, \phi, x, a), \quad J(\theta) = \inf_{\phi \in a(\theta)} J(\theta, \phi) > 0$$

$$L(\theta, \phi) = \sup_{(x, a) \in S} I(\theta, \phi, x, a), \quad L(\theta) = \inf_{\phi \in a(\theta)} L(\theta, \phi).$$

$(\hat{\theta}_n)_{n \in \mathbb{N}}$  désigne désormais une suite d'estimateurs du maximum de vraisemblance, elle est exponentielle consistante.

Comme dans la partie sans contrôle, on note  $h(\theta) = \theta_i$  si  $\theta$  est élément de  $\Theta_i$  ( $i = 1$  ou  $2$ ), et  $a(\theta) = \Theta - h(\theta)$ , et on désigne par  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  une suite d'estimateurs de  $\theta$  qui vérifient pour tout  $n$  :

$$\text{Log} \prod_{k=0}^{n-1} p(\hat{\theta}_n, X_k, A_k, X_{k+1}) = \sup_{\alpha \in a(\hat{\theta}_n)} \text{Log} \prod_{k=0}^{n-1} p(\alpha, X_k, A_k, X_{k+1})$$

Les résultats obtenus sont des résultats asymptotiques.

Sans hypothèses supplémentaires, nous construisons un test séquentiel clos quelle que soit la stratégie ; nous obtenons une majoration (asymptotique) du risque, indépendante de la stratégie. Aucune stratégie n'est particularisée. Avec une hypothèse supplémentaire de récurrence, pour le test de 2 hypothèses simples, nous construisons une procédure (règles d'arrêt et de décision, stratégie) possédant une bonne propriété d'optimalité.

## 1. Etude d'un temps d'arrêt

### 1.1. Préliminaires

#### Lemme 1

$\mathcal{Y}$  étant un espace métrique compact et  $q$  une transition de  $\mathcal{Y}$  dans  $\mathcal{Y}$  telle que l'application  $x \longrightarrow q(x, \cdot)$  soit continue de  $\mathcal{Y}$  dans  $\mathcal{P}(\mathcal{Y})$  ensemble des probabilités sur  $\mathcal{Y}$  muni de la topologie de la convergence étroite ; alors pour toute  $g$  v.a. sur  $\mathcal{Y} \times \mathcal{Y}$  continue, l'application

$$x \longrightarrow q \int g(x) = \int g(x, y) q(x, dy)$$

est continue de  $\mathcal{Y}$  dans  $\mathbb{R}$ .

Démonstration : Facile

Lemme 2

Soit  $(Y_n)_{n \in \mathbb{N}}$  une suite de v.a. définies sur un espace  $(B, \underline{B}, \lambda)$ , intégrables adaptée à une famille croissante  $(\sigma_n)_{n \in \mathbb{N}}$  de sous tribus de  $\underline{B}$  vérifiant : il existe un entier positif tel que pour tout  $n$ , on ait :

$$E(Y_{n \wedge k} / \sigma_{n-1}) \geq m > 0.$$

On définit pour  $b$  réel, la v.a.  $v'_b$  par  $\inf\{n ; y_1 + \dots + y_n \geq b\}$ .

On a : 
$$E(v'_b) \leq \frac{b+k}{m}.$$

Démonstration

Il suffit de montrer le lemme pour des v.a. majorées par  $k$ .

On définit :

$$V_n = Y_1 + \dots + Y_n$$

$$A_n = \sum_{i=1}^n E(Y_i / \sigma_{i-1})$$

$$M_0 = 0 \quad \text{et} \quad M_n = V_n - A_n$$

Puisque  $Y_n$  est intégrable,  $V_n$  et  $M_n$  sont intégrables, et

$$E(M_n / \sigma_{n-1}) = M_{n-1}. \quad (M_n)_{n \geq 0} \text{ est donc une martingale adaptée à la famille } (\sigma_n)_{n \geq 0}.$$

Pour tout  $n$  entier naturel, on a d'après le théorème d'arrêt :

$$E(M_{v'_b \wedge n}) = 0 = E(V_{v'_b \wedge n}) - E(A_{v'_b \wedge n})$$

Comme : 
$$E(A_{v'_b \wedge n}) \geq m E(v'_b \wedge n)$$

et : 
$$\begin{aligned} E(V_{n \wedge v'_b}) &= E(V_n \mathbb{1}_{\{n < v'_b\}}) + E((V_{v'_b-1} + Y_{v'_b}) \mathbb{1}_{\{n \geq v'_b\}}) \\ &\leq b \lambda(n < v'_b) + (b+k) \lambda(n \geq v'_b) \leq (b+k) \end{aligned}$$

On a : 
$$E(v'_b \wedge n) \leq \frac{b+k}{m}$$

$$E(v'_b) = \lim_{n \rightarrow \infty} E(v'_b \wedge n) \leq \frac{b+k}{m}.$$

1.2. Définition d'un temps d'arrêt. Propriété de finitude

Soit  $b$  réel positif et  $\phi$  de  $\Theta$ . On définit :

$$\tau_b = \inf\{n ; S_n(\hat{\theta}_n, \hat{\theta}_n) > b\}$$

$$\tau_{b,\phi} = \inf\{n ; \forall p \geq n, S_p(\theta, \phi) > b\}$$

On a :  $\tau_b \leq \max(T_\theta, \max_{\phi \in a(\theta)} \tau_{b,\phi})$  (1.2)

o Une application directe du lemme 2 permet d'écrire :

Lemme 3

Soit  $\phi$  différent de  $\theta$ . S'il existe un entier naturel  $k_0$  et un réel  $m$ , strictement positifs tels qu'on ait :

$$\inf_{(x,a) \in S} \psi_{k_0}(\theta, \phi, x, a) = \inf_{(x,a) \in S} \int \inf(k_0, \text{Log} \frac{p(\theta, x, a, y)}{p(\phi, x, a, y)}) \Pi(\theta, x, a ; dy) \geq m > 0$$

alors :  $\sup_{x \in \mathcal{D}} \sup_{\delta \in \mathcal{D}} E_{\theta, x}^\delta(\tau_{b,\phi}) \leq \frac{b+k_0}{m}$ .

Montrons que les hypothèses du lemme 3 sont vérifiées.

Le théorème de convergence de Béppe-Lévi entraîne :

$$\lim_k \uparrow \psi_k(\theta, \phi, x, a) = \text{Log} \int \frac{p(\theta, x, a, y)}{p(\phi, x, a, y)} \Pi(\theta, x, a ; dy) = I(\theta, \phi, x, a)$$

Puisque  $p$  est continue et puisque le noyau  $\Pi$  est fellerien, le lemme 1, entraîne que  $I(\theta, \phi, \cdot, \cdot)$  et  $\psi_k(\theta, \phi, \cdot, \cdot)$  sont continues sur le compact  $S$ . Ce qui permet de conclure, grâce au théorème de Dini que la suite  $(\psi_k(\theta, \phi, \cdot, \cdot))_{k \in \mathbb{N}}$  converge uniformément vers  $I(\theta, \phi, \cdot, \cdot)$  sur  $S$ . Donc, pour tout  $\phi$  différent de  $\theta$ , il existe un entier  $k_0$ , tel que pour tout  $k \geq k_0$ ,  $\inf_{(x,a) \in S} \psi_k(\theta, \phi, x, a)$  soit strictement positif.

Grâce à la relation (1.2) et à la consistance exponentielle, on en déduit que les temps d'arrêt  $\tau_b$  sont  $P_{\theta, x}^\delta$  presque sûrement finis pour tout  $x$  et pour toute  $\delta$  de  $\mathcal{D}$ .

1.3. Définition d'une procédure

On note  $v_c$  le temps d'arrêt  $\tau_{-\text{Log } c}$  et  $v_{c,\phi}$  le temps d'arrêt  $\tau_{-\text{Log } c, \phi}$ .  
 Pour une stratégie  $\delta$  de  $\mathcal{D}$ , on désigne par  $\phi_c^\delta$  la procédure suivante :

- sa règle d'arrêt est  $v_c$
- si  $n = v_c$ , on arrête et on décide  $h(\hat{\theta}_n)$
- si  $n < v_c$ , on continue, on contrôle la chaîne avec la stratégie  $\delta$  et on refait une expérience.

Les procédures  $\phi_c^\delta$  sont closes pour toute valeur du coût  $c$  et toute stratégie  $\delta$ .  $\beta_c^\delta(\theta)$  et  $R_c^\delta(\theta)$  désigneront respectivement la probabilité d'erreur et le risque de la procédure  $\phi_c^\delta$  lorsque  $\theta$  est le vrai paramètre.

Toute autre procédure séquentielle de temps d'arrêt  $T$ , de règle de décision  $D$  et de stratégie  $\delta$  sera notée  $(T, D, \delta)$ .

1.4. Propriétés

Théorème 2

Pour tout  $\theta$  de  $\Theta$ , on a :

$$\sup_{\delta \in \mathcal{D}} \beta_c^\delta(\theta) \leq qc.$$

Démonstration

Elle est analogue à celle faite dans [3] (théorème 3.a) ; il suffit de remarquer que pour tout  $\delta$  de  $\mathcal{D}$ ,

$$P_\theta^\delta = \prod_{i=0}^{n-1} \frac{p(\theta, X_i, A_i, X_{i+1})}{p(\phi, X_i, A_i, X_{i+1})} P_\phi^\delta \text{ sur } \mathcal{B}_n.$$

Théorème 3

Pour tout  $\theta$  de  $\Theta$ , on a :

$$\sup_{\delta \in \mathcal{D}} E_\theta^\delta(v_c) \leq \frac{-(1+\sigma(1))\text{Log } c}{J(\theta)}.$$

Démonstration

Soit  $\varepsilon > 0$  et  $N_c = \frac{-(1 + \frac{\varepsilon}{2}) \text{Log } c}{J(\theta)}$ .

Il suffit de prouver qu'il existe  $k$  et  $b$  strictement positifs, tels que pour tout  $n > N_c$ , on ait :

$$\sup_{\delta \in \mathcal{D}} P_{\theta}^{\delta}(v_c = n) \leq k e^{-bn}$$

Le reste de la démonstration est analogue à celle du théorème 3.b de [3].

La relation :  $\{v_c = n\} \subset \{T_{\theta} > n\} \cup \{ \bigcup_{\phi \in \alpha(\theta)} \{v_{c,\phi} = n\} \}$  entraîne pour tout  $n$  de  $N$  :

$$\sup_{\delta \in \mathcal{D}} P_{\theta}^{\delta}(v_c = n) \leq \sup_{\delta \in \mathcal{D}} P_{\theta}^{\delta}(T_{\theta} \geq n) + \sum_{\phi \in \alpha(\theta)} \sup_{\delta \in \mathcal{D}} P_{\theta}^{\delta}(v_{c,\phi} = n)$$

La suite  $(\hat{\theta}_n)_{n \in N}$  étant exponentiellement consistante pour toute stratégie, il existe  $k_1$  et  $b_1$  strictement positifs tels que pour tout  $n$  de  $N$ , on ait :

$$\sup_{\delta \in \mathcal{D}} P_{\theta}^{\delta}(T_{\theta} \geq n) \leq k_1 e^{-b_1 n}$$

D'autre part, d'après le corollaire 1.3. du théorème des grandes déviations [4], il existe  $k_2$  et  $b_2$  strictement positifs, tels que pour tout  $\phi$  de  $\Theta$ , pour tout  $n$  de  $N$ , on ait :

$$\sup_{\delta \in \mathcal{D}} P_{\theta}^{\delta}(S_n(\theta, \phi) \leq \frac{n J(\theta)}{1 + \frac{\varepsilon}{2}}) \leq k_2 e^{-b_2 n}$$

c'est à dire :

$$\sup_{\delta \in \mathcal{D}} P_{\theta}^{\delta}(S_n(\theta, \phi) \leq -\text{Log } c) \leq k_2 e^{-b_2 n},$$

dès que  $n$  est supérieur à  $N_c$ .

Comme :  $v_{c,\phi}$  vaut  $\inf\{n ; \forall p \geq n, S_p(\theta, \phi) > -\text{Log } c\}$ , on a pour toute  $\delta$  de  $\mathcal{D}$  :

$$P_{\theta}^{\delta}(v_{c,\phi} = n) \leq P_{\theta}^{\delta}(S_{n-1}(\theta, \phi) \leq -\text{Log } c)$$

c'est à dire, il existe  $k'_2$  et  $b'_2$  strictement positifs, tels que pour tout  $\phi$  de  $\Theta$ , pour tout  $n > N_n$ , on ait :

$$\sup_{\delta \in \mathfrak{D}} P_{\theta}^{\delta}(v_{c, \phi} = n) \leq k_2' e^{-b_2' n}$$

Ce qui permet de conclure.

Corollaire 1

Pour tout  $\theta$  de  $\Theta$ ,

$$\sup_{\delta \in \mathfrak{D}} R_c^{\delta}(\theta) \leq \frac{-(1 + \sigma(1)) c \text{Log } c}{J(\theta)}$$

Les lemmes suivantes seront utilisés pour demander une propriété d'optimalité.

Lemme 4

Soit  $b$ ,  $0 < b < 1$ . Soit  $(T, D, \delta)$  une procédure séquentielle,  $e(\theta)$  la probabilité d'erreur associée. Si pour tout  $\theta$ ,  $e(\theta)$  est un  $\mathcal{O}(-b \text{Log } b)$ ; on a pour tout  $\epsilon \in ]0, 1[$  et  $(\theta, \phi) \in \Theta^2$ ,

$$P_{\theta}^{\delta}(S_T(\theta, \phi) < -(1 - \epsilon) \text{Log } b) = \mathcal{O}(b^{\epsilon} \text{Log } b)$$

Démonstration

En se rappelant que :

$$P_{\phi}^{\delta} = \prod_{k=0}^{n-1} \frac{p(\phi, X_k, A_k, X_{k+1})}{p(\theta, X_k, A_k, X_{k+1})} P_{\theta}^{\delta} \text{ sur } \mathcal{D}_n$$

la démonstration se fait de la même façon que celle du lemme 3.a de [3].

Lemme 5

Pour tout  $\epsilon > 0$ , pour toute stratégie  $\delta$  de  $\mathfrak{D}$ ,

$$P_{\theta}^{\delta}(\max_{1 \leq m \leq n} \min_{\phi \in \mathfrak{a}(\theta)} S_m(\theta, \phi) \geq n(L(\theta) + \epsilon)) \xrightarrow{n \rightarrow \infty} 0$$

Démonstration

Soit  $\delta$  une stratégie cohérente, et  $\epsilon$  strictement positif.

Pour tout  $m$  de  $N$ , pour tout  $\phi$ ,



$$S_m(\theta, \phi) = (S_m(\theta, \phi) - \sum_{k=0}^{m-1} I(\theta, \phi, X_k, A_k)) + \sum_{k=0}^{m-1} I(\theta, \phi, X_k, A_k)$$

$$S_m(\theta, \phi) = Z_{m,1}(\theta, \phi) + Z_{m,2}(\theta, \phi)$$

$$Z_{m,2}(\theta, \phi) \leq m L(\theta, \phi) P_\theta^\delta \text{ ps}$$

$$\min_{\phi \in a(\theta)} Z_{m,2}(\theta, \phi) \leq m L(\theta) , P_\theta^\delta \text{ ps}$$

Ce qui entraîne :

$$\begin{aligned} P_\theta^\delta (\max_{1 \leq m \leq n} \min_{\phi \in a(\theta)} S_m(\theta, \phi) \geq n(L(\theta) + \epsilon)) &\leq P_\theta^\delta (\max_{1 \leq m \leq n} \min_{\phi \in a(\theta)} Z_{m,1}(\theta, \phi) > n \epsilon) \\ &\leq \sum_{\phi \in a(\theta)} P_\theta^\delta (\max_{1 \leq m \leq n} Z_{m,1}(\theta, \phi) > n \epsilon) \end{aligned}$$

$\{Z_{m,1}(\theta, \phi)\}$  est une martingale adaptée aux tribus  $(\mathcal{D}_n)_{n \in \mathbb{N}}$  sur  $(\Omega, \mathcal{A}, P_\theta^\delta)$ .

D'après l'inégalité de Doob, on a :

$$P_\theta^\delta (\max_{1 \leq m \leq n} \min_{\phi \in a(\theta)} S_m(\theta, \phi) \geq n(L(\theta) + \epsilon)) \leq \sum_{\phi \in a(\theta)} \frac{E_\theta^\delta(Z_{n,1}^2)}{n^2 \epsilon^2}$$

et on vérifie facilement que pour tout  $n \in \mathbb{N}$ ,  $\frac{1}{n} E_\theta^\delta(Z_{n,1}^2)$  est majorée par une constante finie  $M$ .

$$P_\theta^\delta (\max_{1 \leq m \leq n} \min_{\phi \in a(\theta)} S_m(\theta, \phi) \geq n(L(\theta) + \epsilon) \leq \frac{qM}{n\epsilon} .$$

### Lemme 6

Soit  $b$ ,  $0 < b < 1$ . Soit  $(T, D, \delta)$  une procédure séquentielle et  $e(\theta)$  la probabilité d'erreur associée lorsque  $\theta$  est le paramètre. On suppose que pour tout  $\theta$ ,  $e(\theta)$  est un  $\mathcal{O}(-b \text{Log } b)$ . On a :

$$E_\theta^\delta(T) \geq \frac{-(1 + \sigma(1)) \text{Log } b}{L(\theta)} .$$

### Démonstration

On suit la démonstration du lemme 3.c de [3] en utilisant les 2 lemmes ci-dessus.

Conséquence

Une procédure  $(T, D, \delta)$  telle que pour tout  $\theta$ , la probabilité d'erreur  $e(\theta)$  soit un  $\mathcal{O}(-c \text{ Log } c)$ , vérifie : pour tout  $\theta$ ,

$$E_{\theta}^{\delta}(T) \geq \frac{-(1 + \sigma(1)) \text{ Log } c}{L(\theta)} .$$

En se rappelant que l'erreur de  $\phi_c^{\delta}$  est un  $\mathcal{O}(c)$ , on obtient :

Théorème 4

Toute procédure de force au moins égale aux procédures  $\{\phi_c^{\delta}\}_{\delta \in \mathcal{D}}$  dure en moyenne plus longtemps que  $\frac{-(1+\sigma(1))\text{Log } c}{L(\theta)}$  qui est du même ordre de grandeur que  $\frac{-(1 + \sigma(1))\text{Log } c}{J(\theta)}$  qui majore la durée moyenne des procédures  $\{\phi_c^{\delta}\}_{\delta \in \mathcal{D}}$ .

III - Test de deux hypothèses simples pour des chaînes contrôlées "récurrentes"

Ces résultats ne particularisent pas une stratégie ; valables pour toute stratégie, ils ne sont pas très précis. On n'obtient pas de stratégie meilleure que les autres. En ajoutant une hypothèse de récurrence, nous construisons une procédure ayant une bonne propriété d'optimalité pour le test de 2 hypothèses simples : le temps d'arrêt et la règle de décision seront les mêmes que précédemment, mais nous particularisons une stratégie.

1.5. Rappels de quelques résultats

Soit  $s$  une stratégie markoviennes stationnaire, telle que la chaîne associée à  $\Pi_{\theta}^s$  soit récurrente Doeblin, apériodique, de probabilité invariante  $\mu_{\theta}^s$ . L'information de Kullback de  $\Pi(\theta, x, s(x) ; \cdot)$  par rapport à  $\Pi(\phi, x, s(x) ; \cdot)$  est :

$$I_s(\theta, \phi, x) = \int \text{Log } \frac{p(\theta, x, s(x) ; y)}{p(\phi, x, s(x) ; y)} \Pi(\theta, x, s(x) ; dy)$$

On définit :

$$I_s(\theta, \phi) = \int I_s(\theta, \phi, x) d\mu_{\theta}^s(x)$$

et 
$$I_s(\theta) = \min_{\phi \in a(\theta)} I_s(\theta, \phi)$$

$I_s(\theta)$  est strictement positif.

Les théorèmes 3.a et 3.b [3] s'écrivent ici :

$$\beta_c^s(\theta) \leq qc \quad \text{et}$$

$$E_\theta^s(v_c) \leq \frac{-(1 + \sigma(1)) \text{Log } c}{I_s(\theta)} .$$

La stratégie  $s$  semble d'autant meilleure qu'elle est d'information maximum. Nous exploitons cette idée pour le test de 2 hypothèses simples. Donc ici ,  $\mathcal{O} = \{\theta_1, \theta_2\}$  , (on notera  $\theta$  la vraie valeur ,  $\phi_\theta$  l'autre valeur) , et on fait les hypothèses suivantes :

- (1) : Toutes les stratégies markoviennes cohérentes stationnaires sont récurrentes Doeblin, apériodiques. (Pour une telle  $s$  , on note  $\mu_\theta^s$  la probabilité invariante) .
- (2) : Il existe une stratégie markovienne stationnaire déterministe  $s^\theta$  qui permet d'obtenir le gain moyen :  $\lim_n \frac{S_n(\theta, \phi)}{n}$  , optimal.. (On dit que  $s^\theta$  est optimale). On notera  $I(\theta) = I_{s^\theta}(\theta) = \sup_{s \in \mathcal{S}} I_s(\theta)$  .

(1) est vérifié par exemple, s'il existe une mesure positive non nulle sur  $\mathcal{H}$  , telle que  $\inf_{(x,a) \in \mathcal{S}} \Pi(\theta, x, a ; \cdot) > v(\cdot)$  . Si en plus les espaces d'états  $\mathcal{H}$  et d'actions  $\mathcal{A}$  sont finis, alors (2) est vérifiée. (Ces résultats ont été montrés lors de la recherche du gain optimal sous contraintes, dans [2] .

On connaît des critères plus fins sous lesquels la condition (2) est vérifiée ([2]) .

1.6. : Désormais  $s^\theta$  désigne une telle stratégie optimale :

$$\lim_n \frac{1}{n} S_n(\theta, \phi_\theta) \xrightarrow[\text{ps}]{P_\theta^{s^\theta}} I(\theta)$$

et  $P_{\theta}^{\delta}$  ps, pour toute stratégie  $\delta$  cohérente :  $\overline{\lim}_n \frac{1}{n} S_n(\theta, \phi_{\theta}) < I(\theta)$ .

De plus, on a :

$$\underline{a} \quad \beta_c^{s_{\theta}}(\theta) \leq 2c$$

$$\underline{b} \quad E_{\theta}^{s_{\theta}}(v_c) \leq \frac{-(1 + \sigma(1)) \text{Log } c}{I(\theta)}.$$

Nous allons montrer que  $s_{\theta}$  possède une "bonne" propriété d'optimalité.

Lemme 7

Pour tout  $\epsilon > 0$ , pour toute  $\delta$  de  $\mathcal{D}$ ,

$$P_{\theta}^{\delta}(\max_{1 \leq m \leq n} S_m(\theta, \phi_{\theta}) \geq n(I(\theta) + \epsilon)) \xrightarrow{n \rightarrow \infty} 0.$$

Démonstration

Soit  $\delta$  une stratégie de  $\mathcal{D}$ ,  $\epsilon$  et  $\eta$  quelconques fixés strictement positifs. La relation :  $\lim_n \frac{1}{n} S_n(\theta, \phi_{\theta}) \leq I(\theta)$   $P_{\theta}^{\delta}$  ps entraîne qu'il existe un entier  $n_0$  vérifiant :  $P_{\theta}^{\delta}(\max_{n_0 \leq m \leq n} S_m(\theta, \phi_{\theta}) \geq n(I(\theta) + \epsilon)) \leq \eta$ .

De plus il existe  $n_1$  tel que pour tout  $n$  supérieur à  $n_1$ , on ait :

$$P_{\theta}^{\delta}(\max_{0 \leq m \leq n_0} S_m(\theta, \phi_{\theta}) \geq n(I(\theta) + \epsilon)) = 0.$$

Ce qui implique qu'il existe  $n_2(\epsilon, \eta)$ , tel que pour tout  $n$  supérieur à  $n_2$ , on ait :

$$P_{\theta}^{\delta}(\max_{0 \leq m \leq n} S_m(\theta, \phi_{\theta}) \geq n(I(\theta) + \epsilon)) \leq \eta.$$

Lemme 8

Soit  $b$ ,  $0 < b < 1$ , et soit  $(T, D, \delta)$  une procédure telle que la probabilité d'erreur associée  $e(\phi)$  soit un  $\mathcal{O}(-b \log b)$  pour tout  $\phi$ .

On a :

$$E_{\theta}(T) \geq \frac{-(1 + \sigma(1)) \text{Log } b}{I(\theta)}.$$

Démonstration

Elle découle du lemme précédent et du lemme 4. Elle est la même que celle du lemme 6 .

De ce dernier résultat, et en rappelant ceux de a) et b) , on conclut :

Théorème 5

La procédure  $\phi_c^{s^\theta}$  est la plus économique parmi les procédures de force au moins égale.

Mais la stratégie  $s^\theta$  n'est pas calculable puisqu'elle dépend du paramètre  $\theta$  . On considère la suite  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  consistante exponentiellement pour toute stratégie. L'idée est alors de prendre la stratégie  $\hat{s}$  obtenues en remplaçant  $\theta$  dans  $s^\theta$  par son estimateur :  $\hat{s} = (s_n)_{n \in \mathbb{N}}$  où pour tout  $n$  , on a :

$$s_n(X^{(n)}) = s_n^{\hat{\theta}}(X_n) .$$

Le but est de montrer que  $\hat{s}$  est aussi bonne sur  $s^\theta$  . Nous allons prouver que :

$$E_\theta^{\hat{s}}(v_c) \leq \frac{-(1 + \sigma(1)) \text{Log } c}{I(\theta)}$$

ce qui permettra de conclure, (puisque'on a aussi  $\beta^{\hat{s}}(\theta) \leq 2c$ ) .

Pour cela, nous introduisons la notion de stratégies "exponentiellement voisines".

1.7. Stratégies exponentiellement voisines

On suppose ici  $\Theta$  réduit à un point et on omet le paramètre. Les hypothèses topologique sur  $\mathcal{X}$  ,  $\mathcal{A}$  et  $\Pi$  sont inutiles.

Définition

Soient  $d$  et  $\delta$  2 stratégies. On définit la v.a. :

$$T(d, \delta) = \inf \{n ; \forall m \geq n , d_m = \delta_m\}$$

$d$  et  $\delta$  seront dites exponentiellement voisines si il existe  $k$  et  $b$

strictement positifs tels que pour tout  $n$ ,

$$P^d(T(d, \delta) > n) \leq k e^{-bn}$$

et 
$$P^\delta(T(d, \delta) > n) \leq k e^{-bn}$$

Exemple : Revenant au cadre général de cet article, pour toute stratégie  $d^\theta$ , on note  $\hat{d}$  l'estimée de  $d^\theta$  obtenue en remplaçant  $\theta$  par son estimateur, c'est à dire,  $\hat{d}$  est définie par :

$$\hat{d} = \hat{d}_n(X^{(n)}), \text{ où pour tout } n, \hat{d}_n(X^{(n)}) = d_{\hat{\theta}_n}^{\hat{\theta}_n}(X^{(n)}).$$

La consistance exponentielle de la suite  $(\hat{\theta}_n)_{n \in \mathbb{N}}$  pour toute stratégie entraîne que  $d^\theta$  et  $\hat{d}$  sont exponentiellement voisines.

Théorème 6

Soit  $s$  une stratégie markovienne stationnaire récurrente Doeblin, apériodique, de probabilité invariante  $\mu^s$ . Soit  $\delta$  une stratégie de  $\mathcal{D}$  telle que  $\delta$  et  $s$  soient exponentiellement voisines, et soit  $g$  une v.a. bornée sur  $S \times \mathcal{X}$ , non nulle. Pour tout  $\varepsilon > 0$ , on a :

$$\overline{\lim}_n \frac{1}{n} \text{Log} \sup_{x \in \mathcal{X}} P_x^\delta \left( \frac{1}{n} \sum_{p=0}^{n-1} g(X_p, A_p, X_{p+1}) \leq \left( \int g d \mu^s \otimes s \otimes \mathbb{1} \right) - \varepsilon \right) < 0$$

Démonstration

$s$  étant récurrente Doeblin apériodique, il existe  $a > 0$  et  $\rho, 0 < \rho < 1$ , tels que pour tout  $n$  de  $\mathbb{N}$ , on ait :

$$\sup_{x \in \mathcal{X}} \left\| \Pi_n(x, s(x), \cdot) - \mu^s(\cdot) \right\| \leq a \rho^n.$$

$\delta$  et  $s$  étant exponentiellement voisines de  $s$ , il existe  $b$  et  $k$  strictement positifs tels que pour tout  $n$  de  $\mathbb{N}$ , on ait :

$$P_\theta^s(T(s, \delta) > n) \leq k e^{-bn}$$

et 
$$P_\theta^\delta(T(s, \delta) > n) \leq k e^{-bn}$$

Dans la démonstration, on écrira  $T$  au lieu de  $T(s, \delta)$  et on notera  $\theta_n$

l'opérateur de translation de  $(\mathcal{H} \times \mathcal{A})^N : \theta_n((x_p, a_p)_{p \in \mathbb{N}}) = ((x_{p+n}, a_{p+n})_{p \in \mathbb{N}})$

$$\begin{aligned}
 & \sup_{A \in \mathcal{A}} |P_x^\delta(\theta_{n+p}(A)) - P_x^S(\theta_{n+p}(A))| = \\
 & = \sup_{A \in \mathcal{A}} |P_x^\delta(\theta_{n+p}(A) \cap (T > n)) + P_x^\delta(\theta_{n+p}(A) \cap (T \leq n)) \\
 & \quad - P_x^S(\theta_{n+p}(A) \cap (T > n)) - P_x^S(\theta_{n+p}(A) \cap (T < n))| \\
 & \leq P_x^\delta(T > n) + P_x^S(T > n) + \sup_{A \in \mathcal{A}} |P_x^S(\theta_{n+p}(A) \cap (T < n)) - P_x^\delta(\theta_{n+p}(A) \cap (T < n))| \\
 & \leq 2ke^{-bn} + \sup_{A \in \mathcal{A}} |E_x^S(1_{\{T < n\}} E_x^S(1_{\{\theta_{n+p}(A)\}} / \mathcal{B}_n)) - E_x^\delta(1_{\{T < n\}} E_n^\delta(1_{\{\theta_{n+p}(A)\}} / \mathcal{B}_n))| \\
 & \leq 2ke^{-bn} + \sup_{A \in \mathcal{A}} |E_x^S(1_{\{T < n\}} \int \Pi_p^S(X_n, dy) P_y^S(A)) - E_x^\delta(1_{\{T < n\}} \int \Pi_p^S(X_n; dy) P_y^S(A))| \\
 & \leq 2ke^{-bn} + \sup_{A \in \mathcal{A}} |E_x^S(\int \Pi_p^S(X_n, dy) P_y^S(A)) \\
 & \quad - E_x^\delta(\int \Pi_p^S(X_n, dy) P_y^S(A))| + \sup(P_x^\delta(T > n), P_x^S(T > n)) \\
 & \leq 3ke^{-bn} + \sup_{A \in \mathcal{A}} |E_x^S(\int \Pi_p^S(X_n; dy) P_y^S(A)) - E_x^\delta(\int \Pi_p^S(X_n, dy) P_y^S(A))| \\
 & \leq 3ke^{-bn} + \sup_{A \in \mathcal{A}} \sup_{\substack{z \in \mathcal{H} \\ x \in \mathcal{H}}} |\int \Pi_p^S(x, dy) P_y^S(A) - \int \Pi_p^S(z, dy) P_y^S(A)| \\
 & \leq 3ke^{-bn} + 2a\rho^P .
 \end{aligned}$$

On a donc montré que pour tout entier  $n$  et tout entier  $P$ ,

$$\sup_{x \in \mathcal{H}} \sup_{A \in \mathcal{A}} |P_x^\delta(\theta_{n+p}(A)) - P_x^S(\theta_{n+p}(A))| \leq 3ke^{-bn} + 2a\rho^P .$$

En particulier, pour tout  $p$  supérieur à  $n$  :

$$\sup_{x \in \mathcal{H}} \sup_{A \in \mathcal{A}} |P_x^\delta(\theta_{n+p}(A)) - P_x^S(\theta_{n+p}(A))| \leq 3ke^{-bn} + 2a\rho^n$$

c'est à dire, il existe 2 constantes  $k_1$  et  $b_1$  strictement positives, telles

que pour tout  $p$  supérieur à  $n$ , on ait :

$$\sup_{x \in \mathcal{H}} \sup_{A \in \mathcal{A}} |P_x^\delta(\theta_{n+p}(A)) - P_x^s(\theta_{n+p}(A))| \leq k_1 e^{-b_1 n}$$

En particulier ,

$$\text{pour } p = n, \quad \sup_{x \in \mathcal{H}} \sup_{A \in \mathcal{A}} |P_x^\delta(\theta_{2n}(A)) - P_x^s(\theta_{2n}(A))| \leq k_1 e^{-\frac{b_1}{2} 2n}$$

$$\text{pour } p = n+1, \quad \sup_{x \in \mathcal{H}} \sup_{A \in \mathcal{A}} |P_x^\delta(\theta_{2n+1}(A)) - P_x^s(\theta_{2n+1}(A))| \leq k_1 e^{-\frac{b_1}{2} 2n} e^{-\frac{b_1}{2}}$$

D'où il existe  $k_2$  et  $b_2$  2 constante strictement positives telle que pour tout  $n$  de  $\mathbb{N}$ , on ait :

$$\sup_{x \in \mathcal{H}} \sup_{A \in \mathcal{A}} |P_x^\delta(\theta_n(A)) - P_x^s(\theta_n(A))| \leq k_2 e^{-b_2 n}$$

Soit alors  $n$  un entier naturel. Pour tout  $m$  inférieur à  $n$ , pour tout  $x$ , on a :

$$\begin{aligned} P_x^\delta \left( \frac{1}{n} \sum_{p=0}^{n-1} g(X_p, A_p, X_{p+1}) \right) &\leq \int g d \mu^s \otimes s \otimes \Pi - \epsilon \\ &\leq P_x^\delta \left( \frac{1}{n} \sum_{p=m}^{n-1} g(X_p, A_p, X_{p+1}) \right) \leq \int g d \mu^s \otimes s \otimes \Pi - \epsilon + \frac{\|g\|_m}{n} \end{aligned}$$

En utilisant le résultat démontré ci-dessus, on obtient pour tout  $x$ ,

$$\begin{aligned} P_x^\delta \left( \frac{1}{n} \sum_{p=0}^{n-1} g(X_p, A_p, X_{p+1}) \right) &\leq \int g d \mu^s \otimes s \otimes \Pi - \epsilon \\ &\leq k_2 e^{-b_2 n} + P_x^s \left( \frac{1}{n} \sum_{p=m}^{n-1} g(X_p, A_p, X_{p+1}) \right) \leq \int g d \mu^s \otimes s \otimes \Pi - \epsilon + \frac{\|g\|_m}{n} \\ &\leq k_2 e^{-b_2 n} + P_x^s \left( \frac{1}{n} \sum_{p=0}^{n-1} g(X_p, A_p, X_{p+1}) \right) \leq \int g d \mu^s \otimes s \otimes \Pi - \epsilon + \frac{2\|g\|_m}{n} \end{aligned}$$

en particulier prenons  $m = \inf \left( \frac{\epsilon}{4\|g\|}, 1 \right) n$ , on a :



$$P_x^\delta \left( \frac{1}{n} \sum_{p=0}^{n-1} g(X_p, A_p, X_{p+1}) \leq \int \text{gd}(\mu^s \times s \times \Pi) - \epsilon \right) < \left( \int \text{gd} \mu^s \times s \times \Pi - \frac{\epsilon}{2} \right).$$

Puisque la chaîne  $\Pi^s$  est récurrente Doeblin, grâce au théorème 1 de [3], il existe  $k_3$  et  $b_3$  strictement positifs tels que pour tout  $x$  de  $\mathcal{H}$  et pour tout  $n$  de  $N$ , on ait :

$$P_x^s \left( \frac{1}{n} \sum_{p=0}^{n-1} g(X_p, A_p, X_{p+1}) \leq \int \text{gd} \mu^s \times s \times \Pi - \frac{\epsilon}{2} \right) \leq k_3 e^{-b_3 n}$$

Ce qui permet de conclure qu'il existe  $k$  et  $b$  strictement positifs, tels que pour tout  $n$ , on ait :

$$\sup_{x \in \mathcal{H}} P_x^\delta \left( \frac{1}{n} \sum_{p=0}^{n-1} g(X_p, A_p, X_{p+1}) \leq \int \text{gd} \mu^s \times s \times \Pi - \epsilon \right) \leq k e^{-bn}$$

c'est à dire :

$$\overline{\lim} \frac{1}{n} \text{Log} \sup_{x \in \mathcal{H}} P_x^\delta \left( \frac{1}{n} \sum_{p=0}^{n-1} g(X_p, A_p, X_{p+1}) \leq \int \text{gd} \mu^s \times s \times \Pi - \epsilon \right) \leq 0.$$

### 1.8. Optimalité de la stratégie $\hat{s}$

#### Théorème 7

$$\text{Pour tout } \theta, E_\theta^{\hat{s}}(v_c) \leq \frac{-(1+\sigma(1)) \text{Log } c}{I(\theta)}.$$

#### Démonstration

On suit celle du théorème 3.b de [3].

$$\text{Soit } \epsilon \text{ strictement positif et } N_c = \frac{-(1 + \frac{\epsilon}{2}) \text{Log } c}{I(\theta)}$$

$$E_\theta^{\hat{s}}(v_c) \leq E_\theta^{\hat{s}}(v_c \mathbb{1}_{(v_c < N_c)}) + E_\theta^{\hat{s}}(v_c \mathbb{1}_{(v_c > N_c)})$$

$$E_\theta^{\hat{s}}(v_c) \leq N_c + \sum_{n > N_c} n P_\theta^{\hat{s}}(v_c = n).$$

Montrons qu'il existe  $k'$  et  $b'$  strictement positifs tels que pour tout  $n > N_c$ ,

on ait :  $P_{\theta}^{\hat{s}}(v_c = n) \leq k' e^{-b'n}$ .

On a :  $\{v_c = n\} \subset \{T_{\theta} > n\} \cup \{v_{c,\phi_{\theta}} = n\}$ .

$s^{\theta}$  est récurrente doebelin, apériodique, et "exponentiellement voisine" de  $\hat{s}$ .

Le théorème 1.7.1. appliqué à la v.a. :

$$g(\theta, x, a, y) = \text{Log} \frac{P(\theta, x, a, y)}{P(\phi_{\theta}, x, a, y)}$$

entraîne qu'il existe  $k'_2$  et  $b'_2$  strictement positifs tels que pour tout  $n$  de  $N$ , on ait :

$$P_{\theta}^{\hat{s}}(S_n(\theta, \phi_{\theta}) \leq n \frac{I(\theta)}{1 + \frac{\varepsilon}{2}}) \leq k'_2 e^{-b'_2 n}.$$

La relation :  $P_{\theta}^{\hat{s}}(v_{c,\phi_{\theta}} = n) < P_{\theta}^{\hat{s}}(S_{n-1}(\theta, \phi_{\theta}) < -\log c)$  permet d'écrire qu'il existe  $k''_2$  et  $b''_2$  strictement positifs, tels que pour tout  $n > N_c$ , on ait :

$$P_{\theta}^{\hat{s}}(v_{c,\phi_{\theta}} = n) \leq k''_2 e^{-b''_2 n}$$

Cette dernière relation et la consistance exponentielle de la suite  $(\hat{\theta}_n)_{n \in N}$  permettant de conclure qu'il existe  $k'$  et  $b'$  strictement positifs tels que pour tout  $n > N_c$ , on ait :

$$E_{\theta}^{\hat{s}}(v_c) \leq N_c + k' \sum_{n > N_c} n e^{-b'n}$$

On trouve alors facilement la démonstration, comme par exemple au théorème 3.2 de [3].

Le lemme 8 et ce dernier résultat impliquent :

### Théorème 8

La procédure  $\phi_c^{\hat{s}}$  est asymptotiquement la plus économique parmi les procédures de force au moins égale, pour le test de 2 hypothèses simples.

1.9. Exemple : Bandit à 2 bras markovien

On dispose de 2 machines A et B ; les observations possibles après la mise en marche de A ou B sont 0 et 1. Pour chaque machine, l'ensemble  $\{0,1\}$  est une chaîne markovienne. Pour l'une,  $\Pi(0,1) = \alpha_0$  et  $\Pi(1,0) = \alpha_1$  ; pour l'autre,  $\Pi(0,1) = \beta_0$  et  $\Pi(1,0) = \beta_1$ . Les  $\alpha_i$  et  $\beta_i$  (pour  $i$  valant 0 ou 1) sont connus, mais les machines ne sont pas identifiées (les fiches techniques sont mélangées ...) : on se pose le problème de savoir à quelle machine correspond le couple  $(\alpha_0, \alpha_1)$ . Le problème de test que l'on se pose est donc un test de 2 hypothèses simples. L'espace des paramètres est un sous ensemble de  $\mathbb{R}^2$  (la première coordonnée correspondra à la machine A) :

$$\Theta = \{(\alpha, \beta), (\alpha, \beta)\}.$$

On note "A" | "B" l'action de mettre la machine A|B en marche.

a) Les transitions : Pour  $i, j$  valant 0 ou 1 ,

$$\begin{aligned} \Pi((\alpha, \beta), i, \text{"A"}, j) &= \alpha_i \quad \text{si } i \neq j \\ &= 1 - \alpha_i \quad \text{si } i = j \end{aligned}$$

$$\begin{aligned} \Pi((\alpha, \beta), i, \text{"B"}, j) &= \beta_i \quad \text{si } i \neq j \\ &= 1 - \beta_i \quad \text{si } i = j \end{aligned}$$

On pose pour  $i$  valant 0 ou 1

$$m_i = \alpha_i \text{ Log } \frac{\alpha_i}{\beta_i} + (1 - \alpha_i) \text{ Log } \frac{1 - \alpha_i}{1 - \beta_i}$$

$$p_i = \beta_i \text{ Log } \frac{\beta_i}{\alpha_i} + (1 - \beta_i) \text{ Log } \frac{1 - \beta_i}{1 - \alpha_i}$$

b) L'information

$$I((\alpha, \beta), (\beta, \alpha), i, \text{"A"}) = m_i = I((\beta, \alpha), (\alpha, \beta), i, \text{"B"})$$

$$I((\alpha, \beta), (\beta, \alpha), i, \text{"B"}) = p_i = I((\beta, \alpha), (\alpha, \beta), i, \text{"A"})$$

c) Les stratégies déterministes sont au nombre de 4 : On les note :

- \*  $s_1 : s_1(0) = s_1(1) = "A"$
- $s_2 : s_2(0) = s_2(1) = "B"$
- $s_3 : s_3(0) = "A", s_3(1) = "B"$
- $s_4 : s_4(0) = "B", s_4(1) = "A"$

\* On obtient pour les probabilités invariantes des chaînes  $\Pi^{s_i}$  :

$$\mu_{(\alpha, \beta)}^{s_1}(1) = \frac{\alpha_0}{\alpha_0 + \alpha_1}, \quad \mu_{(\beta, \alpha)}^{s_1}(1) = \frac{\beta_0}{\beta_0 + \beta_1}$$

$$\mu_{(\alpha, \beta)}^{s_2}(1) = \frac{\beta_0}{\beta_0 + \beta_1}, \quad \mu_{(\beta, \alpha)}^{s_2}(1) = \frac{\alpha_0}{\alpha_0 + \alpha_1}$$

$$\mu_{(\alpha, \beta)}^{s_3}(1) = \frac{\alpha_0}{\alpha_0 + \alpha_1}, \quad \mu_{(\beta, \alpha)}^{s_3}(1) = \frac{\beta_0}{\beta_0 + \beta_1}$$

$$\mu_{(\alpha, \beta)}^{s_4}(1) = \frac{\beta_0}{\beta_0 + \beta_1}, \quad \mu_{(\beta, \alpha)}^{s_4}(1) = \frac{\alpha_0}{\alpha_0 + \alpha_1}$$

\* On note  $I_i$  l'information  $I((\alpha, \beta), (\beta, \alpha), s_i)$  et  $I_i^*$  l'information  $I((\beta, \alpha), (\alpha, \beta), s_i)$ . On obtient :

$$I_1 = \frac{\alpha_0}{\alpha_0 + \alpha_1} m_1 + \frac{\alpha_1}{\alpha_0 + \alpha_1} m_0 \quad \text{et} \quad I_3 = \frac{\alpha_0}{\alpha_0 + \beta_1} p_1 + \frac{\beta_1}{\alpha_0 + \beta_1} m_0$$

$$I_1^* = \frac{\alpha_0}{\alpha_0 + \alpha_1} p_1 + \frac{\alpha_1}{\alpha_0 + \alpha_1} p_0 \quad \text{et} \quad I_3^* = \frac{\alpha_0}{\alpha_0 + \beta_1} m_1 + \frac{\beta_1}{\alpha_0 + \beta_1} p_0$$

BIBLIOGRAPHIE

- [1] CHERNOFF H. : Sequential design of experiments -  
Ann. Math. Stat., Vol. 30, 755-770 (1959) .
- [2] GEORGIN J.P. : Estimation et contrôle des chaînes de Markov sur des espaces  
arbitraires - Journées de statistique des processus stochas-  
tiques - Proceedings, Grenoble - Lecture notes in Mathematics,  
636 (juin 1977) .
- [3] MAIGRET N. : Majorations de Chernoff et statistique séquentielle pour des  
chaînes de Markov récurrentes au sens de Doebelin - Séminaires  
d'Orsay - Astérisque.
- [4] MAIGRET N. : Majorations de Chernoff pour des chaînes de Markov contrôlées.  
A paraître .

Nelly MAIGRET  
Université Paris Nord  
Mathématiques  
ERA CNRS 532 "Statistique Appliquée"  
Avenue J.B. Clément  
93430 VILLETANEUSE

VITESSES OPTIMALES DE CONVERGENCE

DES ESTIMATEURS

Lucien BIRGÉ

I - Introduction

Etant donné un problème statistique défini par un espace de paramètres  $\Theta$  muni d'une distance  $d$  et par des probabilités  $P_\theta$  sur un espace  $(\Omega^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$ , on peut s'intéresser aux performances des estimateurs  $\hat{\theta}_n$  de  $P_\theta^n$  où  $P_\theta^n$  est la projection de  $P_\theta$  sur  $(\Omega^n, \mathcal{A}^n)$ . En particulier si les  $P_\theta^n$  se comportent bien en fonction de  $n$ , (et si la distance choisie sur  $\Theta$  correspond à une distance entre les  $P_\theta$ ) on peut espérer un résultat du genre  $d(\theta, \hat{\theta}_n) \xrightarrow[n \rightarrow +\infty]{} 0$   $P_\theta$  p.s. ou  $\mathbb{E}_\theta(d(\hat{\theta}_n, \theta)) \xrightarrow[n \rightarrow +\infty]{} 0$ .

On peut alors se demander s'il existe une puissance  $\alpha$  de  $n$  telle que  $\mathbb{E}_\theta[n^\alpha d(\hat{\theta}_n, \theta)]$  reste bornée uniformément en  $n$  et  $\theta$  et quel est le meilleur  $\alpha$  possible.

Le Cam dans [1] et [2] s'est intéressé à ce problème en supposant que les  $P_\theta^n$  étaient des mesures produits  $(P_\theta^1)^{\otimes n}$  et que la distance  $d$  était la distance de Hellinger  $h$  entre les  $P_\theta^1$ . Il a montré que, pourvu que l'espace  $\Theta$  ne soit pas trop gros (de dimension métrique finie et compact), on pouvait construire des estimateurs tels que  $\mathbb{E}_\theta[n h^2(\hat{\theta}_n, \theta)]$  soit borné et que ceci constituait la meilleure vitesse possible. On va ici démontrer un résultat analogue, mais qui englobe certains cas de variables dépendantes et permet de retrouver ces résultats de Le Cam en affaiblissant les hypothèses de compacité.

II - Les hypothèses

On suppose que l'on observe un processus  $X = (X_1, \dots, X_n, \dots)$  de loi  $P_\theta$  sur  $(\Omega^{\mathbb{N}}, \mathcal{A}^{\mathbb{N}})$ . On notera  $P_\theta^n$  la loi jointe des  $n$  premières coordonnées de  $X$  et on supposera que  $\Theta$  est muni d'une pseudo-distance  $d$ , c'est-à-dire d'une application de  $\Theta \times \Theta$  dans  $\overline{\mathbb{R}^+}$  qui vérifie les propriétés suivantes :

D1)  $d(s, t) = d(t, s)$

D2) il existe  $M > 0$  et  $B \geq 2$  tels que pour  $d(s,t)$  et  $d(t,u) \leq M$ , on ait l'inégalité triangulaire  $d(s,u) \leq \frac{B}{2}[d(s,t) + d(t,u)]$ .

De plus la topologie induite par  $d$  sur  $\Theta$  doit être liée à la topologie forte sur les  $P_\Theta^n$  de la manière suivante :

D3) il existe  $\eta$  et  $A_1$  positifs tels que si  $\sqrt{n} d(s,t) \leq \eta$ , on ait :

$$\frac{1}{2} \| |P_s^n - P_t^n| \| \leq A_1 \sqrt{n} d(s,t)$$

D4) il existe une constante  $A_2$  telle que pour tous les couples  $s$  et  $t$  on ait :

$$\Pi(s^n, t^n) = 1 - \frac{1}{2} \| |P_s^n - P_t^n| \| \leq A_2 e^{-nd^2(s,t)}.$$

La notation  $\Pi$  est reprise de Le Cam et on peut remarquer (cf. Kraft [3]) que  $\Pi(P,Q)$  représente la somme des erreurs du meilleur test entre  $P$  et  $Q$ .

(D4) précise donc la vitesse de séparation des hypothèses  $P_s^n$  et  $P_t^n$  en fonction de  $n$ .

Définition : On appellera  $\lambda$ -réseau  $\frac{\lambda}{2}$ -discernable  $R_\lambda$  un ensemble de points de  $\Theta$  vérifiant les propriétés suivantes :

- a) Pour tout  $x$  de  $\Theta$ , il existe  $y$  dans  $R_\lambda$  avec  $d(x,y) \leq \lambda$ .
- b) Pour tous  $x$  et  $y$  dans  $R_\lambda$  on a  $d(x,y) \geq \lambda/2$ .

Le théorème de Zorn permet de démontrer sur tout espace muni d'une pseudo-distance l'existence de  $\lambda$ -réseaux  $\lambda$ -discernables et pour une suite  $\lambda_n$  décroissante l'existence de réseaux  $R_{\lambda_n}$  croissants.

On fait sur  $\Theta$  les hypothèses topologiques suivantes :

Il existe des constantes positives  $\alpha, \delta$  et  $A_4$ ,  $p \geq 1$  telles que l'on ait :

R1) pour tous  $\theta$  dans  $\Theta$   $r$  et  $\lambda$  tels que  $\lambda \leq \delta$  et  $r\lambda \leq \alpha$ , on aura :

$$\text{Cardinal}[R_\lambda \cap \mathcal{B}(\theta, r\lambda)] \leq A_4 r^p$$

R2) pour tout entier  $i \geq 2$  et tout  $\theta$  dans  $\Theta$  on a

$$\text{Cardinal}[R_{\alpha_i} \cap \mathcal{B}(\theta, i\alpha)] \leq e^{Q_i^2}$$

où l'on note  $\mathcal{B}(x,r)$  la boule de centre  $x$  et rayon  $r$ . Quitte à changer la

constante  $Q$  qui intervient dans (R2) on peut toujours remplacer  $\alpha$  par une constante  $\alpha' < \alpha$  ce qui nous permettra de supposer :

$$(1) \quad \alpha < \frac{M}{B^2} .$$

L'hypothèse (R1) dit que  $\Theta$  est localement de dimension métrique finie, (R2) qu'il ne croît pas trop vite à l'infini. Lorsque l'hypothèse (R2) n'est pas vérifiée, on pourra toujours se limiter à une boule dans  $\Theta$ .

Généralement, on pourra vérifier l'hypothèse (R1) en comparant la distance  $d$  à une autre distance  $d'$  sur  $\Theta$  (par exemple euclidienne lorsque  $\Theta$  est une partie de  $\mathbb{R}^n$ ) pour laquelle  $\Theta$  est de dimension métrique finie. En particulier (R1) sera vérifiée si l'on peut trouver des constantes  $a \leq b$  et  $r \leq q$  telles que l'on ait :

$$a[d'(s,t)]^q \leq d(s,t) \leq b[d'(s,t)]^r$$

lorsque  $d(s,t)$  est assez petit. cf [2].

Une première conséquence de ces hypothèses est l'existence du maximum de vraisemblance sur un réseau.

#### Lemme 1

Soit un réseau  $R_\lambda$  de pas  $\lambda$  inférieur à  $\alpha$  et  $n > \frac{Q}{\alpha}$ , sous les hypothèses précédentes l'estimateur du maximum de vraisemblance existe sur le réseau  $R_\lambda$  p.s. lorsque  $s$  est un point de  $R_\lambda$ .

#### Démonstration

Si l'estimateur  $\hat{\theta}$  du maximum de vraisemblance n'existe pas, pour toute boule  $\mathcal{B}(s, k\alpha)$  on peut trouver un point  $t$  de  $R_\lambda$  extérieur à la boule tel que le test de rapport de vraisemblance au seuil  $0$  de  $s$  contre  $t$  rejette  $s$ . On va majorer cette probabilité par la somme des probabilités des erreurs des tests de  $s$  contre  $t$  pour tous les points  $t$  hors de la boule  $\mathcal{B}(s, k\alpha)$ . Pour cela, on découpe l'espace en couronnes  $\mathcal{B}(s, (k+1)\alpha) \setminus \mathcal{B}(s, k\alpha)$ . Si  $t$  est dans la couronne, d'après (D4) l'erreur du test de  $s$  contre  $t$  est au plus  $A_2 e^{-nk^2\alpha^2}$  et le nombre



de points  $t$  de la couronne est majoré par celui de la boule  $\mathcal{B}(s, (k+1)\lambda)$  c'est à dire  $A_4 \left(\frac{\alpha}{\lambda}\right)^p e^{Q(k+1)^2}$ . On obtient finalement :

$$P_s^n[\{\omega | \hat{\theta}(\omega) \text{ n'existe pas}\}] \leq \lim_k \sum_{i=k}^{+\infty} A_4 \left(\frac{\alpha}{\lambda}\right)^p e^{Q(i+1)^2} A_2 e^{-ni^2\alpha^2}.$$

D'après l'hypothèse, pour  $i$  assez grand on aura  $Q(i+1)^2 < ni^2\alpha^2$  et la série du second membre converge, d'où le résultat.

### III - Vitesses maximales d'estimation

On obtient ici un résultat tout à fait analogue à ceux de Le Cam [1] de majoration de la vitesse uniforme de convergence d'un estimateur, pourvu que l'espace  $\Theta$  soit suffisamment riche, en particulier sans point isolé.

#### Proposition 2

Supposons que pour un point  $s$  de  $\Theta$ , il existe une suite  $s_n$  de points de  $\Theta$  différents de  $s$  et tels que  $d(s, s_n)$  tende vers 0. Soit  $K$  un réel positif inférieur à  $\left(\frac{\eta \wedge A_1^{-1}}{2B}\right)^2$  alors, on a pour tout estimateur  $\hat{\theta}$  et tout réel  $r > 0$  :

$$\overline{\lim}_n \sup_{t \in \mathcal{D}(s, r)} P_t^n [n d^2(t, \hat{\theta}) > K] > 0$$

#### Démonstration

Soit  $K'$  tel que  $\eta \wedge A_1^{-1} > K' > 2B \sqrt{K}$ . Par hypothèse, on pourra trouver des entiers  $n$  arbitrairement grands tels qu'il existe un point  $t$  de  $\Theta$  avec

$$r \geq \frac{K'}{\sqrt{n}} > d(s, t) > \frac{2B \sqrt{K}}{\sqrt{n}}.$$

Supposons en outre que  $\sqrt{\frac{K}{n}}$  est inférieur à  $M$  et considérons le test de  $s$  contre  $t$  qui accepte  $s$  si  $d(s, \hat{\theta}_n) \leq d(t, \hat{\theta}_n)$  et  $t$  dans le cas contraire.

D'après (D2) si  $d^2(s, \hat{\theta}_n) < \frac{K}{n}$  alors  $d^2(t, \hat{\theta}_n) > \frac{K}{n} > d^2(s, \hat{\theta}_n)$  puisque  $d(s, t) > 2B \sqrt{\frac{K}{n}}$ .

Donc la somme des erreurs de ce test sera inférieure à

$$P_s^n[\text{nd}^2(s, \hat{\theta}_n) > K] + P_t^n[\text{nd}^2(t, \hat{\theta}_n) > K] \quad \text{et on aura alors :}$$

$$\Pi(P_s^n, P_t^n) \leq \sup_{t \in \mathfrak{B}(s, r)} P_t^n[\text{nd}^2(t, \hat{\theta}_n) > K]$$

Mais d'après (D3), on a  $\Pi(P_s^n, P_t^n) \geq 1 - A_1 \sqrt{n} d(s, t) \geq 1 - A_1 K'$  et donc

$$\overline{\lim}_n \sup_{t \in \mathfrak{B}(s, r)} P_t^n[\text{nd}^2(t, \hat{\theta}) \leq K] > 1 - A_1 K' > 0 .$$

Remarque : Si on suppose que le point  $s$  est tel que pour tout  $r > 0$ , on peut trouver  $t$  avec  $d(s, t) = r$ , le résultat précédent devient :

$$\lim_n \sup_{t \in \mathfrak{B}(s, r)} P_t^n[\text{nd}^2(t, \hat{\theta}) > K] > 0 .$$

### Corollaire 3

Sous les hypothèses de la proposition 2, on a :

$$\overline{\lim}_n \sup_{t \in \mathfrak{B}(s, r)} \mathbb{E}_t[\text{nd}^2(t, \hat{\theta})] > 0 .$$

## IV - Estimateurs atteignant la vitesse optimale

### 1) Construction des estimateurs

Soit  $n$  le nombre de variables observées. Alors, il existe  $b$  dans  $[1, B[$  et  $k, q$  entiers tels que :

$$(2) \quad \sqrt{n} \alpha = b B^k \quad \text{et} \quad (3) \quad B^q \leq \sqrt{n} \frac{M}{Bb} < B^{q+1} .$$

D'après (1)  $k < q$ .

On supposera en outre  $n$  assez grand pour que l'on ait  $k \geq 2$  et  $n \alpha^2 \geq 2Q B^4$  ce qui entraîne :

$$(4) \quad b^2 B^{2k} \geq 2Q B^4 .$$

Etant donné  $a$  tel que

$$(5) \quad a \leq \inf(2^6 \eta, 2^6 \delta, 1)$$

on posera :

$$a_i = \frac{a}{\sqrt{n}} B^{-3(i+1)} \quad b_i = \frac{b}{\sqrt{n}} B^i \quad C = \frac{a}{b} B^{-7}$$

Alors  $\alpha = b_k$ , (3')  $\frac{M}{B^2} < b_q \leq \frac{M}{B}$  et  $a_i \leq C b_i \quad i$ .

Etant donnée une suite croissante de réseaux  $R_i$  de pas  $a_i$  ( $i = 1, \dots, q$ ) on construit une suite  $\hat{\theta}_i$ , des régions  $B_i$ , un estimateur  $\hat{\theta}$  et une variable  $\tau$  comme suit :

$\hat{\theta}_q$  est le maximum de vraisemblance sur  $R_q$  s'il existe, sinon on prend  $\hat{\theta}$  arbitraire et  $\tau = +\infty$ .

On définit quand cela est possible  $B_q = \mathcal{B}(\hat{\theta}_q, b_q)$ . Le processus se poursuit en définissant par récurrence  $\hat{\theta}_i$  le maximum de vraisemblance sur  $B_{i+1} \cap R_i$  si cet ensemble est non vide et  $B_i = B_{i+1} \cap \mathcal{B}(\hat{\theta}_i, b_i)$  (toujours non vide car  $\hat{\theta}_i$  est dans  $B_{i+1}$ ). Si  $B_{i+1} \cap R_i$  est vide, on posera  $\hat{\theta} = \hat{\theta}_{i+1}$  et  $\tau = i+1$ . Si on peut continuer le processus jusqu'au bout on posera  $\hat{\theta} = \hat{\theta}_1$  et  $\tau = 1$ .

On a alors toujours si  $\tau \leq i$   $\hat{\theta}_i \in B_i$   $B_i \subset B_{i+1}$  et  
et diamètre  $B_i \leq B b_i \leq M$ .

## 2) Préliminaires

### Lemme 4

Pour tout  $s$  de  $R_q$ , on a :

$$(6) \quad P_s^n [s \notin B_q] \leq 2A_2 A_4 \left(\frac{b}{a}\right)^p B^{(3q+k+3)p} e^{-\frac{7}{8} b^2 B^{2q}}$$

### Démonstration

Cette expression est majorée par la probabilité qu'il existe un point  $t$  avec  $d(s,t) > b_q$  tel que le test de  $s$  contre  $t$  rejette  $s$ . Pour évaluer cette quantité on fait le même découpage en couronnes que pour le lemme 1 et on majore

le nombre de points de la couronne par le nombre de points de la boule, ce qui donne le résultat :

$$P_s^n [s \notin B_q] \leq A_2 A_4 \left(\frac{\alpha}{a}\right)^p \sum_{\ell=1+q-k}^{+\infty} e^{QB^{2\ell}} e^{-n(b_{k+\ell-1})^2}$$

$$= A_2 A_4 \left(\frac{b}{a}\right)^p B^{(3q+k+3)p} \sum_{\ell=1+q-k}^{+\infty} e^{QB^{2\ell} - b^2 B^{2(k+\ell-1)}}$$

or

$$e^{QB^{2\ell} - b^2 B^{2(k+\ell-1)}} = e^{(Q-b^2 B^{2k-2})B^{2\ell}} \leq e^{-\frac{7}{8} b^2 B^{2k-2\ell-2}}$$

d'après (4) .

Cette série converge très rapidement et on peut la majorer par le double de son premier terme d'où le résultat.

Lemme 5

Pour tout  $s_i$  de  $R_i$  et  $q-1 \geq j \geq i$  , on a :

$$(7) \quad P_{s_i}^n [s_i \in (B_{j+1} - B_j) \cap \tau \leq j] \leq \begin{cases} (A_2 A_4 \left(\frac{b}{a}\right)^p B^{(4j+5)p} e^{-b^2 B^{2j}} & \text{si } j+2 \leq k \\ (A_2 A_4 \left(\frac{b}{a}\right)^p B^{(3j+k+3)p} e^{-\frac{b^2}{2} B^{2j}} & \text{si } j+2 > k \end{cases}$$

Démonstration

Pour tout point de cet ensemble, il existe un point  $\hat{\theta}_j$  de  $R_j$  , inclus dans  $B_{j+1}$  (donc  $d(s_i, \hat{\theta}_j) \leq B b_{j+1}$ ) tel que le test de  $s_i$  contre  $\hat{\theta}_j$  accepte  $\hat{\theta}_j$  et  $d(s_i, \hat{\theta}_j) \geq b_j$  . Donc la probabilité considérée est majorée par  $e^{-n b_j^2}$  fois le nombre de points de la boule de centre  $s_i$  et rayon  $B b_{j+1}$  qui appartiennent au réseau  $R_j$  . On a alors deux cas à considérer :

- si  $j+2 \leq k$  on obtient la majoration :

$$A_2 A_4 \left(\frac{b_{j+2}}{a_j}\right)^p e^{-n b_j^2} = A_2 A_4 \left(\frac{b}{a}\right)^p B^{(4j+5)p} e^{-b^2 B^{2j}}$$

- si  $j+2 > k$  il vient :

$$A_2 A_4 \left(\frac{b_k}{a_j}\right)^p e^{QB^{2(j+2-k)}} e^{-n b_j^2} = A_2 A_4 \left(\frac{b}{a}\right)^p B^{(3j+k+3)p} e^{2j} [QB^{4-2k} - b^2]$$

Comme la majoration obtenue ne dépend pas de  $i$  mais seulement de  $j$ , on notera  $H_j$  le terme du second membre de (7) et  $H_q$  celui de (6).

Lemme 6

$$P_{s_i}^n [d(s_i, \hat{\theta}) > Bb_i] \leq \sum_{j=i}^{q-1} H_j + H_q \quad s_i \in R_i \quad i \leq q$$

Si  $d(s_i, \hat{\theta}) > Bb_i$ , alors  $s_i$  n'appartient pas à  $B_i$  et on peut donc majorer  $P_{s_i}^n [d(s_i, \hat{\theta}) > Bb_i]$  par :

$$\sum_{j=i}^{q-1} P_{s_i}^n [s_i \in B_{j+1} - B_j) \cap \tau \leq j] + P_{s_i}^n [s_i \in B_q]$$

d'où le résultat .

IV - Vitesse atteinte par l'estimateur

Théorème 7

Soit un nombre  $N \geq \frac{B}{4}[B+C]b_q$ , et  $\hat{\theta}$  l'estimateur précédemment construit, il existe une fonction  $K(M)$  ne dépendant pas de  $n$ , et bornée sur tout intervalle  $[\varepsilon, +\infty[$  telle que l'on ait :

$$IE_s [n(d(s, \hat{\theta}) \wedge N)^2] \leq K(M) .$$

Démonstration

On peut écrire :

$$\begin{aligned} IE_s [n(d \wedge N)^2] &= n \int_0^{+\infty} P_s^n [(d \wedge N)^2 > t] dt \\ &= n \int_0^N P_s^n [d > u] 2u du \quad \text{où } d = d(s, \hat{\theta}) \\ &= n(I_1 + I_2 + I_3) \quad \text{avec} \end{aligned}$$

$$I_1 = \int_0^{B/2(B+C)b_1} P_s^n [d > u] 2u du \leq \frac{B^2}{4}(B+C)^2 b_1^2 = \frac{b^2 B^4 (B+C)^2}{4n}$$

$$I_2 = \int_{\frac{B}{2}[B+C]b_q}^N P_s^n [d(s, \hat{\theta}) > a] 2u \, du$$

$$I_3 = \sum_{i=1}^{q-1} \int_{\frac{B}{2}(B+C)b_i}^{B/2(B+C)b_{i+1}} P_s^n [d(s, \hat{\theta}) > u] 2u \, du$$

Au point  $s$  associons une suite  $s_i$ ,  $i = 1, \dots, q$  telle que  $d(s, s_i) \leq a_i \leq Cb_i$  pour tout  $i$ .

Par (D3) et (5), on aura :

$$P_s^n [d(s, \hat{\theta}) > u] \leq P_{s_i}^n [d(s, \hat{\theta}) > u] + A_i \sqrt{n} a_i$$

et d'après (D2) :

$$P_{s_i}^n [d(s, \hat{\theta}) > \frac{B}{2}(B+C)b_i] \leq P_{s_i}^n [d(s_i, \hat{\theta}) > Bb_i] .$$

Ceci permet d'obtenir pour  $I_2$  et  $I_3$  les majorations :

$$I_2 \leq (N^2 - \frac{B^2}{4}(B+C)^2 b_q^2) (A_1 \sqrt{n} a_q + P_{s_q}^n [d(s_q, \hat{\theta}) > Bb_q])$$

$$I_3 \leq \frac{B^2}{4} (B+C)^2 \sum_{i=1}^{q-1} (b_{i+1}^2 - b_i^2) (A_1 \sqrt{n} a_i + P_{s_i}^n [d(s_i, \hat{\theta}) > Bb_i])$$

On obtient donc finalement, en utilisant les lemmes 4 et 6 ainsi que (3) :

$$I_2 \leq [N^2 - B^2 (\frac{B+C}{4n})^2 b^2 B^2] [A_1 a B^{-3(q+1)} + H_q]$$

$$I_3 \leq \frac{B^2 (B+C)^2}{4n} b^2 \sum_{i=1}^{q-1} (B^{2i+2} - B^{2i}) [A_1 a B^{-3(i+1)} + \sum_{j=i}^{q-1} H_j + H_q]$$

Donc :

$$I_2 + I_3 \leq [N^2 - \frac{B^4 (B+C)^2 b^2}{4n}] H_q + [N^2 - \frac{B^2 (B+C)^2 b^2}{4n} B^2 q] A_1 a B^{-3(q+1)} + \frac{B^2 (B^2 - 1) (B+C)^2 b^2}{4n} \left[ \sum_{i=1}^{q-1} A_1 a B^{-i-3} + \sum_{i=1}^{q-1} B^{2i} \sum_{j=i}^{q-1} H_j \right]$$

$$\begin{aligned} &\leq N^2 H_q + \left[ N^2 - \frac{M^2(B+C)^2}{4B^2} \right] A_1 a \frac{B^3 b^3}{M^3 n^{3/2}} \\ &+ \frac{B^2(B^2-1)(B+C)^2 b^2}{4n} \left[ A_1 a \frac{B^{-3} - B^{-q-2}}{B-1} + B^{2 \sum_{j=1}^{q-1} H_j} \frac{B^{2j-1}}{B^2-1} \right] \\ &\leq N^2 H_q + \left[ N^2 - \frac{M^2(B+C)^2}{4B^2} \right] A_1 a \left[ \frac{BB}{M/n} \right]^3 + A_1 a \frac{(B+1)(B+C)^2 b^2}{4nB} \\ &+ \frac{B^4(B+C)^2 b^2}{4n} \sum_{j=1}^{q-1} B^{2j} H_j \end{aligned}$$

(2), (3) et (3') donnent une majoration de  $H_q$  par

$$2A_4 A_2 \left(\frac{b}{a}\right)^p B^3 \left(\frac{\sqrt{n} M}{Bb}\right)^{3p} \left(\frac{\sqrt{n} \alpha}{b}\right)^p e^{-\frac{7}{8} \frac{n M^2}{B^4}} \leq 2A_4 A_2 B^3 \left[\frac{M^3 \alpha}{a B^3}\right]^p n^{2p} e^{-\frac{7}{8} \frac{n M^2}{B^4}}$$

Le lemme 3 va nous donner :

$$\sum_{j=1}^{k-2} B^{2j} H_j = A_2 A_4 \left(\frac{b}{a}\right)^p B^{5p} \sum_{j=1}^{k-2} B^{2j(2p+1)} e^{-b^2 B^{2j}}$$

$$\sum_{j=k-1}^{q-1} B^{2j} H_j = A_2 A_4 \left(\frac{b}{a}\right)^p B^{4p} \sum_{j=k-1}^{q-1} B^{2j(2p+1)} B^{p(k-1-j)} e^{-\frac{b^2}{2} B^{2j}}$$

et donc 
$$\sum_{j=1}^{q-1} B^{2j} H_j \leq A_2 A_4 \left(\frac{b}{a} B^4\right)^p \sum_{j=1}^{q-1} B^{2j(2p+1)} e^{-\frac{B^{2j}}{2}}$$

Si on pose 
$$S = \sum_{j=1}^{+\infty} B^{2j(2p+1)} e^{-\frac{1}{2} B^{2j}} < +\infty,$$

$K_1 = A_2 A_4 \left(\frac{B^5}{a}\right)^p$   $K_2 = 2A_4 A_2 B^3 \left[\frac{\alpha}{\alpha B^3}\right]^p$ , on pourra écrire :

$$\sum_{j=1}^{q-1} B^{2j} H_j \leq K_1 S \quad M_n^2 H_q \leq K_2 M^{3p+2} n^{2p+1} e^{-\frac{7}{8} \frac{n M^2}{B^4}}$$

Si  $K_3$  est le maximum de la fonction  $x^{2p+1} e^{-\frac{7x}{8B^4}}$  on aura

$$M_n^2 H_q \leq K_2 K_3 M^{-p} H(n, M) \quad \text{où } H(n, M) \longrightarrow 0 \text{ si } n \text{ ou } M \longrightarrow +\infty \text{ et } H(n, M) \leq 1$$

pour tout  $n$  et tout  $M$ .

On obtient finalement avec  $K_4 = \frac{B^6(B+C)^2}{4}$

$$\begin{aligned} \mathbb{E}_s [n(d \wedge N)^2] &< K_4 + K_2 K_3 \frac{N^2}{M^{2+p}} H(n, M) + (N^2 - \frac{M^2 K_4}{B^8}) A_1 a \frac{B^6}{M^3 \sqrt{n}} \\ &+ A_1 a K_4 (B+1) B^{-S} + K_1 K_4 S \end{aligned}$$

C.Q.F.D.

Remarques : Si (D2) est vérifiée pour tout M positif, en posant  $N = \frac{M}{-B^4} \sqrt{K_4}$ , on voit pour la formule précédente que  $\mathbb{E}_s [n(d(s, \hat{\theta}) \wedge N)^2]$  admet une borne qui ne dépend pas de N et donc que pour tout entier N, on pourra construire un estimateur  $\hat{\theta}(N)$  qui vérifiera :

$$\mathbb{E}_s [n(d(s, \hat{\theta}(N)) \wedge N)^2] \leq K \quad \text{uniformément en } N \text{ et } n .$$

- Tout ce qui précède peut se transposer aisément en remplaçant  $\sqrt{n}$  par  $n^\gamma$  si on suppose que dans (D3) et (D4) les équations deviennent :

$$\frac{1}{2} | | P_s^n - P_t^n | | < A_1 n^\gamma d(s, t) \quad \Pi(s^n, t^n) < A_2 e^{-n^{2\gamma} d^2(s, t)}$$

le résultat devient alors  $\mathbb{E}_s [ [n^\gamma (d(s, \hat{\theta}) \wedge N) ]^2 ] < K$ . Cependant ce résultat reste vide faute d'exemples de processus vérifiant (D3) et (D4) dans le cas  $\gamma \neq \frac{1}{2}$ .

## V - Applications

### 1) Cas des variables indépendantes

Dans ce cas  $P_\theta^n$  s'écrit  $(P_\theta^1)^{\otimes n}$ . Conformément à Le Cam [1], on posera :

$$\rho(P, Q) = \int \sqrt{dP dQ} \quad h^2(P, Q) = \frac{1}{2} \int (\sqrt{dP} - \sqrt{dQ})^2 = 1 - \rho$$

On a alors (8)  $h^2(P, Q) \leq \frac{1}{2} | | P - Q | | < \sqrt{2} h(P, Q)$

et (9)  $\rho(P^{\otimes n}, Q^{\otimes n}) = \rho^n(P, Q)$ .





On posera  $d^2 = -\log(1-h^2) = -\log \rho$ .

Dans ce cas pour  $h^2 < M' < \frac{1}{2}$  on pourra écrire  $h < d < Bh$  ce qui permet avec (8) d'établir (D2) et (D3). (D4) se déduit de (9) et du fait que  $\Pi$  est inférieur à  $\rho$ .

## 2) Cas des chaîne de Markov

On suppose donnée sur  $(\mathcal{X}, \mathcal{A})$  une famille  $K_\theta(x, dy)$  de noyaux de transition. On notera  $K_\theta^n(x, dy)$  le noyau itéré  $n$  fois de  $K_\theta(x, dy)$ .  $P_\theta$  est alors la loi de la chaîne de Markov associée de loi initiale  $\nu$ .

### Proposition 8

Supposons qu'il existe une mesure  $\mu$  et des réels positifs  $0 < a < A$  tels que toutes les probabilités  $\nu$  et  $K_s(x, dy)$  soient dominées par  $\mu$  pour tous les  $s$  et  $x$ , de densités  $g(y)$  et  $p_s(x, y)$  respectivement et que l'on ait pour tout  $s$  et tous  $x$  et  $x'$  dans  $\mathcal{X}$

$$a \leq \frac{p_s(x, y)}{p_s(x', y)} \leq A \quad \text{et} \quad \frac{g(y)}{p_s(x, y)} \leq A$$

alors il existe sur  $\Theta$  une pseudo-distance vérifiant les hypothèses (D1) à (D4).

### Démonstration

On notera  $q_{s,t}(x, y) = \sqrt{p_s(x, y) p_t(x, y)}$  et on définira les noyaux itérés

$$p_s^n(x, y) = \int p_s(x, x_1) p_s(x_1, x_2) \dots p_s(x_{n-1}, y) d\mu(x_1) \dots d\mu(x_n)$$

et de même  $q_s^n(x, y)$ .

On aura alors  $\rho(P_s^n, P_t^n) = \int q_{s,t}^n(x, y) \nu(dx) \mu(dy)$ .

On définit alors  $d$  par la relation :

$$(10) \quad d^2(s, t) = -\frac{1}{3} \log \left[ \sup_x \int q_{s,t}^2(x, y) \mu(dy) \right].$$

On va tout d'abord vérifier (D3) :

$q_{s,t}(x,y)$  définit sur  $\mathfrak{X}$  un noyau sous-Markovien  $R$  défini par

$$R(x,A) = \int_A q_{s,t}(x,y) \mu(dy)$$

qu'on prolonge en un noyau Markovien sur  $\mathfrak{X} \cup \{\Delta\}$  en posant :

$$R(x, \{\Delta\}) = 1 - R(x, \Omega) \quad \text{si } x \text{ est dans } \Omega \quad \text{et } R(\Delta, \{\Delta\}) = 1 .$$

On aura alors  $R^n(x, \mathfrak{X}) = \int q_{s,t}^n(x,y) \mu(dy)$  , et d'après (10)

$$\inf_x R^2(x, \{\Delta\}) = 1 - e^{-3d^2(s,t)} .$$

Si on appelle  $\beta$  cette quantité, pour tout  $\varepsilon$  positif on pourra trouver un  $x$  tel que :

$$R^2(x, \{\Delta\}) = \int_{\mathfrak{X}} R(x, dy) R(y, \{\Delta\}) + R(x, \{\Delta\}) \leq \beta + \varepsilon .$$

Soit alors un point  $x'$  quelconque de  $\mathfrak{X}$  on aura :

$$\begin{aligned} \int_{\mathfrak{X}} R(x', dy) R(y, \{\Delta\}) &= \int_{\mathfrak{X}} R(x, dy) d \frac{R(x', dy)}{R(x, dy)} R(y, \{\Delta\}) \\ &\leq \frac{A}{a} \int_{\mathfrak{X}} R(x, dy) R(y, \{\Delta\}) \leq \frac{A}{a} (\beta + \varepsilon) . \end{aligned}$$

Comme  $\varepsilon$  est arbitraire , il vient :

$$(11) \quad \sup_x \int_{\mathfrak{X}} R(x, dy) R(y, \{\Delta\}) \leq \frac{A\beta}{a}$$

et par le même raisonnement :

$$(12) \quad \int_{\mathfrak{X}} R(y, \{\Delta\}) \nu(dy) \leq \frac{A\beta}{a} .$$

On peut alors calculer  $R^n(x, \mathfrak{X})$  par récurrence, pour tout  $x$

$$R^2(x, \mathfrak{X}) = 1 - R^2(x, \{\Delta\}) \geq 1 - \frac{A\beta}{a} - R(x, \{\Delta\}) .$$

Si  $R^i(x, \mathfrak{X}) \geq 1 - (i-1) \frac{A\beta}{a} - R(x, \{\Delta\})$  on aura :

$$\begin{aligned} R^{i+1}(x, \mathfrak{X}) &= \int_{\mathfrak{X}} R(x, dy) R^i(y, \mathfrak{X}) \\ &\geq R(x, \mathfrak{X}) - (i-1) \frac{A\beta}{a} - \int_{\mathfrak{X}} R(x, dy) R(y, \{\Delta\}) \\ &\geq 1 - R(x, \{\Delta\}) - i \frac{A\beta}{a} \quad \text{d'après (11)} . \end{aligned}$$

Il s'en suivra que l'on a :

$$\begin{aligned} \rho(P_s^n, P_t^n) &= \int R^n(x, \mathfrak{X}) \nu(dx) \geq 1 - \frac{(n-1)A\beta}{a} - \int R(x, \{\Delta\}) \nu(dx) \\ &\geq 1 - \frac{nA\beta}{a} \quad \text{par (12)} . \end{aligned}$$

En utilisant la relation (8) et en remplaçant  $\beta$  par sa valeur, on obtiendra ainsi :

$$\frac{1}{2} \|P_s^n - P_t^n\| \leq \sqrt{\frac{6An}{a}} d(s, t) .$$

(D4) est immédiat puisque l'on a :

$$\rho(P_s^{2n}, P_t^{2n}) = \int q_{s,t}^{2n}(x, y) \nu(dx) \mu(dy) \leq \left[ \sup_x \int q_{s,t}^2(x, y) \mu(dy) \right]^n \leq e^{-3nd^2(s, t)}$$

comme  $\rho$  décroît avec  $n$  on a la même inégalité pour  $\rho(P_s^{2n+1}, P_t^{2n+1})$  et le résultat s'ensuit puisque  $\Pi$  est inférieur à  $\rho$  .

Pour (D2) on utilise le fait que  $d$  est comparable à la distance de Hellinger.

En effet, on a :

$$\rho(P_s^2, P_t^2) \geq 1 - \frac{2A}{a} [1 - e^{-3d^2(s, t)}]$$

et 
$$\rho(P_s^2, P_t^2) \leq e^{-3d^2(s, t)} .$$

On aura donc :

$$1 - e^{-3d^2(s,t)} \leq H^2(P_s^2, P_t^2) \leq \frac{2A}{a} [1 - e^{-3d^2(s,t)}] .$$

Ce qui montre que  $d$  et  $H$  sont équivalentes tant que  $H$  n'est pas trop grand et donc que  $d$  vérifie bien (D2) pour  $M$  et  $B$  convenablement choisis.

Bibliographie :

- [1] LE CAM L. : Convergence of estimates under dimensionality restrictions  
Annals of Statistics, 1973, vol. 1, n° 1, p. 38-53
  
- [2] LE CAM L. : On local and global properties in the theory of asymptotic normality of experiments  
Stochastic process and related Topics, vol. 1, Acad. Press, 1975
  
- [3] KRAFT C. : Some conditions for consistency and uniform consistency of statistical procedures  
University of California, Publications in Statistics, vol. 2, n° 6, p.125-142
  
- [4] DACUNHA-CASTELLE D. : Vitesse de convergence pour certains problèmes statistiques  
Ecole d'Eté de Probabilités de Saint-Flour, 1977 - Lecture Notes in Maths.  
n° 678, Springer-Verlag.

Lucien BIRGÉ  
Université Paris VII  
Mathématiques  
ERA CNRS 532 "Statistique Appliquée"  
2, Place Jussieu  
75221 PARIS CEDEX 05

QUELQUES OCCURENCES DES GRANDS ECARTS DANS LA LITTERATURE "ELECTRONIQUE"

Gabriel RUGET.

Sous le nom de "Chernoff bound", la majoration exponentielle de la probabilité qu'une somme de  $n$  variables indépendantes dépasse un seuil  $na$  est aussi répandue dans les mathématiques de l'électronique que le théorème central limite ; pour voir un exemple très proche du centre de cet exposé, la transmission des données, on pourra examiner [18] puis [22]\*.

Plus rares sont les utilisations (conscientes ou pas) de théorèmes de grands écarts d'apparence plus sophistiquée quoique équivalents (lois empiriques, marches aléatoires). Mais ceux qui soupèsent avec délicatesse la queue d'une queue (cf. [10]), pourraient utiliser les théorèmes de Borovkov (cf. exposé n° IV, Deshayes - Picard) qui, non seulement peuvent prendre la place de l'identité de Wald, mais permettent d'éviter le recours à [13]: une queue augmente d'une unité selon un processus de renouvellement (loi  $G$  de transformée de Cramer  $\Gamma$ ) et diminue de même (loi  $F$  de transformée de Cramer  $\phi$ , arrivées et départs indépendants) ; l'espérance de  $F$  est inférieure à celle de  $G$ , de sorte qu'en moyenne la queue se vide ; la probabilité d'une excursion, à partir de la taille d'équilibre  $0$ , jusqu'à la taille  $n$  est asymptotiquement exponentielle et l'on

---

\* Une suite de données (à nombre de niveaux fini) passe dans un filtre linéaire (interférence intersymboles), puis est bruitée ; suivant une idée de Forney, on veut la rétablir en filtrant de façon à réduire le nombre de symboles interférant, sans trop augmenter le bruit, puis en décodant séquentiellement (et on peut alors se permettre de décoder au maximum de vraisemblance en utilisant l'algorithme de Viterbi); la ligne de transmission (i.e. le premier filtre) n'étant pas stationnaire, le filtre à la réception est rendu adaptatif; il n'est donc jamais exactement ce qu'on voudrait, et cela provoque un bruit d'interférence intersymboles résiduel que l'on veut majorer pour évaluer les performances du décodeur.

cherche l'exposant. Heuristiquement : si  $C_i$  désignent les temps entre départs successifs,  $P(C_1 + \dots + C_k \simeq kx) \simeq e^{-k\phi(x)}$  donc  
 $P(C_1 + \dots + C_{\alpha T} \simeq T) \simeq e^{-\alpha T\phi(1/\alpha)}$  ; la probabilité que, pendant un temps  $T$ , il y ait  $\alpha$  départs et  $\beta$  arrivées par unité de temps, vaut  
 $\exp - T (\alpha\phi(1/\alpha) + \beta\Gamma(1/\beta))$  ; or le temps pour que, dans ces conditions, la taille de la queue augmente de  $n$ , vaut  $\frac{n}{\beta-\alpha}$  ; l'exposant cherché est donc

$$\min_{\alpha, \beta} \frac{1}{\beta-\alpha} (\alpha\phi(1/\alpha) + \beta\Gamma(1/\beta))$$

En dérivant, on trouve les conditions

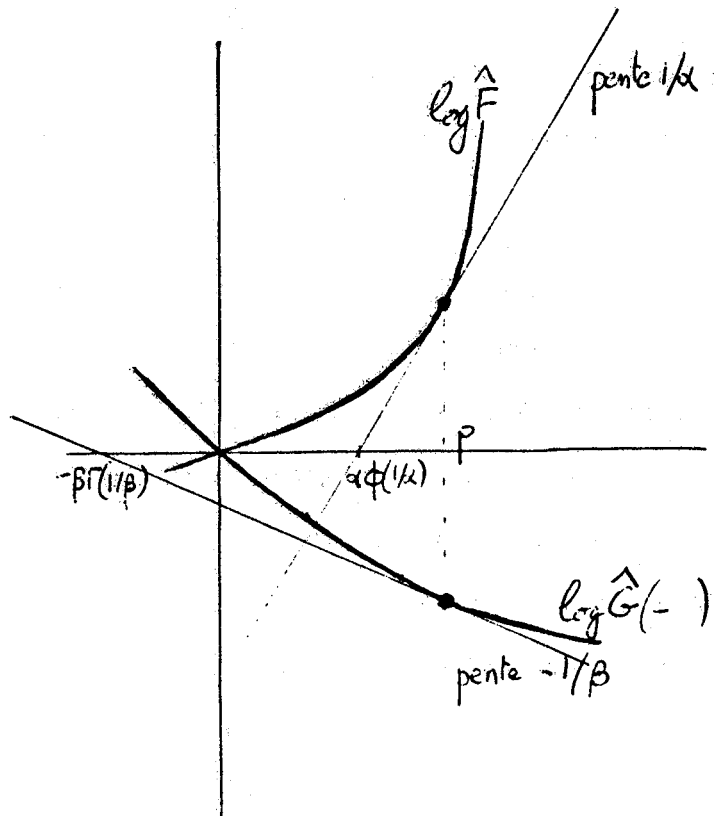
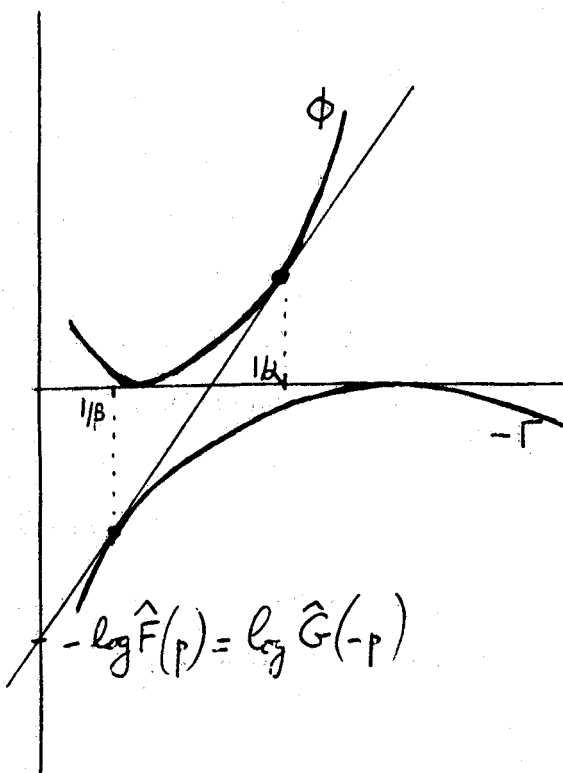
$$\phi(1/\alpha) + \Gamma(1/\beta) = \left(\frac{1}{\alpha} - \frac{1}{\beta}\right) \phi'(1/\alpha) = \left(\frac{1}{\beta} - \frac{1}{\alpha}\right) \Gamma'(1/\beta)$$

Posant  $\phi'(1/\alpha) = -\Gamma'(1/\beta) = p$ , au point solution, les conditions deviennent

$$\phi(1/\alpha) - p \frac{1}{\alpha} = -\Gamma(1/\beta) - p \frac{1}{\beta}$$

ce qui exprime que les tangentes en  $1/\alpha$  à  $\phi$  et en  $1/\beta$  à  $-\Gamma$  sont confondues, d'où la condition déterminant  $p$

$$\hat{F}(p) \hat{G}(-p) = 1$$



Un retour au graphe des transformées de Laplace montre qu'alors

$$\alpha \phi(1/\alpha) + \beta \Gamma(1/\beta) = -\beta \log \hat{G}(-p) - \alpha \log \hat{F}(p)$$

D'où l'exposant  $\log \hat{F}(p)$ .

Cette heuristique peut être transformée en analyse précise comme suit : on définit une nouvelle loi  $\tilde{P}$  pour le processus en décrétant que les temps de service sont de loi  $d\tilde{F} = e^{p \cdot} dF(\cdot)/\hat{F}(p)$  et les temps entre arrivées de loi  $d\tilde{G} = e^{-p \cdot} dG(\cdot)/\hat{G}(p)$ , pour le  $p$  déterminé ci-dessus. On peut évaluer  $P(\Sigma)/\tilde{P}(\Sigma)$ , où  $\Sigma$  désigne l'ensemble des trajectoires du processus issues de 0 (file vide) et atteignant la taille  $n$  sans se vider à aucun instant intermédiaire : on trouve

$$P(\Sigma) / \tilde{P}(\Sigma) = \frac{\alpha}{p} (\hat{F}(p) - 1) \hat{F}(p)^{-n}$$

$\tilde{P}(\Sigma)$ , qui tend vers une limite non nulle lorsque  $n \rightarrow \infty$ , peut être évalué numériquement à peu de frais, ainsi que le temps moyen de retour à l'état vide, on dispose alors du temps moyen de premier passage à l'état  $n$ .

Bien entendu, cette approche permet (sans aboutir à un résultat explicite aussi simple) d'envisager le cas de plusieurs processus de renouvellement indépendants rythmant chacun un processus d'accroissements non plus déterministes, mais i.i.d.

Mais le véritable avantage d'une approche consciente des situations de grands écarts, ce sont les outils disponibles lorsque les accroissements ne sont pas homogènes. Tel est le cas dans la modélisation du système ALOHA (transmission de trains de données de longueur fixe, synchronisés, sur un canal - satellite, fibre de verre ... - d'accès libre pour un grand nombre d'utilisateurs : ceux-ci, qui écoutent en permanence le canal, savent après chaque intervalle de temps si celui-ci a été utilisé par zéro, un, ou plusieurs terminaux ; dans ce dernier cas, tous les messages sont perdus, et doivent être réémis suivant une politique à déterminer. On désire calculer :

1°) Pour les politiques les plus simples, qui ne stabilisent pas le canal, le temps moyen avant déstabilisation (le processus  $N(t)$  suivi est celui du nombre des messages à réémettre ; les politiques de réémission et la loi d'arrivée de nouveaux messages sont choisis de façon à le rendre markovien). Si  $v(n) = E(N(t+1) - N(t) \mid N(t) = n)$  est  $\geq \alpha > 0$  pour  $n$  assez grand, la chaîne est transiente ("instabilité") ; soit  $n_c$  le plus petit  $n$  tel que  $v(n)$  soit  $> 0$  pour  $n \geq n_c$  ; c'est le temps du premier passage à  $n_c$  qu'on appelle temps de déstabilisation.

2°) Pour les politiques plus sophistiquées [7] qui stabilisent le canal (chaîne positive récurrente), le bon critère de comparaison est, non pas le temps moyen nécessaire pour faire passer un message, qui est faible, mais la probabilité que le temps d'attente dépasse un seuil (correspondant à une taille de mémoire tampon). On peut ramener ce problème à l'étude, pour une chaîne de Markov sur  $N$  à "petits" accroissements et petite probabilité d'extinction, de la probabilité de survie d'une trajectoire pendant un temps  $> T$  ( $T$  grand).

Ces méthodes seront par exemple exposées dans [5]. Comme dans l'exemple précédent, une étude heuristique basée sur des théorèmes asymptotiques grossiers, fournit un bon changement de probabilité, après quoi on peut procéder à une simulation (c'est une technique d' "importance sampling") ou à un calcul analytique (approximation par une diffusion).

Terminons cet exposé par une revue rapide d'un rameau récent de la théorie de Shannon : il s'agit de borner des probabilités d'erreur - exponentiellement petites - lorsqu'on transmet sur un canal ou lorsqu'on code une source avec un critère de distorsion, les codes étant par blocs (ce qui tend vers l'infini et rend la probabilité d'erreur, resp. de dépasser un niveau de distorsion imposé, petite, est la taille des blocs). Précisément :

1° Problème : un canal bruité est caricaturé sous les traits d'une matrice  $Q_{k/j}$  de probabilités de transition connue et fixe entre un alphabet d'entrée fini  $J$  et un alphabet de sortie fini  $K$  (on ne prend pas en compte ici



l'interférence intersymboles). Un code par blocs de longueur  $N$  et de rendement  $R$  est la donnée de  $M = e^{NR}$  mots  $a_m$  de longueur  $N$  dans l'alphabet  $J$ , et d'une partition  $Y_m$  de l'ensemble des mots de longueur  $N$  de l'alphabet  $K$ , indexée par le vocabulaire d'entrée retenu. Outre son rendement, le code est apprécié en fonction de la probabilité d'erreur  $\max_{a_m} P$  (la traduction de  $a_m$  par le canal ne tombe pas dans  $Y_m$ ), ou d'une probabilité d'erreur moyenne, ce qui revient au même vu le peu de précision de ce qu'on peut dire de ces probabilités. On cherche une minoration de ces probabilités d'erreur, dont on démontre ensuite l'acuité par des majorations aussi voisines que possible (obtenues en exhibant des codes, même irréalistes, parce que la quantité de calculs à effectuer pour leur décodage serait prohibitive) ; le seul intérêt de ce travail est d'apprécier l'efficacité de codes utilisables. Au mieux, on obtient que la meilleure probabilité d'erreur accessible a un logarithme équivalent à  $-NE(R)$  (lorsque  $N \rightarrow \infty$ ; on précisera  $E(R)$ ) ; la théorie, qui ne prend pas en compte les coûts de calcul et de transmission, ne dit pas si, pour atteindre une probabilité d'erreur donnée, il vaut mieux chercher à se rapprocher de l'optimum à  $N$  modérément grand, ou augmenter  $N$ , voire changer le canal (pour jouer sur  $E(R)$ ) ; il est en effet difficile d'évaluer le coût en calcul sans vraiment particulariser les algorithmes, et la pondération entre les coûts de calcul et ceux de transmission dépend de la technologie, donc est très fluctuante ; il semble qu'actuellement, on puisse se permettre beaucoup de calculs, et des  $N$  grands.

#### Résultats :

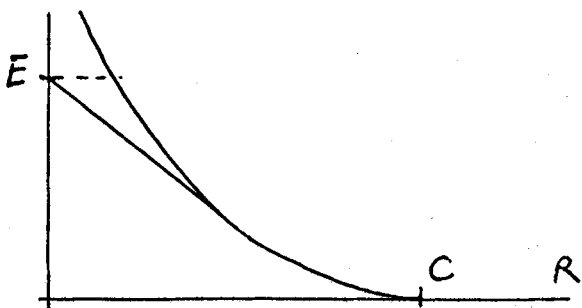
- Shannon s'était d'abord borné à dire que, pour  $R < C = \sup_p \sum_{j,k} p_j Q_{k/j} \log \frac{Q_{k/j}}{\sum_j p_j Q_{k/j}}$ , on peut réduire autant qu'on veut la probabilité d'erreur, et que c'est impossible pour  $R > C$ . Quelques articles (de [21] à [15]) étendent la notion de capacité à des canaux stationnaires, mais ayant une mémoire. Nous engageons plutôt le lecteur à consulter l'article de revue [25], où l'on décrit comment ce type de résultats a pu être étendu à des situations de communication plus générales, par exemple un canal portant deux

communications en sens contraire interférant, ou un codeur devant coder pour plusieurs canaux simultanément.

- Ce sont les minoration exponentielles de probabilité d'erreur qui font appel aux grands écarts (pour les majorations, voir [8]). Dans [2], on trouvera une preuve (par réduction à un problème de test) de ce qu'on peut minorer l'erreur en prenant pour exposant

$$E(R) = \max_{s \geq 0} \max_p [-sR - \log \sum_k (\sum_j p_j Q_{k/j}^{1/1+s})^{1+s}] .$$

Malheureusement, cette minoration n'est optimale que pour les rendements tels que le  $s$  maximisant les expressions ci-dessus soit  $\leq 1$ . Le seul autre résultat dont on dispose est une minoration optimale aux rendements voisins de zéro [23]



$$E = \max_p - \sum_{ik} p_i p_k \log \sum_j \sqrt{Q_{j/i} Q_{j/k}}$$

L'enveloppe de ces deux résultats constitue presque l'état actuel de la minoration (voir tout de même [4]).

On trouvera dans [23] une minoration de la probabilité que le véritable mot d'entrée ne figure pas dans une liste de longueur donnée des mots les plus vraisemblables au regard de la sortie, résultat utilisé dans [14] pour analyser la principale faiblesse du décodage séquentiel (à rapprocher du point 2° évoqué à propos d'ALOHA). Enfin, dans [13], on trouvera des minoration valables pour tous les  $N$ , ce qui est peut-être plus intéressant dans d'autres domaines de la statistique que le codage.

2° Problème : Une source émet une suite infinie de lettres dans un alphabet fini  $J$ , de façon aléatoire, mais au moins stationnaire. On veut traduire cette source dans un (peut-être) autre alphabet  $K$ , et la qualité de cette traduction est sanctionnée par un critère de distorsion "context-free"  $d(j,k)$ . On additionne les distorsions sur les paires de lettres homologues du mot original et du mot traduit.

On fixe la vitesse  $R$  de la traduction : par exemple, si on code par blocs de longueur  $N$  grande, et si le vocabulaire traduction est constitué de  $M$  mots que les statistiques de la source conduisent à utiliser avec la même fréquence,  $R = \frac{1}{N} \log M$  ; de façon plus générale - quoique n'englobant pas le cas précédent, si le procédé de traduction, et donc la traduction, sont stationnaires, c'est l'entropie de la traduction que l'on fixe,  $\lim_{n \rightarrow \infty} \frac{1}{n} H(Y_n)$ , où  $Y_n$  désigne le bloc de longueur  $n$  de la traduction - voir [20] pour comprendre la signification de l'entropie dans le cas ergodique ; en fait, si l'on code par blocs continus (ou par bloc glissant) on fixe directement  $R = \frac{1}{N} \log M$ , car lorsqu'on voudra optimiser la distorsion, l'égalité des fréquences sera ipso facto vérifiée. On veut alors connaître ce qu'on peut faire de mieux comme probabilité qu'un mot source ne puisse être traduit avec une distorsion inférieure à  $D$  (par lettre, en moyenne) : dans le cas du codage par blocs, on veut minimiser

$$P(x \in J^N \mid \min_{y \in \text{code}} \frac{1}{N} \sum d(x_i, y_i) > D) .$$

Comme pour le premier problème, on trouve dans la littérature deux types de résultats : ou bien la source  $S$  est seulement supposée ergodique, voire stationnaire, et on peut donner une distorsion limite  $D^S(R)$  au-dessus de laquelle la probabilité ci-dessus peut être rendue arbitrairement petite (en jouant sur  $N$ , et sur le code ; voir [6], [9], [12]). Ou bien on évalue asymptotiquement cette probabilité, mais après s'être placé dans un cas particulier : celui d'une source émettant des lettres i.i.d. (cf. [2], [19], [24]). Nous allons voir qu'il ne doit pas être difficile de faire un peu mieux.

Le résultat sur  $D^S(R)$ , ou plutôt sur la fonction réciproque  $R^S(D)$ , est le suivant :  $n$  donné, on considère les lois  $\mu_n$  sur  $J^n \times K^n$  de marginale  $\pi_J \mu_n$  sur  $J^n$  la loi de la source, et telles que la distorsion moyenne  $E_{\mu_n} \frac{1}{n} \sum d(x_i, y_i)$  soit inférieure à  $D$ . On regarde le minimum parmi ces lois de  $\frac{1}{n} I(\mu_n)$ , où

$$I(\mu_n) = H(\pi_J \mu_n) + H(\pi_K \mu_n) - H(\mu_n) .$$

Soit  $R_n^S(D)$  ce minimum, alors  $R^S(D) = \lim_{n \rightarrow \infty} R_n^S(D)$ .

Les preuves de ce théorème commencent en examinant le cas où la source émet des lettres i.i.d., cas où on a en fait  $\lim R_n^S(D) = R_1^S(D)$ , et où l'on gagne à exprimer le problème de minimisation sur des probabilités de transition  $Q$  de  $J$  dans  $K$  plutôt que sur des lois jointes

$$R^P(D) = \inf_{Q \in \mathcal{Q}} \sum_{j,k} p_j Q_{jk} \log \frac{Q_{jk}}{\sum_j Q_{jk} p_j}$$

où  $\mathcal{Q} = \{Q \mid D(Q/p) = \sum_{j,k} p_j Q_{jk} d(j,k) \leq D\}$ .

Démontrons (en nous inspirant de [24]) que  $R^P(D)$  est inférieur à l'expression proposée : cela nous donnera l'idée de la formule pour une source markovienne, voire  $n$ -markovienne (la minoration de  $R^S(D)$  est claire, voir [12]).

Avec une probabilité aussi voisine de 1 que l'on veut (en augmentant la taille  $N$  des blocs sources codés), les proportions des lettres  $j$  dans les blocs sources sont arbitrairement proches des  $p_j$ . Nous allons recouvrir l'ensemble des blocs de longueur  $N$  ayant une telle "composition"  $p$  par des boules de rayon  $D$  (au sens de la distorsion) ; les centres de ces boules seront des mots tirés au hasard (lettres i.i.d. de loi  $q$  à choisir ; mots indépendants). Tout tient alors dans le

Lemme : étant donné un mot  $X$  de composition  $p$ , la probabilité que la boule tirée comme indiqué ci-dessus contienne  $X$  est, pour le  $q$  le plus favorable, "de l'ordre de"  $\exp - NR^P(D)$ .

Preuve : la loi de la distorsion est le mélange, dans les proportions  $p_j$ , des lois affectant à  $d(j,k)$  le poids  $q_k$  ; son log Laplace est donc

$\sum_j p_j \log \sum_k q_k \exp sd(j,k)$  ; d'où la transformée de Cramer

$$\sup_{s (\leq 0)} [sD - \sum_j p_j \log \sum_k q_k \exp sd(j,k)] = \inf_{D(Q/p) \leq D} \sum_j p_j Q_{jk} \log \frac{Q_{jk}}{q_k}$$

(pour comprendre cette dernière égalité, penser : en tirant sous  $q$ , quelle est la probabilité d'obtenir un "canal empirique"  $Q$  de distorsion  $\leq D$  ?). La minimi-

sation en  $q$  donne alors le résultat énoncé.

On peut maintenant avancer sans difficulté une majoration pour  $R^S(D)$  valable dans un cas plus général, disons, pour ne pas alourdir les notations, celui d'une source 1-markovienne. Soit  $c$  la loi jointe de deux lettres sources consécutives. Nous nous proposons de tirer des mots de code avec une probabilité de transition bien choisie. Appelons  $M^c$  l'ensemble des lois sur  $J_1 \times J_2 \times K_1 \times K_2$  (avec  $J_1 = J_2 = J$ ,  $K_1 = K_2 = K$ ) de marginale  $c$  sur  $J_1 \times J_2$ , et de même marginale  $\hat{\mu}$  sur  $J_1 \times K_1$  et sur  $J_2 \times K_2$ ; appelons  $\bar{\mu}$  la marginale de  $\mu$  sur  $K_1 \times K_2$ . Pour  $\mu \in M^c$ , posons

$$D(\mu) = \sum \hat{\mu}_{jk} d(j,k)$$

$$I(\mu) = \int d\mu \log \frac{d\mu(k_2/j_1, j_2, k_1)}{\bar{d}\mu(k_2/k_1)}$$

Alors

$$R^S(D) \leq \inf_{D(\mu) \leq D} I(\mu),$$

et l'égalité paraît plausible (sur ce sujet, voir [1] et [11]).

Revenons aux sources i.i.d. Il résulte du lemme qu'il existe un code  $(N,R)$  tel que le pourcentage des mots sources de composition  $p'$  codés avec une distorsion  $\geq D$  soit

$$\begin{aligned} &\leq (1 - \exp N(-R^{p'}(D) - o(N)))^{\exp NR} \\ &\leq \exp - \exp NR \exp N(-R^{p'}(D) - o(N)) \\ &= \exp(-\exp N(R - R^{p'}(D) - o(N))). \end{aligned}$$

Toujours d'après [24], on minimise la probabilité de dépassement de la distorsion  $D'$  à rendement  $R$  donné en mélangeant les codes obtenus ci-dessus pour toutes les compositions  $p'$  telles que  $R^{p'}(D) < R$  (le nombre de  $p'$  à envisager croît comme une puissance de  $N$ , ce qui est négligeable devant la taille de chaque code, qui est exponentielle en  $N$ ). La probabilité étudiée est alors,

à un terme négligeable près, majorée par  $\sum p'_j |R^{p'_j}(D) \geq R$  Probabilité que un N-mot de la source soit de composition  $p'$ , quantité dont  $-1/N$  le logarithme, que nous appellerons  $F(R,D)$ , vaut

$$\inf_{R^{p'}(D) \geq R} \sum p'_j \log \frac{p'_j}{p_j}$$

Deux remarques :

1°) Nous n'avons guère été soigneux parce que nous savons que  $F(R,D)$  dépend continûment de  $R$  ; ceci tient à la convexité de  $F$ , qui ne se voit pas sur la forme indiquée, mais sous une autre expression due à Blahut [3]

$$F(R,D) = \sup_Q \inf_{\substack{I(Q/p') \geq R \\ D(Q,p') \geq D}} I(p'/p)$$

2°) Il est clair que la suite des arguments peut s'étendre aux sources markoviennes, puisque avec l'exposé n° VI, on sait calculer la probabilité qu'un N-mot source ait une loi jointe vérifiant certaines contraintes.

Pour terminer, indiquons comment l'on voit (d'après [20]) que  $\exp - NF(R,D)$  minore, pour tous les codes  $(N,R)$  imaginables, la probabilité de dépassement de la distorsion  $D$ . On part du fait que, si  $q$  est une probabilité sur  $J$ , et  $\mathcal{M}$  est l'ensemble des mots distordus par un code  $(N,R)$ , si  $R < R^q(D)$ , alors  $q(\mathcal{M}) \geq \text{Cste} > 0$ . Utilisant toujours le même argument, pour tout  $p'$  tel que  $R^{p'}(D) > R$ , on a

$$\begin{aligned} p(\mathcal{M}) &\geq p(\mathcal{M} \cap \text{composition} \simeq p') \\ &\geq p'(\mathcal{M} \cap \text{composition} \simeq p') \exp(-N(I(p'/p) + o(N))) \end{aligned}$$

Puisque  $p'(\mathcal{M})$  est minoré, le premier terme l'est aussi, d'où le résultat.

Bibliographie :

- [1] BERGER T. : Explicit bounds to  $R(D)$  for a binary symmetric Markov source,  
IEEE IT 23 , (1977), p. 52
- [2] BLAHUT R. : Hypothesis testing and information theory  
IEEE IT 20, n° 4 (1974), p. 405
- [3] BLAHUT R. : Information bounds of the Fano-Kullback type  
IEEE IT 22, n° 4 (1976), p. 410
- [4] BLAHUT R. : Composition bounds for channel block codes  
IEEE IT 23, n° 6 (1977), p. 656
- [5] COTTRELL M., FORT J.C., MAGOUIRES G.  
Colloque GRETSI, 1979
- [6] DAVISSON L., PURSLEY M. : A direct proof of the coding theorem for  
discrete sources with memory  
IEEE IT 21, n° 3 (1975), p. 310
- [7] FAYOLLE G., GELENBE E., LABETOULLE J. : Stability and optimal control of  
the packet switching broadcast channel  
J.A.C.M., vol. 24, n° 3 (1977), p. 375
- [8] GALLAGER R. : A simple derivation of the coding theorem and some appli-  
cations  
IEEE IT 11, p. 3
- [9] GALLAGER R. : Information theory and reliable communication,  
New York, Wiley 1968
- [10] GOVER D., SHEDLER G. : Approximate models for processor utilization  
in multiprogrammed computer systems,  
Siam J. Comput. vol. 2, n° 3 (1973), p. 183
- [11] GRAY R. : Information rates of autoregressive processes  
IEEE IT 16 (1970), p. 412
- [12] GRAY R., NEUHOFF D., OMURA J. : Process definitions of distortion rate  
functions and source coding theorems  
IEEE IT 21, n° 5 (1975), p. 524

- [13] HAJI R., NEWELL G. : A relation between stationary queue and waiting time distributions  
J. Appl. Probability 8 (1971), p. 617
  
- [14] JACOBS I., BERLEKAMP E. : A lower bound to the distribution of computation for sequential decoding  
IEEE IT 13, n° 2 (1967), p. 167
  
- [15] KIEFFER J. : On sliding block coding for transmission of a source over a stationary nonanticipatory channel  
Inf. and Control 35 (1977), p. 1
  
- [16] KLEINROCK L. : Queuing systems , vol. 2 (computer applications)  
John Wiley & Sons, 1976
  
- [17] LAM S.S. : Packet switching in a multi-access broadcast channel with applications to satellite communication in a computer network,  
Report UCLA-ENG 7429, mars 1974
  
- [18] LUGANNANI R. : Intersymbol interference and probability of error in digital systems  
IEEE IT 15, n° 6 (1969), p. 682
  
- [19] MARTON K. : Error exponent for source coding with a fidelity criterion  
IEEE IT 20, n° 2 (1974), p. 197
  
- [20] ORNSTEIN D. : Ergodic theory, randomness, and dynamical systems  
Yale Mathematical monographs n° 5 (1974)
  
- [21] PARTHASARATHY K. : On the integral representation of the rate of transmission of a stationary channel,  
Illinois J. Math. 5 (1961), p. 299
  
- [22] QURESHI S., NEWHALL E. : An adaptive receiver for data transmission over time-dispersive channels  
IEEE IT 19, n° 4 (1973), p. 448
  
- [23] SHANNON C., GALLAGER R., BERLEKAMP E. : Lower bounds to error probability for coding on discrete memoryless channels I et II  
Inf. and Control 10 (1967), p. 65 et p. 522



