Statistique et analyse des données

Denis Bosq

Lois limites et efficacité asymptotique des tests hilbertiens de dimension finie sous des hypothèses adjacentes

Statistique et analyse des données, tome 8, nº 1 (1983), p. 1-40

<http://www.numdam.org/item?id=SAD_1983__8_1_1_0>

Statistiques et Analyse de données
1983 - Vol. 8 n° 1 pp. 1-40

LOIS LIMITES ET EFFICACITE ASYMPTOTIQUE DES TESTS HILBERTIENS DE DIMENSION FINIE SOUS DES HYPOTHESES ADJACENTES

Denis BOSQ

Université de Lille
Département de Mathématiques

Résumé : *Dans ce travail, on considère des tests Hilbertiens de dimension finie et on détermine la loi limite de ces tests pour des hypothèses adjacentes à l'hypothèse nulle. On en déduit des résultats relatifs à l'efficacité asymptotique de ces tests.*

On obtient notamment des critères pour qu'un test soit asymptotiquement optimal et l'on compare l'efficacité d'un tel test avec celle du test de Neyman-Pearson.

Abstract : *We consider finite dimensional Hilbertian tests and we obtain their limit distributions for hypothesis "adjacent" to the null hypothesis. As an application we obtain results about the asymptotic efficiency of these tests.*

CLASSIFICATION AMS (MOS) 1980-62F10.

Mots clés : *Tests Hilbertiens, Hypothèses adjacentes, Efficacité asymptotique, Test de Neyman-Pearson.*

Les tests d'ajustement Hilbertiens sont des tests sur χ^2 généralisés, faciles à utiliser et souvent plus efficaces que le test du χ^2 usuel.

Ces tests sont définis de la façon suivante : Soit $X_1,\dots,X_n$, n variables aléatoires indépendantes et de même loi, définies sur l'espace probabilisé (Ω,A,P) et à valeurs dans l'espace mesurable $(E,\mathcal{B})$. On cherche à tester $H_o = \{\mu\}$ contre $H_1 = \mathcal{M} - \{\mu\}$ où $\mathcal{M}$ désigne une famille de Probabilités sur $(E,\mathcal{B})$. Pour cela on se donne $K : E\times E \to \mathbb{R}$ mesurable, bornée et telle que $\int K(x,t)\, d\mu(x) = 1$, $t \in E$ et l'on pose

$$S_n(t) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} K(X_i,t)-1\right] \qquad , \quad t \in E$$

Le test Hilbertien de noyau K est alors défini par la région critique

$$||S_n||^2 > w_n$$

où w_n est un nombre donné et où $||.||$ désigne la semi-norme de $L^2(\mu)$.

Le test du χ^2 basé sur la partition $B_1,\dots,B_k$ correspond au noyau

$$K(x,t) = \sum_{j=1}^{k} \left[\mu(B_j)\right]^{-1} 1_{B_j}(x)\ 1_{B_j}(t) \qquad ; \qquad x,t \in E \ .$$

Dans les publications antérieures [1],[2] et [3] nous avons obtenu des résultats sur le comportement asymptotique de S_n lorsque la loi des X_I est μ ou une loi fixe de la contre-hypothèse.

Le présent travail est consacré à l'étude de propriétés asymptotiques locales au voisinage de l'hypothèse nulle : on se donne $(\nu_n,\ n \geq 1) \subset \mathcal{M}$ et "adjacente" à μ. en un sens qui sera précisé et l'on se propose d'étudier le comportement asymptotique du test lorsque les observations sont faites de la façon suivante :

X_{11} est une v.a. de loi ν_1 ;

X_{11} , X_{22} sont des v.a. indépendantes et de même loi ν_2 ;

...

X_{1n} ,..., X_{nn} sont des v.a. indépendantes et de même loi ν_n ;

...

Dans la suite nous déterminons la loi limite du test lorsque le noyau est de dimension finie. Nous faisons une étude analogue pour le test de Neyman-Pearson.

Comme application nous étudions l'efficacité asymptotique des tests Hilbertiens et nous donnons la forme des tests asymptotiquement optimaux.

L'étude de l'efficacité asymptotique de certains tests analogues aux tests Hilbertiens a été effectuée par G. GREGORY dans [4] et [5].

I - CAS D'UN NOYAU FIXE DE DIMENSION FINIE.-

a) Soit k un entier positif donné et soit E un sous-espace vectoriel de $L^2(\mu)$ qui possède les trois propriétés suivantes :

P_1 - $1 \in E$.

P_2 - E est de dimension $k+1$.

P_3 - Les fonctions appartenant à E sont bornées.

On suppose que le noyau K du test Hilbertien utilisé est le noyau reproduisant de E(1). On lui associe la statistique à valeurs dans $L^2(\mu)$ définie par

$$S_n(.) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} K(X_{in},.) - 1\right] \quad , \quad n \geq 1$$

d'où un test Hilbertien de la forme $||S_n||^2 > w_n$.

Remarquons que $S_n(.)$ est à valeurs dans E' sous-espace vectoriel de E orthogonal à 1.

Dans la suite Y_n désignera une v.a. de loi ν_n et on posera $Z_n = K(Y_n,.) - 1$: cette v.a. est à valeurs dans E'. Enfin $\int K(x,.)\, d\nu_n(x) - 1$ est manifestement dans E'.

b) *Adjacence de* (ν_n)

Nous ferons les hypothèses suivantes :

$\boxed{A1}$ $\sqrt{n}\left[\int K(x,.)\, d\nu_n(x) - 1\right]$ converge dans E' (2)

$\boxed{A2}$ L'opérateur de covariance de Z_n converge simplement vers l'identité de E'.

---

(1) Pour la définition d'un noyau reproduisant voir l'appendice.

(2) ou dans $L^2(\mu)$, ce qui est équivalent ici.

Remarque : Dans les applications la limite de $\sqrt{n}\left[\int K(x,.)\, d\nu_n(x) - 1\right]$ sera un élément non nul de E'.

Lemme 1.- 1) A_1 est équivalente à

$\boxed{A'_1}$ Il existe $(e_j \,;\, j = 1,\dots,k)$ base orthonormale de E', telle que $V_n = \sqrt{n}\left(\int e_1\, d\nu_n, \dots, \int e_k\, d\nu_n\right)$ converge dans $\mathbb{R}^k$.

2) A_2 est équivalente à

$\boxed{A'_2}$ Il existe $(e_j \,;\, j = 1,\dots,k)$ base orthonormale de E', telle que $\int e_j e_{j'}\, d\nu_n - \int e_j\, d\nu_n \int e_{j'}\, d\nu_n \to \delta_{jj'}$ $\quad j,j' = 1,\dots,k$.

Démonstration :

1) Si $\sqrt{n}\left[\int K(x,.)\, d\nu_n(x) - 1\right] \to h$ dans E' et si $(e_j \,;\, j = 1,\dots,k)$ désigne une base orthonormale de E', il vient :

$$\left\|\sqrt{n}\int K(x,.)\, d\nu_n(x) - h\right\|^2 = \sum_{j=1}^{k}\left[\sqrt{n}\int e_j\, d\nu_n - \int e_j\, h\, d\mu\right]^2 \to 0$$

donc V_n converge dans $\mathbb{R}^k$. La réciproque est claire.

2) Soit C_n l'opérateur de covariance de Z_n. S'il converge on en déduit que :

$$E\left[\langle Z_n,f\rangle \langle Z_n,g\rangle\right] - E\langle Z_n,f\rangle \,.\, E\langle Z_n,g\rangle = \langle C_n f,g\rangle \to \langle f,g\rangle \qquad f,\, g \in E'\,;$$

Comme $K-1$ est le noyau reproduisant de E' cela s'écrit encore

$$\int f\, g\, d\nu_n - \int f\, d\nu_n \int g\, d\nu_n \longrightarrow \langle f,g\rangle \quad ; \qquad f,\, g \in E'$$

donc A'_2 est vérifiée.

Inversement, A'_2 implique que, pour tout $f \in E'$ et tout $j \in \{1,\dots,k\}$

$$<C_n f, e_j> = \sum_{j'=1}^{k} (\int e_{j'} \, f \, d\mu) <C_n e_{j'}, e_j>$$

$$\to \sum_{j'=1}^{k} (\int e_{j'} \, f \, d\mu) \, \delta_{jj'} = \int e_j \, f \, d\mu = <e_j, f>$$

et on en déduit que C_n converge vers l'identité. ∎

Lemme 1 bis.- Soit $\nu_n = (1+h_n) \cdot \mu + \eta_n$, $h_n \in L^1(\mu)$, $\eta_n \perp \mu$ la décomposition de Lebesgue de ν_n par rapport à μ.

Alors

A_1, A_2 $\Longleftrightarrow$ $Pr_{E'}(\sqrt{n}\, h_n)$ converge dans E' et $Pr_{E''}(h_n) \to 0$ dans E'' où E'' désigne l'espace vectoriel engendré par les éléments de E' et leurs produits deux à deux. (1)

Démonstration : Comme (η_n) est une suite de mesures étrangères à μ, on peut choisir (e_j) telle que

$$\int e_j \, d\eta_n = \int e_j \, e_{j'} \, d\eta_n = 0 \qquad j,j' = 1,\dots,k \;\; ; \;\; n \geq 1 .$$

Alors il est clair que la convergence de V_n dans $\mathbb{R}^k$ équivaut à celle de $Pr_{E'}(\sqrt{n}\, h_n)$ dans E'.

Pour conclure il suffit de remarquer que

$$\int e_j \, e_{j'} \, d\nu_n - \int e_j \, d\nu_n \int e_{j'} \, d\nu_n = \delta_{jj'} + \int e_j \, e_{j'} \, h_n \, d\mu - \int e_j \, h_n \, d\mu \int e_{j'} \, h_n \, d\mu \qquad n \geq 1 .$$

Les détails sont laissés au lecteur. ∎

---

(1) $Pr_{E'}$ est défini par la formule

$$Pr_{E'}(f)(.) = \int K'(.,t) f(t) d\mu(t) \qquad , \quad f \in L^1(\mu)$$

où K' désigne le noyau de E'. $Pr_{E''}$ se définit d'une manière analogue.

c) *La loi limite*.

Proposition 1.- Si A_1 et A_2 sont vérifiées, on a

$$n^{-1/2}\left(\sum_{i=1}^{n} e_1(X_{in}), \ldots, \sum_{i=1}^{n} e_k(X_{in})\right) \xrightarrow{\mathcal{L}} (U_1 + v_1, \ldots, U_k + v_k)$$

où $U_1,\ldots,U_k$ est un échantillon de $\mathcal{N}(0,1)$ et $(v_1,\ldots,v_k)$ la limite de (V_n).

Et, par conséquent

$$\|S_n\|^2 \xrightarrow{\mathcal{L}} \sum_{j=1}^{k} (U_j + v_j)^2$$

Démonstration : D'après le théorème de SAZONOV (cf. [3] p. 15) on a

$$\sup_{C\in\mathcal{C}} \left|P\left[\sqrt{n}(\hat{A}_n - A_n) \in C\right] - P\left[U_n \in C\right]\right| \leq \gamma_n$$

où $\mathcal{C}$ est l'ensemble des convexes mesurables de $\mathbb{R}^k$,

$$\hat{A}_n = \left(\frac{1}{n}\sum_{i=1}^{n} e_1(X_{in}), \ldots, \frac{1}{n}\sum_{i=1}^{n} e_k(X_{in})\right)$$

$$A_n = \left(\int e_1\, d\nu_n, \ldots, \int e_k\, d\nu_n\right)$$

$U_n = (U_{1n}, \ldots, U_{kn})$ est une variable Gaussienne à k dimensions, centrée et de matrice de covariance .

$$C_{\nu_n} = \left(\int e_j e_{j'}\, d\nu_n - \int e_j\, d\nu_n \int e_{j'}\, d\nu_n\right)_{j,j'=1,\ldots,k}$$

enfin $\gamma_n = \dfrac{8\, c_0\, M^3}{\sqrt{n}} \left[\sum_{\ell=1}^{k} \eta_{\ell n}^{-3/2}\right] k^{9/2}$

où $M = \sup\limits_{\substack{x\in E\\ j=1,\ldots,k}} |e_j(x)|$ et où les $\eta_{\ell n}$ sont les valeurs propres de C_{ν_n} (C_{ν_n} est régulière pour n assez grand puisque son déterminant tend vers 1).

Soit maintenant C un convexe fixe, on considère le convexe mesurable $C_n = C - \sqrt{n}\, A_n$. Alors

$$|P(\sqrt{n}(\hat{A}_n - A_n) \in C_n) - P(U_n \in C_n)| \leq \gamma_n$$

soit

$$|P(\sqrt{n}\, \hat{A}_n \in C) - P(U_n + \sqrt{n}\, A_n \in C)| \leq \gamma_n \qquad (1)$$

Maintenant A_1 et A_2 impliquent la convergence de C_{ν_n} vers I_k matrice identique $k \times k$, au sens de la convergence terme à terme. On en déduit la convergence des valeurs propres de C_{ν_n} vers 1. Par conséquent (γ_n) tend vers 0.

Comme U_n tend en loi vers U Gaussienne centrée de matrice de covariance I_k et comme $\sqrt{n}\, A_n \to v$ où $v = (v_1,\ldots,v_k)$, il vient

$$U_n + \sqrt{n}\, \hat{A}_n \xrightarrow{\mathcal{L}} U + v$$

d'où

$$||S_n||^2 \xrightarrow{\mathcal{L}} ||U + v||^2 \qquad \blacksquare$$

Remarque : Bien entendu la loi limite de $||S_n||^2$ ne dépend pas de la base (e_j) choisie : il s'agit d'un χ^2 décentré à k degrés de liberté dont le paramètre de non centralité vaut
$\lambda^2 = ||v||^2 = \lim_{n\to\infty} n \, ||\int K(x,.)\, d\nu_n(x) - 1||^2$.

d) *Une extension.*

Nous allons voir que la deuxième partie de la proposition 1 reste valable sous l'hypothèse :

$\boxed{B}$ $$\|Pr_{E'}(\sqrt{n}\, h_n)\| \to \lambda \quad \text{et} \quad Pr_{E''}(h_n) \to 0 \quad .$$

Il est facile de vérifier les implications

$$A_1 \ , \ A_2 \Longrightarrow B \Longleftrightarrow A_2 \ , \ A_1''$$

où A_1'' est la condition

$\boxed{A_1''}$ $$\left\|\sqrt{n}\left[\int K(x,.)\, d\nu_n(x) - 1\right]\right\| \to \lambda$$

Remarque : Si $\nu_n \prec \mu$ et $h_n \in E'$, B s'écrit simplement

$$\sqrt{n}\ \left\|1 - \frac{d\nu_n}{d\mu}\right\| \to \lambda \quad .$$

Proposition 2.- Sous la condition B , $\|S_n\|^2$ tend en loi vers un χ^2 décentré à k degrés de libertés et admettant λ^2 comme paramètre de décentrage.

Démonstration : La relation (1) obtenue dans la démonstration de la proposition 1 reste valable, et comme les remarques faites ci-dessus montrent que $C_{\nu_n} \to I_k$ on a encore $\gamma_n \to 0$.

Il suffit donc d'étudier la loi limite de

$$\|U_n + \sqrt{n}\, A_n\|^2 = \sum_{j=1}^{k} \left(U_{nj} + \sqrt{n} \int e_j\, d\nu_n\right)^2 \quad .$$

D'après le théorème 1 p. 88 de BARRA ([6]) la fonction caractéristique de cette v.a. s'écrit

$$\varphi_n(t) = \left[\mathrm{Det}(I_k - 2it\ C_{\nu_n})\right]^{-1/2} \cdot \exp \cdot \left[it \sqrt{n}\ A_n'(I_k - 2it\ C_{\nu_n})^{-1} A_n\right]$$

Le 1^er facteur converge vers $\left[\mathrm{Det}(I_k - 2it\ I_k)\right]^{-1/2} = (1-2it)^{-k/2}$.

Posons $$B_n = (I_k - 2it\ C_{\nu_n})^{-1}$$

alors $$B_n \to (1 - 2it)^{-1}\ I_k$$

On se donne maintenant un $\varepsilon > 0$; pour n assez grand

$$\max_{j,j'} |B_n(j,j') - (1 - 2it)^{-1}\ \delta_{jj'}| \leq \varepsilon$$

d'où

$$\left|n \sum_{j,j'} B_n(j,j')\ A_n(j)\ A_n(j') - n \sum_{j,j'} A_n(j)\ A_n(j')\ \delta_{jj'}(1 - 2it)^{-1}\right|$$
$$\leq n\ \varepsilon \sum_{j,j'} |A_n(j)\ A_n(j')| \qquad (2)$$

mais

$$|\sqrt{n}\ A_n(j)| = |\sqrt{n} \int e_j\ d\nu_n| = |\int \sqrt{n}\ h_n\ e_j\ d\mu| \leq ||Pr_{E'}\ \sqrt{n}\ h_n||$$

donc

$$n\ \varepsilon \sum_{j,j'} |A_n(j)\ A_n(j')| \leq \varepsilon\ ||Pr_{E'}\ \sqrt{n}\ h_n||^2\ k^2$$

comme le deuxième terme du premier membre de (2) vaut $(1-2it)^{-1} \sum_j n\ A_n^2(j)$; il converge vers $(1-2it)^{-1}\ \lambda^2$ et il en est alors de même pour le premier c'est-à-dire pour $A_n'\ B_n\ A_n$.

Finalement

$$\varphi_n(t) \to (1-2it)^{-k/2} \exp\left[it(1-2it)^{-1} \lambda^2\right]$$

qui est la f.c. du χ^2 cherché. ∎

e) Efficacité asymptotique du test.

Soit $\alpha \in \left]0,1\right[$, sous $A_1 - A_2$ ou B la puissance asymptotique du test de niveau asymptotique α s'écrit

$$p_\alpha(K) = P\left[\chi^2(k,\lambda^2) > \chi^2_\alpha\right]$$

où χ^2_α est défini par la relation

$$\alpha = P\left[\chi^2(k,0) > \chi^2_\alpha\right]$$

Pour optimiser cette puissance asymptotique, nous allons nous placer dans le cas où

$$\nu_n = \left(1 + \frac{g_n}{\sqrt{n}}\right) . \mu$$

avec $g_n \to g$ faiblement dans $L^2(\mu)$. On supposera que $\|g\| \neq 0$. Remarquons que la convergence faible de g_n vers g et le fait que $\int 1 . g_n \, d\mu = 0$ implique $\int 1 . g \, d\mu = 0$.

On peut se poser le problème sous la forme suivante :

Soit $\mathcal{K}_g$ l'ensemble des noyaux reproduisants sur les sous-espaces de $L^2(\mu)$ qui vérifient les propriétés P_1 , P_2 , P_3 (où k est fixe).

Il s'agit de déterminer $K \in \mathcal{K}_g$ qui maximise $p_\alpha(K)$; or cette quantité peut s'écrire

$$p_\alpha(K) = P[(U_1 + \lambda)^2 + U_2^2 + \ldots + U_k^2 > \chi_\alpha^2]$$

Il suffit alors de maximiser $p_\alpha(K)$ lorsque l'on conditionne par $U_2^2 + \ldots + U_k^2$. On est donc ramené à la minimisation de

$$P[(U_1 + \lambda)^2 \leq \chi_\alpha^2 - (U_2^2 + \ldots + U_k^2)]$$

Si $\chi_\alpha^2 \leq U_2^2 + \ldots + U_k^2$ n'importe quelle valeur de λ convient. Sinon on est ramené à la minimisation de

$$P[-\sqrt{\chi_\alpha^2 - (U_2^2 + \ldots + U_k^2)} - \lambda \leq U_1 \leq \sqrt{\chi_\alpha^2 - (U_2^2 + \ldots + U_k^2)} - \lambda]$$

Comme $U_1 \sim N(0,1)$ et que l'intervalle $(-\sqrt{\ } - \lambda, \sqrt{\ } - \lambda)$ a pour centre λ on en déduit qu'il faut maximiser $|-\lambda|$ c'est-à-dire maximiser

$$\begin{aligned}\lambda^2 &= \lim_{n\to\infty} \left|\left|\sqrt{n} \left[\int K(x,.)\, d\nu_n(x) - 1\right]\right|\right|^2 \\ &= \lim_{n\to\infty} \left|\left|\int K(x,.)\, g_n(x)\, d\mu(x)\right|\right|^2 \\ &= \sum_{j=1}^{k} \left(\int e_j\, g\, d\mu\right)^2\end{aligned}$$

Autrement dit il s'agit de maximiser la norme de la projection de g sur l'espace E de noyau K : ceci a lieu chaque fois que $g \in E$ donc chaque fois que $K = 1 + \frac{g}{||g||} \otimes \frac{g}{||g||} + \sum_{j=2}^{k} f_j \otimes f_j$ où $(1, \frac{g}{||g||}, f_2, \ldots, f_k)$ forment un système orthonormal de $L^2(\mu)$.

On en déduit la :

Proposition 3.- Le test Hilbertien de niveau asymptotique α $(\in]0,1[)$ basé sur le noyau K est asymptotiquement le plus puissant (1) pour toute suite d'alternatives de la forme

---

(1) parmi les tests Hilbertiens dont le noyau est de rang $k+1$.

$$\nu_n = (1 + \frac{g_n}{\sqrt{n}})\ \mu \quad , \qquad n \geq 1 .$$

où (g_n) tend faiblement vers g élément non nul de l'espace de noyau K.

Remarque.- La proposition 3 s'étend au cas où

$$\nu_n = (1 + \frac{g_n}{\sqrt{n}})\ .\ \mu + m_n \quad , \quad m_n \perp \mu$$

où

a) $g_n \to g$ faiblement ,

b) $\int g_n \, d\mu \to 0$.

En effet, on a alors

$$\sqrt{n}\left[\int K(x,.)\ d\nu_n(x) - 1\right] = \int K(x,.)\ g_n(x)\ d\mu(x) + \sqrt{n}\ m_n(E)$$

or $$1 = \nu_n(E) = 1 + \frac{1}{\sqrt{n}} \int g_n \, d\mu + m_n(E)$$

donc b) implique $\sqrt{n}\ m_n(E) \to 0$.

Par ailleurs

$$\int g_n \, d\mu \to 0 = \int g \, d\mu$$

et l'on peut terminer la démonstration comme dans la proposition 3.

δ) Vitesse de convergence vers la loi limite.

L'évaluation de la vitesse de convergence de $||S_n||^2$ vers sa loi limite permet d'apprécier la valeur de la notion d'efficacité asymptotique.

Dans ce paragraphe, $||.||_p$ désigne la norme euclidienne de $\mathbb{R}^p$, $||.||_L$ la norme d'un opérateur linéaire, I_p la matrice unité $p \times p$ et enfin $<.,.>_p$ le produit scalaire de $\mathbb{R}^p$.

Lemme 2.- Soit $P = N(m,C)$ une loi normale dans $\mathbb{R}^k$ dont la matrice de covariance C est régulière et soit Q la loi normale $N(m',I_k)$. Alors

$$\sup_{B\in\mathcal{B}_{\mathbb{R}^k}} |P(B) - Q(B)| \leq c_1 \, ||m'-m||_k + c_2 \, ||C-I_k||_L$$

où les constantes c_1 et c_2 sont données par les formules suivantes :

$$c_1 = \frac{(2\pi)^{-k/2}}{4} (\det C)^{-1/2} \int (||C^{-1}||_L ||x-m||_k + ||x-m'||_k)(e^{-y'} + e^{-y})\, dx$$

et

$$c_2 = \frac{(2\pi)^{-k/2}}{2} (\det C)^{-1/2} \int ||C^{-1}||_L \; ||x-m||_k \; ||x-m'||_k \, (e^{-y'} + e^{-y})\, dx$$

$$+ \max\left(\sum_{h=0}^{k-1} ||C||_L^h \, , \, ||C^{-1}||_L \sum_{h=0}^{k-1} ||C^{-1}||_L^h\right) \times \frac{1}{2} (\det C)^{-1/2}$$

où

$$y = \frac{1}{2} < C^{-1}(x-m) \, , \; x - m >_k \quad \text{et} \quad y' = \frac{1}{2} ||x-m'||_k^2 \; .$$

Démonstration : Notons simplement $<.,.>$ le produit scalaire de $\mathbb{R}^k$ et posons

$$\delta = -\frac{1}{2}\left[<C^{-1}(x-m),\ x-m> - <x-m',\ x-m'>\right]$$

$$= -\frac{1}{2}\left[<C^{-1}(x-m),\ m'-m> + <C^{-1}(x-m),\ x-m'> - <x-m',\ x-m'>\right]$$

$$= -\frac{1}{2}\left[<C^{-1}(x-m),\ m'-m> + <(C^{-1}-I)\ x,\ x-m'> + <m' - C^{-1}\ m\ ,\ x-m'>\right]$$

$$= -\frac{1}{2}\left[<C^{-1}(x-m),\ m'-m> + <(C^{-1}-I)\ x,\ x-m'> + <m'-m,\ x-m'> + <(I-C^{-1})\ m,\ x-m'>\right]$$

Alors, comme

$$||C^{-1}-I||_L = ||C^{-1} - C^{-1}C||_L \leq ||C^{-1}||_L\ .\ ||I-C||_L$$

on a une majoration de la forme

$$|\delta| \leq A(x)\ ||m'-m||_k + B(x)\ .\ ||C-I||_L \tag{1}$$

Maintenant, étudions

$$\Delta = \sup_{B\in\mathcal{B}_{\mathbb{R}^k}} |P(B) - Q(B)| = \frac{1}{2}\,\frac{1}{(2\pi)^{k/2}}\int|(\det\ C)^{-1/2}\ e^{-y} - e^{-y'}|\ dx$$

On a

$$\Delta \leq \frac{(2\pi)^{-k/2}}{2}\int (\det\ C)^{-1/2}\ |e^{-y} - e^{-y'}|\ dx$$

$$+ \frac{(2\pi)^{-k/2}}{2}\ |1 - (\det\ C)^{-1/2}|\int e^{-y'}\ dx = \Delta_1 + \Delta_2\ .$$

Le premier terme se majore en remarquant que $y - y' = \delta$. Donc :

$$|\Delta_1| \leq \frac{(2\pi)^{-k/2}}{2} (\det C)^{-1/2} \int |\delta| \, e^{-y''} \, dx$$

où y'' est dans l'intervalle (y,y') (ou (y',y)), par conséquent

$$e^{-y''} \leq e^{-y'} + e^{-y}$$

on en déduit, compte tenu de (1)

$$\Delta_1 \leq \frac{(2\pi)^{-k/2}}{2} (\det C)^{-1/2} \, ||m'-m||_k \int A(x) \, (e^{-y'} + e^{-y}) \, dx$$

$$+ \frac{(2\pi)^{-k/2}}{2} (\det C)^{-1/2} \, ||I-C||_L \int B(x) \, (e^{-y'} + e^{-y}) \, dx$$

D'autre part

$$\Delta_2 = \frac{1}{2} \, |1 - (\det C)^{-1/2}| = \frac{1}{2(\det C)^{1/2}} \, |(\det C)^{1/2} - 1|$$

$$\leq \frac{1}{2(\det C)^{1/2}} \, |\det C - 1|$$

Nous allons distinguer deux cas :

1) Si $\det C \geq 1$, alors

$$|\det C - 1| = \det C - 1 \leq ||C||_L^k - 1$$

$$\leq (||C||_L - 1) \, (\sum_{h=0}^{k-1} ||C||_L^h)$$

$$\leq ||C-I||_L \, (\sum_{h=0}^{k-1} ||C||_L^h)$$

2) Si $\det C \leq 1$, alors

$$|\det C - 1| = 1 - \det C = 1 - (\det C^{-1})^{-1}$$

$$= \frac{(\det C^{-1}) - 1}{\det C^{-1}} \leq (\det C^{-1}) - 1$$

$$\leq (||C^{-1}||_L^k - 1)$$

$$\leq ||C^{-1} - I||_L \left(\sum_{h=0}^{k-1} ||C^{-1}||^h\right)$$

$$\leq ||C^{-1}||_L \; ||C - I||_L \left(\sum_{h=0}^{k-1} ||C^{-1}||^h\right)$$

En rassemblant les inégalités obtenus on en déduit le résultat annoncé. ∎

Proposition 4.- Les notations et les hypothèses étant celles de la proposition 1, on pose

$$\eta_n = \sup_{1\leq j\leq k} \left|\sqrt{n} \int e_j \, d\nu_n - v_j\right| \quad , \qquad n \geq 1$$

et

$$\varepsilon_n = \sup_{1\leq j\leq k} |C_{\nu_n}(i,j) - \delta_{ij}| \quad , \qquad n \geq 1 \; .$$

Dans ces conditions il existe des constantes c_1' et c_2' telles que

$$\sup_{C\in\mathcal{C}} |P(\sqrt{n}\,\hat{A}_n \in C) - P(U + v \in C)| \leq \gamma_n + c_1' \eta_n + c_2' \varepsilon_n \; .$$

Démonstration : Comme

$$|P(\sqrt{n}\,\hat{A}_n \in C) - P(U + v \in C)| \leq |P(\sqrt{n}\,\hat{A}_n \in C) - P(U_n + \sqrt{n}\,A_n \in C)|$$
$$+ |P(U_n + \sqrt{n}\,A_n \in C) - P(U + v \in C)|$$

il suffit de majorer le deuxième terme du majorant.

Pour cela on utilise le lemme 2 avec $m = \sqrt{n}\, A_n$, $C = C_{\nu_n}$ et $m' = v$.

Comme $U_n + \sqrt{n}\, A_n \longrightarrow U + v$ on a aussi la convergence des moments (cf. par exemple [7]). D'autre part, par continuité

$$\lim_{n\to\infty} ||C_{\nu_n}||_L = \lim_{n\to\infty} ||C_{\nu_n}^{-1}||_L = \lim_{n\to\infty} (\det C_{\nu_n}) = 1$$

donc c_1 et c_2 sont bornées et l'on conclut en remarquant que les normes des e.v. de dimension finie sont équivalentes. ∎

Corollaire 1.- Si $\nu_n = (1 + \frac{g_n}{\sqrt{n}})\,\mu$ où g_n tend fortement vers g dans $L^2(\mu)$, avec $||g_n - g|| = O(n^{-1/2})$ alors

$$\sup_{C\in\mathcal{C}} |P(\sqrt{n}\,\hat{A}_n \in C) - P(U + v \in C)| = O(n^{-1/2}) .$$

Démonstration : D'abord

$$|\sqrt{n} \int e_j\, d\nu_n - \int e_j\, g\, d\mu| = |\int e_j (g_n - g)\, d\mu| \leq ||g_n - g||$$

donc

$$\delta_n = O(n^{-1/2})$$

Ensuite, comme

$$|\int e_j\, d\nu_n| = |\frac{1}{\sqrt{n}} \int e_j\, g_n\, d\mu| \leq \frac{||g_n||}{\sqrt{n}}$$

il vient

$$|\int e_j\, e_{j'}\, d\nu_n - \int e_j\, d\nu_n \int e_{j'}\, d\nu_n - \delta_{jj'}| \leq |\int e_j\, e_{j'}\, d\nu_n - \delta_{jj'}| + \left(\frac{||g_n||}{\sqrt{n}}\right)^2$$

$$\leq |\frac{1}{\sqrt{n}} \int e_j\, e_{j'}\, g_n\, d\mu| + \frac{||g_n||^2}{n}$$

$$\leq \frac{||g_n||}{\sqrt{n}} + \frac{||g_n||^2}{n}$$

donc $\varepsilon_n = O(n^{-1/2})$ et comme il en est de même pour γ_n on a la vitesse de convergence annoncée. ■

II - COMPARAISON AVEC LE TEST DE NEYMAN-PEARSON.-

On se donne une suite de probabilités (ν_n) de la forme

$$\nu_n = (1 + g_n \frac{c_n}{\sqrt{n}}) \cdot \mu \qquad , \qquad n \geq 1$$

où $g_n \in L^2(\mu)$ et $c_n \in \mathbb{R}^+$, $n \geq 1$.

Posons

$$T_n = \sum_{j=1}^{n} \operatorname{Log}(1 + g_n(X_{jn}) \frac{c_n}{\sqrt{n}}) \qquad , \qquad n \geq 1$$

a) Lorsque les X_{jn} sont indépendantes et de loi μ, nous allons voir que T_n a une loi limite sous des hypothèses assez générales.

Pour cela on va utiliser la relation élémentaire

$$\operatorname{Log}(1 + u) = u - \frac{u^2}{2(1+\theta u)^2} \quad \text{où} \quad \theta = \theta(u) \in (-1, +1) .$$

Il vient

$$(1) \qquad c_n^{-1} T_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} g_n(X_{jn}) - \frac{c_n}{2n} \sum_{j=1}^{n} \frac{g_n^2(X_{jn})}{\left[1 + \theta_{jn} \frac{c_n}{\sqrt{n}} g_n(X_{jn})\right]^2}$$

si g_n est bornée et si $\frac{c_n}{\sqrt{n}} \sup |g_n(t)| \to 0$, on a

$$(2) \qquad \frac{1}{(1 + \frac{c_n}{\sqrt{n}} \sup |g_n|)^2} \leq \frac{1}{\left[1 + \theta_{jn} \frac{c_n}{\sqrt{n}} g_n(X_{jn})\right]^2} \leq \frac{1}{(1 - \frac{c_n}{\sqrt{n}} \sup |g_n|)^2}$$

au moins pour n assez grand.

Alors, si $\frac{c_n^2}{n}\left[\int g_n^4\,d\mu - \left(\int g_n^2\,d\mu\right)^2\right] \longrightarrow 0$, (3)

on en déduit que

$$-\frac{c_n}{2n}\sum_{j=1}^{n}\frac{g_n^2(X_{jn})}{\left[1+\theta_{jn}\frac{c_n}{n}g_n(X_{jn})\right]^2}+\frac{c_n}{2}\int g_n^2\,d\mu \xrightarrow[\text{m.q.}]{} 0 \quad .$$

D'autre part, comme $g_n(X_{jn})$ est centrée

$$E\left(\exp\frac{it}{\sqrt{n}}g_n(X_{jn})\right) = 1-\frac{t^2}{2n}Vg_n(X_{jn})$$

$$+ m_3\frac{(it)^3}{3!\,n^{3/2}} + \theta\, m_4\frac{(it)^4}{4!\,n^2} \qquad (4)$$

$|\theta| \leq 1$; où $m_k = E\left[g_n^k(X_{jn})\right]$.

Si $\frac{m_3}{\sqrt{n}}$ et $\frac{m_4}{n} \longrightarrow 0$ il vient

$$E\left(\exp i\frac{t}{\sqrt{n}}\sum_{j=1}^{n}g_n(X_{jn})\right) = \left[1-\frac{t^2}{2n}\int g_n^2\,d\mu + o\left(\frac{1}{n}\right)\right]^n$$

Si $\int g_n^2\,d\mu \longrightarrow \gamma^2$ on en déduit que

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}g_n(X_{jn}) \xrightarrow{\mathcal{L}} N(0,\gamma^2) \quad .$$

Sous toutes les hypothèses précédentes on en déduit que

$$c_n^{-1}\,T_n + \frac{c_n}{2}\int g_n^2\,d\mu \xrightarrow{\mathcal{L}} N(0,\gamma^2)$$

(cf. [8] p. 38).

D'où le

Lemme 3.- Si $\nu_n = (1 + g_n \frac{c_n}{\sqrt{n}})\,\mu$, $n \geq 1$

où $||g_n||^2 \longrightarrow \gamma^2$, $\frac{c_n}{\sqrt{n}} \sup |g_n| \longrightarrow 0$,

$$\frac{c_n^2}{n}\left(\int g_n^4\, d\mu - ||g_n||^2\right) \longrightarrow 0\ ,\ \frac{\int g_n^3}{\sqrt{n}} \longrightarrow 0\ ,\ \frac{\int g_n^4}{n} \longrightarrow 0$$

et si les X_{jn} sont indépendantes <u>et de loi μ</u>.

$$c_n^{-1} \sum_{j=1}^{n} \mathrm{Log}\left(1 + g_n(X_{jn}) \frac{c_n}{\sqrt{n}}\right) + \frac{c_n}{2} ||g_n||^2 \xrightarrow{L} N(0,\gamma^2) \quad .$$

b) Nous allons maintenant étudier le comportement asymptotique de T_n lorsque $L(X_{1n},\ldots,X_{nn}) = \nu_n^{\otimes n}$, $n \geq 1$.

Tout d'abord, si X est de loi ν_n ,

$$E\left[c_n\, g_n^2(X)\right] = c_n \int g_n^2\, d\nu_n = c_n \int g_n^2\, d\mu + \frac{c_n^2}{\sqrt{n}} \int g_n^3\, d\mu \qquad \text{et}$$

$$E\left[c_n^2\, g_n^4(X)\right] = c_n^2 \int g_n^4\, d\mu + \frac{c_n^3}{\sqrt{n}} \int g_n^5\, d\mu$$

Si l'on suppose toujours que $\frac{c_n}{\sqrt{n}} \sup |g_n| \longrightarrow 0$ on aura donc convergence en moyenne quadratique vers 0 de la quantité

$$V_n = -\frac{c_n}{2n} \sum_{j=1}^{n} \frac{g_n^2(X_{jn})}{\left[1 + \theta_{jn} \frac{c_n}{\sqrt{n}} g_n(X_{jn})\right]^2} + \frac{c_n}{2} \int g_n^2\, d\mu + \frac{c_n^2}{\sqrt{n}} \int g_n^3\, d\mu$$

si et seulement si

$$j_n = \frac{1}{n}\left[c_n^2 \int g_n^4\, d\mu + \frac{c_n^3}{\sqrt{n}} \int g_n^5\, d\mu - c_n^2 \left(\int g_n^2\, d\mu\right)^2 - \frac{c_n^4}{n} \int g_n^3\, d\mu \right.$$
$$\left. - 2\,\frac{c_n^3}{\sqrt{n}} \int g_n^2\, d\mu \,.\, \int g_n^3\, d\mu\right]$$

tend vers 0 .

Posons $G_n = \max(1, \sup |g_n|)$ comme

$$|j_n| \leq 2\left(\frac{c_n\, G_n^2}{\sqrt{n}}\right)^2 + 3\left(\frac{c_n\, G_n^{5/3}}{\sqrt{n}}\right)^3 + \left(\frac{c_n}{\sqrt{n}}\, G_n^{3/4}\right)^4$$

si $\dfrac{c_n\, G_n^2}{\sqrt{n}} \longrightarrow 0$ on a $j_n \longrightarrow 0$ (et aussi $\dfrac{c_n}{\sqrt{n}} \sup |g_n| \longrightarrow 0$) .

D'autre part, comme

$$E(g_n(X_{jn})) = \int g_n \left(1 + \frac{c_n}{\sqrt{n}}\, g_n\right) d\mu = \int \frac{c_n}{\sqrt{n}}\, g_n^2\, d\mu$$

on va considérer

$$U_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[g_n(X_{jn}) - \frac{c_n}{\sqrt{n}}\, ||g_n||^2\right]$$

soit φ_n la f.c. de $[..]$:

$$\varphi_n\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n}\left[V g_n(X_{jn})\right] + \frac{m_3}{\sqrt{n}}\, \frac{(it)^3}{3!n} + \theta\, \frac{m_4}{n}\, \frac{(it)^4}{4!n} \quad , \quad |\theta| \leq 1$$

ce qui précède montre que

$$V g_n(X) = E\, g_n^2(X) - (E\, g_n(X))^2 = \int g_n^2\, d\mu + \frac{c_n}{\sqrt{n}} \int g_n^3\, d\mu - \frac{c_n^2}{n} \left(\int g_n^2\, d\mu\right)^2$$

comme $\dfrac{c_n^2}{n} \left(\int g_n^2\, d\mu\right)^2 \leq \left(\dfrac{c_n}{\sqrt{n}}\, G_n^2\right)^2$

la nouvelle condition $\frac{c_n}{\sqrt{n}} G_n^3 \longrightarrow 0$ suffit à assurer que

$$Vg_n(X) = ||g_n||^2 + o(1) \quad .$$

Pour avoir une loi limite il suffit donc que les conditions $||g_n||^2 \to \gamma^2$, $\frac{m_3}{\sqrt{n}} \longrightarrow 0$, $\frac{m_4}{n} \longrightarrow 0$ soient vérifiées.

Or $\qquad E(Y - EY)^3 = EY^3 - 3EY^2 \,.\, EY + 3(EY)^3 - (EY)^3$

et $\qquad |E(g_n^\ell(X))| = |\int g_n^\ell \, d\mu + \frac{c_n}{\sqrt{n}} \int g_n^{\ell+1} \, d\mu|$

$$\leq G_n^\ell \left(1 + \frac{c_n}{\sqrt{n}} G_n\right) \leq 2\, G_n^\ell \quad (n \text{ assez grand}) \ , \quad \ell \geq 1 \ .$$

Par conséquent

$$m_3 = O(G_n^3)$$

de la même façon on peut voir que

$$m_4 = O(G_n^4)$$

d'où les conditions $\frac{G_n^3}{\sqrt{n}} \longrightarrow 0$ et $\frac{G_n^4}{n} \longrightarrow 0$ ce qui se ramène simplement à $\frac{G_n^3}{\sqrt{n}} \longrightarrow 0$ (cf. $\left(\frac{G_n^2}{\sqrt{n}}\right)^2 = \frac{G_n^4}{n}$) posons $C_n = \sup(1, c_n)$

alors la condition

$$\boxed{\frac{C_n \, G_n^3}{\sqrt{n}} \longrightarrow 0}$$

implique à la fois $\frac{c_n \, G_n^3}{\sqrt{n}} \longrightarrow 0$ et $\frac{G_n^3}{\sqrt{n}} \longrightarrow 0$, ainsi que $\frac{c_n \, G_n^2}{\sqrt{n}} \longrightarrow 0$.

D'où le :

Lemme 4.- Si $\nu_n = (1 + g_n \frac{c_n}{\sqrt{n}})\,\mu$, $n \geq 1$

où $||g_n||^2 \to \gamma^2$ et $\frac{C_n^2\, G_n^3}{\sqrt{n}} \longrightarrow 0$ $(C_n = \sup(1, c_n)$

$G_n = \sup(1, \sup |g_n|))$

et si $L(X_{1n},\dots,X_{nn}) = \nu_n^{\otimes n}$, $n \geq 1$, alors

$$c_n^{-1} \sum_{j=1}^{n} \text{Log}(1 + g_n(X_{jn}) \frac{c_n}{\sqrt{n}}) - \frac{c_n}{2} ||g_n||^2 \xrightarrow{L} N(0, \gamma^2) .$$

En effet, ce qui précède montre que $V_n \xrightarrow[m.q.]{} 0$ et aussi que $V_n - \frac{c_n^2}{\sqrt{n}} \int g_n^3\, d\mu \xrightarrow[m.q.]{} 0$ car $\frac{C_n^2\, G_n^3}{\sqrt{n}} \longrightarrow 0$ et d'autre part

que $U_n \xrightarrow{L} N(0,\gamma^2)$ donc $U_n + V_n - \frac{c_n^2}{\sqrt{n}} \int g_n^3\, d\mu \xrightarrow{L} N(0,\gamma^2)$ C.Q.F.D

c) On a alors l'énoncé suivant qui résume les deux lemmes précédents :

Proposition 5.- Si $\nu_n = (1 + \frac{c_n}{\sqrt{n}} \cdot g_n) \cdot \mu$, $n \geq 1$

où
$$||g_n||^2 \longrightarrow \gamma^2$$

et
$$\frac{C_n^2\, G_n^3}{\sqrt{n}} \longrightarrow 0$$

alors 1°) Si $L(X_{1n},\dots,X_{nn}) = \mu^{\otimes n}$, $n \geq 1$

$$c_n^{-1} \sum_{j=1}^{n} \text{Log}(1 + \frac{c_n}{\sqrt{n}} g_n(X_{jn})) + \frac{c_n}{2} ||g_n||^2 \xrightarrow{L} N(0,\gamma^2)$$

2°) Si $L(X_{1n},\dots,X_{nn}) = \nu_n^{\otimes n}$, $n \geq 1$

$$c_n^{-1} \sum_{j=1}^{n} \text{Log}(1 + \frac{c_n}{\sqrt{n}} g_n(X_{jn})) - \frac{c_n}{2} ||g_n||^2 \xrightarrow{L} N(0,\gamma^2) .$$

. Dans le cas particulier où $c_n \equiv 1$, $g_n \equiv g$ on trouve sans aucune condition, dans le 1er cas :

$$\sum_{j=1}^{n} \text{Log}\left(1 + \frac{g(X_{jn})}{\sqrt{n}}\right) \xrightarrow{\mathcal{L}} N\left(-\frac{||g||^2}{2}, ||g||^2\right)$$

et dans le 2ème cas :

$$\sum_{j=1}^{n} \text{Log}\left(1 + \frac{g(X_{jn})}{\sqrt{n}}\right) \xrightarrow{\mathcal{L}} N\left(\frac{||g||^2}{2}, ||g||^2\right)$$

ce qui pourrait s'établir directement à partir de théorèmes de contigüité (cf. [9] p. 12).

Remarque.- Si $\nu_n = \left(1 + \frac{c_n\, g_n}{\sqrt{n}}\right) . \mu + m_n$, $m_n \perp \mu$ les résultats subsistent moyennant la condition supplémentaire $\sum_n c_n \sqrt{n} \int g_n\, d\mu < +\infty$.

En effet, le lemme de Borel Cantelli montre que, si $\mu(E_o) = 1$; $m_n(E_o) = 0$, $n \geq 1$, on a alors

$$P\left(\liminf \bigcap_{j=1}^{n} \{X_{jn} \in E_o\}\right) = 1 \quad .$$

d) *Application à la puissance asymptotique du test de Neyman-Pearson sous des hypothèses adjacentes.*

Le test de Neyman Pearson est de la forme

$$T_n = \sum_{j=1}^{n} \text{Log}\left(1 + \frac{c_n}{\sqrt{n}}\, g(X_{jn})\right) > \text{Log}\ w_n$$

ou encore

$$\frac{1}{\gamma}\left[c_n^{-1}\, T_n + \frac{c_n}{2}\, ||g_n||^2\right] > \frac{1}{\gamma}\left[c_n^{-1}\, \text{Log}\ w_n + \frac{c_n}{2}\, ||g_n||^2\right]$$

il est de niveau asymptotique $\alpha \in]0,1[$ si

$$z_n = \frac{1}{\gamma}\left[c_n^{-1} \text{ Log } w_n + \frac{c_n}{2} ||g_n||^2\right] \longrightarrow N_\alpha \tag{1}$$

où $P(N > N_\alpha) = \alpha$ $(N \sim N(0,1))$.

Si $c_n \to c$, il vient

$$w_n \longrightarrow \exp\left[N_\alpha \gamma_c - \frac{1}{2} \gamma^2 c^2\right]$$

Si (1) est vérifié on peut chercher la puissance asymptotique du test. On sait que, sous la suite d'alternatives,

$$\frac{1}{\gamma}\left[c_n^{-1} T_n - \frac{c_n}{2} ||g_n||^2\right] \xrightarrow{L} N(0,1)$$

la puissance asymptotique est alors la limite de

$$P\left[N > \frac{1}{\gamma}\left[c_n^{-1} \text{ Log } w_n - \frac{c_n}{2} ||g_n||^2\right]\right]$$

soit

$$P\left[N > g_n - \frac{c_n}{\gamma} ||g_n||^2\right]$$

la limite appartient à $]0,1[$ si et seulement si $c_n \to c$, elle vaut alors

$$P(N > N_\alpha - c\gamma)$$

Si $c_n \to +\infty$ la puissance asymptotique est égale à 1.

e) Optimalité asymptotique.

Si, lorsque $L(X_{1n},\ldots,X_{nn}) = \mu^{\otimes n}$

$$P\left(\prod_{j=1}^{n}\left(1+\frac{c_n}{\sqrt{n}}\, g_n(X_{jn})\right) = \lambda\right) = 0 \quad , \quad n \geq 1 \ , \quad \lambda \in \mathbb{R}$$

il existe, pour toute suite (α_n) , $\alpha_n \in \left]0,1\right[$ une suite (S_n) de tests de Neyman-Pearson non randomisés de niveau α_n.

Soit alors (T_n) une suite de tests tels que le niveau de T_n soit α_n , $n \geq 1$. On suppose que $\alpha_n \to \alpha \in \left]0,1\right[$. Soit p_n la puissance de S_n et q_n celle de T_n , on a

$$p_n \geq q_n \quad , \quad n \geq 1$$

d'où

$$\begin{aligned} \limsup q_n \leq \lim p_n &= P(N > N_\alpha - c\gamma) \quad \text{si} \quad c_n \to c \ , \\ &= 1 \qquad\qquad\qquad\quad\ \text{si} \quad c_n \to +\infty \ . \end{aligned}$$

Lorsque $c_n \to c$ (T_n) est <u>asymptotiquement optimale</u> si $q_n \to P(N > N_\alpha - c\gamma)$.

f) Efficacité asymptotique des tests Hilbertiens.

On a vu que le test Hilbertien de noyau fixé et de niveau asymptotique $\alpha \in \left]0,1\right[$ a une puissance asymptotique qui vaut au plus $P\left[\chi^2(k,\gamma^2) > \chi^2_\alpha\right]$ (ceci pour $c_n \equiv 1$ et $\|g_n\|^2 \to \gamma^2$).

L'efficacité asymptotique d'un tel test vaut donc

$$e = \frac{P\left[\chi^2(k,\gamma^2) > \chi^2_\alpha\right]}{P\left[N > N_\alpha - \gamma\right]}$$

<u>Cas particulier</u> : pour $k = 1$, comme $\chi_\alpha = N_{\alpha/2}$ il vient facilement

$$e = \frac{P(|N-\gamma| > N_{\alpha/2})}{P(N > N_\alpha - \gamma)}$$

III - CAS D'UN NOYAU VARIABLE DE DIMENSION FINIE.-

On se donne maintenant une suite $(K_n, n \geq 1)$ de noyaux définis positifs de rangs respectifs k_n, i.e.

$$K_n(x,y) = \sum_{j=0}^{k_n} e_{jn}(x)\, e_{jn}(y) \quad ; \quad n \geq 1 \; ; \; x \, , \, y \in E$$

où $e_{on} \equiv 1$, $n \geq 1$.

Pour utiliser le théorème de SAZONOV nous aurons besoin de minorer les valeurs propres d'une matrice de covariance. C'est l'objet du

Lemme 5.- Soit C une matrice régulière $k \times k$ dont les valeurs propres sont $\lambda_1 \geq \dots \geq \lambda_k$. Alors

$$\lambda_k \geq 1 - \sqrt{\sum_{j,j'} [C(j,j') - \delta_{jj'}]^2} \qquad (1)$$

Démonstration : On vérifie directement que les valeurs propres de $C - I_k$ sont $\lambda_1 - 1 \geq \lambda_2 - 1 \geq \dots \geq \lambda_k - 1$. Donc

$$||C-I_k||_L = \sup_j |\lambda_j - 1|$$

mais $||C-I_k||_L^2 \leq \sum_{j,j'} [C(j,j') - \delta_{jj'}]^2 = ||C-I_k||_{k\times k}^2$,

on en déduit $\quad 1 - \lambda_k \leq \sup_j |\lambda_j - 1| \leq ||C-I_k||_{k\times k} \quad$ d'où (1). ∎

Maintenant, avec des notations analogues à celles de la proposition 1, on a le :

Lemme 6.- Si

1°) $\frac{k_n^{11}}{n} \longrightarrow 0$

2°) $M = \sup_{x\in E, 1\leq j\leq k_n, n\geq 1} |e_{jn}(x)| < +\infty$

3°) $\limsup \left[\| C_{\nu_n} - I_{k_n} \|_{k_n\times k_n} \right] < 1$

Alors, avec des notations claires

$$\sup_{C\in\mathcal{C}_{k_n}} |P(\sqrt{n}\,\hat{A}_n \in C) - P(U_n + \sqrt{n}\,A_n \in C)| \longrightarrow 0 \tag{2}$$

Démonstration : D'après l'inégalité de SAZONOV et 2°) il suffit de montrer que

$$\gamma_n = \frac{8\, c_o\, M^3}{\sqrt{n}}\, k_n^{11/2}\, \eta_{k_n n}^{-3/2} \longrightarrow 0 \qquad (1)$$

Le lemme 5 implique

$$\limsup \eta_{k_n n}^{-3/2} \leq \limsup \left[1 - \| C_{\nu_n} - I_{k_n} \|_{k_n\times k_n} \right]^{-3/2}$$

donc d'après 3°)

$$\limsup \eta_{k_n n}^{-3/2} < +\infty$$

et d'après 1°) $\qquad \gamma_n \longrightarrow 0$

d'où (2). ∎

---

(1) $\eta_{k_n, n}$ désigne la plus petite valeur propre de C_{ν_n} .

Proposition 6.- Sous les conditions suivantes :

a) $k_n \to +\infty$ et $\dfrac{k_n^{11}}{n} \longrightarrow 0$

b) $n \sum_{j=1}^{k_n} \left[\int e_{jn}(x)\, d\nu_n(x)\right]^2 \longrightarrow \ell$

c) $\limsup \| C_{\nu_n} - I_{k_n} \|_{k_n \times k_n} < 1$

d) $M = \sup_{x\in E,\, 1\le j\le k_n,\, n\ge 1} |e_{jn}(x)| < +\infty$

on a :

$$\frac{\|S_n\|^2 - k_n}{\sqrt{2k_n}} \xrightarrow{\mathcal{L}} N(0,1)$$

Démonstration : Les conditions du lemme 2 étant vérifiées il suffit de montrer que :

$$Z_n = \frac{\|U_n + \sqrt{n}\, A_n\|^2 - k_n}{\sqrt{2k_n}} \xrightarrow{\mathcal{L}} N(0,1)$$

or $$\|U_n + \sqrt{n}\, A_n\|^2 = \sum_{j=1}^{k_n} \left(U_{jn} + \sqrt{n}\int e_{jn}\, d\nu_n\right)^2$$

$$= \sum_{j=1}^{k_n} U_{jn}^2 + n \sum_{j=1}^{k_n} \left(\int e_{jn}\, d\nu_n\right)^2$$

$$+ 2 \sum_{j=1}^{k_n} \sqrt{n}\, U_{jn} \int e_{jn}\, d\nu_n$$

donc $Z_n = V_n + W_n$ où

$$V_n = \frac{\sum_{1}^{k_n} (U_{jn}^2 - 1)}{\sqrt{2k_n}} \quad \text{et} \quad W_n = \frac{n \sum_{j=1}^{k_n} (\int e_{jn}\, d\nu_n)^2 + 2\sqrt{n} \sum_{j=1}^{n} U_{jn} \int e_{jn}\, d\nu_n}{\sqrt{2k_n}}$$

Alors $$V_n \xrightarrow{\mathcal{L}} N(0,1)$$

et

$$E\left[\exp(it\sqrt{n} \sum_{j=1}^{n} (\int e_{jn}\, d\nu_n)\, U_{jn}\right] = \exp\left[-2n \sum_{j=1}^{k_n} a_{jn}^2(\nu_n)\, t^2\right]$$

$$\longrightarrow \exp\left[-2\ell\, t^2\right]$$

donc le numérateur de W_n converge en loi et comme $k_n \to +\infty$ $W_n \xrightarrow{\mathcal{L}} 0$ et le résultat s'ensuit. ∎

Remarque : La loi limite est la même que sous l'hypothèse nulle. Les conditions de la proposition suivante permettent d'obtenir une loi limite différente :

Proposition 7.- Si, dans les hypothèses de la proposition 6, on remplace b) par

$$\text{b}')\quad \frac{n}{\sqrt{k_n}} \sum_{j=1}^{k_n} (\int e_{jn}\, d\nu_n)^2 \longrightarrow \ell'$$

il vient :

$$\frac{||S_n||^- - k_n}{\sqrt{2k_n}} \xrightarrow{\mathcal{L}} N(\frac{\ell'}{\sqrt{2}}, 1) \quad .$$

Démonstration : On cherche encore la loi limite de Z_n (lemme 6) mais, d'après une propriété du χ^2 décentré,

$$\sum_1^{k_n} \left[U_{jn} + \sqrt{n}\, a_{jn}(\nu_n)\right]^2 \quad \text{et} \quad \sum_1^{k_n} U_{jn}^2 + n \sum_1^{k_n} a_{jn}^2(\nu_n) \quad \text{ont la même loi .}$$

On est donc, finalement, ramené à la recherche de la loi limite de

$$Z'_n = \frac{\sum_1^{k_n} U_{jn}^2 + n \sum_1^{k_n} a_{jn}^2(\nu_n) - k_n}{\sqrt{2k_n}}$$

et la condition $\frac{n}{\sqrt{2k_n}} \sum_1^{k_n} a_{jn}^2(\nu_n) \longrightarrow \frac{\ell'}{\sqrt{2}}$ permet de conclure . ∎

Exemples :

1) Si $\nu_n = (1 + \frac{g_n}{\sqrt{n}}) \cdot \mu$ la condition b) de la proposition 6 s'écrit :

$$n \sum_1^{k_n} \left(\int e_{jn}\, d\nu_n\right)^2 = ||Pr_{K_n}\, g_n||^2 \longrightarrow \ell$$

D'autre part

$$\sum_{j,j'} \left[\int e_{jn}\, e_{j'n}\, d\nu_n - \int e_{jn}\, d\nu_n \int e_{j'n}\, d\nu_n - \delta_{jj'}\right]^2$$

$$= \sum_{j,j'} \left[\frac{1}{\sqrt{n}} \int e_{jn}\, e_{j'n}\, d\mu - \frac{1}{n} \int e_{jn}\, g_n\, d\mu \int e_{j'n}\, g_n\, d\mu\right]^2$$

$$\leq \frac{2}{n} \sum_{j,j'} \left(\int e_{jn}\, e_{j'n}\, g_n\, d\mu\right)^2 + \frac{2}{n^2} \sum_{j,j'} \left(\int e_{jn}\, g_n\, d\mu \int e_{j'n}\, g_n\, d\mu\right)^2$$

$$\leq \frac{2}{n} ||g_n||^2 \sum_{j,j'} ||e_{jn}\, e_{j'n}||^2 + \frac{2}{n^2} ||Pr_{K_n}\, g_n||^4$$

$$\leq \frac{2}{n} ||g_n||^2 M^2 k_n^2 + o(1)$$

donc la condition $\limsup 2\, M^2 \dfrac{\|g_n\|^2\, k_n^2}{n} < 1$ est suffisante pour que c) soit vérifiée. En particulier il suffit que $\limsup \|g_n\|^2 < +\infty$.

2) Si $\nu_n = (1 + g_n \dfrac{k_n^{1/4}}{\sqrt{n}})\, \mu$.

La condition b') s'écrit

$$\frac{n}{\sqrt{k_n}} \sum_1^{k_n} \left(\int e_{jn}\, d\nu_n\right)^2 = \|Pr_{K_n}\, g_n\|^2 \longrightarrow \ell' .$$

Un calcul analogue au précédent donne

$$\sum_{j,j'} \left[\int e_{jn}\, e_{j'n}\, d\nu_n - \int e_{jn}\, d\nu_n \int e_{j'n}\, d\nu_n\right]^2 \leq 2\, M^2 \frac{k_n^{5/2}}{n} \|g_n\|^2 +$$

$$2\, \frac{k_n}{n^2} \|Pr_{K_n}\, g_n\|^4$$

et la condition c) est vérifiée dès que

$$\limsup 2\, M^2 \frac{k_n^{5/2}}{n} \|g_n\|^2 < 1$$

donc, en particulier, si $\|g_n\| = o\left(\dfrac{\sqrt{n}}{k_n^{5/4}}\right)$.

Remarques :

1) $(e_{jn}) \equiv (e_j)$ est total et si g_n tend fortement vers g on a $\ell' = \|g\|^2$.

En effet

$$\sum_{j=1}^{k_n} \left(\int g_n\, e_j - \int g\, e_j\right)^2 + \sum_{k_n+1}^{\infty} \left(\int g\, e_j\right)^2 \longrightarrow 0$$

donc $Pr_{K_n}\, g_n \to g$.

2) On a des résultats analogues si ν_n admet une partie étrangère à μ.

3) Nadaraja dans [10] a obtenu la loi limite de la proposition 6 sous des conditions plus fortes que les nôtres.

- *Puissance asymptotique* : Soit $\alpha \in]0,1[$.

si $w_n = N_\alpha \sqrt{2k_n} + k_n$ avec $P(N > N_\alpha) = \alpha$ ($N \sim N(0,1)$)

alors

$$P_\mu(||S_n||^2 > w_n) = P_\mu\left(\frac{||S_n||^2 - k_n}{\sqrt{2k_n}} > N_\alpha\right) \longrightarrow P_\mu(N > N_\alpha) = \alpha$$

le test est donc de niveau asymptotique α et, sous les hypothèses de la proposition 7, sa puissance asymptotique s'écrit

$$P(N + \frac{\ell'}{\sqrt{2}} > N_\alpha) \quad .$$

- *Optimisation* : Maximiser la puissance asymptotique revient donc à maximiser ℓ'. Quand $\frac{d\nu_n}{d\mu} = 1 + g_n \frac{k_n^{1/4}}{\sqrt{n}}$ cela revient à maximiser $\lim_{n\to\infty} ||Pr_{K_n} g_n||^2$ (lorsque cette limite existe). Quand $||g_n||^2$ converge on est amené à choisir des espaces E_n tels que $E_n \ni g_n$, ces espaces maximisent $||Pr_{E_n} g_n||$ donc aussi sa limite.

Quand g_n tend fortement vers g, $Pr_{K_n} g_n$ aussi pour n'importe quel choix de K_n de la forme $\sum_1^{k_n} e_j \otimes e_j$ où (e_j) est total dans $L^2(\mu)$. Les tests de ce type sont donc tous asymptotiquement optimaux. Cependant, à distance finie, il paraît préférable de choisir des espaces $E_n \ni g_n$.

- *Efficacité asymptotique* : En appliquant la proposition 5 avec $c_n = k_n^{1/4}$ on voit que, si $||g_n||^2 \to \gamma^2$, la condition

$\frac{k_n^{1/2}}{\sqrt{n}} \sup(1, \sup |g_n|)^3 \to 0$ assure une puissance asymptotique égale à 1 pour le test de Neyman-Pearson de niveau asymptotique α.

L'efficacité asymptotique d'un test Hilbertien (de noyau variable et de dimension finie) optimal sera donc égale à sa puissance asymptotique :

$$e = P(N + \frac{||g||^2}{\sqrt{2}} > N_\alpha) .$$

Suivant la valeur de $||g||$ on peut obtenir n'importe quelle valeur de l'intervalle $[\alpha, 1[$.

Remarques :

- *Formes intrinsèques des hypothèses sur* (ν_n).

La condition b') peut s'écrire sous la forme intrinsèque

$$\boxed{A_1''} \qquad n \, ||K_n'||^{-1}_{L^2(\mu \otimes \mu)} \, || \int K_n'(x,.) \, d\nu_n(x)||^2_{L^2(\mu)} \longrightarrow 0$$

où $K_n' = K_n - 1$,

ce qui apparaît comme une généralisation de la condition A_1 .

Au lieu d'écrire la condition c) sous une forme intrinsèque on peut chercher directement une condition sur les valeurs propres de l'opérateur de covariance de la v.a.

$$Z_n = K_n'(Y_n,.) - \int K_n'(y,.) \, d\nu_n(y)$$

où $Y_n \sim \nu_n$, $n \geq 1$.

Cette v.a. est centrée et à valeurs dans l'e.v. E_n de noyau K_n'.
Soit C_{Z_n} son opérateur de covariance. Alors

$$\langle C_{Z_n} f,g\rangle = E\left[\langle Z_n,f\rangle \langle Z_n,g\rangle\right] \quad ; \quad f,\ g \in E_n$$

$$= \int fg \, d\nu_n - \int f \, d\nu_n \int g \, d\nu_n$$

Soit alors f une fonction propre de C_{Z_n} associée à la valeur propre

$$\lambda = \langle C_{Z_n} f,\ f\rangle = \int f^2 \, d\nu_n - \left(\int f \, d\nu_n\right)^2 = \sigma^2 f(Y_n)$$

Alors si $f_{1n}, \ldots, f_{k_n n}$ désignent les fonctions propres de C_{Z_n} la condition

$$\boxed{A_2''} \qquad \liminf \left[\inf_{1 \leq j \leq k_n} \sigma^2 f_{jn}\right] > 0$$

peut être utilisée à la place de la condition c). Cette condition est beaucoup moins maniable mais plus générale.

– Cas où $\nu_n = \left(1 + g_n \dfrac{k_n^{1/4}}{\sqrt{n}}\right) \mu$, $n \geq 1$.

Alors on a vu que la condition b') pouvait s'écrire

$$||Pr_{K_n} g_n||^2 \longrightarrow \ell'$$

En ce qui concerne la condition A_2'' on peut l'obtenir de la façon suivante :

Soit f de norme 1 appartenant à E_n , alors

$$\sigma^2 f(Y_n) = \int f^2 \left(1 + g_n \frac{k_n^{1/4}}{\sqrt{n}}\right) d\mu - \left[\int f\left(1 + g_n \frac{k_n^{1/4}}{\sqrt{n}}\right) d\mu\right]^2$$

$$= 1 + \frac{k_n^{1/4}}{\sqrt{n}} \int f^2 g_n \, d\mu - \frac{k_n^{1/2}}{n} \left(\int f \, g_n \, d\mu\right)^2 .$$

$$\Longrightarrow 1 - \sigma^2 f(Y_n) \leq \frac{k_n^{1/2}}{n} G_n^2 + \frac{k_n^{1/4}}{\sqrt{n}} G_n$$

d'où, par exemple, la condition

$$\limsup \frac{k_n^{1/4}}{\sqrt{n}} G_n < \frac{1}{2}$$

qui assure A_2'' .

APPENDICE

NOYAUX REPRODUISANTS

Soit H un espace de Hilbert dont les éléments sont des fonctions réelles définies sur un ensemble E donné et dont le produit scalaire est noté $<.,.>$.

Soit K une fonction réelle définie sur $E \times E$, on dit que K est un noyau reproduisant pour H si

a) $\forall t \in E, \quad K(.,t) \in H$

b) $\forall f \in H, \quad t \in E, \qquad <f,K(.,t)> = f(t)$ (propriété de reproduction).

Si K existe il est unique et c'est une fonction symétrique et semi-définie positive. Inversement toute fonction vérifiant ces deux dernières propriétés est le noyau reproduisant d'un espace de Hilbert.

Si H est de dimension finie d il admet un noyau reproduisant K donné par la formule

$$K(x,y) = \sum_{j=1}^{d} e_j(x)\, e_j(y) \qquad ; \qquad x,y \in E$$

où $e_1,\dots,e_d$ désigne une base orthonormale quelconque de H.

Pour des détails sur les espaces à noyau reproduisant, on pourra consulter [7].

B I B L I O G R A P H I E.

[1] BOSQ D. - *Tests d'ajustement Hilbertiens.* Pub. Int. UER Math. Lille 1, 39 p. n° 125 (1978).

[2] BOSQ D. - *Tests Hilbertiens et test du* χ^2. C.R.A.S., A, t. 286 p. 946-48 (1978).

[3] BOSQ D. - *Sur une classe de tests qui contient le test du* χ^2. Pub. Inst. Stat. Univ. Paris XXV fasc 1-2, p. 1-16 (1980).

[4] GREGORY G. - *Large sample theory for U-statistics and tests of fit.* Annals of Stat., Vol. 5 n° 1 p. 110-123 (1977).

[5] GREGORY G. - *On efficiency and optimality of quadratic tests.* Annals of Stat., Vol. 8 n° 1 p. 116-131 (1980).

[6] BARRA J. - *Notion fondamentales de Statistique Mathématique.* Dunod (1971).

[7] NEVEU J. - *Processus Gaussiens.* Montréal.

[8] FOURGEAUD C., FUCHS A. - *Statistique.* Dunod (1967).

[9] RAOULT J.P. - *Présentation de la contiguïté entre deux suites de probabilités.* Séminaire de Statistique de Paris VII - Orsay (1975).

[10] NADARAJA E.A. - *A quadratic measure of the deviation of a density estimator.* Theory of Proba. and Applic. 21 p. 843-50 (1976).