

STATISTIQUE ET ANALYSE DES DONNÉES

RENÉ BLACHER

Divers indices de dépendance fournis par les fonctions orthogonales

Statistique et analyse des données, tome 12, n° 1-2 (1987), p. 1-27

http://www.numdam.org/item?id=SAD_1987__12_1-2_1_0

© Association pour la statistique et ses utilisations, 1987, tous droits réservés.

L'accès aux archives de la revue « Statistique et analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

DIVERS INDICES DE DEPENDANCE FOURNIS PAR LES FONCTIONS ORTHOGONALES

René BLACHER

TIM 3 - IMAG. Université Scientifique et Médicale de Grenoble
BP. 68, 38042 Saint Martin d'Hères Cedex.

Résumé : Dans cet article, nous étudions le rapport entre les coefficients du développement en série de la densité de probabilité d'un vecteur aléatoire à valeurs dans $[0,1]^2$, de lois marginales uniformes avec la dépendance mesurée par les normes L^h classiques. Nous étudions le cas de développements en série de polynômes et en série de fonctions indicatrices d'ensemble. Nous appliquons ces résultats au test Hilbertien d'indépendance, ce qui permettra de préciser quelle "indépendance" il teste réellement. Un résultat particulier se rapportera au choix du nombre de classes dans le test du chi-deux.

Summary : In this paper, we study the connection between the dependence and the coefficients of the series expansion of the probability density of a random vector which takes values in $[0,1]^2$ with uniform marginal distributions. We use these results for the Hilbertian independence test what allows us to precise the "independence" which is really tested. We study the particular case of the chi-squared test.

Mots-clefs : Test Hilbertien, fonctions orthogonales, indépendance.

Indice de classification ISI 06. 010

INTRODUCTION

Lorsque l'on veut tester l'indépendance de deux variables aléatoires X et Y , on a très souvent recours au classique test du chi-deux. Dans ce cas, on doit alors choisir le nombre de classes avec lequel on compte opérer ce test. Or, il arrive parfois que ce choix soit mal effectué et que, par voie de conséquence, l'on tire des conclusions fausses d'un test pourtant correctement réalisé.

Or le test du chi-deux est un test Hilbertien particulier. Celui-ci, rappelons-le, consiste à tester un certain nombre de coefficients du développement en série de fonctions orthogonales de la densité de probabilité d'un vecteur aléatoire (X, Y) . Et le problème qui se pose alors est de savoir combien de ces coefficients on doit tester : dans le cas particulier du test du chi-deux, ce choix est celui du nombre de classes.

En fait, pour résoudre ce problème il faut savoir quelle "indépendance" on veut tester. Nous aurons donc besoin d'indicateurs de dépendance. Nous choisirons ceux fournis par les normes L^h classiques des fonctions de répartition. Nous les étudierons dans une première partie et nous verrons que ce sont de "bons" indicateurs de dépendance.

Dans une seconde partie, nous étudierons le problème du choix du nombre de coefficients à tester. Nous nous placerons dans le cas particulier d'un test d'indépendance de deux variables aléatoires de loi uniforme sur $[0,1]$. On sait d'ailleurs que l'on peut se ramener à cette hypothèse dès que les lois de X et Y ont des fonctions de répartition continues.

Nous mènerons cette étude en choisissant comme bases orthonormales, soit la base des polynômes de Legendre, soit la base des fonctions de Walsh, celles-ci étant des combinaisons linéaires de fonctions indicatrices d'ensemble. C'est dans ce dernier cas que l'on retrouve le fait que le test du chi-deux est un test Hilbertien : les deux tests sont alors identiques.

Enfin, dans une dernière partie, nous tirerons les conclusions des résultats obtenus et nous donnerons quelques conséquences plus pratiques, en particulier sur le choix du nombre de classes dans le test du chi-deux.

Signalons que cette étude a été menée de manière approfondie dans [2], [3] et [5] où l'on retrouve les démonstrations détaillées de tous les résultats donnés ici, ainsi que l'étude du cas de la base orthogonale des fonctions trigonométriques et des tests d'uniformité.

NOTATIONS

Nous noterons (Ω, \mathcal{A}, P) un espace probabilisé et V^U l'ensemble des vecteurs aléatoires, de lois marginales uniformes, applications de (Ω, \mathcal{A}, P) dans $[0,1]^2$.

En général lorsque nous aurons un élément (X, Y) de V^U , nous noterons par Q sa probabilité image, par F sa fonction de répartition et par f sa densité de probabilité éventuelle.

De plus, pour tout couple d'applications de $[0,1]$ dans \mathbb{R} , (g_1, g_2) , nous noterons par $g_1 \times g_2$ l'application de $[0,1]^2$ dans $\mathbb{R} : \forall (x_1, x_2) \in [0,1]^2, g_1 \times g_2(x_1, x_2) = g_1(x_1) \cdot g_2(x_2)$, et par U la fonction de répartition d'un vecteur aléatoire de loi uniforme : $\forall (x, y) \in [0,1]^2, U(x, y) = xy$.

Nous noterons par $\delta_h, h \in \mathbb{N}^* \cup \{+\infty\}$, les distances normées L^h classiques sur les éléments de $L^h([0,1]^2)$:

$$\delta_h(g, g') = \|g - g'\|_{L^h([0,1]^2)} \quad \text{avec la convention}$$

$$\delta_\infty(g, g') = \delta(g, g') = \sup_{(x,y) \in [0,1]^2} |g(x,y) - g'(x,y)| .$$

Ces distances nous fournissent alors toute une suite d'indices de dépendance entre X et Y que nous noterons : $D_h(X, Y) = \delta_h(F, U)$, $D(X, Y) = \delta(F, U)$, pour $h \in \mathbb{N}^* \cup \{+\infty\}$. Nous noterons alors par $D_h^N(X, Y)$ ces indices normalisés (1-1).

Nous noterons par \mathbb{N}^{*2} l'ensemble $\mathbb{N}^{*2} = \mathbb{N}^* \times \mathbb{N}^*$.

De plus, si X est une variable aléatoire, alors $E(X)$ représentera son espérance mathématique.

Enfin les signes $<$, $>$ et $\| \cdot \|$ désigneront le produit scalaire et la norme sur les espaces L^2 .

I: INDICATEURS DE DÉPENDANCE $D_h^N(\cdot, \cdot)$

Ces indices de dépendance ont été étudiés par Schweizer et Wolf dans leur article de 1981 [16]. Nous allons rappeler le résultat principal de celui-ci : ce sont de bons indices de dépendance car ils vérifient les axiomes de Renyi. Nous donnerons aussi quelques nouveaux théorèmes et nous terminerons par des exemples où nous verrons que ces indicateurs ont aussi de bonnes propriétés du point de vue statistique. Nous voulons dire par là, que l'estimation que l'on pourra en donner, à partir d'un échantillon sera a priori satisfaisante.

Rappelons aussi que les distances $\delta_h(F, U)$ ne définissent pas des indicateurs de dépendance que dans le seul cas de V^U . On sait en effet s'y ramener pour tout vecteur aléatoire (X, Y) dont les fonctions de répartition des lois marginales F_1 et F_2 sont continues. Dans ce cas, en effet, $(F_1(X), F_2(Y)) \in V^U$. Et, en fait, ce sont les indicateurs $D_h^N(F_1(X), F_2(Y))$ qui vérifient les axiomes de Renyi.

Nous commençons cette étude par la normalisation.

Théorème 1-1 : Soit $(X, Y) \in V^U$; si nous posons

$$D_h^N(X, Y) = \left[\frac{(h+1)[(2h+1)!]}{(h!)^2} \right]^{\frac{1}{h}} \cdot D_h(X, Y) \quad \text{pour } h \in \mathbb{N}^* ,$$

$$\text{et } D^N(X, Y) = D^N_{\infty}(X, Y) = 4.D(X, Y),$$

alors pour tout $h \in \mathbb{N}^* \cup \{+\infty\}$, on a l'inégalité $D^N_h(X, Y) \leq 1$.

De plus, lorsque $h \in \mathbb{N}^*$, l'égalité est réalisée si et seulement si $X=Y$ p.s. ou $X = 1-Y$ p.s. . Enfin, lorsque $h = +\infty$, l'égalité est réalisée si $X=Y$ p.s. ou $X=1-Y$ p.s. .

Démonstration : Il suffit d'effectuer le calcul de $\delta_h(F, U)$ dans le cas où F est la fonction de répartition de (X, Y) avec $X = Y$ ou $X = 1-Y$, et de se reporter aux axiomes 3 et 5 de Renyi (cf 1-3, 1-4 ou p 882, 883 [16]) et au théorème 2 page 881 de [16]. Signalons que ce calcul est effectué en détail dans la partie C-2 de [2].

Rappelons que ces coefficients ne sont pas indépendants les uns des autres.

Théorème 1-2 : Gardons les notations et hypothèses précédentes. Alors

$$\frac{2}{3} (D(X, Y))^3 \leq D_1(X, Y) \leq D_2(X, Y) \leq D_3(X, Y) \leq \dots \leq D(X, Y).$$

L'inégalité de gauche est une conséquence de la condition de Lipschitz que doivent vérifier les fonctions de répartition des éléments de V^U (cf partie C-5 [2], p 880, 881 [16]). Les autres inégalités sont classiques (th 5, p 214 [12]).

Nous rappelons maintenant les axiomes que Renyi a introduits en 1959 et qu'il estimait nécessaires à de bonnes mesures de dépendance (cf [15], [16]).

Axiomes de Renyi 1-3 : soit $R(.,.)$ un indicateur de la dépendance entre deux variables aléatoires. On dit que $R(.,.)$ vérifie les axiomes de Renyi, si pour tout couple (X, Y) de variables aléatoires de fonctions de répartition marginale continues (cf p 882 [16]), les conditions suivantes sont vérifiées.

- 1) $R(X, Y)$ est défini.
- 2) $R(X, Y) = R(Y, X)$.
- 3) $0 \leq R(X, Y) \leq 1$.
- 4) $R(X, Y) = 0$ si et seulement si X et Y sont indépendantes.
- 5) $R(X, Y) = 1$ si et seulement si X est presque sûrement une fonction strictement monotone de Y et si Y est presque sûrement une fonction strictement monotone de X .
- 6) Si f et g sont presque sûrement deux fonctions strictement monotones, alors $R(f(X), g(Y)) = R(X, Y)$.
- 7) Si la loi de (X, Y) est la loi normale alors $R(X, Y)$ est une fonction strictement croissante de $|r|$, où r représente le coefficient de corrélation linéaire de X et Y .
- 8) Soit (X_n, Y_n) , $n \in \mathbb{N}^*$, une suite de vecteurs aléatoires convergente en loi vers un vecteur (X, Y) alors $\lim_{n \rightarrow \infty} R(X_n, Y_n) = R(X, Y)$.

$$n \rightarrow \infty$$

Remarquons que les éléments de V^U vérifient les axiomes de Renyi.

Théorème 1-4 : Soit $D^N_h(F_1(X), F_2(Y))$ l'indicateur de dépendance entre 2 variables aléatoires X et Y défini pour tout vecteur aléatoire (X, Y) à valeurs dans un sous-ensemble de \mathbb{R}^2 dont les fonctions de répartition marginale F_1 et F_2 sont continues. Alors cet indicateur vérifie les axiomes de Renyi.

La démonstration est donnée en page 883 [16] et complétée par le calcul de la normalisation donnée en 1-1 ci-avant.

Une autre propriété intéressante de D_h^N , c'est que si l'on approxime la mesure image de (X,Y) par une probabilité admettant une densité constante sur des rectangles de côtés parallèles aux axes, on obtient une majoration simple de l'erreur commise. Dans cet article, nous donnerons ce résultat en 2-7 dans le cas particulier d'une partition de $[0,1]^2$ en 4^n carrés égaux.

Bien sûr, tous les indicateurs de dépendance existants n'ont pas cette propriété. C'est le cas, par exemple, des indicateurs d_h , obtenus toujours à partir des distances L_h^n , mais distances, non plus des fonctions de répartition, mais cette fois des densités de probabilité.

Aussi, pour mieux faire ressortir les qualités des D_h^N , nous les comparerons aux d_h dans les exemples que nous allons étudier après avoir donné les définitions de ces nouveaux indicateurs.

1-5 Indicateurs d_h : Dans le cas où (X,Y) admet une densité de probabilité $f \in L^h$, nous définissons d_h par $d_h(X,Y) = \delta_h(f, 1_{[0,1]^2})$, où $1_{[0,1]^2}$ représente la fonction indicatrice de $[0,1]^2$.

Autrement, on peut généraliser cette définition de la façon suivante (cf 1-5 [3]).

Nous notons par \mathcal{P} l'ensemble des partitions boréliennes finies de $[0,1]^2$, $\{B_s\}_{s \in S}$, vérifiant $\mu(B_s) > 0$ - μ , mesure de Lebesgue. Soit $\{B_s\} \in \mathcal{P}$, nous noterons par $f_{\{B_s\}}$ la fonction de $[0,1]^2$ dans \mathbb{R} , constante et égale à $P\{(X,Y) \in B_s\} / \mu(B_s)$ sur chaque B_s . Nous définissons alors d_h par $d_h(X,Y) = \sup_{\{B_s\} \in \mathcal{P}} \sup_s \{\delta_h(f_{\{B_s\}}, 1_{[0,1]^2})\}$.

On peut alors prouver (p 3 [2], 1-13, 1-17 [3]) que $\sup(d_h(X,Y)) = 2$ et $\sup(d_h(X,Y)) = +\infty$ si $h \geq 2$.

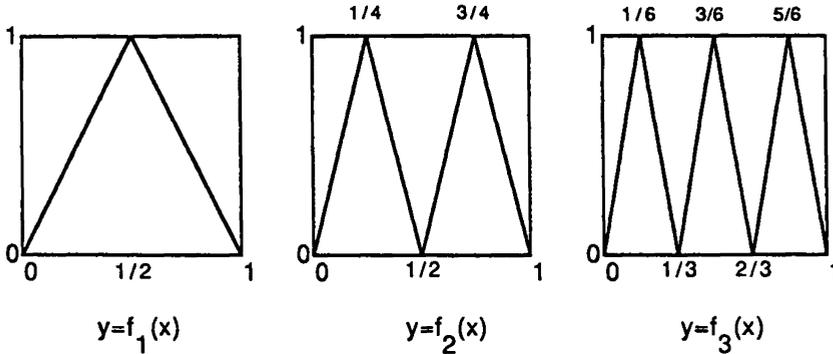
Comme nous nous servirons de ces indicateurs uniquement dans les exemples et seulement pour mieux faire ressortir certaines qualités des D_h^N , nous ne démontrerons pas ces propriétés ici. Remarquons que dans le cas où $\{B_s\}$ est une partition de $[0,1]^2$ en rectangles de côtés parallèles aux axes, et où l'on utilise la probabilité empirique, $d_2(X,Y)$ est alors l'indice classique de dépendance du chi-deux (p 449 [13]).

Nous passons maintenant à l'étude d'exemples.

Exemple 1-6 (cf p 19,20 [2]) : Soit $n \in \mathbb{N}^*$, et soit f_n la fonction définie par $f_n(x) = \overline{2nx}$ s'il existe

$k \in \mathbb{N}$ tel que $x \in [(2k)/(2n), (2k+1)/(2n)[$, et $\overline{-2nx}$ sinon, où pour tout $r \in \mathbb{R}$, \overline{r} désigne

l'unique réel de $[0,1[$ tel que $\overline{r} - r \in \mathbb{Z}$.



Et soit $(X_n, Y_n) \in V^U$ vérifiant $Y_n = f_n(X_n)$. On a alors les égalités suivantes :

$$D_h^N(X_n, Y_n) = \frac{1}{2^n}, \text{ pour tout } h \in \mathbb{N}^* \cup \{+\infty\}, \text{ et pour tout } n \in \mathbb{N}^*.$$

D'autre part, on trouve que le coefficient de corrélation est nul tandis que $d_h(X_n, Y_n) = +\infty$ pour tout $h \geq 2$.

Exemple 1-7 : soit $n \in \mathbb{N}^*$, et soit un entier $j : 1 \leq j \leq 4^n$. Notons par e_j et r_j , le quotient et le reste de la division euclidienne de $j-1$ par 2^n : $j-1 = e_j \cdot 2^n + r_j$. Et soit T_n la bijection de $(1, 2, \dots, 4^n)$ dans lui-même, définie par : $T_n(j) = 2^n \cdot r_j + e_j + 1$.

Notons alors $(X_n, Y_n)_n \in \mathbb{N}^*$ une suite de V^U , définie sur l'espace probabilisé (Ω, \mathcal{A}, P) , et admettant une probabilité concentrée sur les carrés $I_{2^{2n}, j} \times I_{2^{2n}, T_n(j)}$ (cf 2-1), et vérifiant $X_n - j/4^n = Y_n - (T_n(j))/4^n$ sur ces carrés.

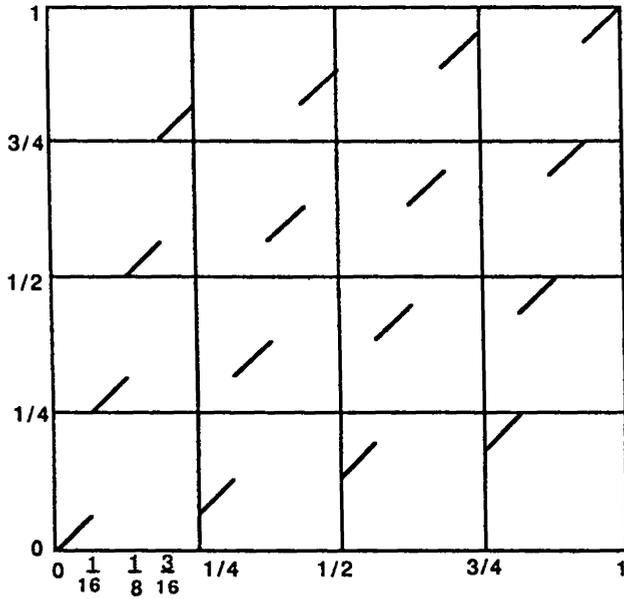
Alors X_n et Y_n sont en bijection, et le théorème 2-7 appliqué aux carrés $I_{2^n, j} \times I_{2^n, j}$ montre que $D^N(X_n, Y_n) \leq 2^{-(n-1)}$.

$$\text{D'autre part, } P((X_n, Y_n) \in I_{2^n, j} \times I_{2^n, j}) - (P(X_n \in I_{2^n, j})) (P(Y_n \in I_{2^n, j})) = 0.$$

Or quelle que soit la méthode avec laquelle nous effectuons nos calculs - à la main ou par ordinateur - ceux-ci sont effectués à 2^{-t} près (ou 10^{-t} près), ce qui veut dire que la famille des boréliens sur $[0, 1]$ dont nous pouvons disposer est la famille finie réunion de tous les intervalles de bornes $k/2^t$ et $(k+1)/2^t$, avec k entier.

Et donc en choisissant $n=t$, on voit que l'on trouvera, étant données les erreurs d'arrondis de nos calculs, pour tout borélien $B_1 \times B_2$:

$$P((X_t, Y_t) \in B_1 \times B_2) = (P(X_t \in B_1)) (P(Y_t \in B_2)) .$$



(X_2, Y_2)

Finalement nous possédons donc deux variables aléatoires X_t et Y_t , dépendantes l'une de l'autre par un bijection et qui pourtant, à cause des arrondis de nos calculs, vérifieront toujours l'égalité de la définition de l'indépendance.

D'autre part, dans ce cas $D_h^N(X_t, Y_t) \approx 0$,

$d_1(X_t, Y_t) = 2$ et $d_h(X_t, Y_t) = +\infty$ si $h \geq 2$.

Or, si nous cherchons à connaître la valeur de ces indicateurs par estimation, nous trouverons probablement, vu l'égalité d'indépendance, que

$D_h^N(X_t, Y_t) = d_h(X_t, Y_t) = 0$.

Donc, pratiquement, l'estimation que nous ferons de l'indice D_h^N sera vraisemblablement raisonnable alors que celle de d_h sera sans doute totalement erronée.

II : RAPPORT ENTRE LA DEPENDANCE ET LES COEFFICIENTS DU DEVELOPPEMENT EN SERIE DE LA DENSITE DE PROBABILITE

Dans cette partie, nous étudions dans le cas de vecteurs $(X,Y) \in V^U$ les rapports entre la dépendance mesurée par les indicateurs D_h^N et les coefficients $a_{i,j}$ du développement en série de la densité de probabilité f par rapport à une base orthonormale $(G_i \times G_j)$:

$$f(x,y) = 1 + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{i,j} G_i(x) G_j(y) \text{ dans } L^2([0,1]^2).$$

En fait, l'étude menée ici ne nécessite pas l'existence d'une densité, mais cela permet d'introduire de façon simple les coefficients $a_{i,j}$.

Nous étudierons le cas où la famille (G_i) est, soit la famille orthonormale des polynômes de Legendre, soit la famille orthonormale des fonctions de Walsh qui sont des combinaisons linéaires de fonctions indicatrices d'ensemble. Et nous allons trouver que ce sont les coefficients $a_{i,j}$ des termes de plus bas degré qui sont les plus importants pour la dépendance, ce qui est un résultat tout à fait naturel. En effet, $a_{1,1}$ sera le coefficient le plus important de la suite $(a_{i,j})$. Or, dans le cas de polynômes de Legendre, $a_{1,1}$ n'est autre que le coefficient de corrélation linéaire. Nous trouvons donc bien un résultat conforme aux usages habituels.

PARTIE II-1 : ASPECT PROBABILISTE

II-1-1 : Cas des fonctions de Walsh

Ces fonctions sont des combinaisons linéaires de fonctions indicatrices de carrés de $[0,1]^2$. Nous les appelons un peu abusivement fonctions de Walsh et l'on pourrait d'ailleurs aussi bien les appeler fonctions de Rademacher (p 6 [1]). En fait, il s'agit d'une autre méthode que celle employée par Walsh [1] pour utiliser les fonctions indicatrices d'ensemble. Mais nous en rejoignons le principe fondamental en imposant à ces fonctions de former une base orthogonale.

Une autre caractéristique de ces fonctions sera que l'approximation d'une probabilité Q de (X,Y) par une probabilité admettant une densité f_n :

$$f_n(x,y) = 1 + \sum_{i=1}^n \sum_{j=1}^n a_{i,j} G_i(x) \cdot G_j(y) \quad ,$$

sera en fait l'approximation classique de Q par la probabilité Q_n admettant une densité constante sur chaque élément de la partition de $[0,1]^2$ en 4^{n+1} carrés égaux. C'est d'ailleurs cette propriété qui permet de prouver l'égalité entre test Hilbertien et test du chi-deux.

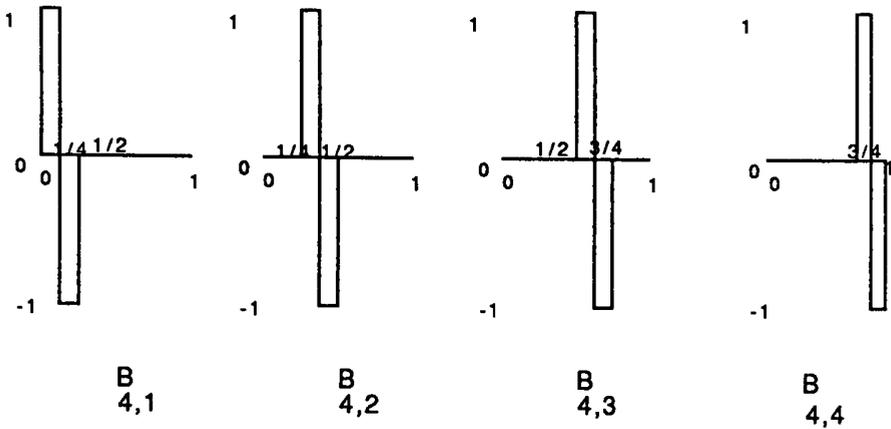
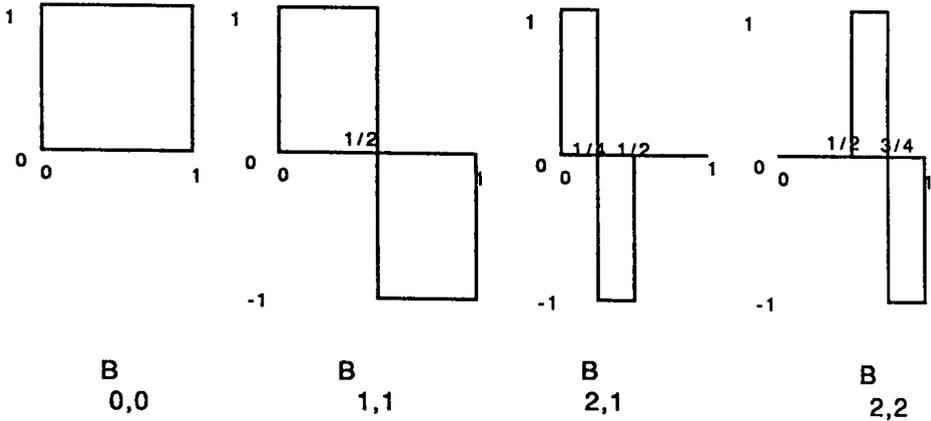
Le résultat que nous obtiendrons alors sera une majoration de l'erreur commise sur $D^N(X,Y)$ si l'on approxime Q par Q_n .

Notations 2-1 : Notons par : G, G_k, G^* et G_k^* , $k \in \mathbb{N}$, les sous-ensembles de \mathbb{N}^2 définis par :

$$G^* = \{ (2^i, j) \in \mathbb{N}^2 \mid i \in \mathbb{N}, j \leq 2^i \} , \quad G_k^* = \{ (2^i, j) \in G^* \mid i \leq k \} ,$$

$$G = G^* \cup \{ (0,0) \} \quad \text{et} \quad G_k = G_k^* \cup \{ (0,0) \} .$$

Alors pour tout $(2^i, j) \in G^*$, nous notons par $I_{2^i, j}$ l'intervalle : $[(j-1)/2^i, j/2^i[$,



et nous définissons les fonctions réelles $B_{0,0}(x) \equiv 1$ et

$$B_{2^i j}, \text{ par : } B_{2^i j}(x) = \begin{cases} 1, & \text{si } x \in I_{2^{i+1}, 2j-1} \\ -1, & \text{si } x \in I_{2^{i+1}, 2j} \\ 0, & \text{sinon.} \end{cases}$$

Nous notons alors par $A_{2^i j}$, les fonctions $B_{2^i j}$ normalisées : $A_{2^i j} = 2^{(i/2)} B_{2^i j}$.

Nous pouvons tout de suite énoncer le théorème suivant.

Théorème 2-2 : la famille $\{A_I \times A_J\}_{(I,J) \in G^2}$ est une base orthonormale de $L^2([0,1]^2)$.

C'est une conséquence simple du fait que les fonctions $A_I \times A_J$ sont engendrées par les fonctions indicatrices de carrés de $[0,1]^2$.

On en déduit l'écriture de la densité.

Théorème 2-3 : soit $(X,Y) \in V^U$ admettant une densité de probabilité, $f \in L^2([0,1]^2)$. Alors, il existe une suite $\{\alpha_{I,J}\}_{(I,J) \in (G^*)^2}$ telle que

$$f = 1 + \sum_{(I,J) \in (G^*)^2} \alpha_{I,J} A_I \times A_J \quad \text{dans } L^2([0,1]^2).$$

Nous obtiendrons alors une approximation de f en limitant l'ordre du développement.

Notations 2-4 : Gardons les hypothèses de 2-3, et soit $n \in \mathbb{N}^*$. Nous notons par f_n la fonction définie par :

$$f_n = 1 + \sum_{(I,J) \in (G_n^*)^2} \alpha_{I,J} A_I \times A_J,$$

$$\text{et par } F_n \text{ sa primitive : } F_n(x,y) = \int_0^x \int_0^y f_n(u,v) \, du \, dv.$$

Nous allons voir que la fonction f_n est la densité d'une probabilité Q_n constante sur chaque élément de la partition de $[0,1]^2$ en 4^{n+1} carrés égaux, et donc, on retrouve ainsi une approximation connue de la probabilité Q de (X,Y) .

D'autre part, on sait qu'une telle approximation est définie même si il n'y a pas de densité.

Généralisation des notations 2-5 : Soit $(X, Y) \in \mathcal{V}^U$, de loi Q sur $[0, 1]^2$, et soit $n \in \mathbb{N}$. Nous notons par $\alpha_{I, J}$, $(I, J) \in G^2$, le réel $\alpha_{I, J} = E(A_I(X).A_J(Y))$, et nous notons toujours par f_n la fonction :

$$f_n = 1 + \sum_{(I, J) \in (G^*_n)^2} \alpha_{I, J} A_I \times A_J ,$$

et par F_n sa primitive.

Donc, même si (X, Y) n'admet pas de densité, F_n , elle, est toujours définie comme la primitive de f_n . De plus, Q_n , la probabilité associée est bien l'approximation de Q dont nous parlions plus haut.

Théorème 2-6 : Nous gardons les notations de 2-5. Alors F_n est la fonction de répartition d'une probabilité Q_n sur $[0, 1]^2$ de lois marginales uniformes, admettant f_n pour densité.

$$\begin{aligned} \text{De plus, sur tout carré } C_{j, j'} = I_{2^{n+1}, j} \times I_{2^{n+1}, j'} \subset [0, 1]^2 , \\ f_n \text{ est constante, non négative, et on a l'égalité } Q_n(C_{j, j'}) = Q(C_{j, j'}) . \end{aligned}$$

La démonstration s'effectue en remarquant que les fonctions indicatrices des intervalles $I_{2^{n+1}, j}$, $j=1, 2, \dots, 2^{n+1}$, et les fonctions de Walsh A_I , $I \in G_n$, engendrent le même sous-espace de $L^2([0, 1])$.

Nous allons maintenant majorer l'erreur commise sur $D^N(X, Y)$ si l'on approxime Q par Q_n .

Théorème 2-7 : Nous conservons les notations de 2-5, et nous notons toujours par F la fonction de répartition de (X, Y) , alors on a l'inégalité $\delta(F, F_n) \leq 2^{-(n+2)}$.

En particulier, si pour tout $(I, J) \in (G^*_n)^2$, on a $\alpha_{I, J} = 0$, on obtient : $D^N(X, Y) \leq 2^{-n}$.

Pour démontrer cette propriété, on remarque d'abord que F et F_n coïncident en chaque sommet de tout carré $C_{j, j'}$ de 2-6. Ensuite, on voit que la condition de Lipschitz que vérifient toutes fonctions de répartition F' et F'' de lois marginales uniformes implique que, si celles-ci coïncident en deux points (x_1, y_0) et (x_2, y_0) , et si, de plus, F' vérifie $F'(x, y_0) = ax + b$ pour tout $x \in [x_1, x_2]$, alors,

$$|F'(x, y_0) - F''(x, y_0)| \leq \frac{1}{4} \cdot (x_2 - x_1), \text{ pour tout } x \in [x_1, x_2].$$

Cette inégalité reste vraie si l'on permute le rôle des x et des y . En l'appliquant alors plusieurs fois dans les carrés $C_{j, j'}$, on trouve la majoration annoncée.

Remarque 2-8 : Signalons que la majoration donnée par le théorème précédent est assez exacte: il est facile (p 18 [2]) de construire des suites $(X_n, Y_n) \in \mathcal{V}^U$ vérifiant

$$D^N(X_n, Y_n) \geq 2^{-n} - 3 \times 4^{-(n+1)}.$$

Remarque 2-9 : Signalons que nous avons également des résultats sur la majoration de la distance à l'uniformité $\delta_1(F, U)$ lorsque l'on ne suppose plus que les lois de X et Y sont uniformes (cf 5-8, 5-12, 5-16 ; [3]).

II-1-2 : Cas des polynômes de Legendre

Nous étudions maintenant le cas où la base $(G_i)_{i \in \mathbb{N}}$ est la famille orthonormale des polynômes de Legendre et nous allons obtenir des résultats analogues à ceux que nous venons de trouver avec les fonctions de Walsh, bien qu'un peu moins fins.

D'autre part, nous allons retrouver le fait que, dans ce cas, les coefficients $a_{i,j}$ du développement en série de la densité sont les coefficients de corrélation d'ordre (i,j) . En particulier, $a_{1,1}$ n'est autre que le coefficient de corrélation linéaire.

Rappelons (p 86 [14]) d'abord la définition classique.

Définition 3-1 : On appelle polynômes de Legendre, les polynômes $L_n, n \in \mathbb{N}$, définis par :

$$L_n(z) = ((n!) \cdot 2^n)^{-1} \cdot \frac{d^n (z^2 - 1)^n}{d z^n}.$$

Cependant, par abus de langage, nous admettrons également la définition suivante.

Définition 3-2 : Nous appellerons aussi polynômes de Legendre, les polynômes $P_i, i \in \mathbb{N}$, et $P_i \times P_j, (i,j) \in \mathbb{N}^2$, définis par :

$$P_i(x) = \sqrt{2i + 1} L_i(2x-1).$$

Comme dans le cas des fonctions de Walsh, nous pouvons énoncer les théorèmes suivants.

Théorème 3-3 : Les polynômes $\{P_i\}_{i \in \mathbb{N}}$ (resp. $\{P_i \times P_j\}_{(i,j) \in \mathbb{N}^2}$) forment une base orthonormale de $L^2([0,1])$ (resp. $L^2([0,1]^2)$).

C'est une conséquence bien connue de la définition des polynômes de Legendre et du théorème de Stone-Weierstrass.

Proposition 3-4 : Soit $(X, Y) \in V^U$, et admettant une densité de probabilité $f \in L^2([0,1]^2)$. Alors, il existe une suite de réels $\{\theta_{i,j}\}_{(i,j) \in \mathbb{N}^{*2}}$ telle que

$$f = 1 + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \theta_{i,j} P_i \times P_j \quad \text{dans } L^2([0,1]^2).$$

Comme précédemment, nous pouvons nous passer de l'existence d'une densité.

3-5 Notation : Soit $(i,j) \in \mathbb{N}^{*2}$. Nous notons par $\theta_{i,j}$ le nombre réel $\theta_{i,j} = E(P_i(X).P_j(Y))$.

Rappelons que l'on a démontré le théorème suivant (4-8 [5], 2-1 [6]).

Théorème 3-6 : Soit $(i,j) \in \mathbb{N}^{*2}$. Alors $\theta_{i,j}$ est le coefficient de corrélation polynômiale d'ordre (i,j) de X et Y. En particulier, $\theta_{1,1} = \rho(X,Y)$, le coefficient de corrélation linéaire.

Et, ce que nous allons étudier, c'est la qualité de l'approximation de la probabilité Q de (X,Y) par la mesure (pas forcément positive : cf 3-24, [3]) admettant la densité :

$$f_n(x,y) = 1 + \sum_{i=1}^n \sum_{j=1}^n \theta_{i,j} P_i(x) P_j(y).$$

Pour cela, nous aurons besoin des coefficients du développement de F en série. Ceux-ci sont liés aux $\theta_{i,j}$ par la relation suivante.

Proposition 3-7 : Gardons les notations et hypothèses précédentes : en particulier, nous notons toujours par F la fonction de répartition de $(X,Y) \in V^U$.

Nous posons alors, pour tout $(i,j) \in \mathbb{Z}^2$:

$$\overline{\theta}_{i,j} = \begin{cases} \theta_{i,j}, & \text{si } i \geq 1 \text{ et } j \geq 1 \\ 0 & \text{sinon,} \end{cases}$$

et

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

Alors, on a l'égalité

$$F - U = \sum_{(i,j) \in \mathbb{N}^2} \beta_{i,j} P_i \times P_j \text{ dans } L^2([0,1]^2).$$

Démonstration : Pour effectuer cette démonstration nous aurons besoin du lemme suivant.

Lemme 3-8 : Pour tout $n \in \mathbb{N}^*$, on a l'égalité :

$$\int_0^x P_n(t) dt = \frac{1}{2\sqrt{2n+1}} \left[\frac{P_{n+1}(x)}{\sqrt{2n+3}} - \frac{P_{n-1}(x)}{\sqrt{2n-1}} \right].$$

Démonstration du lemme : On peut écrire la dérivée L'_n de L_n sous la forme :

$$L'_n(x) = \sum_{i \leq n-1} a_i L_i(x).$$

L'orthogonalité des polynômes permet alors d'affirmer que :

$$a_i = \langle L'_n, \frac{L_i}{\|L_i\|} \rangle \cdot \frac{1}{\|L_i\|}.$$

Rappelons aussi que

$$\int_{-1}^{+1} L'_n(x) \cdot L_i(x) dx = [L_n(x) \cdot L_i(x)]_{-1}^{+1} - \int_{-1}^{+1} L_n(x) \cdot L'_i(x) dx.$$

Comme nous supposons $i < n$, L'_i sera combinaison linéaire des polynômes L_m , $m < n$, et donc

$$\int_{-1}^{+1} L_n(x) \cdot L'_i(x) dx = 0.$$

D'autre part, on sait que $L_n(1)=1$ et $L_n(-1)=(-1)^n$ (p 87 : 128, [14]).

Donc $[L_n(x) \cdot L_i(x)]_{-1}^{+1} = 0$ si $(n+i)$ pair, et 2 sinon.

Enfin, de l'égalité $\|L_i\|^2 = \frac{2i+1}{2}$, on déduira

$$L'_n(x) = \sum_{i \leq n-1 \text{ et } (i+n) \text{ impair}} (2i+1) L_i(x).$$

Pour démontrer le lemme, il suffit alors de dériver chacun des termes de l'égalité annoncée.

Démontrons alors la proposition 3-7. Remarquons que F-U est bornée, et donc que $F-U \in L^2([0,1]^2)$.

Ensuite, nous supposons d'abord que (X,Y) admet une densité de probabilité $f \in L^2([0,1]^2)$. Alors on montre que l'on peut écrire dans $L^2([0,1]^2)$ muni de la topologie de la convergence uniforme :

$$F(x,y) - xy = \sum_{(i,j) \in \mathbb{N}^{*2}} \theta_{i,j} \left(\int_0^x P_i(t) dt \right) \left(\int_0^y P_j(t') dt' \right) =$$

$$\sum_{(i,j) \in \mathbb{N}^{*2}} \frac{\theta_{i,j}}{4\sqrt{(2i+1)(2j+1)}} \left(\frac{P_{i+1}(x)}{\sqrt{2i+3}} - \frac{P_{i-1}(x)}{\sqrt{2i-1}} \right) \left(\frac{P_{j+1}(y)}{\sqrt{2j+3}} - \frac{P_{j-1}(y)}{\sqrt{2j-1}} \right)$$

On en déduit que, pour tout $(i,j) \in \mathbb{N}^{*2}$, on a l'égalité $\langle P_i \times P_j, (F-U) \rangle = \beta_{i,j}$.

Démontrons maintenant le cas où l'on n'a plus l'hypothèse de l'existence d'une densité, et notons par F_n le développement de F en fonction de Walsh jusqu'à l'ordre n (2-5) et par Q_n la probabilité associée sur $[0,1]^2$. Soient $\theta^n_{i,j}$ et $\beta^n_{i,j}$ les notations de $\theta_{i,j}$ et $\beta_{i,j}$ lorsque l'on remplace F par F_n . Alors, il est clair que Q_n converge étroitement vers Q (2-7), et donc que $\theta^n_{i,j}$ converge vers $\theta_{i,j}$, et $\beta^n_{i,j}$ vers $\beta_{i,j}$.

Comme d'autre part, Q_n admet une densité f_n et que $\delta(F,F_n)$ tend vers 0, alors

$$\lim_{n \rightarrow \infty} \beta^n_{i,j} = \langle F-U, P_i \times P_j \rangle,$$

ce qui achève la démonstration.

Nous sommes alors en mesure de majorer la dépendance fournie par les $\theta_{i,j}$, ou ce qui revient au même par les $\beta_{i,j}$.

Théorème 3-9 : Soit $\{M_n\}_{n \in \mathbb{N}}$ la suite :

$$M_0 = \frac{1}{12}, M_1 = \frac{1}{60} \text{ et } M_n = \frac{1}{2(2n+3)(2n-1)} \text{ pour } n \geq 2,$$

alors, pour tout $(i,j) \in \mathbb{N}^2$, on a l'inégalité $(\beta_{i,j})^2 \leq M_i M_j$.

Démonstration : Notons par S_i le polynôme

$$S_i = \frac{P_i}{\sqrt{2i+1}}$$

Supposons par exemple, i et $j \geq 2$. Alors

$$\begin{aligned} & \frac{\theta_{i+1,j+1}}{\sqrt{(2i+3)(2j+3)}} + \frac{\theta_{i-1,j-1}}{\sqrt{(2i-1)(2j-1)}} - \frac{\theta_{i+1,j-1}}{\sqrt{(2i+3)(2j-1)}} - \frac{\theta_{i-1,j+1}}{\sqrt{(2i-1)(2j+3)}} \\ &= E((S_{i+1} - S_{i-1})(X) \cdot (S_{j+1} - S_{j-1})(Y)) \\ &\leq \sqrt{E\{([S_{i+1} - S_{i-1}](X))^2\} \cdot E\{([S_{j+1} - S_{j-1}](Y))^2\}} \end{aligned}$$

d'après l'inégalité de Schwartz. On en déduit les inégalités annoncées.

Des résultats précédents, on déduira facilement la majoration de $D_2^N(X, Y)$ suivante.

Proposition 3-10 : Soit $q \in \mathbb{N}$, $q \geq 2$. Nous notons par E_q le sous-ensemble de \mathbb{N}^{*2} :

$E_q = \{(i, j) \mid i \leq q \text{ et } j \leq q\}$. Supposons que pour tout $(i, j) \in E_q$, $\theta_{i,j} = 0$. Alors on a les inégalités :

$$\begin{aligned} D_2^N(X, Y) &\leq \sqrt{180 \left(\sum_{n=0}^{\infty} M_n \right) \left(\sum_{n=q}^{\infty} M_n \right)} \leq \sqrt{\frac{45}{4} \left(\sum_{n=0}^{\infty} M_n \right) \cdot \text{Log} \left(\frac{2q+1}{2q-3} \right)} \\ &\leq 1,372 \sqrt{\text{Log} \left(\frac{2q+1}{2q-3} \right)} \leq \frac{2,744}{\sqrt{2q-3}} \end{aligned}$$

II-I-3 Cas des fonctions trigonométriques

Signalons que pour le cas des développements en série de Fourier (sinus et cosinus), nous avons obtenu (cf partie II [3], Partie II-2-2 [5]) des résultats tout à fait semblables.

PARTIE II-2 : ASPECT STATISTIQUE

Grâce aux résultats précédents, nous allons maintenant pouvoir étudier les tests d'indépendance de deux variables aléatoires X et Y , $(X, Y) \in \mathcal{V}^U$. Pour cela, nous nous servirons du test Hilbertien de Denis BOSQ. Nous rappellerons d'abord la définition et les propriétés de ce test. Puis nous en déduirons une forme théorique du test d'indépendance. Enfin, nous donnerons quelques résultats numériques.

II-2-1 Test Hilbertien

Nous allons rappeler la définition du test Hilbertien dans les cas qui nous intéressent. Bien sûr, le test Hilbertien est défini sous des conditions beaucoup plus générales et a de nombreuses autres propriétés que celles citées ici. Pour celles-ci, ainsi que pour les démonstrations, nous renvoyons à [9], [10] et [11].

Hypothèses 4-1 : On supposera que $(X, Y) \in \mathcal{V}^U$. D'autre part, la famille (G_i) représentera toujours, soit la famille des polynômes de Legendre, soit la famille des fonctions de Walsh. De plus, nous posons $a_{i,j} = E(G_i(X) \cdot G_j(Y))$.

Notations 4-2 : Soit $(X_q, Y_q)_{q \in \mathbb{N}^*}$ une suite de vecteurs aléatoires indépendants, de même loi que (X, Y) et définis sur le même espace probabilisé (Ω, \mathcal{A}, P) .

Nous notons par D_n^* un sous-ensemble fini non vide de \mathbb{N}^{*2} ou de $(\mathbb{N}^*)^2$ suivant le cas, et nous posons $\text{card}(D_n^*) = k_n - 1$: pratiquement, nous prendrons $D_n^* = E_n$ ou $(G_n^*)^2$.

Soit alors la statistique $\|S_n\|^2 = n \left(\sum_{(i,j) \in D_n^*} [a_{i,j}^n]^2 \right)$,

$$\text{où } a_{i,j}^n = \frac{1}{n} \left(\sum_{q=1}^n G_i(X_q) G_j(Y_q) \right),$$

est un estimateur convergent presque sûrement de $a_{i,j}$.

Alors, nous pouvons énoncer (cf [9], [10]) les propositions suivantes.

Proposition 4-3 : Si $a_{i,j} = 0$ pour tout $(i,j) \in D_n^*$, alors $\|S_n\|^2$ suit asymptotiquement une loi du khi-deux à $k_n - 1$ degrés de liberté.

Proposition 4-4 : S'il existe $(i,j) \in D_n^*$ tel que $a_{i,j} \neq 0$, alors $\|S_n\|^2$ converge presque sûrement vers $+\infty$.

II-2-2 Application au test d'indépendance

Dans le cas où (X, Y) admet une densité $f \in L^2$, nous avons toujours

$$f = 1 + \sum a_{i,j} G_i \times G_j.$$

et donc, lorsque X et Y sont indépendantes tous les $a_{i,j}$ sont nuls. Donc, le test Hilbertien qui teste la nullité d'un nombre fini de $a_{i,j}$ va nous permettre de tester l'indépendance de X et Y .

Aussi nous introduisons les notations suivantes.

Notations 4-5 : Soient $H_0 (D^*_n)$ et $H_1 (D^*_n)$ les hypothèses :

$$H_0 (D^*_n) : \langle\langle \forall (i,j) \in D^*_n, a_{i,j} = 0 \rangle\rangle \quad \text{et} \quad H_1 (D^*_n) : \langle\langle \exists (i,j) \in D^*_n : a_{i,j} \neq 0 \rangle\rangle .$$

De plus, soit $h \in \mathbb{N}^* \cup \{+\infty\}$, nous noterons alors par $C^h_0(\epsilon)$ et C_1 les hypothèses :

$$C^h_0(\epsilon) : \langle\langle D^N_h(X, Y) \leq \epsilon \rangle\rangle \text{ et } C_1 : \langle\langle X \text{ et } Y \text{ ne sont pas indépendantes} \rangle\rangle .$$

Remarquons que l'on peut aussi écrire C_1 sous la forme : $C_1 : \langle\langle D^N_h(X, Y) \neq 0 \rangle\rangle$.

4-6 Test Hilbertien d'indépendance Les résultats de la partie précédente vont maintenant nous permettre de préciser ce qui est réellement testé lorsque l'on utilise le test Hilbertien pour tester l'indépendance.

Supposons donc que nous voulions tester l'indépendance de deux variables aléatoires X et Y telles que $(X, Y) \in \mathbb{V}^U$.

Nous devons donc choisir quelle base employer, puis ensuite, un indicateur de dépendance, c'est à dire que nous devons choisir h . Pratiquement on choisira $h = 2$ ou $+\infty$, bien que, grâce à 1-2, on ait également des résultats sur les autres valeurs possibles de h .

Alors, ces choix étant faits, nous savons qu'il existe une suite réelle $\epsilon(q)$ vérifiant

$$\lim_{q \rightarrow \infty} (\epsilon(q)) = 0$$

telle que l'on puisse exprimer les résultats du test de la façon suivante.

L'acceptation de l'hypothèse $H_0(D^*_n)$ entraîne l'acceptation de l'hypothèse $C^h_0(\epsilon(n))$: $\langle\langle X$ et Y sont indépendantes à $\epsilon(n)$ près $\rangle\rangle$. Et l'acceptation de $H_1(D^*_n)$ implique le rejet de l'indépendance. De plus, ces hypothèses ne sont pas exclusives l'une de l'autre.

II - 2-3 Résultats numériques

4-7 Cas des fonctions de Walsh.

D'après les résultats précédents, nous déduisons que le test Hilbertien d'indépendance sera le test de $H_0 ((G_n^*)^2) \subset C_0^\infty(\varepsilon(n)) : \langle D^N(X,Y) \leq 2^{-n} \rangle$ contre $H_1 ((G_n^*)^2) \subset C_1 : \langle X \text{ et } Y \text{ ne sont pas indépendantes} \rangle$.

Concrètement, si nous notons par $N_{f.o.}$ le nombre de fonctions orthogonales employées, nous trouvons $N_{f.o.} ((G_n^*)^2) = (2^{n+1} - 1)^2$,

et nous aurons donc le tableau suivant.

n	2	3	7
$N_{D(X,Y) \leq}$	1/4	1/8	1/128
$N_{f.o.}$	49	225	65 025

Signalons que dans ce cas on peut prouver (II-5 [5], [8]) l'égalité $\|S_n\|^2 = \chi^2_{X,Y} - \chi^2_X - \chi^2_Y$, où $\chi^2_{X,Y}$ (resp. χ^2_X, χ^2_Y) représente la statistique classique du chi-deux sur $[0,1]^2$ (resp. $[0,1]$) muni de la loi uniforme, et partagé en 4^{n+1} carrés égaux (resp. en 2^{n+1} intervalles égaux). Le théorème 4-3 montre alors que, sous l'hypothèse d'indépendance, $\chi^2_{X,Y} - \chi^2_X - \chi^2_Y$ suit asymptotiquement une loi du chi-deux à $(2^{n+1} - 1)^2$ degrés de liberté.

D'autre part, on peut aussi prouver que χ^2_X est identique à la statistique Hilbertienne sur les $\alpha_{I,(0,0)}$. Ce résultat montre que $\chi^2_{X,Y}$ teste tous les $\alpha_{I,J}$ tels que $(I,J) \in (G_n)^2 \setminus \{(0,0), (0,0)\}$, y compris les $\alpha_{I,(0,0)}$ et $\alpha_{(0,0),J}$ que l'on sait déjà être nuls. Autrement dit, lorsque l'on teste l'indépendance de deux variables aléatoires de lois de probabilité connues au moyen du test du chi-deux, on teste de nouveau ces lois de probabilité.

De plus, il est facile de voir que, dans le cas de lois marginales inconnues, le test d'indépendance du chi-deux teste la nullité des coefficients $\alpha_{I,J}, (I,J) \in (G^*_n)^2$, et de eux seuls. Ce résultat explique la différence de degrés de liberté de la loi asymptotique de $\chi^2_{X,Y}$ selon que les lois marginales sont connues ou inconnues. Enfin, on peut aussi remarquer que le nombre de fonctions orthogonales utilisées est égal au nombre de classes.

4-8 Cas des polynômes de Legendre : Nous gardons les notations et hypothèses de 3-10. Alors, si l'on teste l'indépendance par le test Hilbertien, on teste donc

$$H_0(E_n) \subset C_0^2(\varepsilon(n)) : \ll D_2^N(X,Y) \leq 1,372 \sqrt{\text{Log} \left[\frac{2n+1}{2n-3} \right]} \gg$$

contre $H_1(E_n) \subset C_1 : \ll X \text{ et } Y \text{ ne sont pas indépendantes} \gg$.

Concrètement nous obtenons le tableau suivant

n	10	25	50	100	255
$D_2^N(X,Y) \leq$	0,631	0,393	0,276	0,195	0,122
N f.o.	100	625	2 500	10 000	65 025

4-9 Cas des fonctions trigonométriques Les conclusions que l'on obtient avec les fonctions sinus et cosinus sont absolument identiques à celles obtenues avec les polynômes de Legendre [5].

4-10 Cas du test d'uniformité On obtient le même genre de résultats avec les fonctions de Walsh et la distance $\delta_1(F,U)$ [5].

Remarque 4-11 Nous remarquons, bien sûr, le nombre élevé de fonctions orthogonales nécessaires pour obtenir de faibles majorations de $D_h^N(X,Y)$. Mais cela est assez normal : nous reviendrons sur cette question en III-3.

III CONCLUSION

Avant de tirer les conséquences pratiques des résultats précédents, rappelons que certaines de ces conclusions peuvent s'étendre à bien d'autres cas que ceux étudiés ici : par exemple, au cas des coefficients de corrélation d'ordre supérieur ([5],[6], [7]), ou aux indicateurs $D_h^N(F_1(\cdot), F_2(\cdot))$ (cf partie I) .

III-1 Considérations sur les indicateurs D_h^N

La question qui se pose d'abord est bien sûr de savoir si le choix de ces indicateurs pour mesurer la dépendance est judicieux. Aussi rappelons d'abord qu'ils ont de bonnes propriétés mathématiques. En particulier, ils vérifient les axiomes que Renyi estimait nécessaires à de bonnes mesures de dépendance (1-3).

Rappelons aussi qu'ils ont un bon comportement par rapport à la taille des échantillons : si un échantillon est de taille suffisante, l'estimation que nous pourrons faire de D_h^N sera a priori satisfaisante (cf 1-7).

On peut traduire ce phénomène en disant que les indicateurs D_h^N correspondent à un aspect pratique et quasi-intuitif de la dépendance : le désir que, lorsque des échantillons de taille élevée vérifient l'indépendance, la mesure de la dépendance soit proche de 0 (les coefficients d_h définis en 1-5 ne possèdent donc pas de telles propriétés).

De plus, par rapport au problème étudié dans cet article, rappelons que la connaissance des premiers termes du développement en série de la densité permettra d'obtenir une bonne approximation de la dépendance mesurée par les indicateurs D_h^N .

En particulier, ils seront surtout sensibles aux coefficients $\theta_{i,j}$, par exemple, tels que $i+j$ soit assez petit, et donc, en particulier, à $\theta_{1,1} = \rho(X, Y)$, le coefficient de corrélation linéaire.

A ce sujet, signalons que l'on peut aussi prouver (4-5 [7]) que les coefficients $\theta_{q,1}$ et $\theta_{1,q}$, $q \in \mathbb{N}^*$, mesurent l'existence d'une bijection, donc d'une dépendance très importante. Or, en se reportant à 3-9, on voit que ce sont ces coefficients qui sont les plus importants pour la dépendance mesurée par les D_h^N .

Finalement, ces résultats sont bien conformes aux espérances et aux usages habituels des Statistiques.

Concluons donc en disant que les D_h^N ont de "bonnes" propriétés et que les résultats obtenus justifient a posteriori leur utilisation pour mesurer la dépendance induite par les coefficients des séries.

III-2 : les $\theta_{i,j}$ et les $\alpha_{i,j}$ considérés comme indicateurs de dépendance

Si, en général, on mesure la dépendance par un seul nombre, on peut aussi la mesurer par plusieurs ou par une infinité, ainsi par exemple, par les coefficients de corrélation d'ordre supérieur, les $\theta_{i,j}$.

Rappelons les propriétés de ces coefficients (pour les $\alpha_{1,j}$, se reporter à la page 131 de [5]).

1) Ils sont normalisés : $|\theta_{i,j}| \leq 1$ (cf 4-17 [5], 2-6[6]) : ce résultat est en fait la conséquence de ce que $\theta_{i,j}$ est le coefficient de corrélation linéaire de $P_i(X)$ et $P_j(Y)$.

2) Pour cette même raison (cf 2-13[6]), on peut voir qu'ils mesurent chacun une dépendance particulière : $\theta_{1,1}$, la dépendance linéaire, $\theta_{1,2}, \theta_{2,1}, \theta_{2,2}$, des dépendances quadratiques, etc...

3) On peut aussi prouver (4-5 [7]) que, pris par groupe, ils mesurent aussi d'autres types de dépendance. Ainsi, la colonne des $\{\theta_{i,1}\}_{i \in \mathbb{N}^*}$ mesure l'existence d'une fonction univoque $f \in L^2$ telle que $Y = f(X)$ p.s. .

4) Considérés en tant que paramètres, les $\theta_{i,j}$ sont "presque indépendants" les uns des autres (cf 4-18 [5], 2-14 et 2-15 [6], 4-1 à 4-4 [7]). Nous voulons dire par là qu'ils ne doivent satisfaire qu'à une seule contrainte : que la fonction de répartition qu'ils définissent soit croissante (ou la densité de probabilité positive).

En effet, toute suite $\{\theta'_{i,j}\}_{(i,j) \in \mathbb{N}^* \times 2}$ vérifiant

$$1 + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \theta'_{i,j} P_i \times P_j \geq 0 \quad ,$$
$$\text{et } \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (\theta'_{i,j})^2 < \infty \quad ,$$

définira sur $[0,1]^2$ une probabilité Q' de lois marginales uniformes.

Remarquons que cette indépendance paramétrique se traduit lors du test Hilbertien en degrés de liberté.

Et, en fait, les $\theta_{i,j}$ considérés comme indicateurs de dépendance, s'avèrent être des instruments très fins et très puissants car ils sont capables de détecter presque tous les types de dépendance fonctionnelle possibles (cf l'étude générale des coefficients de corrélation d'ordre supérieur : spécialement dans [6]).

Et finalement, plutôt que de se limiter à un seul indicateur, on les choisira pour mesurer la dépendance, lorsque l'on aura besoin d'une étude fine de celle-ci.

III-3 : Choix de l'ordre du développement

Bien qu'il existe une généralisation de 4-1, 4-2 et 4-3 au cas du test d'un nombre infini de coefficients $a_{i,j}$ (cf [9] et [10]), il est bien clair que concrètement, on devra toujours se limiter à en tester un nombre fini.

Or, supposons que nous voulions tester l'indépendance de deux variables aléatoires, et supposons que le résultat du test nous conduise à accepter l'hypothèse H_0 (D^*). Si nous nous rappelons l'exemple 1-7, on voit que l'on ne pourra même pas en déduire si les deux variables sont indépendantes ou en bijection.

On ne pourra fournir une réponse plus certaine que si l'on accepte H_1 (D^*) : Dans ce cas, on pourra conclure qu'il n'y a pas indépendance.

Finalement, le choix du nombre de coefficients testés ne peut donc être fait qu'en fonction de l'indépendance désirée. Il convient donc de préciser d'abord de quelle "indépendance" on a besoin, et ensuite d'effectuer le choix du nombre de coefficients à tester avec soin.

A ce sujet, faisons les remarques suivantes.

Nombre de fonctions orthogonales nécessaires

Posons nous donc la question de savoir si le grand nombre de fonctions orthogonales nécessaires (cf 4-11) pour obtenir que les coefficients $D_{N_h}^N$ soient assez petits est justifié.

Il semble bien qu'il faille répondre par oui. En effet, si l'on veut une "indépendance de qualité" il faudra faire intervenir un grand nombre de fonctions orthogonales.

Par exemple, si l'on se contente d'une "indépendance $D^N(X,Y) \leq 1/8$ ", il faudra déjà utiliser une partition de $[0,1]^2$ en 256 carrés égaux (i.e. 225 fonctions orthogonales pour le test d'indépendance).

Or, dans ce cas, on considère que $(X,Y) \in V^U$, défini par $Y = 8X$ (cf 1-6) est un couple de variables indépendantes (ce sera aussi le cas dans l'exemple ci-après).

Par contre, si l'on estime que $D^N(X,Y) \leq (1/128) = 0,0078$ est à peine suffisant comme critère d'indépendance, il faudra recourir à une partition de $[0,1]^2$ en au moins 2^{16} carrés ou à $(2^8 - 1)^2$ fonctions orthogonales !

Et rappelons nous aussi que les majorations que nous avons données des indicateurs $D_{N_h}^N(X,Y)$ ne sont pas trop grossières (2-8,3-10). En particulier, la majoration de $D^N(X,Y)$ pour les fonctions de Walsh est presque exacte (2-8).

Donc, si l'on a besoin d'une "indépendance" de qualité, il nous faudra tester un grand nombre de coefficients $a_{i,j}$. Or, il arrive fréquemment que ce nombre ne soit pas choisi assez grand. Donnons un exemple concret, et relativement fréquent, de ce genre d'erreur et des conséquences pratiques qui en découlent.

Exemple du choix d'un nombre de classes insuffisant

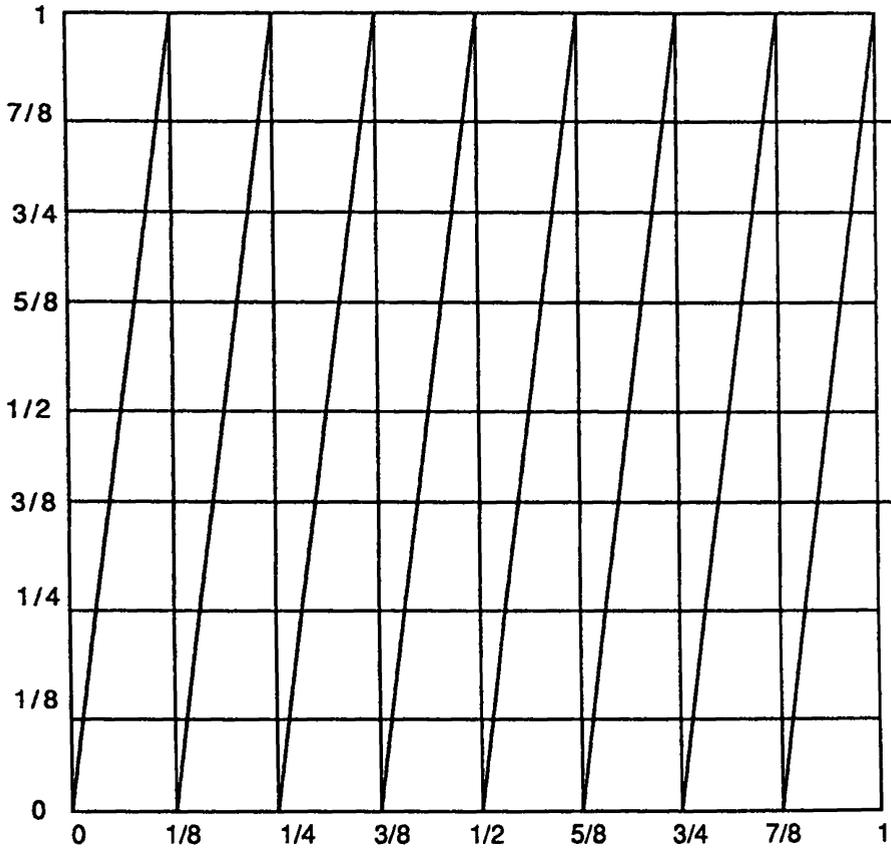
Supposons que nous voulions tester la qualité d'une simulation de la loi uniforme sur $[0,1]$ fournie par la congruence linéaire $T(x) \equiv 8x \pmod{m}$, où $m \in \mathbb{N}^*$.

Pour cela, on décide de tester si deux nombres simulés successifs peuvent être considérés comme indépendants l'un de l'autre en employant le test du chi-deux associé à la loi uniforme sur $[0,1]^2$ et à une partition de $[0,1]^2$ en 64 carrés égaux.

On va donc tester la bonne répartition des points d'un échantillon simulé de taille n dans $[0,1]^2$, c'est à dire les points de la forme

$$\left(\frac{T^{2p}(x')}{m}, \frac{T^{2p+1}(x')}{m} \right),$$

où $p=1,2,\dots,n$ et où x' est choisi au hasard dans \mathbb{N} .



$$f(x) = \overline{8x}$$

D'après ce que nous avons dit (cf 4-7 ou [8]), ce test est identique au test Hilbertien des $\alpha_{I,J}$, $(I,J) \in D^* = G_4 \times G_4 \setminus \{(0,0), (0,0)\}$.

Supposons enfin que l'on ait choisi de prendre des échantillons de taille $n < 400$.

Or, il se peut que la congruence T soit telle que tout échantillon de la forme définie ci-dessus soit convenablement réparti dans les 64 carrés. Il se peut aussi que la qualité uniforme de cette répartition s'améliore si la taille de l'échantillon simulé augmente (à cause des propriétés des congruences linéaires et de leur période effective qu'il suffit de choisir égale à 400 : cf II-2, II-3 [17], 1-7 [4]). On peut prendre par exemple $m = 8^{400} + 1$ cf 2-13 [4].

Dans ce cas, bien sûr, en effectuant le test, nous trouverons que

- 1) l'hypothèse $H_0(D^*)$ est vérifiée,
- 2) la vérification de cette hypothèse s'améliore si la taille de l'échantillon augmente.

Si nous ne prenons pas de précautions, nous déduirons que

- 1) la simulation est bonne,
- 2) la qualité de cette simulation s'améliore lorsque l'on augmente la taille de l'échantillon,

alors que,

- 1) les points étant concentrés sur 8 segments de droite, la simulation peut être considérée comme mauvaise,
- 2) ce défaut apparaît de plus en plus nettement si la taille de l'échantillon augmente.

Nous aurons donc conclu l'opposé de la réalité. Et cela parce que nous avons oublié que nous ne testions que la nullité des premiers coefficients $\alpha_{I,J}$ qui, eux sont effectivement nuls.

Or ce genre d'erreur est relativement fréquent. Ainsi, il se produit assez souvent lors des tests des simulations engendrées par les congruences, et cela, à cause des propriétés des périodes effectives. L'exemple que nous avons donné ici n'a été choisi volontairement aussi grossier que pour visualiser le type d'erreurs commises dans ces cas là.

Conclusion

Lorsque l'on veut tester l'indépendance de deux événements par le test Hilbertien, il faut donc procéder en trois temps.

- 1) Savoir de quelle "indépendance" on a besoin, et ce, en fonction de la nature des événements testés: économie, médecine, etc.
- 2) En fonction de l'"indépendance" désirée, choisir le nombre de coefficients à tester.
- 3) Opérer alors le test.

Et, il faut bien se souvenir que si l'"indépendance" désirée n'a pas été définie avec soin, on peut très bien déduire d'un test correctement effectué des conclusions, non seulement fausses, mais carrément opposées à la réalité, comme c'était le cas dans l'exemple ci-dessus.

Les résultats de cet article permettent donc de préciser quelle "indépendance" est réellement testée, et visualisent la différence entre le test d'indépendance réel et le test d'un nombre fini de

coefficients. Et, c'est bien sûr, de la confusion entre ces deux notions que provient le fait que le choix du nombre de classes, par exemple, laisse parfois à désirer.

Remarques

En ce qui concerne le choix concret du nombre de classes dans les tests de qualité de simulation, il faut donc recourir aux critères de qualité de celles-ci : à ce sujet, se rapporter à [4], [5], [17] et [18], l'exemple ci-dessus, ayant été traité en partie III-2 de [5].

Dans le cas où l'on veut tester une "indépendance maximale", il faut alors choisir le nombre de coefficients à tester en fonction de la taille de l'échantillon utilisé : se reporter alors aux résultats de Bosq sur ce sujet en [9], [10] et [11].

Rappelons nous enfin que le genre d'erreurs signalées ici n'a pas uniquement lieu dans les tests Hilbertiens. La vérification de l'égalité d'indépendance (1-7) due aux arrondis des calculs donnera les mêmes résultats avec n'importe quel test.

BIBLIOGRAPHIE

- [1] BEAUCHAMPS K.G. (1975)
Walsh functions and their applications. Academic Press, London.
- [2] BLACHER R. (1982)
Etude des coefficients de dépendance entre deux variables aléatoires créés à partir des distances mathématiques classiques. I.M.A.G. R.R. n° 289.
- [3] BLACHER R. (1982)
Influence sur la dépendance des coefficients du développement en série de la densité de probabilité. I.M.A.G. R.R. n° 343.
- [4] BLACHER R. (1983)
Quelques propriétés des congruences linéaires considérées comme générateurs de nombres pseudo-aléatoires. I.M.A.G. R.R. n° 345.
- [5] BLACHER R. (1983)
Indicateurs de dépendance entre deux variables aléatoires fournis par le développement en série de la densité de probabilité. Thèse 3ème Cycle, Université Scientifique et Médicale de Grenoble.
- [6] BLACHER R. (1984)
Coefficients de corrélation d'ordre supérieur.
"Statistique et analyse des données" Volume 9, n° 2, p 48-67 .
- [7] BLACHER R. (1984)
Coefficients de corrélation d'ordre (i,j) et variances d'ordre i. TIM3-IMAG R.R. n° 463.
- [8] BLACHER R. (1985)
A new chi squared independence test. TIM3-IMAG R.R. n° 512.

- [9] BOSQ D. (1978)
Tests d'ajustement hilbertiens. Université des Sciences et Techniques de Lille 1,
publication interne de l'U.E.R. de Mathématiques pures et appliquées : n° 125.
- [10] BOSQ D. (1978)
Tests hilbertiens et test du khi-deux. C.R. Acad. Sc. Paris. t. 286 : 946-948 .
- [11] BOSQ D. (1983)
Lois limites et efficacité asymptotique des tests hilbertiens de diverses lois sous des hypothèses
adjacentes. "Statistique et analyse des données", Volume 8 n° 1 , p 1-40 .
- [12] BOURBAKI N. (1965)
Eléments de mathématiques, Livre VI Intégration, Chapitre 4, 2ème édition. Hermann, Paris.
- [13] CAILLEZ F. , PAGES J.P. (1976)
Introduction à l'Analyse des données S.M.A.S.H. , Paris .
- [14] NATANSON I.P. (1965)
Constructive function theory : Volume 2, Approximation in mean.
Frederick Ungar Publishing C°. New-York.
- [15] RENYI A. (1959)
On measures of dependence.
Acta. Math. Acad. Sci. Ungar. 10. 441-451.
- [16] SCHWEIZER B., WOLF E.F. (1981)
On non parametric measures of dependence for random variables.
Annals of Mathematic Statistics, 9, 879-885.
- [17] VAN CUTSEM B. (1980)
Etude de la période de la suite de nombres pseudo-aléatoires engendrée par une congruence
linéaire. Bulletin de l'A.P.M. n° 324.
- [18] VAN CUTSEM B. (1985)
Simulation des lois de probabilité. Statistique et analyse des données : Vol 10, n° 2 , p 63-87.