APPROXIMATION D'UNE PROBABILITE SUR UN ESPACE DE HILBERT ET ANALYSE EN COMPOSANTES PRINCIPALES

Alain ARCONTE-Jérôme MANUCEAU-Claude MARTIAS

Université des Antilles et de la Guyane
U.F.R. Sciences Exactes et Naturelles
Département de mathématiques B.P. 592
97167 Pointe-à-Pitre. GUADELOUPE

Résumé.

Soit $\mathcal{H}$ un espace de Hilbert réel, séparable et Q une probabilité sur $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$, de variance finie. On montre qu'il existe une suite croissante de sous espaces vectoriels de dimensions finies $(V_n)_{n\in \mathbb{N}}$, tels que les probabilités marginales correspondantes $\tilde{Q}_{V_n}$ soient des approximations de Q sur V_n et que la suite $(\tilde{Q}_{V_n})_{n\in \mathbb{N}}$ converge étroitement vers Q. On montre enfin que l'Analyse en Composantes Principales est un cas particulier, du problème traité.

1 Introduction.

Dans l'étude d'une probabilité sur un espace de Hilbert il est parfois utile d'en rechercher une approximation ou une suite d'approximations. Il est naturel de chercher à construire cette dernière à l'aide de probabilités marginales sur des sous-espaces vectoriels de dimension finie.

Les techniques utilisées par l'Analyse en Composantes Principales sont très générales et connues depuis longtemps sous par exemple la dénomination de "décomposition de Kanhunen - Loeve [6]."

Elles ont été développées dans des domaines divers (en particulier dans la théorie du signal) par de très nombreux auteurs : S. CILIBERTO et B. NICOLAENKO [4], N. AUBRY, R. GUYONNET et R. LIMA [1][2], M.KIRBY et D. ARMBRUSTER [7], C. PARDOUX [8], J.D. RODRIGUEZ et L. SIROVICH [9].

La généralisation de l'Analyse en Composantes Principales à un espace de Hilbert réel, séparable donne une dimension nouvelle à cette théorie classiquement considérée comme une méthode statistique descriptive ; dans ce cadre plus large, elle apparaît comme une méthode d'approximation d'une probabilité. Elle peut donc être considérée comme faisant partie de la statistique inférentielle.

2 Approximation d'une probabilité.

Soit $\mathcal{H}$ un espace de Hilbert réel, séparable, $\mathcal{B}(\mathcal{H})$ la tribu borelienne associée et Q une probabilité sur $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. On suppose que sa moyenne

$\mu = E_Q(X)$ existe ($\in \mathcal{H}$), ainsi que sa variance

$$var(Q) = \int_{\mathcal{H}} \|x-\mu\|^2 dQ(x) < +\infty.$$

Cela implique que sa covariance

$$\Gamma = cov(Q) = \int_{\mathcal{H}} (x-\mu) \otimes (x-\mu) dQ(x)$$

est un opérateur nucléaire de $\mathcal{L}(\mathcal{H})$ et que

$$var(Q) = Tr(cov(Q)).$$

Proposition 2.1 *Soit $(Q_n)_n$ une suite de probabilités sur $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ ayant la même moyenne μ. Alors,*

$$(var(Q_n) \underset{n}{\longrightarrow} 0) \Longrightarrow (Q_n \underset{n}{\Longrightarrow} \delta_\mu)$$

où δ_μ est la mesure de Dirac en μ.

Démonstration : $\forall \rho > 0$,

$$var(Q_n) \geq \int_{B(\mu,\rho)^c} \|x-\mu\|^2 dQ_n(x) \geq \rho^2 Q_n(B(\mu,\rho)^c).$$

Si f est une fonction continue bornée de $\mathcal{H}$ dans $I\!R$,

$$E_{Q_n}(|f-f(\mu)|) = \int_{B(\mu,\rho)} |f-f(\mu)| dQ_n + \int_{B(\mu,\rho)^c} |f-f(\mu)| dQ_n.$$

En choisissant ρ tel que $|f-f(\mu)| < \varepsilon$ sur $B(\mu,\rho)$, alors,

$$E_{Q_n}(|f-f(\mu)|) \leq \varepsilon + \frac{2\|f\|}{\rho^2} var(Q_n) \qquad \diamond$$

Remarque : La réciproque est fausse. Pour le voir, il suffit de prendre dans $(I\!R, \mathcal{B}(I\!R))$ la suite des probabilités

$$P_n = \left(1 - \frac{1}{2n^2}\right)\delta_0 + \frac{1}{2n^2}(\delta_{-n} + \delta_{+n})$$

Evidemment $var(P_n) = 1$ et $P_n \Longrightarrow \delta_0$.

Soit V un sous espace vectoriel fermé de $\mathcal{H}$. On sait que $\mathcal{H}$ est isomorphe à $V \times V^{\perp}$ et que $\mathcal{B}(\mathcal{H}) = \mathcal{B}(V) \otimes \mathcal{B}(V^{\perp})$. Soit P_V le projecteur orthogonal sur V et posons

$$Q_V = Q \circ P_V^{-1}.$$

Q_V est une probabilité sur $(V, \mathcal{B}(V))$ et

$$\forall A \in \mathcal{B}(V), Q_V(A) = Q(A \times V^{\perp}).$$

Q_V est donc la probabilité marginale de Q sur V et on voit aisément que

$$varQ = varQ_V + varQ_{V^{\perp}}$$

et que $$suppQ_V = P_V(suppQ).$$

Grâce à la Proposition 2.1, on peut s'attendre à ce que la probabilité $Q_V \otimes \delta_{P_{V^{\perp}}(\mu)}$ soit une "approximation" de Q, d'autant "meilleure" que $varQ_{V^{\perp}}$ est plus petite. La qualité de cette approximation pourra être appréciée par le rapport $\alpha_V = \dfrac{var(Q_V)}{var(Q)}$ qui appartient à $[0,1]$ et vérifie
$\alpha_V = 0 \iff Q = \delta_{P_V(\mu)} \otimes Q_{V^{\perp}}$ et $\alpha_V = 1 \iff Q = Q_V \otimes \delta_{P_{V^{\perp}(\mu)}}$.

Nous allons préciser cette notion d'approximation, dans les propositions 2.2 et 2.3.

Dans tout ce qui va suivre, nous allons noter $(e_i)_{i \in \mathbb{N}^*}$ une base orthonormale de $\mathcal{H}$ qui diagonalise Γ et $(\lambda_i)_{i \in \mathbb{N}^*}$ la suite des valeurs propres correspondantes. On peut supposer que $\forall i \in \mathbb{N}^*, \lambda_i > 0$. En effet, si $ker\Gamma \neq \{0\}$, en posant $V = (ker\Gamma)^{\perp}$, on peut voir que

$$supp(Q) \subset V \times \{P_{V^{\perp}}(\mu)\}$$

et que $$Q = Q_V \otimes \delta_{P_{V^{\perp}}(\mu)}.$$
En outre, nous choisissons l'indéxation de la base de façon à ce que la suite des valeurs propres soit décroissante.
Evidemment $(\lambda_i)_i \in \ell^1_{\mathbb{R}}, \{\lambda_i | i \in \mathbb{N}^*\} \subset [0, \|\Gamma\|]$ et $\lambda_1 = \|\Gamma\|$.

Proposition 2.2

$$\forall n \in \mathbb{N}^*, \ si\ V_n = [\{e_1, \ldots, e_n\}],$$

alors,

$$var(Q_{V_n}) = \sup_{dimV=n} var(Q_V).$$

Démonstration : raisonnons par récurrence.

1) $n = 1$. Soit e un vecteur unitaire de $\mathcal{H}$. Alors $varQ_{[e]} =< e|cov(Q)e >$. Evidemment, ce nombre est maximum lorsque e est un vecteur propre de $covQ$, correspondant à la plus grande valeur propre.

2) Supposons la propriété vraie pour n et démontrons la pour $n+1$. Soit V un sous espace vectoriel de $\mathcal{H}$ de dimension $n+1$. Montrons que

$$varQ_V \leq varQ_{V_{n+1}}.$$

Puisque la codimension de $V_n^\perp$ est n, alors, $V_n^\perp \cap V \neq \{0\}$; soit x un élément unitaire de cet ensemble et $W = [x]^\perp \cap V$. Par hypothèse $varQ_W \leq varQ_{V_n}$ et puisque $covQ$ laisse invariant $V_n^\perp, varQ_{[x]} \leq varQ_{[e_{n+1}]}$. D'où

$$varQ_V = varQ_W + varQ_{[x]} \leq varQ_{V_{n+1}} \qquad \diamond$$

Les V_n ne sont uniques que si les valeurs propres de $covQ$ sont de multiplicité 1. V_n s'appelle, **espace principal** de Q de dimension n.

Corollaire 2.1 *Toute probabilité gaussienne $Q = N(\mu, \Gamma)$ sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, peut s'écrire,*

$$Q = \otimes_{i=1}^n Q_{[e_i]}$$

où $e_1, \ldots, e_n$ est une base orthonormale qui diagonalise Γ, dont les valeurs propres sont $\lambda_1, \ldots, \lambda_n$. Si $\lambda_i \neq 0$, $Q_{[e_i]} = N(P_{[e_i]}(\mu), \lambda_i)$ et si $\lambda_i = 0$, $Q_{[e_i]} = \delta_{P_{[e_i]}}(\mu)$.

$\mathcal{H}$ étant séparable, la base $(e_i)_{i \in \mathbb{N}^*}$ définit un isomorphisme de $\mathcal{H}$ sur $\ell^2_{\mathbf{R}}$. Comme $\ell^2_{\mathbf{R}} = \mathbb{R}^n \times \ell^2_{\mathbf{R}}, \forall n \in \mathbb{N}^*$, on a

$$\mathcal{B}(\ell^2_{\mathbf{R}}) = \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\ell^2_{\mathbf{R}}).$$

Posons

$$\widetilde{Q}_{V_n} = Q_{V_n} \otimes \delta_{P_{V_n^\perp}(\mu)}.$$

Evidemment

$$var\widetilde{Q}_{V_n} = \sum_{i=1}^n \lambda_i \to var(Q) = \sum_{i=1}^\infty \lambda_i.$$

Proposition 2.3

$$\tilde{Q}_{V_n} \underset{n}{\Longrightarrow} Q.$$

Démonstration : $\forall x \in \mathcal{H}$, soit $(x_i)_{i \in \mathbb{N}^*}$ la suite des coordonnées de x dans la base $(e_i)_{i \in \mathbb{N}^*}$ et $(\mu_i)_{i \in \mathbb{N}^*}$ celle de μ. Si f est une fonction continue bornée de $\ell^2_{\mathbb{R}}$ dans $\mathbb{R}$, alors,

$$\begin{aligned} E_{Q_n}(f) &= \int_{\mathbb{R}^n} f(x_1, \ldots, x_n, \mu_{n+1}, \ldots) dQ_{V_n}(x_1, \ldots, x_n) \\ &= \int_{\ell^2} f(x_1, \ldots, x_n, \mu_{n+1}, \ldots) dQ(x_1, \ldots) \\ &= \int_{\ell^2} \varphi_n dQ \end{aligned}$$

où $\varphi_n((x_i)_{i \in \mathbb{N}^*}) = f(x_1, \ldots, x_n, \mu_{n+1}, \ldots)$.

On conclut avec le théorème de la convergence dominée $\diamond$

Dans la proposition 2.3, nous aurions pu remplacer $(V_n)_n$ par une suite quelconque de sous espaces vectoriels de dimensions finies de $\mathcal{H}$, croissante, dont l'union est dense dans $\mathcal{H}$. Le choix des espaces principaux a un double intérêt. D'une part nous connaissons pour tout n $var(\tilde{Q}_{V_n})$ et d'autre part cette suite, d'après la proposition 2.2 est celle qui converge le plus vite vers $var(Q)$.

Remarque : Supposons pour simplifier les notations que $\mu = 0$. $\tilde{Q}_{V_n}$ est une approximation de Q sur $V_n \times \{0\}$ car

i) $supp(\tilde{Q}_{V_n}) \subset V \times \{0\}$

ii) $\forall A \in \mathcal{B}(V)$ $\tilde{Q}_{V_n}(A \times \{0\}) = Q(A \times V_n^{\perp})$

d'où $0 \leq \tilde{Q}_{V_n}(A \times \{0\}) - Q(A \times \{0\}) =\leq Q(\{0\} \times (V_n^{\perp} - \{0\})) \leq$
$\int_{B(0,\rho)-\{0\}} dQ + \int_{(\{0\} \times V_n^{\perp}) \cap B(0,\rho)^c} \frac{\|x\|^2}{\rho^2} d\,Q(x) \leq$
$Q(B(0,\rho) - \{0\}) + \frac{1}{\rho^2} var(\tilde{Q}_{V_n^{\perp}})$

Chaque terme de cette dernière somme pouvant être choisi inférieur ou égal à $\frac{\epsilon}{2}$; le premier en prenant ρ suffisamment petit et le deuxième en prenant n suffisamment grand.

En outre si $\dim[supp(Q)] = k$ alors V_k qui est l'espace principal de dimension k est le seul sous-espace de dimension k qui vérifie $\tilde{Q}_{V_k} = Q$.

Posons $\mathcal{C}_n = \mathcal{B}(\mathbb{R}^n) \times \ell^2 \subset \mathcal{B}(\ell^2), \mathcal{C} = \cup_{n\in\mathbb{N}^*}\mathcal{C}_n$ et $\mathcal{B}^\infty = \sigma(\mathcal{C})$. Si Q est une gaussienne dans $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$, d'après le corollaire 2.1 :
$\forall n \in \mathbb{N}^*, Q_n = \left(\otimes_{i=1}^n Q_{[e_i]}\right) \otimes \delta_{P_{V_n^\perp}(\mu)}$. Evidemment $Q = Q_n$ sur $\mathcal{C}_n$. On peut alors noter $\otimes_{i=1}^\infty Q_{[e_i]}$ la restriction de Q à $\mathcal{C}$.

Lemme 2.1

$$\mathcal{B}^\infty = \mathcal{B}(\ell_2).$$

Démonstration : il est clair que $\mathcal{B}^\infty \subset \mathcal{B}(\ell_2) = \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}(\ell^2)$. Pour démontrer l'inclusion inverse, nous allons montrer que $B(0,1) \in \mathcal{B}^\infty$; plus précisément, nous montrerons que $B(0,1)$ peut-être recouvert par un ensemble dénombrable d'éléments de $\mathcal{B}^\infty$.

Soit $(I_i)_{i\in\mathbb{N}}$ une suite d'intervalles $I_i =]z_i - \rho_i, z_i + \rho_i[$, formant une base pour l'ensemble des ouverts de $]-1,1[$. L'ensemble $\mathcal{D}$ des éléments de la forme $\prod\limits_{n\in\mathbb{N}^*} I_{i_n}$ où $\sum\limits_{n=1}^{\infty} \rho_{i_n}^2 \leq 1 - \sum\limits_{i=1}^{\infty} z_{i_n}^2$ est un sous ensemble dénombrable de $\mathcal{B}^\infty$ dont les éléments sont inclus dans $B(0,1)$.

$$\forall x = (x_i)_{i\in\mathbb{N}^*} \in B(0,1) \text{ et } \forall n \in \mathbb{N}^*, \exists I_{i_n}$$

tel que $$x_n \in I_{i_n} \subset]x_n - \tfrac{1-|x_n|}{2^{n+1}}, x_n + \tfrac{1-|x_n|}{2^{n+1}}[.$$

On voit alors que $x \in \prod_{n=1}^\infty I_{i_n} \in \mathcal{D}$. $\diamondsuit$

Corollaire 2.2 *Pour toute gaussienne Q de $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$, on a*

$$Q = \bigotimes_{i=1}^{\infty} Q_{[e_i]}$$

où $Q_{[e_i]} = N(P_{[e_i]}(\mu), \lambda_i)$ si $\lambda_i \neq 0$ et $Q_{[e_i]} = \delta_{P_{[e_i]}(\mu)}$ si $\lambda_i = 0$.

3 Analyse en Composantes Principales.

Soit $(\Omega, \mathcal{F}, R)$ un espace probabilisé, $\mathcal{H}$ un espace de Hilbert réel séparable et X une variable aléatoire de $(\Omega, \mathcal{F})$ dans $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$ de norme au carré intégrable. Définissons $\tilde{X}$ appliquant $L^2_{\mathbb{R}}(\Omega, \mathcal{F}, R)$ dans $\mathcal{H}$ par

$$\tilde{X}(f) = \int_\Omega X(\omega) f(\omega) dR(\omega).$$

On voit que

$$\forall u \in \mathcal{H}, \forall \omega \in \Omega, \tilde{X}^*(u)(\omega) = < X(\omega) | u >$$

et que $\tilde{X} \circ \tilde{X}^* = covR_X$.

Définition : On appelle *Analyse en Composantes principales de* X la recherche d'une base orthonormale $(e_i)_{i \in \mathbb{N}^*}$ de $\mathcal{H}$ diagonalisant $covR_X$ et rangée par ordre décroissant de leurs valeurs propres correspondantes. Les e_i s'appellent *vecteurs principaux* et les $\varepsilon_i = \frac{1}{\sqrt{\lambda_i}} \tilde{X}^*(e_i)$ qui sont des vecteurs propres de $\tilde{X}^* \circ \tilde{X}$ avec les mêmes valeurs propres, *composantes principales.*

L'avantage de cette définition sur celle proposée dans [5] est que c'est la plus simple possible généralisant la théorie classique de l'A.C.P. où
$\Omega = \{x_1, \ldots, x_n\}$ et $R = \sum_{i=1}^n p_i \delta_{x_i} (p_i > 0, \sum_i p_i = 1)$ [3].

Notons $\sigma(X) = X^{-1}(\mathcal{B}(\mathcal{H})) \subset \mathcal{F}$ et $R_X = Q$. Avec les notations ci-dessus, on a

Proposition 3.1 *$\tilde{X}$ (resp. $\tilde{X}^*$) est une injection continue de $L^2_{\mathbb{R}}(\Omega, \sigma(X), R)$ dans $\mathcal{H}$ (resp. de $\mathcal{H}$ dans $L^2_{\mathbb{R}}(\Omega, \sigma(X), R)$), telle que*

$$\begin{aligned} \forall i \in \mathbb{N}^*, \tilde{X}(\varepsilon_i) &= \sqrt{\lambda_i} e_i \\ \tilde{X}^*(e_i) &= \sqrt{\lambda_i} \varepsilon_i \end{aligned}$$

*En outre $(\varepsilon_i)_{i \in \mathbb{N}^*}$ est une base orthonormale de $L^2(\Omega, \sigma(X), R)$ et
$\forall n \in \mathbb{N}^*, (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), R_{(\sqrt{\lambda_1}\varepsilon_1, \ldots, \sqrt{\lambda_n}\varepsilon_n)})$ est égal à un isomorphisme près à $(V_n, \mathcal{B}(V_n), Q_{V_n})$.*

Démonstration : $P_{V_n}(X) = \sum_{i=1}^n < X | e_i > e_i = \sum_{i=1}^n \sqrt{\lambda_i} \varepsilon_i e_i$ et $Q_{V_n} = Q \circ P_{V_n}^{-1} = R \circ P_{V_n}(X)^{-1} = R_{\sum_{i=1}^n \sqrt{\lambda_i} \varepsilon_i e_i}$ $\diamond$

Ainsi, les n premiers facteurs principaux $\varepsilon_1, \ldots, \varepsilon_n$, permettent de reconstruire Q_n, approximation de R_X. On comprend tout l'intérêt de ces derniers en inférence statistique.

4 Conclusion.

L'Analyse en Composantes Principales est habituellement présentée, comme une **méthode descriptive** de l'étude d'un système statistique [3] ce qui a un intérêt limité. On voit avec la proposition 3.1, qu'elle peut-être aussi considérée comme une méthode d'approximation d'une probabilité inconnue, à laquelle nous avons accès que par l'intermédiaire de l'observation expérimentale de variables aléatoires.

Elle fait donc partie intégrante de la statistique inférentielle.

Références

[1] N. Aubry, R. Guyonnet, R. Lima : Spatio-temporel analysis of complex signals : théorie and applications. J. Stat. Phys. 63, 3 (1991).

[2] N. Aubry, R. Guyonnet, R. Lima : On the turbulence spectra (in preparation).

[3] F. Cailliez et J.P. Pages : Introduction à l'analyse des données. S.M.A.C.H. (1976).

[4] S. Cilibento and B. Nicolaenko : Europhys. Le H 14(4) 303.

[5] J. Dauxois et A. Pousse : Analyses factorielles en calcul des probabilités et en statistique. Thèse. Université Paul Sabatier. Toulouse 1976.

[6] K. Kanhuren : zw Apektral theorie stochatische prozesse.Ann. Acad. Sci. Fernicae Ser. A.1 (1944) 34

[7] M. Kirby and D. Armbryster : Reconstructing phase foce for P.D.E. simulations Preprint (1991).

[8] C. Pardoux : Apport de l'analyse factorielle à l'étude d'un processus. Rev. Stat. App. 1989 XXXVII(4) 41-60.

[9] J.D. Rodriguez and α Sinovich, Physica D. 43 (1990) 77.