SUR L'APPROXIMATION DE FAMILLES DE MESURES PAR DES FAMILLES GAUSSIENNES

BY

LUCIEN LE CAM

TECHNICAL REPORT NO. 38

NOVEMBER 1984

RESEARCH PARTIALLY SUPPORTED BY
NATIONAL SCIENCE FOUNDATION GRANT MCS84-03239

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

Sur l'approximation de familles de mesures
par des familles Gaussiennes

Par Lucien Le Cam
Université de Californie, Berkeley

1. Introduction

Le présent travail est une contribution à l'étude du problème suivant: Soient $X_1, X_2, \ldots, X_n$ des observations dont la loi (simultanée) appartient à une famille $\mathcal{E}_n$ à $\{P_{\theta,n}; \theta \in \Theta_n\}$. Sous quelles conditions, et en quel sens, peut-on affirmer que la famille $\mathcal{E}_n$ est proche d'une famille Gaussienne $\mathcal{G}_n$?

Le terme "proche" sera entendu au sens de la distance Δ introduite dans l'article [11] (Le Cam, 1964). Il ne s'agit donc pas de savoir si telle on telle fonction des observations a une distribution proche d'une loide de Moivre-Laplace mais de savoir si les fonctions de risque disponibles à partir de $\mathcal{E}_n$ sont proches de celles disponibles sur $\mathcal{G}_n$. Les définitions précises de la distance Δ et des familles Gaussiennes sont rappelées en Section 2.

L'étude qui suit est une étude locale, puisque nous ne considérons que des familles $\mathcal{E}_n$ où des paires $(P_{s,n}, P_{t,n})$ extraites de $\mathcal{E}_n$ ne se séparent pas entiérement. Par contre, la situation considérée ici est "non paramétrique" en ce sens que les ensembles Θ_n peuvent ètre totalement arbitraires. Le seule exception à cela est le Théorème 4.3 qui est un théorème global où les Θ_n sont assujettis à des conditions qui restreignent leur dimension métrique.

*Research supported by National Science Foundation Grant MCS84-03239.

Nous avons traité séparément, en Section 4 le cas d'observations indépendantes équidistribuées parce que le problème d'approximation s'y pose de façon très claire, et aussi parce que les espaces de Hilbert qui interviennent dans les approximations Gaussiennes y prennent une forme concréte et agréable. Toutefois les démonstrations principales se font aussi facilement pour le cas d'observations qui ne sont ni indépendantes, ni équidistribuées mais seulement sujettes aux conditions de negligibilité des Sections 5 et 6. Ces démonstrations ne sont donc pas repétées en Section 4.

De façon plus précise, le travail est organisé de la façon suivante. La Section 2 rappelle les définitions nécessaires de distances. Elle contient aussi la définition des familles Gaussiennes et quelques remarques sur la structure des noyaux de covariance qui peuvent intervenir dans l'approximation d'une famille $\mathcal{E}_n$ donnée.

La Section 3 a pour objet de situer le présent travail par rapport à d'autres études qui considérent aussi des approximations Gaussiennes. Le point principal est que ces autres études utilisent des hypothèses, telles que les hypothèses dites LAN, qui sont des salades niçoises mélangeant les conditions d'approximabilité Gaussienne avec des paramétrisations spéciales. Elles ne peuvent donc couvrir le cas général et laissent même échapper des cas très simples où les observations sont indépendantes équidistribuées.

En Section 4 nous démontrons quelques théorèmes d'approximation pour des observations indépendantes équidistribuées et enonçons une conjecture: La condition de Lindeberg qui est évidemment nécessaire pour la validité locale de l'approximation Gaussienne est peut-être aussi suffisante. Enfin nous donnons un théorème d'approximation globale qui semble peu connu, mais très utilisable.

La Section 5 commence l'étude de situations où les observations ne sont ni indépendantes, ni équidistribuées mais sujettes à des conditions inspirées de la condition de Lindeberg de la Section 4. On y donne des résultats qui permettent de tronquer les rapports de vraisemblance et d'obtenir une approximation de type exponentiel semblable à celle utilisée en Section 4. L'utilisation de telles approximations pour l'étude de la construction d'estimateurs y est briévement décrite.

La Section 6 reprend les hypothèses de la Section 5 et montre que, sous une hypothèse de mesurabilité supplémentaire, on peut approcher l'expérience par une expérience "Gaussienne mixte". Il s'agit là d'une famille de mesures dont les rapports de vraisemblance ont la même structure que ceux d'une famille Gaussienne excepté pour le fait que ce qui joue le rôle d'une matrice de covariance est maintenant aléatoire. Si l'on prend les distributions conditionnelles étant données ces matrices, la famille devient Gaussienne.

N'ayant pu démontrer que l'approximation est valable pour la distance forte Δ, nous montrons qu'elle est valable pour des distances obtenues en considérant des problèmes de décision où les décisions possibles sont ou bien en nombre fini fixe, ou situées dans un espace vectoriel de dimension finie fixée, ou dans une boule Hilbertienne quelconque, la fonction de perte étant alors le carré de la norme.

Nous donnerons, dans un travail ultérieur des applications des résultats des Sections 5 et 6.

Le présent article à été lu avec un soin minutieux par un rapporteur anonyme. Je lui dois tous mes remerciements pour avoir corrigé mes calculs et aussi pour avoir, ici et là, amélioré mes bornes.

2. Distances entre expériences - Expériences Gaussiennes

Soit Θ un emsemble. On appelle expérience indexée par Θ toute structure formée d'une tribu A de parties d'un espace X et d'une application $\theta \rightsquigarrow P_\theta$ de Θ dans l'espace des mesures de probabilité sur A. Soient $E = \{P_\theta;\ \theta \in \Theta\}$ et $F = \{Q_\theta;\ \theta \in \Theta\}$ deux expériences indexées par le même Θ mais peut-être portées par des tribus différentes. On a défini dans [11] un écart $\Delta(E,F)$ par une méthode qui, mis à part des considérations techniques de peu d'interét ici, revient à dire ceci: Pour toute fonction de perte W telle que $0 \leq W \leq 1$, pour toute fonction de risque r disponible sur Θ en utilisant l'une des expériences, il existe une fonction de risque r' disponible sur l'autre expérience et telle que $r'(\theta) \leq r(\theta) + \Delta$. Pour définir cet écart, le travail [11] utilise une définition généralisée de ce que l'on doit entendre par régle de décision mais la différence entre cette définition généralisée et la définition habituelle est sans grande importance pour la suite. Cela étant entendu, on peut donner pour la pseudo distance Δ plusieur formules différentes. La formule utilisée ici est la suivante. Pour une mesure de probabilité π à support fini sur Θ et pour une famille $C = \{c(j,\theta);\ j \in J,\ \theta \in \Theta\}$ de nombres $c(j,\theta) \in [0,1]$ soit $\nu(\pi,C,E)$ la norme L_1 (= variation totale) de la mesure $\sup_{j \in J} \sum_\theta c(j,\theta)\pi_\theta P_\theta$. On a alors

$$\Delta(E,F) = \sup_{\pi,C} |\nu(\pi,C,E) - \nu(\pi,C,F)| ,$$

le supremum étant pris sur toutes les probabilités π et tous les systèmes C. La quantité $1 - \nu(\pi,C,E)$ est le risque de Bayes pour la loi a priori π et la fonction de perte $1 - c$.

Nous utiliserons aussi une pseudo-distance Δ_m obtenue exactement de la même façon mais en prenant seulement des systèmes C où la cardinalité de l'ensemble J ne dépasse pas l'entier m.

L'interet de cette distance a été discuté par Torgersen [20].

On dit que E et F sont equivalentes ou qu'elles ont le même type si $\Delta(E,F) = 0$. Sur l'ensemble des types d'expériences indexées par Θ la fonction Δ devient une distance.

Pour un Θ fixé on peut aussi utiliser Δ pour définir une topologie faible: des F convergent faiblement vers un E si pour toute partie finie $F \subset \Theta$ les expériences F_F restreintes à F sont telles que $\Delta(F_F, E_F) \rightarrow 0$.

Une expérience $G = \{G_\theta;\ \theta \in \Theta\}$ est dite Gaussienne si elle satisfait aux deux conditions suivantes:

1) Pour toute paire (s,t) d'éléments de Θ les mesures G_s et G_t sont mutuellement absolument continues.

2) Il existe un $s \in \Theta$ tel que, si $\Lambda(t,s) = \log(dG_t/dG_s)$ et si les distributions sont induites par la mesure G_s, alors le processus $t \leadsto \Lambda(t,s)$ est un processus stochastique Gaussien.

Les proprietés suivantes des expériences Gaussiennes sont bien connues et faciles à établir. Designons par E_s l'espérance mathématique associée à la mesure G_s.

a) Si le processus $t \leadsto \Lambda(t,s)$ est Gaussien pour G_s, il est aussi Gaussien pour tout G_θ. Sa covariance est indépendante de la mesure G_θ utilisée pour induire les distributions.

b) $E_s\Lambda(t,s) = -\frac{1}{2}$ variance de $\Lambda(t,s)$

c) $E_\theta\Lambda(t,s) - E_s\Lambda(t,s)$ est égal à la covariance de $\Lambda(t,s)$ et $\Lambda(\theta,s)$

d) Toute expérience Gaussienne est équivalente à une expérience obtenue de la façon suivante:

Soit H un espace de Hilbert et soit X un processus Gaussien linéaire canonique de H. Cela veut dire que pour tout $y \in H$ on a $E\langle y,X\rangle = 0$ et $E|\langle y,X\rangle|^2 = \|y\|^2$. Ce processus définit sur H une mesure cylindrique G'_0. Soit G'_z le mesure cylindrique distribution du processus $y \rightsquigarrow \langle y,z\rangle + \langle y,X\rangle$. La famille $\{G'_z;\ z \in H\}$ est une expérience Gaussienne. Pour toute expérience Gaussienne $G = \{G_\theta;\ \theta \in \Theta\}$ il existe un espace de Hilbert H et une application ϕ de Θ dans H telle que G soit équivalente à $G' = \{G'_{\phi(\theta)};\ \theta \in \Theta\}$.

La construction d'une représentation de cette nature fait intervenir des espaces dont nous aurons à nous servir à plusieurs reprises.

Soit $M(\Theta)$ l'espace des mesures réelles à support fini sur Θ et soit $M_0 = M_0(\Theta)$ le sous espace linéaire des $\mu \in M(\Theta)$ telles que $\mu(\Theta) = 0$. Puisque les mesures $\{G_\theta: \theta \in \Theta\}$ sont mutuellement absolument continues, pour toute paire (s_1,s_2) d'éléments de Θ et pour tout $\mu \in M_0$ les variables aléatoires $\int \Lambda(t,s_i)\mu(dt)$, $i=1,2$ sont presque surement égales. Pour une expérience Gaussienne la variance $V(\mu)$ de $\int \Lambda(t,s)\mu(dt)$ ne dépend pas de la loi G_θ utilisée pour induire les distributions. C'est le carré, $V(\mu) = \|\mu\|_G^2$ d'une certaine seminorme de type Hilbertien sur M_0. On peut identifier à zéro les μ tels que $\|\mu\|_G = 0$ puis compléter l'espace pour obtenir un espace de Hilbert H. La représentation de Θ dans H se fait en prenant un point arbitraire s comme origine et en envoyant Θ dans M_0 par l'application $\theta \rightsquigarrow \delta_\theta - \delta_s$ où δ_θ est la mesure de Dirac à θ.

Nous aurons besoin des faits suivants, pour des éléments λ et μ de M_0.

e) La covariance de $\int \Lambda(t,s)\mu(dt)$ et $\int \Lambda(t,s)\lambda(dt)$ est égale à $-\frac{1}{2}\iint \|\delta_t - \delta_u\|_G^2 \lambda(du)\mu(dt)$

f) $\|\delta_t - \delta_u\|^2 = -8 \log \int \sqrt{dG_u dG_t}$

Il résulte de ces égalités que la structure de l'expérience Gaussienne G est entiérement determinée par les affinités $\int \sqrt{dG_u dG_t}$.

L'objet principal du présent travail est d'étudier l'approximation de certaines expériences $E = \{P_\theta;\ \theta \in \Theta\}$ par des expériences Gaussiennes. La question se pose donc de savoir quelles expériences Gaussiennes peuvent être considérées comme approximations possibles. Les résultats disponibles dans cette direction sont très maigres. L'un d'entre eux peut s'exprimer comme suit:

LEMME 2.1. *Posons* $q^2(s,t) = -8 \log \int \sqrt{dP_s dP_t}$ *et supposons qu'il existe un nombre* $b \in]1,\infty[$ *tel que* $q^2(s,t) \leq b$ *pour toute paire d'éléments de* Θ. *Alors pour tout* $\varepsilon > 0$ *il existe un nombre* δ *dépendant seulement de* b *et* ε *tel que, si l'expérience Gaussienne* $G = \{G_\theta;\ \theta \in \Theta\}$ *satisfait à* $\Delta(E,G) < \delta$, *on a*

$$\left| q^2(s,t) + 8 \log \int \sqrt{dG_s dG_t} \right| < \varepsilon$$

pour toute paire (s,t) *d'éléments de* Θ.

Démonstration. Pour des expériences binaires $\{P_s, P_t\}$, l'affinité $\int \sqrt{dP_s dP_t}$ est une fonction continue pour la distance Δ. Voir, par exemple [20]. Le résultat est une conséquence immédiate de cette continuité et de la continuité uniforme des logarithmes en dehors d'un voisinage de zéro.

Le résultat du Lemme 2.1 peut se mettre sous une forme plus suggestive. Prenons pour cela l'ensemble B_1 des $\mu \in M_0$ dont la variation totale ne dépasse pas l'unité. On peut definir sur M_0 une forme quadratique K par la formule $K(\mu) = -\frac{1}{2}\iint q^2(s,t)\mu(ds)\mu(dt)$. Pour toute paire (λ,μ) d'éléments de B_1 on doit avoir $\left|K(\lambda-\mu) - \|\lambda-\mu\|_G^2\right| < \varepsilon$.

Il est donc tentant d'utiliser la forme K pour définir une approximation Gaussienne. Toutefois ceci n'est généralement pas possible. En effet la forme K n'a aucune raison d'être positive. Le plus souvent la fonction q ne satisfait même pas aux inégalités triangulaires. On sera donc forcé d'utiliser d'autres formes.

Il y a d'autres problèmes dont la solution nous échappe encore. Supposons par exemple que G^1 et G^2 soient deux expériences Gaussiennes toutes les deux très proches de l'expérience E. Alors $\Delta(G^1,G^2)$ doit être petit. Chacun des G^i définit sur M_0 une seminorme Hilbertienne $\|\mu\|_{G^i}$. Le raisonnement précédent montre que, au moins sur B_1, les deux normes $\|\mu\|_{G^i}$ doivent différer peu. On pourait dire que la correspondance canonique entre B_1 muni de $\|\cdot\|_{G^1}$ et B_1 muni de $\|\cdot\|_{G^2}$ est "presque isométrique," mais ce terme peut-être entendu en des sens très différents. En voici deux:

Une correspondance Γ est isométrique à ε près si pour toutes paires $(x_i,y_i) \in \Gamma$ on a $\left| \|x_1-x_2\|_{G^1} - \|y_1-y_2\|_{G^2} \right| < \varepsilon$.

Une correspondance Γ est à distance ε d'une isométrie s'il existe une isométrie J telle que la distance de Hausdorff entre les graphes Γ et J est inférieure à ε.

Pour cette deuxième définition il est plus naturel de se placer dans le produit des espaces de Hilbert H_i complétions de M_0 pour les seminormes attachées aux G^i. Soit donc Θ_i l'image de Θ dans H_i par l'application $\theta \rightsquigarrow \delta_\theta - \delta_s$.

Sous la condition $q^2 \leq b$ du Lemme 2.1 avec $\Delta(G^i,E) < \delta/2$ la correspondance canonique Γ entre Θ_1 et Θ_2 est bien isométrique à ε-près mais nous ne savons pas si elle est à petite distance d'une isométrie.

Dans l'ordre inverse il est facile de démontrer que si Γ est à petite distance d'une isométrie alors $\Delta(G^1,G^2)$ doit être petite. En effet soient x et y deux éléments de M_0 ou des espaces H_i les mesures G_x et G_y correspondantes sont telles que

$$\frac{1}{2}\|G_x-G_y\| \leq 2\left\{1 - \exp\{-\frac{1}{8}\|x-y\|^2_{G^i}\}\right\}^{1/2}$$
$$\leq \frac{1}{\sqrt{2}}\|x-y\|_{G^i} .$$

On peut donc remplacer Θ_2 par un ensemble isométrique à Θ_1 sans changer beaucoup l'expérience G^2. Une telle modification ne semble par possible sous la condition faible d'être isométrique à ε près.

Il existe bien entendu des cas où les deux définitions sont essentiellement équivalentes (cas précompacts par exemple) mais cela ne va pas très loin.

Le problèmes est d'un certain interêt pour nos questions d'approximations Gaussiennes. En effet, supposons qu'il existe des expériences Gaussiennes G^i telles que, suivant un filtre, les distances $\Delta(E,G^i)$ tendent vers zéro. Nous savons que les normes correspondantes $\|\cdot\|_{G^i}$ doivent avoir une différence tendant vers zéro sur les boules du type B_1 mais nous ne savons pas plus.

En particulier, il est concevable, quoique peu probable, que les familles satisfaisant à une condition de Lindeberg de la Section 4 admettent des approximations Gaussiennes mais que ce ne soient pas celles qui sont décrites dans la Section 4.

3. L'utilisation d'approximations Gaussiennes

La raison même de l'introduction de la distance Δ est celle qui a déjà été donnée en Section 2: Si deux expériences $\mathcal{E}_i$ sont telles que $\Delta(\mathcal{E}_1, \mathcal{E}_2) \leq \varepsilon$, alors, pour toute fonction de perte W telle que $0 \leq W \leq 1$, pour toute fonction de risque r disponible sur l'un des $\mathcal{E}_i$, il existe une fonction de risque r' disponible sur l'autre expérience telle que $|r-r'| \leq \varepsilon$. On peut donc transférer des connaissances valables pour un $\mathcal{E}_i$ en connaissances approximatives pour l'autre.

Il se trouve que les expériences Gaussiennes sont peut-être celles qui ont été le plus étudiées en statistique. On connait donc les propriétés de leurs fonctions de risque dans beaucoup de cas particuliers. Si donc on sait que $\Delta(\mathcal{E}, \mathcal{G})$ est petite, et si l'on connait pour $\mathcal{G}$, par exemple, une borne supérieure ou une borne inférieure des risques minimax, cela donne une connaissance analogue, approchée, pour $\mathcal{E}$.

On doit noter toutefois que notre connaissance des propriétés des expériences Gaussiennes, n'est pas tellement bonne et que cette méthode n'est pas toujours facilement utilisable.

Les cas les mieux connus sont ceux où la représentation canonique de Θ dans l'espace de Hilbert associé à $\mathcal{G} = \{G_\theta; \theta \in \Theta\}$ est un sous espace linéaire. Dans des cas plus généraux on peut tout de même obtenir certains résultats. Par exemple, sous des conditions sur la dimension métrique, on peut utiliser les bornes supérieures de Le Cam [13] ou de Birgé [3]. Ces bornes sont bien applicables ici puisque $\mathcal{G}$ est infiniment divisible.

On peut aussi obtenir des bornes inférieures en utilisant le lemme de Fano (voir Ibragimov-Has'minskii [7], page 323) ou les résultats d'Assouad [1]. Le lemme de Fano est particuliérement aisé à appliquer ici puisque (avec les notations de la Section 2) le nombre de Kullback-Leibler pour une paire (G_s,G_t) est égal à $\frac{1}{2}\|\delta_s-\delta_t\|_G^2$, alors qu'il pourrait fort bien être infini pour la paire (P_s,P_t).

Il y a grand nombre de cas où la résolution du problème statistique devient aisée, ou même quelquefois triviale, si l'on utilise les approximations Gaussiennes de cette manière. Ceci s'applique, par exemple, à l'étude des estimateurs BAN et des tests $C(\alpha)$ de Neyman, aux résultats de Chernoff de 1954 [4] ou aux résultats de Begun, Hall, Huang et Wellner [2].

La méthode même remonte au travail fondamental de Wald [21] paru en 1943. Il est vrai que Wald utilise des familles hetéroscédastiques qui ne sont donc pas Gaussiennes en notre sens, mais nous verrons (Section 4, Théorème 4.3) que dans son cas on peut aussi utiliser des familles Gaussiennes en notre sense, donc homoscédastiques.

Une grande partie de la littérature statistique utilise des moyens beaucoup plus faibles. On se contente de regarder des suites $\{\mathcal{E}_n\}$, $n=1,2,\ldots$ d'expériences $\mathcal{E}_n = \{P_{\theta,n};\ \theta \in \Theta\}$ indexées par un ensemble Θ <u>indépendant</u> <u>de</u> n dans des cas où les $\mathcal{E}_n$ convergent <u>au</u> <u>sens</u> <u>faible</u> vers une expérience Gaussienne. L'ensemble Θ fixe peut-être différent de l'ensemble utilisé pour la paramétrisation originale. Souvent c'est un espace Euclidien rattaché à l'espace original par des transformations linéaires (par exemple la transformation $\theta = t\sqrt{n}$).

La convergence faible permet déjà d'obtenir certains résultats. Elle suffit à établir le théorème du minimax asymptotique de Hájek et Le Cam (voir [14] pour la forme générale et quelques références). Elle suffit aussi à établir le théorème de convolution de Hájek [6]. Notons toutefois que cette technique a des points faibles.

D'une part elle ne peut servir à obtenir des bornes supérieures pour les fonctions de risque. D'autre part elle est intimement liée au fait que Θ doit rester fixe et que l'on doit entiérement passer à la limite. Il n'est pas toujours possible ou désirable de fixer Θ. Par exemple on peut vouloir accroitre le nombre des paramétres quand le nombre d'observations croît. Il est certainement très souvent indésirable de faire simplement "tendre n vers l'infini".

Il est bien vrai qu'un résultat du type $\Delta(\mathcal{E},\mathcal{G}) < \varepsilon$ pour un expérience Gaussienne $\mathcal{G}$ nous dit que les fonctions de risque de $\mathcal{E}$ sont proches de celles de $\mathcal{G}$ mais ne nous dit rien sur la manière d'obtenir ces fonctions. Toutefois il existe une théorie fort étendue de la construction d'estimateurs, tests, ou intervalles de confiance dans de tels cas. Cette théorie a surtout été exposée dans le cadre des conditions dites LAN dont il nous faut dire quelques mots.

La partie essentielle des conditions LAN de Le Cam (1960) pourrait être resumée comme suit: Prenons une suite d'expériences $\mathcal{E}_n = \{P_{\theta,n};\ \theta \in \Theta_n\}$ où <u>tout</u> dépend de n. Puisque tout va dépendre de n, il est plus simple de l'omettre pour simplifier les notations.

Nous parlerons donc d'une suite d'expériences $\mathcal{E} = \{P_\theta;\ \theta \in \Theta\}$. Il y aura bien quelques termes qui ne dépendent pas de n, auquel cas ce sera dit explicitement ou indiqué en disant qu'ils sont <u>fixes</u>.

Soit $q \geq 0$ la fonction définie par $q^2(s,t) = -8 \log \int \sqrt{dP_s dP_t}$. Considérons les conditions suivantes:

(A0) On dispose d'estimateurs auxiliaires $\hat{\theta}$ et on ne considère que les $\theta \in \Theta$ tels que $q(\hat{\theta},\theta)$ reste borné en probabilité pour P_θ.

(A1) Des points θ satisfont à la condition d'approximation Gaussienne locale (A1) si pour tout réel <u>fixe</u> b il existe des expériences Gaussiennes $G_{\theta,b} = \{G_t;\ t \in \Theta,\ q(t,\theta) \leq b\}$ dont la distance à $\{P_t;\ t \in \Theta,\ q(t,\theta) \leq b\}$ tend vers zéro.

(A2) Des θ satisfont à (A2) s'ils satisfont à (A1) pour des expériences Gaussiennes dont les espaces de Hilbert associés ont une dimension linéaire bornée indépendamment de n.

Les conditions LAN de [10] ne sont pas enoncées sous cette forme pour une raison très simple: Elles ont été écrites en Décembre 1957 alors que la distance Δ n'a été introduite qu'en Décembre 1958. Toutefois on peut s'assurer sans grande difficulté que la construction d'estimateurs asymptotiquement exhaustifs de [10] dépend essentiellement seulement des conditions (A0), (A1) et (A2). Elle dépend naturellement beaucoup de (A0). Cette condition a disparu dans ce qu'il est maintenant convenu d'appeler les conditions LAN (voir par exemple Ibragimov-Has'minskii). Sans une condition telle que (A0) on ne peut espèrer obtenir que des résultats tout à fait locaux.

Les conditions LAN de [10] ou [7] font aussi intervenir d'autres espaces. En effet on suppose que la paramétrisation originale des expériences est une paramétrisation Euclidienne fixe et après avoir changé les espaces par des transformations linéaires, on supppose que ces espaces Euclidiens s'appliquent linéairement sur les espaces de Hilbert de la condition (A1). C'est là une restriction accidentelle due à des raisons purement historiques. Toutefois elle élimine certains cas intéressants. Nous en verrons un ci-dessous, mais il y en a de plus intéressants où la paramétrisation initiale est sur un espace fonctionnel de dimension infinie quoique (A2) soit bien satisfaite.

Pour la condition (A0) voir les remarques à la fin de la Section 4.

D'autres conditions souvent utilisées font intervenir des "espaces tangents". Les exemples principaux de telles conditions sont ceux de Koshevnik et Levit [9] et Pfanzagl et Wefelmeyer [18]. On considère une tribu $\mathcal{A}$ et une famille $\mathcal{P}$ de mesures de probabilité sur $\mathcal{A}$. Soit p_0 un élément de $\mathcal{P}$. Les auteurs cités considèrent des applications continues $t \rightsquigarrow p_t$ de $[0,1]$ dans $\mathcal{P}$. Koshevnik et Levit plongent $\mathcal{P}$ dans un espace de Hilbert $\mathcal{H}$ en utilisant une distance h définie par $h^2(p,q) = 4\int(\sqrt{dp} - \sqrt{dq})^2$. On prend p_0 comme origine de $\mathcal{H}$. L'application $t \rightsquigarrow p_t$ donne alors une application $t \rightsquigarrow \sqrt{dp_t} - \sqrt{dp_0}$ dans $\mathcal{H}$ et l'on prend, si elle existe, la demi-droite tangente à cette courbe à l'origine de $\mathcal{H}$. L'ensemble de ces demi-droites est un cône S appelé espace tangent de $\mathcal{P}$ à p_0.

La définition de Pfanzagl et Wefelmeyer est d'un aspect assez différent mais elle revient au même. (Dans [19], Pfanzagl et Wefelmeyer font observer, avec quelque justice, que leur définition est plus facilement adaptable aux problèmes où on s'intéresse à des vitesses de convergence.)

Prenons alors n observations indépendantes équidistribuées dont les distributions individuelles sont données par des mesures $p_{\theta/\sqrt{n}}$, $\theta \in \mathbb{R}$ où $t \rightsquigarrow p_t$ est un des arcs différentiables en zéro décrits ci-dessus. Soit $P_{\theta,n}$ la mesure produit correspondante. On vérifie aisément que les expériences $\{P_{\theta,n};\ 0 \leq \theta \leq b\}$ pour b fixe tendent vers une expérience Gaussienne. Ceci donne, enfin de compte, une expérience Gaussienne $\mathcal{G} = \{G_s;\ s \in S\}$ indexée par le cône tangent S.

La technique des espaces tangents a semble-t-il deux buts. L'un est de remplacer un ensemble compliqué par un espace tangent plus simple.

L'autre est de fournir des approximations Gaussiennes. Cette technique soulève pourtant beaucoup de problèmes.

Soit en effet b un nombre fixe. Soit $\mathcal{P}_n$ la partie de $\mathcal{P}$ formée des p tels que

$$H_n^2(p,p_0) = 4n \int (\sqrt{dp} - \sqrt{dp_0})^2 \leq b^2 .$$

La norme H_n ainsi définie envoie $\mathcal{P}_n$ sur une certaine partie, soit S_n, d'un espace de Hilbert $\mathcal{H}_n$. On peut aussi considérer que le cone tangent S est une partie de $\mathcal{H}_n$ et prendre la partie $S(b)$ de S située dans la boule $B(b) = \{x;\ x \in \mathcal{H}_n,\ \|x\| \leq b\}$. A l'espace $\mathcal{H}_n$ correspond de façon canonique une expérience Gaussienne $\mathcal{G}_n$. Cela donne plusieurs expériences.

L'une d'entre elles, soit $\mathcal{E}_n$ est l'expérience produit pour n observations de lois individuelles contenues dans $\mathcal{P}_n$. De plus, soit $\mathcal{G}'_n$ et $\mathcal{G}''_n$ les restrictions respectives de $\mathcal{G}_n$ à S_n et $S(b)$. Pour $\mathcal{E}_n$ nous considérerons que l'expérience est indexée par l'image S_n de $\mathcal{P}_n$ dans $\mathcal{H}_n$. On peut alors se demander si $\Delta(\mathcal{G}'_n,\mathcal{E}_n) \to 0$ et si $\mathcal{G}'_n$ et $\mathcal{G}''_n$ sont des reparamétrisations d'expériences Gaussiennes proches l'une de l'autre.

Pour la deuxième question l'approximation de $S(b)$ par des arcs différentiables semble donner une réponse affirmative, mais ce n'est qu'une illusion. En effet $\mathcal{G}'_n$ et $\mathcal{G}''_n$ peuvent être très loin d'être presque isométriques, même au sens le plus faible décrit en Section 2. La condition voulue pour cela est la condition utilisée par Chernoff en 1954 [4], à savoir que la distance de Hausdorff entre S_n et $S(b)$ dans $\mathcal{H}_n$ tende vers zéro. Sous cette condition S_n et $S(b)$ peuvent être mis en correspondance par un graphe qui est proche du graphe d'une isométrie.

Cela donne une reparamétrisation pour laquelle G'_n et G''_n deviennent bien proches l'une de l'autre. C'est un cas très spécial. En effet il est parfaitement possible que S soit l'espace de Hilbert H out entier. Alors $S(b)$ est la boule de rayon b de H_n. Toutefois pour $b \geq 1$ et pour une mesure p_0 diffuse, une telle boule contient toujours des éléments z tels que la boule de rayon $1/4$ centrée à z est disjointe de l'image de P.

Prenons par exemple l'espace $L_{2,0}(p)$ de la Section 4 avec p égale à la mesure de Lebesgue sur $[0,1]$. Toute fonction bornée inférieurement de $L_{2,0}(p)$ est "éligible" pour n assez grand au sens défini dans cette Section 4. Soit g_n la fonction égale à $\frac{1}{\sqrt{2}}\,2^n$ sur $[0,\frac{1}{2^{2n}}]$, à $-\frac{1}{\sqrt{2}}\,2^n$ sur $(\frac{1}{2^{2n}},\frac{2}{2^{2n}}]$ et à zéro autrement. C'est un élément de $L_{2,0}(p)$ de norme unité, mais il n'ya pas d'élément éligible dans la boule de centre g_n et de rayon $\frac{1}{\sqrt{2}}[1-\frac{2\sqrt{2n}}{2^n}]$ quoique l'espace tangent soit l'espace $L_{2,0}(p)$ tout entier.

En ce qui concerne la convergence de $\Delta(E_n,G'_n)$ vers zéro, nous n'avons pas à l'heure actuelle de réponse complète. Ce sera l'objet de la Section 4 qui suit. On ne peut avoir une telle convergence que si les S_n satisfont à une condition de Lindeberg uniforme, ce qui n'est pas le cas en général.

Le problème principal associé à l'usage des cônes tangents est qu'il fait intervenir un espace qui n'a pas d'analogue si les observations sont indépendantes mais non équidistribuées, et encore moins si elles ne sont pas indépendantes. Même dans le cas d'observations indépendantes équidistribuées ni les conditions LAN, ni les cônes tangents ne peuvent couvrir tous les cas intéréssants.

Prenons par exemple la famille de translation $\{p_s ; s \in \mathbb{R}\}$ dont les densités par rapport à la mesure de Lebesgue sur la droite sont données par $[1-|x-s|]^+$. Soit $P_{\theta,n}$ la mesure produit pour n observations tirées de $[1-|x-\frac{\theta}{\sqrt{n \log n}}|]^+$. Les conditions LAN sont satisfaites.

Il y a bien une expérience Gaussienne G telle que $\{P_{\theta,n} ; \theta \in \mathbb{R}\}$ converge vers G. Par contre les arcs $s \rightsquigarrow p_s$ n'ont pas de dérivée à zéro et les images $\sqrt{n}(\sqrt{dp_{\theta/\sqrt{n \log n}}} - \sqrt{dp_0})$ ne convergent pas dans l'espace de Hilbert.

Un autre exemple est fourni pas les densités $c \exp\{-|x-s|^\alpha\}$ avec α fixé, $\alpha \in (0,\frac{1}{2})$. Ici les espaces tangents sont réduits à l'origine de l'espace de Hilbert et les conditions LAN ne sont pas satisfaites. Même la condition (A2) de la présente section n'est pas satisfaite. Pourtant si l'on prend les expériences produits $\{P_{s,n} ; s \in \mathbb{R}\} = E_n$ (avec s dans $\mathbb{R}$ tout entier), il y a bien des expériences Gaussiennes G_n telles que $\Delta(E_n, G_n) \to 0$.

Cela résulte par exemple d'une petite modification du Théorème 4.3.

Pour toutes ces raisons, il semble bien nécessaire de reconsidérer de plus près le problème d'approximation, ce que nous allons faire maintenant.

4. Observations indépendantes équidistribuées

Nous nous occuperons ici surtout de suites d'observations indépendantes équidistribuées mais donnerons aussi quelques indications sur le cas de variables indépendantes non équidistribuées. Tout cela sera fait dans un cadre purement local, mis à part le Théorème 4.3.

Pour chaque entier n, soit $A_{j,n}$; $j \in I_n$ une famille de tribus sur des ensembles $X_{j,n}$. Soit Θ_n un ensemble et, pour chaque paire (j,n), soit $p_{\theta,j,n}$; $\theta \in \Theta_n$ une famille de probabilités sur $A_{j,n}$. On considère l'expérience produit $E_n = \{P_{\theta,n};\ \theta \in \Theta_n\}$ avec $P_{\theta,n}$ égal à la mesure produit $\prod_j p_{\theta,j,n}$.

Pour le cas équidistribué, on suppose que tous les systèmes $\{X_{j,n}, A_{j,n}, p_{\theta,j,n}\}$ sont des copies d'un même $\{X_{0,n}, A_{0,n}, p_{\theta,0,n}\}$ et que l'ensemble d'indices I_n est l'ensemble des entiers $I_n = \{1,2,\ldots,n\}$. Même dans ce cas nous supposerons toujours que toutes les pieces des systèmes en question dépendent de l'entier n.

Puisque n se trouve partout, il n'est pas nécessaire de le mettre. Il sera donc supprimé toutes les fois que cela sera possible pour alléger les notations.

Si quelque terme est indépendant de n, cela sera indiqué en disant qu'il est fixe ou en employant des qualificatifs analogues.

Soit donc $E_n = \{P_\theta;\ \theta \in \Theta\}$ un produit de n copies d'une même expérience $E_1 = \{p_\theta;\ \theta \in \Theta\}$. La question se pose de savoir s'il existe des expériences Gaussiennes G_n telles que $\Delta(E_n, G_n)$ tende vers zéro quand $n \to \infty$. Nous considèrerons ici le cas local soumis à la restriction suivante:

Soit h la distance de Hellinger définie par $h^2(s,t) = \frac{1}{2}\int (\sqrt{dp_s} - \sqrt{dp_t})^2$.

(B) Il existe un nombre $b_0 \in (0,\infty)$ fixe tel que $\sup\{nh^2(s,t);$ $(s,t) \in \Theta_n \times \Theta_n,\ n=1,2,\ldots\} \leq b_0$.

Il revient au même de dire que, pour des suites arbitraires $\{(s,t)\}$, $(s,t) \in \Theta_n \times \Theta_n$, les mesures P_s et P_t ne se séparent pas entiérement. Alors, si $\Delta(E_n, G_n) \to 0$, il en est de même pour les paires $\{G_s, G_t\}$ de $G_n = \{G_\theta;\ \theta \in \Theta\}$. Ces paires sont donc contigües et il en est de même des paires $\{P_s\}$, $\{P_t\}$.

On ne perd donc pas vraiment en généralité si l'on impose la condition suivante.

(C) Les paires $\{P_s\}$, $\{P_t\}$, $(s,t) \in \Theta \times \Theta$ sont contigües. Il y a un $p \in \{p_\theta;\ \theta \in \Theta\}$ tel que tous les p_θ soient absolument continues par rapport à p.

Sous ces conditions il est commode d'introduire un espace de Hilbert $L_{2,0}(p)$ qui n'est pas exactement l'espace associé au carré $4nh^2$ issu de la distance de Hellinger mais qui en diffère très peu. L'espace $L_{2,0}(p)$ est l'espace des classes d'équivalence de l'espace $\mathcal{L}_{2,0}(p)$ de fonctions g telles que

(1) $\int g\, dp = 0$

et (2) $\|g\|^2 = \int g^2 dp < \infty$

Toute mesure de probabilité q dominée par p admet une représentation sous la forme

$$dq = [1 - c(\frac{\|g\|}{2\sqrt{n}}) + \frac{g}{2\sqrt{n}}]^2 dp$$

où $g \in \mathcal{L}_{2,0}(p)$ est tel que

(3) $1 - c[\frac{\|g\|}{2\sqrt{n}}] + \frac{g}{2\sqrt{n}} \geq 0$ et $\|g\|^2 \leq 4n$

et où c est l'application de $[0,1]$ dans $[0,1]$ définie par $1 - c(x) = \sqrt{1-x^2}$.

A noter que si la paire p, q était fixe l'élément $g \in \mathcal{L}_{2,0}(p)$ dépendrait encore de n. Tout élément $g \in \mathcal{L}_{2,0}(p)$ qui satisfait à la condition de positivité (3) sera dit <u>éligible</u> (ou éligible à l'étape n, si besoin est).

Au lieu de prendre un ensemble Θ quelconque, on peut supposer que Θ est une partie de $L_{2,0}(p)$ ou, si l'on veut, de $\mathcal{L}_{2,0}(p)$. La distinction entre fonctions et leurs classes d'équivalence étant pénible pour les notations, nous ne la ferons pas. Elle interviendrait pourtant de façon assez sérieuse pour la définition des processus empiriques et du bruit blanc Gaussien utilisés ci dessous, mais il est facile de se convaincre que, même là, cela n'a pas d'importance vraie. En effet la formule donnée en Section 2 pour la distance Δ ne fait intervenir que des termes du type $\|\sup_{j\in J} \sum_g C(j,g)\pi_g P_g\|$ où π est à support fini. Si donc $\Delta(\mathcal{E}_n,\mathcal{G}_n) > \varepsilon$ il y a une partie dénombrable de Θ telle que $\Delta(\mathcal{E}'_n,\mathcal{G}'_n) > \varepsilon$ pour les expériences $\mathcal{E}'_n$ et $\mathcal{G}'_n$ obtenues en remplaçant Θ par cette partie dénombrable. Il est alors visible que le choix d'un élément $g \in \mathcal{L}_{2,0}(p)$ dans sa classe d'équivalence est sans importance.

Si $\Delta(\mathcal{E}_n,\mathcal{G}_n) \to 0$ et si (B) est satisfaite, pour toutes paires (s,t) de Θ les logarithmes $\log \frac{dP_t}{dP_s}$ doivent avoir des lois qui, quand n devient grand, sont approximativement Gaussiennes. Par contiguïté ces lois normales doivent aussi être du type $\mathcal{N}(-\frac{1}{2}\sigma^2,\sigma^2)$. Il en résulte (voir [12] par exemple) que les $g \in \Theta$ doivent satisfaire à une condition du type de Lindeberg qui peut s'exprimer comme suit:

(L) Pour tout $\varepsilon > 0$ fixe

$$\sup_{g \in \Theta} \int g^2 I[|g| > \varepsilon\sqrt{n}]dp$$

tends vers zéro quand $n \to \infty$.

(La condition doit bien être uniforme puisqu'elle doit être satisfaite pour toute suite $\{g_n\}$, $g_n \in \Theta_n$.)

Au moment de la présente rédaction nous ne savons pas si les conditions (B), (C) et (L) sont suffisantes pour entrainer l'existence d'expériences Gaussiennes G_n telles que $\Delta(E_n, G_n)$ tende vers zéro. C'est toutefois une conjecture qui est supportée par les résultats qui suivent.

Nous supposerons désormais que (B), (C) et (L) sont satisfaites. La condition (B) peut alors être mise sous la forme

(B') Il existe un b fixe tel que $\|g\|^2 \leq b$ pour tout n et tout $g \in \Theta_n$.

Une autre condition nécessaire pour l'existence d'expériences Gaussiennes G_n telles que $\Delta(E_n, G_n) \to 0$ est l'existence d'expériences infiniment divisibles F_n telles que $\Delta(E_n, F_n) \to 0$. Ici il n'y a aucun problème. En effet soit F_n l'expérience Poissonisée associée à E_n. Elle est obtenue en tirant d'abord un entier N d'une loi de Poisson telle que $E\,N = n$ et en faisant N observations au lieu de n.

PROPOSITION 4.1 *Si les conditions* (B) *et* (C) *sont satisfaites alors la distance* $\Delta(E_n, F_n)$ *entre* E_n *et sa Poissonisée* F_n *tend vers zéro quand* $n \to \infty$.

Démonstration. Les expériences F_n et E_n sont encadrées par les expériences F_n^- et F_n^+ où l'on prend respectivement $m = \min(n,N)$ et $M = \max(n,N)$ observations. Ces deux expériences peuvent être representées sur le même espace mesurable, produit infini de copies de $(\mathcal{X}_n, \mathcal{A}_n)$. Pour $\mathcal{F}_n^+$ on prend pour densités les produits $f_\theta^+ = \prod_{j=1}^{M} \frac{dp_\theta}{dp}$. Pour $\mathcal{F}_n^-$ on prend les mêmes produits mais on remplace $\frac{dp_\theta}{dp}$ par 1 entre m et M. L'affinité entre les mesures correspondantes P_n^- et P_n^+ est donc $\rho_n = \int \left(\prod_{j=m+1}^{M} \frac{dp_\theta}{dp} \right)^{1/2} dP^+$. Si ρ est l'affinité entre p et p_θ pour une observation ceci n'est autre que $E\rho^{(M-m)}$ mais $M-m$ est de l'ordre de grandeur de $\sqrt{n}$. Puisque $\rho^n \geq 1 - e^{-b}$ ne tend pas vers zéro $E\rho^{(M-m)}$ tends vers un. D'où le résultat.

Au prix de quelques complications de notation pour la démonstration on pourrait tout aussi bien se passer de la condition (C).

Remarquons toutefois que le résultat n'est pas correct sans la restriction (B). En effet, soit $\{\phi_k;\ k = 1,2,\ldots\}$ la suite des fonctions de Rademacher sur $[0,1]$ et soient $(1+\phi_k)$ des densités par rapport à la mesure de Lebesgue. On peut montrer que la distance entre le produit $\mathcal{E}_n$ correspondant et sa Poissonisation reste bornée inférieurement par au moins $\frac{1}{4}$. Une proposition analogue à la Proposition 4.1 serait très utile pour le cas de variables indépendantes nonéquidistribuées, mais un tel résultat n'a été démontré que sous des conditions très spéciales.

Montrons d'abord qui'il est possible de tronquer les $g \in \Theta$ et de se ramener au cas où ils sont bornés uniformement sur Θ (mais peut-être pas indépendamment de n).

Soit un g <u>éligible</u> de $\mathcal{L}_{2,0}(p)$. Posons $g' = gI[|g| \leq \varepsilon\sqrt{n}]$ et $g^* = g' - \int g'dp$. Soit p_g la mesure définie par

$$dp_g = \left[1 - c\left(\frac{\|g\|}{2\sqrt{n}}\right) + \frac{g}{2\sqrt{n}}\right]^2 dp$$

et soit P_g la mesure produit de n copies de p_g. On définit P_{g^*} de la même manière.

LEMME 4.1. On suppose que $\|g\|^2 \leq b$, que $\int g^2 I[|g| > \varepsilon\sqrt{n}]dp < \varepsilon < \frac{1}{2}$ et que $b \leq 3n$.

Alors g^* est éligible,

$$nh^2(g,g^*) \leq \frac{1}{2}\|g-g'\|^2 \leq \frac{\varepsilon}{2}$$

et

$$\|P_g - P_{g^*}\| \leq 2\sqrt{\varepsilon} .$$

Démonstration. Par construction $|g^*| \leq 2\varepsilon\sqrt{n}$. Donc g^* sera éligible si $1 - c(\frac{\|g^*\|}{2\sqrt{n}}) \geq \varepsilon$. Mais $\|g^*\|^2 \leq \|g'\|^2 \leq \|g\|^2 \leq b$. Il en résulte que g^* est certainement éligible si $1 - c(\frac{1}{2}\sqrt{b/n}) \geq \varepsilon$, ce qui donne le résultat voulu puisque $\{1 - c[\frac{1}{2}\sqrt{b/n}]\}^2 = 1 - \frac{1}{4}\frac{b}{n} \geq \frac{1}{4}$.

Puisque g^* est éligible, la distance de Hellinger $h(g,g^*)$ est donnée par

$$2h^2(g,g^*) = \{c[\frac{\|g^*\|}{2\sqrt{n}}] - c(\frac{\|g\|}{2\sqrt{n}})\}^2 + \frac{1}{4n}\|g-g^*\|^2 .$$

Le terme $\|g-g^*\|^2$ est la variance de $g-g'$. Donc $\|g-g^*\|^2 \leq \|g-g'\|^2 \leq \varepsilon$ par hypothèse. Pour le terme qui fait intervenir la fonction c notons que

$$c[\frac{\|g^*\|}{2\sqrt{n}}] - c[\frac{\|g\|}{2\sqrt{n}}] = \frac{1}{4n}[\|g\|^2 - \|g^*\|^2]A$$

avec $A = \left\{\sqrt{1 - \frac{\|g^*\|^2}{4n}} + \sqrt{1 - \frac{\|g\|^2}{4n}}\right\}^{-1} \leq \frac{1}{2}\frac{1}{\sqrt{1 - \frac{b}{4n}}} \leq 1$. On a donc

$$\left|c[\frac{\|g^*\|}{2\sqrt{n}}] - c(\frac{\|g\|}{2\sqrt{n}})\right| \leq \frac{1}{4n}\|g-g^*\|[\|g\|+\|g^*\|]$$
$$\leq \frac{\sqrt{b}}{2n}\|g-g^*\| \leq \frac{\sqrt{b}}{2n}\sqrt{\varepsilon} .$$

Ceci donne

$$\left|c\left(\frac{\|g^*\|}{2\sqrt{n}}\right) - c\left(\frac{\|g\|}{2\sqrt{n}}\right)\right|^2 \leq \frac{b}{4n^2}\,\varepsilon < \frac{3}{4n}\,\varepsilon\ .$$

On a donc bien

$$nh^2(g,g^*) \leq \frac{1}{2}\,\varepsilon\ .$$

Il en résulte que

$$\frac{1}{2}\int(\sqrt{dP_g} - \sqrt{dP_{g^*}})^2 \leq \frac{1}{2}\,\varepsilon$$

et $\|P_g - P_{g^*}\| \leq 2\sqrt{\varepsilon}$ d'après les inégalités entre distance de Hellinger et distances L_1.

On pourrait, bien sûr, se passer du Lemme 4.1 pour établir certains des résultats qui suivent. Toutefois notre but est de remplacer les P_g par d'autres mesures Q_g dont les densités mutuelles ont exactement la même forme que des densités de mesures Gaussiennes. Pour cela il nous faudra utiliser des résultats d'intégrabilité uniforme tels que ceux donnés par le lemme suivant.

LEMME 4.2. <u>Soient</u> X_j, $j=1,2,\ldots$ <u>des variables indépendantes telles que</u> $EX_j = 0$ <u>et</u> $X_j \leq a$ <u>et que</u> $\sum E(X_j^2) = \sigma^2$. <u>Alors, pour tout</u> $t \geq 0$ <u>on a</u>

$$E\, e^{t\sum_j X_j} \leq \exp\{\frac{\sigma^2 t^2}{2}\, e^{ta}\}$$

<u>et</u>

$$E\exp\{\sum_j X_j - \frac{1}{2}\sigma^2\} \leq \exp\{\frac{\sigma^2}{2}(e^{\varepsilon}-1) + (e^a-1)\varepsilon'\}$$

<u>avec</u> $\varepsilon' = \sum_j E\, X_j^2 I[X_j > \varepsilon]$.

Démonstration. Posons $\phi_j(t) = Ee^{tX_j}$. On a

$$E\, e^{\Sigma t X_j} = \prod_j \phi_j(t) \le \exp\{\sum_j [\phi_j(t)-1]\} .$$

Si M désigne la mesure définie par $M(A) = \sum_j \Pr[X_j \in A]$ ceci donne

$$\log E\, e^{\Sigma t X_j} \le \frac{t^2}{2}\int x^2 M(dx) + \int rt(x) M(dx)$$

où $r(x)$ est le reste $r(x) = e^x - 1 - x - \frac{x^2}{2} = x^2\int_0^1 (1-\xi)(e^{\xi x}-1)d\xi$. Ce reste a le même signe que x. On a donc bien

$$\int r(tx)M(dx) \le \frac{t^2}{2}\,\sigma^2(e^{ta}-1)$$

et

$$\int r(x)M(dx) \le \frac{1}{2}\,\sigma^2(e^{\varepsilon}-1) + (e^a-1)\int x^2 I[x > \varepsilon]M(dx) ,$$

ce qui donne les résultats enoncés.

En particulier, si $|X_j| \le \varepsilon$ et $E\, X_j = 0$ on aura

$$E\, \exp\{\sum_j X_j - \frac{1}{2}\sigma^2\} \le \exp\{\frac{1}{2}\sigma^2(e^{\varepsilon}-1)\} .$$

On peut noter que la même borne s'applique à des sommes $\sum_j \sum_{i=1}^{\nu_j} X_{j,i}$ où les $X_{j,i}$ sont des copies indépendantes de la variable X_j et où les ν_j sont des variables de Poisson, indépendantes entre elles et indépendantes des $X_{j,i}$, telles que $E\, \nu_j = 1$. En effet la première ligne de la démonstration revient à remplacer $\sum X_j$ par la somme Poissonisée.

Considérons maintenant des mesures Q_g définies comme suit. Soient ω_j; $j=1,\ldots,n$ les variables indépendantes sous observation. Soit S_n le processus empirique défini par

$$S_n(A) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [I_A(\omega_j)-p(A)] \ .$$

On notera $\langle g,S_n\rangle$ l'intégrale $\int g \, dS_n$. Prenons pour Q_0 la mesure P produit de n copies de notre mesure de base p. On peut alors définir une mesure Q_g par

$$\frac{dQ_g}{dQ_0} = \exp\{\langle g,S_n\rangle - \tfrac{1}{2}\|g\|^2\}$$

pour tout $g \in L_{2,0}(p)$. Il n'est pas dit que ce soit une mesure finie.

PROPOSITION 4.2. <u>Soit</u> $g \in L_{2,0}(p)$ <u>éligible et tel que</u> $\|g\|^2 \leq b$ <u>et</u> $|g| \leq \varepsilon\sqrt{n}$, <u>avec</u> $\varepsilon \leq \log 2$, $b\varepsilon \leq \log 2$.

<u>Alors</u> $\|P_g - Q_g\| \leq K\sqrt{b\varepsilon}$ <u>pour un coefficient</u> $K \leq 5$.

La proposition est presque un cas particulier de la proposition suivante.

PROPOSITION 4.3. <u>Soit</u> $\{g_k;\ k=1,2,\ldots\}$ <u>une suite d'éléments de</u> $L_{2,0}(p)$, <u>tous éligibles, tels que</u> $\|g_k\| \leq b$ <u>et</u> $|g_k|^2 \leq \varepsilon\sqrt{n}$. <u>Supposons de plus que les variables</u> $g_k(\omega_1)$ <u>soient mutuellement indépendantes sous</u> p. <u>Alors si</u> $f = \sum_k \alpha_k g_k$ <u>avec</u> $\sum_k \alpha_k^2 \leq 1$ <u>est éligible et si</u> $\max(\varepsilon, b\varepsilon) \leq 1$ <u>on a</u>

$$\|P_f - Q_f\| \leq 2[5b\varepsilon + \frac{b^2}{n}]^{1/2}.$$

<u>Démonstration</u>. Les variables $\frac{1}{\sqrt{n}}\alpha_k g_k(\omega_j)$, $j=1,2,\ldots,n;\ k=1,2,\ldots$ sont mutuellement indépendantes et bornées en valeur absolue par ε. On peut donc leur appliquer le Lemme 4.2. Avec $Q_f = \exp\{\langle f, S_n\rangle - \frac{1}{2}\|f\|^2\}$, $\langle f, S_n\rangle = \frac{1}{\sqrt{n}}\sum_k \sum_j \alpha_k g_k(\omega_j)$ ceci donne

$$\|Q_f\| \leq \exp\{(e^{\varepsilon}-1)\frac{1}{2n}\sum_k \sum_j \alpha_k^2 E g_k^2(\omega_j)\}$$

$$= \exp\{\frac{1}{2}(e^{\varepsilon}-1)\|f\|^2\} \leq \exp\{\frac{1}{2}(e^{\varepsilon}-1)b\} \leq 1 + 2\varepsilon b,$$

puisque, pour $0 \leq x \leq 1$, on a $e^x \leq 1 + 2x$.

Calculons maintenant l'affinité entre P_f et Q_f. C'est un produit de n terms tous égaux à

$$A(n) = \int [1 - c(\frac{\|f\|}{2\sqrt{n}}) + \frac{f}{2\sqrt{n}}]\exp\{\frac{1}{2\sqrt{n}}f - \frac{1}{4n}\|f\|^2\}dp$$

$$= \exp\{-\frac{1}{4n}\|f\|^2\}B(n)$$

où

$$B(n) = [1-c(\frac{\|f\|}{2\sqrt{n}})][1+\frac{1}{8n}\|f\|^2] + \frac{\|f\|^2}{4n} + \frac{1}{16n\sqrt{n}}\int f^3 dp$$
$$+ \int [1-c(\frac{\|f\|}{2\sqrt{n}}) + \frac{f}{2\sqrt{n}}] r(\frac{f}{2\sqrt{n}}) dp$$

avec $r(x) = e^x - 1 - x - \frac{x^2}{2}$, comme dans le Lemme 4.2. Puisque les g_k sont indépendantes et centrées

$$\int f^3 dp = E(\sum_k \alpha_k g_k) = \sum_k \alpha_k^3 E g_k^3$$

donc
$$\left|\int f^3 dp\right| \leq b\varepsilon\sqrt{n} .$$

Pour le premier terme, on peut écrire

$$1 - c(x) = 1 - \frac{x^2}{2} - \frac{x^4}{4} \frac{1}{(1+\sqrt{1-x^2})^2}$$
$$\geq 1 - \frac{x^2}{2} - \frac{x^4}{2} .$$

Donc $[1-c(x)](1+\frac{x^2}{2}) \geq 1 - (\frac{x^2}{2})^2 - \frac{x^4}{2} - \frac{x^6}{4}$. Ceci donne

$$[1-c(\frac{\|f\|}{2\sqrt{n}})][1+\frac{1}{8n}\|f\|^2] \geq 1 - \frac{3}{64n^2}\|f\|^4 - \frac{1}{2^8}\frac{1}{n^3}\|f\|^6 .$$

Le terme $\int [1-c(\frac{\|f\|}{2\sqrt{n}}) + \frac{f}{2\sqrt{n}}] r(\frac{f}{2\sqrt{n}}) dp$ peut être minoré de la manière suivante. Remarquons d'abord que $xr(x) \geq 0$. Donc le terme $\int \frac{f}{2\sqrt{n}} r(\frac{f}{2\sqrt{n}}) dp$ est positif. Il suffira donc de borner inférieurement $\int r(\frac{f}{2\sqrt{n}}) dp$. Mais on a

$$r(\frac{f}{2\sqrt{n}}) = \exp\left\{\frac{1}{2\sqrt{n}} \sum_k \alpha_k g_k\right\} - 1 - \frac{1}{2\sqrt{n}} \sum \alpha_k g_k - \frac{1}{8n}(\sum \alpha_k g_k)^2 .$$

Tenant compte du fait que $E\, g_k = 0$ pour les espérances mathématiques prisent sous p, on obtient

$$E\, r(\frac{f}{2\sqrt{n}}) = \prod_k \left\{1 + \frac{1}{8n}\alpha_k^2 \|g_k\|^2 + Er(\frac{\alpha_k g_k}{2\sqrt{n}})\right\} - 1 - \frac{1}{8n}\|f\|^2 .$$

De plus $Er(\frac{\alpha_k g_k}{2\sqrt{n}}) \geq \frac{\alpha_k^2 \|g_k\|^2}{8n}(e^{-\alpha_k \varepsilon} - 1)$. Les termes du produit sont donc supérieurs à l'unité et on a

$$E\, r(\frac{f}{2\sqrt{n}}) \geq \sum_k \frac{\alpha_k^2 \|g_k\|^2}{8n}(e^{-\alpha_k \varepsilon} - 1)$$
$$\geq - \frac{\|f\|^2}{8n} \varepsilon .$$

Mettant les termes ensemble, il vient

$$B(n) \geq 1 + \frac{\|f\|^2}{4n} - \frac{b\varepsilon}{16n} - \frac{\varepsilon \|f\|^2}{8n} - \frac{3}{64n^2}\|f\|^4 - \frac{1}{2^8 n^3}\|f\|^6$$
$$\geq 1 + \frac{\|f\|^2}{4n} - \frac{3b\varepsilon}{16n} - \frac{36}{n^2 2^8} b^2 ,$$

en utilisant l'inégalité $b \leq 4n$ pour le dernier terme.

Finalement en utilisant l'inégalité $e^{-x}(1+x) \geq 1 - x^2$ pour $x \geq 0$, on obtient

$$A(n) \geq 1 - \frac{3b\varepsilon}{16n} - \frac{36b^2}{2^8 n^2} - \frac{\|f\|^4}{16n^2}$$
$$\geq 1 - \frac{3b\varepsilon}{16n} - \frac{9}{64n^2} b^2 .$$

Il en résulte que l'affinité $\int \sqrt{dP_f dQ_f} = \rho$ est supérieure à

$$1 - \frac{3b\varepsilon}{16} - \frac{13b^2}{64n}.$$

L'inégalité de la proposition résulte alors de l'inégalité générale

$$\frac{1}{2}\|P-Q\| \leq \{\|\frac{1}{2}(P+Q)\|^2 - \rho^2\}^{1/2}$$

et du fait que $\|Q\| \leq 1+2b\varepsilon$ comme nous l'avons vu plus haut.

En effet on aura

$$\left|\left\|\frac{1}{2}[P_f+Q_f]\right\| - \rho\right| \leq \frac{3}{16}b\varepsilon + \frac{13b^2}{64n} + b\varepsilon \ .$$

Ceci donne

$$\left|\frac{1}{2}\|P_f+Q_f\| - \rho\right| \leq \frac{19}{16}b\varepsilon + \frac{13b^2}{64n}$$

$$\frac{1}{2}\|P_f+Q_f\| + \rho \leq 2[1+b\varepsilon] \leq 4$$

et le résultat voulu après un peu de calcul.

La Proposition 4.2 s'obtient de la même façon excepté que l'on remplace des termes tels que $\frac{1}{n^2}\|f\|^4$ par $\frac{1}{n}\|g\|^2\varepsilon^2$. Nous démontrerons d'ailleurs un résultat analogue pour un cas plus général en Section 5.

Disons qu'une partie éligible T de $L_{2,0}(p)$ satisfait à $L(b,\varepsilon)$ si pour tout $g \in T$ on a $\|g\|^2 \leq b$ et $\int g^2 I[|g| > \varepsilon\sqrt{n}]dp < \varepsilon$.

Ce que nous venons d'établir peut être exprimé ainsi: T possède une partie T' telle que pour tout $\bar{g} \in T$ il y a un $g \in T'$ tel que $\|P_{\bar{g}}-P_g\| \leq 2\sqrt{\varepsilon}$. De plus pour tout $g \in T'$ on a $\|Q_g-P_g\| \leq K\sqrt{b\varepsilon}$.

Considérons maintenant l'expérience Gaussienne G canoniquement attachée à l'espace $L_{2,0}(p)$. On peut la représenter par des mesures G_g telles que $\frac{dG_g}{dG_0} = \exp\{\langle g,W\rangle - \frac{1}{2}\|g\|^2\}$ où W est un bruit blanc Gaussien attaché à $L_{2,0}(p)$, c'est à dire où les $\langle g,W\rangle$ sont des variables Gaussiennes telles que $E\langle g,W\rangle = 0$ et $E|\langle g,W\rangle|^2 = \|g\|^2$.

Les rapports de vraisemblance $\frac{dQ_g}{dQ_0} = \exp\{\langle g,S_n\rangle - \frac{1}{2}\|g\|^2\}$ ont très exactement la même structure que ceux de l'expérience Gaussienne G. On peut toujours représenter les processus linéaires $g \rightsquigarrow \langle g,S_n\rangle$ et $g \rightsquigarrow \langle g,W\rangle$ par des éléments aléatoires à valeur dans le dual algébrique de $L_{2,0}(p)$, ou si l'on préfère, dans le dual algébrique de l'espace de Hilbert $L_{2,0}(p)$. Nous n'insisterons pas sur cette question car nous avons déjà noté plus haut qu'il suffit de regarder des parties dénombrables D de $L_{2,0}(p)$, auquel cas les processus $g \rightsquigarrow \langle g,S_n\rangle$ et $g \rightsquigarrow \langle g,W\rangle$ peuvent tous deux être considérés comme processus à trajectoires dans $\mathbb{R}^D$. Le processus $g \rightsquigarrow \langle g,S_n\rangle$ y a là une loi F_0 (sous p) et le processus $g \rightsquigarrow \langle g,W\rangle$, la loi Gaussienne G_0 de même covariance.

Considérons donc par exemple le cas $\mathbb{R}^D$. Puisque les rapports de vraisemblance sont donnés pour les Q_g et pour les G_g par la même expression $\exp\{\langle g,X\rangle - \frac{1}{2}\|g\|^2\}$, $X \in \mathbb{R}^D$ on voit que les solutions de Bayes associées à $\{Q_g\}$ et $\{G_g\}$ ont aussi des expressions identiques en temps que fonctions définies sur $\mathbb{R}^D$. Il est vrai que les Q_g ne sont pas des mesures de probabilité mais cela est sans grande importance puisque $\|Q_g-P_g\|$ est petit, donc aussi $|\|Q_g\|-1|$ et encore $\|\|Q_g\|^{-1}Q_g - P_g\| \leq 2\|Q_g-P_g\|$.

Il est bien tentant de conclure que les <u>risques</u> de Bayes sont aussi presque les mêmes et donc que la distance Δ entre E_n et l'expérience Gaussienne associée est petite. Toutefois la situation est plus compliquée: les mesures F_0 et G_0 sont le plus souvent disjointes, donc le fait que les solutions de Bayes soient les mêmes ne nous dit rien puisque l'ensemble où on les regarde pour F_0 est disjoint de celui qu'il faut faire intervenir pour G_0.

(Pour voir que F_n et G sont souvent disjointes prenons le cas de la suite $\{g_k ; k=1,2,\dots\}$ de la Proposition 4.3 et supposons $\|g_k\| = 1$. Alors $\sup_k |\langle g_k, S_n\rangle| \leq \varepsilon\sqrt{n}$, mais $\sup_k |\langle g_k, W\rangle| = \infty$ presque surement.)

La partie n'est pas perdue pour autant. En effet considérons les expressions $\|\sup_{j\in J} \sum_g C(j,g)\pi_g Q_g\|$ qui entrent dans la formule de la Section 2 pour l'écart Δ. Ici elles prennent la forme

$$E \sup_{j\in J} \sum_g C(j,g)\pi_g \exp\{\langle g,X\rangle - \tfrac{1}{2}\|g\|^2\}$$

pour une espérance prise pour Q_0 ou pour G_0 suivant le cas. Or les fonctions

$$X \rightsquigarrow \phi(X) = \sup_{j\in J} \sum_g C(j,g)\pi_g \exp\{\langle g,X\rangle - \tfrac{1}{2}\|g\|^2\}$$

sont des fonctions convexes très particulières. Il n'est donc pas exclu que les différences $\int\phi(x)F_0(dx) - \int\phi(x)G_0(dx)$ soient petites. Il est vrai que pour ces fonctions ϕ les fonctions de variable réelles $\xi \rightsquigarrow \phi[(1-\xi)X+\xi Y]$ sont intégrales de fonctions qui peuvent présenter des discontinuités. Toutefois il est possible de les rendre infiniment dérivables sans les modifier beaucoup. En effet soit Θ une partie de $L_{2,0}(p)$ telle que les espérances mathématiques $E \exp\{\langle g,X\rangle - \tfrac{1}{2}\|g\|^2\}$ soient proches de l'unité.

Soit $Z = g \rightsquigarrow \langle g,Z\rangle$ un processus Gaussien linéaire tel que $E\langle g,Z\rangle = 0$ et $E|\langle g,Z\rangle|^2 = \sigma^2\|g\|^2$ et soit F la loi de Z. Soit $\mathcal{F}$ l'expérience où pour g la distribution de Z a une densité $\exp\{\langle g,Z\rangle - \tfrac{1}{2}\sigma^2\|g\|^2\}$ par rapport à F. Soit $\mathcal{E}_n \times \mathcal{F}$ l'expérience qui consiste à faire $\mathcal{E}_n$ puis F indépendamment. En mettent $\mathcal{E}_n$ et $\mathcal{E}_n \times F$ sur le même espace on voit que $\Delta(\mathcal{E}_n, \mathcal{E}_n \times F) \leq \sigma \sup\{\|g\|: g \in \Theta\}$. Si donc $\sup\{\|g\|: g \in \Theta\} \leq b_1$ on peut, à une erreur σb_1 près, remplacer les fonctions $X \rightsquigarrow \phi(X)$ par les fonctions $X \rightsquigarrow \psi(X) = E\{\phi(X+Z)|X\}$. Ceci donne bien des fonctions indéfiniment dérivables.

Toutefois nous n'avons pu démontrer que les conditions (B), (C) et (L) sont suffisantes pour assurer l'existence d'une expérience Gaussienne $\mathcal{G}_n$ telle que $\Delta(\mathcal{E}_n, \mathcal{G}_n) \to 0$. Pour énoncer quelques résultats positifs, introduisons les définitions suivantes.

DÉFINITION 1. Soit $\Theta_n \subset L_{2,0}(p)$. Pour tout $\varepsilon > 0$ soit $N_n(\varepsilon)$ le plus petit nombre d'ensembles de diamétres $\leq \varepsilon$ nécessaires pour couvrir Θ_n. La suite $\{\Theta_n\}$ est dite <u>asymptotiquement précompacte</u> si pour tout ε fixe $\limsup_n N_n(\varepsilon) < \infty$.

DÉFINITION 2. Pour chaque n soit $\{g_k: k = 1,2,\ldots\}$ une partie de $L_{2,0}(p)$ (dépendant de n). On suppose que les g_k sont mutuellement indépendantes sous P_0, que $\|g_k\|^2 \leq b$ pour un b fixe et que les familles $\{g_k: k = 1,2,\ldots\}$ satisfont à la condition de Lindeberg (L). Soit Θ_n l'ensemble des $f \in L_{2,0}(p)$ de la forme $f = \sum_k \alpha_k g_k$ pour des coefficients tels que $\sum \alpha_k^2 \leq b$. On dira que Θ_n est une boule à base de Lindeberg indépendante.

THÉORÈME 4.1. <u>Soit Θ_n une partie éligible de $L_{2,0}(p)$ satisfaisant aux conditions (B), (C) et (L). Soit $\mathcal{E}_n$ l'expérience produit correspondante et soit $\mathcal{G}_n$ la restriction à Θ_n de l'expérience Gaussienne canonique de $L_{2,0}(p)$. Alors $\Delta(\mathcal{E}_n, \mathcal{G}_n) \to 0$ dans les cas suivants</u>:

1) Les Θ_n sont asymptotiquement précompacts.

2) Les Θ_n sont contenus dans des boules à bases de Lindeberg indépendantes.

3) Les Θ_n sont contenus dans boules à base de Lindeberg formées d'éléments de $L_{2,0}(p)$ deux à deux disjoints.

Démonstration. Le cas (1) est bien connu. On démontre d'abord le résultat en supposant que la cardinalité de Θ_n reste bornée. On passe au cas général par la technique de Lindae (voir par exemple [15]).

Pour le deuxième cas procédons de la façon suivante. Pour la base g_k, soit $F_{n,k}$ la distribution cumulative de $\langle g_k, S_n\rangle$ et soit $\Phi_{n,k}$ la distribution Gaussienne de $\langle g_k, W\rangle$. Soit $X_k = F^{-1}_{n,k}\Phi_{n,k}[\langle g_k, W\rangle]$.

La distribution de X_k est exactement la même que celle de $\langle g_k, S_n\rangle$. Puisque les variables sont indépendantes, alors pour $\sum \alpha_k^2 < \infty$, $\sum_k \alpha_k X_k$ a la même distribution que $\sum_k \alpha_k \langle g_k, S_n\rangle$. Soit donc, pour $f = \sum \alpha_k g_k$ dans boule Θ_n, une mesure P'_f qui a pour densité par rapport à G_0 l'expression $\exp\{\sum_k \alpha_{k,f} X_k - \frac{1}{2}\|f\|^2\}$. Ceci donne une expérience E'_n équivalente à l'expérience $F_n = \{Q_f : f \in \Theta_n\}$.

Considérons alors les différences du type

$$\begin{aligned} D_n = {} & E \sup_j \sum_g C(j,f)\pi_f \exp\{\sum \alpha_{k,f} X_k - \tfrac{1}{2}\|f\|^2\} \\ & - E \sup_j \sum_f C(j,f)\pi_f \exp\{\langle f, W\rangle - \tfrac{1}{2}\|f\|^2\} \end{aligned}$$

où les π sont des probabilités à support fini sur Θ_n. On a toujours

$$|D_n| \leq \sum_f \pi_f e^{-\frac{1}{2}\|f\|^2} E|\exp\{\sum \alpha_{k,f} X_k\} - \exp\langle f, W\rangle| \; .$$

D'après le Lemme 4.2, les exponentielles de cette différence sont équiintégrables. Pour un $\varepsilon > 0$ fixé on peu donc trouver un nombre a fixe tel que $E|\exp\{\langle f,W\rangle\} - \exp\{a\wedge\langle f,W\rangle\}| < \varepsilon$ et de même pour $(\sum_k \alpha_{k,f} X_k) \wedge a$.

Il existe donc une constante K fixe telle que

$$|D_n| \leq 2\varepsilon + K \sum_f \pi_f E|\sum \alpha_{k,f} X_k - \langle f,W\rangle|$$

ou encore

$$|D|_n \leq 2\varepsilon + K\left\{\sum_f \pi_f E\left|\sum_k \alpha_{k,f}[X_k - \langle g_k,W\rangle]\right|^2\right\}^{1/2} .$$

Le résultat sera donc acquis si nous montrons que $\sum_k \alpha^2_{k,f} E|X_k - \langle g_k,W\rangle^2|$ tend vers zéro, ou encore que $E|X_k - \langle g_k,W\rangle|^2$ tend vers zéro uniformément en k. Ici les variables X_k^2 et $|\langle g_k,W\rangle|^2$ sont équiintégrables. Pour celles dont les variances restent supérieures à un ε fixe, le théorème de Lindeberg montre bien que $X_k - \langle g_k,W\rangle$ tend vers zéro uniformément en probabilité. Les autres valeurs de k donnent $E|X_k - \langle g_k,W\rangle|^2 \leq 4\, E\, X_k^2$. D'où le résultat. Ceci finit la démonstration pour le cas (2). Pour le cas (3), remplaçons les $\mathcal{E}_n$ par les expériences Poissonisées $\mathcal{F}_n$ qui leur correspondent. Les processus S_n sont alors remplaçés par des processus S_n^* tels que $S_n^*(A) = \frac{1}{\sqrt{n}} \sum_{j=1}^{N} I_A(\omega_j)$. Il n'est pas besoin de les centrer puisque les éléments de $L_{2,0}(p)$ ont des espérances nulles. Les proprietés d'équiintégrabilité utilisées ci dessus sont encore valables pour les $\langle f,S_n^*\rangle$. En effet, c'est par cet intermédiaire qu'elles ont été démontrées dans le Lemme 4.2.

Pour le processus Poissonisé une suite $\{g_k\}$ formée d'éléments deux à deux disjointes donne des variables aléatoires $\langle g_k,S_n^*\rangle$ qui sont indépendantes. On peut donc faire le même raisonnement que pour le cas (2). Ceci complète la démonstration.

Remarque 1. Les cas (2) et (3) de ce Théorème font intervenir des boules qui n'ont peut être pas un interèt considérable pour les besoins pratiques. Toutefois il n'était pas évident a priori que l'on puisse se passer de conditions de précompacité. Les boules du cas (2) sont des boules entières dans un sous espace linéaire de $L_{2,0}(p)$. Si par exemple p est sans atomes, ce peuvent bien être des boules de dimension infinie, comme on le voit en prenant pour les g_k les fonctions de Rademacher sur $[0,1]$. De même, le cas (3) permet d'utiliser des partitions quelconques des ensembles sousjácents X_n, mais pour que les conditions de Lindeberg ne mênent pas à des expériences triviales il faut qu'au moins certains ensembles des partitions aient des probabilités qui décroissent moins rapidement que $1/n$.

Remarque 2. La démonstration pour le cas (2) repose sur la technique suivante. Pour des probabilités π_f, fixes pour un n donné mais dépendant de n, on a fabriqué une distribution conjointe de (S_n,W) telle que

$$E \sum_f \pi_f |\langle f,S_n\rangle - \langle f,W\rangle|^2$$

tend vers zéro quand n tend vers l'infini. Cette distribution simultanée des (S_n,W) était ici prise indépendante des π. Même si elle en dépendait, son existence entrainerait que $\Delta(E_n,G_n) \to 0$. L'existence de telles distributions pour (S_n,W) est une condition plus forte que la condition $\Delta(E_n,G_n) \to 0$. C'est visible à l'endroit où nous avons interchangé les signes $|\cdot|$ et E. On aurait toutefois pu penser que ces distributions conjointes existent toujours sous les conditions (B), (C) et (L). Il n'en est rien.

Prenons par exemple $[0,1]$ avec la mesure de Lebesgue. Soit $\{g_k;\ k=1,2,\ldots\}$ une base orthonormale de $L_{2,0}(p)$ formée de fonctions $|g_k|$ toutes inférieures à une constante fixée b. Il en existe bien, par exemple une base formée de sinus et cosinus. Supposons que, quelque soient les probabilités $\pi_{k,n}$, il existe distribution pour la paire (S_n,W) telle que

$$d_n = E\sum_k \pi_{k,n}|\langle g_k,S_n\rangle-\langle g_k,W\rangle|^2$$

tende vers zéro. On peut reécrire les $\pi_{k,n}$ sous la forme $\pi_{k,n} = \beta^2_{k,n}$ pour des $\beta_{k,n}$ positifs ou négatifs. La quantité d_n n'est alors autre que $d_n = E|\langle f_n,S_n\rangle-\langle f_n,W\rangle|^2$ pour $f_n = \sum_k \beta_{k,n}g_k$. Si d_n tends vers zéro, la distance entre la loi de $\langle f_n,S_n\rangle$ et la loi Gaussienne de $\langle f_n,W\rangle$ tends vers zéro. Toutefois il existe bien des f_n tels que $\|f_n\| \leq 1$ qui ne satisfont pas aux conditions de Lindeberg. Un tel couplage de S_n et W n'est donc pas possible, même si on le fait dépendre des probabilités π. La méthode de démonstration utilisée dans le Théorème 1 ne s'applique donc pas au cas de boules sur des bases orthogonales, ce qui aurait bien entrainé immédiatement la suffisance des conditions (B), (C) et (L). Des suites formées de variables indépendantes g_k ne peuvent jamais engendrer $L_{2,0}(p)$ tout entier. En effet si $E\,g_k = 0$, alors g_1g_2 est orthogonal à tous les g_k. Le problème d'approximation sous (B), (C) et (L) reste donc ouvert.

Puisqu'il en est ainsi, il est peut être bon de montrer que, toutefois, (B), (C) et (L) extrainment bien une approximabilité pour des distances plus faibles.

La distance Δ était obtenue en regardant <u>tous</u> les problèmes de statistique où la fonction de perte W est telle que $0 \leq W \leq 1$. On peut limiter la classe de problèmes. Par exemple Torgersen [20] a utilisé une distance Δ_m définie pas les problèmes où le statisticien prend ses décisions dans un ensemble ayant au plus m éléments. Denotant $\nu(\pi,C,E)$ la norme de $\sup_{j\in J} \sum_g C(j,g)\pi_g P_g$, la distance $\Delta_m(E,G)$ est égale à

$$\sup_{\pi,C} |\nu(\pi,C,E)-\nu(\pi,C,G)|$$

pour un supremum pris sur toutes les π et sur tous les systèmes $\{C(j,g): j \in J, g \in \Theta\}$ où l'ensemble J a au plus m éléments.

Une autre distance Δ_H peut être définie en regardant les problèmes d'estimation Hilbertiens. De façon précise, soit U la boule unité d'un espace de Hilbert. Soit $\theta \rightsquigarrow \phi(\theta)$ une application quelconque de Θ dans U. On considère les problèmes d'estimation où les décisions possibles sont les éléments de U et où la fonction de perte est $W(\theta,t) = \|\phi(\theta)-t\|^2$. La distance $\Delta_H(E,G)$ est au plus ε si et seulement si pour toute boule U, toute application ϕ, une fonction de risque r disponible sur E peut être reproduite à ε-près sur G et inversement.

Avec les notations précédentes pour un $\Theta \subset L_{2,0}(p)$ la distance $\Delta_H(E,G)$ est la différence maximum entre $\|\inf_x \sum_g \|x-\phi(g)\|^2 \pi_g P_\theta\|$ et la norme correspondante sur G, pour tout choix de π et d'applications ϕ dans une boule unité d'un Hilbert. Il n'est d'ailleurs pas besoin de traiter toutes ces boules. Puisque ni E ni G ne peuvent séparer des éléments équivalents de $L_{2,0}(p)$ et puisque la cardinalité de l'image par ϕ d'une partie de l'espace de Hilbert $L_{2,0}(p)$ ne peut excéder celle de $L_{2,0}(p)$, on peut se borner à la boule unité U de $L_{2,0}(p)$ lui-même.

THÉORÈME 4.2. *Soit* $\Theta_n \subset L_{2,0}(p)$ *satisfaisant aux conditions* (B), (C) *et* (L). *Soit* E_n *l'expérience produit associée à* Θ_n *et soit* G_n *la restriction à* Θ_n *de l'expérience Gaussienne canonique* G *de* $L_{2,0}(p)$.

Alors quand $n \to \infty$, *les distances* $\Delta_H(E_n,G_n)$ *et* $\Delta_m(E_n,G_n)$ *tendent vers zéro pour tout entier* m *fixe*.

Démonstration. Nous ne donnerons pas la démonstration ici, puisque le résultat est un cas particulier du résultat plus général etabli en Section 6, Théorèmes 6.1 et 6.3.

Les résultats établis ou énoncés ci dessus ne font intervenir que des ensembles qui satisfont à la condition (B). Ce sont donc des ensembles "assez petits", tout au moins si on les regarde dans l'espace des distributions individuelles. Il est clair que, pour des questions d'estimation par exemple, on voudrait bien avoir des approximations sur des ensembles "beaucoup plus grands". Nous ne savons obtenir de tels résultats que par un procédé de recollement déjà décrit dans [13]. (Le Lemme 1 de [13] est évidemment faux. Il manque dans son énoncé et sa démonstration un nombre de signes Σ. Toutefois le théorème de recollement, Théorème 1 de [13], est correct, à part peut-être multiplication de la borne donnée par un facteur 16.)

Afin de montrer qu'il est possible d'obtenir de tels résultats considérons un problème où les observations sont indépendantes, mais pas nécessairement équidistribuées. Autrement dit, considérons des produits $P_{\theta,n} = \prod_j p_{\theta,j,n}$ comme tout au début de cette Section, avec $\theta \in \Theta_n$ et $j \in I_n$.

Ici il n'y a plus d'espace commode analogue aux $L_{2,0}(p)$ utilisables pour les variables équidistribuées. Nous définirons donc des espaces de Hilbert comme suit.

Puisque n est partout ci n'est pas nécessaire de l'écrire. Il sera donc omis le plus souvent.

Soit h_j la distance de Hellinger définie sur Θ par $h_j^2(s,t) = \frac{1}{2}\int(\sqrt{dp_{s,j}} - \sqrt{dp_{t,j}})^2$. Soit $H^2(s,t) = \sum_{j\in J} h_j^2(s,t)$.

Soit M_0 l'espace des mesures réelles à support fini sur Θ et telles que $\mu(\Theta) = 0$. Posons

$$\|\mu\|_H^2 = -4\iint H^2(s,t)\mu(ds)\mu(dt) .$$

On obtient un espace de Hilbert H en prenant le quotient de M_0 par l'espace $\{\mu: \|\mu\|_H = 0\}$ et en complétant pour la norme $\|\mu\|_H$.

Considérons les conditions suivantes pour les expériences produits $E_n = \{P_{\theta,n}; \theta\in\Theta_n\}$, mais avec l'indice n omis des notations.

(A1) <u>Si</u> $H^2(s,t)$ <u>reste borné (indépendamment de</u> n) <u>alors</u> $\sup_j h_j^2(s,t)$ <u>tends vers zéro quand</u> $n\to\infty$.

(A2) <u>Si</u> $H^2(s,t)$ <u>reste borné, alors pour tout</u> ε <u>fixe</u>

$$\sum_j (p_{s,j}+p_{t,j})\left\{\left|\frac{dp_{t,j}}{dp_{s,j}} - 1\right| > \varepsilon\right\}$$

<u>tends vers zéro</u>.

Cette condition est équivalente à une condition de Lindeberg pour les racines $\sqrt{dp_{t,j}/dp_{s,j}}$ avec la condition supplémentaire que la somme des masses de parties singulières tendent vers zéro.

(A3) <u>Les</u> Θ <u>ont une dimension dominée au sens suivant. Pour tout</u> ε <u>fixe, soit</u> $N_n(\varepsilon)$ <u>le nombre minimum d'ensembles de diamètre</u> $x \geq \varepsilon$ <u>nécessaires pour couvrir des ensembles de diamètre</u> $2x$ <u>dans</u> (Θ,H).

Alors il existe une fonction $\varepsilon \to \nu(\varepsilon) \in (0,\infty)$ *indépendante de* n *telle que* $N_n(\varepsilon) \leq \nu(\varepsilon)$ *pour tout* n.

THÉORÈME 4.3. *Supposons les conditions* (A1), (A2) *et* (A3) *satisfaites.*

Soit G *la restriction à* Θ *de l'expérience Gaussienne canonique attachée à l'espace de Hilbert* H *défini ci dessus. Alors* $\Delta(E,G) \to 0$ *quand* $n \to \infty$.

Démonstration. Les théorèmes de Birgé [3] (où même ceux de [13]) entrainent l'existence d'estimateurs $\hat{\theta}$ tels que $\sup_{\theta} E_{\theta}H^2(\hat{\theta},\theta)$ reste borné indépendamment de n par un nombre $b < \infty$. Cela est vrai pour E et pour G.

Pour toute boule $B(a) = \{t: H(s,t) \leq a\}$ centrée à un point s et de rayon a fixé, les conditions (A1), (A2) et (A3) entrainent bien que $\Delta(E_{B(a)}, G_{B(a)})$ tend vers zéro pour les restrictions des expériences E et G à $B(a)$. Cela résulte par exemple des formules données dans [12] et de la précompacité uniforme des boules $B(a)$. Il y a donc aussi une suite $\{a_n\}$ tendant vers l'infini telle que $\Delta[E_{B(a_n)}, G_{B(a_n)}]$ tende encore vers zéro uniformément pour toutes ces boules.

Cela permet d'utiliser le théorème de raccordement de [13]. Le résultat peut donc être considéré comme acquis.

Remarque 1. A noter que l'expérience G est une expérience Gaussienne au sens de notre définition, c'est à dire homoscédastique. Les conditions (A1), (A2) et (A3) sont en particulier satisfaites dans le système utilisé par Wald en 1943 ([21]). Wald utilise les espaces de Hilbert locaux (de dimension vectorielle bornée) définis par les matrices d'information de Fisher. Cela lui donne une variété Riemannienne et des approximations Gaussiennes heteroscédastiques.

Ici l'approximation est homoscédastique, mais indexée par une partie d'un espace de Hilbert de dimension vectorielle infinie en général. Même localement, l'espace en question peut fort bien ne pes être approximable par un espace de dimension vectorielle finie. La condition (A3) dit bien que Θ a pour chaque ε fixe une dimension finie, mais c'est une dimension <u>métrique</u> et non pas une dimension vectorielle.

<u>Remarque 2</u>. Nous avons déjà mentioné l'exemple des densités de translation $C \exp\{-|x-\theta|^{\alpha}\}$ pas rapport à la mesure de Lebesgue sur la droite et pour un α fixé, $\alpha \in (0,\frac{1}{2})$. Quand θ varie dans $\mathbb{R}$ et pour des observations indépendantes équidistribuées, ces familles ne satisfont pas tout à fait à la condition (A3). Toutefois on peut aisement se procurer des estimateurs θ^* tels que $|\theta^*-\theta|$ reste borné en probabilité. On se ramène donc de suite au cas d'une partie bornée de la droite. Pour une telle partie les conditions (A1), (A2) et (A3) sont bien satisfaites. On obtient la droite toute entière par recollement. L'expérience produit $\mathcal{E}_n = \{P_{\theta,n};\ \theta \in \mathbb{R}\}$ est donc bien approximable par l'expérience Gaussienne correspondante. Ici l'espace $\mathcal{H}$ pourrait être remplacé par un autre espace de Hilbert obtenu en mettant sur $\mathbb{R}$ une distance proportionelle à $\sqrt{n}|s-t|^{\beta}$, $\beta = \frac{1}{2}(1+2\alpha)$ au lieu des normes habituelles telles que $n^{1/2\beta}|s-t|$.

Ceci montre bien que les Θ_n utilisables dans ce Théorème 4.3 sont "beaucoup plus grands" que ceux des Théorème 4.1 et 4.2 en ce sens que leur diamètres ne sont pas limités. Ils sont par contre beaucoup plus "éffilés" localement que les boules à bases indépendantes du Théorème 4.1.

5. Extension au cas d'observations dépendantes. Formules préliminaires.

Considérons une tribu A de parties d'un ensemble X et une filtration

$$A_1 \subset A_2 \subset \cdots \subset A_k \subset A_{k+1} \subset \cdots \subset A$$

donnée par une famille croissante finie de sous tribus de A, la dernière étant A elle même. Soient P_i; $i = 0,1$ deux mesures de probabilité sur A. Si P_1 est dominée par P_0 on peut écrire sa densité sous la forme

$$\frac{dP_1}{dP_0} = \prod_k (1+X_k)$$

où les X_k sont A_k mesurables et tels que $X_k \geq -1$. Pour obtenir cette représentation on peut procéder comme suit. Soit E_k^i l'opérateur espérance conditionelle étant donné k pour la mesure P_i. On pose $\gamma_0 = 1$, $\gamma_k = E_k^0 \frac{dP_1}{dP_0}$, $f_k = \frac{\gamma_k}{\gamma_{k-1}}$ et $X_k = f_k - 1$.

Les X_k sont des accroissements de martingales pour P_0. En effet $E_{k-1}^0 X_k = 0$. Si T est A_k mesurable, on aura $E'_{k-1} T = E_{k-1}^0 T(1+X_k)$.

Dans la suite on posera $\sigma_k^2 = E_{k-1}^0 X_k^2$. Considérons alors les restrictions suivantes, où il est supposé que ε est petit, par exemple $\varepsilon < \frac{1}{10}$.

(R1) $\sum_k (P_0+P_1)\{|X_k| > \varepsilon\} \leq \varepsilon$

(R2) Il y a un nombre $b < \infty$ tel que $\sum_k \sigma_k^2 \leq b$

Comme déjà signalé en Section 4, la condition (R1) est équivalente à la condition de Lindeberg pour les variables $\sqrt{1+X_k} - 1$ si ces variables sont indépendantes. Nous l'avons imposée ici par analogie. Toutefois nous ne savons pas si elle est nécessaire lorsque les variables sont dépendantes.

La condition (R2) est une condition analogue à la condition (B) de la Seetion 4. Elle va empêcher les mesures P_0 et P_1 de se séparer. Toutefois, puisque le b <u>n'est pas aléatoire</u>, elle entraine aussi beaucoup d'autre choses.

Nous avons supposé ici que la filtration $\{A_k\}$ est donnée par une famille finie de tribus. On pourrait utiliser tout aussi bien une famille dénombrable où A est la tribu engendrée par $\bigcup_k A_k$. Nous ne le ferons pas, car cela complique un peu les raisonnements sans apporter grand chose aux résultats.

Remarquons qu'ici on utilise les densités $1+X_k$ au lieu des racines carrées de la Section 4. Cela n'a pas d'importance réelle si, comme nous allons le faire ci-dessous, on suppose les X_k uniformément bornés. Les X_k sont des différences de martingales, ce qui les rend plus facilement maniables que les $\sqrt{1+X_k} - 1$. Pour que les hypothèses faites ici deviennent comparables à celles de la Section 4 il faut les appliquer aux variables tronquées $\frac{g^*}{\sqrt{n}}$ du Lemme 4.1. Cela fait, les résultats des Sections 5 et 6 deviennent plus généraux que les résultats correspondants de la Section 4.

Aux deux conditions (R1) et (R2) nous ajouterons les deux restrictions suivantes:

(R3) $X_k \leq 1$

(R4) $E^1 \sum_k E^0_{k-1} I[X_k < -\frac{1}{2}] < \varepsilon$

Ces deux conditions supplémentaires sont "presque" des conséquences de (R1) et (R2) en ce sens que si (R1) et (R2) sont satisfaites il est possible de modifier les X_k, obtenant de nouvelles variables X'_k qui satisfont à des conditions voisines de (R3) et (R4) sans pour cela modifier beaucoup les produits $\prod_k (1+X_k)$. Le but de la présente Section est de démontrer celà et d'obtenir une formule exponentielle analogue à la formule de la Proposition 4.2. Il est bien évident que le nombre 1 de (R3) pourrait être remplacé par une constante positive quelconque. Les variables tronquées du Lemme 5.2 seront aussi utilisées pour les résultats de la Section 6.

LEMMA 5.1. _Supposons que les_ X_k _vérifient la condition_ (R1). _Soit_ $X^*_k = X_k I[X_k \leq \varepsilon]$ _et soit_ $a_k = -E^0_{k-1} X^*_k$. _Soit_ P^*_1 _la mesure dont la densité par rapport à_ P_0 _est le produit_ $\prod_k [1+X^*_k+a_k]$. _Alors_ $\| P^*_1 - P_1 \| \leq 2\varepsilon$ _et_ $X^*_k + a_k \leq 1+\varepsilon$.

Démonstration. Les variables a_k sont positives. Le rapport $f = \frac{dP^*_1}{dP_1}$ est donc supérieur à $g = \prod_k (1+X^*_k)(1+X_k)^{-1}$. Sous la condition (R1) on a $P_1[g \neq 1] \leq \varepsilon$ donc

$$\|P_1^*-P_1\| \le 2\int[1-(1\wedge f)]dP_1$$
$$\le 2\int[1-1\wedge g]dP_1 \le 2\varepsilon \quad .$$

Puisque $a_k = -E^0_{k-1}X^*_k$ on a bien $0 \le a_k \le 1$ et $X^*_k + a_k \le 1+\varepsilon$. (En fait (R1) entraine que $(E^0+E^1)(\sum_k a_k) \le 2\varepsilon$.)

Supposons maintenant que (R1), (R2) et (R3) sont satisfaites et prenons un nombre $a \ge 2$. Soit I_k l'indicateur de l'ensemble

$$\sup_{\nu\le k-1}\prod_{j\le\nu}(1+X_j) \le a \; .$$

Posons $X^a_k = I_k X_k$. Par construction on a encore $X^a_k \ge -1$ et $E^0_{k-1}X^a_k = 0$. Donc le produit $\prod_k(1+X^a_k)$ est une densité par rapport à P_0 d'une mesure de probabilité P^a_1.

LEMMA 5.2. <u>Si les conditions</u> (R1), (R2) <u>et</u> (R3) <u>sont satisfaites on a</u>

(i) $\sup_k \prod_{j\le k}(1+X^a_j) \le 2a$,

(ii) $\|P^a_1-P_1\| \le 2\,\dfrac{e^b}{a}$.

<u>Démonstration</u>. Soit k le premier entier tel que $\prod_{j\le k}(1+X_j) > a$. On a donc $\prod_{j\le k-1}(1+X_j)(1+X_k) \le a(1+X_k) \le 2a$, par (R3). Ceci donne la première inégalité. Pour la seconde, notons que $\prod_k(1+X_k)$ et $\prod_k(1+X^a_k)$ sont toujours égaux excepté sur l'ensemble A où $\sup_k\prod_{j\le k}(1+X_j) > a$. Il suffira donc de montrer que $P_1(A)$ et $P^a_1(A)$ sont tous deux inferieurs à e^b/a. Pour P_1 notons que

$$E^1_{k-1}(1+X_k) = E^0_{k-1}(1+X_k)^2 = 1+\sigma^2_k \; .$$

De même si E^a_{k-1} désigne une espérance conditionelle pour P^a_1 on a

$$E^a_{k-1}(1+X_k) = E^0_{k-1}(1+X_k)(1+X^a_k) = 1+E^0_{k-1}X^2_k I_k \; .$$

Ceci est égal à $1+I_k\sigma_k^2 \leq 1+\sigma_k^2$ puisque I_k est A_{k-1} mesurable. Il en résulte que les produits $\prod_{j\leq k}(1+X_j)\exp\{-\sigma_j^2\}$ sont des surmartingales positives pour P_1 et pour P_1^a. Pour de telles surmartingales on aura bien

$$\text{Prob}\{\sup_k \prod_{j\leq k}(1+X_j)\exp\{-\sigma_j^2\} \geq \gamma\} \leq \frac{1}{\gamma}\,.$$

Puisque $\sum_j \sigma_j^2 \leq b$, ceci donne le résultat voulu en prenant $\gamma = ae^{-b}$.

Nous aurons aussi besoin, en Section 6, de résultats un peu plus précis. L'un d'entre eux est le suivant:

LEMME 5.3. <u>Si les conditions</u> (R1), (R2) <u>et</u> (R3) <u>sont satisfaites, alors pour tout entier</u> $n \geq 1$ <u>on a</u>

$$P_0\{\sup_k \prod_{j\leq k}(1+X_j) \geq a\} \leq \frac{1}{a^n}\exp\{(2^n-n-1)b\}\,.$$

<u>Démonstration</u>. On peut écrire $(1+X_k)^n = 1+nX_k + \sum_{s=2}^{n}\binom{n}{s}X_k^s \leq 1+nX_k + \sum_{s=2}^{n}\binom{n}{s}X_k^2$, puisque $|X_k| \leq 1$. Donc $E_{k-1}^0(1+X_k)^n \leq 1+(2^n-1-n)\sigma_k^2$. Il en résulte que les produits

$$\prod_{j\leq k}(1+X_j)^n\exp\{-(2^n-1-n)\sigma_j^2\}$$

forment une surmartingale. La conclusion s'en déduit comme auparavant.

<u>Remarque 1</u>. On aurait aussi pu utiliser le théorème de Dubins et Freedman [5] qui dit que $\exp\{\lambda\sum_{j\leq k}X_j - \phi(\lambda)\sum_{j\leq k}\sigma_j^2\}$ est une surmartingale pour $\phi(\lambda) = \exp\{\lambda\}-1-\lambda$. La formule donnée ci-dessus sera plus commode en Section 6. Sa méthode de démonstration montre aussi qu'on aurait pu améliorer les bornes du Lemme 5.2.

Remarque 2. La condition (R4) est peu agréable parce qu'elle mélange les espérances conditionnelles sous P_0 avec des espérances sous P_1. Il est bien vrai que, pour $\varepsilon \leq \frac{1}{2}$, et pour des variables indépendantes elle est une conséquence de (R1). Dans le cas général nous ne savons pas ce qu'il en est. Toutefois en remplaçant P_1 par P_1^a on a bien

$$\int\left\{\sum_k E_{k-1}^0 I[X_k < -\tfrac{1}{2}]\right\} dP_1^a \leq 2a\int \sum_k E_{k-1}^0 I[X_k < -\tfrac{1}{2}] dP_0 \leq 2a\varepsilon ,$$

d'après (R1), si $\varepsilon \leq \frac{1}{2}$. Il est donc possible de considérer que l'addition de (R3) et (R4) aux conditions essentielles (R1) et (R2) ne fait pas vraiment perdre de généralité.

Pour terminer cette section, nous allons démontrer une formule exponentielle analogue à celle de la Proposition 4.2.

THÉORÈME 5.1. Soit Q_1 la mesure dont la densité par rapport à P_0 est

$$\exp\{\sum_k [X_k - \tfrac{1}{2}\sigma_k^2]\} .$$

Alors sous les conditions (R1)-(R4) il existe une constante $K(b)$, dépendant seulement de b, telle que si $b\varepsilon < 1-\varepsilon$ on ait

$$\|Q_1 - P_1\| \leq K(b)\sqrt{\varepsilon} .$$

La démonstration va être divisée en plusieurs étapes. Nous montrerons d'abord que P_1 peut être remplacée par la mesure P_2 définie par

$$\frac{dP_2}{dP_0} = \prod_k e^{-\frac{1}{2}\sigma_k^2}[1 + X_k + \frac{X_k^2}{2}] .$$

Cette mesure P_2 n'est pas une probabilité, mais elle est inférieure à la mesure P_3 définie par

$$\frac{dP_3}{dP_0} = \prod_k \left[\frac{1}{1+\sigma_k^2/2}\right]\left[1+X_k+\frac{X_k^2}{2}\right] .$$

Puisque P_3 est une probabilité, on a bien $\|P_2\| \leq 1$.

LEMME 5.4. <u>On suppose que les conditions</u> (R1)-(R4) <u>sont satisfaites, que</u> $\varepsilon < \frac{1}{4}$.

<u>Alors</u> $\|P_2-P_1\| \leq [7+3b]\sqrt{\varepsilon}$.

<u>Démonstration</u>. Posons $Y_k = \frac{X_k^2}{1+X_k}$. Le logarithme $\Lambda = \log \frac{dP_2}{dP_1}$ prend alors la forme

$$\Lambda = \sum_k \left\{\log\left(1+\frac{Y_k}{2}\right) - \frac{1}{2}\sigma_k^2\right\} .$$

Posons encore $Z_k = Y_k I[|X_k| \leq \varepsilon]$. Puisque Y_k et Z_k sont tous les deux positifs, ceci donne

$$\log\left[1+\frac{Y_k}{2}\right] \geq \log\left[1+\frac{Z_k}{2}\right] \geq \frac{Z_k}{2} - \frac{1}{2}\left(\frac{Z_k}{2}\right)^2 .$$

De plus $|X_k| \leq \varepsilon$ entraine $Y_k \leq \frac{\varepsilon^2}{1-\varepsilon}$. Considérons d'alors la somme $\sum Z_k^2$. On a

$$E_{k-1}^1 Z_k^2 \leq \frac{\varepsilon^2}{1-\varepsilon} E_{k-1}^1 Z_k \leq \frac{\varepsilon^2}{1-\varepsilon} E_{k-1}^1 Y_k = \frac{\varepsilon^2}{1-\varepsilon}\,\sigma_k^2 .$$

Donc

$$E^1 \sum_k \frac{1}{2}\left(\frac{Z_k}{2}\right)^2 \leq \frac{\varepsilon^2}{8(1-\varepsilon)}\, b .$$

Pour décrire le comportement du terme $\sum Z_k$ posons $s_k^2 = E_{k-1}^1 Z_k$. Pour la mesure P_1 les terms $Z_k - s_k^2$ sont les accroissements d'une martingale $\{M_k\}$. Les variances conditionnelles de cette martingale satisfont à

$$E^1_{k-1}|Z_k-s_k^2|^2 \leq \frac{2\varepsilon^2}{1-\varepsilon} E^1_{k-1}|Z_k-s_k^2| \leq 4\,\frac{\varepsilon^2}{1-\varepsilon}\, s_k^2 \ .$$

De plus $s_k^2 = E^1_{k-1}Z_k \leq E^1_{k-1}Y_k = \sigma_k^2$. On a donc $\sum_k s_k^2 \leq \sum_k \sigma_k^2 \leq b$ et pour $M = \sum_k [Z_k - s_k^2]$, $E^1M^2 \leq 4\,\frac{\varepsilon^2}{1-\varepsilon}\, b$.

Il reste à examiner la somme $V = \sum_k [\sigma_k^2 - s_k^2]$. On peut écrire

$$\begin{aligned}
\sigma_k^2 - s_k^2 &= E^1_{k-1}Y_k I[|X_k| > \varepsilon] \\
&= E^0_{k-1}X_k^2 I[|X_k| > \varepsilon] \\
&= E^0_{k-1}X_k^2 I[X_k > \varepsilon] + E^0_{k-1}X_k^2 I[-\tfrac{1}{2} \leq X_k < -\varepsilon] \\
&\quad + E^0_{k-1}X_k^2 I[X_k < -\tfrac{1}{2}] \ .
\end{aligned}$$

Les trois termes du membre droit peuvent être bornés comme suit

(a) $E^0_{k-1}X_k^2 I[X_k > \varepsilon] \leq E^1_{k-1}X_k^2 I[X_k > \varepsilon]$

(b) $E^0_{k-1}X_k^2 I[-\frac{1}{2} \leq X_k < -\varepsilon] \leq 2E^1_{k-1}X_k^2 I[-\frac{1}{2} \leq X_k < -\varepsilon]$

Finalement, d'après la condition (R4)

(c) $E^1 \sum_k E^0_{k-1}X_k^2 I[X_k < -\frac{1}{2}] \leq \varepsilon$.

Il vient donc

$$E^1V \leq \varepsilon + 2\sum_k P_1[|X_k| > \varepsilon] \leq 3\varepsilon \ .$$

Posant $T = \frac{1}{2}V + 4\sum_k \frac{1}{2}(\frac{Z_k}{2})^2$, on voit que $\Lambda \geq \frac{1}{2}M - T$ où M est la variable finale d'une martingale telle que $E^1M = 0$ et $E^1(\frac{1}{2}M)^2 \leq \frac{\varepsilon^2}{1-\varepsilon}\, b$ et où T est une variable positive telle que

$$E^1T \leq \frac{\varepsilon^2 b}{8(1-\varepsilon)} + \frac{3}{2}\,\varepsilon \quad .$$

Les inégalités de Tchebychev et Markov donnent alors

$$P_1\{\tfrac{1}{2}M \leq -\sqrt{\varepsilon}\} \leq \frac{b\varepsilon}{1-\varepsilon} ,$$

$$P_1[T \geq \sqrt{\varepsilon}] \leq \tfrac{3}{2}\sqrt{\varepsilon} + \frac{b\varepsilon\sqrt{\varepsilon}}{8(1-\varepsilon)} .$$

Ceci donne

$$P_1[\Lambda \leq -2\sqrt{\varepsilon}] \leq \tfrac{3}{2}\sqrt{\varepsilon} + \frac{b\varepsilon}{8(1-\varepsilon)}[8 + \sqrt{\varepsilon}]$$

et pour $\varepsilon \leq \frac{1}{4}$

$$P_1[\Lambda \leq -2\sqrt{\varepsilon}] \leq [\tfrac{3}{2} + \tfrac{17}{12}b]\sqrt{\varepsilon} .$$

D'après les inégalités usuelles, si $f = \frac{dP_2}{dP_1}$ on a

$$\|P_2 - P_1\| = 2\int[1 - (1 \wedge f)]dP_1 + \|P_2\| - \|P_1\| .$$

Ici $\|P_2\| \leq \|P_1\| = 1$. Il vient donc

$$\begin{aligned}\|P_2 - P_1\| &\leq 2\{[1 - \exp\{-2\sqrt{\varepsilon}\}] + [\tfrac{3}{2} + \tfrac{17}{12}b]\sqrt{\varepsilon}\} \\ &\leq 7\sqrt{\varepsilon} + \tfrac{17}{6}b\sqrt{\varepsilon} .\end{aligned}$$

D'où le résultat.

Le raisonnement s'applique tout aussi bien à la mesure de probabilité $P_3 \geq P_2$. On a donc aussi:

COROLLAIRE. <u>Sous les hypothèses du Lemme 6 on a</u>

$$\|P_3 - P_1\| \leq [7 + 3b]\sqrt{\varepsilon} .$$

Passons maintenant à la différence entre les mesures P_2 et Q_1.

LEMME 5.5. <u>On suppose que les conditions</u> (R1)-(R4) <u>sont satisfaites et que</u> $\varepsilon \le \frac{1}{4}$. <u>Alors il existe une constante</u> $K_1(b)$, <u>dépendant seulement de</u> b, <u>telle que</u>

$$\|Q_1 - P_2\| \le K_1(b)\sqrt{\varepsilon}\ .$$

<u>Démonstration</u>. Posons $A_1 = 1$ et

$$A_k = \prod_{j<k} [1 + X_j + \frac{X_j^2}{2}] \exp\{-\frac{1}{2}\sigma_j^2\}$$

$$B_k = \prod_{j>k} \exp[X_j - \frac{1}{2}\sigma_j^2]$$

$$\begin{aligned} \Delta_k &= \exp\{-\frac{1}{2}\sigma_k^2\}[\exp\{X_k\} - (1 + X_k + \frac{X_k^2}{2})] \\ &= \exp\{-\frac{1}{2}\sigma_k^2\}\{X_k^2 \int_0^1 (1-\xi)(e^{\xi X_k} - 1)d\xi\}\ . \end{aligned}$$

Alors la différence des rapports de vraisemblance est égale à

$$D = \frac{dQ_1}{dP_0} - \frac{dP_2}{dP_0} = \sum_k A_k \Delta_k B_k\ ,$$

donc

$$\|Q_1 - P_2\| = E^0|D| \le \sum_k E^0 A_k |\Delta_k| B_k\ .$$

Soit γ le nombre $\gamma = e - 2.5$. D'après le résultat de Dubins et Freedman [5], l'inégalité $X_k \le 1$ entraine que les variables $\exp \sum_{j\le k} [X_j - \frac{1}{2}\sigma_j^2 - \gamma\sigma_j^2]$ forment une surmartingale sous P_0. On a donc

$$E_k^0 B_k \exp\{-\gamma \sum_{j>k} \sigma_j^2\} \le 1$$

et donc, suivant (R2), $E_k^0 B_k \le \exp\{\gamma b\}$. Ceci donne

$$E^0|D| \leq e^{\gamma b} \sum_k E^0 A_k|\Delta_k|$$
$$= e^{\gamma b} E^0\{\sum_k A_k E^0_{k-1}|\Delta_k|\}$$

puisque A_k est $\mathcal{A}_{k-1}$ mesurable. La dernière expression peut aussi s'écrire sous la forme

$$e^{\gamma b}\int\{\sum_k E^0_{k-1}|\Delta_k|\}dP_2 \ .$$

La formule de Taylor montre que

$$E^0_{k-1}|\Delta_k| \leq (e^{\varepsilon}-1)\sigma_k^2 + (e-1)E^0_{k-1}X_k^2 I[|X_k| > \varepsilon]$$
$$\leq (e^{\varepsilon}-1)\sigma_k^2 + (e-1)\sigma_k^2 \ .$$

On a donc en particulier

$$\sum_k E^0_{k-1}|\Delta_k| \leq (e^{\varepsilon}-1)b + (e-1)b \leq 2(e-1)b \ .$$

D'après le Lemme 5.4 ceci entraine

$$\int[\sum_k E^0_{k-1}|\Delta_k|]dP_2 \leq \int[\sum_k E^0_{k-1}|\Delta_k|]dP_1 + 2(e-1)b[7+3b]\sqrt{\varepsilon} \ .$$

Il suffira donc de borner $\int[\sum_k E^0_{k-1}|\Delta_k|]dP_1$, ou encore la quantité plus grande

$$(e^{\varepsilon}-1)b + (e-1)\int\{\sum_k E^0_{k-1}X_k^2 I[|X_k| > \varepsilon]\}dP_1 \ .$$

Au cours de la démonstration du Lemme 5.4 nous avons vu que le terme intégral dans cette formule est borné par 3ε. Il vient donc

$$E^0|D| \leq e^{\gamma b}\{(e^{\varepsilon}-1)b + 3\varepsilon(e-1) + 2(e-1)b(7+3b)\sqrt{\varepsilon}\} \ .$$

On a donc

$$\|P_2-Q_1\| \leq K(b,\varepsilon)\sqrt{\varepsilon}$$

où

$$K(b,\varepsilon) = e^{\gamma b}\{2(e-1)b(7+3b) + [3(e-1) + be^{1/4}]\sqrt{\varepsilon}\} ,$$

ce qui entraine le résultat désiré.

Pour démontrer le Théorème 5.1 il suffit maintenant d'écrire

$$\|Q_1-P_1\| \leq \|Q_1-P_2\| + \|P_2-P_1\| .$$

Au lieu d'une paire $\{P_0,P_1\}$ de mesures sur A, on peut considérer une expérience $E = \{P_\theta;\ \theta \in \Theta\}$ où Θ est un ensemble arbitraire. Supposons choisi un θ_0 particulier dans Θ et posons $P_0 = P_{\theta_0}$. Si P_0 domine chacun des P_θ on a bien une représentation

$$\frac{dP_\theta}{dP_0} = \prod_k [1 + X_k(\theta)]$$

comme auparavant.

Supposons alors que chacune des P_θ satisfait aux conditions (R1)-(R4). A chaque P_θ on peut alors associer la mesure Q_θ dont la densité par rapport à $P_0 = Q_0$ est donnée par

$$\exp\{\sum_k [X_k(\theta) - \frac{1}{2}\sigma_k^2(\theta)]\}$$

où l'on a posé $\sigma_k^2(\theta) = E_{k-1}^0 X_k^2(\theta)$. Le Théorème 5.1 dit alors que

$$\sup_\theta \|P_\theta - Q_\theta\| \leq K(b)\sqrt{\varepsilon}$$

donc aussi, si l'on veut normaliser

$$\sup_\theta \|P_\theta - \|Q_\theta\|^{-1} Q_\theta\| \leq 2K(b)\sqrt{\varepsilon} .$$

La famille $\{Q_\theta;\ \theta \in \Theta\}$ posséde une structure bien particulière. En effet, soit M_0 l'espace des mesures réelles μ à support fini sur Θ et telles que $\mu(\Theta) = 0$. On peut plonger Θ dans M_0 par l'application $\theta \rightsquigarrow \delta_\theta - \delta_{\theta_0}$ où δ_θ est la mesure de Dirac à θ.

Soit alors $L_k(\mu) = \int X_k(\theta)\mu(d\theta)$ et soit $B_k(\mu) = E^0_{k-1}[L_k(\mu)]^2$. Finalement soit $L(\mu) = \sum_k L_k(\mu)$ et $B(\mu) = \sum_k B_k(\mu)$. Considérons les mesures $\bar{Q}_\mu$ dont la densité par rapport à $Q_0 = P_0$ est donnée par

$$\frac{d\bar{Q}_\mu}{dQ_0} = \exp\{L(\mu) - \frac{1}{2}B(\mu)\} \quad .$$

Ce sont bien des mesures finies puisque l'inégalité de Dubins et Freedman donne

$$E^0 \exp\{L(\mu) - \frac{1}{2}B(\mu)\} \leq \exp\{\frac{b}{2}\|\mu\|_1 \exp[\|\mu\|_1 - 1]\}$$

avec $\|\mu\|_1$ égal à la variation totale de μ. La famille $\{Q_\theta;\ \theta \in \Theta\}$ est donc la restriction à Θ d'une famille $\{\bar{Q}_\mu;\ \mu \in M_0\}$ qui présente sur M_0 une structure tout à fait analogue à celle d'une famille Gaussienne, excepté que la forme quadratique B est une forme aléatoire.

Notons que $E^0_{k-1}[L_k(\mu)]^2 = B_k(\mu)$ implique que $L_k(\mu)$ s'annule presque surement en même temps que $B_k(\mu)$. Ceci veut donc dire qu'en faisant tendre ε vers zéro, on obtient des familles $\{P_\theta;\ \theta \in \Theta\}$ qui sont B-approximables au sens de [15] Chapitre 11.

Ce sont des familles où les logarithmes des rapports de vraisemblance admettent une approximation quadratique et où les propriétés statistiques sont, en première approximation, entièrement déterminées par le comportement des formes quadratiques B.

Pour de telles familles on dispose de nombre de théorèmes qui proviennent principalement du fait que 1) les distributions <u>a posteriori</u>

associées aux $\{Q_\theta;\ \theta \in \Theta\}$ ont la même structure que si l'expérience en question était Gaussienne et 2) sous des conditions supplémentaires sur la dimension des espaces (aléatoires) de Hilbert associés aux formes B, et sur l'existence d'estimateurs adéquats, on peut procéder à la construction par solution d'équations linéaires de "variables de centrage" Z telles que

$$\frac{d\bar{Q}_\mu}{dQ_0} = \exp\{-\frac{1}{2}[B(Z-\mu) - B(Z)]\}$$

et obtenir pour ces variables des propriétés analogues à celles que l'on attribue souvent au maximum de vraisemblance.

Ceci ne demande rien de plus localement que les conditions (R1)-(R4) et des conditions de dimensionalité.

Toutefois, il existe une classe particulière de familles du type $\{\bar{Q}_\mu;\ \mu \in M_0\}$ appelées expériences <u>Gaussiennes</u> <u>mixtes</u>. Ce sont celles où, conditionellement étant donné les $B(\mu)$, les variables $L(\mu)$ sont Gaussiennes d'espérance zéro et variance $B(\mu)$.

On ne peut espérer que cette condition supplémentaire soit satisfaite sous les seules restrictions (R1)-(R4). Elle est pourtant satisfaite asymptotiquement dans de nombreux cas, comme l'a démontré P. Jeganathan [8]. Dans les cas paramétriques classiques, on utilise des transformations de l'espace des paramétres par des homotéthies telles que la multiplication par $\sqrt{n}$, ou des transformations linéaires plus compliquées. Dans de tels cas, les expériences limites satisfont à des conditions d'invariance par translation qui entrainent que, si elles ont la structure quadratique décrite ci-dessus, elles sont aussi "Gaussiennes mixtes". Toutefois, les raisonnements de ce type entrainent seulement des convergences <u>faibles</u> vers les expériences Gaussiennes mixtes. Nous donnerons, ci-dessous, des conditions qui entrainent des convergences plus fortes.

6. Variables dépendantes. Approximations Gaussiennes mixtes.

Dans cette section nous considérerons une famille $\mathcal{E} = \{P_\theta;\ \theta \in \Theta\}$ où les P_θ sont toutes dominées par $P_0 = P_{\theta_0}$. On suppose que les P_θ sont données sur une tribu $\mathcal{A}$, la dernière dans la filtration $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots \subset \mathcal{A}_k \subset \cdots \subset \mathcal{A}$. On écrit $\frac{dP_\theta}{dP_0} = \prod_k [1 + X_k(\theta)]$ comme expliqué en fin de la Section 5 et on construit les formes linéaires L et quadratiques B comme cela a été décrit.

Nous supposerons que chaque P_θ satisfait aux conditions (R1)-(R4) de la Section 5 et supposerons en plus que la condition suivante est satisfaite.

Supposons que la filtration démarre par une tribu $\mathcal{A}_0 \subset \mathcal{A}_1$ au lieu de par $\mathcal{A}_1$.

(R5) Sur $\mathcal{A}_0$ toutes les P_θ coincident avec P_0. Pour chaque $\mu \in \mathcal{M}_0$, la valeur $B(\mu)$ de la forme quadratique B est $\mathcal{A}_0$-mesurable.

Le condition (R5) est une condition très forte. Elle est satisfaite dans certains cas spéciaux tels que celui d'observations prises sequentiellement avec temps d'arrêt indépendant des observations elles-mêmes et quelques autres exemples analogues. Au prix de quelques modifications des démonstrations, on peut remplacer (R5) par des conditions applicables à beaucoup de problèmes plus généraux.

Une des modifications possible est la suivante:

(R'5) Sur $\mathcal{A}_0$ toutes les P_θ coincident avec P_0. Il existe une forme quadratique B^* définie sur $\mathcal{M}_0$ telle que

1) pour chaque $\mu \in \mathcal{M}_0$, $B^*(\mu)$ est $\mathcal{A}_0$-mesurable,
2) la différence B^*-B est une forme positive,
3) Si $\mu = \delta_t - \delta_s$, $(s,t) \in \Theta\times\Theta$ alors $B^*(\mu) - B(\mu) \leq \varepsilon$.

Une condition de cette espèce est souvent satisfaite dans les cas où on considère une filtration donnée par un suite infinite $A_1 \subset A_2 \subset \cdots \subset A_k \subset \cdots$, notre A étant un des termes de la suite $\{A_k\}$ et B^* étant la limite des formes définies par les A_k successives quand k tend vers l'infini.

Une autre modification possible de (R5) est la suivante. Posons $B_k(\mu) = E^0_{k-1}\left|\int X_k(\theta)\mu(d\theta)\right|^2$, de sorte que $B = \sum_k B_k$.

(R"5) <u>Sur</u> A_0 <u>toutes les</u> P_θ <u>coincident avec</u> P_0. <u>Il existe des formes quadratiques positives</u> B^*_k <u>telles que</u>

1) <u>pour chaque</u> μ, $B^*_k(\mu)$ <u>est</u> A_{k-1} <u>mesurable et</u> $\sum_k B^*_k(\mu)$ <u>est</u> A_0<u>-mesurable</u>

2) <u>Si</u> $\mu = \delta_t - \delta_s$, $(s,t) \in \Theta\times\Theta$, <u>alors</u> $\sum_k |B^*_k(\mu) - B_k(\mu)| \leq \varepsilon$.

Les constructions et démonstrations qui suivent pourraient se faire sous (R"5) en substituant les B^*_k aux B_k. Sous (R'5), il suffit de remplacer des sommes $\sum_{j>k} B_j$ par les différences $B^* - \sum_{j\leq k} B_j$. Cela ne change rien à l'essentiel des démonstrations. Nous utiliserons donc (R5) pour des raisons de simplicité de notation.

Sous cette condition (R5) la forme B est la somme de formes B_k données par $B_k(\mu) = E^0_{k-1}\left|\int X_k(\theta)\mu(d\theta)\right|^2$. La somme $\sum_{j\leq k} B_j$ est donc A_{k-1} mesurable. Sous la condition (R5) il en est donc de même de la somme $\sum_{j>k} B_j$. Ceci va nous permettre de construire une expérience "Gaussienne mixte" associée aux formes B_k.

Les constructions et les démonstrations utilisent seulement les lois sous P_0 des processus aléatoires $(k,\theta,\mu) \rightsquigarrow [X_k(\theta), B_k(\mu)]$. On peut donc, au lieu de travailler sur un espace quelconque, travailler sur un espace produit $\{\Omega, A'_0\} \times \{X', A'\}$ avec sur (Ω, A'_0) une mesure de probabilité P et sur (X', A') une filtration $A'_1 \subset A'_2 \subset \cdots \subset A'$. De plus, en prenant pour X' un espace de trajectoires pour les processus décrits plus haut, on peut supposer que la mesure de base P_0 a été desintégrée sous la forme $P_0(d\omega,dx) = P(d\omega)F_\omega(dx)$ où $\omega \rightsquigarrow F_\omega$ est une distribution conditionnelle régulière des $x' \in \mathcal{X}'$ étant donné $\omega \in \Omega$, c'est à dire, étant donné $\mathcal{A}'_0$.

La condition (R5) revient alors à dire que la forme aléatoire B est une application $\mathcal{A}'_0$ mesurable $\omega \rightsquigarrow B^\omega$ dans l'espace des formes quadratiques positives sur $\mathcal{M}_0$. De même les formes aléatoires B_k sont des applications $\mathcal{A}'_0 \times \mathcal{A}'_{k-1}$ mesurables, $(\omega,x) \rightsquigarrow B_k^{\omega,x}$ telles que $B^\omega = \sum_k B_k^{\omega,x}$.

Sous ces conditions fixons un ω et construisons l'espace de Hilbert $\mathcal{H}^\omega$ complétion de $\mathcal{M}_0$ pour la norme issue de B^ω. Pour chaque k et x on a $B_k^{\omega,x}(\mu) \leq B^\omega(\mu)$. Il existe donc une application linéaire symétrique $C_k^{\omega,x}$ de $\mathcal{H}^\omega$ dans lui-même telle que le produit scalaire $\langle\lambda,C_k^{\omega,x}\mu\rangle$ pris dans $\mathcal{H}^\omega$ est le produit scalaire associé à $B_k^{\omega,x}$. L'application $C_k^{\omega,x}$ posséde une racine carrée symétrique de type positif, soit $\Gamma_k^{\omega,x}$, telle que $\langle\lambda,C_k^{\omega,x}\mu\rangle \equiv \langle\Gamma_k^{\omega,x}\lambda,\Gamma_k^{\omega,x}\mu\rangle$. Cette application $\Gamma_k^{\omega,x}$ est bien déterminée.

Supposons maintenant que nous possédons aussi un stock inépuisable de variables aléatoires Gaussiennes centrées, toutes indépendantes et indépendantes des observations dans $\mathcal{A}'_0 \times \mathcal{A}'$. A partir de ces variables construisons une suite <u>indépendante</u> $\{\eta_k\}$ de processus Gaussiens canoniques de $\mathcal{H}^\omega$. L'image $\Gamma_k^{\omega,x}\eta_k$ est un processus Gaussien tel que

$$E_{k-1}|\langle\lambda,\Gamma_k^{\omega,x}\eta_k\rangle|^2 = B_k^{\omega,x}(\lambda) .$$

<u>Ici</u> E_{k-1} <u>représente une espérance mathématique conditionelle pour la tribu</u> $\mathcal{A}_{k-1}$ <u>engendrée par</u> $\mathcal{A}'_0 \times \mathcal{A}'_{k-1}$ <u>et par les processus</u> η_j <u>d'indice</u> $j < k$.

Cela étant on peut définir pour tout $\mu \in M_0$ une mesure $\bar{G}_\mu$ dont la densité (par rapport à P_0 multiplié par la distribution des processus Gaussiens) est

$$\frac{d\bar{G}_\mu}{d\bar{G}_0} = \exp\{\sum_k [\langle \mu, \Gamma_k^{\omega,x} \eta_k \rangle - \frac{1}{2} B_k^{\omega,x}(\mu)]\}$$

avec quelques abus de notation qui ne devraient pas prêter à confusion.

Une remarque essentielle est la suivante:

<u>Pour un</u> (ω,x) <u>donné, la variable aléatoire</u> $\sum_{j>k} \langle \mu, \Gamma_j^{\omega,x} \eta_j \rangle$ <u>a une distribution Gaussienne centrée de variance</u> $\sum_{j>k} B_j^{\omega,x}(\mu)$ (donc A_{k-1} mesurable).

Il en résulte que pour toute fonction A_{k-1} mesurable ϕ on aura $E_{k-1}\phi[\sum_{j>k} \langle \mu, \Gamma_j^{\omega,x} \eta_j \rangle] = E_{k-1}\phi[\langle \mu, Z \rangle]$ où Z est obtenu à partir d'un processus Gaussien canonique de H^ω par application d'une transformation A_{k-1} mesurable.

La construction qui vient d'être décrite donne bien une certaine expérience Gaussienne mixte $\{\bar{G}_\mu;\ \mu \in M_0\}$. On peut la restreindre à Θ en posant $G_\theta = \bar{G}_\mu$ pour $\mu = \delta_\theta - \delta_{\theta_0}$. Cela donne alors une expérience $G = \{G_\theta;\ \theta \in \Theta\}$ indexée par Θ.

On peut conjecturer que sous les hypothèses (R1)-(R5) l'expérience G va peu différer de l'expérience $E = \{P_\theta;\ \theta \in \Theta\}$. Toutefois nous ne savons démontrer cela que pour les distances Δ_m et Δ_H introduites dans la Section 4, on encore pour la distance Δ elle-même mais sous des conditions de précompacité.

Rappelons que la distance $\Delta_m(E,G)$ peut être définie comme suit. Soit $f_\theta = \frac{dP_\theta}{dP_0}$ et $g_\theta = \frac{dG_\theta}{dG_0}$. Soient $C(j,\theta)$ des nombres $C(j,\theta) \in [0,1]$ et soit $\pi = \{\pi_\theta;\ \theta \in \Theta\}$ une mesure de probabilité à support fini sur Θ.

Alors $\Delta_m(E,G)$ est le supremum des différences

$$\left| E \sup_{j \in J} \sum_{\theta} C(j,\theta)\pi_\theta f_\theta - E \sup_{j \in J} \sum_{\theta} C(j,\theta)\pi_\theta g_\theta \right|$$

où les C et π varient sans restriction dans leur domaine excepté que la cardinalité de l'ensemble J ne peut excéder m. Ici la première espérance mathématique est prise sous F_0, la seconde sous G_0.

THÉORÈME 6.1. Supposons que les conditions (R1)-(R5) soient toutes satisfaites et que G soit l'expérience Gaussienne mixte décrite ci dessus. Alors il existe des coefficients $K(b)$ dépendant seulement de b tels que

$$\Delta_m(E,G) \leq m^2 K(b)\varepsilon^{1/12} .$$

Le théorème sera la conséquence d'une série de lemmes demontrés ci dessous. Notons encore que la démonstration montre aussi que la condition très forte (R5) n'a pas besoin d'être exactement satisfaite. Il suffit qu'elle le soit d'une manière assez approchée comme indiqué dans la modification (R"5). En fait nous utiliserons les mesures tronquées P_θ^a données par le Lemme 5.2. Pour ces mesures (R5) n'est satisfaite que de façon approchée. La valeur de a sera choisie ultérieurement.

Reprenant les applications $\sum_\theta C(j,\theta)\pi_\theta f_\theta$, nous allons considérer qu'elles transforment les rapports de vraisemblance en variables aléatoires à valeurs dans $\mathbb{R}^J$. On munira $\mathbb{R}^J$ de sa norme Euclidienne habituelle donnée par $|z|^2 = \sum_{j \in J} |z_j|^2$. Les fonctions coordonnées ℓ_j sont alors des fonctions Lipschitziennes telles que $|\ell_j(z)-\ell_j(z')| \leq |z-z'|$. Il en est donc de même des fonctions $\sup_j \ell_j$ qui interviennent dans la définition de Δ_m.

Sur $\mathbb{R}^J$ on peut approcher toute fonction Lipschitzienne bornée par des fonctions indéfiniment dérivables.

En effet, soit ψ définie sur $\mathbb{R}^J$ telle que $|\psi(z)| \leq a$ et $|\psi(z)-\psi(z')| \leq |z-z'|$. Soit Z une variable Gaussienne à valeurs dans $\mathbb{R}^J$ et telle que sa densité soit $\frac{1}{(2\pi)^{m/2}} \exp\{-\frac{1}{2}\|z\|^2\}$ pour une norme $\|z\|$ telle que $\|z\| = \frac{1}{\sigma}|z|$. On supposera $\sigma < 1$.

LEMME 6.1. <u>Soit</u> ψ <u>satisfaisant aux conditions</u> $|\psi| \leq c$ <u>et</u> $|\psi(z)-\psi(z')| \leq |z-z'|$. <u>Soit</u> $\phi(z) = E\,\psi(z+Z)$. <u>Alors</u>

1) $|\phi(z)-\psi(z)| \leq \sigma\sqrt{m}$,

2) <u>la dérivée seconde</u> $\ddot{\phi}$ <u>de</u> ϕ <u>est bornée au sens que</u>

$$|u'\ddot{\phi}(z)v| \leq \frac{2c}{\sigma^2}|u|\,|v| \;,$$

3) <u>la derivée</u> $\ddot{\phi}$ <u>est Lipschitzienne</u>:

$$|u[\ddot{\phi}(z)-\ddot{\phi}(z')]v| \leq \frac{2}{\sigma^2}|u|\,|v|\,|z-z'| \quad .$$

<u>Démonstration</u>. On a

$$E|\psi(z+Z)-\psi(z)| \leq E|Z| = \sigma E\|Z\| \quad .$$

Toutefois $E\|Z\|^2 = m$. D'où le premier résultat. Pour le second notons que la dérivée seconde $\ddot{\phi}$ peut s'écrire

$$u'\ddot{\phi}(z)v = E\;\psi(z+Z)[\langle z,u\rangle\langle z,v\rangle-\langle u,v\rangle]$$

pour les produits scalaires associés à la norme $\|\cdot\|$. On a donc

$$|u'\ddot{\phi}(z)v| \leq cE\{|\langle Z,u\rangle\langle Z,v\rangle| + |\langle u,v\rangle|\}$$

et

$$E|\langle Z,u\rangle\langle Z,v\rangle| \leq \|u\|\|v\| \quad .$$

On a aussi

$$|u'[\ddot{\phi}(z)-\ddot{\phi}(z')]v| \leq |z-z'|[\|u\|\|v\| + |u||v|]$$

d'où le résultat.

Nous reviendrons plus loin sur le problème créé par le fait que la fonction $\sup_j \ell_j(z)$ est bien Lipschitzienne mais non bornée. Pour le moment considérons des rapports de vraisemblance f_θ et g_θ et une application dans $\mathbb{R}^J$ donnée par $S = \sum_\theta C(\theta)\pi_\theta f_\theta^a$ où les $C(\theta)$ ont des coordonnées $C_j(\theta)$ telles que $0 \leq C_j(\theta) \leq 1$ et où $\pi = \{\pi_\theta\}$ est une probabilité à support fini sur Θ.

Nous allons prendre pour les f_θ^a des rapports $\dfrac{dP_\theta^a}{dP_0} = \Pi[1 + X_k^a(\theta)]$ où les X_k^a sont des versions tronquées des $X_k(\theta)$ suivant le procédé du Lemme 5.2 de telle sorte que si $U_k(\theta) = \prod_{i<k} [1 + X_i^a(\theta)]$ on ait $\sup_k U_k(\theta) \leq 2a$.

Posons aussi $W_k(\theta) = \exp\{\sum_{i>k} \langle\mu_\theta, \Gamma_i \eta_i\rangle - \frac{1}{2}\sigma_i^2(\theta)\}$ et $V_k(\theta) = \exp\{\langle\mu_\theta, \Gamma_k\eta_k\rangle - \frac{1}{2}\sigma_k^2(\theta)\} - 1$, avec $\mu_\theta = \delta_\theta - \delta_{\theta_0}$.

À ces variables nous allons appliquer la méthode de Lindeberg. Pour cela, on se servira des notations suivantes

a) $S_k = \sum_\theta C(\theta)\pi_\theta U_k(\theta)[1 + X_k^a(\theta)]W_k(\theta)$,

b) $T_k = \sum_\theta C(\theta)\pi_\theta U_k(\theta)[1 + V_k(\theta)]W_k(\theta)$,

c) $T = \sum_\theta C(\theta)\pi_\theta g_\theta$,

d) $R_k = \sum_\theta C(\theta)\pi_\theta U_k(\theta)W_k(\theta)$,

e) $Z_k = T_k - R_k = \sum_\theta C(\theta)\pi_\theta U_k(\theta)W_k(\theta)V_k(\theta)$,

f) $Y_k = S_k - R_k = \sum_\theta C(\theta)\pi_\theta U_k(\theta)W_k(\theta)X_k^a(\theta)$.

Pour toute fonction ϕ on peut écrire

$$\begin{aligned}\phi(T)-\phi(S) &= \sum_k [\phi(T_k)-\phi(S_k)] \\ &= \sum_k \{[\phi(T_k)-\phi(R_k)] - [\phi(S_k)-\phi(R_k)]\} .\end{aligned}$$

Si ϕ est une fonction bornée dont la derivée seconde est bornée et Lipschitzienne, un terme tel que $\phi(T_k)-\phi(R_k)$ possède un developpement de Taylor

$$\begin{aligned}\phi(T_k)-\phi(R_k) &= \dot{\phi}(R_k)(T_k-R_k) + \frac{1}{2}(T_k-R_k)'\ddot{\phi}(R_k)(T_k-R_k) \\ &\quad + \frac{1}{2}(T_k-R_k)'[\ddot{\phi}(R_k^*)-\ddot{\phi}(R_k)](T_k-R_k)\end{aligned}$$

où R_k^* est situé entre T_k et R_k.

Si K est un nombre tel que

$$|u'\ddot{\phi}(z)v - u'\ddot{\phi}(z')v| \leq K|z-z'||u||v| \quad ,$$

le terme du troisième ordre est inférieur en valeur absolue à $\frac{1}{2}K|Z_k|^3$. Nous allons d'abord montrer que ces termes peuvent être négligés.

LEMME 6.2. <u>Sous les conditions</u> (R1)-(R5), <u>il existe un nombre</u> $K(b)$, <u>ne dépendant que de</u> b, <u>tel que</u>

$$\sum_k E^0|Z_k|^3 \leq 8a^3m^{3/2}K(b)\sqrt{\varepsilon} \quad .$$

<u>Démonstration</u>. La norme $|Z_k|$ est au plus égale à $\sqrt{m}$ fois la norme maximum des coordonnées. On a donc

$$|Z_k| \leq (\sqrt{m})\sum_\theta \pi_\theta U_k(\theta)W_k(\theta)|V_k(\theta)| \quad .$$

Ceci entraine

$$|Z_k|^3 \leq m^{3/2} \sum_\theta \pi_\theta U_k^3(\theta) W_k^3(\theta) |V_k(\theta)|^3 .$$

Prenant l'espérance conditionnelle E_{k-1}, on voit que

$$E_{k-1}|Z_k|^3 \leq m^{3/2} \sum_\theta \pi_\theta U_k^3(\theta) E_{k-1}|V_k(\theta)|^3 E_k W_k^3(\theta) .$$

Ici $E_k W_k^3(\theta)$ est dèjà A_{k-1} mesurable. La variable $W_k(\theta)$ peut se mettre sous la forme $W_k(\theta) = \exp\{\xi - \frac{1}{2}\beta^2\}$ où ξ est conditionellement, distribué suivant $N(0,\beta^2)$ et où $\beta^2 = \sum_{j>k} \sigma_j^2(\theta)$ est A_{k-1} mesurable. On a donc $E_k W_k^3(\theta) = \exp\{3\beta^2\} \leq \exp\{3b\}$.

Pour borner $E_{k-1}|V_k(\theta)|^3$ on peut mettre $V_k(\theta)$ sous une forme analogue, soit $V_k(\theta) = \exp\{\eta - \frac{1}{2}\sigma^2\} - 1$ avec η Gaussienne centrée de variance $\sigma^2 = \sigma_k^2(\theta)$. On a

$$\begin{aligned} E_{k-1}|V_k(\theta)|^3 &= E_{k-1}V_k^3(\theta) + 2E_{k-1}[V_k^3(\theta)]^- \\ &= (e^{3\sigma^2}-1) - 3(e^{\sigma^2}-1) + 2E_{k-1}[V_k^3(\theta)]^- . \end{aligned}$$

De plus $[V_k(\theta)]^- \leq [\eta - \frac{1}{2}\sigma^2]^-$. Puisque $\sigma^2 \leq 1$, on en déduit qu'il existe un coefficient K_1 tel que $E_{k-1}|V_k(\theta)|^3 \leq K_1 \sigma_k^3(\theta)$.

On a aussi $U_k(\theta) = \prod_{j<k} [1 + X_j^a(\theta)] \leq 2a$. Il vient donc

$$\sum_k E_{k-1}|Z_k|^3 \leq 8a^3 m^{3/2} K_1 \sum_{\theta,k} \pi_\theta \sigma_k^3(\theta) .$$

Dans cette expression on peut remplacer $\sum_k \sigma_k^3(\theta)$ par la borne $b \max_k \sigma_k(\theta)$. Notons alors que

$$\sigma_k^2(\theta) \leq \varepsilon^2 + E_{k-1} I[|X_k(\theta)| > \varepsilon] .$$

Donc

$$|E \max_k \sigma_k(\theta)|^2 \leq E \max_k \sigma_k^2(\theta) \leq \varepsilon^2 + \varepsilon \quad ,$$

d'après (R1). Cela entraine bien le résultat desiré.

Un raisonnement tout à fait analogue mais un peu plus simple montre que les différences Y_k jouissent d'une propriété analogue, à savoir:

LEMME 6.3. <u>Sous les conditions</u> (R1)-(R5) <u>on a</u>

$$\sum_k E|Y_k|^3 \leq 8a^3m^{3/2}K(b)\varepsilon \ .$$

En effet $E_{k-1}|X_k^a(\theta)|^3 \leq \varepsilon\sigma_k^2(\theta) + E_{k-1}\{|X_k^a(\theta)| > \varepsilon\}$.

Il résulte de ces deux lemmes que si l'on veut montrer que $E[\phi(S)-\phi(T)]$ est petit, pour une fonction ϕ satisfaisant aux conditions énoncées, il suffira de montrer que les différences

$$\sum_k E[\dot{\phi}(R_k)Z_k - \dot{\phi}(R_k)Y_k]$$

et

$$\sum_k E[Z_k'\ddot{\phi}(R_k)Z_k - Y_k'\ddot{\phi}(R_k)Y_k]$$

sont petites.

Pour la première différence cela est immédiat puisque $E_{k-1}Z_k = E_{k-1}Y_k = 0$ et R_k est $\mathcal{A}_{k-1}$-mesurable. Pour les termes du second order il faudra montrer que les matrices $E_{k-1}Z_kZ_k'$ et $E_{k-1}Y_kY_k'$ différent peu. C'est bien exact mais demande un peu de travail. Nous allons commencer par calculer ces matrices.

Un terme de $E_{k-1}Z_kZ_k'$ se met sous la forme

$$\sum_s \sum_t C_i(s)C_j(t)\pi_s\pi_t U_k(s)U_k(t)[E_{k-1}V_k(s)V_k(t)] \times [E_kW_k(s)W_k(t)]$$

avec un terme $E_kW_k(s)W_k(t)$ qui est $\mathcal{A}_{k-1}$-mesurable. Conditionnellement le produit $V_k(s)V_k(t)$ est de la forme

$$V_k(s)V_k(t) = [\exp\{\xi_s - \frac{1}{2}\sigma_k^2(s)\} - 1][\exp\{\xi_t - \frac{1}{2}\sigma_k^2(t)\} - 1]$$

où les (ξ_s, ξ_t) ont une certaine distribution Gaussienne centrée. Ceci donne

$$E_{k-1}V_k(s)V_k(t) = \exp\{\mathrm{Cov}_k(\xi_s, \xi_t)\} - 1$$

où Cov_k est la covariance conditionnelle de ξ_s et ξ_t. Mais, par construction, cette covariance n'est autre que la covariance conditionnelle

$$\sigma_k(s,t) = E_{k-1}X_k(s)X_k(t) .$$

On a donc

$$\begin{aligned} E_{k-1}V_k(s)V_k(t) &= \exp\{\sigma_k(s,t)\} - 1 \\ &= \sigma_k(s,t) + r_k(s,t) , \end{aligned}$$

avec

$$r_k(s,t) = \exp\{\sigma_k(s,t)\} - 1 - \sigma_k(s,t) .$$

Le terme correspondant de la matrice $E_{k-1}Y_kY_k'$ prend la forme

$$\sum_s \sum_t C_i(s)C_j(t)\pi_s\pi_t U_k(s)U_k(t)E_kW_k(s)W_k(t) \times E_{k-1}X_k^a(s)X_k^a(t) .$$

Ici $E_{k-1}X_k^a(s)X_k^a(t) = I_k(s)I_k(t)\sigma_k(s,t)$ pour les indicateurs $I_k(s)$ et $I_k(t)$ qui produisent les troncations des $X_k(s)$ et $X_k(t)$ respectivement.

On a encore $U_k(s)$ et $U_k(t)$ au plus égaux à $2a$ et $E_kW_k(s)W_k(t) \leq \exp\{b\}$. Ceci conduit au résultat suivant.

Soit $M_{k,i,j}$ le terme d'indice i, j de la matrice $M_k = E_{k-1}Z_kZ_k'$ et soit $M^a_{k,i,j}$ le terme correspondant de la matrice $E_{k-1}Y_kY_k'$.

LEMME 6.4. Sous les conditions (R1)-(R5) on a

$$\sum_k E|M_{k,i,j} - M^a_{k,i,j}| \leq 4a^2 be^{2b}[\varepsilon(1+\varepsilon) + 2\,\frac{e^{(2^n-1-n)b}}{a^n}] \;,$$

pour tout entier $n \geq 1$.

Démonstration. D'après le calcul précedent la différence $D_k = |M_{k,i,j} - M^a_{k,i,j}|$ est au plus égale à

$$4a^2 e^{2b}\{\sum_s \sum_t \pi_s \pi_t r_k(s,t) + \sum_s \sum_t \pi_s \pi_t [I - I_k(s) I_k(t)] |\sigma_k(s,t)|\} \;.$$

Bornons $|\sigma_k(s,t)|$ par $\sigma_k(s)\sigma_k(t)$. On a alors

$$\sum_k [I - I_k(s) I_k(t)] |\sigma_k(s,t)| \leq [\sum_k \sigma_k(s)\sigma_k(t)] \max_k [I - I_k(s) I_k(t)]$$
$$\leq b \max_k [I - I_k(s) I_k(t)] \;.$$

D'après le Lemme 5.3 on a

$$E \max_k [I - I_k(s) I_k(t)] \leq 2\,\frac{e^{(2^n-1-n)b}}{a^n}$$

pour tout entier n.

Pour le terme en $r_k(s,t) = \exp\{\sigma_k(s,t)\} - 1 - \sigma_k(s,t)$ on peut borner $r_k(s,t)$ par $\sigma_k^2(s,t) \leq \sigma_k^2(s)\sigma_k^2(t)$. Ceci donne

$$\sum_k r_k(s,t) \leq \sum_k \sigma_k^2(s)\sigma_k^2(t)$$
$$\leq b \max_k \sigma_k^2(t) \;.$$

Or nous avons vu, dans le démonstration du Lemme 6.2, que $E \max_k \sigma_k^2(t) \leq \varepsilon(1+\varepsilon)$. Ceci donne bien le résultat énoncé.

En rassemblant les résultats établis ci dessus on obtient le corollaire suivant:

COROLLAIRE. *Soit* ϕ *une fonction définie sur* $\mathbb{R}^J$ *telle que les termes de la seconde derivée* $\ddot{\phi}$ *soient tous bornées par un nombre* A_1. *Supposons aussi que* $\ddot{\phi}$ *satisfasse à la condition de Lipschitz* $|u'\ddot{\phi}(z)v-u'\ddot{\phi}(z')v| \leq A_2|z-z'||u||v|$.

Alors, sous les conditions (R1)-(R5) *on aura*

$$|E[\phi(S)-\phi(T)]| \leq 4m^2A_1be^{2b}a^2\{\varepsilon(1+\varepsilon)+2\frac{e^{(2^n-1-n)b}}{a^n}\}$$
$$+ 8m^{3/2}A_2K(b)a^3(\varepsilon+\sqrt{\varepsilon})$$

où K(b) *dépends seulement de* b.

(En fait $K(b) = K_0+K_1b$ pour des constantes universelles K_0 et K_1.)

Il est maintenant facile de compléter la démonstration du Théorème 6.1. Pour cela notons que l'expérience $\mathcal{E}$ elle-même donne une application S_0 des rapports de vraisemblance $f_\theta = \frac{dP_\theta}{dP_0}$ dans $\mathbb{R}^J$ par

$$S_{0,j} = \sum C_j(\theta)\pi_\theta f_\theta \quad .$$

De même $\mathcal{E}^a$ obtenu par troncation donne l'application S de coordonnées $\sum C_j(\theta)\pi_\theta f_\theta^a$ et l'expérience Gaussienne mixte $\mathcal{G}$ donne $\sum C_j(\theta)\pi_\theta g_\theta$. Les fonctions dont nous comparons les espérances sont simplement les fonctions $\psi(z) = \sup_j \ell_j(z)$, maximum des coordonnées de $\mathbb{R}^J$. Ces fonctions ne sont pas bornées. Pour les borner remplaçons ψ par ψ_γ défini par $\psi_\gamma(z) = \sup_j [\gamma \wedge \ell_j(z)]$.

LEMME 6.5. <u>Sous les conditions</u> (R1)-(R5) <u>on a</u>

$$E|\psi(S)-\psi_\gamma(S)| \le \frac{e^{(2^n-1-n)b}}{\gamma^{n-1}}$$

$$E|\psi(T)-\psi_\gamma(T)| \le \frac{e^{\frac{1}{2}n(n-1)b}}{\gamma^{n-1}}$$

<u>pour tout entier</u> n.

<u>Démonstration</u>. En fait on a toujours $f_\theta^a \le 2a$. La première inégalité n'a donc d'intérêt que si $\gamma > 2a$, mais elle s'applique aussi aux rapports non tronqués f_θ. Pour le montrer notons que chaque coordonnée de S_0 est toujours inférieure à $\sum \pi_\theta f_\theta$. On a donc

$$|\psi(S_0)-\psi_\gamma(S_0)| \le \sum_\theta \pi_\theta f_\theta I[f_\theta > \gamma] .$$

D'après la démonstration du Lemme 5.3 on a $E\ f_\theta^n \le e^{(2^n-1-n)b}$, donc

$$E\ f_\theta I[f_\theta \ge \gamma] \le \frac{e^{(2^n-1-n)b}}{\gamma^{n-1}} .$$

De même, pour les variables Gaussiennes mixtes on aura $E\ g_\theta^n \le \exp \frac{1}{2}n(n-1)b$. Ceci donne les inégalités voulues.

Remplaçons maintenant les $\psi_\gamma(z)$ par $\psi_{\gamma,\sigma}(z) = E\ \psi_\zeta(z+Z)$ où Z a la distribution normale centrée de densité proportionnelle à $\exp\{-\frac{1}{2\sigma^2}|z|^2\}$. D'après le Lemme 6.1 on aura $|\psi_{\gamma,\sigma}(z)-\psi_\gamma(z)| \le \sigma\sqrt{m}$.

La dérivée seconde $\ddot{\psi}_{\gamma,\sigma}$ satisfait à une condition de Lipschitz de coefficient $A_2 = \frac{2}{\sigma^2}$. De plus cette dérivée seconde est bornée par $\frac{2\gamma}{\sigma^2}$. Le corollaire énoncé après le Lemme 6.4 donne donc

$$|E\,\psi_\gamma(S) - E\,\psi_\gamma(T)| \le 2\sigma\sqrt{m} + \frac{8m^2\gamma}{\sigma^2}\, be^{2b} a^2 \{\varepsilon(1+\varepsilon) + \frac{e^{(2^n-1-n)b}}{a^n}\}$$
$$+ 16\,\frac{m^{3/2}}{\sigma^2}\, K(b) a^3 [\varepsilon + \sqrt{\varepsilon}] \ .$$

Dans cette formule on peut prendre $n = 12$, $a = \varepsilon^{-1/12}$, $\sigma = \varepsilon^{1/12}$, $\gamma = \varepsilon^{-1/12}$. On aura alors

$$|E\,\psi_\gamma(S) - E\,\psi_\gamma(T)| \le m^2 K_2(b)\varepsilon^{1/12}$$

pour une certaine fonction K_2 qui dépend seulement de b. Pour repasser aux fonctions non tronquées, utilisons le Lemme 6.5, avec un entier n (différent de celui qui a été pris ci dessus) égal à 2. Cela ajoutera une différence d'ordre $\varepsilon^{1/12}$.

Enfin il nous faut passer des mesures tronquées P_θ^a aux mesures P_θ elles-mêmes. D'après le Lemme 5.2 cela ajoute encore une différence de l'ordre $\frac{2e^b}{a}$ donc aussi d'ordre $\varepsilon^{1/12}$. Le Théorème est donc démontré.

Le Théorème 6.1 dit que les expériences $\mathcal{E}$ et $\mathcal{G}$ sont proches pour certains problèmes statistiques. Toutefois, à part le cas $m = 2$ qui est celui des tests, ces problèmes ne sont peut-être pas d'importance majeure. Introduisons donc d'autres classes de problèmes, plus proches de ceux qui sont souvent traités dans la littérature. Pour cela considérons un nombre D fixé et l'espace vectoriel $\mathbb{R}^D$. Sur $\mathbb{R}^D$ soit $x \rightsquigarrow |x|$ une norme et soit U la boule unité de $\mathbb{R}^D$ pour cette norme. Considérons des fonctions de perte $L(x,\theta)$ définies sur $U \times \Theta$ et une fonction $y \rightsquigarrow \alpha(y)$ définie sur $[0,2]$ et telle que $\alpha(y) \to 0$ si $y \to 0$.

Soit $\mathcal{L}(\alpha,D)$ la classe des fonctions de perte L telles que $0 \le L(x,\theta) \le 1$ et $|L(x,\theta)-L(x',\theta)| \le \alpha[|x-x'|]$ pour tout θ et toute paire (x,x') d'éléments de U.

Cette classe dépend du module de continuité α, de la dimension D et du choix de la norme. Toutefois puisque sur les $\mathbb{R}^D$ les normes sont toutes équivalentes, le choix de la norme est de peu d'intérêt, pourvu que ce soit la même norme qui définisse la boule U et les bornes $\alpha[|x-x'|]$ pour les différences de valeurs des fonction de perte.

Nous appellerons $L^*(\alpha,D)$ la classe des fonctions de perte qui peuvent être obtenues ainsi mais pour tout espace vectoriel $\mathbb{R}^{D'}$ avec $D' \leq D$, c'est à dire que $L^*(\alpha,D) = \cup\{L(\alpha,D'),\ D' \leq D\}$.

Maintenant nous pouvons definir une distance $\Delta_{\alpha,D}(E,G)$ comme suit: C'est la borne inférieure des nombres $\varepsilon \geq 0$ tels que pour chaque fonction de perte de $L^*(\alpha,D)$ toute fonction de risque achevable sur l'une des expériences peut aussi être obtenue sur l'autre à un ε près.

Notons que la classe $L^*(\alpha,\Delta)$ est assez riche. Par exemple si $\alpha(y) = y$, elle comprend la classe des fonctions de perte $\frac{1}{2}|x-\phi(\theta)|$ où ϕ est une application tout à fait arbitraire dans une $U \subset R^{D'}$, $D' \leq D$. A un coéfficient numérique près elle comprend aussi les fonctions de perte $[\frac{1}{2}|x-\phi(\theta)|]^2$ pour des normes Euclidiennes ou non, mais bien sûr avec x et $\phi(\theta)$ dans U.

Pour appliquer le Théorème 6.1, considérons non pas une expérience donnée E mais une famille $\{E_\varepsilon;\ \varepsilon \in]0,1[\}$, $E_\varepsilon = \{P_{\theta,\varepsilon};\ \theta \in \Theta_\varepsilon\}$ de telles expériences.

<u>Ici absolument tout dépendra de ε à moins que le contraire soit specifié de façon explicite</u>.

Nous allons supposer ce qui suit

Hypothèse (H). <u>Il existe un $b \in]1,\infty[$ fixe, indépendant de ε tel que E_ε satisfasse aux conditions (R1)-(R5) pour la valeur $\varepsilon \in (0,1)$ qui entre dans ces conditions</u>.

THÉORÈME 6.2. Supposons b, α et D fixes, indépendants de ε et supposons que E_ε satisfasse à l'hypothèse H, énoncée ci dessus. Soit G_ε l'expérience Gaussienne mixte associée.

Alors $\Delta_{\alpha,D}(E_\varepsilon,G_\varepsilon)$ tend vers zéro quand $\varepsilon \to 0$.

Démonstration. Soit γ un nombre $\gamma \in (0,1)$ fixe. Soit δ, fixe tel que $0 \leq y \leq \delta$ entraine $\alpha(y) < \gamma$. Pour une boule unité de type $U \subset \mathbb{R}^D$ soit $\{u_j;\ j \in J\}$ une partie maximale de U sujétte aux conditions que $|u_i-u_i| > \delta$ si $i \neq j$. Notons que la cardinalité d'un tel ensemble est bornée par un nombre m fixe, indépendant du choix de la norme.

Mais alors pour une fonction L de la classe $L(\alpha,D)$ on aura pour tout $x \in U$ un u_j tel que $|L(x,\theta)-L(u_j,\theta)| \leq \alpha[|x-u_j|] < \gamma$. A γ près on peut donc remplacer la fonction L par une fonction L' où le statisticien n'a qu'un nombre fini $\leq m$ de décisions possibles. Or nous venons de voir que $\Delta_m(E_\varepsilon,G_\varepsilon) \leq K(m,b)\varepsilon^{1/12}$. Donc pour γ fixé on aura

$$\limsup_{\varepsilon\to 0} \Delta_{\alpha,D}(E_\varepsilon,G_\varepsilon) \leq 2\gamma \quad .$$

Ce qui démontre le résultat énoncé.

Dans une telle situation on peut même se permettre de laisser varier D avec ε pourvu que D ne croisse pas trop rapidement. En effet, la borne du Théorème 6.1 a la forme $m^2K(b)\varepsilon^{1/12}$. Si par exemple on prend $\alpha(y) = y$ on peut prendre des dimensions $D(\varepsilon)$ quelconques pourvu que $D(\varepsilon) \leq o|\log \varepsilon|$, par exemple. En effet, le nombre de points d'un réseau à distance γ dans $U \subset \mathbb{R}^D$ est de l'ordre de $\frac{1}{\gamma^D}$.

Ici les espaces vectoriels sont de dimension finie. Nous allons maintenant montrer qu'il est possible d'obtenir un résultat analogue pour la distance Δ_H associée aux problèmes où la fonction de perte a la forme

$\|t-\psi(\theta)\|^2$ pour la norme d'un espace de Hilbert H quelconque et pour une application quelconque ψ de Θ dans la boule unité de H.

La démonstration suit les mêmes étapes que la démonstration du Théorème 6.1, mais les calculs sont un peu différents.

Soit $E = \{P_\theta ; \theta \in \Theta\}$ une expérience où toutes les P_θ sont dominées par $P_0 = P_{\theta_0}$. Soit $f_\theta = \frac{dP_\theta}{dP_0}$ et soit π une probabilité à support fini sur Θ. Si ψ est une application de Θ dans un espace de Hilbert H, pour la fonction de perte $\|t-\psi(\theta)\|^2$, le risque de Bayes prend la forme $E^0 \frac{1}{2} \frac{Q(f)}{L(f)}$ où $Q(f)$ est la forme quadratique

$$Q(f) = \sum_s \sum_t \|\psi(s)-\psi(t)\|^2 \pi_s \pi_t f_s f_t$$

et $L(f)$ est la forme linéaire

$$L(f) = \sum_s \pi_s f_s \, .$$

Si ψ prend ses valeurs dans la boule unité de H on a toujours $Q(f) \leq 4|L(f)|^2$. Considérons alors le vecteur $S = S(f)$ dont la première coordonnée est $Q(f)$ et la seconde coordonnée est $L(f)$. L'expérience Gaussienne mixte donne des formes analogues où f est remplacé par le vecteur g des rapports $g_\theta = \frac{dG_\theta}{dG_0}$. Pour montrer que

$$\left| E \frac{1}{2} \frac{Q(f)}{L(f)} - E \frac{1}{2} \frac{Q(g)}{L(g)} \right|$$

est petit, il suffira de montrer que les lois de $S = S(f)$ et $T = S(g)$ diffèrent peu. En effet $\frac{1}{2} \frac{Q(f)}{L(f)} \leq 2L(f)$ et les $L(f)$ sont équiintégrables, d'après (R2). De même pour $S(g)$.

Nous allons faire la démonstration en suivant pas à pas les étapes de la démonstration du Théorème 6.1 et en utilisant les mêmes notations que pour la démonstration du Lemme 6.2.

Les termes U_k, X_k^a, W_k, V_k ayant la même signification que pour le Lemme 6.2, posons

$$M_k(s,t) = \|\psi(s)-\psi(t)\|^2 \pi_s \pi_t U_k(s) U_k(t) W_k(s) W_k(t) \ .$$

Le vecteur correspondant au S_k du Lemme 6.2 aura maintenant pour coordonnées

$$S_{k,1} = \sum_s \sum_t M_k(s,t)[1+X_k^a(s)][1+X_k^a(t)]$$

et

$$S_{k,2} = \sum_s \pi_s U_k(s) W_k(s)[1+X_k^a(s)] \ .$$

Les coordonnées du terme qui correspond à T_k s'obtiennent en remplaçant $1+X_k^a(s)$ et $1+X_k^a(t)$ respectivement par $1+V_k(s)$ et $1+V_k(t)$. De même le vecteur appelé R_k pour le Lemme 6.2 s'obtient en remplaçant $X_k^a(s)$ et $X_k^a(t)$ par zéro.

La différence appelée Y_k a pour première coordonnée

$$Y_{k,1} = 2 \sum_s \sum_t M_k(s,t) X_k^a(s) + \sum_s \sum_t M_k(s,t) X_k^a(s) X_k^a(t) \ .$$

Pour toute fonction ϕ qui est deux fois dérivable on a encore

$$\phi(S) - \phi(T) = \sum_k [\phi(S_k) - \phi(T_k)]$$

et, par exemple,

$$\phi(S_k) - \phi(R_k) = \dot{\phi}(R_k) Y_k + \frac{1}{2} Y_k' \ddot{\phi}(R_k^*) Y_k$$

avec R_k^* intermédiaire entre R_k et S_k, exactement comme pour le Lemme 6.2, mais alors que les différences Y_k du Lemme 6.2 sont fonctions linéaires des X_k^a, celles qui apparaissent ici ont un terme linéaire et un terme quadratique. Autrement dit $Y_k = \bar{Y}_k + \bar{\bar{Y}}_k$ où $\bar{Y}_k$ est la fonction linéaire

$$\bar{Y}_k = \begin{cases} \sum_s \sum_t M_k(s,t) X_k^a(s) \\ \sum_s \pi_s U_k(s) W_k(s) X_k^a(s) \end{cases}$$

et où $\bar{\bar{Y}}_k$ est le vecteur dont la première coordonnée est $\sum_s \sum_t M_k(s,t) X_k^a(s) X_k^a(t)$, la deuxième étant égale à zéro. On peut donc mettre la différence $\phi(S_k)-\phi(R_k)$ sous la forme $A_k + B_k + C_k$ avec

$$A_k = \dot{\phi}(R_k)\bar{Y}_k + \frac{1}{2}\bar{Y}_k'\ddot{\phi}(R_k)\bar{Y}_k + \dot{\phi}_1(R_k) \sum_s \sum_t M_k(s,t) X_k^a(s) X_k^a(t)$$

ou $\dot{\phi}_1$ est la première coordonnée de $\dot{\phi}$ et

$$B_k = \frac{1}{2}\bar{Y}_k'[\ddot{\phi}(R_k^*)-\ddot{\phi}(R_k)]\bar{Y}_k$$
$$C_k = \bar{Y}_k'\ddot{\phi}(R_k^*)\bar{\bar{Y}}_k + \frac{1}{2}\bar{\bar{Y}}_k'\ddot{\phi}(R_k^*)\bar{\bar{Y}}_k \;.$$

Pour les différences $\phi(T_k)-\phi(R_k)$ on a des expressions analogues, soient $\tilde{A}_k$, $\tilde{B}_k$, $\tilde{C}_k$ où les X_k^a sont remplacés par des V_k et les termes $\bar{Y}_k$ et $\bar{\bar{Y}}_k$ sont devenus $\bar{Z}_k$ et $\bar{\bar{Z}}_k$.

LEMME 6.6. Il existe un nombre $\bar{K}(b)$ dépendant seulement de b tel que si (R1)-(R5) sont satisfaites on a

$$\sum_k E|\bar{Y}_k|^3 \le a^6 \bar{K}(b)\varepsilon$$
$$\sum_k E|\bar{Z}_k|^3 \le a^6 \bar{K}(b)\sqrt{\varepsilon} \;.$$

Démonstration. On peut écrire

$$|\bar{Y}_k| \le 6 \sum_s \sum_t \pi_s \pi_t U_k(s) U_k(t) W_k(s) W_k(t) |X_k^a(s)|$$

$$= 6[\sum_t \pi_t U_k(t)W_k(t)][\sum_s \pi_s U_k(s)W_k(s)|X_k^a(s)|] .$$

Pour $\bar{Z}_k$ on a une expression analogue où X_k^a a été remplacé par V_k. Le résultat est donc une conséquence des calculs des Lemmes 6.2 et 6.3.

Ce lemme permet de borner la somme des termes B_k.

Pour les termes appelés C_k ci-dessus on peut procéder de façon analogue. La matrice $\bar{\bar{Y}}_k\bar{Y}_k'$ n'a que deux termes différents de zéro et $\bar{\bar{Y}}_k\bar{\bar{Y}}_k'$ n'a qu'un seul terme non nul.

On obtient ainsi un résultat du type suivant:

LEMME 6.7. _Soit_ L_2 _une borne pour les termes de la derivée seconde_ $\ddot{\phi}$. _Alors, sous les hypothèses_ (R1)-(R5), _il existe des constantes_ $K_1(b)$ _telles que_

$$\sum_k E|C_k| \le L_2 a^4 K_1(b)\varepsilon$$

$$\sum_k E|\tilde{C}_k| \le L_2 a^4 K_1(b)\sqrt{\varepsilon} .$$

Il suffira donc de regarder ce qui se passe pour les termes appelés A_k et $\tilde{A}_k$. Ici on a encore $E_{k-1}\bar{Y}_k = E_{k-1}\bar{Z}_k = 0$. Pour les termes quadratiques on peut procéder exactement comme pour le Lemma 6.4 et obtenir un résultat du type suivant:

LEMME 6.8. _Soit_ L_1 _une borne pour les termes de la derivée_ $\dot{\phi}$. _Alors, sous_ (R1)-(R5), _il existe un nombre_ $K_2(b)$ _tel que_

$$\sum_k |E(A_k-\tilde{A}_k)| \le (L_1+L_2)a^4K_2(b)[\varepsilon + \frac{e^{(2^n-1-n)b}}{a^n}]$$

pour tout entier $n \ge 1$.

Rassemblant les différents termes on obtient le résultat suivant.

THÉORÈME 6.3. Soit Δ_H la distance obtenue en considérant tous les problèmes d'estimation d'applications ψ de Θ dans des boules unité d'espaces de Hilbert avec fonction de perte égale au carré de norme.

Si les conditions (R1)-(R5) sont satisfaites on a

$$\Delta_H(E,G) \leq K^*(b)\varepsilon^{\frac{1}{22}}$$

pour un coéfficient $K^*(b)$ qui dépend seulement de b.

Démonstration. Soient $S^{(1)}$ et $S^{(2)}$ les coordonnées du vecteur S défini par les expériences tronquées E^a. Le risque pour E^a est de la forme $E\,\frac{1}{2}\,\frac{S^{(1)}}{S^{(2)}}$ avec $S^{(1)} \leq 4(S^{(2)})^2$. Pour les expériences E^a on a toujours $S^{(2)} \leq 2a$. On peut aussi tronquer les rapports de vraisemblance de G à 2a à une erreur d'ordre $\frac{e^{3b}}{a}$ près. Puisque le rapport $S^{(1)}/S^{(2)}$ est toujours inférieur à $4S^{(2)}$ l'ensemble où $S^{(2)} \leq a^{-1}$ ne peut contribuer que $4a^{-1}$ au plus. De même pour T.

Considerons donc la fonction $(x,y) \rightsquigarrow \phi(x,y) = \frac{x}{y}$ dans l'ensemble où $0 < a^{-1} \leq y \leq 2a$ et $0 \leq x \leq 4y^2$ et supposons $a > 8$.

Dans cet ensemble la fonction ϕ a des dérivées premières et secondes bornées par a^2. De plus la dérivée seconde satisfait à une condition de Lipschitz avec un coefficient de Lipschitz L_3 inférieur à $6a^4$. Dans le Lemme 6.8 on peut prendre $n = 11$. Rassemblant les termes obtenus dans les Lemmes (6.6) à (6.8) on obtient une borne du type $K_3(b)[\frac{1}{a}+a^{10}\sqrt{\varepsilon}+a^6\varepsilon]$. Ceci devient $K^*(b)\varepsilon^{1/22}$ en prenant $a = \varepsilon^{-1/22}$. D'où l'énoncé.

Il est à remarquer que le calcul précédent est bien grossier et qu'il devrait être possible d'améliorer la puissance de ε qui entre dans cette formule.

Au cours de la Section 4 nous avons énoncé des théorèmes de convergence (Théorème 4.2) faisant intervenir les distances Δ_m et Δ_H. Ils sont évidemment des conséquences des Théorèmes 6.1 et 6.3 démontrés ici. Une démonstration directe pour le cas des observations indépendantes de la Section 4 peut se faire de façon un peu plus simple. En effet il n'est pas besoin de faire intervenir les variables tronquées X_k^a. Dans des termes tels que les termes $\sum_k \sum_\theta \pi_\theta U_k^3(\theta) W_k^3(\theta) |X_k(\theta)|^3$ que nous avons rencontrés pour le Lemme 6.2 et ses analogues, les U_k, W_k et X_k sont alors des variables indépendantes, ce qui permet de borner les espérances mathématiques de façon un peu plus simple qu'ici. Toutefois la marche générale du raisonnement disponible demeure la même. Pour obtenir de meilleurs résultats, et peut-être montrer que la distance $\Delta(\mathcal{E},\mathcal{G})$ elle même est petite il semble nécessaire d'étudier de plus près le comportement de applications du type $f \rightsquigarrow \sup_{j \in J} \sum_\theta C_j(\theta) \pi_\theta f_\theta$ en fonction du vecteur $f = \{f_\theta;\ \theta \in \Theta\}$. On sent bien, intuitivement, que plus l'ensemble J est grand plus ces fonctions sont "régulières". Il devrait donc être possible d'améliorer de beaucoup le Théorème 6.1. Toutefois, à l'heure actuelle, nous n'avons pas réussi à le faire.

1. Assouad, P. (1983). "Deux remarques sur l'estimation." C.R. Acad. Sc. Paris, vol. 296, pp. 1021-1024.

2. Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A. (1983). "Information and asymptotic efficiency in parametric-non parametric models." Ann. Stat., vol. 11, pp. 432-452.

3. Birgé, L. (1983). "Approximation dans les espaces métriques et théorie de l'estimation." Z. Wahrscheinlichkeitstheorie verw. Gebiete, vol. 65, pp. 181-237.

4. Chernoff, H. (1954). "On the distribution of the likelihood ratio." Ann. Math. Statist., vol. 25, pp. 573-578.

5. Dubins, L. and Freedman, D. (1966). "On the expected value of a stopped martingale." Ann. Math. Statist., vol. 37, pp. 1505-1509.

6. Hájek, J. (1970). "A characterization of the limiting distributions of regular estimates." Z. Wahrscheinlichkeitstheorie verw. Gebiete, vol. 14, pp. 323-330.

7. Ibragimov, I. A. and Has'minskii, R. Z. (1979). Statistical estimation, asymptotic theory. English translation (1981). Springer Verlag.

8. Jeganathan, P. (1982). "On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal." Sankhyā, Series A, vol. 44, pp. 173-212.

9. Koshevnik, Yu. A. and Levit, B. Ya. (1976). "On a non parametric analogue of the information matrix." Theor. Probab. Appl., vol. 21, pp. 738-753.

10. Le Cam, L. (1960). "Locally asymptotically normal families of distributions." Univ. of California Pub. Statistics, vol. 3, pp. 37-98.

11. Le Cam, L. (1964). "Sufficiency and approximate sufficiency." Ann. Math. Statist., vol. 35, pp. 1419-1455.

12. Le Cam, L. (1969). Théorie asymptotique de la décision statistique. Presses de l'Université de Montréal, Montréal, Québec.

13. Le Cam, L. (1975). "On local and global properties in the theory of asymptotic normality of experiments." In Stochastic Processes and related topics. M. Puri, ed. Academic Press. pp. 13-54.

14. Le Cam, L. (1979). "On a theorem of J. Hájek." In The Jaroslav Hájek memorial volume. Acad. Proha, pp. 119-135.

15. Le Cam, L. (1985). Asymptotic Methods in Statistical Decision Theory. Monograph. To be published.

16. Neyman, J. (1949). "Contributions to the theory of the χ^2-test." Proc. Berk. Symp. Math. Stat. Prob., pp. 239-273.

17. Neyman, J. (1954). "Sur une famille de tests asymptotiques des hypothèses statistiques composées." Trabajos de Estadistica, vol. 5, pp. 161-168.

18. Pfanzagl, J. and Wefelmeyer, W. (1982). Contributions to a general asymptotic statistical theory. Springer Verlag.

19. Pfanzagl, J. and Wefelmeyer, W. (1984). Contributions to a general asymptotic statistical theory. Vol. II. Preprint.

20. Torgersen, E. N. (1970). "Comparison of experiments when the parameter space is finite." Z. Wahrscheinlichkeitstheorie verw. Gebiete, vol. 16, pp. 216-249.

21. Wald, A. (1943). "Tests of statistical hypotheses concerning several parameters when the number of observations is large." Trans. Amer. Math. Soc., vol. 34, pp. 426-482.