Higher Order Correlation coefficients

René Blacher

Higher Order Correlation Coefficients

René BLACHER

Laboratory LJK
Université Joseph Fourier
Grenoble
France

Summary : We continue the study of higher order correlation coefficients. These coefficients complement the classical correlation coefficient by measuring some dependences more and more sharply. We prove that they can detect the existence of functional dependence of type g(X)=g'(Y). We obtain the asymptotic distribution of empirical higher order correlation coefficients.

Summary : On continue l'étude des coefficients de corrélation d'odre supérieur. Ces coefficients complètent le coefficient de corrélation classique en mesurant des dépendances de plus en plus fines. Nous montrons qu'ils peuvent détecter toute dépendance du type g(X)=g'(Y).

Key Words : Correlation coefficients, canonical analysis, functional dependence, conditional probability, orthogonal functions, Hilbertian test.

1 Introduction

Orthogonal polynomials have many interesting geometrical applications in Probability and Statistics. So they have introduced higher order correlation coefficients and higher order variances (cf [17], [2], [8], [10], [9], [7], [6], [13]). They also have introduced new assumptions for the central limit theorem (cf [7]). One can also obtain the distributions of quadratic forms, Gaussian or not Gaussian, and simple methods of calculation of these laws (cf [12]).

In this paper we continue the study of higher order correlation coefficients wich generalize and complement the classical correlation coefficient.

Notations 1.1 *let (X,Y), $X \in \mathbb{R}^p$, $Y \in \mathbb{R}^q$, be a random vector defined on a probability space $(\Omega, \mathcal{A}, P)$. We denote by $\mathcal{Q}$, μ, μ', and $F_{X,Y}$, F_X, F_Y the laws and the distribution functions of (X,Y), X and Y respectively. We denote by $\{P_i\}_{i=0,1,...}$ and $\{Q_j\}_{j=0,1,...}$ two families of orthonormal functions of $L^2(\mathbb{R}^p, \mu)$ and $L^2(\mathbb{R}^q, \mu')$, respectively, such that $P_0 \equiv 1$ and $Q_0 \equiv 1$.*

For every $i > 0$ and $j > 0$, we set

$$\rho_{i,j} = \rho_{i,j}(X,Y) = \mathbb{E}\{P_i(X)Q_j(Y)\} \ ,$$

where $\mathbb{E}$ denotes the expectation .

For example, **at first, we suppose that p=q=1 and that $\{P_i\}_{i=0,1,...}$ and $\{Q_j\}_{j=0,1,...}$ are the families of orthonormal polynomials**.

Then, the $\rho_{i,j}$'s measure polynomial dependences. In particular $\rho_{1,1}$ measures the linear dependence. Indeed, $P_1(x) = \frac{x-\mathbb{E}\{X\}}{\sigma(X)}$ and $Q_1(y) = \frac{y-\mathbb{E}\{Y\}}{\sigma(Y)}$ where $\sigma^2(.)$ is the variance. Therefore $\rho_{1,1}$ is the classical linear correlation coefficient ρ. In the same way, $\rho_{1,2}$, $\rho_{2,1}$ and $\rho_{2,2}$ measure quadratic dependences.

Now, we suppose moreover that $\{P_i\}_{i=0,1,...}$ and $\{Q_j\}_{j=0,1,...}$ are bases of $L^2(\mathbb{R},\mu)$ and $L^2(\mathbb{R},\mu')$, respectively. Therefore, when (X,Y) has a density function f with respect to the product measure $\mu \otimes \mu'$ such that $f \in L^2(\mathbb{R}^q, \mu \otimes \mu')$ then,

$$f(x,y) = 1 + \sum_{i>0,j>0} \rho_{i,j} P_i(x) Q_j(y)$$

in $L^2(\mathbb{R}^q, \mu \otimes \mu')$.

Clearly X and Y are independent if $\rho_{i,j} = 0$ for all $(i,j) \in \{1,2,...\} \otimes \{1,2,...,\}$. Moreover ρ, the classical linear correlation coefficient, is naturally the first term of a sequence of correlation coefficients : $\rho = \rho_{1,1}$.

Obviously the $\rho_{i,j}$'s generalize and complement the classical correlation coefficient. So we call them "higher order polynomial correlation coefficients" or more simply "higher order correlation coefficients".

These higher order coefficients have been introduced by Blacher in 1983 in [8]. In this report, we continue their study and we develop the obtained before results in [6] and [2].

At first when (X,Y) have not density function, we can write

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) + \sum_{i>0,j>0} \rho_{i,j} \Big(\int_{-\infty}^{x} P_i.d\mu \Big) \Big(\int_{-\infty}^{y} Q_j.\mu' \Big)$$

for all $(x,y) \in \mathbb{R}^2$.

Moreover, we can write also the conditional expectation, the conditional probability and the conditional density with the $\rho_{i,j}$'s. For example let $\mathbb{E}\{Y|X=x\}$ be the conditional expectation of Y given X=x. Then,

$$\mathbb{E}\{Y|X=x\} = \mathbb{E}\{Y\} + \sigma(Y) \sum_{i>0} \rho_{i,j} P_i(x) \ \ in \ L^2(\mathbb{R},\mu) \ .$$

Moreover, the normality of the classical correlation coefficient is generalized in a interesting way :

$$\sum_{i>0} \rho_{i,j}^2 \leq 1 \ .$$

Moreover,

$$\sum_{i>0} \rho_{i,j}^2 = 1$$

if and only if there exists $g \in L^2(\mathbb{R},\mu)$ such that $Q_j(Y) = g(X)$ a.s.

As a matter of fact the most part of these properties holds when P_i and Q_j are not orthonormal polynomials and for any p and q. Then, **now we dot suppose that P_i and Q_j are orthonormal polynomials any more**.

Now the $\rho_{i,j}$'s defines orthogonal projections.

Notations 1.2 *let $\mathbb{P}$ and $\mathbb{Q}$ be the subspaces of $L^2(\mathbb{R}^{p+q}, \mathcal{Q})$ generated by the functions $(x,y) \mapsto P_i(x)$, i=1,2,.... and $(x,y) \mapsto Q_j(y)$, i=1,2,.... , respectively.*

We define the matrix ρ by $\rho = \{\{\rho_{i,j}\}\}$, $i > 0$ and $j > 0$.

Then, ρ and ${}^t\rho$ are matrices of orthogonal projection associated to the bases (of $\mathbb{P}$ and $\mathbb{Q}$) $\{P_i\}$ and $\{Q_j\}$.

Then, if $\{P_i\}$ and $\{Q_j\}$ are bases of $L^2(\mathbb{R}^p, \mu)$ and $L^2(\mathbb{R}^q, \mu')$, there exists $g \in L^2(\mathbb{R}^p, \mu)$ and $g' \in L^2(\mathbb{R}^q, \mu')$ such that $g(X) = g'(Y)$ a.s. if and only if 1 is an eigenvalue of ${}^t\rho\rho$. In this case, g' is an eigenvector of ${}^t\rho\rho$.

Now, when canonical analysis exists, canonical correlation coefficients are a particular case of higher order correlation coefficients. For example, let us suppose that $F_{X,Y}$ is ϕ^2-bounded, i.e. $F_{X,Y}$ has a density function f with respect to $\mu \otimes \mu'$, $f \in L^2(\mathbb{R}^2, \mu \otimes \mu')$. Then, the pair of canonical function η_i, ξ_i is a pair of eigenelements of operators of orthogonal projection Π, Π^* associated to ${}^t\rho$, ρ. Moreover, ρ_i, the associated canonical correlation coefficient is the associated eigenvalue. Indeed,

$$\mathbb{E}\{\xi_i(X)\eta_j(Y)\} = \rho_i \delta_{i,j}$$

(where $\delta_{i,j}$ is the Kronecker delta). Finally, η_i is also an eigenvector of ${}^t\rho\rho$ and ρ_i^2 is the associated eigenvalue.

Clearly, the canonical corelation coefficients are a particular case of the higher order correlation coefficients : ρ_i is the correlation coefficient of order (i,i) associated to ξ_i and η_i.

Moreover,

$$f(x,y) = 1 + \sum_{i>0} \rho_i \xi_i(x) \eta_i(y)$$

and $\phi^2 + 1 = \int f^2 . d(\mu \otimes \mu')$.

Now, we understood that, often, one can use canonical correlation coefficients in the general case. We recall some results about this question in section 4.

Now a good point of orthogonal functions is that we can obtain easily estimators of the previous coefficients.

Notations 1.3 *Let $\{(X_\ell, Y_\ell)\}_{\ell \in \mathbb{N}}$, $X_\ell \in \mathbb{R}^p$, $Y_\ell \in \mathbb{R}^q$, be an I.I.D. sequence of random vectors defined on $(\Omega, \mathcal{A}, P)$ with $(X_0, Y_0) = (X, Y)$. For all $n \in \mathbb{N}^*$, we denote by $\mathcal{Q}_n$, μ_n, μ'_n the empirical probabilities associated to $\{(X_\ell, Y_\ell)\}_{\ell=1,2,...,n}$, $\{X_\ell\}_{\ell=1,2,...,n}$, $\{Y_\ell\}_{\ell=1,2,...,n}$.*

For example, $\mu_n(E) = \frac{1}{n}.card\{1 \leq \ell \leq n | X_\ell \in E\}$.

In order to show how we obtain estimators of the $\rho_{i,j}$'s, *we suppose again that $\{P_i\}$ and $\{Q_j\}$ are the families of orthonormal polynomials with p=q=1.* In this case, we can use the empirical orthogonal polynomials.

Let P_i^n and Q_j^n be the orthonormal polynomials associated to μ_n and μ'_n. Then, for every x and y, $P_i^n(x) \overset{a.s.}{\to} P_i(x)$ and $Q_j^n(y) \overset{a.s.}{\to} Q_j(y)$. Indeed, in order to obtain P_i^n and Q_j^n from P_i and Q_j, it is enough to replace moments by empirical moments. For example, $P_1^n(x) = \frac{x - \mathbb{E}_n(X)}{\sigma_n(X)}$ and $Q_1^n(y) = \frac{y - \mathbb{E}_n(Y)}{\sigma_n(Y)}$ where $\mathbb{E}_n(X)$ and $\sigma_n(X)$ are the empirical expectation and the empirical standard deviation.

Clearly by using theses empirical polynomials, we obtain estimators of the $\rho_{i,j}$'s. Indeed, we define $\hat{\rho}_{i,j}^n$ by

$$\hat{\rho}_{i,j}^n = \frac{1}{n}\Big[\sum_{\ell=1}^n P_i^n(X_\ell) Q_j^n(Y_\ell)\Big] = \int P_i^n(x) Q_j^n(y) \mathcal{Q}_n(dx, dy) \ .$$

Then, $\hat{\rho}^n_{i,j} \stackrel{a.s.}{\to} \rho_{i,j}$.

Then, we call $\hat{\rho}^n_{i,j}$ the empirical correlation coefficient of order (i,j). In particular, $\hat{\rho}^n_{1,1}$ is the classical empirical correlation coefficient.

As a matter of fact, we use the $\hat{\rho}^n_{i,j}$'s when we do not know the marginal distributions. Otherwise, we can use the $\rho^n_{i,j}$'s defined as $\rho^n_{i,j} = \int P_i(X)Q_j(Y).\mathcal{Q}_n(dx,dy)$.

Then, we obtain easily estimators of all functions which we have introduced above. For example

$$\hat{f}^n(x,y) = 1 + \sum_{i,j=1}^{h_n} \hat{\rho}^n_{i,j} P^n_i(x) Q^n_j(y)$$

where h_n is an increasing sequence of integers. Then, $\hat{f}^n$ converges almost surely to f when h_n is correctly choosen.

The proofs of these results are simple : it suffices to apply the geometrical properties of empirical orthogonal polynomials. We set $P^n_i = P_i + \sum_{s=0}^i \epsilon_{i,s} P_s$. Then, $\epsilon_{i,s} \stackrel{a.s.}{\to} 0$. We deduce easily the above convergences.

But the geometrical properties of empirical orthogonal functions are mainly interesting in order to obtain asymptotic distributions. Indeed,

$$\epsilon_{i,s} = -\int P_i P_s.d\mu_n + o_p(n^{-1/2}) \ if \ s < i \ \ and \ \ 2\epsilon_{i,i} = 1 - \int P_i P_i.d\mu_n + o_p(n^{-1/2}) \ ,$$

where $o_p(.)$ denotes the sthocastics "o" [1] .

The form of these results is quite remarkable. Indeed, the elementary properties of orthogonal functions show that

$$\epsilon_{i,s} = \int P^n_i P_s.d\mu \ \ if \ s < i \ \ and \ \ \epsilon_{i,i} = \int P^n_i P_i.d\mu - 1 \ .$$

Now, we can generalize these results to other orthogonal families. Indeed, the above geometrical properties holds when orthonormals functions are built up by the Gram Schmidt process.

Then, we can write easily the asymptotical distributions of all above estimators by using orthogonal projection.

Notations 1.4 *let h and k* $\in \mathbb{N}^*$. *We denote by* $\overline{\Pi}^*[P_i]$ *and* $\overline{\Pi}[Q_j]$ *the orthogonal projections of* $(x,y) \mapsto P_i(x)$ *and* $(x,y) \mapsto Q_j(y)$ *onto the subspaces of* $L^2(\mathbb{R}^{p+q}, \mathcal{Q})$ *generated by the functions* $(x,y) \mapsto Q_j(y)$ *, j=0,1,2,.....,k-1, and* $(x,y) \mapsto P_i(x)$*, i=0,1,....,h-1, respectively. Then, we set*

$$C_{i,j}(x,y) = P_i(x)Q_j(y) - \frac{\rho_{i,j}}{2}[P_i(x)^2 + Q_j(y)^2] - Q_j(y)\overline{\Pi}^*[P_i](y) - P_i(x)\overline{\Pi}[Q_j](x) \ .$$

Then, we shall prove that the random matrix $\sqrt{n}\{\{\hat{\rho}^n_{s,t} - \rho_{s,t}\}\}$, $(s,t) \in \{1,2,....,h\} \times \{1,2,....,k\}$ has asymptotically a normal distribution with mean 0 and variance matrix

$$\{\{\mathbb{E}\{C_{i,j}(X,Y)C_{i',j'}(X,Y)\}\}\} \ .$$

[1] According to [19] page 8, section 1.2.5, we write $X_n = o_p(Z_n)$ for two sequences of random variable X_n and Z_n, if $X_n/Z_n \stackrel{p}{\to} 0$.

We shall see that this result is more simple than the old classical results. For example, by using this method, we find immediately that the asymptotic variance of $\hat{\rho}^n_{1,1}$ is equal to

$$\Big(1+\frac{\rho_{1,1}^2}{2}\Big)\mathbb{E}\{P_1(X)^2Q_1(Y)^2\}+\rho_{1,1}^2\frac{\mathbb{E}\{P_1(X)^4\}+\mathbb{E}\{Q_1(Y)^4\}}{4}-\rho_{1,1}\Big[\mathbb{E}\{P_1(X)Q_1(Y)^3\}+\mathbb{E}\{P_1(X)^3Q_1(Y)\}\Big].$$

This result is the good point of empirical orthogonal functions : by simple proofs, we obtain *explicitly each term* of asymptotic matrices in a geometrical form.

With these results, we have Hilbertian independence tests. Indeed, by the central Limit Theorem we know its asymptotical distribution. In particular, when X and Y are independent,

$$n\Big[\sum_{i=1}^{h}\sum_{j=1}^{k}(\rho^n_{i,j})^2\Big]$$

has asymptotically a chi squared distribution with hk degrees of freedom. We deduce an independence test : this test is particular case of the Hilbertian test of Bosq (cf [14]).

2 Elementary properties

At first, we have the following result.

Theorem 2.1 *We suppose that $\{P_i\}$ and $\{Q_j\}$ are two bases of $L^2(\mathbb{R}^p,\mu)$ and $L^2(\mathbb{R}^q,\mu')$, respectively. Let $g\in L^2(\mathbb{R}^p,\mu)$ and $g'\in L^2(\mathbb{R}^q,\mu')$. Then,*

$$\int g(x)g'(y)\mathcal{Q}(dx,dy)=\Big(\int g.d\mu\Big)\Big(\int g'.d\mu'\Big)+\sum_{i>0,j>0}\rho_{i,j}\Big(\int gP_i.d\mu\Big)\Big(\int g'Q_j.d\mu'\Big)\ .$$

Proof We can write $g=\sum_i\alpha_iP_i$ in $L^2(\mathbb{R}^p,\mu)$ and $g'=\sum_j\beta_jQ_j$ in $L^2(\mathbb{R}^q,\mu')$.

Therefore, $g(x)=\sum_i\alpha_iP_i(x)Q_0(y)$ and $g'(y)=\sum_j\beta_jP_0(x)Q_j(y)$ in $L^2(\mathbb{R}^{p+q},\mathcal{Q})$. Then, by the continuity of the scalar product,

$$\int g(x)g'(y)\mathcal{Q}(dx,dy)=\sum_{i,j}\alpha_i\beta_j\rho_{i,j}\ ,$$

with $\rho_{0,0}=1$ and $\rho_{s,0}=\rho_{0,s}=0$ if $s\neq 0$. ■.

Therefore that proves that we can write $F_{X,Y}$ as in the introduction :

$$F_{X,Y}(x,y)=F_X(x)F_Y(y)+\sum_{i>0,j>0}\rho_{i,j}\Big(\int_{-\infty}^{x}P_i.d\mu\Big)\Big(\int_{-\infty}^{y}Q_j.d\mu'\Big)\ .$$

More generally we have the following theorem.

Theorem 2.2 *We suppose that the hypotheses of theorem 2.1 hold. Let E and F be two Borel sets of $\mathbb{R}^p$ and $\mathbb{R}^q$. Then,*

$$P\{X\in E,Y\in F\}=\mu(E)\mu'(F)+\sum_{i>0,j>0}\rho_{i,j}\Big(\int_E P_i.d\mu\Big)\Big(\int_F Q_j.d\mu'\Big)\ .$$

In particular, we have the following theorem

Theorem 2.3 *We suppose that the hypotheses of theorem 2.1 hold. Then, X and Y are independent if and only if* $\rho_{i,j} = 0$ *for all* $i \geq 1$ *and* $j \geq 1$

Now, it is easy to prove the following theorem.

Theorem 2.4 *We suppose that the hypotheses of theorem 2.1 hold.*

Then, there exists a probability density function f with respect to $\mu \otimes \mu'$ *such that* $f \in L^2(\mathbb{R}^q, \mu \otimes \mu')$ *if and only if* $\sum_{i>0,j>0} \rho_{i,j}^2 < +\infty$. *Moreover, under this hypothesis,*

$$f(x, y) = 1 + \sum_{i>0,j>0} \rho_{i,j} P_i(x) Q_j(y) \quad in \quad L^2(\mathbb{R}^{p+q}, \mu \otimes \mu') \ .$$

Now we generalize the normality of the linear correlation coefficient.

Theorem 2.5 *For all* $j \geq 1$, $\sum_{i \geq 1} \rho_{i,j}^2 \leq 1$. *Moreover,* $\sum_{i \geq 1} \rho_{i,j}^2 = 1$ *if and only if there exists* $g \in \mathbb{P}$ *such that* $Q_j(Y) = g(X)$ *a.s.*

Proof It is enough to apply the elementary properties of the orthogonal projection of $(x, y) \mapsto Q_j(y)$ onto $\mathbb{P}$. ■.

As a matter of fact ρ is a matrix of orthogonal projection.

Theorem 2.6 *The pair of operators* Π Π^* *is given by the matrices* ρ, ${}^t\rho$ *with respect to the bases* $\{Q_j\}_{j\geq 1}$, $\{P_i\}_{i\geq 1}$.

Proof Again, it is enough to apply the elementary properties of the orthogonal projection. Indeed, Π Π^* are bounded. Then, the theory of infinite matrices is the simple generalization of the finite case (cf Akhiezer parag 26, Smirnov ch 5, Weidmann ch 6.3). ■.

The following theorem shows that empirical measures are always ϕ-bounded.

Theorem 2.7 *We suppose that* μ *is concentrated in h+1 distinct points. Then, (X,Y) has a density function with respect to* $\mu \otimes \mu'$: $f \in L^2(\mathbb{R}^{p+q}, \mu \otimes \mu')$.

Proof We can suppose that $\{P_i\}$ and $\{Q_j\}$ are bases of $L^2(\mathbb{R}^p, \mu)$ and $L^2(\mathbb{R}^q, \mu')$ (th 3-9 Weidmann). Moreover, $\{P_i\}$ is a finite family : i=0,1,....,h. Then, it is enough to apply theorems 2.5 and 2.4. ■.

3 Functional dependence

We know that $P_i(x) = Q_j(Y)$ a.s. if and only if $\rho_{i,j} = 1$ and that $Q_j(Y) = g(X)$ a.s. if and only if $\sum_{i\geq 1} \rho_{i,j}^2 = 1$.

In order to generalize these results, we need operators of orthogonal projection. In particular, we recall that we denote by $\mathbb{P}$ and $\mathbb{Q}$ are the subspaces of $L^2(\mathbb{R}^{p+q}, \mathcal{Q})$ generated by $(x, y) \mapsto P_i(x)$, i=1,2,.... and $(x, y) \mapsto Q_j(y)$, i=1,2,.... Then, we need operators Π and Π^*.

Notations 3.1 *We denote by* Π *and* Π^* *the operators of orthogonal projections of* $\mathbb{Q}$ *onto* $\mathbb{P}$ *and* $\mathbb{P}$ *onto* $\mathbb{Q}$, *respectively.*

Then, $\rho = \{\{\rho_{i,j}\}\}$ is the matrix of orthogonal projection associated to Π with respect to the bases (of $\mathbb{P}$ and $\mathbb{Q}$) $\{P_i\}$ and $\{Q_j\}$. Morover ${}^t\rho$ is the matrix associated to Π^* (${}^t\rho$ is the transpose of ρ).

Now, when q=1, when $\{P_i\}$ is a basis of $L^2(\mathbb{R}, \mu)$, and when $Q_1(y) = \frac{y-\mathbb{E}(Y)}{\sigma(Y)}$, there exists $g \in \mathbb{P}$, such that Y=g(X) a.s. if and only if $\sum_{i>0} \rho_{i,1}^2 = 1$. Moreover, in this case

$$Y = \mathbb{E}\{Y\} + \sigma(Y)\Big[\sum_{i>0} \rho_{i,1} P_i(X)\Big] \;\; a.s.$$

As a matter of fact, this result is also a particular case of the following theorems.

Theorem 3.1 *Let λ be an eigenvalue of Π, Π^*. Then, $-1 \leq \lambda \leq 1$. Moreover, 1 or -1 is an eigenvalue if and only if there exists $g \in \mathbb{P}$, $g \neq 0$, and $g' \in \mathbb{Q}$, $g' \neq 0$ such that g(X)=g'(Y) a.s. Under this hypothesis, g and g' are eigenelements associated to eigenvalue 1 (or -1). Moreover $g = \Pi(g')$ and $g' = \Pi^*(g)$.*

On the other hand all eigenvalues are equal to 0 when X and Y are independent.

Proof This result is a corollary of theorems 1-2 and 1-5 of [15]. ■

Theorem 3.2 *Let ν be an eigenvalue of ${}^t\rho\rho$. Then, $0 \leq \nu \leq 1$.*

Moreover, 1 is an eigenvalue if and only if there exists $g \in \mathbb{P}$, $g \neq 0$, and $g' \in \mathbb{Q}$, $g' \neq 0$ such that g(X)=g'(Y) a.s. Under this hypothesis, g' is an eigenvector associated to eigenvalue 1. Moreover $g = \Pi(g')$ and $g' = \Pi^*(g)$.

On the other hand all eigenvalues are equal to 0 when X and Y are independent.

Proof We know that λ is an eigenvalue of Π, Π^* if and only if λ^2 is an eigenvalue of $\Pi \circ \Pi^*$ (cf ch 7 of [21] or 3-14 of [6]). ■

In particular, we can generalize theorem 1.4 of Lancaster.

Proposition 3.1 *We suppose that $\{P_i\}$ and $\{Q_j\}$ are two bases of $L^2(\mathbb{R}^p, \mu)$, and $L^2(\mathbb{R}^q, \mu')$, respectively. Then, there exists a measurable function g such that Y=g(X) a.s. if and only if ${}^t\rho\rho = \mathbb{I}$ where $\mathbb{I}$ is the identity matrix.*

Proof Let $\gamma \in L^2(\mathbb{R}^q, \mu')$.

At first, we suppose there exists g such that Y=g(X) a.s. Then, $\gamma(Y) = \gamma[g(X)]$ and $\gamma \circ g \in L^2(\mathbb{R}^p, \mu)$. Then, $\Pi(\gamma) = \gamma \circ g$ and $\Pi^*(\gamma \circ g) = \gamma$, i.e. $\Pi^* \circ \Pi$ is the identity operator.

Now, we suppose ${}^t\rho\rho = \mathbb{I}$. Then, $||\gamma|| = ||\Pi^* \circ \Pi(\gamma)|| \leq ||\Pi(\gamma)|| \leq ||\gamma||$ where $||.||$ is the norm of $L^2(\mathbb{R}^{p+q}, \mathcal{Q})$. Then, $\Pi(\gamma) = \gamma$ in $L^2(\mathbb{R}^{p+q}, \mathcal{Q})$.

Therefore, when $Y_s \in L^2(\Omega, \mathcal{A}, P)$, $(Y = (Y_1, Y_2, ..., Y_q))$, there exists $g^s \in L^2(\mathbb{R}^p, \mu)$ such that $Y_s = g^s(X)$ a.s.

When $Y_s \notin L^2(\Omega, \mathcal{A}, P)$, we use a partition $\{\mathcal{O}_m\}$ of Ω : we define $Y_s^m \in L^2(\Omega, \mathcal{A}, P)$ by $Y_s^m(\omega) = Y_s(\omega)$ when $\omega \in \Omega_m$ and 0 if not. ■

4 Canonical correlation coefficients

At first, we suppose that $F_{X,Y}$ is ϕ^2-bounded. Then, the pair of canonical function η_i, ξ_i is a pair of eigenelements of Π, Π^* and ρ_i, the canonical correlation coefficient, is the associated eigenvalue : $\mathbb{E}\{\xi_i(X)\eta_j(Y)\} = \rho_i\delta_{i,j}$. Then, η_i is also an eigenvector of ${}^t\rho\rho$ and ρ_i^2 is the associated eigenvalue.

Moreover, $f(x,y) = 1 + \sum_{i>0} \rho_i \xi_i(x)\eta_i(y)$ and $\phi^2 + 1 = \int f^2.d(\mu \otimes \mu')$.

Now, in the case of canonical analysis of countable type, we obtain still an orthonormal basis of canonical functions $\{\xi_i\}$ and $\{\eta_i\}$. Indeed, we recall the following theorem.

Theorem 4.1 *We suppose that there exists* $\{\xi_i\}_{i\geq 1}$ *and* $\{\eta_j\}_{j\geq 1}$ *two orthonormal bases of* $\mathbb{P}$ *and* $\mathbb{Q}$ *respectively such that* $\mathbb{E}\{\xi_i(X)\eta_j(Y)\} = \rho_i\delta_{i,j}$. *Then,* ξ_i, η_i *is a pair of eigenelements of* Π, Π^* *with associated eingenvalue* ρ_i.

In this case, $(\{\xi_i\}, \{\eta_i\}, \{\rho_i\})$ is a canonical analysis of countable type of $\mathbb{P}$ and $\mathbb{Q}$ (cf [16]).

In this case, if $\mathbb{P} = L^2(\mathbb{R}^p, \mu)$ and $\mathbb{Q} = L^2(\mathbb{R}^q, \mu')$,

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) + \sum_{i>0} \rho_i \Big(\int_{-\infty}^{x} \xi_i.d\mu \Big) \Big(\int_{-\infty}^{y} \eta_i.d\mu' \Big)$$

for all $(x, y) \in \mathbb{R}^{p+q}$.

In the general case, Dauxois and Pousse have generalized the definition of canonical analysis by using Π and Π^*. Unfortunately the canonical correlation coefficients are not always defined.

However, in practice, it is not important that the theoretical ρ_i exist. Indeed, we calculate the eigenelements $\overline{\xi}_i = \overline{\xi}_i(h)$ and $\overline{\eta}_i = \overline{\eta}_i(h)$ and the eigenvalues $\overline{\rho}_i = \overline{\rho}_i(h)$ of a finite matrix $\overline{\rho}(h) = \{\{\rho_{i,j}\}\}$, i and $j \in \{1, 2, ..., h\}$.

Now, by theorem 2.2, we can specify this approximation.

Theorem 4.2 *We suppose that the hypotheses of theorem 2.2 hold. For all* $h \in \mathbb{N}^*$, *we denote by* $\overline{\mathbb{P}}^h$ *and* $\overline{\mathbb{Q}}^h$ *the subspaces of* $L^2(\mathbb{R}^{p+q}, \mathcal{Q})$ *generated by* $(x, y) \mapsto P_i(x)$, *i=1,2,....,h, and* $(x, y) \mapsto Q_j(y)$, *j=1,2,....,h, respectively.*

Then, there exists $\{\overline{\xi}_i^h\}_{i=1,2,...,h}$ *and* $\{\overline{\eta}_j^h\}_{j=1,2,...,h}$ *two orthonormal bases of* $\overline{\mathbb{P}}^h$ *and* $\overline{\mathbb{Q}}^h$ *respectively, such that* $\mathbb{E}\{\overline{\xi}_i^h(X)\overline{\eta}_j^h(Y)\} = \overline{\rho}_i^h\delta_{i,j}$ *for all* $(i, j) \in \{1, 2, ..., h\}^2$.

Moreover,

$$\boldsymbol{P}\{X \in E, Y \in F\} = \mu(E)\mu'(F) + Lim_{h\to\infty} \Big[\sum_{i=1}^{h} \overline{\rho}_i^h \Big(\int_E \overline{\xi}_i^h.d\mu \Big) \Big(\int_F \overline{\eta}_i^h.d\mu' \Big) \Big] \ .$$

In particular,

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) + Lim_{h\to\infty} \Big\{ \sum_{i=1}^{h} \overline{\rho}_i^h \Big(\int_{-\infty}^{x} \xi_i^h.d\mu \Big) \Big(\int_{-\infty}^{y} \eta_i^h.d\mu' \Big) \Big\}$$

for all $(x, y) \in \mathbb{R}^{p+q}$.

5 Conditional Probabilities

At first, because the conditional expctation is an orthogonal projection, we have the following theorem (cf Lancaster th 1-2).

Theorem 5.1 *We suppose that* $\{P_i\}$ *is a basis of* $L^2(\mathbb{R}^p, \mu)$. *We suppose also that q=1,* $\mathbb{E}\{Y^2\} < +\infty$, $Q_1(y) = \frac{y-\mathbb{E}\{Y\}}{\sigma(Y)}$. *Then,* $\mathbb{E}\{Y|X = x\} \in L^2(\mathbb{R}^p, \mu)$.

Moreover,

$$\mathbb{E}\{Y|X = x\} = \mathbb{E}\{Y\} + \sigma(Y) \sum_{i>0} \rho_{i,1} P_i(x) \ \in L^2(\mathbb{R}^p, \mu) \ .$$

We deduce a series expansion for the conditional probability.

Theorem 5.2 *We suppose that the hypotheses of theorem 2.2 hold. Let $P\{Y \in F|X = x\}$ be the conditional probability of $Y \in F$ given X=x. Then, $P\{Y \in F|X = x\} \in L^2(\mathbb{R}^p, \mu)$. Moreover,*

$$P\{Y \in F|X = x\} = \mu'\{F\} + \sum_{i>0} \Big[\sum_{j>0} \rho_{i,j} \Big(\int_F Q_j.d\mu' \Big) \Big] P_i(x) \ .$$

Proof By theorem 5.1,

$$P\{Y \in F|X = x\} = \mu'\{F\} + \sum_{i>0} \Big[\int_{Y \in F} P_i(X).dP \Big] P_i(x) \ .$$

By theorem 5.1 again ,

$$\int_{Y \in F} P_i(X).dP = \int_F \mathbb{E}\{P_i(X)|Y = y\}.\mu'(dy) = \int \mathbb{1}_F(y) \Big(\sum_{j>0} \rho_{i,j} Q_j(y) \Big) \mu'(dy) \ .$$

By the continuity of scalar product, we deduce 5-2. ■

Now the dependence density is also a conditional density

Theorem 5.3 *We suppose that there exists a probability density function f of (X,Y) with respect to $\mu \otimes \mu'$. Let μ'^x be the conditional distribution of Y given X=x. Let f^x be the function defined by $f^x(y) = f(x,y)$. Then, f^x is μ-almost surely the probability density function of μ'^x with respect to μ'.*

Proof This theorem is proved by the same way as the classical theorem for the Lebesgue measure. ■

In [6], we named f "dependence density" because this density defines completely dependence between X and Y. Now, if $f \in L^2$, we can specify this result.

Theorem 5.4 *We suppose that the hypotheses of theorem 2.4 hold. Then, $f^x \in L^2(\mathbb{R}^q, \mu')$ μ-almost surely.*

Moreover, for all $x \in \mathbb{R}^p$, such that f^x is defined and $f^x \in L^2(\mathbb{R}^q, \mu')$,

$$f^x(y) = 1 + \sum_{j>0} H_j^x Q_j(y) \ \ in \ L^2(\mathbb{R}^q, \mu'),$$

and for all $j > 0$, $H_j^x = \sum_{i>0} \rho_{i,j} P_i(x) \in L^2(\mathbb{R}^p, \mu)$.

Proof Let $\mathcal{D}$ be set of $x \in \mathbb{R}^p$ such that f^x is a density and $f^x \in L^2(\mathbb{R}^q, \mu')$.
When $x \in \mathcal{D}$, we set $f^x = \sum_{j \geq 0} H_j^x Q_j$ in $L^2(\mathbb{R}^q, \mu')$.
When $x \notin \mathcal{D}$, we set $H_0^x = 1$ and $H_j^x = 0$ if $j > 0$.
Then, $\mu(\mathcal{D}) = 1$. Therefore,

$$\begin{aligned} \int [H_j^x]^2.\mu(dx) &= \int \Big[\int f(x,y) Q_j(y).\mu'(dy) \Big]^2.\mu(dx) \\ &\leq \int \Big[f(x,y)^2.\mu'(dy) \Big] \Big[\int Q_j(y)^2.\mu'(dy) \Big].\mu(dx) \\ &= \int \Big[f(x,y)^2.\mu'(dy) \Big].\mu(dx) < \infty \ . \end{aligned}$$

Therefore, $H_j^x \in L^2(\mathbb{R}^p, \mu)$ and

$$\int H_j^x P_i(x).\mu(dx) = \int \Big[\int f(x,y) Q_j(y).\mu'(dy) \Big] P_i(x) \mu(dx) = \rho_{i,j} \ . \ ■$$

6 Estimation

In this section, we use again $\{(X_\ell, Y_\ell)\}_{\ell\in\mathbb{N}}$, an I.I.D. sequence of random vectors. We denote by $\mathcal{Q}_n$, μ_n, μ'_n the empirical probabilities associated to $\{(X_\ell, Y_\ell)\}_{\ell=1,2,...,n}$, $\{X_\ell\}_{\ell=1,2,...,n}$, $\{Y_\ell\}_{\ell=1,2,...,n}$.

For example suppose that P_i^n and Q_j^n are the orthonormal polynomials associated to μ_n and μ'_n, then, $P_1^n(x) = \frac{x-\mathbb{E}_n(X)}{\sigma_n(X)}$ and $Q_1^n(y) = \frac{y-\mathbb{E}_n(Y)}{\sigma_n(Y)}$.

Moreover, $\hat{\rho}_{i,j}^n = \int P_i^n(x)Q_j^n(y)\mathcal{Q}_n(dx,dy)$ and $\rho_{i,j}^n = \int P_i(X)Q_j(Y).\mathcal{Q}_n(dx,dy)$ are estimators of $\rho_{i,j}$.

Now, we generalize these results to other orthogonal families. Indeed, the above geometrical properties holds when orthonormals functions are built up by the Gram Schmidt process. Then, we recall the hypotheses which are necessary in order to build up empirical orthogonal functions.

Notations 6.1 *Let $\{Z_\ell\}_{\ell\in\mathbb{N}}$ be an IID sequence defined on $(\Omega, \mathcal{A}, P)$. Let m be the law of Z_0. For all $n \in \mathbb{N}^*$, we denote by m_n the empirical probability associated to $\{Z_\ell\}_{\ell=1,2,....,n}$.*

Let $z_0, z_1, ..., z_h$ be h+1 real variables. We set $z = (z_0, z_1, ..., z_h)$ and we identify z_s with the function $z \mapsto z_s$. We suppose that $\int z_s^2.dm < +\infty$ for all $s \in 0, 1, ..., h$ and that $z_0, z_1, ..., z_h$ are lineraly independent in $L^2(\mathbb{R}^{h+1}, m)$.

Moreover, let $<,>$ and $||.||$, (resp $<,>_n$ and $||.||_n$) be the scalar product and the norm of $L^2(\mathbb{R}^{h+1}, m)$ (resp $L^2(\mathbb{R}^{h+1}, m_n)$).

Under these hypotheses, we can define orthogonal functions

Notations 6.2 *For all $z \in \mathbb{R}^{h+1}$, we set $\tilde{A}_0(z) = A_0(z) = \tilde{A}_0^n(z) = A_0^n(z) = 1$ and for $h \geq j > 0$, let*

$$\tilde{A}_j(z) = z_j - \sum_{s=0}^{j-1} < z_j, A_s > A_s(z),$$

$$\tilde{A}_j^n(z) = z_j - \sum_{s=0}^{j-1} < z_j, A_s^n >_n A_s^n(z),$$

$$A_j(z) = \frac{\tilde{A}_j(z)}{||\tilde{A}_j||}.$$

$$A_j^n(z) = \frac{\tilde{A}_j^n(z)}{||\tilde{A}_j^n||_n} \quad if\ ||\tilde{A}_j^n||_n \neq 0, \quad A_j^n(z) = 0 \quad if\ ||\tilde{A}_j^n||_n = 0.$$

For example, in order to obtain orthogonal polynomials, we orthogonalize $1, x, x^2, ..., x^h$ in $L^2(\mathbb{R}^p, \mu)$ by the Gram schmidt Process. Moreover, we build up the empirical orthonormal polynomials by orthogonalizing $1, x, x^2, ..., x^h$ in $L^2(\mathbb{R}^p, \mu_n)$. More generally, we can obtain estimators of P_i if $\{P_i\}$ is built up by the Gram Schmidt Process.

Hypotheses 6.1 *In this section, we suppose that $\{P_i\}$, i=0,1,...,h, and $\{Q_j\}$, j=0,1,...,k, are two families of orthonormal functions. Therefore ρ is an $h \times k$ matrix.*

We suppose also that, for all $\ell \in \mathbb{N}$, $Z_\ell = \phi(X_\ell)$ where ϕ is a measurable function. We suppose also that, $P_i(x) = A_i[\phi(x)]$ for all $i \in \{0, 1, ..., h\}$. Then, we define P_i^n by $P_i^n(x) = A_i^n[\phi(x)]$.

We suppose that the corresponding assumptions hold for the Q_j's : $Q_j(y) = B_j[\gamma(y)]$, $Q_j^n(y) = B_j^n[\gamma(y)]$.

For example if $\phi(X) = (1, X, X^2, ..., X^h)$, $\{P_i\}$ is the family of orthonormal polynomials.

Then, the P_i^n's are estimators of the P_i's.

Theorem 6.1 *For all $j \in \{0, 1, ..., h\}$, we set*

$$\tilde{P}_j^n = \tilde{P}_j + \sum_{s=0}^{j} \tilde{\epsilon}_{j,s}^n P_s \quad \text{and} \quad P_j^n = P_j + \sum_{s=0}^{j} \epsilon_{j,s}^n P_s \ .$$

Then, for all $i \in \{0, 1, ..., h\}$ and for all $s \in \{0, 1, ..., i\}$, $\epsilon_{i,s} \stackrel{a.s.}{\to} 0$.

Proof In order to prove this result, we use the same method as in 3-3 of [3]. ■

Remark that $\tilde{\epsilon}_{j,j}^n = 0$, i.e. $\tilde{P}_j^n = \tilde{P}_j + \sum_{s=0}^{j-1} \tilde{\epsilon}_{j,s}^n P_s$.

Then, we can define higher order empirical correlation coefficients.

Notations 6.3 *For all $(i, j) \in \{0, 1, ..., h\} \times \{0, 1, ..., k\}$, we set*

$$\hat{\rho}_{i,j}^n = \int P_i^n(x) Q_j^n(y).\mathcal{Q}_n(dx, dy) \quad \text{and} \quad \rho_{i,j}^n = \int P_i(x) Q_j(y).\mathcal{Q}_n(dx, dy) \ .$$

Moreover, we set $\hat{\rho}^n = \{\{\hat{\rho}_{i,j}^n\}\}_{(i,j)\in\{0,1,...,h\}\times\{0,1,...,k\}}$ and $\rho^n = \{\{\rho_{i,j}^n\}\}_{(i,j)\in\{0,1,...,h\}\times\{0,1,...,k\}}$.

Then, these coefficients are estimators of the $\rho_{i,j}$'s.

Theorem 6.2 *For all $(i, j) \in \{0, 1, ..., h\} \times \{0, 1, ..., k\}$, $\hat{\rho}_{i,j}^n \stackrel{a.s.}{\to} \rho_{i,j}$ and $\rho_{i,j}^n \stackrel{a.s.}{\to} \rho_{i,j}$.*

Proof This theorem is deduced from theorem 6.1. ■.

We obtain also estimators of the canonical correlation coefficients.

Theorem 6.3 *For all $(i, j) \in \{0, 1, ..., h\} \times \{0, 1, ..., k\}$, let (ξ_i, η_j), $(\hat{\xi}_i^n, \hat{\eta}_j^n)$, (ξ_i^n, η_j^n) be the eigenelements of $({}^t\rho, \rho)$, $({}^t\hat{\rho}^n, \hat{\rho}^n)$, $({}^t\rho^n, \rho^n)$, respectively, with associated eigenvalues ρ_i, $\hat{\rho}_i$, ρ_i^n.*

Then, for all $x \in \mathbb{R}^p$, and for all $i \in \{0, 1, ..., h\}$, $\hat{\xi}_i^n(x) \stackrel{a.s.}{\to} \xi_i(x)$ and $\xi_i^n(x) \stackrel{a.s.}{\to} \xi_i(x)$. Moreover, $\hat{\rho}_i^n \stackrel{a.s.}{\to} \rho_i$ and $\rho_i^n \stackrel{a.s.}{\to} \rho_i$.

Proof This theorem is deduced from theorem 6.1. ■.

Remark 6.1 *In theorem 6-7 of [2], we have written that $\hat{\xi}_i^n(x) \stackrel{a.s.}{\to} \tau_i$ and $\xi_i^n(x) \stackrel{a.s.}{\to} \tau_i$. It was an error. Moreover, we had not defined τ_i in this paper : τ_i is defined in part I, theorem 3-23 of [6].*

Finally, we can also obtain estimators of the dependence density and of conditional probability. For example, we have the following theorem.

Theorem 6.4 *We suppose that the hypotheses of theorem 2.4 hold.*

Then, there exists two increasing sequences of integers $\{h_n\}$ and $\{k_n\}$ such that $\int (f - \hat{f}^n)^2.d(\mu \otimes \mu') \stackrel{a.s.}{\to} 0$ where

$$\hat{f}^n(x, y) = 1 + \sum_{i=1}^{h_n} \sum_{j=1}^{k_n} \hat{\rho}_{i,j}^n P_i^n(x) Q_j^n(y) \ .$$

Proof We know that

$$\int (f - \hat{f}^n)^2.d(\mu \otimes \mu') = \sum_{i=1}^{h_n} \sum_{j=1}^{k_n} \left[\hat{\rho}_{i,j}^n - \rho_{i,j}\right]^2 + \sum_{i>h_n \ or \ j>k_n} \rho_{i,j}^2 \ . \ \blacksquare$$

In the same way, we can also obtain estimators of $\mathbb{E}\{Y|X=x\}$, $P\{Y \in F|X=x\}$ and f^x. In particular $\mathbb{E}^n\{Y|X=x\}$ is the OLSE (cf [5]). We shall study these results later.

In order to obtain the asymptotical distribution of the $\hat{\rho}_{i,j}^n$'s, we recall the theorem 1 of [5] (cf also theorem 11, page 23 of [13]).

Theorem 6.5 *We suppose that $\mathbb{E}\{P_i(X)^4\} < +\infty$ for i=0,1,...,h. Then, for all $i \in \{0,1,...,h\}$, $\epsilon_{i,s} = -\int P_i P_s.d\mu_n + o_p(n^{-1/2})$ if $i > s$ and $\epsilon_{i,i} = \frac{1-\int P_i^2.d\mu_n}{2} + o_p(n^{-1/2})$.*

Now we recall that we set $C_{i,j}(x,y) = P_i(x)Q_j(y) - \frac{\rho_{i,j}}{2}[P_i(x)^2 + Q_j(y)^2] - Q_j(y)\overline{\Pi}^*[P_i](y) - P_i(x)\overline{\Pi}[Q_j](x)$ where $\overline{\Pi}^*[P_i]$ and $\overline{\Pi}[Q_j]$ are the orthogonal projections of $P_i(x)$ and $Q_j(y)$ onto the subspaces of $L^2(\mathbb{R}^{p+q},\mathcal{Q})$ generated by the functions $(x,y) \mapsto Q_j(y)$, j=0,1,.....,k-1, and $(x,y) \mapsto P_i(x)$, i=0,1,....,h-1.

Then, we can prove the following theorem.

Theorem 6.6 *We suppose that $\mathbb{E}\{P_i(X)^4\} < +\infty$ for all $i \in \{0,1,...,h\}$ and $\mathbb{E}\{Q_j(Y)^4\} < +\infty$ for all $j \in \{0,1,...,k\}$.*

Then, $\sqrt{n}(\hat{\rho}^n - \rho)$ has asymptotically a normal distribution with mean 0 and covariance matrix $\big\{\{\mathbb{E}\{C_{i,j}(X,Y)C_{i',j'}(X,Y)\}\}\big\}$.

Proof We set

$$P_i^n = P_i + \sum_{s=0}^{i} \epsilon_{i,s} P_s \quad and \quad Q_j^n = Q_j + \sum_{t=0}^{j} \epsilon'_{j,t} Q_t \ ,$$

and we use theorem 6.5.

Then, $\epsilon_{i,s}\epsilon'_{j,t} = o_p(n^{-1/2})$ and

$$\epsilon_{i,s} \int P_s(x)Q_j(y).\mathcal{Q}_n(dx,dy) = \epsilon_{i,s}\mathbb{E}\{P_s(X)Q_j(Y)\} + o_p(n^{-1/2}) \ .$$

We deduce that

$$\sqrt{n}(\hat{\rho}_{i,j}^n - \rho_{i,j})$$

$$\begin{array}{cc} = & \int P_s(x)Q_j(y).\mathcal{Q}_n(dx,dy) - \frac{\rho_{i,j}}{2}\Big[\int P_i^2.d\mu_n + \int Q_j^2.d\mu'_n\Big] \\ - & \sum_{s=0}^{i-1} \mathbb{E}\{P_s(X)Q_j(Y)\} \int P_i P_s.d\mu_n \\ - & \sum_{t=0}^{j-1} \mathbb{E}\{P_i(X)Q_t(Y)\} \int Q_j Q_t.d\mu'_n \\ + & o_p(n^{-1/2}) \end{array}$$

$$= \int C_{i,j}.d\mathcal{Q}_n + o_p(n^{-1/2}) \ .$$

By the holder inequality, $\mathbb{E}\{C_{i,j}(X,Y)^2\} < +\infty$. Then, it is enough to apply the central limit theorem. ■

Therefore, when X and Y are independent $\sqrt{n}\hat{\rho}^n$ has asymptically a normal distribution with mean 0 and covariance matrix the identity matrix.

Now, in some cases, on can obtain asymptotic distribution of some functional estimators.

Theorem 6.7 *We suppose that the hypotheses of theorems 2.4 and 6.6 hold. We write* $\theta^n = \{\{\theta^n_{i,j}\}\}_{(i,j)\in\{0,1,...,h-1\}\times\{0,1,...,k-1\}}$ *where*

$$\hat{f}^n(x,y) = \sum_{i=0}^{h_n}\sum_{j=0}^{k_n} \theta^n_{i,j} P_i(x) Q_j(y) \ .$$

Moreover, in this theorem, we set $\rho = \{\{\rho_{i,j}\}\}_{(i,j)\in\{0,1,...,h-1\}\times\{0,1,...,k-1\}}$ *with* $\rho_{0,0} = 1$ *and* $\rho_{0,i} = \rho_{i,0} = 0$ *if* $i > 0$.

Then, $\sqrt{n}(\theta^n - \rho)$ *has asymptotically a normal distribution with mean 0 and covariance matrix* $\big\{\{\mathbb{E}\{M_{i,j}(X,Y)M_{i',j'}(X,Y)\}\}\big\}$ *where* $M_{i,j}(x,y) = P_i(x)Q_j(y) - P_i(x)\overline{\Pi}'[Q_j](x) - Q_j(y)\overline{\Pi}[P_i](y) - \rho_{i,j}$.

Proof We set $\epsilon^n = \{\{\epsilon_{i,s}\}\}_{(i,s)\in\{0,1,....,h-1\}^2}$ with $\epsilon_{i,s} = 0$ if $s > i$.

Then

$${}^t(P^n_0, P^n_1, ...P^n_{h-1}) = {}^t(P_0, P_1, ...P_{h-1}) + \epsilon^n\ {}^t(P_0, P_1, ...P_{h-1}) \ .$$

Of course, we can write equivalent equalities for the Q_j's : $Q^n_j = Q_j + \sum_{s=0}^{k-1} \beta^n_{j,s} Q_s$ and

$${}^t(Q^n_0, Q^n_1, ...Q^n_{k-1}) = {}^t(Q_0, Q_1, ...Q_{k-1}) + \beta^n\ {}^t(Q_0, Q_1, ...Q_{k-1}) \ .$$

Let $\mathbb{I}_h$ be the idendity matrix (h,h). Then,

$$\theta^n = \big(\mathbb{I}_h + {}^t\epsilon^n + \epsilon^n + {}^t\epsilon^n\epsilon^n\big)\rho^n\big(\mathbb{I}_k + {}^t\beta^n + \beta^n + {}^t\beta^n\beta^n\big) \ .$$

Let $\mathcal{R}^n = \{\{\int P_r P_t.d\mu_n\}\}_{(r,t)\in\{0,1,....,h-1\}^2}$ and $\mathcal{T}^n = \{\{\int Q_r Q_t.d\mu_n\}\}_{(r,t)\in\{0,1,....,k-1\}^2}$. Then, by using theorem 6.5

$$\begin{aligned}\theta^n &= \rho^n + ({}^t\epsilon^n + \epsilon^n)\rho + \rho({}^t\beta^n + \beta^n) + o_p(n^{-1/2}) \\ &= \rho^n + (\mathbb{I}_h - \mathcal{R}^n)\rho + \rho(\mathbb{I}_k - \mathcal{T}^n) + o_p(n^{-1/2}) \ .\end{aligned}$$

We deduce the theorem. ■.

Now, one can suppose that $h, k \to \infty$. Indeed, let $\overline{\Pi}[P_i]^\infty$ be the orthogonal projection of $P_i(x)$ on the subspace generated by the functions $Q_j(y)$, $j \in \mathbb{N}$ (if $\{Q_j\}$ is a basis of $L^2(\mathbb{R}, \mu')$, this subspace is obviously $L^2(\mathbb{R}, \mu')$). Then, we know that $\overline{\Pi}[P_i](y) \to \overline{\Pi}[P_i]^\infty$ as $h \to \infty$ in $L^2(\mathbb{R}^q, \mu')$.

Then, one can obtain various assumptions such that

$$\mathbb{E}\{M_{i,j}(X,Y)M_{i',j'}(X,Y)\} \to \mathbb{E}\{M^\infty_{i,j}(X,Y)M^\infty_{i',j'}(X,Y)\}$$

where $M^\infty_{i,j}(x,y) = P_i(x)Q_j(y) - P_i(x)\overline{\Pi}'[Q_j]^\infty(x) - Q_j(y)\overline{\Pi}[P_i]^\infty(y) - \rho_{i,j}$ and, therefore, the asymptotic distribution of

$$\hat{f}^n(x,y) = 1 + \sum_{i=1}^{h_n}\sum_{j=1}^{k_n} \hat{\rho}^n_{i,j} P^n_i(x) Q^n_j(y)) \ .$$

Moreover, one can obtain the same type of results for the other estimators introduced in this report. In particular for the OLSE. We shall study these problems later.

Thes results are more simple than the old classical results. Indeed in order to obtain asymptotic distributions, one could also use the following theorem (th A, p 122 of [19]).

Theorem 6.8 *Let $U_n = (U_{n,1}, U_{n,2}, ..., U_{n,k}) \in \mathbb{R}^k$ be a random vector asymptotically normal with mean $\overline{m}$ and covariance matrix $b_n\Sigma$ when $b_n \to 0$. Let $g(u) = [g_1(u), g_2(u), ..., g_r(u)] \in \mathbb{R}^r$, $u = (u_1, u_2, ..., u_k)$, be a vector valued function for which each component functions $g_i(u)$ is real valued and has a nonzero differential at $u = \overline{m}$.*

Then, $g(U_n)$ is asymptotically normal with mean $g(\overline{m})$ and covariance matrix $b_n^2 D\Sigma^t D$ where

$$D = \Big[\frac{\partial g_i}{\partial u_j}\Big|_{u=\overline{m}}\Big]_{r\times k} .$$

This theorem was the key of many problems on symptotical distributions. But the obtained formulae my be complicated. For example, in order to obtain the asymptotic distribution of the empirical linear correlation coefficient, we set $\hat{\rho}_{1,1}^n = g(U)$ (cf [19] p 126) where

$$U = \Big(\mathbb{E}_n(X), \mathbb{E}_n(Y), \int x^2.\mu_n(dx), \int y^2.\mu'_n(dy), \int xy.\mathcal{Q}_n(dx, du)\Big)$$
$$= \frac{1}{n}\Big(\sum_{\ell=1}^n X_\ell, \sum_{\ell=1}^n Y_\ell, \sum_{\ell=1}^n X_\ell^2, \sum_{\ell=1}^n Y_\ell^2, \sum_{\ell=1}^n X_\ell Y_\ell\Big) .$$

and

$$g(u_1, u_2, u_3, u_4, u_5) = \frac{u_5 - u_1 u_2}{\sqrt{(u_3 - u_1^2)(u_4 - u_2^2)}} .$$

Now, $b_n = \frac{1}{n}$,

$$\overline{m} = \Big(\mathbb{E}\{X\}, \mathbb{E}\{Y\}, \mathbb{E}\{X^2\}, \mathbb{E}\{Y^2\}, \mathbb{E}\{XY\}\Big)$$

and

$$\Sigma = \begin{pmatrix} \mathbb{E}\{X^2\} - \mathbb{E}\{X\}\mathbb{E}\{X\}) \; , \; \mathbb{E}\{XY\} - \mathbb{E}\{X\}\mathbb{E}\{Y\}) \; , \; \mathbb{E}\{X^3\} - \mathbb{E}\{X\}\mathbb{E}\{X^2\}) \; , \; \mathbb{E}\{XY^2\} - \mathbb{E}\{X\}\mathbb{E}\{Y^2\}) \; , \; \mathbb{E}\{X^2Y\} - \mathbb{E}\{X^2\}\mathbb{E}\{Y\}) \\ \mathbb{E}\{XY\} - \mathbb{E}\{X\}\mathbb{E}\{Y\}) \; , \; \mathbb{E}\{Y^2\} - \mathbb{E}\{Y\}\mathbb{E}\{Y\}) \; , \; \mathbb{E}\{X^2Y\} - \mathbb{E}\{X^2\}\mathbb{E}\{Y\}) \; , \; \mathbb{E}\{Y^3\} - \mathbb{E}\{Y\}\mathbb{E}\{Y^2\}) \; , \; \mathbb{E}\{XY^2\} - \mathbb{E}\{Y\}\mathbb{E}\{XY\}) \\ \mathbb{E}\{X^3\} - \mathbb{E}\{X\}\mathbb{E}\{X^2\}) \; , \; \mathbb{E}\{X^2Y\} - \mathbb{E}\{X^2\}\mathbb{E}\{Y\}) \; , \; \mathbb{E}\{X^4\} - \mathbb{E}\{X^2\}\mathbb{E}\{X^2\}) \; , \; \mathbb{E}\{X^2Y^2\} - \mathbb{E}\{X^2\}\mathbb{E}\{Y^2\}) \; , \; \mathbb{E}\{X^3Y\} - \mathbb{E}\{X^3\}\mathbb{E}\{XY\}) \\ \mathbb{E}\{XY^2\} - \mathbb{E}\{X\}\mathbb{E}\{Y^2\}) \; , \; \mathbb{E}\{Y^3\} - \mathbb{E}\{Y\}\mathbb{E}\{Y^2\}) \; , \; \mathbb{E}\{X^2Y^2\} - \mathbb{E}\{X^2\}\mathbb{E}\{X^2\}) \; , \; \mathbb{E}\{Y^4\} - \mathbb{E}\{Y^2\}\mathbb{E}\{Y^2\}) \; , \; \mathbb{E}\{XY^3\} - \mathbb{E}\{Y^2\}\mathbb{E}\{XY\}) \\ \mathbb{E}\{X^2Y\} - \mathbb{E}\{X\}\mathbb{E}\{XY\}) \; , \; \mathbb{E}\{XY^2\} - \mathbb{E}\{X\}\mathbb{E}\{XY\}) \; , \; \mathbb{E}\{X^3Y\} - \mathbb{E}\{XY\}\mathbb{E}\{X^2\}) \; , \; \mathbb{E}\{XY^3\} - \mathbb{E}\{XY\}\mathbb{E}\{Y^2\}) \; , \; \mathbb{E}\{X^2Y^2\} - \mathbb{E}\{XY\}\mathbb{E}\{XY\}) \end{pmatrix}$$

Then the form of asymptotic variance of $\hat{\rho}_{1,1}^n$ is complicated. For example,

$$\frac{\partial g}{\partial u_1} = \frac{[u_5 - u_1u_2]'[\sqrt{(u_3 - u_1^2)(u_4 - u_2^2)}] - [u_5 - u_1u_2][\sqrt{(u_3 - u_1^2)(u_4 - u_2^2)}]'}{|(u_3 - u_1^2)(u_4 - u_2^2)|}$$
$$= \frac{-u_2\sqrt{(u_3 - u_1^2)(u_4 - u_2^2)} - (1/2)[u_5 - u_1u_2][(-2u_1^2)(u_4 - u_2^2)]\big\{(u_3 - u_1^2)(u_4 - u_2^2)\big\}^{-1/2}}{|(u_3 - u_1^2)(u_4 - u_2^2)|}$$
$$= \frac{-u_2}{\sqrt{(u_3 - u_1^2)(u_4 - u_2^2)}} + \frac{[u_5 - u_1u_2][u_1^2(u_4 - u_2^2)]}{\big\{(u_3 - u_1^2)(u_4 - u_2^2)\big\}^{3/2}} .$$

Therefore

$$\frac{\partial g}{\partial u_1}(\overline{m}) = \frac{-\mathbb{E}\{Y\}}{\sqrt{\big(\mathbb{E}\{X^2\} - \mathbb{E}\{X\}^2\big)\big(\mathbb{E}\{Y^2\} - \mathbb{E}\{Y\}^2\big)}} + \frac{\mathbb{E}\{X\}^2\big[\mathbb{E}\{XY\} - \mathbb{E}\{X\}\mathbb{E}\{Y\}\big]\big(\mathbb{E}\{Y^2\} - \mathbb{E}\{Y\}^2\big)}{\big\{\big(\mathbb{E}\{X^2\} - \mathbb{E}\{X\}^2\big)\big(\mathbb{E}\{Y^2\} - \mathbb{E}\{Y\}^2\big)\big\}^{3/2}} .$$

Now it is needed to write also $\frac{\partial g}{\partial u_s}(\overline{m})$ for s=2,3,4,5.

Then, a priori, by using theorem 6.8, one see that the asymptotic variance of $\hat{\rho}^n_{1,1}$ is a rational fraction of square roots of linear combination of polynomials in $\mathbb{E}\{X\}$, $\mathbb{E}\{X^2\}$, $\mathbb{E}\{X^2\}$, $\mathbb{E}\{Y^2\}$ and $\mathbb{E}\{XY\}$.

Then the calculations in order to obtain explicitly this variance are very complicated. If they are done, we see that the writing of this variance is indeed complicated. However, this is only the variance of the correlation coefficient of order (1,1). Imagine what it will be for other correlation coefficients.

In contrast, if we use the theorem 6.6, we saw that this variance is given in a simple geometric form. Moreover the computation is also much simpler. Indeed, by using theorem 6.6, we find immediately that the asymptotic variance of $\hat{\rho}^n_{1,1}$ is equal to

$$\Big(1+\frac{\rho_{1,1}^2}{2}\Big)\mathbb{E}\{P_1(X)^2Q_1(Y)^2\}+\rho_{1,1}^2\frac{\mathbb{E}\{P_1(X)^4\}+\mathbb{E}\{Q_1(Y)^4\}}{4}-\rho_{1,1}\Big[\mathbb{E}\{P_1(X)Q_1(Y)^3\}+\mathbb{E}\{P_1(X)^3Q_1(Y)\}\Big].$$

This result is the good point of empirical orthogonal functions : by simple proofs, we obtain *explicitly each term* of asymptotic matrices in a geometrical form.

7 Hilbertian independence test

Because we have the asymptotic distribution of ρ^n, we can deduce a Hilbertian independence test.

Theorem 7.1 *We suppose that the hypotheses of theorem 6.6 hold. We set*

$$||\hat{S}_n||^2 = n\Big[\sum_{i=1}^{h}\sum_{j=1}^{k}(\hat{\rho}^n_{i,j})^2\Big]$$

$$||S_n||^2 = n\Big[\sum_{i=1}^{h}\sum_{j=1}^{k}(\rho^n_{i,j})^2\Big]\ .$$

Then, if X and Y are independent, $||\hat{S}_n||^2$ and $||S_n||^2$ have asymptotically a chi squared distribution with hk degrees of freedom.

Moreover, il there exists $(i,j)\in\{1,2,...,h\}\times\{1,2,...,k\}$ such that $\rho_{i,j}\neq 0$, then, $||\hat{S}_n||^2 \stackrel{a.s.}{\rightarrow} +\infty$ and $||S_n||^2 \stackrel{a.s.}{\rightarrow} +\infty$.

We point out that Bosq has studied the asymptotic power of the test associated to $||S_n||^2$ when $h=h(n)\rightarrow+\infty$ and $k=k(n)\rightarrow+\infty$ (cf [14]). Moreover, the power of the test associated to $||\hat{S}_n||^2$ is studied in [6].

8 Examples

8.1 Polynomial correlation coefficients

We suppose that $\{P_i\}$ and $\{Q_j\}$ are the families of orthonormal polynomials (with p=q=1). Then, the $\rho_{i,j}$'s measure polynomials dependences. Thus, $\rho_{1,2}$, $\rho_{2,1}$ and $\rho_{2,2}$ measure quadratic dependences. For example, $Y=aX^2+bX+c$ if and only if

$$\rho_{1,1}^2+\rho_{2,1}^2=1.$$

Now, when $h, k \to \infty$, we know that $F^{h,k}_{X,Y}(x,y) \to F_{X,Y}(x,y)$ where

$$F^{h,k}_{X,Y}(x,y) = F_X(x)F_Y(y) + \sum_{i=1}^{h}\sum_{j=1}^{k}\rho_{i,j}\Big(\int_{-\infty}^{x} P_i.d\mu\Big)\Big(\int_{-\infty}^{y} Q_j.d\mu'\Big)\ .$$

For example if X and Y have the uniform distribution U([0,1]), then, $P_i = Q_i = L_i$, the Legendre polynomials of degree i, and if $\rho_{i,j} = 0$ for $i \le k$ and $j \le k$, then, by [11] and [8],

$$\mathcal{D}(X,Y) \le 1,372\sqrt{log\Big(\frac{2k+1}{2k-3}\Big)}\ ,$$

where

$$\mathcal{D}(X,Y) \le \sqrt{90\Big(\int\int\big[F_{X,Y}(x,y) - F_X(x)F_Y(y)\big]\mu(dx)\mu'(dy)\Big)}\ .$$

As a matter of fact, $\mathcal{D}(X,Y)$ is an standardized indicator of dependence which checks the axioms of Renyi (cf [22], or 1-5 [8]).

We recall also that $\{P_i\}$ may be not a basis of $L^2(\mathbb{R},\mu)$ (cf Natanson p 149-150). But $\{P_i\}$ is a basis when the supporting set of μ is bounded or also when X has a Normal or a Gamma distribution (cf Natanson).

8.2 Hermite correlation coefficients

We are in a particular case of polynomials correlation coefficients. When $X \sim N(0,1)$ and $Y \sim N(0,1)$, the orthonormal polynomials are the Hermite polynomials H_i. They are a basis of $L^2(\mathbb{R},\mu)$.

In particular, when (X,Y) is a normal vector $(X,Y) \sim N_2(0,C)$, $H_i(x)$ and $H_i(y)$ are the canonical functions with ρ^i for associated correlation coefficient with $\rho = \rho_{1,1}$, i.e. $\rho_i = \rho^i$, $\rho_{i,j} = \rho_i\delta_{i,j}$ (cf Lancaster 3-5-2). There is a single dependence parameter ρ. Therefore, if $\rho_{1,1} = \rho = 0$, $\rho_{i,j} = 0$ and X and Y are independent.

Now, if we suppose only that $X \sim N(0,1)$ and $Y \sim N(0,1)$, we have a countable number of dependence parameters, the $\rho_{i,j}$'s. In this case, $\rho_{1,1}$ may be equal to 0 even if X and Y are not independent.

Now we suppose that $(X,Y) \in \mathbb{R}^4$ is a nonsingular normal vector, $X = (X_1, X_2, X_3)$, $X_s \sim N(0,1)$, $Y \sim N(0,1)$. We suppose that X_1, X_2 and X_3 are not independent. Then, the family $\{H_{i_1}(X_1)H_{i_2}(X_2)H_{i_3}(X_3)\}$ is not orthogonal.

But we know that there exists a matrix Λ such that ${}^t(U_1,U_2,U_3) = \Lambda^t(X_1,X_2,X_3)$ where (U_1,Y), U_2 and U_3 are independent and $U_s \sim N(0,1)$ (Csaki Fisher page 39-43). So we set $\mathcal{H}_{i_1,i_2,i_3}(x_1,x_2,x_3) = H_{i_1}(u_1)H_{i_2}(u_2)H_{i_3}(u_3)$ where ${}^t(u_1,u_2,u_3) = \Lambda^t(x_1,x_2,x_3)$. Then, $\{\mathcal{H}_{i_1,i_2,i_3}\}$ is an orthonormal basis of $L^2(\mathbb{R}^3,\mu)$.

Moreover $\rho_{(i_1,i_2,i_3),J} = \mathbb{E}\{\mathcal{H}_{i_1,i_2,i_3}(X)H_J(Y)\} = \overline{\rho}^{i_1}\delta_{i_1,J}\delta_{i_2,0}\delta_{i_3,0}$ where $\overline{\rho}$ is the linear correlation coefficient of U_1 and Y.

8.3 Spearmann correlation coefficients

We suppose that F_X and F_Y are continue with p=q=1. We know that $\big(F_X(X),F_Y(Y)\big) \in [0,1]^2$. Moreover, $F_X(X)$ and $F_Y(Y)$ have the uniform distribution μ_u.

Let $\{L_i\}$ be the family of the Legendre orthonormal polynomials. Then $\{L_i\}$ is an orthonormal basis of $L^2(\mathbb{R},\mu_u)$ (cf Natanson).

We know that ρ^S, the Spearmann correlation coefficient of (X,Y) is equal to the linear correlation coefficient of $F_X(X)$ and $F_Y(Y)$. Then, we denote by $\rho^S_{i,j}$ the polynomial correlation coefficient of order (i,j) of $F_X(X)$ and $F_Y(Y)$: $\rho^S_{i,j} = \mathbb{E}\{L_i[F_X(X)]L_j[F_Y(Y)]\}$. Clearly the $\rho^S_{i,j}$'s generalize and complement $\rho^S = \rho^S_{1,1}$. Then, we can call them "higher order Spearmann correlation coefficient".

We set $\tilde{\rho}_{i,j} = \rho_{i,j}$ if $i \geq 1$ and $j \geq 1$ and 0 if not. Then, by 3-7 of [11], one can write

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\beta_{i,j}L_i[F_X(x)]L_j[F_Y(y)] \ ,$$

where

$$\beta_{i,j} = \frac{\frac{\tilde{\rho}_{i+1,j+1}}{\sqrt{(2i+3)(2j+3)}} + \frac{\tilde{\rho}_{i-1,j-1}}{\sqrt{(2i-1)(2j-1)}} - \frac{\tilde{\rho}_{i-1,j+1}}{\sqrt{(2i-1)(2j+3)}} - \frac{\tilde{\rho}_{i+1,j-1}}{\sqrt{(2i+3)(2j-1)}}}{4\sqrt{(2i+1)(2j+1)}} \ .$$

Then, as in section 8.1, one can deduce

$$\mathcal{D}(X,Y) \leq 1,372\sqrt{log\Big(\frac{2k+1}{2k-3}\Big)} \ ,$$

if $\rho^S_{i,j} = 0$ for $i \leq k$ and $j \leq k$.

8.4 Haar Correlation coefficient

Let $\{E_i\}_{i=0,1,\ldots,h}$ and $\{F_j\}_{j=0,1,\ldots,k}$ be two partitions of $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. If P_i and Q_j are linerar combinations of the indicator functions $\mathbb{1}_{E_s}$ and $\mathbb{1}_{F_t}$, we obtain systems of Haar. Moreover, in order to obtain orthonormal basis, it is enough to use sequences of partitions judiciously chosen.

For example, when $\mu = \mu_u$ the lebesgue measure on [0,1[, we can use Haar special system $\{W_i\}$: $W_0 \equiv 1$, and for $i = 2^{k'} + m$, $m < 2^{k'}$, $k', m \in \mathbb{N}$, $W_i(x) = 1$ if $\frac{2m}{2^{k'+1}} \leq x < \frac{2m+1}{2^{k'+1}}$, $W_i(x) = -1$ if $\frac{2m+1}{2^{k'+1}} \leq x < \frac{2m+2}{2^{k'+1}}$ and $W_i(x) = 0$ if not. Afterwards, we normalize the W_i's.

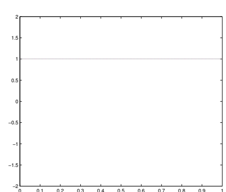

Figure 1: W_0

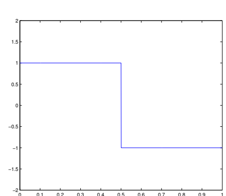

Figure 2: W_1

Now, when, we truncate the series expansion of P (cf theorem 2.2), we obtain an approximation of $\mathcal{Q}$. This approximation has a density function with respect to $\mu \otimes \mu'$:

$$f^{h,k}(x,y) = 1 + \sum_{i=1}^{h}\sum_{j=1}^{k}\rho_{i,j}P_i(x)Q_j(y) \ .$$

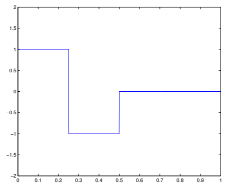

Figure 3: W_2

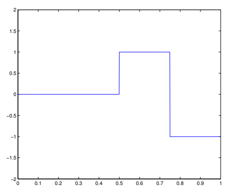

Figure 4: W_3

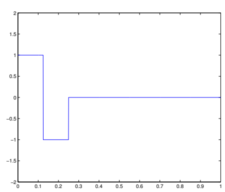

Figure 5: W_4

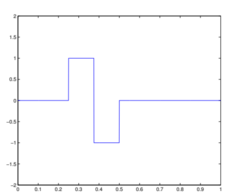

Figure 6: W_5

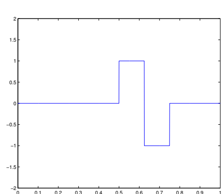

Figure 7: W_6

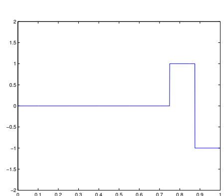

Figure 8: W_7

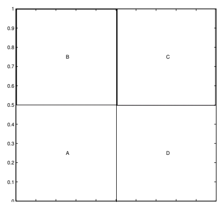

Figure 9: $\rho_{1,1} = \sqrt{2}[\mathcal{Q}(A) + \mathcal{Q}(C) - \mathcal{Q}(B) - \mathcal{Q}(D)]$

Clearly f is constant on every rectangle $E_i \times F_j$: $f(x,y) = \frac{\mathcal{Q}(E_i \times F_j)}{\mu(E_i)\mu'(F_j)}$. In the same way,

$$\mathbf{P}^{h,k}\{Y \in F | X = x\} = \mu'(F) + \sum_{i=1}^{h} \Big[\sum_{j=1}^{k} \rho_{i,j} \int_F Q_j.d\mu' \Big] P_i(x)$$

is an approximation of the conditional probability. Thus, we can choose $\{F_j\}$ such that there exists j with $F = F_j$. Then, when $x \in E_i$,

$$\mathbf{P}^{h,k}\{Y \in F | X = x\} = \frac{\mathcal{Q}(E_i \times F)}{\mu(E_i)} = \mathbf{P}\{Y \in F | X \in E_i\} \ .$$

Moreover, when p=1, we can suppose that the E_i's are intervals. Then, $\mathbf{P}\{Y \in F | X = x\}$ is the limit in L^2 of $\frac{\mathcal{Q}(E_i \times F)}{\mu(E_i)}$, $x \in E_i$, when the length $|E_i|$ of E_i converges to 0 :

$$\mathbf{P}\{Y \in F | X = x\} = lim_{|E_i| \to 0,\ x \in E_i} \Big(\frac{\mathcal{Q}(E_i \times F)}{\mu(E_i)} \Big) \ \ in \ L^2(\mathbb{R}, \mu) \ .$$

In order to estimate P_i and Q_j, we remark that hypotheses 6.1 hold : we orthogonalize $\{\mathbb{1}_{E_i}\}$ and $\{\mathbb{1}_{F_j}\}$ with respect to μ_n and μ'_n. For example, we can choose $\phi = (1, \mathbb{1}_{E_0}, \mathbb{1}_{E_1}, ..., \mathbb{1}_{E_{h-1}})$. Then, we estimate f by

$$f^n(x,y) = 1 + \sum_{i=1}^{h} \sum_{j=1}^{k} \rho^n_{i,j} P^n_i(x) Q^n_j(y)$$

or

$$\hat{f}^n(x,y) = 1 + \sum_{i=1}^{h} \sum_{j=1}^{k} \hat{\rho}^n_{i,j} P^n_i(x) Q^n_j(y) \ .$$

Let $n_{i,j}$ (resp, $n_{i.}$, $n_{.j}$) be the number of (X_ℓ, Y_ℓ) which belongs to $E_i \times F_j$, (resp, $E_i \times \mathbb{R}^q$, $\mathbb{R}^p \times F_j$) when $1 \le \ell \le n$. Then, $\hat{f}^n$ and f^n are constant on each $E_i \times F_j$:

$$\hat{f}^n(x,y) = n \ \frac{n_{i,j}}{n_{i.}n_{.j}} \quad and \quad f^n(x,y) = \frac{n_{i,j}}{n\mu(E_i)\mu'(F_j)} \ .$$

Moreover, when $x \in E_i$,

$$\mu'_n(F_j) + \sum_{s=1}^{h} \Big[\sum_{t=1}^{k} \hat{\rho}^n_{s,t} \int_F Q_t.d\mu'_n \Big] P^n_s(x) = \frac{n_{i,j}}{n_{i.}} \ ,$$

the empirical probability of $Y \in F_j$ given $X \in E_i$.

Now, the Hilbertian independence tests are the chi squared independence tests. When the marginal distributions are unknown, the following equalities hold :

$$\hat{\chi}^2_{X,Y} = n \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{\big(n_{i,j} - \frac{n_{i.}n_{.j}}{n}\big)^2}{n_{i.}n_{.j}} = n \int\int (\hat{f}^n - 1)^2.d\mu_n d\mu'_n = n \sum_{i=1}^{h} \sum_{j=1}^{k} (\hat{\rho}^n_{i,j})^2 \ .$$

Then, the chi squared independence test with estimation of parameters is an particular Hilbertian independence test.

In the case where the marginal distribution are known, we have proved that (with classical notations : cf [4])

$$n\sum_{i=1}^{h}\sum_{j=1}^{k}(\rho_{i,j}^{n})^2$$

$$=\sum_{i=0}^{h}\sum_{j=0}^{k}\frac{\big(n_{i,j}-n\mu(E_i)\mu'(F_j)\big)^2}{n\mu(E_i)\mu'(F_j)}-\sum_{i=0}^{h}\frac{\big(n_{i.}-n\mu(E_i)\big)^2}{n\mu(E_i)}-\sum_{j=0}^{k}\frac{\big(n_{.j}-n\mu'(F_j)\big)^2}{n\mu'(F_j)}\ .$$

$$=\chi^2_{X,Y}-\chi^2_X-\chi^2_Y\ .$$

The independence test is the restricted chi squared test. Therefore, the Hilbertian independence test is more powerful than the chi squared test because this one tests again that the marginal distributions are μ and μ' (cf [4]).

References

[1] AKHIEZER N.I. , GLAZMAN I.M. (1966) Theory of linear operators in Hilbert space vol 1. Second printing, Frederick Ungar Publishing Co, New York.

[2] BLACHER R. (1993) Higher Order Correlation Coefficients. Statistics 25, 1-15.

[3] BLACHER R. (1984) Coefficients de corrélation d'ordre supérieur. Statistique et Analyse des Données, 9, n^o2, p 48-67.

[4] BLACHER R. (1988) A new form for the chi squared independence test. Statistics 19 519-536.

[5] BLACHER R. (1989) Loi asymptotique des estimateurs des moindres carrés en régression linéaire. Compte rendu de l'académie des sciences de Paris. t 309, série I, p 629-632.

[6] BLACHER R. (1990) Quelques applications des fonctions orthogonales en probabilité et statistique. Thèse de l'université Joseph Fourier. Grenoble.

[7] BLACHER R. (1995) Central limit theorem by polynomial dependence coefficients. Journal of computational and applied mathematics 57, 45-56.

[8] BLACHER R. (1983) Indicateurs de dépendance fournis par le développement en série de la densité de probabilité. Thèse 3 Cycle, Université Joseph Fourier de Grenoble. Hal archives ouvertes : http://tel.archives-ouvertes.fr

[9] BLACHER R. (1984) Coefficient de correlation d'ordre (i,j) et variances d'ordre i. Rapport de recherche 463, Laboratoire TIM3 IMAG, Université de Grenoble.

[10] BLACHER R. (1984) Coefficient de correlation d'ordre (i,j) et variances d'ordre i. Symposium International de Bar Le Duc, Polynomes orthogonaux et applications, Springer Verlag, 1171, 475-486.

[11] BLACHER R. (1987) Divers indice de dépendance fournis par les fonctions orthogonales. Statistique et Analyse des donnés, vol 12, n^o 1-2, p 1-27.

[12] BLACHER R. (2003) Multivariate quadratic forms of random vectors. Journal of multivariate analysis 87, 2-23.

[13] BLACHER R. (2012) Higher Order variance and Gauss Jacobi Quadrature http://hal.archives-ouvertes.fr/hal-00708152/

[14] BOSQ.D (1983) Lois limites et efficacité asymptotique des tests Hilbertiens de diverses lois sous des hypothèses adjacentes. Statistique et Analyse des Données, 8 n^o1, p 1-40.

[15] CSAKI P. , FISHER J. (1963) On the general notion of maximal correlation. Magyar tud. Akad. Mat. Kuato int. Kozl, 8, 27-51.

[16] DAUXOIS J. POUSSE A. (1975) Une extension de l'analyse canonique. Quelques applications. Ann. Inst. Henri Poincaré, vol XI , $n^o 4$, p 355-379.

[17] LANCASTER H. O. (1960) Orthogonal models for contingency tables. Developments in statistics. Academic Press, New York.

[18] NATANSON I.P. (1965) Constructive function theory, vol 2, Approximation in mean. Frederick ungar Publishing Co, New York

[19] SERFLING R.J. (1980) Approximation theorems of mathematical statistics. Wiley, New York

[20] SMIRNOV (1961) Linear algebra and group theory . McGraw-Hill Company Inc, New York, The Maple Press Company, York

[21] WEIDMANN J. (1980) Linear operators in Hilbert space. Springer-Verlag. New York.

[22] SCHWEIZER B. WOLF E.F. (1981) On non parametric measures of dependence for random variables. The annals of statistics, vol 9 $n^o 4$, 879-885.

[23] CHIHARA. (1978) An introduction to orthogonal polynomials Gordon and Breach, New York

[24] FREUD Geza. (1971) Orthogonal polynomials. Pergamon Press

[25] HIDA T. (1980) Brownian motion. Springer, Berlin

[26] SZEGO G. (1959) Orthogonal polynomials. Americal Mathematical Society, Providence.