Distances euclidiennes sur les mesures signées et application à des théorèmes de Berry-Esséen

Ch. Suquet

Abstract

Starting from an integral representation of reproducing kernels, we define an inner product on bounded signed measures. The space of measures is embedded in a reproducing kernel Hilbert space and in a L^2 space. With a suitable choice of the kernel, the Euclidean associated distance on signed measures metrizes the weak convergence of probability measures. Using this framework, we obtain some rates of convergence in the CLT under independence (univariate and multivariate cases) or positive dependence (univariate case).

Résumé

A partir d'une représentation intégrale de noyaux reproduisants, on définit un produit scalaire sur l'espace des mesures signées. On injecte isométriquement cet espace de mesures dans un espace autoreproduisant et dans un espace L^2. Pour un choix convenable du noyau, la distance euclidienne associée métrise la convergence étroite des mesures de probabilité. Le cadre fonctionnel ainsi mis en place nous permet d'obtenir des vitesses de convergence dans le théorème central limite sous des hypothèses d'indépendance (cas uni et multivarié) ou de dépendance positive (cas univarié).

Received by the editors June 1994
Communicated by M. Hallin
AMS Mathematics Subject Classification : 60B10, 60F05, 60E10.
Keywords : Reproducing kernel Hilbert space. Signed measure. Probability metric. Berry Esséen's theorem. Positive dependence.

1 Introduction

L'étude des produits scalaires sur l'espace des mesures signées avec application à l'estimation fonctionnelle a son origine dans les travaux de Ch. Guilbart ([6], 1978). Une idée clé était d'utiliser une représentation fonctionnelle des mesures en les plongeant dans un espace de Hilbert autoreproduisant H_K dont le noyau K vérifie $K(x,y) = \langle \delta_x \,,\, \delta_y \rangle$. Nous avons repris cette idée en nous limitant à des noyaux K définis comme ci dessous dans l'exemple 1 de façon à pouvoir construire les mesures aléatoires comme variables aléatoires hilbertiennes dont les observations soient des intégrales des fonctions de H_K (Suquet [17], [18]). Notre construction explicite des noyaux s'est révélée mal adaptée à l'étude d'un problème comme la vitesse de convergence dans le théorème central limite. Pour cela il est commode de disposer d'espaces autoreproduisants invariants par translations. D'autre part la représentation intégrale d'un noyau particulier associé au mouvement brownien a été utilisée récemment pour l'obtention de principes d'invariance dans L^2 (Oliveira 1990 [12] , Oliveira et Suquet 1993 [13], [14]). Cela nous a incité à adopter une représentation intégrale des noyaux permettant de travailler simultanément avec un espace autoreproduisant et avec un espace L^2. Dans cet article, nous commençons par présenter un cadre unificateur pour l'étude des trois types de problèmes évoqués ci-dessus. On montre ensuite comment cette approche peut être utilisée pour établir des théorèmes de Berry Esséen sous des hypothèses d'indépendance (cas uni et multivarié) ou de dépendance positive (cas univarié).

Soient $\mathcal{X}$ un espace métrique et $\mathcal{M}$ l'espace des mesures signées sur sa tribu borélienne. On pose :

$$K(x,y) = \int_U r(x,u)\bar{r}(y,u)\rho(du), \qquad x,y \in \mathcal{X}, \tag{1}$$

où ρ est une mesure positive sur l'espace U et où la fonction $r : \mathcal{X} \times U \mapsto \mathbf{C}$ vérifie :

$$\sup_{x\in\mathcal{X}} \|r(x,.)\|_{L^2(\rho)} < +\infty. \tag{2}$$

Alors K est un noyau reproduisant borné. On notera H_K l'espace autoreproduisant associé. Voici deux exemples montrant la portée de la représentation (1) :
Exemple 1 : On définit K par :

$$K(x,y) = \sum_{i\in\mathbf{N}} f_i(x)f_i(y), \qquad \sum_{i\in\mathbf{N}} f_i(x)^2 < +\infty, \quad x,y \in \mathcal{X}, \tag{3}$$

où les f_i sont des fonctions réelles sur $\mathcal{X}$. Alors K admet la représentation (1) en prenant pour ρ la mesure de comptage sur $\mathbf{N}$ et en définisant les fonctions r par : $r(x,i) = f_i(x)$. Si $\mathcal{X}$ est séparable, on obtient ainsi en raison du théorème de Mercer tous les noyaux reproduisants continus réels.
Exemple 2 : On prend $\mathcal{X} = \mathbf{R}$, $r(x,u) = \exp(ixu)$ et pour ρ une mesure bornée sur $\mathbf{R}$. On a alors une représentation du type Bochner pour les noyaux stationnaires continus :

$$K(x,y) = \int_{\mathbf{R}} \mathrm{e}^{i(x-y)u}\rho(du) = \Gamma(x-y). \tag{4}$$

2 Distances sur les mesures

2.1 Représentation intégrale des éléments de H_K

La représentation intégrale (1) du noyau K permet d'expliciter une isométrie entre les espaces de Hilbert H_K et $L^2(\rho)$. Ce résultat est bien connu dans le contexte des processus gaussiens, il s'agit alors de l'isométrie entre l'espace gaussien et l'espace autoreproduisant associé au noyau de covariance du processus, cf. par exemple Neveu [8].

Proposition 1 *(i) h est une fonction de H_K si et seulement si il existe une fonction g de $L^2(\rho)$ telle que $h(x) = \int_U g(u)\bar{r}(x,u)\,\rho(du)$.*

(ii) Cette représentation est unique si l'on impose à g d'appartenir à E s.e.v. fermé de $L^2(\rho)$ engendré par $\{r(x,.),\ x \in \mathcal{X}\}$.

(iii) Elle définit alors une isométrie Ψ de H_K sur E.

Preuve : Soit G_K le s.e.v. de H_K formé des combinaisons linéaires finies des $K(x,.)$. Considérons l'égalité fonctionnelle dans G_K :

$$h = \sum_{i=1}^{n} a_i K(x_i,.) = \sum_{j=1}^{m} b_j K(y_j,.).$$

On en déduit :

$$\forall x \in \mathcal{X} \qquad \int_U \sum_{i=1}^{n} a_i r(x_i,u)\bar{r}(x,u)\rho(du) = \int_U \sum_{j=1}^{m} b_j r(y_j,u)\bar{r}(x,u)\rho(du).$$

En notant $g(u) = \sum_{i=1}^{n} a_i r(x_i,u)$ et $g'(u) = \sum_{j=1}^{m} b_j r(y_j,u)$, les fonctions g et g' sont dans E et leur différence est orthogonale à E. On a donc $g = g'$ ρ-presque partout. Cette remarque préliminaire garantit la cohérence de la définition de l'application linéaire Ψ par :

$$\Psi : G_K \to E \qquad h = \sum_{i=1}^{n} a_i K(x_i,.) \longmapsto \Psi(h) = \sum_{i=1}^{n} a_i r(x_i,.).$$

On remarque alors que :

$$\begin{aligned} \|h\|_K^2 = \sum_{i,j=1}^{n} a_i a_j K(x_i,x_j) &= \int_U \sum_{i,j=1}^{n} a_i a_j r(x_i,u)\bar{r}(x_j,u)\rho(du) \\ &= \int_U \Big|\sum_{i=1}^{n} a_i r(x_i,u)\Big|^2 \rho(du). \end{aligned}$$

L'application Ψ injecte donc isométriquement G_K dans un s.e.v. dense de E. Elle admet un prolongement unique noté encore Ψ qui réalise une isométrie de H_K sur E. De plus pour toute $h \in H_K$,

$$\forall x \in \mathcal{X}, \quad h(x) = \int_U \Psi(h)(u)\bar{r}(x,u)\rho(du).$$

Soient maintenant h un élément quelconque de H_K et (h_m) une suite d'éléments de G_K convergeant pour la topologie forte de H_K vers h. Par isométrie, son image (g_m) par Ψ est une suite de Cauchy dans $L^2(\rho)$ et converge donc fortement dans cet espace vers une fonction g. Par autoreproduction, (h_m) converge aussi ponctuellement vers h sur $\mathcal{X}$:

$$\forall x \in \mathcal{X} \quad h(x) = \lim_{m\to+\infty} \int_U g_m(u)\bar{r}(x,u)\rho(du).$$

Par convergence faible dans $L^2(\rho)$ on a d'autre part

$$\lim_{m\to+\infty} \int_U g_m(u)\bar{r}(x,u)\rho(du) = \int_U g(u)\bar{r}(x,u)\rho(du),$$

ce qui achève la preuve de (ii) et (iii). Enfin pour vérifier (i), il suffit de remarquer que

$$\forall x \in \mathcal{X},\ \forall g \in L^2(\rho) \quad \int_U g(u)\bar{r}(x,u)\rho(du) = \int_U \pi_E(g)(u)\bar{r}(x,u)\rho(du),$$

où π_E désigne la projection orthogonale sur E. ∎

Remarque : Dans l'exemple 1, les fonctions f_i sont dans l'espace H_K. Pour le vérifier, il suffit d' utiliser (i) avec pour g l'indicatrice du singleton $\{i\}$. Ce résultat était obtenu de manière moins immédiate dans [17].

2.2 Produit scalaire sur l'espace des mesures

Notre objectif est de définir un produit scalaire sur l'espace $\mathcal{M}$ des mesures signées sur la tribu borélienne de $\mathcal{X}$ en plongeant $\mathcal{M}$ dans H_K ou dans $L^2(\rho)$ grâce à la proposition 1. Cette opération n'est pas possible pour n'importe quel noyau K. Intuitivement, il faut que l'espace H_K soit assez riche en fonctions. Cela nous amène à rajouter une hypothèse de caractérisation des mesures par le noyau K. Pour expliciter cette hypothèse, nous avons besoin du résultat préliminaire suivant :

Proposition 2 *Si la fonction r vérifie (2) et si μ est une mesure signée sur la tribu borélienne de $\mathcal{X}$, alors pour ρ-presque tout u la fonction $r(.,u)$ est μ-intégrable.*

Preuve : Il suffit d'étudier la cas où μ est positive bornée. En utilisant (2) on a

$$\forall (x,y) \in \mathcal{X}^2, \quad \int_U |r(x,u)\bar{r}(y,u)|\,\rho(du) \le \|r(x,.)\|_{L^2(\rho)}\|r(y,.)\|_{L^2(\rho)} \le C < +\infty.$$

Comme $\mu(\mathcal{X})$ est fini, on en déduit

$$\int_{\mathcal{X}^2}\int_U |r(x,u)\bar{r}(y,u)|\,\rho(du)\mu\otimes\mu(dx,dy) < +\infty.$$

Une fonction ρ-intégrable étant finie ρ-presque partout, il en résulte via le théorème de Fubini-Tonnelli sur $\mathcal{X}^2 \times U$ que pour ρ-presque tout u

$$\int_{\mathcal{X}^2} |r(x,u)\bar{r}(y,u)|\,\mu\otimes\mu(dx,dy) < +\infty,$$

d'où en utilisant le théorème de Fubini-Tonnelli sur $\mathcal{X}^2$

$$\int_{\mathcal{X}} |r(x,u)|\, \mu(dx) < +\infty.$$

■

Nous supposerons désormais vérifiée l'hypothèse suivante de caractérisation des mesures signées :

(**CM**) Si $\int_{\mathcal{X}} r(x,u)\mu(dx) = 0$ pour ρ-presque tout u, alors $\mu = 0$.

Nous sommes maintenant en mesure d'appliquer la formule de Guilbart pour définir un produit scalaire sur $\mathcal{M}$.

Proposition 3 *Si les fonctions r et ρ vérifient (2) et (***CM***), alors*

$$\langle \mu\,,\, \nu \rangle_K = \int_{\mathcal{X}^2} K(x,y)\, \mu \otimes \nu(dx,dy) \tag{5}$$

définit un produit scalaire sur $\mathcal{M}$.

Preuve : Les mesures μ et ν sont à variation totale finie et K est borné donc l'intégrale dans (5) a bien un sens. La bilinéarité de $\langle .\,,\,. \rangle_K$ est évidente. Pour la définie positivité, on remarque avec les mêmes justifications qu'à la proposition précédente pour l'interversion des intégrations que

$$\int_{\mathcal{X}^2} K\, d\mu \otimes \mu = \int_U \left| \int_{\mathcal{X}} r(x,u)\mu(dx) \right|^2 \rho(du) \tag{6}$$

et on conclut grâce à l'hypothèse de caractérisation des mesures. ■

2.3 Injection isométrique de $\mathcal{M}$ dans H_K

Les fonctions les plus simples de l'espace fonctionnel H_K sont les combinaisons linéaires $h = \sum_{i=1}^n a_i K(x_i, .)$. Ces fonctions peuvent s'écrire sous la forme $h = \int K(x,.)\, d\mu$ où μ est une combinaison linéaire de masses de Dirac. Si on remplace μ par une mesure quelconque, on obtient encore un élément de H_K. Plus précisément, on a le résultat suivant :

Théorème 1 *Si les fonctions r et ρ vérifient (2) et (***CM***), $(\mathcal{M}, \langle .\,,\,. \rangle_K)$ s'injecte isométriquement sur un sous-espace vectoriel dense de H_K par l'application*

$$\Theta :\ \mathcal{M} \longrightarrow H_K \qquad \mu \longmapsto \int_{\mathcal{X}} K(x,.)\, \mu(dx).$$

Pour toute $h \in H_K$ et toute $\mu \in \mathcal{M}$, on a

$$\langle h\,,\, \Theta(\mu) \rangle_K = \int_{\mathcal{X}} h\, d\mu \qquad \langle \Theta(\mu)\,,\, h \rangle_K = \int_{\mathcal{X}} \bar{h}\, d\mu.$$

Preuve : En invoquant à nouveau le théorème de Fubini, on peut écrire :

$$\Theta(\mu)(y) = \int_U \bar{r}(y,u) \int_{\mathcal{X}} r(x,u)\mu(dx)\rho(du). \tag{7}$$

Considérons l'espace vectoriel $H_1 = \Theta(\mathcal{M})$. Θ est clairement une surjection linéaire de $\mathcal{M}$ sur H_1. Pour vérifier l'injectivité de Θ, on remarque que d'après (6) et (7),

$$\int_{\mathcal{X}} \Theta(\mu)(y)\mu(dy) = \int_{\mathcal{X}^2} K\, d\mu \otimes \mu = \|\mu\|_K^2.$$

La nullité de $\Theta(\mu)$ implique donc celle de μ. On définit maintenant sur H_1 un produit scalaire $\langle .\,,\,. \rangle_{H_1}$ par

$$\langle g\,,\,h \rangle_{H_1} = \left\langle \Theta^{-1}(g)\,,\,\Theta^{-1}(h) \right\rangle_K .$$

On montre alors que l'espace préhilbertien $(H_1, \langle .\,,\,. \rangle_{H_1})$ admet un complété *fonctionnel* et que ce complété est exactement H_K. La preuve détaillée en est donnée (dans un cadre moins général) dans [17]. Cette preuve repose sur les trois lemmes ci dessous. Une fois ces lemmes établis, la preuve dépend uniquement du fait que K est un noyau reproduisant borné. Nous nous contenterons donc de vérifier que les lemmes restent valides avec notre nouvelle construction de K.

Lemme 1 *L'espace H_1 contient toutes les fonctions $K(x,.)$ et vérifie la propriété d'autoreproduction pour le noyau K.*

Preuve : $K(x,.) = \Theta(\delta_x)$ où δ_x est la mesure de Dirac au point x. Pour toute $f = \Theta(\mu)$ de H_1 on a

$$\begin{aligned} \langle f\,,\,K(x,.) \rangle_{H_1} &= \langle \mu\,,\,\delta_x \rangle_K = \int_{\mathcal{X}^2} K\, d\,\mu \otimes \delta_x \\ &= \int_{\mathcal{X}} K(y,x)\mu(dy) = \Theta(\mu)(x) = f(x). \end{aligned}$$

■

Lemme 2 *Pout toute $f \in H_1$, toute $\mu \in \mathcal{M}$, $\langle f\,,\,\Theta(\mu) \rangle_{H_1} = \int_{\mathcal{X}} f\, d\mu$.*

Preuve : Si $f = \Theta(\nu)$, on a d'après (7) :

$$f = \int_U \bar{r}(.,u) \int_{\mathcal{X}} r(x,u)\nu(dx)\rho(du).$$

On en déduit :

$$\begin{aligned} \int_{\mathcal{X}} f\, d\mu &= \int_{\mathcal{X}} \int_U \bar{r}(y,u) \int_{\mathcal{X}} r(x,u)\nu(dx)\,\rho(du)\mu(dy) \\ &= \int_{\mathcal{X}^2} \int_U r(x,u)\bar{r}(y,u)\,\rho(du)\nu \otimes \mu(dx,dy) \\ &= \langle \nu\,,\,\mu \rangle_K = \langle f\,,\,\Theta(\mu) \rangle_{H_1} . \end{aligned}$$

■

Lemme 3 *La convergence d'une suite dans l'espace fonctionnel H_1 implique sa convergence uniforme sur $\mathcal{X}$. On a en effet :*

$$\|\Theta(\mu)\|_\infty \leq \sup_{\mathcal{X}^2} |K|^{1/2} \|\Theta(\mu)\|_{H_1}. \tag{8}$$

Preuve : C'est immédiat avec la propriété d'autoreproduction et l'inégalité de Cauchy-Schwarz pour le produit scalaire de H_1. ∎

2.4 Représentation des mesures dans $L^2(\rho)$

Comme H_K est isométrique par Ψ à un s.e.v. fermé E de $L^2(\rho)$, l'application $\Psi \circ \Theta$ injecte isométriquement $\mathcal{M}$ dans $L^2(\rho)$. L'image de la mesure μ dans $L^2(\rho)$ peut se lire sur la relation (7). Plus précisément, on a :

Proposition 4 *Pour $\mu \in \mathcal{M}$, notons $\zeta_\mu = \Psi \circ \Theta(\mu)$. Alors ζ_μ est l'élément de $L^2(\rho)$ donné par :*

$$\zeta_\mu(u) = \int_{\mathcal{X}} r(x,u)\, \mu(dx) \qquad (u \in U). \tag{9}$$

Preuve : Notons provisoirement $g(u)$ le second membre de (9). La fonction g est bien dans $L^2(\rho)$ puisque d'après (6) :

$$\int_U |g|^2\, d\rho = \int_{\mathcal{X}^2} K\, d\mu \otimes \mu \leq \sup_{\mathcal{X}^2} |K| \cdot |\mu| (\mathcal{X})^2 < +\infty.$$

La relation (7) nous permet d'affirmer que $g = \Psi(\Theta(\mu))$ si l'on prouve que g est dans E (cf. Prop. 1). Soit f un élément quelconque de l'orthogonal de E. On a au moins formellement :

$$\int_U f(u)\bar{g}(u)\, \rho(du) = \int_{\mathcal{X}} \int_U f(u)\bar{r}(x,u)\, \rho(du)\, \mu(dx) = 0 \tag{10}$$

puisque f est orthogonale à tous les $r(x,.)$. L'interversion des intégrations dans (10) se justifie par les majorations suivantes :

$$\left| \int_U |f(u)| \int_{\mathcal{X}} |\bar{r}(x,u)|\ |\mu|\, (dx) \rho(du) \right|^2 \leq \|f\|^2_{L^2(\rho)} \int_U \left| \int_{\mathcal{X}} |\bar{r}(x,u)|\ |\mu|\, (dx) \right|^2 \rho(du)$$

$$\begin{aligned}
\int_U \left| \int_{\mathcal{X}} |\bar{r}(x,u)|\ |\mu|\, (dx) \right|^2 \rho(du) &\leq \int_U \int_{\mathcal{X}^2} |r(x,u)\bar{r}(y,u)|\, |\mu| \otimes |\mu|\, (dx,dy) \rho(du) \\
&\leq \int_{\mathcal{X}^2} \int_U |r(x,u)\bar{r}(y,u)|\, \rho(du)\, |\mu| \otimes |\mu|\, (dx,dy) \\
&\leq |\mu|\, (\mathcal{X})^2 \sup_{x \in \mathcal{X}} \|r(x,.)\|^2_{L^2(\rho)} < +\infty.
\end{aligned}$$

∎

2.5 Exemples

Exemple 1 : On reprend l'exemple 1 donné en introduction. Le noyau s'écrit :

$$K(x,y) = \sum_{i=0}^{+\infty} f_i(x) f_i(y).$$

Pour avoir un noyau uniformément borné, on suppose que les f_i vérifient la condition $\sum_{i=0}^{+\infty} \|f_i\|_\infty^2 < +\infty$. L'hypothèse de caractérisation des mesures s'écrit ici :
Si $\int_{\mathcal{X}} f_i \, d\mu = 0$ *pour tout* i, *alors* $\mu = 0$. Dans [17] et [18], cette construction a été utilisée pour l'étude de problèmes de convergence de suites de mesures aléatoires. Dans cet exemple, $L^2(\rho) = \ell^2(\mathbf{N})$. On vérifie facilement que :

$$\|\Theta(\mu)\|_K^2 = \sum_{i=0}^{+\infty} \left(\int_{\mathcal{X}} f_i \, d\mu \right)^2 = \sum_{i=0}^{+\infty} \left| \langle f_i \, , \, \Theta(\mu) \rangle_K \right|^2$$

et que cette relation s'étend à H_K par densité. Les fonctions f_i constituent donc une structure oblique (*frame*) tendue de borne 1 (cf. [4]). Si on a de plus $\|f_i\|_K = 1$ pour tout i, alors les f_i constituent une base orthonormée de H_K.

Exemple 2 : On prend $\mathcal{X} = U = \mathbf{R}^d$, $r(x,u) = \exp(i \langle x \, , \, u \rangle)$. On choisit pour ρ une mesure positive bornée. On obtient ainsi tous les noyaux continus stationnaires. Les éléments de H_K sont les transformées de Fourier (relativement à ρ) des éléments de $L^2(\rho)$:

$$h \in H_k \quad \text{ssi} \quad \exists g \in L^2(\rho) \quad h(x) = \int_{\mathbf{R}^d} \exp(-i \langle x \, , \, u \rangle) g(u) \, \rho(du).$$

La fonction ζ_μ est tout simplement la fonction caractéristique de la mesure μ.

Exemple 3 : Prenons $\mathcal{X} = U = [0, +\infty[$, $r(x,u) = \exp(-xu)$ et pour ρ une mesure positive bornée à support dans $[0, +\infty[$. Le noyau K s'écrit :

$$K(x,y) = \int_0^{+\infty} \mathrm{e}^{-(x+y)u} \, \rho(du).$$

Les éléments de H_K sont les transformées de Laplace des fonctions de $L^2(\rho)$:

$$h \in H_k \quad \text{ssi} \quad \exists g \in L^2(\rho) \quad h(x) = \int_0^{+\infty} \mathrm{e}^{-xu} g(u) \, \rho(du).$$

La fonction ζ_μ est la transformée de Laplace de la mesure μ.

Exemple 4 : On prend $\mathcal{X} = U = [0,1]$, $r(x,u) = \mathbb{1}_{[0,x]}(u)$. On choisit pour ρ la mesure de Lebesgue. Alors :

$$K(x,y) = \int_0^1 \mathbb{1}_{[0,x]}(u) \mathbb{1}_{[0,y]}(u) \, \lambda(du) = \min(x,y).$$

L'espace H_K est l'espace autoreproduisant du mouvement brownien. Il est à noter que cet espace caractérise seulement les mesures sans masse ponctuelle en 0. Pour remédier à cet inconvénient, il suffit de prendre $\rho = \lambda + \delta_0$, on aura alors $K(x,y) = 1 + \min(x,y)$. Les fonctions de H_K sont les primitives des éléments de $L^2(\rho)$:

$$h(x) = h(0) + \int_0^x g(u)\,\lambda(du), \qquad g \in L^2(\rho),$$

avec $h(0) = 0$ si $\rho = \lambda$. La fonction ζ_μ est la queue de la mesure μ. Dans [19], on donne une classification des mesures en fonction de la régularité de leur représentation dans H_K. En prenant $r(x,u) = \mathbb{1}_{[x,1]}(u)$, le noyau devient $K(x,y) = 2 - \max(x,y)$ et ζ_μ est la fonction de répartition de la mesure μ. Ce noyau a été utilisé dans [12], [13] et [14] pour obtenir des principes d'invariance dans $L^2([0,1])$.

3 La topologie induite sur les probabilités

Si l'espace $\mathcal{X}$ est compact et le noyau K continu, alors la topologie induite sur $\mathcal{M}$ est celle de la convergence faible des mesures (cf. Guilbart [6]). Si $\mathcal{X}$ est métrique séparable, il existe un noyau continu K ayant la même propriété ([6]). Si $\mathcal{X}$ est métrique localement compact et si le noyau K est continu, la topologie induite est en général moins fine que la topologie faible ([6]). En se limitant aux mesures de probabilité, on a néanmoins le résultat suivant :

Théorème 2 *Soit $\mathcal{X}$ métrique séparable localement compact, si K est continu sur $\mathcal{X}^2$, si pour tout $x \in \mathcal{X}$, $K(x,.)$ tend vers 0 à l'infini (au sens du compactifié d'Alexandrov), si μ_n, $n \in \mathbf{N}$ et μ sont des probabilités sur $\mathcal{X}$, alors les trois conditions suivantes sont équivalentes :*
(i) $\Theta(\mu_n) \longrightarrow \Theta(\mu)$ *au sens* $\sigma(H_K, H_K)$
(ii) $\|\mu_n - \mu\|_K \longrightarrow 0$
(iii) μ_n *converge étroitement vers* μ.

Preuve : Notons $\mathcal{C}_0(\mathcal{X})$ l'espace de Banach des fonctions continues sur $\mathcal{X}$ tendant vers 0 à l'infini. On commence par un résultat de densité.

Lemme 4 *Si K vérifie l'hypothèse de caractérisation des mesures et si $K(x,.)$ est dans $\mathcal{C}_0(\mathcal{X})$ pour tout $x \in \mathcal{X}$, alors H_K est dense dans $\mathcal{C}_0(\mathcal{X})$.*

Preuve : On sait que $\mathcal{M}$ peut être identifié au dual de $\mathcal{C}_0(\mathcal{X})$ (théorème de Riesz). La nullité pour tout x de l'intégrale $\int_{\mathcal{X}} K(x,y)\,\mu(dy)$ implique celle de μ en raison de la caractérisation des mesures. Donc les fonctions $K(x,.)$ forment une famille totale dans $\mathcal{C}_0(\mathcal{X})$. D'autre part, grâce à la propriété d'autoreproduction, la convergence pour la topologie de H_K implique la convergence uniforme sur $\mathcal{X}$ (K étant borné sur $\mathcal{X}^2$). Donc toutes les fonctions de H_K sont dans $\mathcal{C}_0(\mathcal{X})$ et H_K est dense dans $\mathcal{C}_0(\mathcal{X})$. ■

Pour prouver le théorème, nous adoptons le schéma :$(i) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i)$, la dernière implication étant évidente.

$(i) \Rightarrow (iii)$: en raison du théorème 1, (i) signifie :

$$\forall h \in H_K, \qquad \int_{\mathcal{X}} h\,d\mu_n \longrightarrow \int_{\mathcal{X}} h\,d\mu, \quad n \to +\infty.$$

Soit $f \in \mathcal{C}_0(\mathcal{X})$, la densité de H_K dans $\mathcal{C}_0(\mathcal{X})$ et le fait que μ_n et μ soient des probabilités impliquent la convergence de $\int_{\mathcal{X}} f\, d\mu_n$ vers $\int_{\mathcal{X}} f\, d\mu$.

$(iii) \Rightarrow (ii)$: On utilise le lemme élémentaire suivant dont nous omettrons la preuve :

Lemme 5 *Dans un espace de Hilbert H, si h_n converge vers h au sens $\sigma(H,H)$ et si $\|h_n\|$ converge vers $\|h\|$, alors h_n converge fortement vers h.*

Comme (iii) implique la convergence $\sigma(H_K, H_K)$ de $\Theta(\mu_n)$ vers $\Theta(\mu)$, il suffit d'établir que $\|\Theta(\mu_n)\|_K$ converge vers $\|\Theta(\mu)\|_K$. Or :

$$\|\Theta(\mu_n)\|_K^2 = \int_{\mathcal{X}^2} K\, d\,\mu_n \otimes \mu_n.$$

La convergence étroite de μ_n vers μ implique celle de $\mu_n \otimes \mu_n$ vers $\mu \otimes \mu$ (cf. Parthasarathy [15]). On peut donc conclure avec la continuité de K. ∎

Voici un exemple où le théorème s'applique. On prend $\mathcal{X} = U = \mathbf{R}^d$, on choisit pour ρ une mesure à densité $\rho = f.\lambda$ où $f \in L^1(\mathbf{R}^d, \lambda)$. Alors le noyau

$$K(x,y) = \int_{\mathbf{R}^d} \exp(iu(x-y)) f(u)\, \lambda(du)$$

est continu. La fonction

$$K(x,.) : \quad y \to K(x,y) = \int_{\mathbf{R}^d} \mathrm{e}^{-iuy}(\mathrm{e}^{iux} f(u))\, \lambda(du)$$

est la transformée de Fourier d'une fonction de L^1 donc est dans $\mathcal{C}_0(\mathcal{X})$.

4 Approximation normale sous indépendance

4.1 Le théorème de Berry-Esséen classique

Soient $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes centrées ayant des moments d'ordre 3 finis. On pose :

$$S_n = \sum_{j=1}^n X_j, \qquad \sigma_j^2 = \mathrm{E}\, X_j^2, \qquad s_n^2 = \mathrm{E}(S_n^2) = \sum_{j=1}^n \sigma_j^2, \qquad S_n^* = s_n^{-1} S_n.$$

La vitesse de convergence de la loi de S_n^* vers la loi gaussienne standard nous est donnée par le théorème de Berry-Esséen (cf. [7]) :

Théorème 3 *Il existe une constante universelle C telle que :*

$$\sup_{x \in \mathbf{R}} |P(S_n^* \leq x) - P(Z \leq x)| \leq C s_n^{-3} \sum_{j=1}^n \mathrm{E}\, |X_j|^3 \tag{11}$$

où Z suit la loi $N(0,1)$.

Le premier membre de (11) s'écrit aussi $\|F_n^* - G\|_\infty$ en notant F_n^* la fonction de répartition de S_n^* et G celle de $N(0,1)$. C'est donc la distance de Kolmogorov entre les deux lois.

Depuis 50 ans, une abondante littérature a été consacrée à la vitesse de convergence dans le théorème central limite. On pourra consulter à ce sujet Hall [7]. D'autres distances que celle de Kolmogorov ont été utilisées, nous renvoyons pour cela à Rachev [16]. Notre objectif est d'établir des résultats du type Berry-Esséen en remplaçant la distance de Kolmogorov par des distances hilbertiennes sur les mesures.

Dans le cadre fonctionnel que nous avons développé à la section 2, les distances correspondant à l'exemple 2 sont bien adaptées à l'étude de l'approximation gaussienne. En raison de l'isométrie entre H_K et $L^2(\rho)$, ces distances s'interprètent aussi comme des distances L^2 entre fonctions caractéristiques. Si on s'en tient à l'aspect calculatoire, il est sans doute plus simple d'adopter ce dernier point de vue. Néanmoins, nous avons choisi de présenter notre premier résultat en travaillant avec l'espace H_K parce que cela permet d'introduire naturellement le choix des noyaux de l'exemple 2. En outre l'interprétation de la distance entre lois à l'aide de H_K nous donne des majorations des quantités $|\mathrm{E}\, h(S_n^*) - \mathrm{E}\, h(Z)|$ pour $h \in H_K$. Cela provient de ce que notre distance d_K possède la ζ-structure au sens de Zolotarev (cf. Rachev [16]). On peut remarquer que l'une des distances utilisées par Maejima et Rachev [11] est un cas particulier de distance L^2 entre fonctions caractéristiques ([16], p. 430). Pour autant que nous sachions, l'étude systématique de ces distances ne semble pas avoir été faite.

4.2 Lemmes préliminaires

Lemme 6 (cf. Araujo, Giné [1]) *Soit $\mathcal{F}$ une famille de fonctions réelles mesurables bornées sur $\mathbf{R}^d$. Soient μ_k, ν_k $k = 1, \ldots, n$ des probabilités sur $\mathbf{R}^d$. Alors si $\mathcal{F}$ est globalement invariante par translations :*

$$\sup_{h\in\mathcal{F}} \left| \int_{\mathbf{R}^d} h\, d(\mu_1 * \cdots * \mu_n - \nu_1 * \cdots * \nu_n) \right| \leq \sum_{k=1}^{n} \sup_{h\in\mathcal{F}} \left| \int_{\mathbf{R}^d} h\, d(\mu_k - \nu_k) \right|.$$

On souhaite exploiter ce lemme avec une famille $\mathcal{F}$ de fonctions de H_K pour laquelle $\|\Theta(\mu)\|_K = \sup_{h\in\mathcal{F}} |\int_{\mathbf{R}^d} h\, d\mu|$. On choisira donc pour $\mathcal{F}$ la boule unité de H_K ou d'un s.e.v. dense. L'hypothèse d' invariance par translations de $\mathcal{F}$ impose alors la stationnarité du noyau :

Lemme 7 *Soit H_K un espace autoreproduisant séparable de fonctions définies sur un groupe additif dont la boule unité soit globalement invariante par translations. Alors les translations sont des isométries de H_K et le noyau est de la forme $K(x, y) = \Gamma(x - y)$.*

Preuve : Pour $a \in \mathcal{X}$, on définit la translation T_a par $T_a h(x) = h(x-a)$. L'invariance globale de la boule unité entraîne que pour tout $h \in H_K$, et tout $a \in \mathcal{X}$, $\|T_a h\|_K \leq \|h\|_K$. On en déduit :

$$\|h\|_K = \|T_{-a} T_a h\|_K \leq \|T_a h\|_K \leq \|h\|_K,$$

d'où $\|T_a h\|_K = \|h\|_K$ et T_a est bien une isométrie de H_K. Soit $(f_i,\, i \in \mathbf{N})$ une base hilbertienne de H_K. En appliquant le théorème de Mercer successivement avec la base $(T_a f_i,\, i \in \mathbf{N})$ puis avec $(f_i,\, i \in \mathbf{N})$, on obtient :

$$K(x,y) = \sum_{i\in\mathbf{N}} T_a f_i(x) T_a f_i(y) = \sum_{i\in\mathbf{N}} f_i(x-a) f_i(y-a) = K(x-a, y-a).$$

En particulier pour $a = -y$, il vient $K(x,y) = K(x-y,0) = \Gamma(x-y)$. ∎

D'autre part, on voudrait avoir un noyau métrisant la convergence étroite des mesures de probabilité. Finalement, le théorème 2, les lemmes 6 et 7 nous conduisent au choix des noyaux de l'exemple 2 :

$$K(x,y) = \int_{\mathbf{R}^d} \mathrm{e}^{i\langle u\,, x-y\rangle} q(u)\, \lambda(du), \qquad q > 0 \ \ \lambda\text{-p.p.}, \quad q \in L^1(\mathbf{R}^d). \tag{12}$$

La condition de stricte positivité de q λ-presque partout sur $\mathbf{R}^d$ assure la caractérisation des mesures. Le lemme suivant, énoncé dans le cas $d = 1$ nous permettra de contrôler les développements à l'ordre 3.

Lemme 8 *Soit K un noyau défini par (12). On suppose que la fonction q vérifie :*

$$\int_{\mathbf{R}} u^6 q(u)\, \lambda(du) = a^2 < +\infty \qquad (a > 0).$$

On note G_K le s.e.v. de H_K formé des combinaisons linéaires finies des $K(x,.)$. On a alors :

$$\forall h \in G_K, \qquad \|h^{(3)}\|_\infty \leq a\|h\|_K. \tag{13}$$

Preuve : Pour x fixé, la fonction $h = K(x,.)$ admet l'écriture intégrale :

$$h(y) = K(x,y) = \int_{\mathbf{R}} \mathrm{e}^{iu(x-y)} q(u)\, \lambda(du) \qquad (y \in \mathbf{R}).$$

En utilisant le théorème de dérivation sous le signe somme de Lebesgue, on voit que h est C^3 et :

$$h^{(3)}(y) = \int_{\mathbf{R}} \mathrm{e}^{iu(x-y)} iu^3 q(u) \lambda(du) \qquad (y \in \mathbf{R}).$$

Par linéarité ceci s'étend à tout G_K. Si $h = \sum_{j=1}^n b_j K(x_j,.)$, on a :

$$h^{(3)}(y) = \int_{\mathbf{R}} \Big(\sum_{j=1}^n b_j \mathrm{e}^{iux_j} \sqrt{q(u)} \Big) \Big(iu^3 \sqrt{q(u)} \mathrm{e}^{-iuy} \Big)\, \lambda(du).$$

On en déduit :

$$\forall y \in \mathbf{R}, \qquad |h^{(3)}(y)|^2 \leq \int_{\mathbf{R}} \Big| \sum_{j=1}^n b_j \mathrm{e}^{iux_j} \Big|^2 q(u)\, \lambda(du) \int_{\mathbf{R}} u^6 q(u)\, \lambda(du). \tag{14}$$

La première intégrale dans (14) s'écrit encore :

$$\int_{\mathbf{R}} \sum_{j,k=1}^n b_j \bar{b}_k \mathrm{e}^{iu(x_j - x_k)} q(u)\, \lambda(du) = \sum_{j,k=1}^n b_j \bar{b}_k K(x_j, x_k) = \|h\|_K^2.$$

∎

4.3 Le cas indépendant univarié

Nous pouvons maintenant énoncer un théorème de Berry-Esséen pour la distance euclidienne associée au noyau K dans le cas où $d = 1$ et où les variables aléatoires X_i sont indépendantes, centrées et pourvues de moments d'ordre 3 finis.

Théorème 4 *Soit K un noyau vérifiant $K(x,y) = \int_{\mathbf{R}} \exp(iu(x-y))q(u)\,\lambda(du)$, la fonction q étant intégrable, strictement positive λ-presque partout sur $\mathbf{R}$ et telle que :*

$$\int_{\mathbf{R}} u^6 q(u)\,\lambda(du) = a^2 < +\infty \qquad (a > 0).$$

On note d_K la distance euclidienne sur $\mathcal{M}$ associée au noyau K et $\mathcal{L}(S_n)$ la loi de la variable aléatoire S_n. On a alors :

$$d_K\Big(\mathcal{L}(S_n), N(0,s_n^2)\Big) \leq \frac{a}{6}\left(\sum_{j=1}^{n} \mathrm{E}\left|X_j^3\right| + \frac{2\sqrt{2}}{\sqrt{\pi}}\sum_{j=1}^{n}\sigma_j^3\right)$$

et la distance d_K métrise la convergence étroite des mesures de probabilité.

Corollaire 1 *Si $S_n^* = s_n^{-1}S_n$, alors :*

$$d_K\Big(\mathcal{L}(S_n^*), N(0,1)\Big) \leq \frac{a}{6}\left(\sum_{j=1}^{n} \mathrm{E}\left|X_j^3\right| + \frac{2\sqrt{2}}{\sqrt{\pi}}\sum_{j=1}^{n}\sigma_j^3\right) s_n^{-3}.$$

Corollaire 2 *Si les variables aléatoires X_j sont indépendantes identiquement distribuées avec $\mathrm{E}\,|X_1^3| = \tau < +\infty$ et $\mathrm{E}\,X_1^2 = \sigma^2$, alors :*

$$d_K\Big(\mathcal{L}(S_n^*), N(0,1)\Big) \leq \frac{a}{6}\left[\frac{2\sqrt{2}}{\sqrt{\pi}} + \frac{\tau}{\sigma^3}\right] n^{-1/2}.$$

Preuve : Il résulte du théorème 2 que la convergence selon d_K d'une suite de probabilités vers une probabilité est la convergence étroite. Par ailleurs, il est immédiat de vérifier que les translations T_a sont des isométries de H_K laissant invariante la boule unité B_K de G_K. Notons que la distance $d_K(\mu,\nu)$ peut s'écrire :

$$\|\Theta(\mu-\nu)\|_K = \sup_{h\in B_K} |\langle h\,,\,\Theta(\mu-\nu)\rangle_K| = \sup_{h\in B_K}\left|\int_{\mathbf{R}} h\,d\mu - \int_{\mathbf{R}} h\,d\nu\right|. \qquad (15)$$

En raison de l'invariance de B_K par translations, (15) montre que l'on n'a pas perdu de généralité en supposant les variables centrées dans l'énoncé du théorème. De plus le lemme 6 s'applique.

Soient alors $Y_1,\ldots,Y_n$ des variables aléatoires indépendantes avec Y_j de loi gaussienne $N(0,\sigma_j^2)$, $j = 1,\ldots n$. D'après le lemme 6 et (15), on a :

$$d_K\Big(\mathcal{L}(S_n), N(0,s_n^2)\Big) \leq \sum_{j=1}^{n}\sup_{h\in B_K}\left|\mathrm{E}\Big(h(X_j) - h(Y_j)\Big)\right|.$$

Les éléments de B_K étant des fonctions C^3, on peut écrire :

$$h(u) = h(0) + h'(0)u + h''(0)\frac{u^2}{2} + h^{(3)}(\alpha u)\frac{u^3}{6}, \qquad \alpha \in [0,1],\ h \in B_K,$$

d'où grâce au lemme 8 :

$$\left|\mathrm{E}\Big(h(X_j) - h(Y_j)\Big)\right| \leq \frac{a}{6}\left(\mathrm{E}\left|X_j^3\right| + \mathrm{E}\left|Y_j^3\right|\right), \qquad h \in B_K. \tag{16}$$

Un calcul élémentaire nous donne :

$$\mathrm{E}\left|Y_j^3\right| = \int_0^{+\infty}\sqrt{\frac{2}{\pi}}\frac{1}{\sigma_j}t^3\exp\Big(\frac{-t^2}{2\sigma_j^2}\Big)\,dt = \sqrt{\frac{8}{\pi}}\sigma_j^3.$$

et on achève la démonstration en sommant sur j dans (16). ■

4.4 Le cas indépendant multivarié

Le théorème 4 se généralise au cas multivarié de la manière suivante :

Théorème 5 *Soient $X_j = (X_j^{(1)}, \dots, X_j^{(d)})$, $j = 1, \dots, n$ des vecteurs aléatoires indépendants, à valeurs dans $\mathbf{R}^d$, centrés ayant des moments absolus d'ordre 3 finis. On note K un noyau défini par $K(x,y) = \int_{\mathbf{R}^d}\exp(i\langle u\,,\,x-y\rangle)q(u)\,\lambda(du)$, la fonction q étant intégrable, strictement positive λ-presque partout sur $\mathbf{R}^d$ et vérifiant :*

$$a_{klm}^2 = \int_{\mathbf{R}^d}(u_k u_l u_m)^2\, q(u_1, \dots, u_d)\, d\lambda(u_1, \dots, u_d) < +\infty, \qquad k, l, m = 1, \dots, d.$$

On désigne par d_K la distance euclidienne sur $\mathcal{M}$ associée au noyau K et par $\mathcal{L}(S_n)$ la loi de la variable aléatoire $S_n = \sum_{j=1}^n X_j$. Alors d_K métrise la convergence étroite des mesures de probabilité et on a :

$$d_K\Big(\mathcal{L}(S_n), N(0,\Gamma_n)\Big) \leq \sum_{j=1}^n\sum_{k,l,m=1}^d \frac{a_{klm}}{6}\Big(\mathrm{E}|X_j^{(k)}X_j^{(l)}X_j^{(m)}| + \mathrm{E}|Y_j^{(k)}Y_j^{(l)}Y_j^{(m)}|\Big),$$

où les Y_j sont des vecteurs aléatoires indépendants, gaussiens centrés tels que X_j et Y_j aient même matrice de covariance ($j = 1, \dots, n$) et Γ_n désigne la matrice de covariance de la somme S_n.

Dans le cas particulier où les X_j sont de même loi, on retrouve bien pour la somme normalisée $n^{-1/2}S_n$ la vitesse de convergence en $n^{-1/2}$ vers $N(0,\Gamma_1)$.

Preuve : On va cette fois utiliser l'espace $L^2(\rho)$ de préférence à H_K. En raison de l'isométrie entre ces deux espaces et du choix du noyau, d_K s'interprète comme la distance $L^2(\rho)$ entre fonctions caractéristiques :

$$d_K\Big(\mathcal{L}(X), \mathcal{L}(Y)\Big)^2 = \int_{\mathbf{R}^d}|\mathrm{E}(\exp(i\langle u\,,\,X\rangle)) - \mathrm{E}(\exp(i\langle u\,,\,Y\rangle))|^2\, q(u)\,\lambda(du).$$

On commence par rappeler deux lemmes élémentaires utilisés dans la démonstration du théorème central limite par les fonctions caractéristiques.

Lemme 9 *Pour tout $t \in \mathbf{R}$, on a :* $\mathrm{e}^{it} = 1 + it - \frac{t^2}{2} + \frac{t^3}{6} R(t)$*, où la fonction* R *vérifie :* $\|R\|_\infty = 1$.

Preuve : Cela résulte immédiatement de l'inégalité bien connue (cf. Feller [5])

$$\left| \mathrm{e}^{it} - 1 - it + \frac{t^2}{2} \right| \leq \frac{|t^3|}{6}, \qquad t \in \mathbf{R}.$$

■

Lemme 10 *Si $z_1, \ldots, z_n$ et $z'_1, \ldots, z'_n$ sont des nombres complexes de module inférieur ou égal à 1, on a :*

$$\left| \prod_{j=1}^n z_j - \prod_{j=1}^n z'_j \right| \leq \sum_{j=1}^n \left| z_j - z'_j \right|.$$

Preuve : cf. Billingsley [2] p. 309. ■

Pour alléger les écritures, nous notons $R_X(u)$ pour $R(\langle u\,,\,X\rangle)$ et $\|\ \|_2$ pour la norme de $L^2(\rho)$. En utilisant le lemme 9, on vérifie que si X et Y sont deux vecteurs aléatoires de $\mathbf{R}^d$ ayant même espérance et même matrice de covariance, la distance d_K entre leurs lois se réduit à :

$$\begin{aligned} d_K(\mathcal{L}(X), \mathcal{L}(Y)) &= \frac{1}{6} \left\| \mathrm{E}(\langle\,.\,,\,X\rangle^3\, R_X) - \mathrm{E}(\langle\,.\,,\,Y\rangle^3\, R_Y) \right\|_2 \\ &\leq \frac{1}{6} \left\| \mathrm{E}\,|\langle\,.\,,\,X\rangle|^3 \right\|_2 + \frac{1}{6} \left\| \mathrm{E}\,|\langle\,.\,,\,Y\rangle|^3 \right\|_2. \end{aligned} \tag{17}$$

Soient $\{X_j,\ j = 1, \ldots n\}$ n vecteurs aléatoires de $\mathbf{R}^d$ ayant des moments absolus d'ordre 3 finis. On choisit les Y_j, $j = 1, \ldots n$ indépendants, gaussiens centrés et tels que pour chaque j, Y_j ait même matrice de covariance que X_j. Alors d'après (17) et le lemme 10, on obtient :

$$d_K\Big(\mathcal{L}(S_n), \mathcal{L}(\sum_{j=1}^n Y_j)\Big) \leq \frac{1}{6} \sum_{j=1}^n \left(\left\| \mathrm{E} \left| \langle\,.\,,\,X_j\rangle \right|^3 \right\|_2 + \left\| \mathrm{E} \left| \langle\,.\,,\,Y_j\rangle \right|^3 \right\|_2 \right)$$

et on termine en remarquant que :

$$\begin{aligned} \left\| \mathrm{E}\,|\langle\,.\,,\,X\rangle|^3 \right\|_2 &= \left(\int_{\mathbf{R}^d} \left| \mathrm{E} \left| \sum_{k=1}^d u_k X^{(k)} \right|^3 \right|^2 q(u_1, \ldots, u_d)\, \lambda(du) \right)^{\frac{1}{2}} \\ &\leq \sum_{k,l,m=1}^d \mathrm{E}|X^{(k)} X^{(l)} X^{(m)}| \left(\int_{\mathbf{R}^d} (u_k u_l u_m)^2\, q(u)\, \lambda(du) \right)^{\frac{1}{2}}. \end{aligned}$$

■

5 Approximation normale et dépendance positive

5.1 Le théorème de Wood

Définition 1 *Les v.a.r.* $(X_n)_{n\geq 1}$ *sont dites :*

a) deux à deux positivement quadrant-dépendantes (PQD) *si pour toute paire d'indices distincts* i, j *on a :*

$$\forall s,t \in \mathbf{R} \qquad P(X_i > s, X_j > t) \geq P(X_i > s)P(X_j > t).$$

b) linéairement positivement quadrant-dépendantes (LPQD) *si pour toute paire* A, B *de parties finies disjointes de* $\mathbf{N}^*$ *et tous réels positifs* $(t_i, i \geq 1)$, *les variables* $\sum_{i\in A} t_i X_i$ *et* $\sum_{i\in B} t_i X_i$ *sont PQD.*

c) associées *si pour tout* n, *tout choix d'indices* $i_1, \dots, i_n$ *et toute paire* f, g *de fonctions croissantes sur* $\mathbf{R}^n$, *on a :*

$$\operatorname{Cov}\left(f(X_{i_1}, \dots, X_{i_n}), g(X_{i_1}, \dots, X_{i_n})\right) \geq 0.$$

Il est clair que si les v.a.r. sont associées, elles sont LPQD.

Newman (1980, [9]) a démontré un théorème central limite pour des v.a. associées. Birkel (1993, [3]) donne un principe d'invariance pour des v.a. LPQD. Oliveira et Suquet, utilisant le noyau évoqué à l'exemple 4 ont obtenu un principe d'invariance LPQD dans $L^2([0,1])$ (1993, [14]). La vitesse de convergence dans le théorème central limite sous dépendance positive a été étudiée par Wood (1983) qui a prouvé le théorème de type Berry-Esséen suivant :

Théorème 6 (Wood, [20]) *Soit* $(X_n)_{n\geq 1}$ *une suite de v.a. associées vérifiant les trois hypothèses suivantes :*

(i) Pour tout n, X_n *est centrée, de variance finie strictement positive, et possède un moment d'ordre 3.*

(ii) La suite $(X_n)_{n\geq 1}$ *est fortement stationnaire : pour tous entiers* $m, j, k_1, \dots,$ k_m, *les vecteurs* $(X(k_1), \dots, X(k_m))$ *et* $(X(k_1 + j), \dots, X(k_m + j))$ *ont même loi.*

(iii) $A^2 = \operatorname{E} X_1^2 + 2\sum_{k=2}^{+\infty} \operatorname{Cov}(X_1, X_k) < +\infty.$

Alors pour $n = m.k$ *:*

$$|F_n(x) - N_A(x)| \leq 16\sigma_k^4 m \frac{A^2 - \sigma_k^2}{9\pi\rho_k^2} + \frac{3\rho_k}{\sigma_k^3} m^{1/2}$$

où $N_A(x)$ *désigne la fonction de répartition de la loi gaussienne centrée de variance* A^2, *et où l'on a noté* $\bar{S}_n = n^{-1/2}(X_1 + \dots + X_n)$, $\sigma_n^2 = \operatorname{E}(\bar{S}_n^2)$ *et* $\rho_n = \operatorname{E}\left|\bar{S}_n^3\right|$.

L'examen de la démonstration du théorème de Wood montre qu'il se généralise sans difficulté au cas des variables LPQD. En utilisant une distance hilbertienne, nous prouvons ci-dessous un théorème de type Berry-Esséen pour des variables LPQD sous des hypothèses plus générales que celles de Wood.

5.2 Approximation normale selon d_K : le cas LPQD

Soit $(X_l)_{l\geq 1}$ une suite LPQD. Pour $n = mk$ on découpe la somme S_n normalisée en m blocs de longueur k renormalisés par leur longueur :

$$\bar{S}_n = \frac{1}{\sqrt{n}}\sum_{l=1}^{n} X_l, \qquad Y_j = \frac{1}{\sqrt{j}}\sum_{(j-1)k<l\leq jk} X_l, \qquad \bar{S}_n = \frac{1}{\sqrt{m}}\sum_{j=1}^{m} Y_j. \tag{18}$$

On pose pour alléger les écritures : $\sigma_n^2 = \mathrm{E}\,\bar{S}_n^2, \quad \sigma_{k,j}^2 = \mathrm{E}\,Y_j^2, \quad \tau_{k,j} = \mathrm{E}\left|Y_j^3\right|.$

Théorème 7 *Soit $(X_l)_{l\geq 1}$ une suite de v.a. LPQD centrées et K un noyau de la forme :*

$$K(x,y) = \int_{\mathbf{R}} \mathrm{e}^{iu(x-y)} q(u)\,\lambda(du), \qquad \int_{\mathbf{R}} u^6 q(u)\,\lambda(du) < +\infty.$$

Alors pour $n = mk$, on a :

$$d_K(\mathcal{L}(\bar{S}_n), N(0,\sigma_n^2)) \leq \frac{a}{6}\Big(\sum_{j=1}^{m}\tau_{k,j} + \frac{2\sqrt{2}}{\sqrt{\pi}}\sum_{j=1}^{m}\sigma_{k,j}^3\Big)m^{-\frac{3}{2}} + b\Big(\sigma_n^2 - \frac{1}{m}\sum_{j=1}^{m}\sigma_{k,j}^2\Big),$$

les constantes positives a et b étant définies par :

$$a^2 = \int_{\mathbf{R}} u^6 q(u)\,\lambda(du), \qquad b^2 = \int_{\mathbf{R}} u^4 q(u)\,\lambda(du).$$

Corollaire 3 *Si de plus la suite $(X_n)_{n\geq 1}$ est faiblement stationnaire au second ordre, i.e. les matrices de covariance des lois de dimension finie sont invariantes par translation d'indices, alors :*

$$d_K(\mathcal{L}(\bar{S}_n), N(0,\sigma_n^2)) \leq \frac{a}{6}\left(m^{-3/2}\sum_{j=1}^{m}\tau_{k,j} + \frac{2\sqrt{2}}{\sqrt{\pi}}\sigma_k^3 m^{-1/2}\right) + b(\sigma_n^2 - \sigma_k^2).$$

Preuve : L'idée de la démonstration est d'approximer $\bar{S}_n$ par une somme normalisée du type $m^{-1/2}\sum_{j=1}^{m} Z_j$ où les Z_j ont même loi que les Y_j mais sont indépendantes. De façon à pouvoir exploiter un lemme de Newman sur les fonctions caractéristiques de v.a. positivement dépendantes, on va travailler ici avec l'espace $L^2(q.\lambda)$ de préférence à H_K. On utilise donc le découpage suivant :

$$d_K(\mathcal{L}(\bar{S}_n), N(0,\sigma_n^2)) = \left(\int_{\mathbf{R}}\left|\mathrm{E}(\exp(iu\bar{S}_n)) - \exp\Big(-\frac{\sigma_n^2 u^2}{2}\Big)\right|^2 q(u)\,\lambda(du)\right)^{1/2}$$

est majoré par $I_1 + I_2 + I_3$ où :

$$\begin{aligned}
I_1^2 &= \int_{\mathbf{R}}\Big|\mathrm{E}(\exp(iu\bar{S}_n)) - \prod_{j=1}^{m}\mathrm{E}\Big(\exp(\frac{iu}{\sqrt{m}}Y_j)\Big)\Big|^2 q(u)\,\lambda(du),\\
I_2^2 &= \int_{\mathbf{R}}\Big|\prod_{j=1}^{m}\mathrm{E}\Big(\exp(\frac{iu}{\sqrt{m}}Y_j)\Big) - \exp\Big(-\frac{u^2}{2m}\sum_{j=1}^{m}\sigma_{k,j}^2\Big)\Big|^2 q(u)\,\lambda(du),\\
I_3^2 &= \int_{\mathbf{R}}\Big|\exp\Big(-\frac{u^2}{2m}\sum_{j=1}^{m}\sigma_{k,j}^2\Big) - \exp\Big(-\frac{\sigma_n^2 u^2}{2}\Big)\Big|^2 q(u)\,\lambda(du).
\end{aligned}$$

Majoration de I_1 : On utilise l'inégalité de Newman démontrée initialement pour des variables associées, mais dont la preuve s'étend immédiatement aux variables LPQD :

Lemme 11 (Newman, [10]) *Si les variables Y_j sont associées,*

$$\Big|\mathrm{E}\Big(\exp(i\sum_{j=1}^m t_jY_j)\Big) - \prod_{j=1}^m \mathrm{E}(\exp(it_jY_j))\Big| \le \frac{1}{2}\sum_{\substack{j,l=1\\ j\neq l}}^m |t_j|\,|t_l|\,\mathrm{Cov}(Y_j,Y_l).$$

Les Y_j définies par (18) sont LPQD comme combinaisons linéaires à coefficients positifs des variables LPQD X_l. En appliquant l'inégalité de Newman avec la décomposition (18), on obtient :

$$I_1 \le \sum_{\substack{j,l=1\\ j\neq l}}^m \frac{1}{2m}\mathrm{Cov}(Y_j,Y_l)\left(\int_{\mathbf{R}} u^4 q(u)\,\lambda(du)\right)^{1/2} = \frac{b}{2}\Big(\sigma_n^2 - \frac{1}{m}\sum_{j=1}^m \sigma_{k,j}^2\Big). \tag{19}$$

Majoration de I_2 : Il suffit d'utiliser le théorème 4 en traitant les variables $m^{-1/2}Y_j$ comme si elles étaient indépendantes. On en déduit :

$$I_2 \le \frac{a}{6}\left(\sum_{j=1}^m \tau_{k,j} + \frac{2\sqrt{2}}{\sqrt{\pi}}\sum_{j=1}^m \sigma_{k,j}^3\right) m^{-3/2}. \tag{20}$$

Majoration de I_3 : Pour $s,t\ge 0$, on a $|\exp(-t)-\exp(-s)| \le |t-s|$, ce qui nous amène à majorer I_3 par :

$$I_3 \le \frac{1}{2}\left(\sigma_n^2 - \frac{1}{m}\sum_{j=1}^m \sigma_{k,j}^2\right)\left(\int_{\mathbf{R}} u^4 q(u)\,\lambda(du)\right)^{1/2}. \tag{21}$$

Remarquons que l'on a toujours $m^{-1}\sum_{j=1}^m \sigma_{k,j}^2 \le \sigma_n^2$ parce que les covariances des Y_j sont positives. En additionnant les inégalités (19), (20) et (21), on obtient le résultat souhaité. La vérification du corollaire est immédiate. ■

6 Remarques

En présentant cette petite étude du théorème de Berry-Esséen, nous n'avions bien évidemment pas la présomption de concurrencer les travaux classiques sur la question. Notre but était de montrer que l'on pouvait retrouver par des techniques élémentaires l'analogue de quelques résultats connus obtenus par des méthodes plus sophistiquées. Une suite naturelle de ce travail serait d'envisager des développements du type Cramér-Edgeworth à partir du développement des fonctions caractéristiques ou par des méthodes de projection orthogonale permettant d'exploiter le caractère euclidien des distances utilisées.

Pour conclure, disons quelques mots sur le choix de la fonction q et donc des constantes qui apparaissent dans les majorations. Pour clarifier cette question, nous partons de la majoration suivante qui donne une application immédiate de nos résultats à l'approximation d'intégrales :

$$|\mathrm{E}\,h(X)-\mathrm{E}\,h(Y)| \leq \|h\|_K d_K(\mathcal{L}(X),\mathcal{L}(Y)), \qquad h \in H_K. \tag{22}$$

On se limite dans ce qui suit au cas indépendant univarié. Notons Z une variable gaussienne standard. Le corollaire 1 et (22) nous donnent :

$$|\mathrm{E}\,h(S_n^*)-\mathrm{E}\,h(Z)| \leq \frac{a}{6}\|h\|_K \left(\sum_{j=1}^n \mathrm{E}\left|X_j^3\right| + \frac{2\sqrt{2}}{\sqrt{\pi}}\sum_{j=1}^n \sigma_j^3\right) s_n^{-3}. \tag{23}$$

Comme le premier membre de (23) ne dépend pas du noyau K, il est naturel de chercher à minimiser le deuxième relativement à K. Autrement dit, h étant une fonction fixée appartenant à l'un des espaces autoreproduisants dont le noyau vérifie les hypothèses du théorème 4, comment choisir la fonction q de façon à minimiser $a\|h\|_K$? Si K est l'un de ces noyaux, en notant $g=\Psi(h)$ comme à la proposition 1, on doit minimiser :

$$A(q)=a^2\|h\|_K^2=\int_{\mathbf{R}}|g(u)|^2 q(u)\,\lambda(du)\int_{\mathbf{R}} u^6 q(u)\,\lambda(du), \tag{24}$$

où g et q sont reliées à h par : $h(x)=\int_{\mathbf{R}}\exp(-iux)g(u)q(u)\,\lambda(du)$. La fonction gq est dans $L^1(\mathbf{R})$ et h est à un facteur de normalisation prés sa transformée de Fourier. Comme la transformation de Fourier dans $L^1(\mathbf{R})$ est injective, le produit gq dans (24) reste égal à une fonction fixe f ne dépendant que de h. Notons au passage qu'on ne gagne rien à remplacer q par une homothétique cq avec $c<1$ puisque g est alors remplacée par g/c. En effectuant le changement de fonction inconnue $q(u)=|f(u)u^{-3}|p(u)$ dans (24), il vient :

$$A(q)=\int_{\mathbf{R}}\frac{|u^3 f(u)|}{p(u)}\,\lambda(du)\int_{\mathbf{R}}|u^3 f(u)|p(u)\,\lambda(du).$$

Grâce à l'inégalité de Cauchy-Schwarz, $A(q)^{1/2}$ est minorée par $\int_{\mathbf{R}}|u^3 f(u)|\lambda(du)$ et ce minorant est effectivement atteint lorque $p(u)\equiv 1$. Donc le choix optimal de q est $q(u)=|f(u)u^{-3}|$ où f est définie par $h(x)=\int_{\mathbf{R}}\exp(-iux)f(u)\,\lambda(du)$. ■

References

[1] A. Araujo and E. Giné. *The central limit theorem for real and Banach valued random variables.* Wiley (1980).

[2] P. Billingsley. *Probability and measure.* Wiley (1979).

[3] T. Birkel. A functional central limit theorem for positively dependent random variables, *J. Multiv. Anal.* 44 (1993), 314–320.

[4] I. Daubechies. *Ten lectures on wavelets.* SIAM (1992).

[5] W. Feller *An introduction to probability theory and its applications,* vol. II. Wiley (1966).

[6] Ch. Guilbart. *Etude des produits scalaires sur l'espace des mesures. Estimation par projection. Tests à noyaux,* Thèse d'Etat, Lille, 1978.

[7] P. Hall. *Rates of convergence in the central limit theorem.* Pitman (1982).

[8] J. Neveu. *Processus aléatoires gaussiens.* Presses de l'Université de Montréal (1968).

[9] C.M. Newman. Normal fluctuations and the FKG inequalities. *Comm. Math. Phys.* 74 119–128.

[10] C.M. Newman, A.L. Wright. An invariance principle for certain dependent sequences. *Ann. Probab.* 9 (1981), 671–675.

[11] M. Maejima, S. T. Rachev. An ideal metric and the rate of convergence to a self-similar process. *Ann. Probab.,* 15, 702–727.

[12] P.E. Oliveira. Invariance principles in L^2. *Comment. Math. Univ. Carolinae* 31 (1990), 357–366.

[13] P.E. Oliveira, Ch. Suquet. Auto-reproducing spaces and invariance principles in L^2. Preprint, *Pub. IRMA* 32-III Lille (1993).

[14] P.E. Oliveira, Ch. Suquet. An invariance principle under positive dependence. Preprint, *Pub. IRMA* 34-I Lille (1994).

[15] K.R. Parthasarathy. *Probability measures on metric spaces.* Academic Press (1967).

[16] S. T. Rachev. *Probability metrics and the stability of stochastic models.* Wiley (1991).

[17] Ch. Suquet. Une topologie préhilbertienne sur l'espace des mesures à signes bornées. *Pub. Inst. Stat. Univ. Paris,* XXXV 2 51–77 (1990).

[18] Ch. Suquet. Convergences stochastiques de suites de mesures aléatoires à signe considérées comme variables aléatoires hilbertiennes. *Pub. Inst. Stat. Univ. Paris* XXXVII 1-2 71–99 (1993).

[19] Ch. Suquet. Représentation de mesures signées dans l'espace autoreproduisant du mouvement brownien *Pub. IRMA* 32-II Lille (1993).

[20] T. E. Wood. A Berry-Esséen theorem for associated random variables. *Ann. Probab.* 11 (1983), 1042–1047.

Charles Suquet
Laboratoire de Statistique et Probabilités
Bât. M2, U.S.T.L., Cité Scientifique
F-59655 Villeneuve d'Ascq Cedex, France.
e-mail : suquet@alea.univ-lille1.fr