



Analyses factorielles des distributions marginales de processus

Rachid Boumaza

► **To cite this version:**

Rachid Boumaza. Analyses factorielles des distributions marginales de processus. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1999. Français. tel-00004806

HAL Id: tel-00004806

<https://tel.archives-ouvertes.fr/tel-00004806>

Submitted on 18 Feb 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE JOSEPH FOURIER - GRENOBLE 1
SCIENCES & GEOGRAPHIE

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE JOSEPH FOURIER

Discipline : **MATHEMATIQUES APPLIQUEES**

Présentée et soutenue publiquement

le 4 janvier 1999

par

BOUMAZA Rachid

ANALYSES FACTORIELLES
DES DISTRIBUTIONS MARGINALES DE PROCESSUS

Directeur de thèse

YCART Bernard

COMPOSITION DU JURY

Mme **ROBERT Claudine** Présidente
M. **CELEUX Gilles**
M. **ESCOUFIER Yves**
M. **SAPORTA Gilbert**
M. **YCART Bernard**

Table des matières

Introduction	5
1 Mesure d'affinité L^2 de deux densités de probabilité et propriétés asymptotiques	11
1.1 Introduction	11
1.2 Mesure d'affinité L^2 de deux densités d'une même famille exponentielle	12
1.3 Mesure d'affinité L^2 de deux densités uniformes	14
1.4 Mesure d'affinité L^2 de deux densités gaussiennes	14
1.4.1 Calcul de cette mesure d'affinité	14
1.4.2 Visualisation de la distance induite par cette mesure d'affinité	16
1.5 Normalité asymptotique du produit scalaire de deux densités gaussiennes	18
1.5.1 Normalité asymptotique de $\langle f^{(n)}, g \rangle$	18
1.5.2 Normalité asymptotique de $\langle f^{(n)}, g^{(n)} \rangle$	24
1.5.3 Conclusion	26
1.6 Normalité asymptotique du vecteur aléatoire $(\langle f^{(n)}, g_q \rangle)$	26
1.7 Étude empirique des vitesses de convergence dans le cas gaussien	28
1.7.1 Convergence de $\langle f^{(n)}, g^{(n)} \rangle$ vers $\langle f, g \rangle$	29
1.7.2 Convergence de $\langle f^{(n)}, f \rangle$ vers la normalité	29
1.7.3 Estimation paramétrique et estimation non paramétrique	32
2 Analyse en Composantes Principales de densités gaussiennes	35
2.1 Hypothèses et position du problème	35

2.2	Solution	36
2.3	Ecriture matricielle de \mathbf{W}	37
2.4	Reconstitution des densités de probabilité	38
2.5	ACP des fonctions caractéristiques	39
2.6	ACP normée et ACP centrée	39
2.6.1	ACP normée	39
2.6.2	ACP centrée	40
2.7	Estimation et convergence	40
2.8	Comparaisons avec STATIS Dual	41
2.8.1	ACP des distributions gaussiennes et STATIS Dual	41
2.8.2	ACP de développements en série de fonctions caractéristiques et STATIS Dual	42
2.9	Conclusion	43
3	Règle de décision probabiliste en analyse discriminante de distributions gaussiennes	45
3.1	Introduction	45
3.1.1	Données et notations	45
3.1.2	Règle bayésienne	46
3.2	Mise en œuvre de la règle bayésienne d'affectation	48
3.2.1	Estimation de la densité gaussienne de la période q	48
3.2.2	Probabilités à priori	48
3.2.3	Première règle d'affectation	48
3.2.4	Seconde règle d'affectation	49
3.2.5	Cas particulier de deux groupes et d'égalité des matrices de variance	49
3.3	Conclusion	51
4	Aspects géométriques de l'analyse discriminante de distributions gaussiennes	53
4.1	Règle d'affectation géométrique	53
4.1.1	Méthode 1	54

4.1.2	Méthode 2	54
4.1.3	Conclusion	55
4.2	Les représentations graphiques classiques de l'analyse discriminante sont-elles possibles?	55
4.2.1	Centrage du nuage des densités	55
4.2.2	Inertie et décomposition	56
4.2.3	Critère optimisé par les représentations graphiques	57
4.3	Propriétés des représentations “sous contraintes”	59
4.3.1	Propriétés des opérateurs \mathbf{V}° et \mathbf{W}°	59
4.3.2	Intérêt des représentations graphiques	60
4.4	Ecriture matricielle de l'analyse discriminante sous contraintes	61
4.4.1	Matrices d'inertie et critère	61
4.4.2	Solution matricielle	62
4.5	Mise en œuvre des calculs	63
4.5.1	Produits scalaires entre densités	63
4.5.2	Matrice M des produits scalaires entre “densités” centrées	63
4.5.3	Matrice V d'inertie totale	63
4.5.4	Matrice B d'inertie inter	64
4.5.5	Matrice W d'inertie intra	64
4.5.6	Coordonnées des densités non datées	65
4.6	Une autre solution matricielle	65
4.7	Conclusion	68

5 Application de l'ACP de densités gaussiennes à quelques processus particuliers 71

5.1	Processus solution d'une équation différentielle stochastique bilinéaire	71
5.1.1	Rappels	71
5.1.2	Exemple	72
5.2	Mouvement brownien	78
5.2.1	Introduction	78

5.2.2	Cas où $p = 1, \mu = 1, \Sigma = 1, t_1 = 1, \forall k, \Delta_k = 1, T = 26$	79
5.2.3	Etude empirique de la convergence de l'ACP	80
5.2.4	Variation de p, μ, Σ	82
5.2.5	Comparaison de l'ACP classique et de l'ACP de densités : cas d'une promenade aléatoire	86
5.2.6	Variation de T et t_1	86
5.2.7	Variation de la suite (Δ_k)	90
5.2.8	Conclusion	93
5.3	Processus d'Ornstein-Uhlenbeck	94
5.3.1	Présentation	94
5.3.2	ACP normée des densités	94
5.4	Données cardiologiques	95
5.4.1	Les données	95
5.4.2	ACP normée des densités	96
5.4.3	Evolution des moyennes temporelles	98
5.4.4	Conclusion	99
5.5	Exemple (Lavit 1988)	101
6	Application de l'analyse discriminante à un problème de datation	103
6.1	Présentation des données	103
6.1.1	Les données	103
6.1.2	Caractère gaussien des données	105
6.1.3	Caractère évolutif des données	108
6.2	Validation empirique des règles d'affectation sur les châteaux datés	112
6.3	Affectation des châteaux et conclusion	112
	Conclusion	119
	Bibliographie	123

Introduction

Les données auxquelles on s'intéresse ici sont du type données ternaires (three-way data):

$$\textit{instants} \times \textit{individus} \times \textit{variables} ,$$

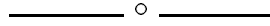
auxquelles sont consacrés de nombreux travaux dont on peut trouver des synthèses (Kroonenberg 1983, Escoufier 1985, Coppi et Bolasco 1989, Kiers 1991), après les deux articles (Tucker 1966, Escoufier 1973) fondateurs de deux approches différentes.

Ces données sont des tableaux $(n_t \times p)$ indicés par t , où à chaque instant t ($t \in \mathcal{T} = \{1, \dots, T\}$) on dispose d'un échantillon de taille n_t d'un vecteur aléatoire à p dimensions: ainsi à chaque instant on observe les mêmes variables quantitatives mais pas nécessairement sur les mêmes individus.

L'objectif est de décrire de façon globale ces données pour en apprécier qualitativement l'évolution: le temps (dans le cas où t fait référence au temps) n'intervenant que comme élément d'interprétation. Ce type de données peut être, par exemple, décrit au moyen de deux analyses distinctes, une portant sur les moyennes, l'autre portant sur les matrices de variance. Notre souci en proposant la méthode dite analyse en composantes principales (ACP) de distributions multidimensionnelles est de disposer d'une analyse globale qui prenne en compte aussi bien les moyennes que les variances / covariances. Cette méthode consiste à associer à chaque tableau t un objet qui est une **densité de probabilité** et d'en faire une ACP à la manière dont procède la méthode STATIS Dual dans sa première étape.

On s'intéressera plus particulièrement au cas où ces densités sont gaussiennes tout en indiquant la voie permettant l'extension de cette analyse à des densités

d'une même famille exponentielle ou à des densités uniformes.



Aux données précédentes on ajoute une variable (qualitative) à Q catégories définie sur l'ensemble des instants, connue sur une partie de ces instants et inconnue sur les autres instants. C'est une situation à laquelle on a été confronté en archéologie à la demande d'un ami, J.M. Rudrauf¹, avec qui nous avons co-dirigé un mémoire d'ingénieur². Sur chaque château (t) J.M. Rudrauf a mesuré p paramètres quantitatifs sur n_t pierres ayant servi à l'édifier. Pour certains de ces châteaux, dits datés, la science historique nous dit la période de sa construction (Y est la variable qui à t associe sa période de construction), par contre pour les autres châteaux l'objectif est de les dater. On a ici un problème de discrimination avec le but de prédire cette variable.

En effet, classiquement, "le but des méthodes de discrimination consiste à prédire une variable qualitative à k catégories à l'aide de p prédicteurs, généralement numériques" (Saporta 1990) mais qui peuvent être qualitatifs (Diday et al 1982, Tomassone et al 1988). Ces méthodes comportent deux étapes :

- une étape d'apprentissage sur un premier ensemble d'individus dont on connaît la catégorie ainsi que les valeurs des p prédicteurs, étape consistant à expliquer la variable qualitative au moyen des p autres variables ;
- une étape de prédiction consistant en l'affectation de chaque individu d'un autre ensemble d'individus dont on ne connaît que les réponses aux p prédicteurs, à une catégorie de la variable qualitative en se basant sur une règle soit de type géométrique, soit de type probabiliste si les p prédicteurs sont gaussiens, conditionnellement à chaque catégorie.

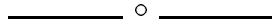
De plus, l'analyse discriminante offre la possibilité de visualiser les individus en optimisant un critère (rapport de l'inertie inter et de l'inertie intra).

¹J.M. Rudrauf. *L'appareil à bossage en grès*. Notes personnelles et communication au séminaire "Les techniques de construction dans l'architecture médiévale en Alsace, 1997".

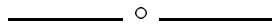
²M. Dlim. *L'analyse en composantes principales : implémentation et application à un problème de datation*. Institut d'informatique, Tizi-Ouzou, 1985.

On va développer cette démarche dans le cas où les “individus” sont des densités de probabilité gaussiennes, éléments d’un espace de Hilbert de dimension infinie, en proposant une règle de chaque type. Matusita (1966) a déjà abordé cette problématique, prolongée par de nombreux travaux incluant même un mélange de variables qualitatives et quantitatives (voir McLachlan 1992 pour une bibliographie exhaustive) et récemment par Bar-Hen et Daudin (1998).

On montre en quoi la visualisation des individus par l’analyse discriminante classique n’est pas possible du fait de la non inversibilité de l’opérateur d’inertie et on propose une solution assez “naturelle” pour lever cette indétermination donnant lieu à une représentation des instants (à partir des densités qui leur sont associées).



On a donc été amené à définir la mesure d’affinité L^2 de deux densités de $L^2(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p}, \nu)$, le produit scalaire de ces deux densités, et à étudier ses propriétés. Cette mesure d’affinité est à la base des analyses factorielles qui seront présentées ici. Nous avons conservé le terme générique d’analyse factorielle couramment utilisé en analyse des données faute d’avoir trouvé un terme plus approprié. En effet dans ces analyses on ne tentera pas d’interpréter les “facteurs” en fonction des données traitées mais on essaiera au travers des représentations fournies par ces analyses de déceler comment se manifeste l’évolution temporelle et comment s’en servir pour faire de la prédiction.



Le premier chapitre est constitué par l’étude de la mesure d’affinité L^2 . Dans un premier temps on la calcule pour quelques types de densités tout en proposant une méthode d’estimation non paramétrique de cette mesure d’affinité pouvant permettre son extension à d’autres types de densités. Dans un deuxième temps on présente le cas gaussien et on établit la normalité asymptotique de l’estimateur paramétrique de cette mesure d’affinité. Enfin dans un dernier temps on illustre

la vitesse de convergence des estimations précédentes par des simulations.

La présentation de l'analyse en composantes principales de distributions gaussiennes multidimensionnelles fait l'objet du second chapitre. On y montre aussi que l'ACP des densités ou des fonctions caractéristiques sont équivalentes et que lorsque le nombre de distributions est fini, cette analyse conduit à un problème matriciel dont la solution est classique. On propose une estimation convergente de cette analyse puis on étudie les différences / ressemblances avec la méthode STATIS dual. On montre en particulier que STATIS dual appliquée à des variables centrées est une ACP de développements limités des fonctions caractéristiques de ces mêmes variables.

Le troisième chapitre utilise le caractère asymptotiquement gaussien de la mesure d'affinité L^2 de deux densités gaussiennes estimées pour proposer l'analyse discriminante de densités gaussiennes. Cette technique, apparentée à l'analyse discriminante à but décisionnel (Saporta 1990), consiste en deux règles d'affectation d'une densité f , estimée au vu d'un échantillon, à une densité parmi Q densités supposées connues. Ces deux règles sont basées sur un critère de type maximum de vraisemblance. On montre de plus que dans un cas particulier une des deux règles est équivalente à une affectation suivant un critère de type distance minimum, cette distance étant celle induite par la mesure d'affinité L^2 .

Ce critère fera la transition vers le quatrième chapitre qui comprend l'analyse discriminante avec règles géométriques. On montre que la recherche d'axes discriminants maximisant le rapport de l'inertie inter et de l'inertie intra mène à une infinité de solutions et qu'en ajoutant une contrainte cela conduit à une solution unique en un certain sens. Cette situation aboutit à un problème matriciel dont on donne la solution. De plus cette solution fournit des représentations graphiques des densités visualisant leurs distances respectives.

Ces analyses seront mises en œuvre dans les deux derniers chapitres.

Le cinquième chapitre est centré sur l'analyse en composantes principales sur données concrètes et sur données de processus gaussiens. Pour les données de processus on traitera plusieurs exemples de processus solutions d'équations différentielles stochastiques, et en particulier le mouvement brownien et le processus

d'Ornstein-Uhlenbeck. L'ACP du mouvement brownien sera particulièrement détaillée et fournira l'occasion de comparer l'ACP de densités et l'ACP classique dans le cas d'une promenade aléatoire. Pour les données concrètes, on appliquera l'ACP de densités à des données fournies par J.L. Bosson et à des données déjà traitées dans (Lavit 1988) par la méthode STATIS.

Enfin le sixième et dernier chapitre avant la conclusion concerne l'analyse discriminante et son application aux données archéologiques fournies par J.M. Rudrauf avec un objectif de datation.

Chapitre 1

Mesure d'affinité L^2 de deux densités de probabilité et propriétés asymptotiques

1.1 Introduction

Considérons deux densités de probabilité f et g sur $(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p})$ de carré intégrable par rapport à une mesure de référence ν sur $(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p})$. Elles sont donc dans l'espace de Hilbert $L^2(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p}, \nu)$, noté H_ν , de produit scalaire $\langle \cdot, \cdot \rangle_\nu$ et de norme associée $\| \cdot \|_\nu$.

Si ν est la mesure de Lebesgue, cet espace est noté $L^2(\mathbb{R}^p)$ et l'indice ν sera omis dans les notations H , $\langle \cdot, \cdot \rangle$ et $\| \cdot \|$.

On appelle mesure d'affinité L^2 entre les densités f et g le produit scalaire de ces deux densités (Qannari 1983) :

$$\langle f, g \rangle_\nu = \int_{\mathbb{R}^p} f(x) g(x) \nu(dx) \quad (1.1)$$

Cette mesure d'affinité bilinéaire et non négative appartient à la classe que Krzanowski a nommé classe des mesures d'affinité de Bhattacharyya en référence à la mesure d'affinité proposée par ce dernier (McLachlan 1992, pp.22-23) :

$$\int_{\mathbb{R}^p} \sqrt{f(x)} \sqrt{g(x)} dx . \quad (1.2)$$

Ce produit scalaire entre \sqrt{f} et \sqrt{g} dans $L^2(\mathbb{R}^p)$ induit la distance de Matusita ou de Hellinger ; l'avantage de cette dernière définition est que toute densité a pour norme 1. La mesure d'affinité L^2 quant à elle présente l'avantage de la linéarité,

ce qui permet le développement d'analyses factorielles de type linéaire et donne la possibilité de l'estimer aisément dans les cas où les densités sont estimées par des combinaisons linéaires (1.4).

Avant d'examiner de façon plus détaillée le cas où f et g sont des densités gaussiennes et ν la mesure de Lebesgue, on procède au calcul de la mesure d'affinité L^2 dans le cas où f et g sont de la même famille exponentielle, puis celui où f et g sont des densités de lois uniformes.

1.2 Mesure d'affinité L^2 de deux densités d'une même famille exponentielle

Rappelons tout d'abord les définitions classiques (pour une référence générale, voir Barndorff-Nielsen 1978). Soient ν une mesure sur \mathbb{R}^p et ϕ une fonction de \mathbb{R}^k dans \mathbb{R}^p , on appelle famille exponentielle engendrée par (ν, ϕ) la famille des lois de probabilité absolument continues par rapport à ν , de densité :

$$\frac{1}{K(\theta)} e^{\langle \theta, \phi(x) \rangle_{\mathbb{R}^p}} \quad \theta \in \Theta$$

où $\langle, \rangle_{\mathbb{R}^p}$ est un produit scalaire sur \mathbb{R}^p et Θ est l'ensemble des valeurs θ telles que :

$$K(\theta) = \int_{\mathbb{R}^p} e^{\langle \theta, \phi(x) \rangle_{\mathbb{R}^p}} \nu(dx)$$

converge.

La mesure d'affinité L^2 de deux densités f et g de cette famille vaut :

$$\frac{1}{K(\theta_f) K(\theta_g)} \int_{\mathbb{R}^p} e^{\langle \theta_f + \theta_g, \phi(x) \rangle_{\mathbb{R}^p}} \nu(dx) . \quad (1.3)$$

La plupart des familles de lois de probabilité classiques sont des cas particuliers de familles exponentielles, et il existe de nombreux processus qui ont la propriété de maintenir à chaque instant leur loi marginale dans une même famille exponentielle (Ycart 1989, 1992a, 1992b).

Sans être exhaustif, le tableau (Tab. 1.1) donne la valeur de (1.3) pour quelques familles de lois.

TAB. 1.1 - Mesures d'affinité pour quelques types de familles exponentielles.

Famille	Mesure de référence (ν)	Mesure d'affinité L^2 ($\langle f, g \rangle_\nu$)
Binomiale $\mathcal{B}(n, p_f)$ $\mathcal{B}(n, p_g)$	$\nu(x) = \binom{n}{x}$ $x \in \{0, 1, \dots, n\}$	$((1 - p_f)(1 - p_g) + p_f p_g)^n$
Poisson $\mathcal{P}(\theta_f)$ $\mathcal{P}(\theta_g)$	$\nu(x) = \frac{1}{x!}$ $x \in \mathbb{N}$	$\frac{e^{\theta_f \theta_g}}{e^{\theta_f + \theta_g}}$
Gamma $G(r_f, a_f)$ $G(r_g, a_g)$	Mesure de Lebesgue sur \mathbb{R}	$\frac{a_f^{r_f+1} a_g^{r_g+1}}{(a_f + a_g)^{r_f+r_g+1}} \frac{\Gamma(r_f+r_g+1)}{\Gamma(r_f+1)\Gamma(r_g+1)}$
Normale $N(\theta_f, \Sigma)$ $N(\theta_g, \Sigma)$	$\frac{d\nu}{dx}(x) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{ \Sigma ^{\frac{1}{2}}} e^{-\frac{1}{2}x'\Sigma^{-1}x}$ $x \in \mathbb{R}^p$	$e^{\theta_f'\Sigma^{-1}\theta_g}$

1.3 Mesure d'affinité L^2 de deux densités uniformes

Hors des familles exponentielles, d'autres cas peuvent avoir un intérêt pratique. Par exemple supposons données f et g les densités des lois de probabilité uniformes sur \mathcal{D}_f et \mathcal{D}_g deux domaines ouverts et bornés de \mathbb{R}^p . On note $\mathbb{1}_{\mathcal{D}}$ l'indicatrice du domaine \mathcal{D} et $vol(\mathcal{D})$ son volume :

$$f = \frac{1}{vol(\mathcal{D}_f)} \mathbb{1}_{\mathcal{D}_f} \quad g = \frac{1}{vol(\mathcal{D}_g)} \mathbb{1}_{\mathcal{D}_g} .$$

La mesure d'affinité L^2 de f et g est :

$$\langle f, g \rangle = \frac{vol(\mathcal{D}_f \cap \mathcal{D}_g)}{vol(\mathcal{D}_f) vol(\mathcal{D}_g)} .$$

L'intérêt pratique du cas uniforme est qu'il permet d'estimer la mesure d'affinité L^2 de deux densités quelconques f et g de $L^2(\mathbb{R}^p)$. En effet ces densités quelconques f et g peuvent être estimées de façon non paramétrique par des combinaisons linéaires \tilde{f} et \tilde{g} de densités uniformes :

$$\tilde{f} = \sum_i \alpha_i \frac{1}{vol(\mathcal{D}_f^i)} \mathbb{1}_{\mathcal{D}_f^i} \quad \tilde{g} = \sum_j \beta_j \frac{1}{vol(\mathcal{D}_g^j)} \mathbb{1}_{\mathcal{D}_g^j} . \quad (1.4)$$

Le produit scalaire $\langle f, g \rangle$ peut donc être estimé par :

$$\langle \tilde{f}, \tilde{g} \rangle = \sum_i \sum_j \alpha_i \beta_j \frac{vol(\mathcal{D}_f^i \cap \mathcal{D}_g^j)}{vol(\mathcal{D}_f^i) vol(\mathcal{D}_g^j)} .$$

On reprendra cette question au paragraphe 1.7.3.

1.4 Mesure d'affinité L^2 de deux densités gaussiennes

1.4.1 Calcul de cette mesure d'affinité

Soient f et g deux densités gaussiennes $N(\mu, \Sigma)$ et $N(m, V)$ respectivement, la mesure d'affinité L^2 de ces deux densités est donnée par la proposition suivante.

Proposition 1.1

$$\langle f, g \rangle = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\Sigma + V|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mu-m)'(\Sigma+V)^{-1}(\mu-m)} . \quad (1.5)$$

On peut obtenir ce produit scalaire en utilisant un calcul déjà fait par Kailath (McLachlan 1992, p.23). On donne ci-dessous un calcul direct.

On note φ_f et φ_g les fonctions caractéristiques associées à f et g . Ces fonctions caractéristiques sont dans $L^2_{\mathbb{C}}(\mathbb{R}^p)$ et le théorème de Plancherel (Vinograd 1987) montre que :

$$\langle \varphi_f, \varphi_g \rangle_{L^2_{\mathbb{C}}(\mathbb{R}^p)} = (2\pi)^p \langle f, g \rangle . \quad (1.6)$$

On a :

$$\langle \varphi_f, \varphi_g \rangle_{L^2_{\mathbb{C}}(\mathbb{R}^p)} = \int_{\mathbb{R}^p} \varphi_f(u) \bar{\varphi}_g(u) du = \int_{\mathbb{R}^p} e^{iu'\mu - \frac{1}{2}u'\Sigma u} e^{-iu'm - \frac{1}{2}u'Vu} du ,$$

soit encore :

$$\int_{\mathbb{R}^p} e^{iu'(\mu-m) - u'(\frac{\Sigma+V}{2})u} du = \int_{\mathbb{R}^p} e^{M(u)} du .$$

On pose :

$$S = \frac{\Sigma + V}{2} . \quad (1.7)$$

La matrice S étant définie positive, on fait le changement de variables :

$$v = S^{\frac{1}{2}}u$$

où $S^{\frac{1}{2}}$ est une matrice telle que : $S^{\frac{1}{2}}S^{\frac{1}{2}} = S$.

En notant a le vecteur $S^{-\frac{1}{2}}(\mu-m)$ de composantes $\alpha_1, \dots, \alpha_p$, la quantité $M(u)$ s'écrit $(iv'a - v'v)$ c'est-à-dire $\sum_{j=1}^p (i\alpha_j v_j - v_j^2)$. L'intégrale précédente devient :

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{j=1}^p e^{-(v_j^2 - i\alpha_j v_j)} |S|^{-\frac{1}{2}} dv_1 \dots dv_p .$$

On a :

$$\forall \alpha \in \mathbb{R}, \quad \int_{-\infty}^{\infty} e^{-(x^2 - i\alpha x)} dx = \sqrt{\pi} e^{-\frac{\alpha^2}{4}} ;$$

ce résultat découlant de :

$$\int_{-\infty}^{\infty} e^{-(x-i\alpha)^2} dx = \sqrt{\pi}$$

qui s'obtient en intégrant la fonction holomorphe $e^{-(z-i\alpha)^2}$ sur le contour rectangulaire du plan complexe délimité par les points A, B, C et D de coordonnées respectives $(-R, 0)$, $(-R, i\alpha)$, $(R, i\alpha)$, $(R, 0)$ et en faisant tendre R vers l'infini.

Ainsi :

$$\langle \varphi_f, \varphi_g \rangle_{L^2_{\mathbb{R}^p}} = |S|^{-\frac{1}{2}} \prod_{j=1}^p (\sqrt{\pi} e^{-\frac{\alpha_j^2}{4}}) = |S|^{-\frac{1}{2}} \pi^{\frac{p}{2}} e^{-\frac{\sum_{j=1}^p \alpha_j^2}{4}},$$

et en remplaçant $\sum_{j=1}^p \alpha_j^2$ par sa valeur $(\mu - m)'S^{-1}(\mu - m)$, il vient :

$$\langle \varphi_f, \varphi_g \rangle = (2\pi)^{\frac{p}{2}} \frac{1}{|\Sigma + V|^{\frac{1}{2}}} e^{-\frac{1}{4}(\mu - m)'S^{-1}(\mu - m)}$$

d'où, compte tenu de (1.6) et de la notation (1.7) :

$$\langle f, g \rangle = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\Sigma + V|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mu - m)'(\Sigma + V)^{-1}(\mu - m)}.$$

□

1.4.2 Visualisation de la distance induite par cette mesure d'affinité

La distance carrée entre les densités de probabilité f et g est obtenue à partir de $\|f - g\|^2$. D'après (1.5), on a :

$$\|f - g\|^2 = \frac{1}{(2\sqrt{\pi})^p} \left(\frac{1}{|\Sigma|^{\frac{1}{2}}} + \frac{1}{|V|^{\frac{1}{2}}} - \frac{2}{|\frac{\Sigma+V}{2}|^{\frac{1}{2}}} e^{-\frac{1}{2}\|\mu - m\|_{(\Sigma+V)^{-1}}^2} \right).$$

où $\|\mu - m\|_{(\Sigma+V)^{-1}}$ désigne la norme du vecteur $(\mu - m)$ dans \mathbb{R}^p muni de la métrique de matrice $(\Sigma + V)^{-1}$.

Cette fonction qu'on ne peut représenter dans le cas général, car dépendant des trois paramètres Σ , V et $(\mu - m)$, sera précisée dans le cas particulier où V vaut $c\Sigma$ ($c > 0$). Elle dépend dans ce cas du paramètre c et de d la distance entre μ et m pour la métrique de matrice Σ^{-1} :

$$\|f - g\|^2 = \frac{1}{(2\sqrt{\pi})^p} \frac{1}{|\Sigma|^{\frac{1}{2}}} \underbrace{\left(1 + \frac{1}{c^{\frac{p}{2}}} - \frac{2}{\left(\frac{1+c}{2}\right)^{\frac{p}{2}}} e^{-\frac{1}{2}\frac{d^2}{1+c}} \right)}_{h(c,d)}$$

Pour $p = 4$, les graphiques (Fig. 1.1) représentent la fonction h sur le pavé $[0.5, 2] \times [0, 4.5]$, ainsi que ses coupes à $(d = 0)$ et $(c = 1)$. Ainsi la distance entre les densités de probabilité apparaît dans le cas gaussien comme une synthèse de deux différences, celle entre les moyennes et celle entre les matrices de variance.

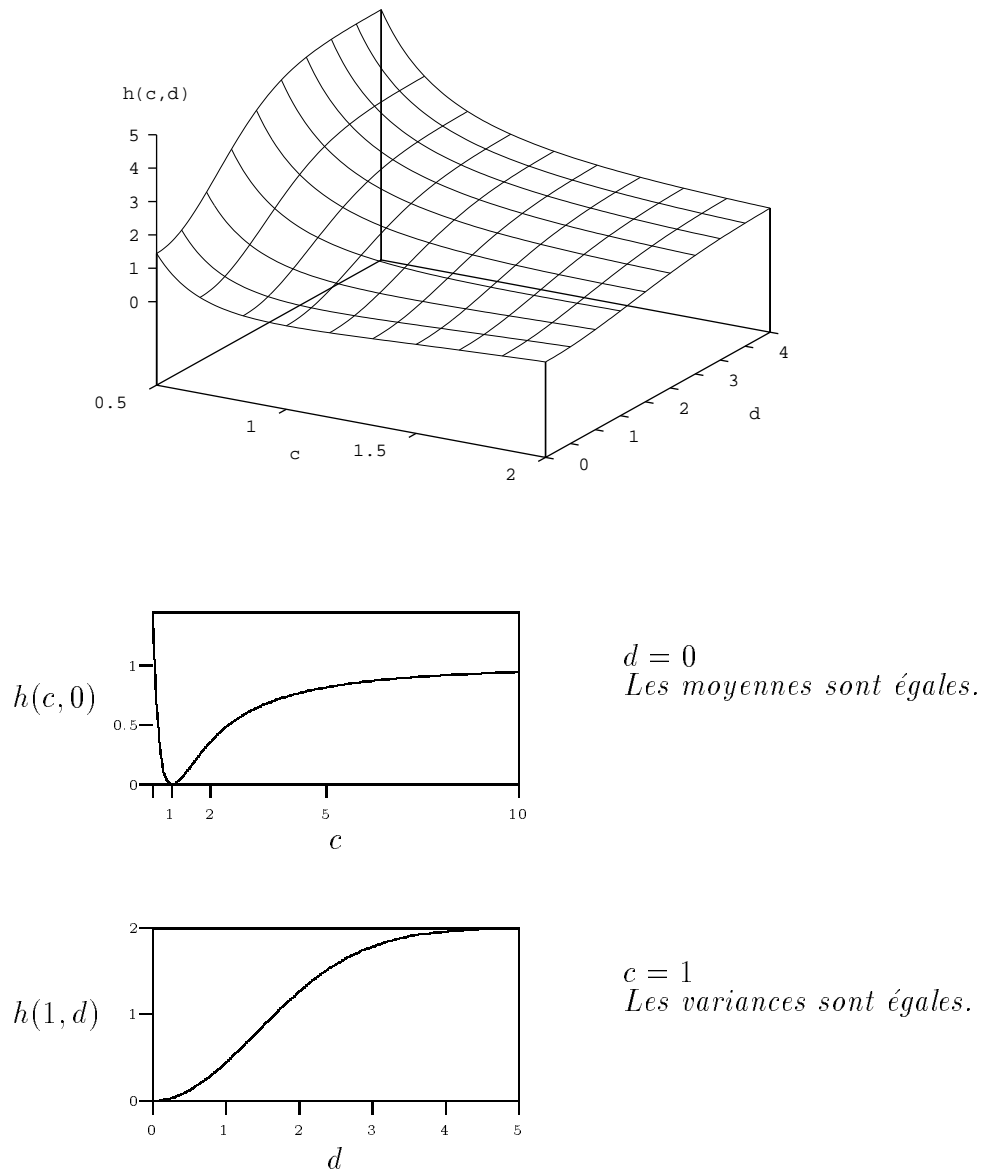


FIG. 1.1 - Graphe de la fonction h et coupes à $d = 0$ et $c = 1$.

1.5 Normalité asymptotique du produit scalaire de deux densités gaussiennes

1.5.1 Normalité asymptotique de $\langle f^{(n)}, g \rangle$

Enoncé

Soient X et Y deux p -vecteurs aléatoires gaussiens $N(\mu, \Sigma)$ et $N(m, V)$ non dégénérés de densités respectives f et g . Le produit scalaire de ces deux densités est donné par la formule (1.5).

Supposons donnés un n -échantillon X_1, \dots, X_n de X et l'estimateur $f^{(n)}$ de f :

$$f^{(n)}(z) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|S|^{\frac{1}{2}}} e^{-\frac{1}{2}(z-\bar{X})'S^{-1}(z-\bar{X})} \quad (1.8)$$

où \bar{X} et S sont les estimateurs du maximum de vraisemblance de μ et Σ :

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \\ S &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' . \end{aligned} \quad (1.9)$$

En notant $\xrightarrow{p.s.}$ la convergence presque sûre, $\xrightarrow{\mathcal{P}}$ la convergence en probabilité et $\xrightarrow{\mathcal{L}}$ la convergence en loi lorsque n croît indéfiniment, on a :

$$\bar{X} \xrightarrow{p.s.} \mu \quad (1.10)$$

$$\sqrt{n}(\bar{X} - \mu) \rightsquigarrow N(0, \Sigma) \quad (1.11)$$

$$S \xrightarrow{p.s.} \Sigma \quad (1.12)$$

$$\sqrt{n} \text{vec}(S - \Sigma) \xrightarrow{\mathcal{L}} N(0, (I + C)(\Sigma \otimes \Sigma)) . \quad (1.13)$$

Dans cette dernière relation (Romain 1997) il s'agit de la loi de Gauss à p^2 dimensions, vec désigne la vectorisation d'une matrice, C désigne la matrice de commutation, I la matrice identité d'ordre p et \otimes le produit de Kronecker (voir Fang et Zhang 1990, p.11 et suivantes).

De (1.10) et (1.12) il vient :

$$\langle f^{(n)}, g \rangle \xrightarrow{p.s.} \langle f, g \rangle . \quad (1.14)$$

Pour simplifier les notations, on notera ψ la fonction qui à un vecteur ν de \mathbb{R}^p et à une matrice carrée d'ordre p définie positive K associe :

$$\psi(\nu, K) = \frac{1}{|K + V|^{\frac{1}{2}}} e^{-\frac{1}{2}(\nu-m)'(K+V)^{-1}(\nu-m)}, \quad (1.15)$$

de sorte que :

$$\psi(\mu, \Sigma) = (2\pi)^{\frac{p}{2}} \langle f, g \rangle$$

et :

$$\psi(\bar{X}, S) = (2\pi)^{\frac{p}{2}} \langle f^{(n)}, g \rangle .$$

De (1.10) et (1.12) il vient :

$$\psi(\bar{X}, S) \xrightarrow{p.s.} \psi(\mu, \Sigma) \quad (1.16)$$

et il existe un voisinage de (μ, Σ) sur lequel ψ est parfaitement définie car l'ensemble des matrices symétriques définies positives constitue un ouvert. De plus ψ est indéfiniment différentiable sur ce voisinage. Pour n suffisamment grand la probabilité que (\bar{X}, S) appartienne à ce voisinage est égale à 1.

Théorème 1.1 *En notant δ le vecteur $(\mu-m)$, , la matrice $(\Sigma+V)^{-1}$, $diag\{\delta\}$ la matrice diagonale des composantes du vecteur δ , $diag\{[,^{-1}]_{ii}^c\}$ la matrice diagonale des cofacteurs ii de la matrice $^{-1}$ et $a[\mu, \Sigma, m, V]$ le nombre :*

$$|, |e^{-\delta'\Gamma\delta} \left(\delta', \Sigma, \delta + 2\text{tr} \left[\left((-, +, \delta\delta', + \frac{|,|}{2}diag\{[,^{-1}]_{ii}^c\} - \frac{1}{2}diag\{\delta\}^2)\Sigma \right)^2 \right] \right)$$

alors :

$$\sqrt{n} \left(\langle f^{(n)}, g \rangle - \langle f, g \rangle \right) \xrightarrow{\mathcal{L}} N \left(0, \frac{a[\mu, \Sigma, m, V]}{(2\pi)^p} \right).$$

En pratique, on aura à considérer l'hypothèse nulle d'égalité des densités f et g .

On notera alors $a[m, V]$ le nombre $a[m, V, m, V]$ et :

$$a[m, V] = \frac{1}{2|2V|} \text{tr} \left[\left(-I + \frac{1}{|2V|}diag\{[2V]_{ii}^c\}V \right)^2 \right], \quad (1.17)$$

car $\delta = 0$ et $^{-1} = \frac{1}{2}V^{-1}$.

Étapes de la démonstration

La démonstration du théorème 1.1 suit les étapes suivantes. On veut montrer que :

$$\sqrt{n} (\psi(\bar{X}, S) - \psi(\mu, \Sigma)) \xrightarrow{\mathcal{L}} N(0, a[\mu, \Sigma, m, V]).$$

Pour cela, on utilise un développement limité à l'ordre 1 de ψ au voisinage de (μ, Σ) .

$$\begin{aligned} \psi(\nu, K) - \psi(\mu, \Sigma) &= \sum_i \frac{\partial \psi}{\partial \nu_i}(\mu, \Sigma) (\nu_i - \mu_i) \\ &+ \sum_i \sum_j \frac{\partial \psi}{\partial K_{ij}}(\mu, \Sigma) (K_{ij} - \Sigma_{ij}) \\ &+ o(\|\nu - \mu\|_p + \|K - \Sigma\|_{p \times p}) \end{aligned}$$

où ν_i (resp. μ_i) est la i -ième composante du vecteur ν (resp. μ), K_{ij} (resp. Σ_{ij}) est le terme ij de la matrice K (resp. Σ), $\|\cdot\|_p$ et $\|\cdot\|_{p \times p}$ sont des normes quelconques respectivement dans l'espace \mathbb{R}^p et l'espace des matrices carrées d'ordre p , o est une fonction négligeable devant x quand x tend vers zéro.

La quantité $\sqrt{n} (\psi(\bar{X}, S) - \psi(\mu, \Sigma))$ se décompose donc en trois termes dont on devra étudier le comportement asymptotique :

$$\begin{aligned} \alpha_1(\bar{X}) &= \sqrt{n} \sum_i \frac{\partial \psi}{\partial \nu_i}(\mu, \Sigma) (\bar{X}_i - \mu_i) \\ \alpha_2(S) &= \sqrt{n} \sum_i \sum_j \frac{\partial \psi}{\partial K_{ij}}(\mu, \Sigma) (S_{ij} - \Sigma_{ij}) \\ \alpha_3(\bar{X}, S) &= \sqrt{n} o(\|\bar{X} - \mu\|_p + \|S - \Sigma\|_{p \times p}) \end{aligned}$$

On montrera successivement les trois résultats suivants.

Lemme 1.1

$$\sqrt{n} \sum_i \frac{\partial \psi}{\partial \nu_i}(\mu, \Sigma) (\bar{X}_i - \mu_i) \rightsquigarrow N(0, (\nabla_\mu \psi)' \Sigma (\nabla_\mu \psi))$$

où $\nabla_\mu \psi$ désigne le gradient de ψ par rapport à ν évalué en (μ, Σ) .

Lemme 1.2

$$\sqrt{n} \sum_i \sum_j \frac{\partial \psi}{\partial K_{ij}}(\mu, \Sigma) (S_{ij} - \Sigma_{ij}) \xrightarrow{\mathcal{L}} N(0, 2\text{tr} [((\nabla_\Sigma \psi) \Sigma)^2])$$

où $\nabla_\Sigma \psi$ désigne la matrice du gradient de ψ par rapport à K évalué en (μ, Σ) .

Lemme 1.3

$$\sqrt{n} o(\|\bar{X} - \mu\|_p + \|S - \Sigma\|_{p \times p}) \xrightarrow{\mathcal{P}} 0.$$

Etant dans un cadre gaussien, les statistiques \bar{X} et S sont indépendantes ; les statistiques intervenant dans les deux premiers lemmes sont donc indépendantes et comme la somme de deux variables gaussiennes indépendantes est encore gaussienne, on obtient :

$$\alpha_1(\bar{X}) + \alpha_2(S) + \alpha_3(\bar{X}, S) \xrightarrow{\mathcal{L}} N\left(0, (\nabla_\mu \psi)' \Sigma (\nabla_\mu \psi) + 2\text{tr} \left[((\nabla_\Sigma \psi) \Sigma)^2 \right] \right).$$

Il ne restera plus qu'à effectuer le calcul de la variance.

Lemme 1.4

$$(\nabla_\mu \psi)' \Sigma (\nabla_\mu \psi) + 2\text{tr} \left[((\nabla_\Sigma \psi) \Sigma)^2 \right] = a[\mu, \Sigma, m, V]$$

Démonstrations des lemmes

On donne dans l'ordre la démonstration des lemmes 1.1 à 1.4.

Démonstration du lemme 1.1 Il découle immédiatement de la normalité du vecteur $\sqrt{n}(\bar{X} - \mu)$ et de l'écriture matricielle de α_1 :

$$\alpha_1(\bar{X}) = (\nabla_\mu)' (\sqrt{n}(\bar{X} - \mu)).$$

□

Démonstration du lemme 1.2 On a :

$$\alpha_2(S) = (\text{vec}(\nabla_\Sigma \psi))' (\sqrt{n} \text{vec}(S - \Sigma)).$$

On utilise la normalité asymptotique du vecteur $\sqrt{n} \text{vec}(S - \Sigma)$ (1.13) qu'on multiplie à gauche par le vecteur $(\text{vec}(\nabla_\Sigma \psi))'$. On obtient la normalité asymptotique souhaitée, avec pour variance :

$$(\text{vec}(\nabla_\Sigma \psi))' ((I + C) \Sigma \otimes \Sigma) (\text{vec}(\nabla_\Sigma \psi))$$

que l'on calcule ci-après :

$$\begin{aligned}
&= (\text{vec}(\nabla_{\Sigma}\psi))'(I + C) (\text{vec}(\Sigma(\nabla_{\Sigma}\psi))) \\
&\qquad\qquad\qquad (\text{Fang et Zhang 1990, p.12 (1.4.4)}) \\
&= (\text{vec}(\nabla_{\Sigma}\psi))'(\text{vec}(\Sigma(\nabla_{\Sigma}\psi)\Sigma) + \text{vec}((\Sigma(\nabla_{\Sigma}\psi)\Sigma)')) \\
&\qquad\qquad\qquad (\text{Fang et Zhang 1990, p.13 (1.4.7)}) \\
&= (\text{vec}(\nabla_{\Sigma}\psi))'(2 \text{vec}(\Sigma(\nabla_{\Sigma}\psi)\Sigma)) \\
&\qquad\qquad\qquad ((\nabla_{\Sigma}\psi) \text{ est symétrique}) \\
&= 2 \text{tr}[(\nabla_{\Sigma}\psi)\Sigma(\nabla_{\Sigma}\psi)\Sigma] .
\end{aligned}$$

□

Démonstration du lemme 1.3 On écrit $\sqrt{n} o(\|\bar{X} - \mu\|_p + \|S - \Sigma\|_{p \times p})$ sous la forme :

$$Z_n = \sqrt{n} (\|\bar{X} - \mu\|_p + \|S - \Sigma\|_{p \times p}) \frac{o(\|\bar{X} - \mu\|_p + \|S - \Sigma\|_{p \times p})}{\|\bar{X} - \mu\|_p + \|S - \Sigma\|_{p \times p}} . \quad (1.18)$$

La loi de $\sqrt{n} \|\bar{X} - \mu\|_p$ ne dépend pas de n (1.11).

De (1.13) et de la continuité de la fonction norme on déduit la convergence en loi de $\sqrt{n} \|S - \Sigma\|_{p \times p}$.

Ces deux variables étant indépendantes leur somme converge en loi.

La variable aléatoire $\|\bar{X} - \mu\|_p + \|S - \Sigma\|_{p \times p}$ converge presque sûrement vers zéro.

Comme $o(x)/x$ tend vers zéro quand x tend vers zéro, il vient que :

$$\frac{o(\|\bar{X} - \mu\|_p + \|S - \Sigma\|_{p \times p})}{\|\bar{X} - \mu\|_p + \|S - \Sigma\|_{p \times p}}$$

converge presque sûrement vers zéro et donc converge en probabilité.

Ainsi la suite de terme général Z_n (1.18) apparaît comme le produit d'une suite de variables aléatoires convergente en loi et d'une suite de variables aléatoires convergente en probabilité vers zéro, d'où le résultat souhaité. □

Démonstration du lemme 1.4 Il s'agit essentiellement de calculer $\nabla_{\mu}\psi$ et $\nabla_{\Sigma}\psi$. On rappelle l'expression de $\psi(\nu, K)$:

$$\psi(\nu, K) = \frac{1}{|K + V|^{\frac{1}{2}}} e^{-\frac{1}{2}(\nu-m)'(K+V)^{-1}(\nu-m)} .$$

Le gradient de ψ par rapport à ν évalué en (μ, Σ) est :

$$\begin{aligned}\nabla_{\mu}\psi &= -\frac{1}{2}\psi(\mu, \Sigma) \nabla_{\mu}(\nu - m)'(K + V)^{-1}(\nu - m) \\ &= -\frac{1}{2}\psi(\mu, \Sigma) 2(\Sigma + V)^{-1}(\mu - m) .\end{aligned}$$

Avec les notations du théorème 1.1 ce gradient devient :

$$\nabla_{\mu}\psi = -|\cdot|, \frac{1}{2}e^{-\frac{1}{2}\delta'\Gamma\delta}, \delta . \quad (1.19)$$

Le gradient de ψ par rapport à Σ évalué en (μ, Σ) est :

$$\nabla_{\Sigma}\psi = -\frac{1}{2}\frac{1}{|\Sigma + V|}\nabla_{\Sigma}|K + V| \psi(\mu, \Sigma) - \frac{1}{2}\psi(\mu, \Sigma) \nabla_{\Sigma}(\nu - m)'(K + V)^{-1}(\nu - m) . \quad (1.20)$$

Le gradient $\nabla_{\Sigma}|K + V|$ est donné dans (Fang et Zhang 1990, p.17 (5)).

$$\begin{aligned}\nabla_{\Sigma}|K + V| &= 2|\Sigma + V|(\Sigma + V)^{-1} - \text{diag}\{[\Sigma + V]_{ii}^c\} \\ &= \frac{2}{|\cdot|}, -\text{diag}\{[\cdot, \cdot^{-1}]_{ii}^c\}.\end{aligned} \quad (1.21)$$

Pour calculer le gradient $\nabla_{\Sigma}[(\nu - m)'(K + V)^{-1}(\nu - m)]$, on calcule la dérivée partielle de $(\nu - m)'(K + V)^{-1}(\nu - m)$ par rapport à la composante K_{ij} de K .

On désigne par E_{ij} la matrice d'ordre p dont tous les termes sont nuls hormis le terme d'indice ij qui vaut 1 .

$$\begin{aligned}\frac{\partial[(\nu - m)'(K + V)^{-1}(\nu - m)]}{\partial K_{ij}} &= (\nu - m)'\frac{\partial[(K + V)^{-1}]}{\partial K_{ij}}(\nu - m) \\ &= \begin{cases} -(\nu - m)'(K + V)^{-1}E_{ii}(K + V)^{-1}(\nu - m) \\ \qquad \qquad \qquad \text{si } i = j \\ -2(\nu - m)'(K + V)^{-1}E_{ij}(K + V)^{-1}(\nu - m). \\ \qquad \qquad \qquad \text{si } i \neq j \end{cases}\end{aligned}$$

Cette dernière égalité est l'extension de la formule de dérivation (Fang et Zhang 1990, p.15 (8)) au cas d'une matrice symétrique.

En remplaçant (ν, K) par (μ, Σ) et avec les notations du théorème 1.1, on obtient le terme ij de la matrice du gradient $\nabla_{\Sigma}[(\nu - m)'(K + V)^{-1}(\nu - m)]$:

$$\left[\nabla_{\Sigma}[(\nu - m)'(K + V)^{-1}(\nu - m)]\right]_{ij} = \begin{cases} -\delta', E_{ii}, \delta & \text{si } i = j \\ -2\delta', E_{ij}, \delta & \text{si } i \neq j . \end{cases}$$

En remarquant que δ' , E_{ij} , δ est égal au terme d'indice ij de la matrice $(, \delta)(, \delta)'$, il vient :

$$\nabla_{\Sigma}[(\nu - m)'(K + V)^{-1}(\nu - m)] = -2, \delta\delta', - (diag\{\delta\})^2. \quad (1.22)$$

De (1.21) et (1.22) on obtient (1.20) la matrice $\nabla_{\Sigma}\psi$ recherchée :

$$-\frac{1}{2}|, |(\frac{2}{|,|}, - diag\{[,^{-1}]_{ii}^c\}) \psi(\mu, \Sigma) - \frac{1}{2}\psi(\mu, \Sigma) (-2, \delta\delta', - (diag\{\delta\})^2),$$

soit :

$$|, |\frac{1}{2} e^{-\frac{1}{2}\delta'\Gamma\delta} \left(-, +, \delta\delta', + \frac{|,|}{2}diag\{[,^{-1}]_{ii}^c\} - \frac{1}{2}(diag\{\delta\})^2 \right). \quad (1.23)$$

Le calcul de $(\nabla_{\mu}\psi)'\Sigma(\nabla_{\mu}\psi) + 2tr [((\nabla_{\Sigma}\psi)\Sigma)^2]$ découle de (1.19) et (1.23) et vaut :

$$|, |e^{-\delta'\Gamma\delta} \left(\delta', \Sigma, \delta + 2tr \left[\left((-, +, \delta\delta', + \frac{|,|}{2}diag\{[,^{-1}]_{ii}^c\} - \frac{1}{2}diag\{\delta\}^2)\Sigma \right)^2 \right] \right),$$

d'où le lemme 1.4. \square

1.5.2 Normalité asymptotique de $\langle f^{(n)}, g^{(n)} \rangle$

Soient f et g les densités des lois $N(\mu, \Sigma)$ et $N(m, V)$, X_1, \dots, X_n un échantillon de loi parente $N(\mu, \Sigma)$ indépendant d'un échantillon Y_1, \dots, Y_n de loi parente $N(m, V)$. On estime les paramètres μ , Σ , m et V respectivement par \bar{X} , S_x , \bar{Y} et S_y (1.9) puis les densités f et g par $f^{(n)}$ et $g^{(n)}$ (1.8).

On cherche à obtenir la distribution asymptotique de $\langle f^{(n)}, g^{(n)} \rangle$. On simplifie l'énoncé en supposant les tailles d'échantillon égales. On pourrait les supposer proportionnelles à une même quantité n tendant vers l'infini, ce qui modifierait la variance asymptotique.

Théorème 1.2

$$\sqrt{n}(\langle f^{(n)}, g^{(n)} \rangle - \langle f, g \rangle) \xrightarrow{\mathcal{L}} N(0, b[\mu, \Sigma, m, V])$$

où $b[\mu, \Sigma, m, V]$ vaut :

$$\frac{|\Gamma|e^{-\delta'\Gamma\delta}}{(2\pi)^p} \left(\delta', \Sigma, \delta + 2tr \left[\left((-, +, \delta\delta', + \frac{|\Gamma|}{2}diag\{[,^{-1}]_{ii}^c\} - \frac{1}{2}diag\{\delta\}^2)\Sigma \right)^2 \right] \right) \\ + \\ \frac{|\Gamma|e^{-\delta'\Gamma\delta}}{(2\pi)^p} \left(\delta', V, \delta + 2tr \left[\left((-, +, \delta\delta', + \frac{|\Gamma|}{2}diag\{[,^{-1}]_{ii}^c\} - \frac{1}{2}diag\{\delta\}^2)V \right)^2 \right] \right)$$

La démonstration peut s'obtenir soit de manière analogue à celle du théorème 1.1 soit directement comme suit.

On a :

$$\langle f^{(n)} - f, g^{(n)} - g \rangle = \langle f^{(n)}, g^{(n)} \rangle - \langle f^{(n)}, g \rangle - \langle f, g^{(n)} \rangle + \langle f, g \rangle ,$$

d'où on obtient :

$$\begin{aligned} \sqrt{n} (\langle f^{(n)}, g^{(n)} \rangle - \langle f, g \rangle) &= \sqrt{n} \langle f^{(n)} - f, g^{(n)} - g \rangle \\ &+ \sqrt{n} (\langle f^{(n)}, g \rangle - \langle f, g \rangle) \\ &+ \sqrt{n} (\langle f, g^{(n)} \rangle - \langle f, g \rangle). \end{aligned}$$

Les lois asymptotiques des variables aléatoires :

$$\begin{aligned} \sqrt{n} (\langle f^{(n)}, g \rangle - \langle f, g \rangle) , \\ \sqrt{n} (\langle f, g^{(n)} \rangle - \langle f, g \rangle) \end{aligned}$$

sont données par le théorème 1.1. De plus ces deux variables étant indépendantes la loi limite de leur somme est donc gaussienne de paramètres la somme de leurs paramètres respectifs ; sa moyenne est nulle et sa variance est :

$$\frac{a[\mu, \Sigma, m, V]}{(2\pi)^p} + \frac{a[m, V, \mu, \Sigma]}{(2\pi)^p} . \quad (1.24)$$

La variable aléatoire $\sqrt{n} \langle f^{(n)} - f, g^{(n)} - g \rangle$ se réécrit :

$$\frac{1}{\sqrt{n}} \langle \sqrt{n} (f^{(n)} - f), \sqrt{n} (g^{(n)} - g) \rangle . \quad (1.25)$$

Les deux variables aléatoires $\sqrt{n} (f^{(n)} - f)$ et $\sqrt{n} (g^{(n)} - g)$ sont indépendantes et convergentes en loi, leur produit scalaire est donc convergent en loi. En multipliant ce produit scalaire par $\frac{1}{\sqrt{n}}$ qui tend vers 0 on est assuré de la convergence en loi de (1.25) vers 0.

Ainsi la loi asymptotique de $\sqrt{n} (\langle f^{(n)}, g^{(n)} \rangle - \langle f, g \rangle)$ est donc gaussienne de moyenne nulle et de variance $b[\mu, \Sigma, m, V]$ égale à (1.24), soit :

$$\begin{aligned} &\frac{|\Gamma| e^{-\delta' \Gamma \delta}}{(2\pi)^p} \left(\delta', \Sigma, \delta + 2 \text{tr} \left[\left((-, +, \delta \delta', + \frac{|\Gamma|}{2} \text{diag}\{[,^{-1}]_{ii}^e\} - \frac{1}{2} \text{diag}\{, \delta\}^2) \Sigma \right)^2 \right] \right) \\ &+ \\ &\frac{|\Gamma| e^{-\delta' \Gamma \delta}}{(2\pi)^p} \left(\delta', V, \delta + 2 \text{tr} \left[\left((-, +, \delta \delta', + \frac{|\Gamma|}{2} \text{diag}\{[,^{-1}]_{ii}^e\} - \frac{1}{2} \text{diag}\{, \delta\}^2) V \right)^2 \right] \right) \end{aligned} \quad \square$$

1.5.3 Conclusion

Comme application immédiate du théorème 1.1 on peut envisager l'utilisation de la statistique

$$T = \langle f^{(n)}, g \rangle$$

comme statistique de test de l'hypothèse

$$H_0 : f = g$$

contre

$$H_1 : f \neq g .$$

La loi asymptotique de T sous H_0 est donnée par le théorème 1.1 d'où un calcul immédiat de la région de rejet.

De manière analogue le théorème 1.2 conduit à la définition d'un test de comparaison de deux échantillons gaussiens. La statistique de test est alors :

$$\langle f^{(n)}, g^{(n)} \rangle .$$

Nous n'avons pas exploré cette voie, assez éloignée des objectifs des techniques factorielles développées ici. Il n'est d'ailleurs pas certain qu'elle apporte un plus par rapport aux tests habituels.

1.6 Normalité asymptotique du vecteur aléatoire ($\langle f^{(n)}, g_q \rangle$)

Pour tout q de $\{1, \dots, Q\}$, soit g_q la densité de la loi $N(m_q, V_q)$ et $f^{(n)}$ l'estimateur de la densité f de $N(\mu, \Sigma)$. Les résultats obtenus au paragraphe précédent peuvent être étendus au cas vectoriel. Les démonstrations étant très proches, on ne les détaillera pas.

Théorème 1.3 *Quand n tend vers l'infini, la suite des vecteurs aléatoires :*

$$\sqrt{n} \left((\langle f^{(n)}, g_1 \rangle, \dots, \langle f^{(n)}, g_Q \rangle) - (\langle f, g_1 \rangle, \dots, \langle f, g_Q \rangle) \right)$$

converge en loi vers $N(0, \frac{1}{(2\pi)^p} A[\mu, \Sigma, (m.), (V.)])$ et en posant pour tous les indices l, j, k de $\{1, \dots, Q\}$:

$$\begin{aligned} \delta_l &= (\mu - m_l), \\ ,_l &= (\Sigma + V_l)^{-1}, \\ \Delta_l &= - ,_l + ,_l \delta_l \delta_l' ,_l + \frac{\Gamma_l}{2} \text{diag}\{[,_l^{-1}]_{ii}^c\} - \frac{1}{2} (\text{diag}\{ ,_l \delta_l\})^2, \end{aligned}$$

le terme général $A[\mu, \Sigma, (m.), (V.)]_{jk}$ de la matrice de variance est égal à :

$$| ,_j ,_k | \frac{1}{2} e^{-\frac{1}{2}(\delta_j' \Gamma_j \delta_j + \delta_k' \Gamma_k \delta_k)} (\delta_j' ,_j \Sigma ,_k \delta_k + 2\text{tr}[\Delta_j \Sigma \Delta_k \Sigma]) \quad (1.26)$$

Par analogie avec (1.15), on pose :

$$\Psi(\nu, K) = (\psi_1(\nu, K), \dots, \psi_Q(\nu, K))$$

de sorte que :

$$\Psi(\mu, \Sigma) = (2\pi)^{\frac{p}{2}} \langle f^{(n)}, g_1 \rangle, \dots, \langle f^{(n)}, g_Q \rangle.$$

En écrivant le développement limité de Ψ au voisinage de (μ, Σ) , on obtient :

$$\begin{aligned} \sqrt{n} \begin{pmatrix} \psi_1(\nu, K) - \psi_1(\mu, \Sigma) \\ \vdots \\ \psi_Q(\nu, K) - \psi_Q(\mu, \Sigma) \end{pmatrix} &= \begin{pmatrix} (\nabla_\mu \psi_1)' \\ \vdots \\ (\nabla_\mu \psi_Q)' \end{pmatrix} \sqrt{n} (\nu - \mu) \\ &+ \begin{pmatrix} (\text{vec}[\nabla_\Sigma \psi_1])' \\ \vdots \\ (\text{vec}[\nabla_\Sigma \psi_Q])' \end{pmatrix} \sqrt{n} \text{vec}(K - \Sigma) \\ &+ \sqrt{n} \begin{pmatrix} o_1(\|\nu - \mu\|_p + \|K - \Sigma\|_{p \times p}) \\ \vdots \\ o_Q(\|\nu - \mu\|_p + \|K - \Sigma\|_{p \times p}) \end{pmatrix} \end{aligned}$$

Lorsque ν et K sont remplacés par \bar{X} et S et que n tend vers l'infini, ce vecteur aléatoire converge en loi vers une gaussienne centrée et de matrice de covariance $A[\mu, \Sigma, (m.), (V.)]$ de terme général jk égal à :

$$(\nabla_\mu \psi_j)' \Sigma (\nabla_\mu \psi_k) + 2\text{tr}[(\nabla_\Sigma \psi_j) \Sigma (\nabla_\Sigma \psi_k) \Sigma]. \quad (1.27)$$

Les gradients (1.19) et (1.23) permettent d'obtenir le terme général (1.27). Il est égal à (1.26). \square

On utilisera le théorème 1.3 dans le cas particulier où $f = g_q$. La matrice de variance asymptotique $A[m_q, V_q, (m.), (V.)]$ que l'on notera $A[m_q, V_q]$, s'obtient en remplaçant dans (1.26) :

$$\begin{aligned} \delta_l \quad \text{par} \quad \delta_{ql} &= (m_q - m_l), \\ ,_l \quad \text{par} \quad ,_{ql} &= (V_q + V_l)^{-1}, \\ \Delta_l \quad \text{par} \quad \Delta_{ql} &= - ,_{ql} + ,_{ql} \delta_{ql} \delta_{ql}' + \frac{|E_{ql}|}{2} \text{diag}\{[,_{ql}^{-1}]_{ii}^c\} - \frac{1}{2} (\text{diag}\{ ,_{ql} \delta_{ql} \})^2. \end{aligned} \tag{1.28}$$

Dans le cas où les matrices de variance V_q sont toutes égales à V ce qui entraîne que les densités g_q sont toutes de même norme, alors les matrices $,_l$ sont toutes égales à $\frac{V^{-1}}{2}$ et l'écriture de la loi conjointe ne dépend plus que des différences entre moyennes δ_{ql} .

1.7 Etude empirique des vitesses de convergence dans le cas gaussien

Cette étude concerne la mesure d'affinité L^2 de deux densités gaussiennes f et g lorsque ces deux densités sont estimées au vu d'un échantillon de taille n . Cette étude sera menée dans trois directions différentes.

La première a pour objet de montrer la vitesse de la convergence de la statistique $\langle f^{(n)}, g^{(n)} \rangle$ vers $\langle f, g \rangle$ où $f^{(n)}$ et $g^{(n)}$ sont les estimateurs paramétriques (1.8) de f et g respectivement.

La seconde montre la vitesse de convergence de $\sqrt{n} (\langle f^{(n)}, f \rangle - \langle f, f \rangle)$ vers la normalité.

Enfin la troisième compare les vitesses de convergence des deux modes d'estimation paramétrique et non paramétrique.

1.7.1 Convergence de $\langle f^{(n)}, g^{(n)} \rangle$ vers $\langle f, g \rangle$

On considère deux variables aléatoires X et Y de lois $N(\mu, \Sigma)$ et $N(m, V)$ avec :

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$m = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad V = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} .$$

La mesure d'affinité L^2 de leurs densités notées f et g vaut d'après (1.5) :

$$\langle f, g \rangle = \frac{1}{2\pi} \frac{1}{\sqrt{5}} e^{-\frac{3}{10}} \approx 0.0527 . \quad (1.29)$$

On simule n réalisations indépendantes x_1, \dots, x_n de X et y_1, \dots, y_n de Y . On calcule les moyennes empiriques (\bar{x}, \bar{y}) et les variances empiriques (s_x, s_y) . Les densités $\hat{f}^{(n)}$ et $\hat{g}^{(n)}$ des lois $N(\bar{x}, s_x)$ sont des estimations de f et g . Leur mesure d'affinité L^2 est une estimation de (1.29).

Les graphiques (Fig. 1.2) donnent les résultats obtenus pour n variant de 5 à 40 dans deux cas. Le premier graphique est celui où les simulations démarrent toujours avec la même graine et donc la réalisation x_n (resp. y_n) est ajoutée aux réalisations précédentes de X (resp. Y). Le second graphique est celui où toutes les simulations sont indépendantes les unes des autres.

Si la dimension des vecteurs aléatoires p est petite (ici $p = 2$), il apparaît que pour une taille d'échantillon relativement faible, de l'ordre de la dizaine, l'estimation de $\langle f, g \rangle$ par $\langle \hat{f}^{(n)}, \hat{g}^{(n)} \rangle$ est assez raisonnable.

1.7.2 Convergence de $\langle f^{(n)}, f \rangle$ vers la normalité

Soit f la densité de X variable aléatoire gaussienne centrée réduite, on étudie la rapidité de la convergence en loi de $\langle f^{(n)}, f \rangle$ vers une gaussienne en procédant comme suit.

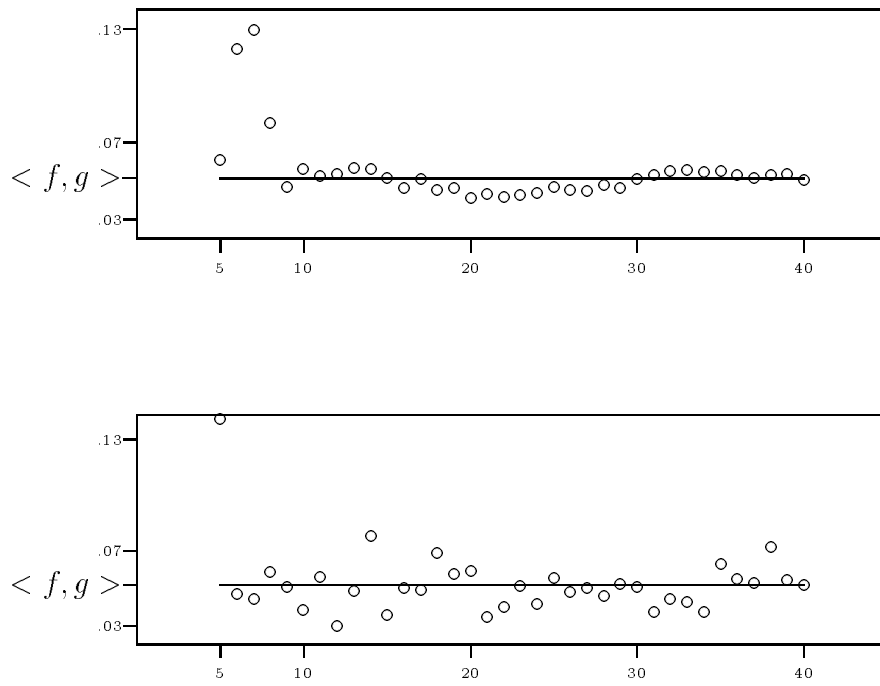


FIG. 1.2 - Représentation de $\langle \hat{f}^{(n)}, \hat{g}^{(n)} \rangle$ en fonction de n .

- Faire pour $n = 5, 10, 15, 20, 25, 30, 35, 40$:
 - Répéter 200 fois :
 - Répéter 1000 fois :
 - Simulation de n réalisations de X .
 - Calcul de la moyenne empirique (\bar{x}), de la variance empirique (s) et de la mesure d'affinité L^2 des densités des lois $N(\bar{x}, s)$ et $N(0, 1)$.
 - finRépéter
 - Test de normalité des 1000 réalisations de $\langle f^{(n)}, f \rangle$ en utilisant le test de Kolmogorov¹ avec un risque de 5% (Saporta 1990, p.338).
 - finRépéter
 - Compter le nombre de fois où l'hypothèse de normalité est rejetée.
- Afficher les résultats.

Ces résultats sont donnés dans le tableau suivant :

n	5	10	15	20	25	30	35	40
Nombre de rejets	101	31	21	21	14	16	16	11
Pourcentage de rejets	50.5	15.5	10.5	10.5	7	8	8	5.5

Le procédé décrit précédemment signifie que si l'hypothèse de normalité de la statistique $\langle f^{(n)}, f \rangle$ est acceptable, les pourcentages de rejets par le test d'ajustement de Kolmogorov devraient être de l'ordre du risque de première espèce considéré, c'est-à-dire 5%.

¹Dans la version antérieure le test du χ^2 avait été utilisé. Suite à une suggestion de Gilbert Saporta ce test a été remplacé par celui de Kolmogorov.

A partir des résultats obtenus, on peut donc accepter la normalité de la statistique $\langle f^{(n)}, f \rangle$ pour n de l'ordre de la trentaine.

1.7.3 Estimation paramétrique et estimation non paramétrique

On considère la densité f de X de loi normale centrée réduite. On simule n réalisations x_1, \dots, x_n de X , on estime f de deux manières différentes puis on mesure l'affinité de ces estimations avec f :

- Paramétriquement : on calcule comme précédemment la moyenne empirique (\bar{x}), la variance empirique (s). On note $\hat{f}^{(n)}$ la densité de la loi $N(\bar{x}, s)$ puis on calcule $\langle \hat{f}^{(n)}, f \rangle$.
- Non paramétriquement : on note $\tilde{f}^{(n)}$ l'estimation de f par l'histogramme des n observations x_1, \dots, x_n . L'histogramme est constitué de K classes D_1, \dots, D_K de même amplitude h égale à $(x_{\max} - x_{\min})/K$. La densité ainsi estimée s'écrit :

$$\tilde{f}^{(n)} = \sum_{k=1}^K \frac{f_k}{h} \mathbf{1}_{D_k} ,$$

où les f_k sont les fréquences relatives des classes. En notant F la fonction de répartition de X , la mesure d'affinité L^2 entre $\tilde{f}^{(n)}$ et f est égale à :

$$\langle \tilde{f}^{(n)}, f \rangle = \sum_{k=1}^K \frac{f_k}{h} (F(x_{\min} + kh) - F(x_{\min} + (k-1)h)) .$$

Les valeurs de $\langle \hat{f}^{(n)}, f \rangle$ et de $\langle \tilde{f}^{(n)}, f \rangle$ sont des estimations de $\langle f, f \rangle$, la norme carrée de f , qui est égale à :

$$\langle f, f \rangle = \frac{1}{2\sqrt{\pi}} \approx 0.282 .$$

Les résultats obtenus pour $K = 20$ et n variant de 5 à 40 sont reproduits sur les graphiques (Fig. 1.3). Le premier graphique est le cas où les simulations à n fixé sont démarrées avec la même graine. Le second graphique est obtenu par des simulations indépendantes les unes des autres.

Tout d'abord on constate que les résultats obtenus ne contredisent pas ce qui a été dit au paragraphe 1.7.1 à propos de la taille d'échantillon minimale raisonnable

pour estimer la mesure d'affinité L^2 de deux densités gaussiennes.

D'autre part on constate que les deux méthodes d'estimation conduisent à des résultats sensiblement équivalents.

Au paragraphe 1.3, on a abordé la question de l'estimation de la mesure d'affinité L^2 de deux densités quelconques (non nécessairement gaussiennes) de $L^2(\mathbb{R}^p)$ en les estimant de manière non paramétrique par des histogrammes. Dans ce qui précède on a montré que dans le cas gaussien les estimations paramétrique et non paramétrique sont équivalentes bien qu'on ait utilisé l'estimation non paramétrique dite "naïve" (Silverman 1986, pp.11-12).

On peut bien évidemment estimer ces densités quelconques par noyaux gaussiens et l'estimation de la mesure d'affinité L^2 peut s'en déduire facilement du fait de la bilinéarité de cette mesure d'affinité.

Bien que nous n'ayons pas exploré cette voie nous pensons qu'elle peut s'avérer fructueuse pour étendre les analyses descriptives (analyse en composantes principales et analyse discriminante géométrique) qui seront développées dans les chapitres suivants (chap. 2 et chap. 4).

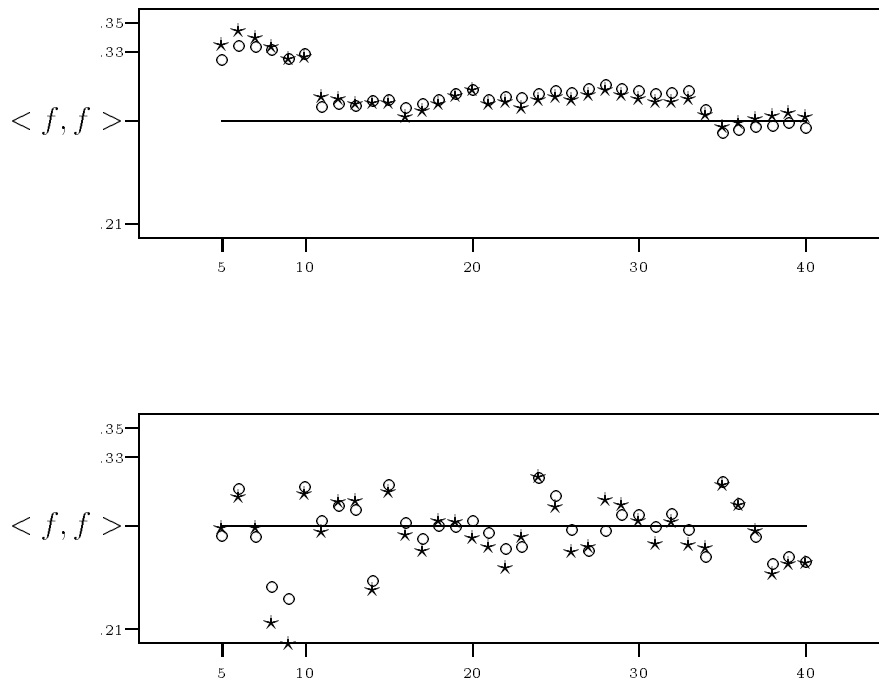


FIG. 1.3 - Représentation de $\langle \hat{f}^{(n)}, f \rangle$ (par \circ) et de $\langle \tilde{f}^{(n)}, f \rangle$ (par \star).

Chapitre 2

Analyse en Composantes Principales de densités gaussiennes

2.1 Hypothèses et position du problème

Soient X_1, \dots, X_T des vecteurs aléatoires tous de distribution de Gauss non dégénérée à p dimensions, de moyennes μ_1, \dots, μ_T et de matrices de variance $\Sigma_1, \dots, \Sigma_T$ respectivement ; les densités de probabilité f_1, \dots, f_T :

$$f_t(x) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_t)' \Sigma_t^{-1}(x-\mu_t)}$$

sont de carré intégrable ; elles constituent donc un nuage \mathcal{F} dans l'espace de Hilbert $L^2(\mathbb{R}^p)$, noté H , de produit scalaire $\langle \cdot, \cdot \rangle$ et de norme associée $\| \cdot \|$.

L'objectif de la méthode proposée étant d'obtenir une représentation approchée de ce nuage, en première étape on cherche g_1 dans H , de norme unité, tel que la quantité :

$$\sum_{t=1}^T \|\mathbf{P}_{g_1}(f_t) - f_t\|^2$$

soit minimale, \mathbf{P}_{g_1} désignant le projecteur orthogonal sur le sous-espace vectoriel de H engendré par g_1 .

Ce critère des moindres carrés est encore équivalent à la maximisation de :

$$\sum_{t=1}^T \|\mathbf{P}_{g_1}(f_t)\|^2 . \tag{2.1}$$

Puis on itère sous contraintes d'orthonormalité : à l'étape k , on cherche g_k dans

$L^2(\mathbb{R}^p)$, de norme unité, orthogonal à g_1, \dots, g_{k-1} , tel que la quantité :

$$\sum_{t=1}^T \|\mathbf{P}_{g_k}(f_t) - f_t\|^2$$

soit minimale.

Le problème ainsi posé est un problème d'ACP, dont la solution est détaillée ci-dessous. Pour l'écrire, nous adoptons le cadre formel de l'ACP d'opérateur compact proposé par (Dauxois et Pousse, 1976) en adaptant les calculs et résultats aux données précédentes.

2.2 Solution

Soit \mathbf{U} l'opérateur de \mathbb{R}^T , muni du produit scalaire classique, dans H défini par :

$$\mathbf{U}u = \sum_{t=1}^T u_t f_t ;$$

en identifiant les espaces de Hilbert séparables H et \mathbb{R}^T à leur dual respectif, l'opérateur adjoint \mathbf{U}^* de \mathbf{U} est :

$$\begin{aligned} \mathbf{U}^* : H &\longrightarrow \mathbb{R}^T \\ g &\longmapsto \mathbf{U}^*g = (\langle f_1, g \rangle_H, \dots, \langle f_T, g \rangle_H) \end{aligned} \quad (2.2)$$

car :

$$\langle \mathbf{U}u, g \rangle = \sum_{t=1}^T u_t \langle f_t, g \rangle = \langle u, \mathbf{U}^*g \rangle_{\mathbb{R}^T} ,$$

$\langle , \rangle_{\mathbb{R}^T}$ désignant le produit scalaire usuel dans \mathbb{R}^T .

Avec ces notations, on remarque que la quantité (2.1) à maximiser s'écrit $\|\mathbf{U}^*g_1\|_{\mathbb{R}^T}^2$; en effet g_1 étant de norme unité :

$$\mathbf{P}_{g_1}(f_t) = \langle f_t, g_1 \rangle g_1$$

et donc :

$$\sum_{t=1}^T \|\mathbf{P}_{g_1}(f_t)\|^2 = \sum_{t=1}^T \langle f_t, g_1 \rangle^2 = \|\mathbf{U}^*g_1\|_{\mathbb{R}^T}^2 .$$

Le problème posé revient donc à chercher g de norme unité qui maximise $\|\mathbf{U}^*g\|_{\mathbb{R}^T}^2$ puis à itérer sous contraintes d'orthonormalité. Comme :

$$\|\mathbf{U}^*g\|_{\mathbb{R}^T}^2 = \langle \mathbf{U}^*g, \mathbf{U}^*g \rangle_{\mathbb{R}^T} = \langle g, \mathbf{U} \circ \mathbf{U}^*g \rangle ,$$

la solution est obtenue en faisant l'analyse spectrale de l'opérateur V égal à $\mathbf{U} \circ \mathbf{U}^*$ qui est autoadjoint, positif et de rang fini, ou encore de l'opérateur $\mathbf{U}^* \circ \mathbf{U}$ noté \mathbf{W} qui a les mêmes valeurs propres non nulles que \mathbf{V} ; de plus si u de \mathbb{R}^T est vecteur propre normé de \mathbf{W} associé à la valeur propre non nulle λ alors :

$$g = \frac{\mathbf{U}u}{\sqrt{\lambda}} \quad (2.3)$$

est un vecteur propre normé de \mathbf{V} associé à la même valeur propre λ .

2.3 Ecriture matricielle de \mathbf{W}

Si e_1, \dots, e_T désigne la base canonique de \mathbb{R}^T , chaque $\mathbf{U}^* f_t$ s'écrira comme combinaison linéaire de ces vecteurs, et dans cette base l'endomorphisme \mathbf{W} de \mathbb{R}^T a pour matrice, notée aussi W :

$$W = (\langle f_s, f_t \rangle)_{(s,t) \in T \times T} \quad (2.4)$$

car :

$$\mathbf{W}e_t = \mathbf{U}^* \circ \mathbf{U}e_t = \mathbf{U}^* f_t = \sum_{s=1}^T \langle f_s, f_t \rangle e_s .$$

Le terme général $\langle f_s, f_t \rangle$ de la matrice W est, d'après (1.5), égal à :

$$\langle f_s, f_t \rangle = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\Sigma_s + \Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mu_s - \mu_t)'(\Sigma_s + \Sigma_t)^{-1}(\mu_s - \mu_t)}. \quad (2.5)$$

Les éléments propres normalisés $(\lambda_1, u_1), \dots, (\lambda_K, u_K)$ de W correspondant aux valeurs propres non nulles et rangées dans l'ordre décroissant permettent d'obtenir en utilisant l'expression (2.3), les vecteurs propres g_1, \dots, g_K de \mathbf{V} .

Ces fonctions de $L^2(\mathbb{R}^p)$ sont des combinaisons linéaires des densités de probabilité gaussiennes f_t mais ne sont pas des densités de probabilité sauf éventuellement g_1 . En effet deux densités de probabilité donc positives ne peuvent être orthogonales pour le produit scalaire de $L^2(\mathbb{R}^p)$.

Quant à l'éventualité que g_1 soit une densité, elle pourrait se produire si par exemple la somme des composantes de u_1 qui sont nécessairement de même signe, vaut $\sqrt{\lambda_1}$; en effet la matrice W a tous ses éléments positifs, son premier vecteur propre u_1 aura donc des composantes toutes de même signe (Théorème de Frobenius) que l'on choisit positif, et g_1 calculé par (2.3) apparaît comme une combinaison convexe de densités de probabilité.

2.4 Reconstitution des densités de probabilité

Pour tout t la densité de probabilité f_t peut se décomposer suivant le système $(g_k)_{k=1, \dots, K}$:

$$f_t = \sum_{k=1}^K \langle f_t, g_k \rangle g_k . \quad (2.6)$$

La coordonnée de chaque f_t suivant g_k étant $\langle f_t, g_k \rangle$, ces coordonnées sont donc les composantes du vecteur $\mathbf{U}^* g_k$ (2.2), égal à $\sqrt{\lambda_k} u_k$ (2.3), et donc (2.6) devient :

$$f_t = \sum_{k=1}^K \sqrt{\lambda_k} u_{kt} g_k , \quad (2.7)$$

où u_{kt} désigne la t -ième composante du vecteur u_k .

On peut obtenir une représentation approchée du nuage \mathcal{F} sur un sous-espace de dimension réduite en tronquant la décomposition précédente et calculer les aides à l'interprétation (Volle, 1981). Ces aides sont les suivantes :

- La qualité globale de l'ACP : elle se mesure par la somme des proportions d'inertie expliquée par les premiers axes retenus, chaque axe k expliquant une proportion égale à $\frac{\lambda_k}{tr[W]}$ où $tr[W]$ désigne la trace de la matrice W et vaut $\sum_{k=1}^K \lambda_k$ qui est l'inertie totale du nuage par rapport à l'origine.

- La qualité de représentation de f_t suivant g_k : elle se mesure par le rapport

$$\frac{\|\mathbf{P}_{g_k}(f_t)\|^2}{\|f_t\|^2} .$$

D'après (2.5), $\|f_t\|^2$ vaut :

$$\frac{1}{(2\sqrt{\pi})^p |\Sigma_t|^{\frac{1}{2}}} , \quad (2.8)$$

d'où :

$$\frac{\|\mathbf{P}_{g_k}(f_t)\|^2}{\|f_t\|^2} = \lambda_k (u_{kt})^2 (2\sqrt{\pi})^p |\Sigma_t|^{\frac{1}{2}} .$$

- Les contributions relatives à l'inertie : la valeur propre λ_k mesurant l'inertie (par rapport à l'origine) du nuage obtenu par projection de \mathcal{F} sur l'axe k . On mesure l'importance de chaque f_t dans la détermination du facteur g_k par sa contribution relative à cette inertie qui vaut $(u_{kt})^2$.

2.5 ACP des fonctions caractéristiques

Plutôt que de représenter la variable X_t par sa densité de probabilité f_t dans $L^2(\mathbb{R}^p)$, il aurait été possible de la représenter par sa fonction caractéristique φ_t dans l'espace de Hilbert $L^2_{\mathbb{C}}(\mathbb{R}^p)$. Cette ACP conduit aux mêmes représentations. En effet le théorème de Plancherel (Vinograd 1987) montre que :

$$\langle \varphi_s, \varphi_t \rangle = (2\pi)^p \langle f_s, f_t \rangle ; \quad (2.9)$$

ainsi la matrice W correspondante à cette ACP est égale à celle définie en (2.4) au coefficient $(2\pi)^p$ près.

2.6 ACP normée et ACP centrée

La présentation de l'ACP précédente a considéré le nuage des densités de probabilité sans aucune transformation.

2.6.1 ACP normée

En fait il est possible de réduire ce nuage en normant à 1 au sens de la norme L^2 les fonctions f_t , c'est-à-dire en considérant les fonctions \overline{f}_t (ces fonctions ne sont plus en général des densités de probabilité) :

$$\overline{f}_t = \frac{f_t}{\|f_t\|},$$

soit :

$$\overline{f}_t(x) = \frac{1}{\pi^{\frac{p}{4}}} \frac{1}{|\Sigma_t|^{\frac{1}{4}}} e^{-\frac{1}{2}(x-\mu_t)' \Sigma_t^{-1}(x-\mu_t)}.$$

L'ACP dite normée conduit à diagonaliser la matrice \overline{W} de terme général \overline{W}_{st} égal à :

$$\begin{aligned} \overline{W}_{st} &= \langle \overline{f}_s, \overline{f}_t \rangle \\ &= 2^{\frac{p}{2}} \frac{|\Sigma_s|^{\frac{1}{4}} |\Sigma_t|^{\frac{1}{4}}}{|\Sigma_s + \Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mu_s - \mu_t)' (\Sigma_s + \Sigma_t)^{-1} (\mu_s - \mu_t)}. \end{aligned} \quad (2.10)$$

Cette normalisation conserve dans H les angles entre les densités mais déforme leurs distances.

Si on considère leur mesure d'affinité de Bhattacharyya (1.2). Elle vaut (McLach-

lan 1992, p.23) :

$$2^{\frac{p}{2}} \frac{|\Sigma_s|^{\frac{1}{4}} |\Sigma_t|^{\frac{1}{4}}}{|\Sigma_s + \Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{4}(\mu_s - \mu_t)'(\Sigma_s + \Sigma_t)^{-1}(\mu_s - \mu_t)}$$

qui est très proche de $\langle \bar{f}_s, \bar{f}_t \rangle$ calculé ci-dessus (2.10).

2.6.2 ACP centrée

Une autre transformation possible, qui respecte les distances, est le centrage du nuage \mathcal{F} des densités f_t en opérant une translation amenant son centre de gravité

$$\frac{1}{T} \sum_{t=1}^T f_t ,$$

noté f_G , sur l'origine de l'espace H . Ces nouvelles fonctions f_t° sont égales à :

$$f_t^\circ = f_t - f_G .$$

La matrice W° à diagonaliser pour obtenir l'ACP centrée est obtenue à partir de W . En effet le terme général W_{st}° de W° étant égal à $\langle f_s^\circ, f_t^\circ \rangle$, son calcul se déduit facilement de (2.5).

2.7 Estimation et convergence

En pratique on ne connaît pas les paramètres μ_t et Σ_t de la distribution (gaussienne) de la variable aléatoire X_t ($t = 1, \dots, T$); si pour tout t on dispose d'un n_t -échantillon X_{t1}, \dots, X_{tn_t} de X_t , on estime ces paramètres par \bar{X}_t et S_t les estimateurs du maximum de vraisemblance respectivement de μ_t et Σ_t (1.9).

On note $n = \inf n_t$ la plus petite taille d'échantillon et $W^{(n)}$ l'estimateur de W obtenu en remplaçant, dans le terme général W_{ts} (2.5) de W , les paramètres μ_t , μ_s , Σ_s et Σ_t par leur estimateur respectif. La convergence presque sûre de ces estimateurs lorsque n croît indéfiniment assure la convergence presque sûre de chaque terme $W_{ts}^{(n)}$ de la matrice $W^{(n)}$ vers le terme respectif W_{ts} de W et donc la convergence uniforme presque sûre de $W^{(n)}$ vers W .

Ainsi l'ACP des densités estimées obtenue par l'analyse spectrale de $W^{(n)}$ est convergente (Dauxois et al. 1982) et la représentation de ces densités est une bonne approximation de la représentation des densités parentes. Au paragraphe

5.2.3 on donnera une idée de cette convergence en présentant l'ACP de T densités parfaitement connues et l'ACP des estimations de ces mêmes densités par la méthode du maximum de vraisemblance sur des échantillons simulés.

2.8 Comparaisons avec STATIS Dual

2.8.1 ACP des distributions gaussiennes et STATIS Dual

Si les données dont on dispose sont telles qu'elles sont présentées en introduction, la méthode STATIS Dual (L'hermier des Plantes 1976, Glaçon 1981, Lavit 1988) peut les décrire en utilisant soit les matrices de variance (Σ_t), soit les matrices de corrélation (R_t); cependant cette description ne tient pas compte des moyennes (μ_t) des variables observées, dans le calcul du compromis à l'étape de l'interstructure. La description des moyennes nécessite une étude séparée. La méthode proposée reste dans l'esprit de la première étape de STATIS Dual, la différence réside dans les objets associés à chaque tableau : ici l'objet associé est f_t , la densité de probabilité de la distribution $\mathcal{N}(\mu_t, \Sigma_t)$ ce qui conduit à diagonaliser la matrice W de terme général (2.5), tandis que dans STATIS Dual l'objet associé est Σ_t (ou R_t) ce qui conduit à diagonaliser la matrice D de terme général :

$$D_{st} = \text{tr}(\Sigma_s \Sigma_t) . \quad (2.11)$$

Si on dispose d'un programme informatique de mise en oeuvre de STATIS Dual, la réalisation de l'ACP de distributions gaussiennes multidimensionnelles ne nécessite que l'adjonction d'un module permettant de calculer l'inverse et le déterminant d'une matrice symétrique définie positive afin d'évaluer la matrice à diagonaliser de terme général W_{ts} (2.5), le coefficient $(2\pi)^{\frac{p}{2}}$ étant bien évidemment inutile; le reste du programme reste sans changement.

Si on utilise le premier facteur principal u_1 de l'ACP des densités de probabilité pour obtenir la densité de probabilité compromis :

$$f_c = \sum_{t=1}^T \alpha_t f_t ,$$

la moyenne et la matrice de variance compromis sont respectivement égales à :

$$\begin{aligned}\mu_c &= \sum_{t=1}^T \alpha_t \mu_t , \\ \Sigma_c &= \sum_{t=1}^T \alpha_t (\Sigma_t + (\mu_t - \mu_c)(\mu_t - \mu_c)') .\end{aligned}$$

On rappelle que $\sum_{t=1}^T \alpha_t$ vaut 1 car les T composantes u_{11}, \dots, u_{1T} du vecteur u_1 sont toutes de même signe et on prend α égal à $\frac{u_1}{\sum_{t=1}^T u_{1t}}$.

2.8.2 ACP de développements en série de fonctions caractéristiques et STATIS Dual

La fonction caractéristique φ_t du vecteur aléatoire X_t dont les composantes seront notées X_{t1}, \dots, X_{tp} se décompose au voisinage de zéro (formule de Taylor à l'ordre q) comme suit :

$$\varphi_t(z) = 1 + \sum_{r=1}^q \frac{i^r}{r!} \sum_{j_1=1}^p \dots \sum_{j_r=1}^p z_{j_1} \dots z_{j_r} E[X_{tj_1} \dots X_{tj_r}] + o(\|z\|^q) .$$

Pour $q = 2$ et si le vecteur X_t est centré ($\mu_t = 0$), cela donne :

$$\varphi_t(z) = 1 - \frac{1}{2} \underbrace{\sum_{j=1}^p \sum_{k=1}^p z_j z_k E[X_{tj} X_{tk}]}_{\text{accolade}} + o(\|z\|^2)$$

La quantité soulignée par l'accolade se développe en :

$$\begin{aligned}& z_1^2 E[X_{t1}^2] + z_1 z_2 E[X_{t1} X_{t2}] + \dots + z_1 z_p E[X_{t1} X_{tp}] \\ & + z_2 z_1 E[X_{t2} X_{t1}] + z_2^2 E[X_{t2}^2] + \dots + z_2 z_p E[X_{t2} X_{tp}] \\ & \dots \\ & + z_p z_1 E[X_{tp} X_{t1}] + z_p z_2 E[X_{tp} X_{t2}] + \dots + z_p^2 E[X_{tp}^2] .\end{aligned}$$

A tout développement en série d'ordre 2 au voisinage de zéro d'une fonction caractéristique centrée, on peut associer un vecteur Φ_t de \mathbb{R}^{p^2} dont les composantes sont les coefficients des polynômes du second degré

$$z_j z_k , \quad 1 \leq j, k \leq p ,$$

les polynômes étant répétés uniquement pour faciliter la numérotation et les écritures.

Si \mathbb{R}^{p^2} est muni du produit scalaire $\langle \cdot, \cdot \rangle_{\mathbb{R}^{p^2}}$ de matrice identité alors :

$$\langle \Phi_s, \Phi_t \rangle_{\mathbb{R}^{p^2}} = \frac{1}{4} \sum_{j=1}^p \sum_{k=1}^p E[X_{sj}X_{sk}] E[X_{tj}X_{tk}]$$

ce qui est le terme général de la matrice à diagonaliser si on procède à l'ACP des T "individus" Φ_1, \dots, Φ_T de \mathbb{R}^{p^2} . Ce terme est encore égal au quart de la trace (2.11) du produit des matrices de covariance Σ_s et Σ_t :

$$\langle \Phi_s, \Phi_t \rangle_{\mathbb{R}^{p^2}} = \frac{1}{4} \text{tr}(\Sigma_s \Sigma_t) = \frac{1}{4} D_{st} .$$

Ainsi si les vecteurs X_1, \dots, X_T sont centrés, STATIS Dual est équivalente à l'ACP des développements en série d'ordre 2 de leurs fonctions caractéristiques dans \mathbb{R}^{p^2} muni de la métrique identité; STATIS Dual serait donc dans le cas centré une forme d'approximation d'une ACP de fonctions caractéristiques qui peut être rapprochée de celle définie au paragraphe (2.5), équivalente à l'ACP des densités, mais qui n'est cependant pas la même.

2.9 Conclusion

Dans ce chapitre, on a principalement proposé deux analyses en composantes principales : celle des densités de probabilité, équivalente à celle des fonctions caractéristiques, et celle des développements limités de fonctions caractéristiques de variables centrées, équivalente à Statis dual sur matrices de variance. On a présenté aussi la variante normée et la variante centrée de la première analyse. Cependant, d'autres analyses descriptives sont possibles, car les opérations de centrage et de normalisation peuvent aussi être opérées sur les données initiales. Sans développer outre mesure cet aspect on peut dire que le choix de telle ou telle méthode est conditionné par la réponse à la question :

quelle évolution veut-on visualiser?

Celle des moyennes et variances / covariances? Celle des moyennes et corrélations?

Celle des variances / covariances? Celle des corrélations? Celle des moyennes?

Au chapitre 5 on illustrera l'ACP normée de densités en centrant notre étude sur l'évolution des moyennes et variances / covariances. Le choix de l'ACP normée est guidé par la volonté de faciliter la lecture comparée des différentes représentations

et présente l'avantage de visualiser les qualités de représentation des densités.

Chapitre 3

Règle de décision probabiliste en analyse discriminante de distributions gaussiennes

3.1 Introduction

3.1.1 Données et notations

\mathcal{T} et \mathcal{T}' sont deux ensembles finis de cardinaux respectifs T et T' :

$$\mathcal{T} = \{1, \dots, T\}$$

$$\mathcal{T}' = \{T + 1, \dots, T + T'\}.$$

Pour tout t de $\mathcal{T} \cup \mathcal{T}'$, f_t est un vecteur de l'espace de Hilbert séparable H muni de son produit scalaire noté \langle, \rangle .

Les vecteurs f_t ($t \in \mathcal{T}$) constituent un nuage \mathcal{F} de points dans H qu'on partitionne en Q sous-nuages $\mathcal{F}_1, \dots, \mathcal{F}_Q$ au vu d'une variable Y , qualitative à Q modalités, définie sur \mathcal{T} ; la variable Y engendre aussi une partition de \mathcal{T} en Q classes $\mathcal{T}_1, \dots, \mathcal{T}_Q$, de cardinaux respectifs T_1, \dots, T_Q .

Les centres de gravité G_1, \dots, G_Q de $\mathcal{F}_1, \dots, \mathcal{F}_Q$, définis par:

$$\forall q \in \{1, \dots, Q\} \quad G_q = \frac{1}{T_q} \sum_{t \in \mathcal{T}_q} f_t; \quad (3.1)$$

constituent eux aussi un nuage \mathcal{G} de H .

3.1.2 Règle bayésienne

De façon générale la règle bayésienne d'affectation à une catégorie q considère les probabilités conditionnelles :

$$P[Y = q | X = o] \quad (q = 1, \dots, Q) , \quad (3.2)$$

où o est une observation de la variable aléatoire X . On affecte l'individu observé à la catégorie q qui réalise le maximum de ces probabilités. Les calculer par la formule de Bayes suppose qu'on connaisse les probabilités $P[Y = q]$ et la densité conditionnelle $d_X^{Y=q}$ de X sachant $[Y = q]$.

En pratique, on aura à effectuer une transformation de la variable X en $Z = \varphi(X)$. On calcule alors des probabilités du type :

$$P[Y = q | Z = \varphi(o)] \quad (q = 1, \dots, Q) . \quad (3.3)$$

Leur calculabilité est subordonnée à la connaissance des probabilités à priori et des densités conditionnelles $d_{\varphi(X)}^{Y=q}$ de Z sachant $[Y = q]$.

Le choix de φ dépend du contexte. Il peut être dicté :

1. Par des considérations liées aux données pour donner du sens aux observations ou à l'activité de modélisation. Par exemple un PNB (produit national brut), un taux d'inflation ... ne s'observent pas directement mais sont le résultat d'une transformation φ d'observations diverses o . Dans le contexte des applications qui font l'objet de la dernière partie de cette thèse, la fonction φ associe une densité de probabilité estimée à un tableau de données.
2. Par des considérations d'ordre mathématique. Par exemple la transformation logarithme utilisée en biologie qui permet de rendre les données traitées gaussiennes. C'est ce qu'on fera dans le traitement des données cardiologiques au paragraphe 5.4.2.
3. Par des considérations liées à l'analyse discriminante comme par exemple la position d'une fonction $\xi(o)$ de l'observation o par rapport à Q centres de gravité G_1, \dots, G_Q des Q catégories de Y . Cette position peut être mesurée

par le vecteur $\varphi(o)$ égal à :

$$(\varphi_1(o), \dots, \varphi_Q(o)) = (\langle \xi(o), G_1 \rangle, \dots, \langle \xi(o), G_Q \rangle) .$$

C'est le point de vue qu'on adopte dans ce chapitre.

On introduira une variante de ce qui précède consistant à considérer les probabilités :

$$P[Y = q | Z_q = \varphi_q(o)] \quad (q = 1, \dots, Q). \quad (3.4)$$

Elles sont plus simples à calculer et c'est d'ailleurs ce qui justifie cette variante. Un autre aspect est la simplicité de l'interprétation puisqu'elle mesure la vraisemblance que l'observation o soit dans la catégorie q en ne regardant que sa position par rapport au centre de gravité G_q .

Cependant en procédant ainsi, on perd l'interprétation "maximum de vraisemblance" de la décision prise (la somme des probabilités (3.4) n'est pas nécessairement égale à 1).

On peut tout de même justifier cette procédure par analogie avec la règle d'affectation géométrique de l'analyse discriminante classique qui calcule les distances aux centres de gravité en utilisant des métriques locales. On n'affecte pas l'observation o à la classe qui réalise :

$$\min_q D(o, G_q)$$

mais à celle qui réalise :

$$\min_q D_q(o, G_q) .$$

Ce type de règle est censé tenir compte non seulement de la position de o par rapport aux centres de classe mais aussi de la forme de chaque classe (Saporta 1990, p.417).

3.2 Mise en œuvre de la règle bayésienne d'affectation

3.2.1 Estimation de la densité gaussienne de la période q

Pour tout t de \mathcal{T} , on dispose d'un échantillon χ_t de taille n_t du vecteur aléatoire X_t supposé gaussien de densité f_t . Le modèle que nous postulons est que les densités $\{f_t, t \in \mathcal{T}_q\}$ d'une même classe q sont toutes égales ; ainsi le centre de gravité G_q est une densité gaussienne de paramètres (m_q, V_q) . On l'estime en remplaçant ses paramètres par leurs estimations du maximum de vraisemblance calculées sur l'échantillon de taille N_q :

$$N_q = \sum_{t \in \mathcal{T}_q} n_t$$

qui est la réunion de tous les échantillons χ_t ($t \in \mathcal{T}_q$).

3.2.2 Probabilités à priori

Pour tout q de $\{1, \dots, Q\}$ on note π_q la probabilité d'appartenance à la catégorie q :

$$\pi_q = P[Y = q].$$

L'estimation de ces probabilités dépend du contexte des données et on peut trouver dans (Mardia 1979, p.304, McLachlan 1992, p.9, Celeux et Nakache 1992, p.19) des méthodes d'estimation. En l'absence de toute connaissance a priori sur le phénomène étudié, on retiendra soit l'égalité de ces probabilités ($\pi_q = 1/Q$), soit la proportionnalité aux effectifs des classes ($\pi_q = T_q/T$).

3.2.3 Première règle d'affectation

On affecte une observation (dans notre cas un instant ou sa densité f) à la catégorie qui réalise :

$$\max_q \pi_q \mathbf{L}_q(z), \quad (3.5)$$

où :

$$z = (\langle \hat{f}, G_1 \rangle \dots \langle \hat{f}, G_Q \rangle)'$$

\hat{f} désignant l'estimation de f à partir d'un échantillon de taille n , c'est-à-dire la densité de la loi $N(\bar{x}, s)$ et \mathbf{L}_q est la densité de la loi donnée par (1.28); si \vec{G}_q note le vecteur (colonne) de \mathbb{R}^Q :

$$\vec{G}_q = (\langle G_q, G_1 \rangle \dots \langle G_q, G_Q \rangle)'$$

l'expression (3.5) devient :

$$\max_q \pi_q \frac{1}{|A[m_q, V_q]|^{\frac{1}{2}}} e^{-\frac{n(2\pi)^p}{2} (z - \vec{G}_q)' A[m_q, V_q]^{-1} (z - \vec{G}_q)} . \quad (3.6)$$

3.2.4 Seconde règle d'affectation

En rappelant que les justifications à cette variante ont été abordées en introduction à ce chapitre, la démarche est identique à la précédente et sous l'hypothèse que f est dans la catégorie q , la loi de $\langle f^{(n)}, G_q \rangle$ est obtenue directement à partir du théorème 1.1. Ainsi la densité f dont on a une réalisation \hat{f} est affectée à la catégorie :

$$\arg \max_q \pi_q \ell_q(z_q) \quad (3.7)$$

où z_q est le nombre $\langle \hat{f}, G_q \rangle$ et ℓ_q est la densité de la loi $N(\|G_q\|^2, \frac{a[m_q, V_q]}{n(2\pi)^p})$ et donc (3.7) devient :

$$\arg \max_q \pi_q \frac{1}{a[m_q, V_q]^{\frac{1}{2}}} e^{-\frac{1}{2} \frac{n(2\pi)^p}{a[m_q, V_q]} (\langle \hat{f}, G_q \rangle - \|G_q\|^2)^2} . \quad (3.8)$$

Cette seconde variante sera étudiée dans le cas particulier d'égalité des probabilités a priori $(\pi_q)_q$ et des matrices de variance $(V_q)_q$.

3.2.5 Cas particulier de deux groupes et d'égalité des matrices de variance

L'égalité des matrices de variance ($V_1 = V_2 = V$), entraîne, d'une part l'égalité des normes des centres de gravité ($\|G_1\| = \|G_2\| = \|\mathbf{G}\|$), et d'autre part l'égalité des nombres $a[m_1, V_1]$ et $a[m_2, V_2]$. Ainsi, en passant au logarithme l'expression

(3.8) devient :

$$\arg \min_q (\langle \hat{f}, G_q \rangle - \|\mathbf{G}\|^2)^2 \quad (3.9)$$

Cette règle est intéressante par les analogies qui peuvent être faites avec l'analyse discriminante linéaire.

Proposition 3.1 *Dans le cas d'égalité des probabilités à priori et des matrices de variance ($V_q = V$, $q = 1, 2$), si $|s| > |V|$ alors (3.7) est équivalente à :*

$$\arg \min_q \|\hat{f} - G_q\|. \quad (3.10)$$

Puisque $|s| > |V|$, on déduit de (2.8) que :

$$\|\hat{f}\| < \|\mathbf{G}\| ,$$

et l'inégalité de Schwarz donne :

$$\langle \hat{f}, G_q \rangle \leq \|\hat{f}\| \|G_q\| < \|\mathbf{G}\|^2 .$$

Les deux nombres $\langle \hat{f}, G_q \rangle - \|\mathbf{G}\|^2$ ($q = 1, 2$) étant donc négatifs, il vient que (3.9) est équivalent à :

$$\arg \max_q \langle \hat{f}, G_q \rangle . \quad (3.11)$$

La mesure d'affinité L^2 entre deux densités étant positive, on en déduit que (3.11) est équivalent à (3.10), car :

$$\|\hat{f} - G_q\|^2 = \|\hat{f}\|^2 - 2 \langle \hat{f}, G_q \rangle + \|G_q\|^2 .$$

Ayant déjà montré l'équivalence de (3.9) et de (3.7), ceci termine la démonstration de la proposition. \square

On remarquera que maximiser les produits scalaires (3.11) entre \hat{f} et les centres de classe $(G_q)_q$ revient à maximiser les cosinus et donc à minimiser l'angle entre \hat{f} et les $(G_q)_q$. Il reste à voir comment opère cette règle sur les observations initiales.

Proposition 3.2 *Dans le cas d'égalité des probabilités à priori et des matrices de variance ($V_q = V$, $q = 1, 2$), si $|s| > |V|$ alors (3.7) est équivalente à :*

$$\arg \min_q (\bar{x} - m_q)'(s + V)^{-1}(\bar{x} - m_q) . \quad (3.12)$$

On a vu que (3.7) est équivalente à (3.11). En utilisant l'expression du produit scalaire entre densités (2.5) que l'on passe au logarithme l'équivalence de (3.7) et (3.12) est démontrée. \square

Ces deux propositions font apparaître les analogies avec l'analyse discriminante linéaire. D'une part on est passé d'une règle probabiliste (3.7) à une règle géométrique (3.10); d'autre part l'affectation de f à un groupe dépend de la distance (pour la métrique $(s + V)^{-1}$) de la moyenne (\bar{x}) à la moyenne de chaque groupe.

3.3 Conclusion

Dans ce chapitre, on a proposé deux règles d'affectation de type probabiliste en s'appuyant sur les lois asymptotiques de certaines variables aléatoires, qu'on appelle géométriques car basées sur le produit scalaire, et on a vu que dans un cas particulier une des règles probabilistes est équivalente à une règle strictement géométrique. Ceci ouvre un angle d'approche différent pour réaliser l'affectation d'une densité gaussienne estimée à une catégorie.

Ces aspects géométriques seront développés dans le chapitre suivant.

Chapitre 4

Aspects géométriques de l'analyse discriminante de distributions gaussiennes

On peut présenter ces aspects dans un cadre plus général que celui des densités gaussiennes. Cependant pour conserver une certaine cohérence à la présentation de l'analyse discriminante de densités gaussiennes, on s'est restreint à ce cadre. En conclusion à ce chapitre on abordera les extensions possibles au cas de densités non gaussiennes.

4.1 Règle d'affectation géométrique

Les notations sont celles définies au paragraphe (3.1.1). Le principe est le suivant. On affecte une densité f_s quelconque à la catégorie qui réalise :

$$\min_q D(f_s, q) .$$

La problématique revient à choisir la distance D entre une densité et une catégorie. Dans le cas qui nous intéresse, bien que d'autres possibilités existent, on se restreindra à la distance au centre de gravité G_q définie au moyen du produit scalaire classique de $H = L^2(\mathbb{R}^p)$:

$$D(f_s, q) = \|f_s - G_q\|. \quad (4.1)$$

Le comportement de cette distance entre deux densités gaussiennes a déjà été étudié au paragraphe (1.4.2), il reste à préciser comment estimer cette distance

lorsqu'on dispose d'un échantillon χ_s , de taille n_s , de loi parente caractérisée par f_s . On propose deux méthodes.

La première ne comporte aucune hypothèse supplémentaire et est basée sur l'explicitation de (4.1) en fonction des produits scalaires $\langle f_s, f_t \rangle$ ($t \in \mathcal{T}_q$) puis l'estimation de leurs paramètres par la méthode du maximum de vraisemblance.

La seconde, déjà abordée au paragraphe 3.2.1 consiste à faire l'hypothèse que les densités f_t ($t \in \mathcal{T}_q$) sont toutes égales à une même densité qui est donc G_q . On est alors conduit à estimer les paramètres (m_q, V_q) de cette loi à partir de l'échantillon composé de la réunion des χ_t ($t \in \mathcal{T}_q$) qui sera de taille $N_q = \sum_{t \in \mathcal{T}_q} n_t$.

4.1.1 Méthode 1

La distance (4.1) se déduit de :

$$\begin{aligned} \|f_s - G_q\|^2 &= \langle f_s - G_q, f_s - G_q \rangle \\ &= \langle f_s - \frac{1}{T_q} \sum_{t \in \mathcal{T}_q} f_t, f_s - \frac{1}{T_q} \sum_{r \in \mathcal{T}_q} f_r \rangle \\ &= \|f_s\|^2 - \frac{2}{T_q} \sum_{t \in \mathcal{T}_q} \langle f_s, f_t \rangle + \frac{1}{T_q^2} \sum_{t \in \mathcal{T}_q} \sum_{r \in \mathcal{T}_q} \langle f_t, f_r \rangle \end{aligned} \quad (4.2)$$

les termes intervenant dans (4.2) ont, rappelons-le, l'expression (2.5) où $\mu_s, \mu_t, \mu_r, \Sigma_s, \Sigma_t$ et Σ_r sont remplacés par leurs estimations.

4.1.2 Méthode 2

Dans ce cas, (4.1) se calcule directement avec (2.5) puisque G_q est gaussienne $N(m_q, V_q)$; m_q et V_q sont estimés de façon habituelle (on note encore χ_t une réalisation de l'échantillon χ_t):

$$\begin{aligned} \widehat{m}_q &= \frac{1}{N_q} \sum_{t \in \mathcal{T}_q} \sum_{x \in \chi_t} x \\ \widehat{V}_q &= \frac{1}{N_q} \sum_{t \in \mathcal{T}_q} \sum_{x \in \chi_t} (x - \widehat{m}_q)(x - \widehat{m}_q)' \end{aligned}$$

soit encore :

$$\widehat{m}_q = \frac{1}{N_q} \sum_{t \in \mathcal{T}_q} n_t \bar{x}_t$$

$$\widehat{V}_q = \frac{1}{N_q} \sum_{t \in \mathcal{T}_q} n_t (s_t + (\bar{x}_t - \widehat{m}_q)(\bar{x}_t - \widehat{m}_q)') .$$

4.1.3 Conclusion

Le choix entre les deux méthodes ne peut que dépendre de l'acceptation ou non de l'égalité des densités appartenant à une même catégorie.

On peut toutefois adopter une attitude empirique en utilisant les techniques utilisées dans la littérature classique de l'analyse discriminante et basées sur les taux de mauvais classements (misclassification) soit sur un échantillon-test si les effectifs T_q sont grands, soit en pratiquant $(T - 1)$ analyses discriminantes de $(T - 1)$ densités et en réaffectant la densité n'ayant pas participé à l'analyse.

4.2 Les représentations graphiques classiques de l'analyse discriminante sont-elles possibles ?

4.2.1 Centrage du nuage des densités

On munit chaque G_q d'un poids $\frac{T_q}{T}$; le centre de gravité de \mathcal{G} est aussi le centre de gravité f_G de \mathcal{F} :

$$f_G = \frac{1}{T} \sum_{t=1}^T f_t = \frac{1}{T} \sum_{q=1}^Q T_q G_q . \quad (4.3)$$

On procède au centrage des nuages \mathcal{F} et \mathcal{G} en les translatant comme suit :

$$\forall t \in \{1, \dots, T\} \quad f_t^\circ = f_t - f_G \quad (4.4)$$

$$\forall q \in \{1, \dots, Q\} \quad G_q^\circ = G_q - f_G = \frac{1}{T_q} \sum_{t \in \mathcal{T}_q} f_t^\circ \quad (4.5)$$

Les nouveaux nuages notés \mathcal{F}° et \mathcal{G}° ont pour centre de gravité l'origine de H et chaque sous nuage \mathcal{F}_q° ainsi obtenu a pour centre de gravité G_q° .

On supposera que les vecteurs $(f_t)_{t \in \mathcal{T}}$ sont linéairement indépendants. Ils engendrent un sous-espace vectoriel de dimension T qui contient aussi les vecteurs f_G , G_q , G_q° , f_t° .

Si \mathcal{T} est $\{1, \dots, T\}$ et F° désigne le sous espace engendré par $f_1^\circ, \dots, f_T^\circ$ alors

F° est de dimension $(T - 1)$. Les systèmes de vecteurs :

$$\begin{aligned} & \{f_1^\circ, \dots, f_{T-1}^\circ\} \\ & \{f_1^\circ - f_T^\circ, \dots, f_{T-1}^\circ - f_T^\circ\} \end{aligned} \quad (4.6)$$

sont des bases de F° et f_T° s'écrit :

$$f_T^\circ = - \sum_{t=1}^{T-1} f_t^\circ . \quad (4.7)$$

Ce sous-espace F° contient les centres de gravité $G_1^\circ, \dots, G_Q^\circ$; si T ne prend pas la modalité q ($Y(T) \neq q$) les composantes de G_q° dans la base $\{f_1^\circ, \dots, f_{T-1}^\circ\}$ sont données par (4.5), et dans le cas contraire ($Y(T) = q$) elles s'obtiennent par :

$$G_q^\circ = -\frac{1}{T_q} \sum_{r \in T - \mathcal{I}_q} f_r^\circ . \quad (4.8)$$

4.2.2 Inertie et décomposition

Définitions

Par analogie avec l'analyse discriminante classique on nomme :

- opérateur d'inertie totale du nuage \mathcal{F}° l'endomorphisme \mathbf{V}° de H qui s'écrit :

$$\mathbf{V}^\circ = \frac{1}{T} \sum_{t \in \mathcal{T}} f_t^\circ \otimes f_t^\circ , \quad (4.9)$$

où $x \otimes y$ est l'endomorphisme de rang un qui à tout h de H associe le vecteur $\langle x, h \rangle y$. En remplaçant f_T° par (4.7) il devient :

$$\mathbf{V}^\circ = \frac{1}{T} \sum_{t=1}^{T-1} (f_t^\circ - f_T^\circ) \otimes f_t^\circ . \quad (4.10)$$

- opérateur d'inertie inter et opérateur d'inertie intra les endomorphismes \mathbf{B}° et \mathbf{W}° de H définis par :

$$\mathbf{B}^\circ = \frac{1}{T} \sum_{q=1}^Q T_q G_q^\circ \otimes G_q^\circ \quad (4.11)$$

$$\mathbf{W}^\circ = \frac{1}{T} \sum_{q=1}^Q \sum_{t \in \mathcal{I}_q} (f_t^\circ - G_q^\circ) \otimes (f_t^\circ - G_q^\circ) . \quad (4.12)$$

Propriétés

Les opérateurs \mathbf{V}° , \mathbf{B}° et \mathbf{W}° , sont autoadjoints positifs.

Les images $\text{Im}\mathbf{V}^\circ$, $\text{Im}\mathbf{W}^\circ$ et $\text{Im}\mathbf{B}^\circ$ sont des sous-espaces vectoriels de F° . L'opérateur d'inertie totale se décompose en inertie inter et inertie intra :

$$\mathbf{V}^\circ = \mathbf{B}^\circ + \mathbf{W}^\circ . \quad (4.13)$$

Interprétation géométrique

Soient h , de norme unité dans H , engendrant le sous-espace Δh et $\mathbf{P}_{\Delta h}$ le projecteur orthogonal sur Δh , l'inertie (par rapport à l'origine O) du nuage $\hat{\mathcal{F}}^\circ$, obtenu par projection de \mathcal{F}° sur Δh est :

$$\frac{1}{T} \sum_{t \in \mathcal{T}} \|\mathbf{P}_{\Delta h} f_t^\circ\|_H^2 = \langle h, Vh \rangle .$$

De même que l'inertie du nuage $\hat{\mathcal{G}}^\circ$ obtenu par projection de \mathcal{G} sur Δh est :

$$\frac{1}{T} \sum_{q=1}^Q \|\mathbf{P}_{\Delta h} G_q^\circ\|_H^2 = \langle h, Bh \rangle .$$

L'inertie de chaque sous-nuage $\hat{\mathcal{F}}_q^\circ$ par rapport à son centre de gravité est :

$$\frac{1}{T_q} \sum_{t \in \mathcal{T}_q} \|\mathbf{P}_{\Delta h} f_t^\circ - \mathbf{P}_{\Delta h} G_q^\circ\|_H^2 = \langle h, \frac{1}{T_q} \sum_{t \in \mathcal{T}_q} (f_t^\circ - G_q^\circ) \otimes (f_t^\circ - G_q^\circ) h \rangle .$$

On vérifie facilement que $\langle h, \mathbf{W}^\circ h \rangle$ mesure l'inertie moyenne, chaque inertie étant pondérée par $\frac{T_q}{T}$.

La relation (4.13) qui donne ici :

$$\langle h, \mathbf{V}^\circ h \rangle = \langle h, \mathbf{B}^\circ h \rangle + \langle h, \mathbf{W}^\circ h \rangle ,$$

s'interprète donc comme en analyse discriminante classique, à savoir la dispersion totale du nuage projeté sur Δh est la somme de la dispersion des centres de gravité et de la dispersion moyenne de chaque sous-nuage par rapport à son centre de gravité.

4.2.3 Critère optimisé par les représentations graphiques

L'objectif de l'analyse factorielle discriminante est, d'un point de vue descriptif, la recherche d'un axe Δh de H orienté de manière à maximiser l'éclatement

du nuage $\widehat{\mathcal{G}}$ tout en minimisant l'éclatement de chaque sous-nuage $\widehat{\mathcal{F}}_q$ autour de son centre de gravité; ce qui se traduit en la recherche de h_1 maximisant :

$$\frac{\langle h, \mathbf{B}^\circ h \rangle}{\langle h, \mathbf{W}^\circ h \rangle}, \quad (4.14)$$

ou :

$$\frac{\langle h, \mathbf{B}^\circ h \rangle}{\langle h, \mathbf{V}^\circ h \rangle}. \quad (4.15)$$

puis on itère sous contrainte d'orthonormalité sur les h_k .

L'analyse discriminante classique — la dimension de H est finie — postule que les données sont telles que \mathbf{V}° est inversible et la solution est obtenue à partir de la diagonalisation de la matrice de l'opérateur $\mathbf{V}^{\circ-1} \mathbf{B}^\circ$ et la solution est unique (à une rotation près dans les sous-espaces propres associés aux valeurs propres multiples).

Dans notre cas l'opérateur \mathbf{V}° étant de rang fini, il ne peut être inversible et la solution ne peut être unique. En effet on peut montrer que si h réalise le maximum alors pour tout h' de $F^{\circ\perp}$, $h'' = h + h'$ réalise le même maximum. En effet $\mathbf{B}^\circ h'$ et $\mathbf{V}^\circ h'$ étant nuls il vient :

$$\frac{\langle h, \mathbf{B}^\circ h \rangle}{\langle h, \mathbf{V}^\circ h \rangle} = \frac{\langle h'', \mathbf{B}^\circ h'' \rangle}{\langle h'', \mathbf{V}^\circ h'' \rangle}.$$

On est dans la même situation qu'en analyse discriminante classique où le nombre de prédicteurs est plus grand que le nombre d'individus observés. Ce problème a été abordé dans le cadre de l'analyse discriminante avec règle probabiliste par des techniques dites de régularisation (Friedman 1989, Mkhadri et al. 1997). En pratique ce qui nous intéresse ce sont les éléments de H qui se construisent à partir des données, ici les vecteurs f_t (ou les f_t°), de manière linéaire: $\sum_t \alpha_t f_t$ (ou $\sum_t \alpha_t f_t^\circ$).

Ainsi plutôt que de chercher h dans H maximisant (4.15), on le cherche dans F° , ce qui conduit à un problème matriciel et rend la solution unique (à une rotation près que l'on précisera).

4.3 Propriétés des représentations “sous contraintes”

4.3.1 Propriétés des opérateurs \mathbf{V}° et \mathbf{W}°

Proposition 4.1 *La restriction $\mathbf{V}_{F^\circ}^\circ$ de \mathbf{V}° à F° est un automorphisme de F° .*

L'opérateur $\mathbf{V}_{F^\circ}^\circ$ est bien un endomorphisme de F° ; en effet $\text{Im } \mathbf{V}^\circ$ est inclus dans F° et donc à fortiori $\text{Im } \mathbf{V}_{F^\circ}^\circ$. Soit f de F° tel que $\mathbf{V}^\circ f = 0$, alors (4.13) donne :

$$\sum_{t=1}^{T-1} \langle f_t^\circ - f_T^\circ, f \rangle f_t^\circ = 0 .$$

Les vecteurs $(f_t^\circ)_{t=1, T-1}$ étant linéairement indépendants, les $(T-1)$ coefficients $\langle f_t^\circ - f_T^\circ, f \rangle$ sont donc nuls.

Ainsi, le vecteur f de F° est nul, car il est orthogonal à tous les vecteurs $f_t^\circ - f_T^\circ$ ($t = 1, T-1$) qui constituent une base de F° (4.6); $\mathbf{V}_{F^\circ}^\circ$ est donc injective.

L'opérateur $\mathbf{V}_{F^\circ}^\circ$ est un endomorphisme injectif de l'espace F° de dimension finie, cela établit que $\mathbf{V}_{F^\circ}^\circ$ est un automorphisme de F° . \square

Proposition 4.2 *L'opérateur \mathbf{W}° est de rang $(T-Q)$ et $\ker \mathbf{W}_{F^\circ}^\circ$ est de dimension $(Q-1)$.*

La démonstration de cette proposition découle du lemme suivant.

Lemme 4.1 *Soit $\{x_1, \dots, x_I\}$ un système de vecteurs d'un espace de Hilbert engendrant un sous-espace χ alors le rang de ce système, le rang de l'opérateur $\sum_i x_i \otimes x_i$ et le rang de sa restriction à χ sont égaux.*

L'opérateur \mathbf{W}° écrit en (4.12) a donc pour rang, le rang du système de vecteurs :

$$\bigcup_{q=1}^Q \{f_t^\circ - G_q^\circ ; t \in \mathcal{T}_q\}$$

égal à :

$$\bigcup_{q=1}^Q \{f_t - G_q ; t \in \mathcal{T}_q\}.$$

Le système $\{f_t - G_q ; t \in \mathcal{T}_q\}$ est de rang $(T_q - 1)$. En rappelant que les vecteurs $\{f_t ; t \in \mathcal{T}\}$ sont linéairement indépendants, on peut montrer que :

$$\text{rg } \mathbf{W}^\circ = \sum_{q=1}^Q (T_q - 1) = T - Q ,$$

d'où :

$$\dim \ker \mathbf{W}_{F^\circ}^\circ = Q - 1 .$$

□

Proposition 4.3 *L'opérateur \mathbf{B}° est de rang $Q - 1$.*

En vertu du lemme 4.1 son rang est égal au rang du système $\{G_1^\circ, \dots, G_Q^\circ\}$. En remarquant que ce système est lié ($\sum_q T_q G_q^\circ = 0$) et qu'en lui ôtant un élément (par exemple le centre de la classe qui contient f_T°) il devient libre, le résultat est immédiat. □

4.3.2 Intérêt des représentations graphiques

Des deux premières propositions précédentes et de (4.13) il vient que le maximum de (4.15) vaut 1 et est atteint dès que h est pris dans $\ker \mathbf{W}^\circ$ et avec les itérations sous contraintes d'orthonormalité, le maximum est alors atteint $(Q - 1)$ fois. L'inertie intra étant nulle dans ce cas, les densités d'une même catégorie se projettent sur l'axe Δh en un point unique.

On est dans un cas trivial de parfaite discrimination visuelle des densités datées (pour lesquelles la variable Y est connue). C'est donc une situation analogue à (Carlier and Gueguen 1994, p.245) puisque, dans le vocabulaire individus x variables — ici les individus sont les $(T - 1)$ “densités” centrées et les variables sont les $(T - 1)$ axes de coordonnées dans F° , axes qui sont ces mêmes “densités” centrées —, on a $(T - 1)$ individus dans un espace de dimension $(T - 1)$.

L'intérêt de la représentation des centres de gravité et des densités non classées $\{f_s ; s \in \mathcal{T}'\}$ sur les $(T - Q)$ premiers axes est qu'elle permet de visualiser :

- les distances respectives entre les centres de gravité et donner une idée graphique du pouvoir séparateur de la règle d'affectation géométrique ;

- les distances de chaque densité aux différents centres de gravité et adopter, si les qualités de représentation sont bonnes, une règle d’affectation visuelle.

De la troisième proposition, il vient que les $(Q - 1)$ autres axes sont dans le noyau de \mathbf{B}° et que sur ces axes on peut avoir une idée visuelle de la dispersion de chaque sous-nuage \mathcal{F}_q° autour de son centre G_q° .

On a pu voir que techniquement la contrainte imposée, rechercher des axes discriminants dans le sous-espace F° engendré par les “densités” centrées, a levé une indétermination mathématique, a conduit à un problème matriciel qu’on traitera ci-après mais a aussi conduit à une indétermination dans le choix des axes : les axes correspondant à la même valeur propre ont tous le même pouvoir séparateur.

4.4 Ecriture matricielle de l’analyse discriminante sous contraintes

4.4.1 Matrices d’inertie et critère

On munit F° de la base $\{f_1^\circ, \dots, f_{T-1}^\circ\}$, l’exclusion de la densité d’indice T est arbitraire ; dans cette base les opérateurs $\mathbf{V}_{F^\circ}^\circ$, $\mathbf{B}_{F^\circ}^\circ$ et $\mathbf{W}_{F^\circ}^\circ$ auront pour matrice V , B et W respectivement ; le produit scalaire de H induit sur F° le produit scalaire de matrice M de terme général M_{st} :

$$\forall s, t \in \{1, \dots, T - 1\} \quad M_{st} = \langle f_s^\circ, f_t^\circ \rangle . \quad (4.16)$$

L’espace \mathbb{R}^{T-1} sera muni de la métrique de matrice M . Toute fonction h_k de F° s’écrivant :

$$\sum_{t=1}^{T-1} h_{k,t} f_t^\circ \quad (4.17)$$

sera représentée par un vecteur colonne, noté aussi h_k , de \mathbb{R}^{T-1} dont les composantes sont les coefficients $h_{k,1}, \dots, h_{k,T-1}$.

Rechercher h_1 normé dans F° est équivalent à rechercher h_1 dans (\mathbb{R}^{T-1}, M) de M -norme unité et le critère défini en (4.15) s’écrit :

$$\mathcal{I}(h) = \frac{h' M B h}{h' M V h} , \quad (4.18)$$

enfin la contrainte d'orthonormalité sur les h_k dans F° devient contrainte de M -orthonormalité sur les h_k de \mathbb{R}^{T-1} .

Les opérateurs \mathbf{V} et \mathbf{B} étant autoadjoints positifs, \mathbf{V}_{F° et \mathbf{B}_{F° le sont aussi ; ainsi les matrices MB , MV et MV^{-1} sont symétriques, MB est semi-définie positive, MV et MV^{-1} sont définies positives.

On notera que la maximisation du critère (4.18) est l'analogie de :

$$\max_a \frac{a'MBMa}{a'MVMa}$$

de l'analyse factorielle discriminante classique avec p variables et n individus (Saporta 1990, p.406), puis itérations de M -orthonormalité.

4.4.2 Solution matricielle

Compte tenu des propriétés énoncées au paragraphe (4.3.1), pour trouver les h_k , il suffit de chercher des bases M -orthonormales des noyaux des matrices W et B .

Les fonctions $f_1^\circ, \dots, f_T^\circ$ peuvent être représentées de manière approchée en les projetant sur un des plans engendrés par les h_k . Pour tout s de $\{1, \dots, T-1\}$ la coordonnée de f_s° suivant h_k est égale à $\langle f_s^\circ, h_k \rangle$, c'est-à-dire à la s -ième composante du vecteur colonne :

$$C_k = M h_k . \quad (4.19)$$

La "densité" f_T° s'écrit en fonction des f_s° (4.7) et donc sa coordonnée est l'opposé de la somme des coordonnées précédentes. Les coordonnées des densités non datées seront calculées au paragraphe 4.5.6.

La qualité de représentation de chaque fonction f_s° par sa projection \widehat{f}_s° se mesure par le rapport des normes :

$$\frac{\|\widehat{f}_s^\circ\|}{\|f_s^\circ\|} .$$

Le carré du numérateur est égal à C_{ks}^2 et le carré du dénominateur est le s -ième terme diagonal de la matrice M , calculée ci-après (4.5.2).

4.5 Mise en œuvre des calculs

4.5.1 Produits scalaires entre densités

On calcule les produits scalaires $\langle f_s, f_t \rangle$ par la formule (2.5) pour tout s de $\mathcal{T} \cup \mathcal{T}'$ et tout t de \mathcal{T} , et on les range dans une matrice D à deux blocs $D^{(1)}$ et $D^{(2)}$, le premier est symétrique et correspond aux densités datées, le second contient le croisement des densités non datées avec les densités datées :

$$D = \begin{pmatrix} D^{(1)} \\ D^{(2)} \end{pmatrix} = \begin{pmatrix} \langle f_1, f_1 \rangle & \dots & \langle f_1, f_T \rangle \\ \vdots & & \vdots \\ \langle f_T, f_1 \rangle & \dots & \langle f_T, f_T \rangle \\ \langle f_{T+1}, f_1 \rangle & \dots & \langle f_{T+1}, f_T \rangle \\ \vdots & & \vdots \\ \langle f_{T+T'}, f_1 \rangle & \dots & \langle f_{T+T'}, f_T \rangle \end{pmatrix} \quad (4.20)$$

4.5.2 Matrice M des produits scalaires entre “densités” centrées

$D_{.t}^{(1)}$ et $D_t^{(1)}$ désignent indifféremment la somme de la t -ème colonne ou t -ème ligne de $D^{(1)}$, car $D^{(1)}$ est symétrique. La somme de tous les termes de $D^{(1)}$ sera notée $D_{..}^{(1)}$. En utilisant les expressions (4.3) et (4.4) on peut calculer le terme général (4.16) de la matrice M :

$$\forall s, t \in \{1, \dots, T-1\} \quad M_{st} = D_{st}^{(1)} - \frac{1}{T}D_{s.}^{(1)} - \frac{1}{T}D_{.t}^{(1)} + \frac{1}{T^2}D_{..}^{(1)} \quad (4.21)$$

4.5.3 Matrice V d'inertie totale

La matrice V est celle de l'opérateur $\mathbf{V}_{F^\circ}^\circ$ dans la base $\{f_1^\circ, \dots, f_{T-1}^\circ\}$. L'image de f_t° par \mathbf{V}° peut être calculée à partir de l'expression (4.10) de \mathbf{V}° et le terme général V_{st} de V s'écrit :

$$V_{st} = \frac{1}{T} \langle f_s^\circ - f_T^\circ, f_t^\circ \rangle ,$$

soit :

$$V_{st} = \frac{1}{T} (D_{st}^{(1)} - D_{Tt}^{(1)} - \frac{1}{T}D_{s.}^{(1)} + \frac{1}{T^2}D_{T.}^{(1)}) , \quad (4.22)$$

ou encore :

$$V_{st} = \frac{1}{T}(M_{st} + M_{.t}) , \quad (4.23)$$

$M_{.t}$ désignant la somme de la t -ième colonne de M .

4.5.4 Matrice B d'inertie inter

La matrice B est celle de l'opérateur $\mathbf{B}^\circ_{F^\circ}$ dans la base $\{f_1^\circ, \dots, f_{T-1}^\circ\}$; l'image de f_t° par \mathbf{B}° peut être calculée à partir de (4.11) :

$$\mathbf{B}^\circ f_t^\circ = \frac{1}{T} \sum_{q=1}^Q T_q \langle G_q^\circ, f_t^\circ \rangle G_q^\circ .$$

On y remplace G_q° par (4.5) :

$$\mathbf{B}^\circ f_t^\circ = \frac{1}{T} \sum_{q=1}^Q T_q \langle G_q^\circ, f_t^\circ \rangle \left(\frac{1}{T_q} \sum_{s \in \mathcal{I}_q} f_s^\circ \right) .$$

En notant $Y(s)$ le numéro de la modalité de la variable Y prise par s , $\mathbf{B}^\circ f_t^\circ$ s'écrit :

$$\frac{1}{T} \sum_{s \in \mathcal{T}} \langle G_{Y(s)}^\circ, f_t^\circ \rangle f_s^\circ .$$

En isolant le terme $\langle G_{Y(T)}^\circ, f_t^\circ \rangle f_T^\circ$ qui en utilisant (4.7) devient :

$$\sum_{r=1}^{T-1} \langle G_{Y(T)}^\circ, f_t^\circ \rangle f_r^\circ ,$$

puis en regroupant les coefficients de chaque f_s° ($s = 1, \dots, T-1$), on obtient :

$$\mathbf{B}^\circ f_t^\circ = \frac{1}{T} \sum_{s=1}^{T-1} \langle G_{Y(s)}^\circ - G_{Y(T)}^\circ, f_t^\circ \rangle f_s^\circ . \quad (4.24)$$

Le terme général B_{st} de la matrice B est donc nul si $Y(s) = Y(T)$ et est égal à :

$$\frac{1}{T} \left[\left(\frac{1}{T_{Y(s)}} + \frac{1}{T_{Y(T)}} \right) \sum_{r \in \mathcal{I}_{Y(s)}} \mathbf{M}_{rt} + \frac{1}{T_{Y(T)}} \sum_{r \notin \mathcal{I}_{Y(s)} \cup \mathcal{I}_{Y(T)}} \mathbf{M}_{rt} \right] \quad (4.25)$$

dans le cas contraire.

4.5.5 Matrice W d'inertie intra

La décomposition (4.13) de l'opérateur d'inertie totale en inertie inter et inertie intra s'écrit aussi matriciellement, d'où :

$$W = V - B .$$

4.5.6 Coordonnées des densités non datées

Ces densités sont d'abord translitées :

$$f_{T+r}^\circ = f_{T+r} - f_G \quad r = 1, \dots, T' .$$

La coordonnée de f_{T+r}° sur l'axe discriminant engendré par h_k est :

$$\langle f_{T+r}^\circ, h_k \rangle$$

et vaut d'après (4.17) :

$$\sum_{t=1}^{T-1} h_{k,t} \langle f_{T+r}^\circ, f_t^\circ \rangle .$$

Ainsi, si N désigne la matrice de taille (T', T) de terme général :

$$N_{rt} = \langle f_{T+r}^\circ, f_t^\circ \rangle , \quad (4.26)$$

le vecteur $N h_k$ a pour composantes les coordonnées des densités non datées sur l'axe engendré par h_k .

Le terme général (4.26) se calcule à partir de la matrice D (4.20) :

$$N_{rt} = D_{rt}^{(2)} - \frac{1}{T} D_{r.}^{(2)} - \frac{1}{T} D_{.t}^{(1)} + \frac{1}{T^2} D_{..}^{(1)} .$$

4.6 Une autre solution matricielle

On propose ici une autre manière de calculer les matrices V et B qui, grâce à une renumérotation des densités, simplifie de beaucoup ce calcul.

On note F et G les matrices respectives des vecteurs $\{f_s^\circ - f_T^\circ ; s = 1, \dots, T-1\}$ et $\{G_{Y(s)}^\circ - G_{Y(T)}^\circ ; s = 1, \dots, T-1\}$ dans la base $\{f_1^\circ, \dots, f_{T-1}^\circ\}$.

Lemme 4.2 *La matrice F est symétrique, inversible et :*

$$F = \begin{pmatrix} 2 & 1 & \dots & \dots & 1 \\ 1 & 2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 2 & 1 \\ 1 & \dots & \dots & 1 & 2 \end{pmatrix} . \quad (4.27)$$

$$F^{-1} = \frac{1}{T} \begin{pmatrix} T-1 & -1 & \dots & \dots & -1 \\ -1 & T-1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & T-1 & -1 \\ -1 & \dots & \dots & -1 & T-1 \end{pmatrix} \quad (4.28)$$

L'écriture de la matrice F découle directement de l'écriture (4.7) de f_T° qu'on rappelle ci-dessous :

$$f_T^\circ = - \sum_{t=1}^{T-1} f_t^\circ .$$

Son inverse se calcule aisément. □

On vérifie alors la relation suivante.

Proposition 4.4

$$V = \frac{1}{T} FM . \quad (4.29)$$

Lemme 4.3 *La matrice G est symétrique de terme général :*

$$G_{ts} = \begin{cases} 0 & \text{si } Y(s) = Y(T) \\ & \text{(s ou t prennent la même modalité que T)} \\ \frac{1}{T_{Y(s)}} + \frac{1}{T_{Y(T)}} & \text{si } Y(s) = Y(t) \neq Y(T) \\ & \text{(s et t prennent une même modalité, autre que celle de T)} \\ \frac{1}{T_{Y(T)}} & \text{si } Y(T) \neq Y(s) \neq Y(t) \neq Y(T) \\ & \text{(s, t et T prennent trois modalités différentes)} \end{cases}$$

Si s et T prennent la même modalité q de la variable Y , alors $G_{Y(s)}^\circ$ et $G_{Y(T)}^\circ$ sont égaux et leur différence est donc nulle.

Dans le cas contraire, on note q la modalité prise par T et q_1 celle prise par s .

On calcule d'abord $G_{Y(T)}^\circ$:

$$\begin{aligned} G_q^\circ &= \frac{1}{T_q} \sum_{r \in \mathcal{I}_q} f_t^\circ && \text{cf. (4.5)} \\ &= \frac{1}{T_q} \left(\sum_{r \in \mathcal{I}_q, r \neq T} f_t^\circ + f_T^\circ \right) \\ &= \frac{1}{T_q} \left(\sum_{r \in \mathcal{I}_q, r \neq T} f_t^\circ - \sum_{t=1}^{T-1} f_t^\circ \right) && \text{cf. (4.7)} \\ &= -\frac{1}{T_q} \sum_{t \in T - \mathcal{I}_q} f_t^\circ \quad . \end{aligned}$$

La t -ième composante de $G_{Y(T)}^\circ$ vaut :

$$\begin{cases} -\frac{1}{T_{Y(T)}} & \text{si } Y(t) \neq Y(T) \\ 0 & \text{sinon.} \end{cases}$$

On calcule $G_{Y(s)}^\circ$:

$$G_{q_1}^\circ = \frac{1}{T_{q_1}} \sum_{t \in \mathcal{I}_{q_1}} f_t^\circ .$$

La t -ième composante de $G_{Y(s)}^\circ$ vaut :

$$\begin{cases} -\frac{1}{T_{Y(s)}} & \text{si } Y(t) = Y(s) \\ 0 & \text{sinon.} \end{cases}$$

□

Remarque Si on ordonne les densités de telle sorte que les T_1 premières sont dans la modalité 1, les T_2 suivantes sont dans la modalité 2, etc, alors les T_1 premières colonnes de G sont égales, les T_2 suivantes aussi, etc ; de plus les $(T_Q - 1)$ dernières colonnes sont nulles. A titre d'illustration on donne l'allure des

colonnes 1 et $(T_1 + 1)$ de la matrice symétrique G :

$$\begin{array}{ccc}
 \uparrow & \frac{1}{T_1} + \frac{1}{T_Q} & \frac{1}{T_Q} \\
 T_1 & \vdots & \vdots \\
 \downarrow & \frac{1}{T_1} + \frac{1}{T_Q} & \frac{1}{T_Q} \\
 \uparrow & \frac{1}{T_Q} & \frac{1}{T_2} + \frac{1}{T_Q} \\
 T_2 & \vdots & \vdots \\
 \downarrow & \frac{1}{T_Q} & \frac{1}{T_2} + \frac{1}{T_Q} \\
 \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots \\
 \uparrow & \frac{1}{T_Q} & \frac{1}{T_Q} \\
 T_{Q-1} & \vdots & \vdots \\
 \downarrow & \frac{1}{T_Q} & \frac{1}{T_Q} \\
 \uparrow & 0 & 0 \\
 (T_Q - 1) & \vdots & \vdots \\
 \downarrow & 0 & 0
 \end{array}$$

Proposition 4.5

$$B = \frac{1}{T}GM .$$

L'expression (4.24) de B fournit immédiatement l'écriture (4.29) de la proposition. \square

4.7 Conclusion

Tout ce qui a été dit dans ce chapitre (règles d'affectation et représentations graphiques sous contraintes) peut être étendu au cas d'objets quelconques qui ne seraient pas nécessairement des densités gaussiennes ni même des densités. La seule hypothèse nécessaire est que ces objets en nombre fini (T) soient des vecteurs linéairement indépendants dans un espace de Hilbert de dimension supérieure ou

égale à T .

Le premier exemple est celui de T densités de $L^2(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p}, \nu)$ qui comprend donc les cas de densités d'une même famille exponentielle ou de densités uniformes et dont les mesures d'affinité L^2 ont été calculées aux paragraphes 1.2 et 1.3 respectivement.

Le second exemple est celui de matrices de variance (ou de corrélation) linéairement indépendantes plongées dans \mathbb{R}^{p^2} muni du produit scalaire trace (2.11).

Chapitre 5

Application de l'ACP de densités gaussiennes à quelques processus particuliers

5.1 Processus solution d'une équation différentielle stochastique bilinéaire

Parmi les processus gaussiens dont la loi est explicitement calculable, les solutions d'équations différentielles stochastiques constituent une classe particulièrement variée. On utilisera ici les résultats de Le Breton et Musiela (1983) dont on rappelle quelques aspects ci-dessous.

5.1.1 Rappels

On considère l'équation différentielle stochastique suivante :

$$\begin{cases} dX_t = [A_0(t)X_t + a_0(t)] dt + \sum_{j=1}^d [A_j(t)X_t + a_j(t)] dB_t^j, & t \geq 0 \\ X_0 = X(0) \end{cases} \quad (5.1)$$

où :

- $B = (B^1, \dots, B^d)$ est le mouvement brownien standard de dimension d ;
- A_0, \dots, A_d sont des matrices carrées d'ordre p ;
- a_0, \dots, a_d sont des vecteurs à p composantes ;

- l'état initial $X(0)$ est un vecteur aléatoire, indépendant de B , de moyenne $\mu(0)$ et de matrice de covariance $\Sigma(0)$.

Sous certaines conditions d'intégrabilité des composantes des matrices A et des vecteurs a , Le Breton et Musiela caractérisent les processus solution de cette équation et donnent les fonctions moyenne et covariance comme solution d'équations différentielles ordinaires. On se limitera au cas où les matrices A_1, \dots, A_d sont nulles. La solution de (5.1) est alors un processus gaussien dès que $X(0)$ est gaussien (Gard 1988, p.120). La moyenne μ_t de X_t est solution de :

$$\dot{\mu}_t = A_0(t)\mu_t + a_0(t), \quad t \geq 0; \quad \mu_0 = \mu(0).$$

La variance Σ_t de X_t est solution de :

$$\dot{\Sigma}_t = A_0(t)\Sigma_t + \Sigma_t A_0(t)' + \sum_{j=1}^d a_j(t)a_j(t)', \quad t \geq 0; \quad \Sigma_0 = \Sigma(0).$$

Parmi les cas particuliers les plus célèbres, on trouve le mouvement brownien lui-même et le processus d'Ornstein-Uhlenbeck. On les examinera séparément dans les sections suivantes, après avoir traité un exemple illustratif de (5.1).

5.1.2 Exemple

Soient :

$$\left\{ \begin{array}{l} d = p = 2 \\ A_0(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad A_1(t) = (0) \quad A_2(t) = (0) \\ a_0(t) = (0) \quad a_1(t) = \begin{pmatrix} e^{\frac{t}{2}} \\ e^{\frac{t}{2}} \end{pmatrix} \quad a_2(t) = \begin{pmatrix} e^{\frac{t}{2}} \\ -e^{\frac{t}{2}} \end{pmatrix}. \end{array} \right. \quad (5.2)$$

L'équation (5.1) a pour solution un processus $(X_t)_{t \geq 0}$ tel que pour tout t :

$$\begin{aligned} X_t &\rightsquigarrow N(\mu_t, \Sigma_t), \\ \mu_t &= \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \end{aligned} \quad (5.3)$$

$$\Sigma_t = \begin{pmatrix} 2e^t & 0 \\ 0 & 2e^t \end{pmatrix}. \quad (5.4)$$

Pour tout instant t , on note f_t la densité de probabilité de X_t . On procède à l'ACP des densités f_0, \dots, f_{25} de ce processus. On rappelle que cette ACP conduit à diagonaliser la matrice W dont le terme général est (2.5), soit ici :

$$\langle f_s, f_t \rangle = \frac{1}{4\pi} \frac{1}{e^s + e^t} e^{-\frac{1}{2} \frac{1 - \cos(s-t)}{e^s + e^t}}.$$

Quant à l'ACP normée, elle s'obtient en diagonalisant la matrice \overline{W} (2.10) de terme général :

$$\langle \overline{f}_s, \overline{f}_t \rangle = 2 \frac{e^{\frac{s+t}{2}}}{e^s + e^t} e^{-\frac{1}{2} \frac{1 - \cos(s-t)}{e^s e^t}}. \quad (5.5)$$

Ce nombre est le cosinus de l'angle formé par les deux densités f_s et f_t dans $L^2(\mathbb{R}^p)$.

Les trois premiers éléments propres de \overline{W} (ACP normée) ou de W (ACP non normée) fournissent une représentation approchée des densités sur les trois premiers axes principaux et les pourcentages d'inertie expliquée par ces axes.

On rappelle que la représentation obtenue à partir de \overline{W} , elle déforme le moins possible les angles entre densités ; quant à la représentation obtenue par la diagonalisation de W déforme le moins possible les distances. On a délibérément opté pour l'ACP normée car elle offre l'avantage de visualiser les qualités de représentation des densités : cette qualité est d'autant meilleure que le point de représentation est proche du cercle de rayon unité.

Au cours du temps l'évolution de la moyenne (5.3) du processus est périodique et l'évolution de la matrice de variance (5.3) est exponentielle. L'évolution des densités devrait en être la résultante comme on l'avait déjà visualisé au paragraphe 1.4.2.

Les graphiques (Fig. 5.1) donnent les projections des densités sur les trois premiers plans principaux.

L'évolution des densités sur le premier plan principal est régulière. Cependant cette régularité ne fait pas apparaître la périodicité à laquelle on aurait pu s'attendre. Ceci est dû au fait que la variance, croissant exponentiellement, est prépondérante ; l'évolution de la moyenne est en quelque sorte "noyée", "diluée"

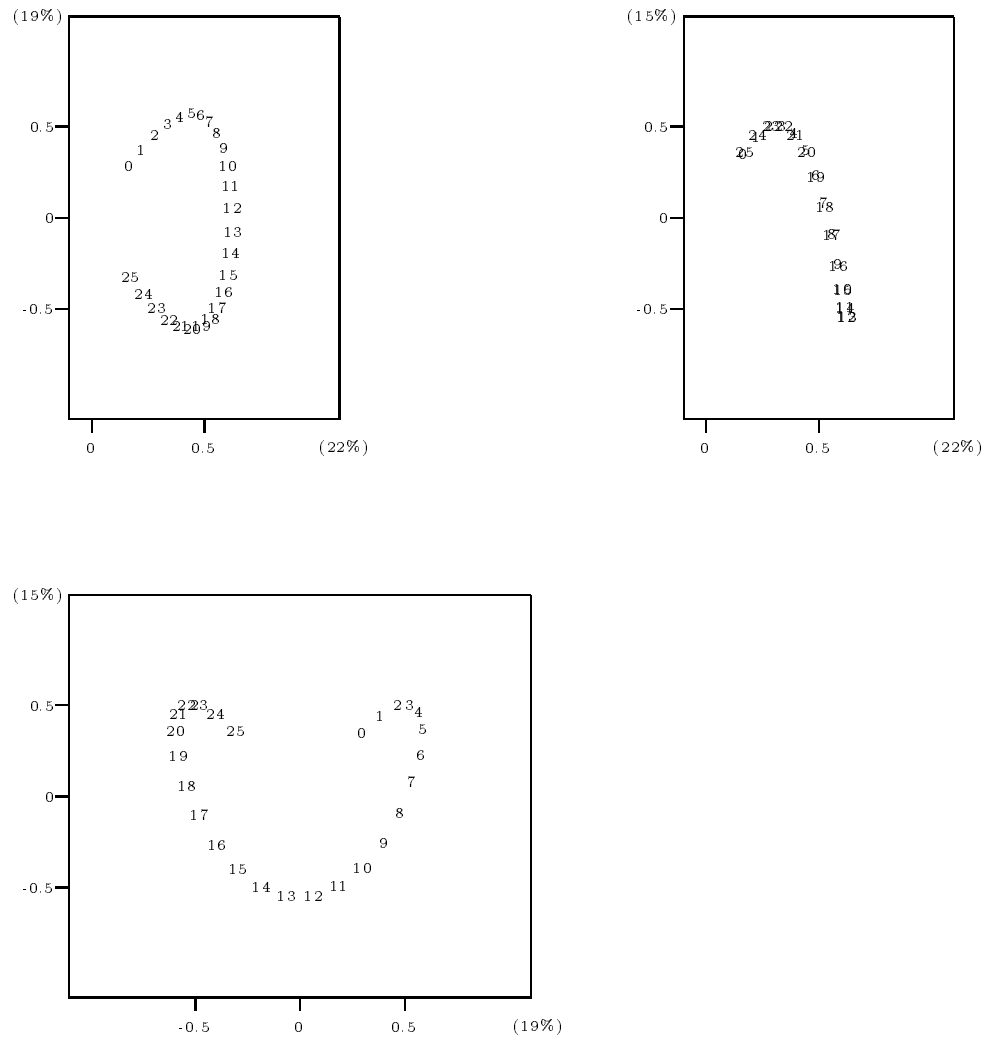


FIG. 5.1 - Les trois premiers plans principaux de l'équation (5.2)

dans l'évolution de la variance.

Pour bien montrer cet aspect, on a fait l'ACP normée en considérant un processus de moyenne constante et de variance (5.4); ce processus est solution de l'équation différentielle considérée en (5.2) à la différence près que la matrice A_0 est nulle :

$$\begin{cases} d = p = 2 \\ A_0(t) = (0) & A_1(t) = (0) & A_2(t) = (0) \\ a_0(t) = (0) & a_1(t) = \begin{pmatrix} e^{\frac{t}{2}} \\ e^{\frac{t}{2}} \end{pmatrix} & a_2(t) = \begin{pmatrix} e^{\frac{t}{2}} \\ -e^{\frac{t}{2}} \end{pmatrix}. \end{cases} \quad (5.6)$$

Les représentations des densités sur les premiers plans principaux sont quasiment les mêmes (Fig. 5.2).

Puis on a considéré un autre processus dont la moyenne est identique à (5.3) mais dont la variance ne croît pas aussi rapidement que (5.4). Ce processus est solution de l'équation différentielle stochastique (5.1) pour :

$$\begin{cases} d = p = 2 \\ A_0(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} & A_1(t) = (0) & A_2(t) = (0) \\ a_0(t) = (0) & a_1(t) = \begin{pmatrix} \frac{1}{10}e^{\frac{t}{50}} \\ \frac{1}{10}e^{\frac{t}{50}} \end{pmatrix} & a_2(t) = \begin{pmatrix} \frac{1}{10}e^{\frac{t}{50}} \\ -\frac{1}{10}e^{\frac{t}{50}} \end{pmatrix}. \end{cases} \quad (5.7)$$

Sa variance est égale à :

$$\Sigma_t = \begin{pmatrix} \frac{1}{2}e^{\frac{t}{25}} & 0 \\ 0 & \frac{1}{2}e^{\frac{t}{25}} \end{pmatrix}.$$

L'ACP normée donne des représentations des densités (Fig. 5.3) reflétant la périodicité de la moyenne; ce caractère périodique est encore plus net en opérant un zoom sur le nuage des densités sur le premier plan principal (Fig. 5.4).

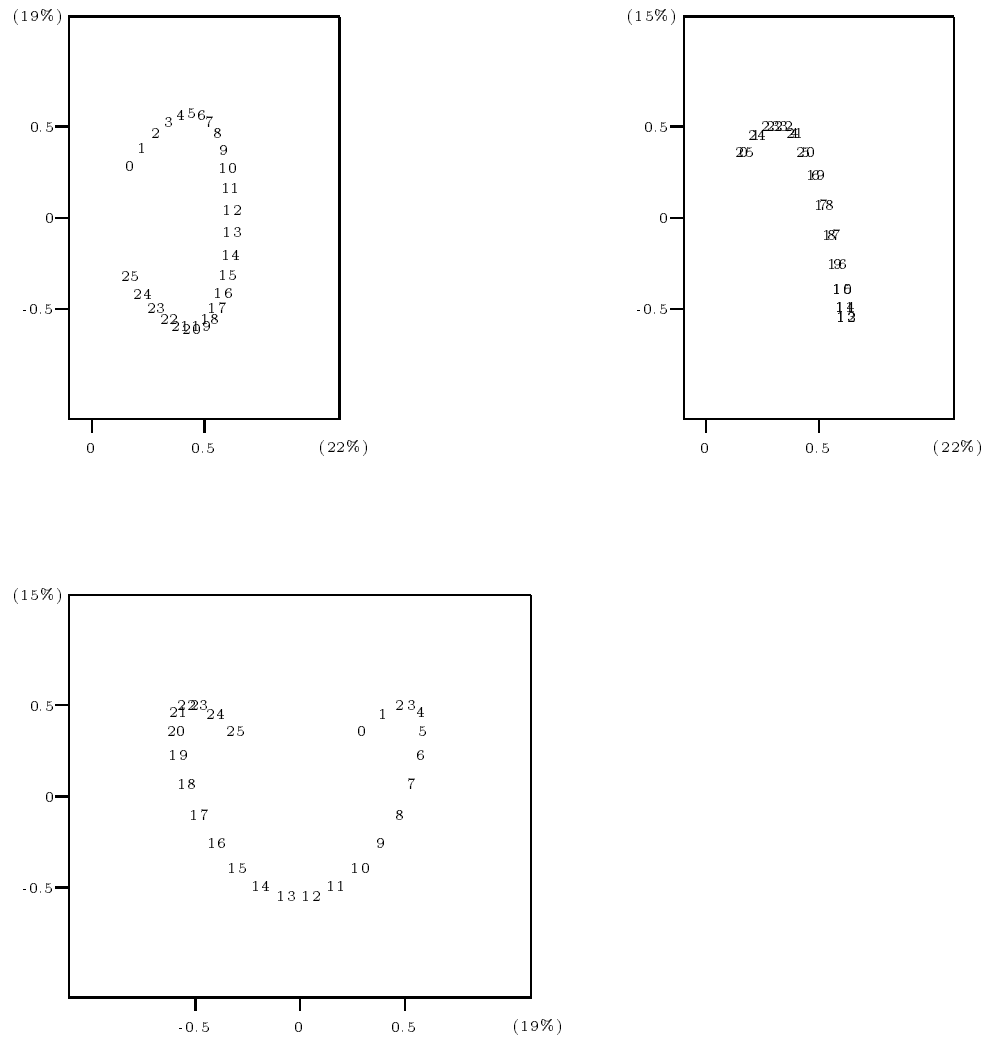


FIG. 5.2 - Les trois premiers plans principaux de l'équation (5.6)

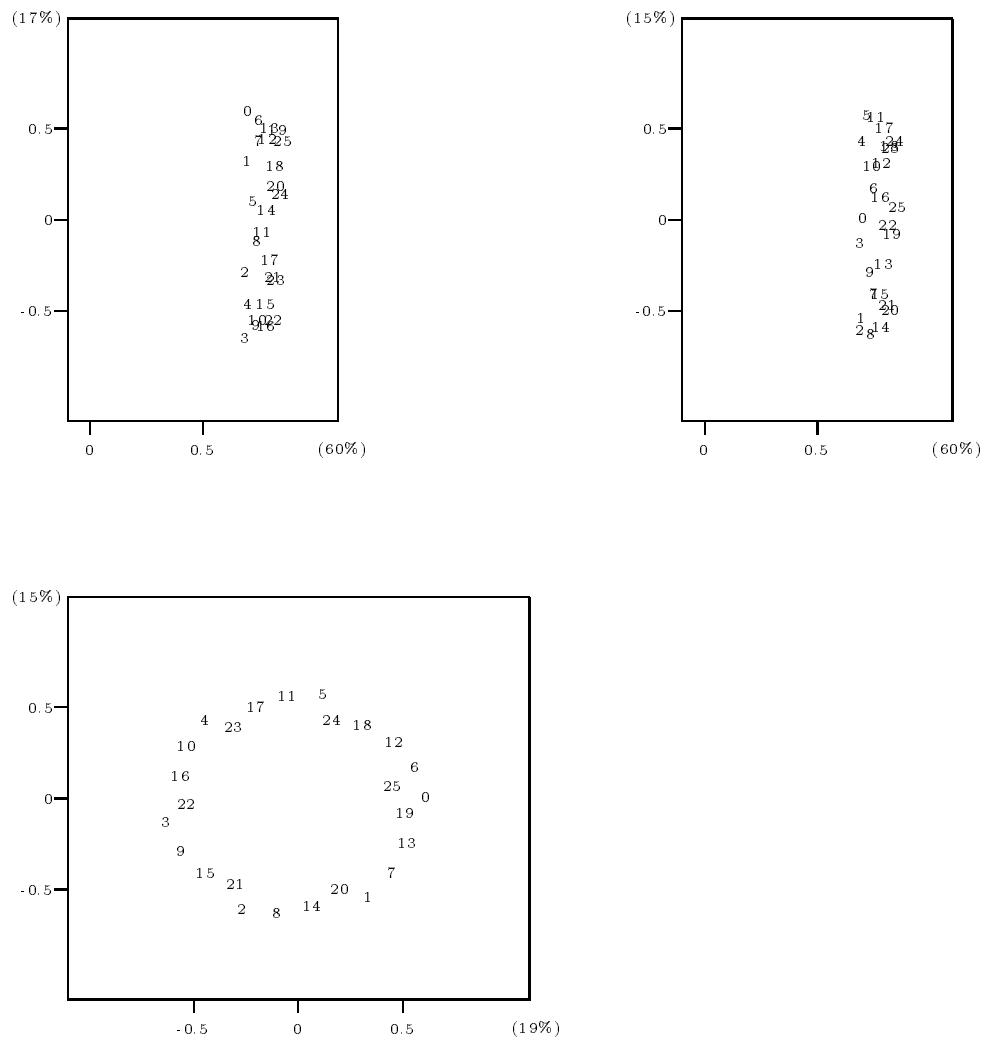


FIG. 5.3 - Les trois premiers plans principaux de l'équation (5.7)

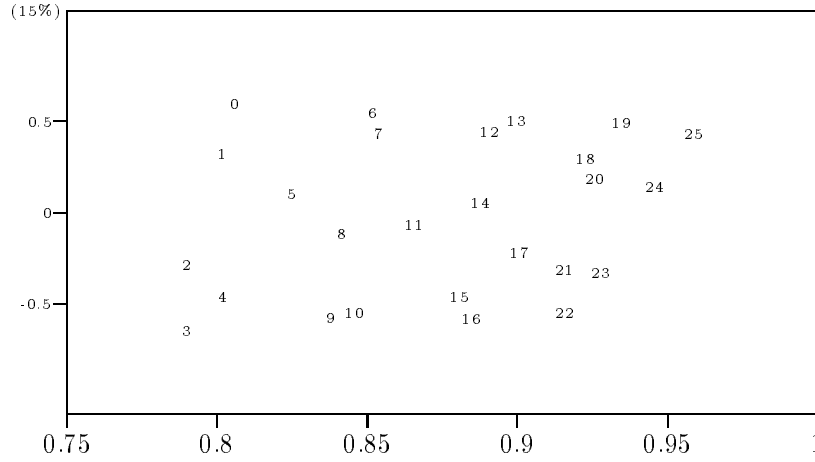


FIG. 5.4 - Zoom sur le premier graphique de Fig. 5.3

5.2 Mouvement brownien

5.2.1 Introduction

Soit $\{B_t, t \geq 0\}$ un mouvement brownien de paramètre de dérive μ ($\mu \in \mathbb{R}^p$) et de paramètre de diffusion Σ (Σ matrice d'ordre p , symétrique, définie positive).

On note f_t la densité de probabilité de B_t qui est donc de loi $N(t\mu, t\Sigma)$; l'ACP des T densités f_{t_1}, \dots, f_{t_T} conduit à diagonaliser la matrice W d'ordre T dont le terme général W_{kl} est :

$$W_{kl} = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \frac{1}{(t_k + t_l)^{\frac{p}{2}}} e^{-\frac{1}{2}\|\mu\|_{\Sigma}^2 \frac{(t_k - t_l)^2}{t_k + t_l}};$$

quant à l'ACP normée, elle s'obtient en diagonalisant la matrice \overline{W} :

$$\overline{W}_{kl} = \left\langle \frac{f_{t_k}}{\|f_{t_k}\|}, \frac{f_{t_l}}{\|f_{t_l}\|} \right\rangle = \sqrt{\frac{2\sqrt{t_k t_l}}{t_k + t_l}}^p e^{-\frac{1}{2}\|\mu\|_{\Sigma}^2 \frac{(t_k - t_l)^2}{t_k + t_l}}. \quad (5.8)$$

La distance entre densités est :

$$\|f_{t_k} - f_{t_l}\|^2 = \frac{1}{(2\sqrt{\pi})^p} \frac{1}{|\Sigma|^{\frac{1}{2}}} \left(\frac{1}{t_k^{\frac{p}{2}}} + \frac{1}{t_l^{\frac{p}{2}}} - \frac{2}{\left(\frac{t_k + t_l}{2}\right)^{\frac{p}{2}}} e^{-\frac{1}{2}\|\mu\|_{\Sigma}^2 \frac{(t_k - t_l)^2}{t_k + t_l}} \right).$$

la caractéristique principale du nuage.

La matrice \overline{W} (Tab. 5.1) fait apparaître, d'une part, que les cosinus des angles formés par deux densités consécutives dans le temps sont proches de 1 et varient régulièrement de 0.82 (entre f_1 et f_2) à 0.99 (entre f_{25} et f_{26}) et, d'autre part, que plus l'écart entre instants s et t est grand plus le cosinus entre densités f_s et f_t est faible, voire nul comme indiqué en (Tab. 5.1).

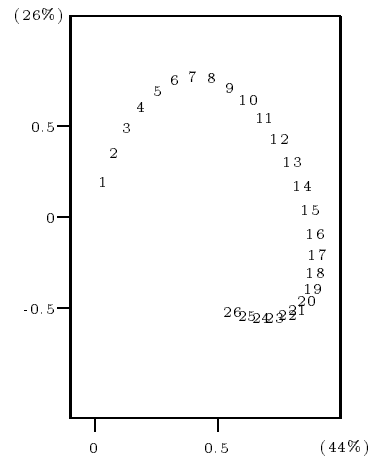


FIG. 5.5 - Cas de référence.

5.2.3 Etude empirique de la convergence de l'ACP

Les valeurs des paramètres définissant le mouvement brownien sont ceux du cas de référence. Pour chaque instant t on simule un échantillon de B_t de taille n , puis on calcule la moyenne et la variance empirique. Puis on procède à l'ACP des densités estimées qui donne le premier plan principal. Puis on fait croître n . Les graphiques (Fig. 5.6) sont ces premiers plans principaux lorsque n prend les valeurs 5, ..., 40.

On notera que pour une taille d'échantillon de 5 on a déjà des pourcentages d'inertie expliquée et une forme de nuage proches de ceux de la limite qui est donnée par le graphique (Fig. 5.5). De plus au vu de ces simulations, on peut avancer que la convergence est assez rapide.

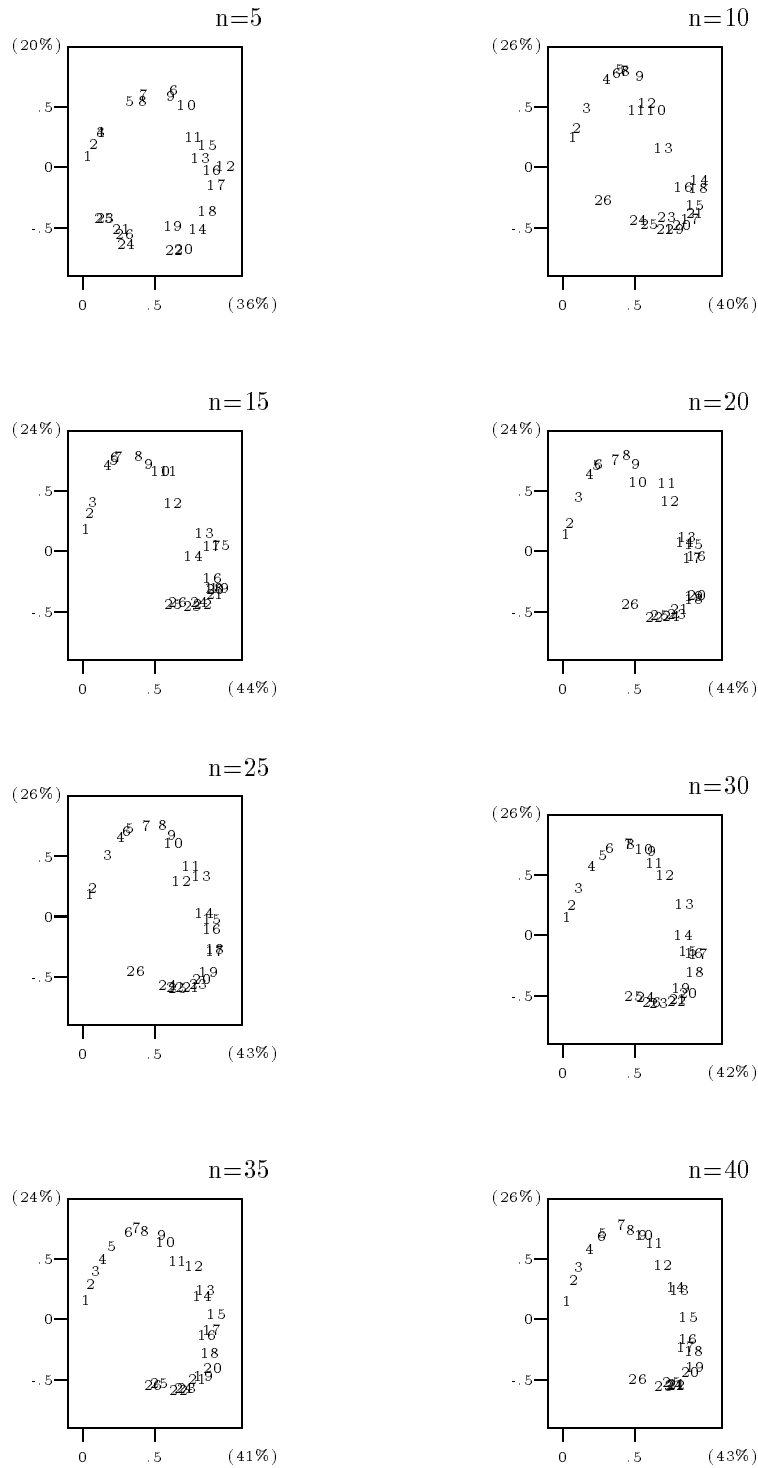


FIG. 5.6 - Premier plan principal pour $n \in \{5, 10, 15, 20, 25, 30, 35, 40\}$

Nous n'avons pas procédé à l'ACP des densités estimées non paramétriquement, car on a déjà remarqué au paragraphe 1.7.3 que les mesures d'affinité entre densités estimées paramétriquement et non paramétriquement étaient équivalentes.

5.2.4 Variation de p , μ , Σ

En observant l'expression de \overline{W} dont le terme général est donné en (5.8), on constate que si μ et Σ varient tout en conservant la quantité $\|\mu\|_{\Sigma^{-1}}$ alors \overline{W} est invariante et par là-même, l'ACP conduit aux mêmes résultats que précédemment. On remarquera au passage que la Σ^{-1} -norme de μ est une généralisation au cas d'une variable statistique vectorielle de l'inverse du coefficient de variation classique d'une variable réelle.

Lorsque $\|\mu\|_{\Sigma^{-1}}$ tend vers zéro on se rapproche du cas où la dérive est nulle (Fig. 5.7). La figure suivante (Fig. 5.8) donne les représentations lorsque $\|\mu\|_{\Sigma^{-1}}$ augmente; la matrice \overline{W} est de la même forme que dans le tableau (Tab. 5.1) avec une baisse régulière des cosinus entre f_1 et f_2 , et une remarquable stabilité des cosinus entre f_{25} et f_{26} (Tab. 5.2).

Si p , la dimension du mouvement brownien, augmente tout en conservant $\|\mu\|_{\Sigma^{-1}}$ constant alors la matrice \overline{W} a toujours l'allure précédente (Tab. 5.1) et les cosinus des angles (f_1, f_2) et (f_{25}, f_{26}) sont donnés dans le tableau (Tab. 5.3). Les représentations (Fig. 5.9) sur les deux premiers plans principaux et les pourcentages d'inertie expliquée sont sensiblement les mêmes.

TAB. 5.2 - Variation de \overline{W} lorsque $\|\mu\|_{\Sigma^{-1}}$ varie.

$\ \mu\ _{\Sigma^{-1}}^2$	0	.1	.5	1	2	5	10
$\overline{W}(1, 2) = \cos(f_1, f_2)$.97	.95	.89	.82	.70	.42	.18
$\overline{W}(25, 26) = \cos(f_{25}, f_{26})$	1	1	1	.99	.98	.95	.91

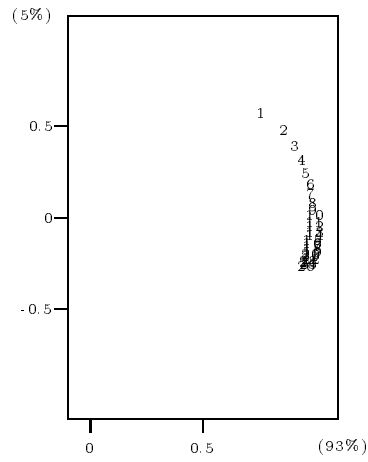


FIG. 5.7 - $\mu = 0$

TAB. 5.3 - Allure de \overline{W} lorsque p varie.

p	1	2	3	10	100
$\overline{W}(1, 2) = \cos(f_1, f_2)$.82	.80	.77	.63	.04
$\overline{W}(25, 26) = \cos(f_{25}, f_{26})$.99	.99	.99	.99	.98

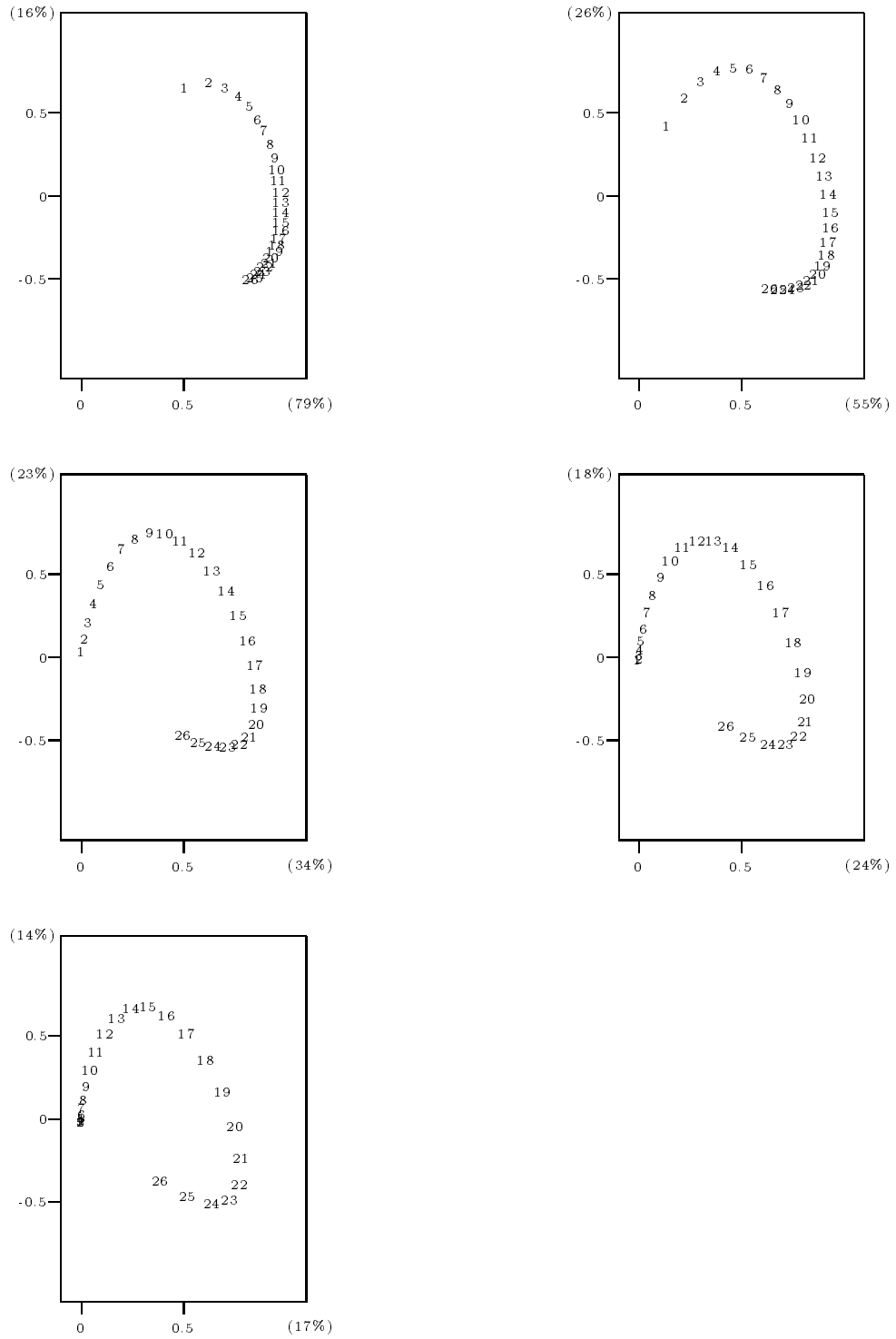


FIG. 5.8 - $\|\mu\|_{\Sigma^{-1}}^2 \in \{0.1, 0.5, 2, 5, 10\}$

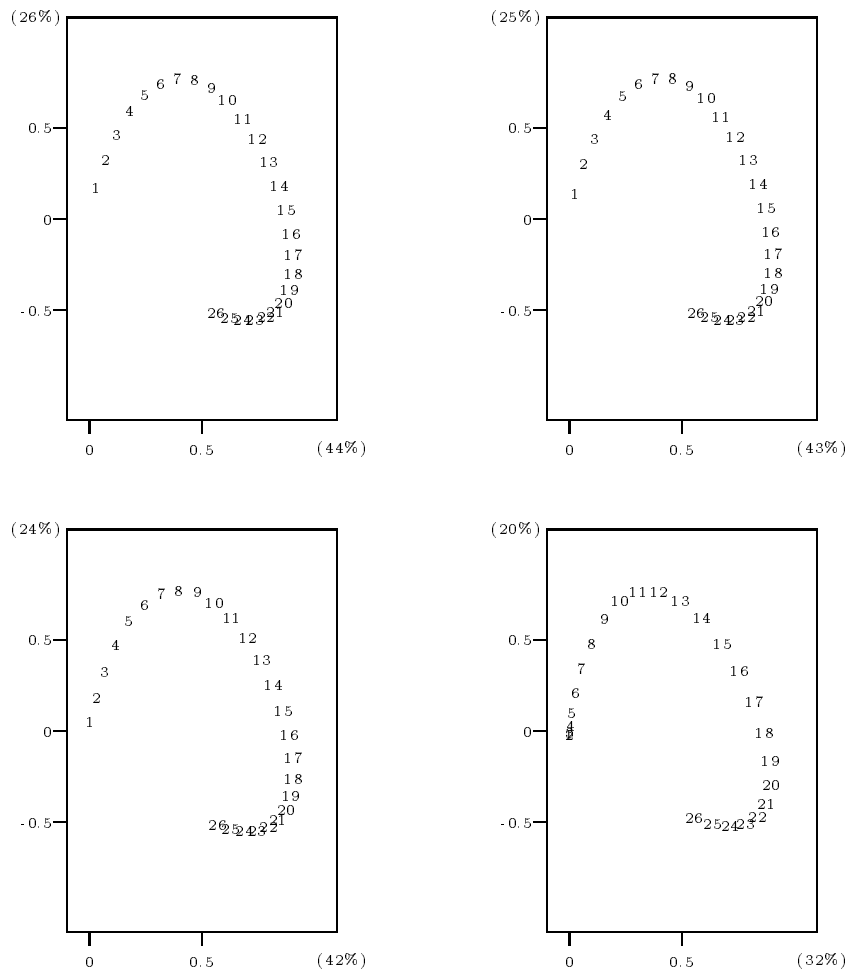


FIG. 5.9 - $\mu = \frac{1}{\sqrt{p}} \mathbf{1}_{R^p}$; $p \in \{2, 3, 10, 100\}$

5.2.5 Comparaison de l'ACP classique et de l'ACP de densités : cas d'une promenade aléatoire

Comme cas particulier du mouvement brownien on considère une promenade aléatoire réelle (Y_t) construite à partir du bruit blanc (ε_t) gaussien de variance σ^2 :

$$\forall t \geq 0, \quad Y_t = \sum_{j=0}^t \varepsilon_j .$$

Les variables aléatoires Y_t ($t = 0, \dots, T$) sont centrées et pour tout t la variance de Y_t est $(t + 1)\sigma^2$. Dans cet exemple, on a pris le cas (arbitraire) T égal à 25, cependant la configuration reste la même pour T plus grand comme on le verra au paragraphe 5.2.6.

L'ACP classique de ces variables conduit à diagonaliser la matrice de corrélation dont le terme général est :

$$R_{ij} = \frac{\min(i, j)}{\sqrt{i} \sqrt{j}},$$

donne la représentation (Fig. 5.10 a) sur le premier cercle des corrélations.

L'ACP normée des densités de probabilité est obtenue en diagonalisant la matrice de terme général :

$$\overline{W}_{ij} = \sqrt{\frac{2\sqrt{ij}}{i+j}};$$

la représentation des densités des variables Y_t sur le premier plan principal est donnée par (Fig. 5.10 b), cette figure étant la reprise de (Fig. 5.7) à une autre échelle.

Dans les deux cas l'évolution temporelle des variables Y_t est bien marquée ; on peut même avancer que la régularité est meilleure pour l'ACP normée des densités.

5.2.6 Variation de T et t_1

Rajouter (resp. enlever) des instants d'échantillonnage revient à ajouter (resp. enlever) des lignes et des colonnes à la matrice \overline{W} (Tab. 5.1). Des expériences numériques menées à partir du cas de référence (Fig. 5.5) montrent une forme analogue des nuages de densités sur les plans principaux et une diminution (resp.

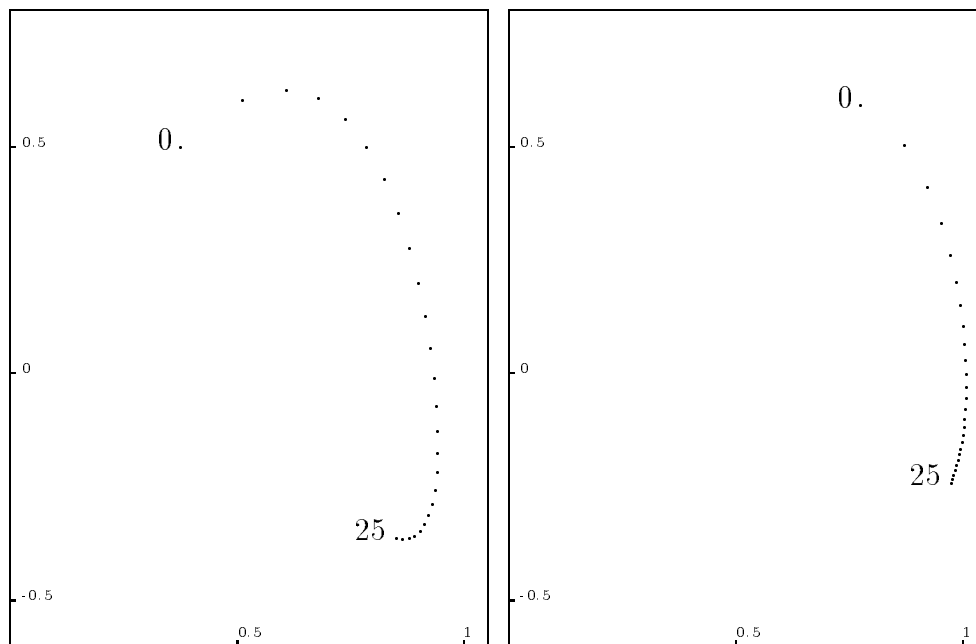
(a) ACP classique des Y_t .(b) ACP des densités des Y_t .

FIG. 5.10 - Promenade aléatoire

augmentation) des pourcentages d'inertie expliquée (Fig. 5.11).

Lorsque t_1 augmente, par exemple à 10, tout en considérant le même nombre de densités ($T = 26$), cela revient à extraire les lignes et colonnes numérotées $t_1, t_1 + 1, \dots, t_1 + 25$ — donc 10, 11, \dots , 35 — d'une des matrices \overline{W} précédentes. Il y a donc un tassement des densités qui augmente de façon importante le pourcentage d'inertie expliquée par le premier axe (Fig. 5.12).

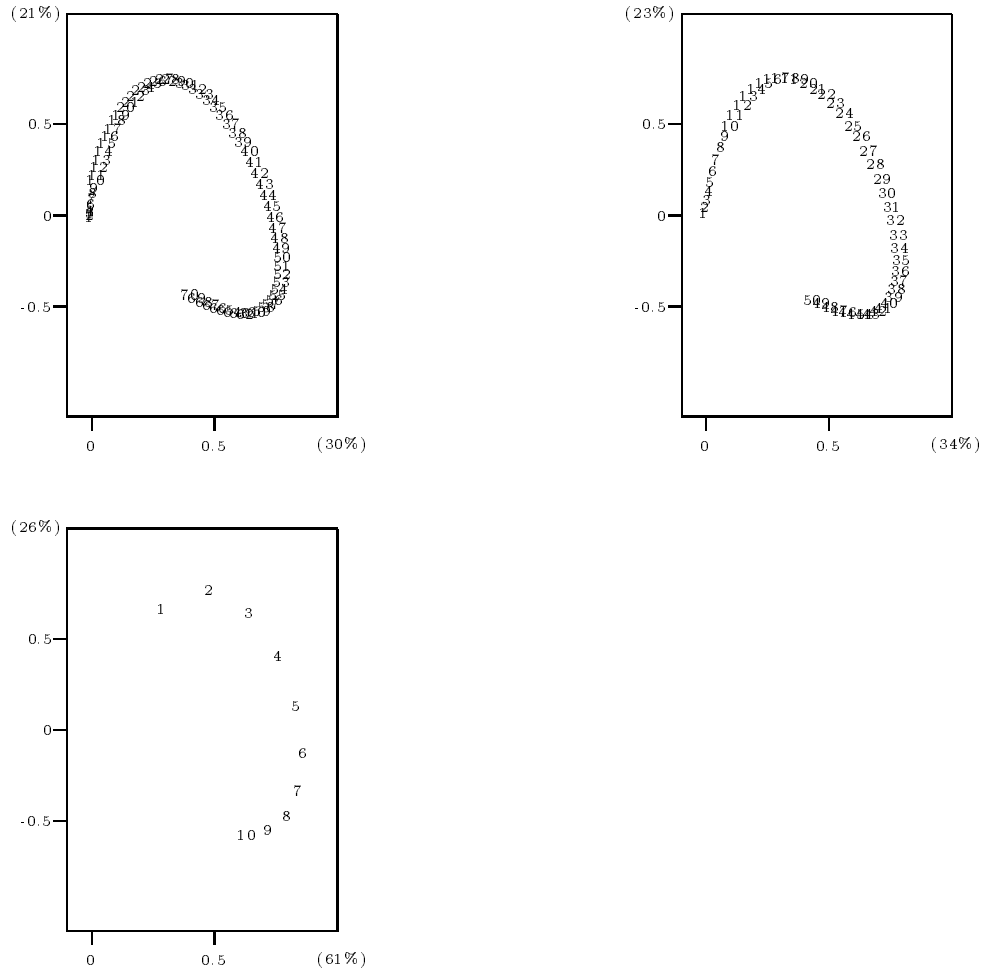


FIG. 5.11 - $T \in \{ 70 , 50 , 10 \}$

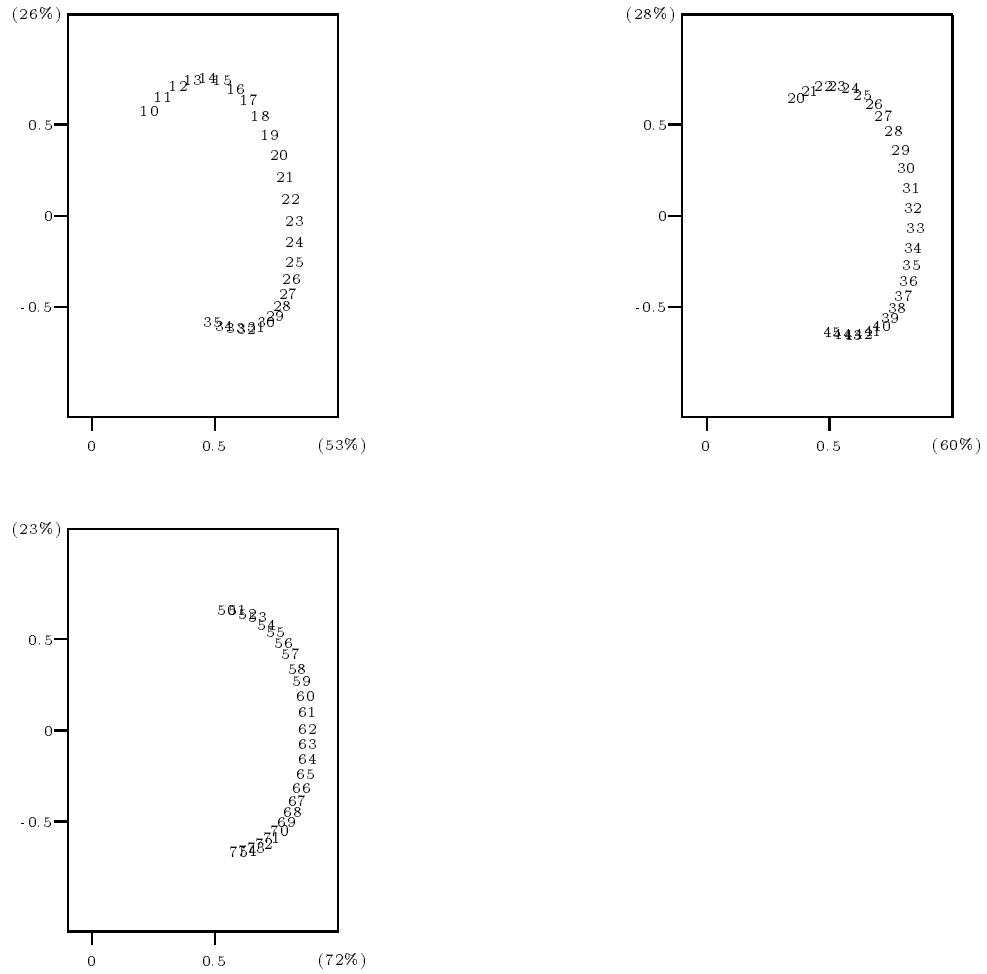


FIG. 5.12 - $t_1 \in \{ 10 , 20 , 50 \}$

5.2.7 Variation de la suite (Δ_k)

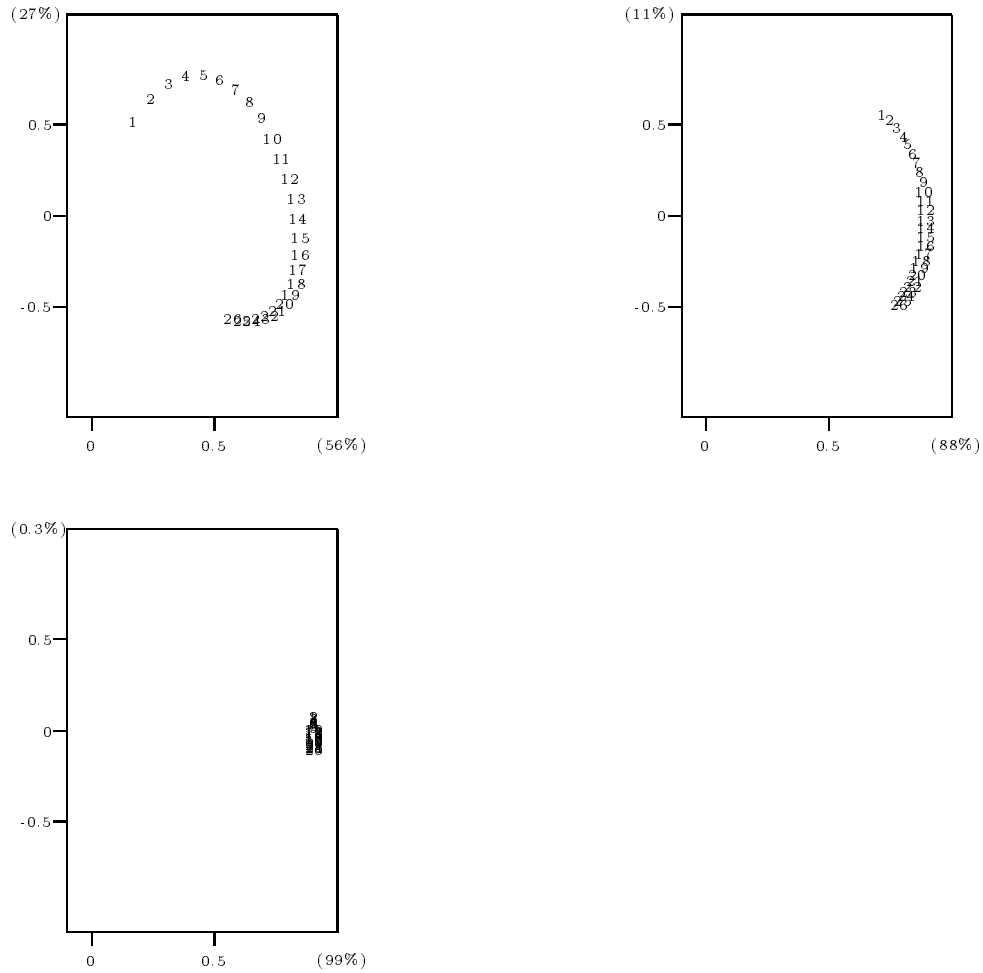


FIG. 5.13 - $\Delta_k \in \{0.5, 0.1, 0.01\}$

Dans tous les cas précédents on a incrémenté le temps d'une unité. Si l'incrément diminue cela revient, comme lorsque t_1 augmente, à opérer un tassement des densités et on retrouve des résultats identiques (Fig. 5.13).

Lorsque l'incrément entre instants consécutifs n'est plus constant toutes les régularités qui sont apparues sont caduques ; à titre d'exemple on tire au hasard 9 instants d'observation différents t_2, \dots, t_{10} entre 2 et 50 auxquels on adjoint

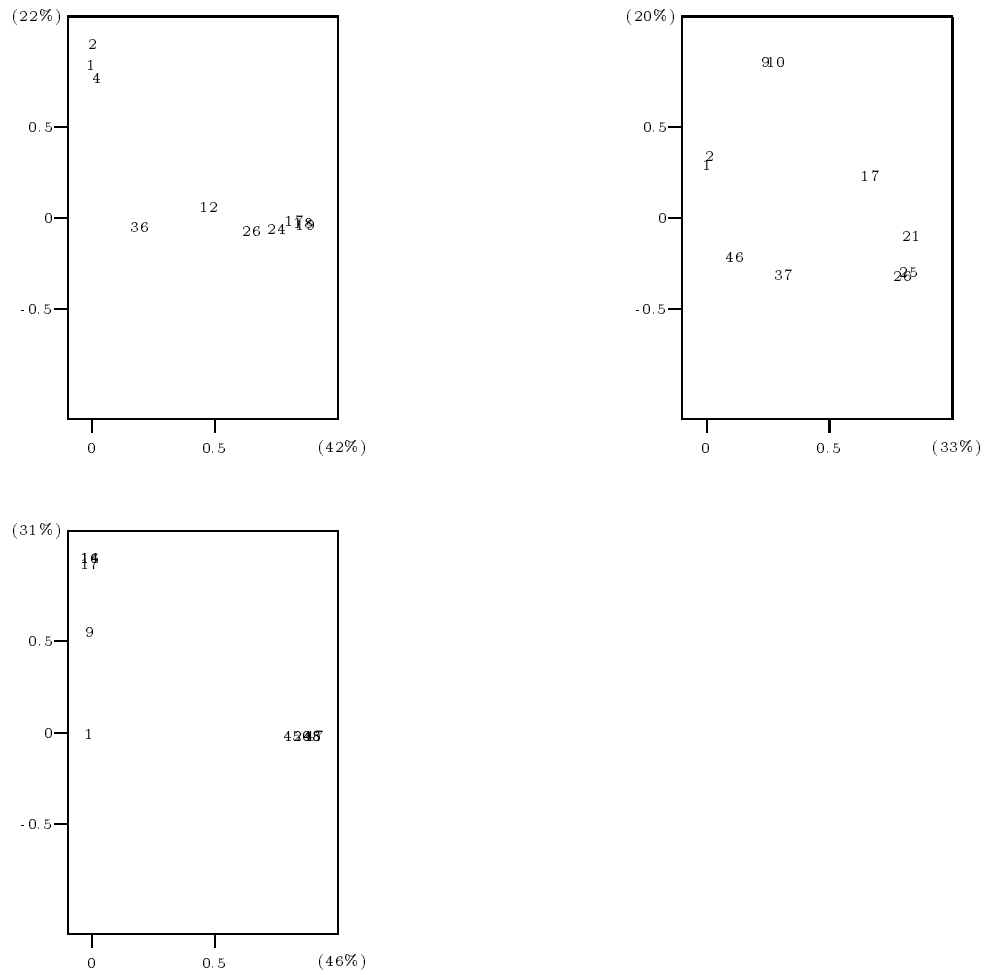


FIG. 5.14 - Les instants pris au hasard sont respectivement pour chaque graphique :

$$\{1, 2, 4, 12, 17, 18, 19, 24, 26, 36\}$$

$$\{1, 2, 9, 10, 17, 21, 25, 26, 37, 46\}$$

$$\{1, 9, 14, 16, 17, 42, 45, 47, 48, 50\}$$

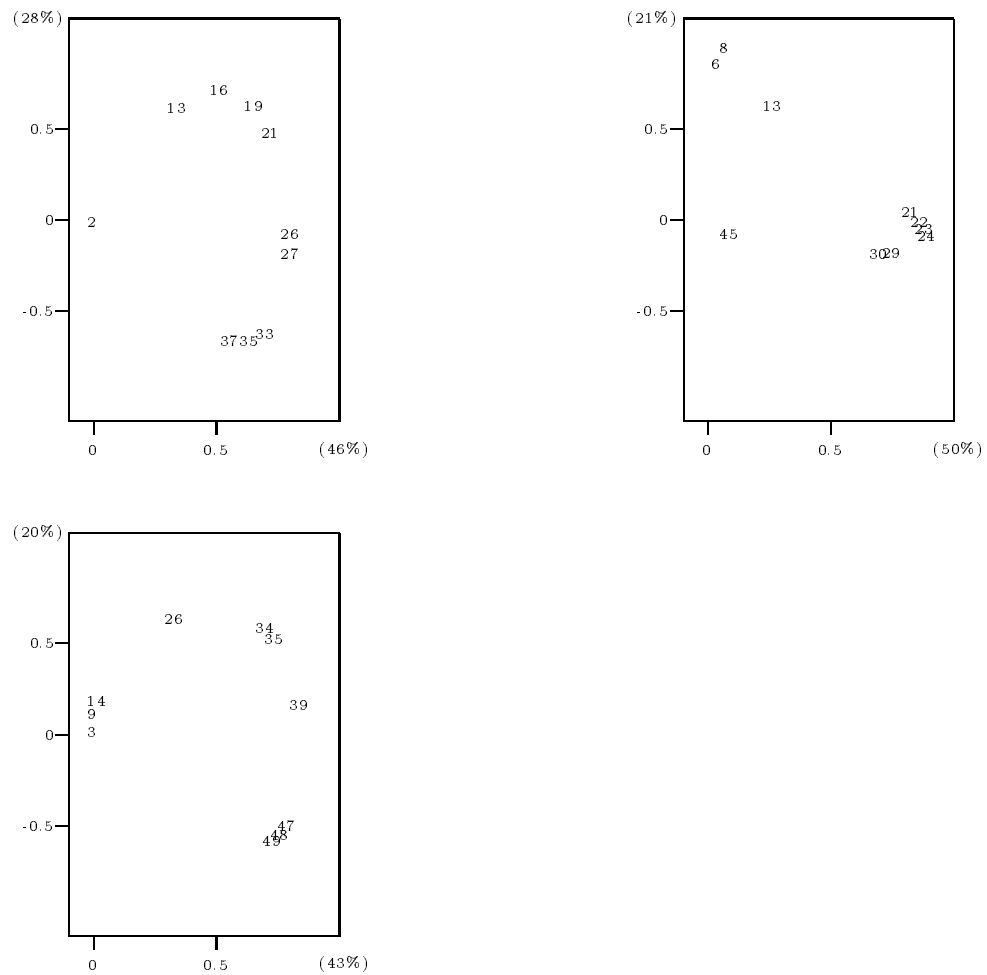


FIG. 5.15 - Les instants pris au hasard sont respectivement pour chaque graphique :

$\{2, 13, 16, 19, 21, 26, 27, 33, 35, 37\}$

$\{6, 8, 13, 21, 22, 23, 24, 29, 30, 45\}$

$\{3, 9, 14, 26, 34, 35, 39, 47, 48, 49\}$

l'origine $t_1 = 1$, puis on fait l'ACP des 10 densités $f_{t_1}, \dots, f_{t_{10}}$; la figure (Fig. 5.14) montre les résultats pour trois simulations.

Enfin on termine par trois simulations (Fig. 5.15) pour lesquelles l'origine t_1 est elle aussi prise au hasard.

Pour tous ces exemples les pourcentages d'inertie expliquée sont très variables et les formes des nuages ne sont pas identiques; mais ce qui peut apparaître comme un inconvénient de la méthode présente en fait un avantage, car comme on peut le constater sur les figures précédentes (Fig. 5.14, Fig. 5.15) le premier plan principal sépare relativement bien les groupes de densités proches dans le temps. L'intérêt de cette remarque est que si pour des données quelconques on ne connaît pas l'instant d'observation on peut raisonnablement l'affecter au groupe de densités dont il est le plus proche.

5.2.8 Conclusion

Si pour le mouvement brownien on a une évolution temporelle "linéaire" des paramètres (μ_t, Σ_t) de la densité gaussienne f_t , puisque :

$$\forall t \in \{1, \dots, T\} : \mu_t = t\mu \text{ et } \Sigma_t = t\Sigma ;$$

le nuage des densités a présenté un continuum régulier bien que non linéaire.

Et pour le cas où le continuum n'est plus évident (incrément du temps variable) alors la méthode présente un intérêt pour faire de la prévision en affectant une période de temps à des données observées mais dont on ignore l'instant d'observation.

5.3 Processus d'Ornstein-Uhlenbeck

5.3.1 Présentation

Comme cas particulier de l'équation (5.1) on considère l'équation différentielle stochastique :

$$\begin{aligned} dX_t &= -\alpha X_t dt + \sigma dB_t \\ X_0 &= x \end{aligned}$$

Sa solution est le processus d'Ornstein-Uhlenbeck. C'est un processus gaussien dont les paramètres μ_t et Σ_t dans le cas unidimensionnel ($p = 1$) sont :

$$\begin{aligned} \mu_t &= x e^{-\alpha t} \\ \Sigma_t &= \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) . \end{aligned} \tag{5.9}$$

C'est un processus qui a la particularité de converger rapidement vers une position d'équilibre pour de grandes valeurs de α positif. Ceci est assez net en observant les paramètres μ_t et Σ_t .

Comment cette constatation se manifeste-telle au niveau des représentations des densités fournies par l'ACP ?

5.3.2 ACP normée des densités

L'ACP normée des densités conduit à diagonaliser la matrice de terme général \overline{W}_{st} égal à :

$$\sqrt{2} \sqrt{\frac{\sqrt{(1 - e^{-2\alpha s})(1 - e^{-2\alpha t})}}{2 - e^{-2\alpha s} - e^{-2\alpha t}}} e^{-\frac{\alpha x^2}{\sigma^2} \frac{(e^{-\alpha s} - e^{-\alpha t})^2}{2 - e^{-2\alpha s} - e^{-2\alpha t}}} . \tag{5.10}$$

La figure (Fig. 5.16) donne la représentation des 26 densités considérées sur le premier plan principal dans le cas où :

$$\sigma = 1 \quad \alpha = 0.1 \quad x = 1 .$$

Comme l'indique cette figure on atteint une position d'équilibre très rapidement, pratiquement à partir de l'instant 10 toutes les densités se projettent en un seul point du plan principal qui notons-le explique 96% de l'inertie du nuage des 26 densités.

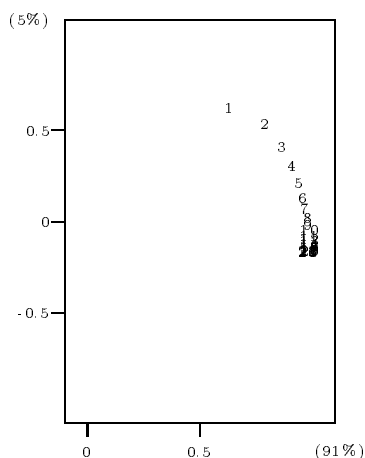


FIG. 5.16 - Processus d'Ornstein-Uhlenbeck

5.4 Données cardiologiques

5.4.1 Les données

Les données¹ concernent 162² opérés du cœur (mise en place de pontages) pour lesquels on a observé plusieurs variables dont 5 paramètres biologiques quantitatifs mesurant la souffrance cellulaire, aussi bien la souffrance globale de tout l'organisme que celle plus particulière du muscle myocardiaque :

- CPK MB,
- CPK,
- Myoglobine,
- Troponine I,

¹Le Dr. Jean-Luc Bosson a cordialement mis ces données à notre disposition par l'intermédiaire de Claudine Robert dont j'ai souvent "squatté" le bureau ; qu'ils en soient vivement remerciés d'autant qu'ils ne trouveront pas ici les réponses souhaitées pour prévenir les risques d'infarctus chez les malades ayant subi des pontages cardiaques.

²Le tableau de données comportait à l'origine 224 individus ; on n'a retenu que les 162 opérés pour lesquels les 5 paramètres biologiques étaient disponibles aux 7 instants d'observation.

– Troponine T,

à 7 instants différents :

1. **I**: induction en salle d'opération,
2. **C**: pendant la circulation sanguine extra corporelle,
3. **D**: après le déclampage aortique,
4. **R**: admission en réanimation,
5. **J 20H**: le jour de l'opération à 20 heures,
6. **J + 1**: à 20 heures le lendemain,
7. **J + 2**: à 20 heures le surlendemain.

Hormis pour les trois derniers instants, le temps écoulé entre deux instants consécutifs ne présente aucune régularité et varie d'un malade à l'autre.

5.4.2 ACP normée des densités

Caractère gaussien des données

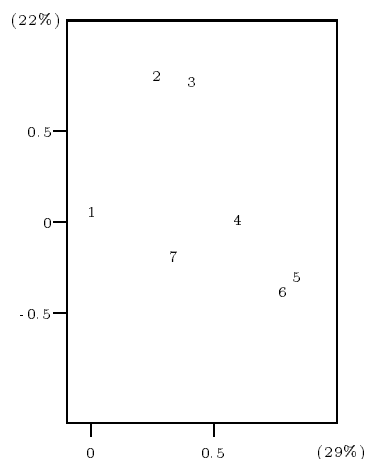
Les histogrammes de chaque variable à chaque instant suggèrent des distributions asymétriques, ce qui a nécessité une transformation logarithmique. Les nouvelles variables (logarithme népérien des paramètres biologiques CPK, CPK MB et Myoglobine) présentent des distributions symétriques et unimodales que nous considérerons gaussiennes. L'hypothèse de log-normalité des deux autres paramètres (Troponine I et T) est beaucoup moins cohérente surtout aux trois premiers instants mais pour conserver l'ensemble des données disponibles nous l'avons admise, en accord avec J.L. Bosson.

Remarque

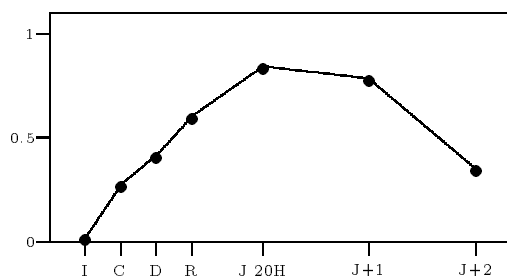
L'objectif recherché par le Dr. Bosson était de définir un score de souffrance cellulaire du myocarde due à l'opération et ce à partir des paramètres biologiques.

L'ACP de ces données ne répond pas directement à cette question. Néanmoins les données fournies constituent un bon exemple de données ternaires pour illustrer cette analyse et pour l'intérêt exploratoire des représentations que permet cette analyse.

Evolution de la souffrance cellulaire



(a) Premier plan principal



(b) Premier axe principal

FIG. 5.17 - Représentation des densités

L'évolution des densités sur le premier plan principal (Fig. 5.17 a) visualise bien l'évolution de la souffrance cellulaire, plus particulièrement le premier axe qui est l'axe vertical de la figure (Fig. 5.17 b).

Cette souffrance est en augmentation constante des instants **1** (induction) à **5** (le soir de l'opération) puis un infléchissement entre l'instant **5** (le soir de l'opération) et l'instant **6** (le lendemain de l'opération) et enfin une diminution plus nette de la souffrance cellulaire entre les instants **6** et **7** (deux jours après l'intervention), ce qui signifie un début de retour à la normale.

5.4.3 Evolution des moyennes temporelles

TAB. 5.4 - Moyennes et écart-types des paramètres biologiques.

	I	C	D	R	J 20H	J+1	J+2
CPK MB	0.8 (0.5)	2.2 (0.4)	2.9 (0.6)	4.3 (0.7)	4.4 (0.7)	3.7 (0.7)	2.2 (0.7)
CPK	3.4 (0.5)	4.3 (0.7)	4.8 (0.7)	6.0 (0.5)	6.8 (0.6)	6.8 (0.5)	6.2 (0.6)
Myoglobine	3.1 (1.2)	5.3 (1.1)	5.8 (1.1)	6.5 (0.9)	7.1 (0.9)	6.8 (1)	5.3 (1.4)
Troponine I	-2.1 (0.7)	-1.9 (0.9)	-1.5 (1.3)	-0.9 (1.5)	-0.3 (1.4)	-0.4 (1.4)	-1.2 (1.3)
Troponine T	-2.0 (0.8)	-1.7 (1.2)	-1.2 (1.5)	-0.5 (1.6)	0.0 (1.7)	-0.0 (1.7)	-0.9 (1.5)

La moyenne temporelle de chaque variable est l'ensemble des moyennes calculées pour chacun des 7 instants (Tab. 5.4). On représente graphiquement ces valeurs par trois courbes (Fig. 5.18) : celle des moyennes, les deux autres s'en déduisent en ajoutant et retranchant l'écart-type de l'instant correspondant. On note que les courbes moyennes de la souffrance cellulaire vont toutes dans le même sens, à savoir croissantes de l'induction en salle d'opération jusqu'au soir à 20 heures puis, à l'exception de CPK MB, à peu près stables entre le soir de l'intervention et le lendemain puis enfin décroissantes au surlendemain.

5.4.4 Conclusion

Les deux descriptions (Fig. 5.17 et Fig. 5.18) de la souffrance cellulaire sont donc similaires mais il était souhaitable de vérifier que les représentations obtenues par l'ACP des densités n'entrent pas en contradiction avec ce que pourraient donner d'autres analyses des données.

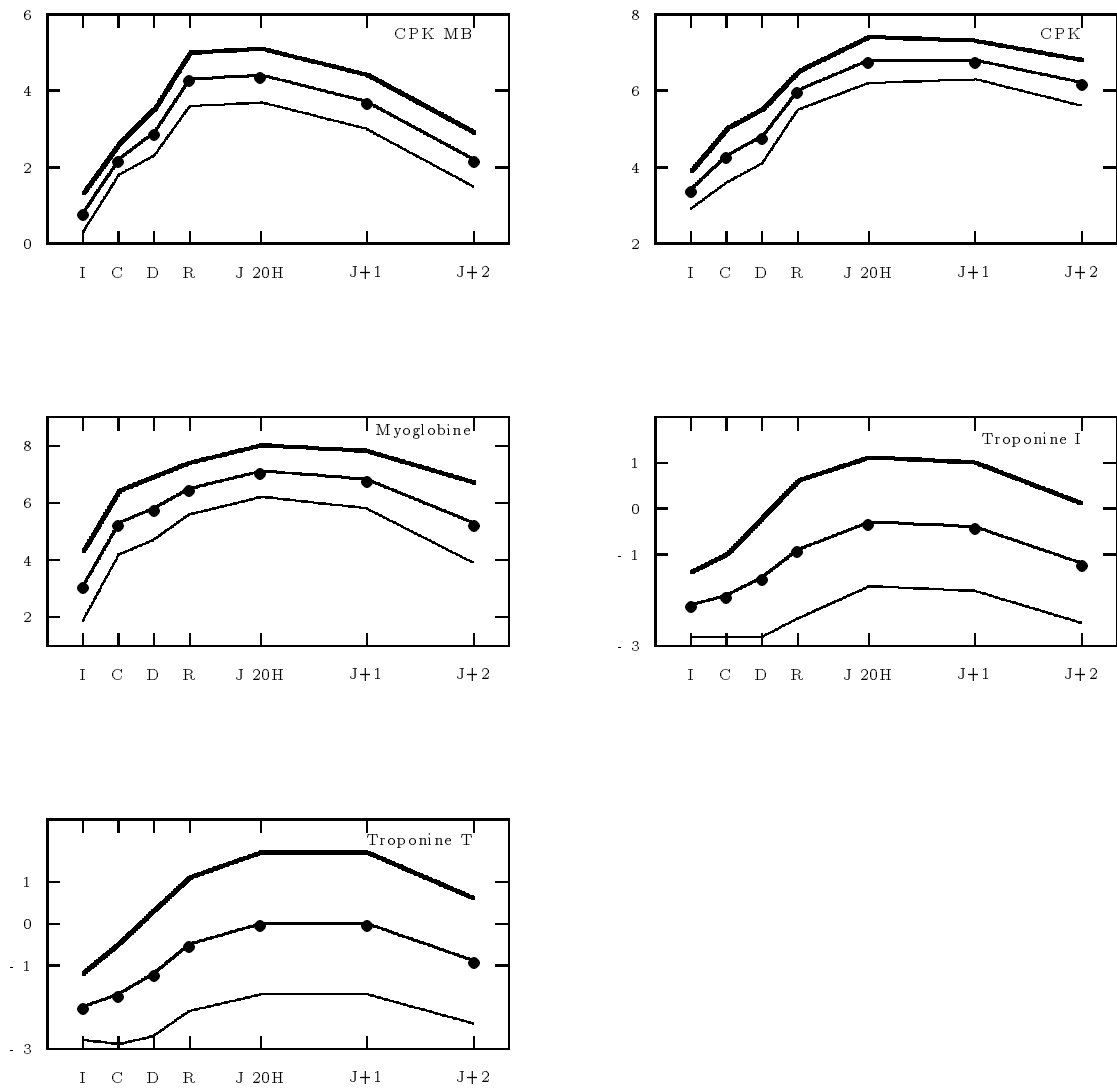
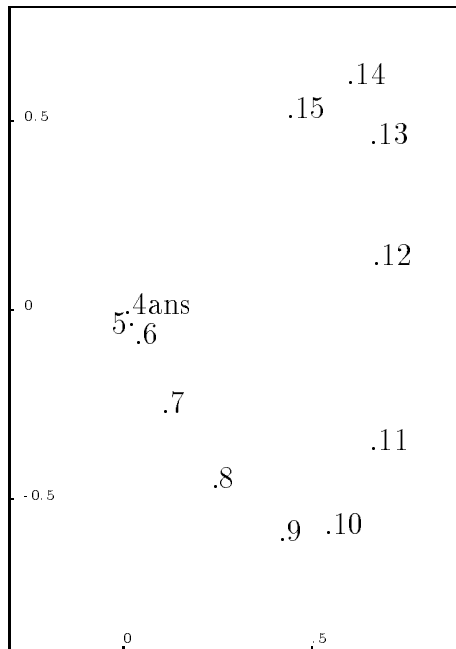
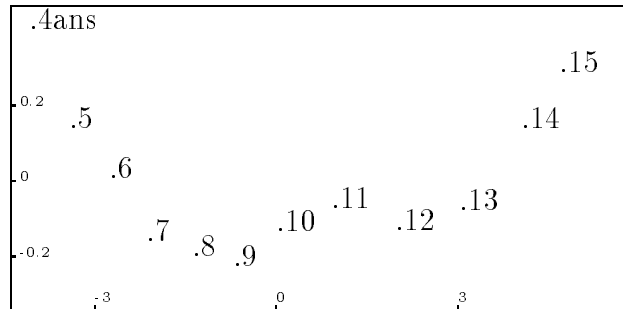


FIG. 5.18 - Evolution des moyennes temporelles des paramètres biologiques.

5.5 Exemple (Lavit 1988)

Nous reprenons les données publiées et analysées de façon détaillée dans (Lavit 1988) au moyen, entre autres, de la méthode STATIS Dual appliquée aux matrices de corrélation. Les données décrivent le suivi de la morphologie (poids, taille, buste, périmètres crânien, thoracique, du bras gauche, du mollet gauche, largeur du bassin) de 30 filles de 4 à 15 ans.

Les premiers plans principaux obtenus par l'application de trois analyses différentes sont donnés dans les graphiques (Fig. 5.19). Le premier plan principal de l'ACP normée des densités de probabilité (Fig. 5.19 b) donne une représentation des densités qui dénote une régularité dans l'évolution morphologique; cette évolution apparaît comme une synthèse des évolutions qui peuvent être observées aussi bien au niveau des moyennes (Fig. 5.19 a) que des matrices de variance (Fig. 5.19 c).



(b) ACP normée des densités.

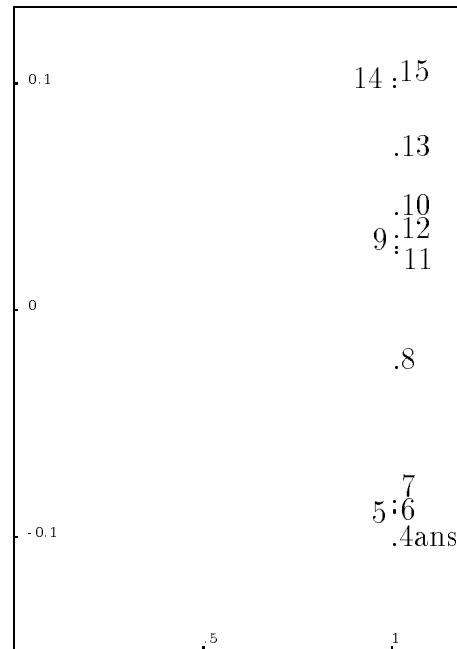
(c) STATIS Dual sur matrices
de variance.

FIG. 5.19 - Exemple (Lavit 1988)

Chapitre 6

Application de l'analyse discriminante à un problème de datation

6.1 Présentation des données

6.1.1 Les données

Les données de châteaux qui seront présentées ont été patiemment et méthodiquement relevées par J.M. Rudrauf¹ (Rudrauf 1987) durant de nombreuses années. D'autres paramètres ont aussi été notés mais ne relevant du champ d'application de l'analyse discriminante de densités, ils ne seront pas présentés. Pour un ou plusieurs éléments de maçonnerie d'un château, des mesures ont été relevées sur n_t pierres prises au hasard parmi les pierres de cet élément t . On dispose de 118 éléments de maçonnerie de 71 châteaux. Par exemple pour le château de Petit-Geroldseck quatre éléments de maçonnerie font partie de l'étude. Il s'agit du donjon, de l'enceinte ouest, de la courtine du logis sud et celle du logis nord : ces éléments portent respectivement les numéros 52, 53, 109 et 110 du tableau (Tab. 6.1). Pour simplifier le langage, l'élément de maçonnerie d'indice t sera appelé château t . De plus, quand on imagine le temps nécessaire à

¹Jean Michel Rudrauf avec qui nous nous apprêtons à réaliser une synthèse de divers travaux sur des châteaux d'Alsace pour publication, nous a fait l'amitié de relire ce chapitre et de le corriger avec le soin qui lui est habituel. Les lecteurs intéressés par l'architecture médiévale et plus particulièrement par les châteaux d'Alsace pourront s'adresser à lui, nous nous ferons un plaisir de transmettre le courrier.

construire un château, on se doute bien que les éléments de ce château n'ont pas toujours été édifiés à la même période, d'où la nécessité de les traiter élément par élément.

Les châteaux proprement dits ne seront pas présentés pour, d'une part, éviter d'alourdir l'exposé et, d'autre part, éviter de conclure trop hâtivement, chaque château nécessitant une étude particulière qui sort du cadre de ce travail.

Les pierres mesurées sont situées entre le sol et une hauteur d'homme. Les p mesures sont prises sur le parement extérieur du château. Ces mesures sont la hauteur, la largeur, le liseré et le bossage (Fig.6.1). Les pierres d'angle et les pierres pour lesquelles une de ces 4 mesures manque (pierre burinée ou pierre cassée) n'ont pas été prises en compte.

Le tableau (Tab. 6.1) donne pour les $T = 117$ châteaux retenus (le château 104

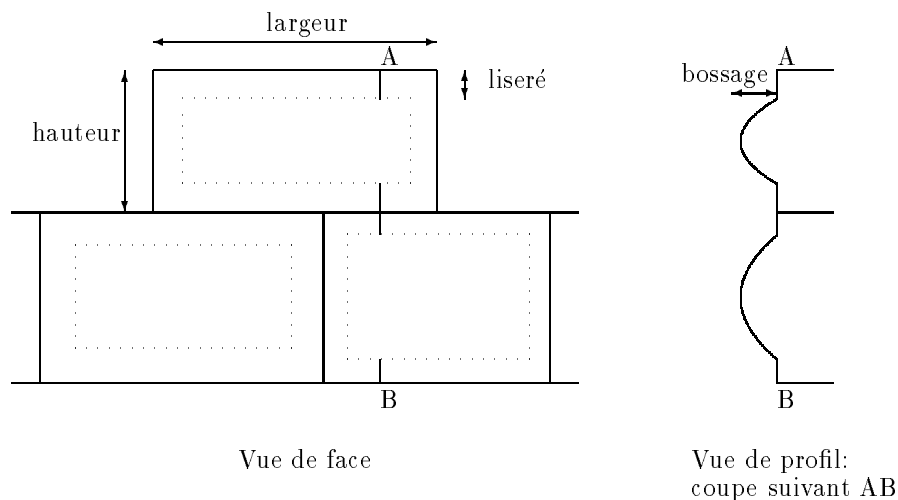


FIG. 6.1 - Mur de château vu de face et de profil

de variance dégénérée n'est pas retenu) le nombre de pierres mesurées, la période d'édification et les moyennes empiriques des p variables observées.

La détermination de la période d'édification d'un château n'est pas toujours

aisée et il n'est nullement ici question d'entrer dans les controverses qu'il peut y avoir à ce sujet (Rudrauf 1987, p.89). En effet il arrive que les historiens s'accordent à un moment donné sur la date d'apparition d'un château que des fouilles archéologiques postérieures remettent en cause. La période retenue dans le tableau (Tab. 6.1) a été reprise des notes personnelles de J.M. Rudrauf tout en soulignant que ce découpage en quatre premières tranches de 35 années puis une tranche de 70 années est un choix arbitraire.

Pour chaque période on a indiqué dans (Tab. 6.2) le nombre de châteaux édifiés au cours de cette période, le nombre de pierres correspondant. La moyenne et la matrice de variance sont calculées à partir de cet échantillon de pierres. Il y a 40 châteaux datés. Les châteaux soulignés dans le tableau (Tab. 6.1) n'ont pas été inclus. Ils seront systématiquement traités comme points supplémentaires :

- Les châteaux 99, 101 et 102 sont d'une période postérieure à 1350 et ne font pas partie de la période étudiée.
- Les autres châteaux soulignés ont un nombre de pierres trop faible (inférieur à 10) pour pouvoir estimer la mesure d'affinité L^2 entre densités ainsi qu'on l'a vu au paragraphe 1.7.1.

6.1.2 Caractère gaussien des données

Si pour tous les châteaux les tailles d'échantillon nous sont apparues trop faibles pour procéder à des tests d'ajustement, nous avons admis ce caractère. Par contre nous avons procédé à ces tests sur les données des 5 périodes et ces tests se sont révélés significatifs.

Ainsi les densités correspondant à des châteaux d'une même période ne peuvent être considérées égales : ce qui est la principale justification théorique (paragraphe 3.2.1 et 4.1) de la mise en œuvre des deux règles probabilistes proposées au chapitre 3 et de la deuxième règle géométrique (paragraphe 4.1.2).

Cependant, nous les avons tout de même appliquées. On montrera les résultats auxquels on aboutit et on les comparera à la première règle géométrique (paragraphe 4.1.1).

TAB. 6.1 - Période d'édification des châteaux et leurs moyennes.

Château	Période	Nombre de pierres n_t	Moyenne				Ecart-type			
			hauteur	largeur	liseré	bossage	hauteur	largeur	liseré	bossage
1	1	25	57.0	75.4	2.5	10.9	4.3	19.6	0.7	2.8
2	1	12	41.5	62.9	2.5	8.6	7.5	9.5	0.4	3.7
3	1	24	36.9	58.8	2.8	6.5	6.3	11.6	0.5	1.7
4	1	14	34.1	52.1	1.7	7.7	3.2	10.8	0.9	3.0
5	1	16	42.1	60.3	3.2	11.3	4.7	11.4	0.8	3.3
<u>6</u>	1	6	64.7	74.3	2.5	10.0	4.5	10.5	0.2	3.2
7	1	12	47.5	61.2	4.0	3.7	1.6	12.4	1.0	1.0
8	1	18	31.2	52.8	3.0	7.4	3.9	10.8	0.8	1.7
9	1	14	40.6	57.0	2.4	8.9	6.4	13.7	0.4	1.8
10	1	42	48.9	69.7	3.0	9.7	10.8	20.8	0.8	4.0
11	3	40	41.0	67.3	4.0	8.6	7.0	15.6	0.6	2.2
12	2	46	34.0	58.4	3.5	9.0	4.2	12.1	0.5	2.2
13	2	20	36.1	61.6	4.1	6.9	5.5	13.4	0.3	1.7
14	2	29	33.5	46.7	2.8	6.7	3.9	8.5	0.4	2.3
15	2	14	41.8	54.2	3.5	8.6	7.7	7.5	0.5	2.2
16		9	34.7	58.8	4.1	9.1	8.5	8.4	0.3	2.3
<u>17</u>	2	5	30.0	52.6	2.8	6.8	5.3	15.0	0.4	0.9
18	3	20	33.1	50.6	3.5	5.3	3.0	9.1	0.8	2.0
19		20	30.9	44.0	2.8	6.0	3.5	6.8	0.4	2.1
20	3	16	27.7	46.4	3.7	6.8	8.0	8.7	0.8	2.2
21	3	15	33.3	53.1	4.0	6.8	3.8	7.8	0.7	1.5
<u>22</u>	2	6	29.7	49.3	2.8	6.1	2.7	8.5	0.2	0.8
23	4	18	32.5	53.7	4.2	6.6	8.1	9.4	0.5	1.5
24		48	27.6	51.1	4.6	4.8	2.5	9.8	0.5	1.2
25	4	26	37.2	52.5	4.3	8.0	4.1	8.1	0.4	2.3
26	4	23	29.3	49.9	3.3	6.4	2.9	12.0	0.4	1.2
27	4	40	32.0	48.1	3.5	6.9	7.1	9.2	0.5	1.8
28		20	29.3	54.4	3.6	6.0	3.2	10.6	0.3	1.7
<u>29</u>	4	8	28.7	55.1	3.0	6.6	3.7	17.6	0.5	1.4
30	4	17	37.6	59.6	3.0	2.8	6.3	7.5	1.7	2.3
31	4	15	33.1	66.7	3.1	5.4	1.9	19.4	0.5	1.8
32		33	28.6	51.2	4.3	6.1	4.3	8.5	0.6	1.6
33		11	30.6	50.9	3.7	6.4	3.3	11.4	0.4	1.7
34		39	31.0	49.2	3.6	6.3	3.0	8.5	0.5	1.4
<u>35</u>	5	8	25.6	47.8	5.5	4.8	0.4	11.6	0.6	1.0
36	5	15	32.7	62.1	4.3	5.6	2.3	22.4	1.0	0.6
<u>37</u>	5	9	22.3	44.7	3.1	5.3	1.4	7.6	0.8	1.6
38	5	38	37.5	55.0	3.5	7.5	4.6	14.3	0.5	1.8
39	5	13	29.8	43.8	4.8	5.4	3.8	7.1	0.4	1.0
<u>40</u>	5	7	40.0	61.9	5.1	7.1	1.7	9.5	1.2	1.2

Château t	Période $Y(t)$	Nombre de pierres n_t	M o y e n n e				E c a r t - t y p e			
			hauteur	largeur	liseré	bossage	hauteur	largeur	liseré	bossage
41	4	28	33.9	61.0	4.4	7.1	3.4	12.6	0.7	1.5
<u>42</u>	5	5	28.8	68.0	5.7	8.2	8.7	27.6	1.9	3.6
43		23	46.9	78.8	3.2	8.3	5.9	16.9	0.4	2.5
44		12	27.3	52.1	3.5	4.7	3.2	8.2	0.4	1.3
45		20	33.0	58.7	2.8	9.0	3.1	10.4	0.6	1.7
46		21	45.8	55.1	2.5	10.7	6.9	15.1	0.4	3.6
47		11	38.9	63.5	3.4	8.3	4.6	12.5	0.9	2.7
48		15	32.3	57.7	2.9	7.7	1.8	14.8	0.5	1.4
49		10	35.9	62.7	4.1	5.5	5.5	8.0	1.0	1.6
50		27	30.0	51.9	3.7	7.4	3.6	15.2	0.4	2.0
51		16	36.9	58.6	2.4	7.8	7.2	11.5	0.4	2.1
52		35	31.3	56.0	3.7	7.2	3.4	10.8	0.4	2.1
53		29	29.2	51.6	3.5	7.6	4.5	12.4	0.4	1.9
54		11	37.2	51.7	3.1	7.3	5.0	12.3	0.9	2.9
55		15	26.0	53.3	3.9	4.5	3.2	12.4	0.8	1.4
56	1	46	51.3	71.6	3.0	12.3	5.1	14.2	0.8	3.4
57		31	24.6	62.3	3.1	6.4	2.6	14.1	0.3	1.1
58		24	28.2	51.3	3.3	5.8	2.7	10.0	0.9	1.3
59		17	34.6	59.5	2.2	8.1	3.0	12.3	0.4	1.7
60	1	49	35.7	56.2	2.9	7.4	6.2	12.3	0.4	1.8
61		23	30.6	66.7	3.0	9.7	4.4	14.4	0.3	2.2
62	1	20	47.7	68.5	2.9	10.6	7.6	15.3	0.4	1.8
63	4	11	26.9	47.6	3.4	4.8	4.3	9.1	0.4	0.9
64		14	29.6	55.7	3.7	6.4	3.9	10.9	0.7	2.1
65		22	32.1	57.7	4.0	6.0	4.0	8.7	0.4	1.1
66		37	30.8	50.0	4.0	7.5	5.1	10.5	0.5	1.7
<u>67</u>	4	5	34.8	58.4	3.4	8.8	6.0	16.4	1.3	1.4
68		29	33.6	61.3	3.0	7.9	5.8	13.1	0.4	2.1
69		16	34.8	62.4	3.6	6.4	3.9	12.3	0.9	2.3
70	4	25	34.6	55.6	5.1	5.7	6.3	9.8	0.7	1.4
71	4	14	26.4	42.3	3.5	4.3	2.1	12.9	0.5	0.6
72		15	26.4	64.3	3.1	5.4	4.0	13.4	0.4	1.7
73		8	46.2	60.9	4.6	8.7	0.9	7.3	0.3	1.7
74		29	32.2	55.3	2.9	7.0	3.3	10.6	0.4	1.6
75		12	34.9	62.3	3.8	5.0	3.5	10.2	1.9	3.9
76		36	31.1	58.8	4.5	6.3	2.7	12.3	0.5	1.4
77		22	31.3	54.1	3.0	5.1	4.4	12.9	0.9	2.7
78		10	28.0	47.4	4.2	6.3	2.6	7.1	0.7	1.3
79		15	36.5	50.5	3.2	5.9	1.8	7.3	0.8	2.0
80		25	39.6	63.6	2.5	9.2	6.7	15.5	0.3	2.0
81	4	13	27.4	46.1	3.0	6.3	4.0	7.2	0.4	1.3
82	4	16	40.3	73.8	4.5	5.4	3.7	12.0	0.7	1.6
83		11	26.8	56.2	3.8	6.1	2.6	13.1	0.4	2.1
84		8	29.0	58.7	3.4	7.2	1.0	11.2	0.3	1.3
85	2	35	32.4	50.6	2.3	7.5	6.0	10.1	0.5	2.3

Château t	Période $Y(t)$	Nombre de pierres n_t	Moyenne				Ecart-type			
			hauteur	largeur	liseré	bossage	hauteur	largeur	liseré	bossage
86		24	38.1	60.3	3.4	9.2	4.3	15.2	0.5	2.6
87		28	30.0	57.5	4.4	6.0	3.1	13.3	0.4	1.0
88		11	37.6	56.4	3.3	8.5	9.7	7.6	0.4	1.4
89		11	47.5	59.5	3.9	7.0	4.6	10.9	0.4	2.3
90	2	11	37.3	61.5	3.7	6.0	3.5	15.5	0.2	0.9
91		9	41	64.8	4.9	6.1	4.7	16.9	0.5	1.2
<u>92</u>	3	5	27.2	56.2	3.2	6.2	1.8	9.3	0.2	0.7
93		21	29.4	71.6	5.4	6.3	1.2	17.8	0.83	1.4
94		7	29.4	58.3	3.7	7.7	2.9	15.6	0.2	1.8
95		21	31.6	50.7	3.8	7.2	5.7	8.5	0.3	1.7
96		17	37.2	67.6	6.0	8.7	4.3	14.5	1.5	2.6
97		16	25.9	54.2	3.1	6.3	4.5	8.2	0.4	1.6
98	4	27	29.1	51.1	3.6	7.4	1.8	6.3	0.4	1.8
<u>99</u>	6	12	38.7	59.4	6.2	5.8	2.4	11.2	0.9	1.7
100		11	36.4	69.8	4.0	6.1	5.2	26.9	1.2	2.5
<u>101</u>	6	22	28.4	44.3	3.9	5.8	4.9	7.1	0.6	1.5
<u>102</u>	6	10	32.0	59.6	3.3	5.7	4.1	14.6	0.8	1.5
103		20	30.8	58	3.5	5.3	2.4	13.5	0.7	1.42
(104)		4	33.0	40.8	2.5	6.0	<u>0.0</u>	3.7	0.3	0.7
105		15	36.3	50.7	4.3	5.8	6.7	15.1	0.6	1.0
106		17	38.8	70.9	5.2	8.0	8.4	18.5	0.7	2.4
107		16	37.9	69.9	5.2	7.6	7.9	18.6	0.7	1.8
108		16	33.0	53.5	3.5	8.0	5.2	11.5	0.6	2.1
109		25	31.4	49.1	2.9	6.8	3.7	11.5	0.5	1.8
110		15	24.5	48.7	2.4	7.8	0.4	9.0	0.2	1.7
111	3	17	36.9	55.6	4.0	8.1	3.1	9.8	0.3	1.6
112		8	28.9	68.7	4.3	5.7	7.0	17.0	0.5	1.0
113		7	25.3	57.1	3.2	5.1	1.4	12.3	0.5	1.4
114		15	28.5	50.1	2.4	7.2	3.8	8.2	0.5	1.1
115		17	26.3	63.2	3.8	7.8	6.6	17.3	0.6	2.8
116		25	28.7	51.4	3.8	7.0	4.4	12.6	0.4	1.7
117		7	33.7	62.3	3.6	8.0	2.3	34.8	0.5	1.7
118	1	20	45.0	61.7	2.9	4.1	5.2	10.9	0.6	1.5

6.1.3 Caractère évolutif des données

L'objectif poursuivi est de dater des (éléments de maçonnerie de) châteaux à partir de mesures sur leurs pierres. L'hypothèse forte sur laquelle s'appuie cette démarche est que l'évolution temporelle des 4 paramètres mesurés sur les pierres est régulière.

Ce caractère évolutif peut être observé de façon nette au niveau des moyennes des variables liseré et bossage (Tab. 6.2) : le liseré ayant tendance à augmenter,

TAB. 6.2 - Moyenne et variance par période.

Période	Nombre de châteaux	Nombre de pierres	Moyenne				Variance									
1-Entre 1140 et 1175	13	312	43.8	63.5	2.9	8.8	97.8	81.9	266	0.44	-0.17	0.64	13.2	13.2	-0.23	13.7
2-Entre 1175 et 1210	6	155	34.8	54.7	3.2	7.7	33.8	12.1	155	1.05	1.81	0.59	3.46	2.90	0.05	5.77
3-Entre 1210 et 1245	5	108	35.9	57.3	3.9	7.4	55.1	29.3	206	1.27	0.22	0.55	8.79	4.66	0.40	5.56
4-Entre 1245 et 1280	13	273	32.6	54.2	3.8	6.2	39.0	20.0	171	1.90	2.16	0.92	1.61	-1.15	0.42	4.64
1-Entre 1280 et 1350	3	66	34.9	54.4	4	6.7	26.4	24.3	277	-1.19	-1.94	0.68	5.62	8.03	-0.38	3.24

le bossage à diminuer, ce qui est assez bien connu.

Pour les deux autres variables, la différence est aussi nette : la première période est particulière par de grandes valeurs de la hauteur et de la largeur.

Pour les matrices de variance, la lecture des évolutions est beaucoup moins aisée aussi bien au niveau des variances des variables que de leurs covariances : le bossage a par exemple une variance qui a tendance à baisser mais ceci n'est pas le cas pour le liseré. Ces évolutions très irrégulières ne seront pas sans incidence sur les résultats fournis par l'analyse en composantes principales et l'analyse discriminante des densités.

Bien que le pourcentage d'inertie expliqué par le premier plan principal (Fig. 6.2) soit faible (32%) ce plan reflète assez bien une certaine évolution temporelle. En effet, on peut remarquer que les châteaux de la période 1 et ceux des autres périodes sont disposés de part et d'autre de la ligne (a). De plus, on peut noter que les châteaux de la troisième période sont situés à droite entre les lignes (b) et (c) et les châteaux des périodes 4 et 5 sont soit en zone centrale, entre les lignes (a) et (b), soit dans la zone inférieure, en dessous de (c). Quant aux châteaux de la deuxième période ils se répartissent quasi verticalement dans ces différentes zones.

Cette évolution peut aussi se voir dans la disposition des centres de gravité, notés G , et des densités, notées g , associés aux périodes ; ces points ayant été traités comme points supplémentaires.

Cependant cette évolution ne permet pas de séparer efficacement les périodes, ce qui se ressentira très nettement au niveau de l'analyse discriminante. Aux chapitres 3 et 4, on a introduit quatre règles d'affectation. Avant de procéder à leur application aux châteaux non datés on va dans un premier paragraphe appliquer ces règles sur les châteaux datés.

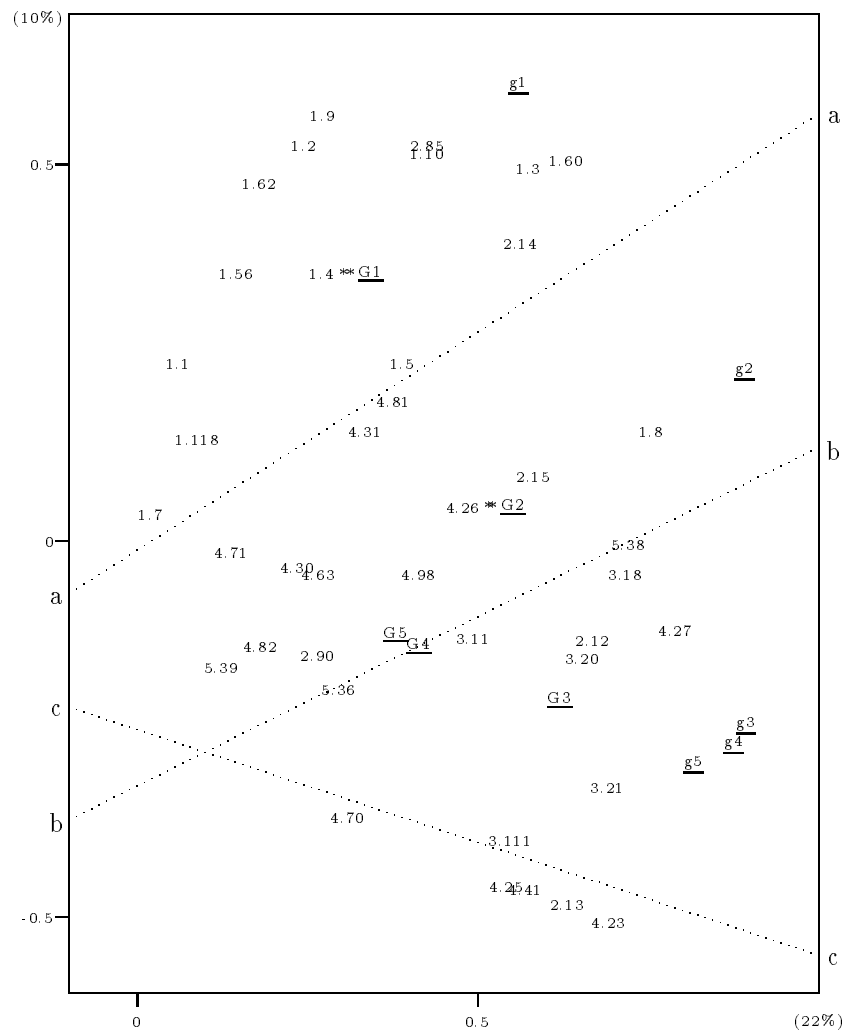


FIG. 6.2 - Premier plan principal de l'ACP des châteaux datés : un château est désigné par son numéro précédé de sa période d'édification.

6.2 Validation empirique des règles d'affectation sur les châteaux datés

Comme cela est habituel en analyse discriminante lorsqu'on est en présence d'un petit nombre d'individus, ici 40, on procède à 40 analyses sur les châteaux datés en omettant à chaque analyse un des 40 châteaux datés qu'on affecte au terme de la dite analyse. Puis on compte le nombre de fois où cette affectation est erronée.

Les tableaux (Tab. 6.3) indiquent que le maximum que l'on puisse espérer ne peut dépasser les 50% de bien classés et ce avec la règle géométrique 1 "distance au centre de gravité" (paragraphe 4.1.1).

Cependant, si les pourcentages de bon classement ne sont pas très élevés les erreurs importantes de datation (erreur absolue de plus d'une période) sont faibles, notamment pour la règle géométrique 1 : la proportion de bon classement (en admettant une erreur absolue d'une période) est de 32 sur 40.

A notre avis ce type d'erreur est admissible par le fait même que la datation d'un château est connue dans les meilleurs des cas à une dizaine d'années près.

Compte tenu de tout ce qui précède, on opte pour la méthode d'affectation avec la règle géométrique 1 pour "prédire" la période d'édification des éléments de maçonnerie non datés.

6.3 Affectation des châteaux et conclusion

Prenons tout d'abord la précaution de dire ici que les résultats donnés ci-après sont d'un apport certain certes, mais dans le cadre de l'hypothèse d'une évolution temporelle des mesures des pierres.

Le tableau (Tab. 6.4) donne l'affectation à une période de chaque élément de maçonnerie par la distance minimale au centre de gravité de chaque période. Les centres de gravité sont déterminés à partir des 40 châteaux qu'on a considérés datés et le mode de calcul des distances est le même pour tous les châteaux, datés ou non. C'est ce qui explique que la proportion de châteaux datés bien affectés passe de 21 à 27 sur 40.

 TAB. 6.3 - Tableaux des classements par règle.

REGLE PROBABILISTE 1

Période d'affectation				
1	2	3	4	5
7	6			
	2	3		1
			2	3
	2	3	3	5
	1		2	

Bien classés : 12 sur 40

REGLE PROBABILISTE 2

Période d'affectation				
1	2	3	4	5
9	3		1	
4		1		1
4			1	
4	2	4	2	1
			3	

Bien classés : 11 sur 40

REGLE GEOMETRIQUE 1

Période d'affectation				
1	2	3	4	5
12	1			
2	1	3		
1	1	2	1	
3	1	3	6	
1	2			

Bien classés : 21 sur 40

REGLE GEOMETRIQUE 2

Période d'affectation				
1	2	3	4	5
10	3			
1	1	3		1
1			2	2
1	2	3	5	2
	1		2	

Bien classés : 16 sur 40

Nous ne commenterons pas davantage ce tableau. Cela dépasse largement le cadre d'un travail mathématique mais c'est là que tout commence pour un travail d'histoire médiévale que nous pensons continuer avec Jean Michel Rudrauf.

TAB. 6.4 - Châteaux, leurs éléments de maçonnerie, leurs datations.

Nom	Élément	Numéro <i>t</i>	Période	Affectation
Arnsberg (Grand-)	donjon	43		1
Arnsberg (Grand-)	enceinte extérieure	44		4
Balbronn	tour-habitat	45		1
Bilstein-Aubure	donjon	11	3	1
Bilstein-Urbeis	donjon	46		1
Birkenfels	donjon	23	4	3
Birkenfels	logis	70	4	4
Burgstall	enceinte	80		1
Dagsburg / Eguisheim	donjon	1	1	1
Dreistein occidental	bouclier	32		4
Dreistein occidental	enceinte nord logis	116		4
Dreistein oriental	enceinte ouest logis	24		5
Falkenstein	enceinte	48		1
Falkenstein	tour adossée au rocher	<u>101</u>	6	4
Fleckenstein	tour du puits	49		2
Fleckenstein	tour d'escalier	<u>102</u>	6	2
Fleckenstein	enceinte	103		4
Fleckenstein	donjon	(104)		
Frankenburg	enceinte	3	1	1
Frankenburg	donjon	50		4
Freudeneck	enceinte ouest haut-château	105		3
Freudeneck	enceinte nord-est haut-château	106		1
Freudeneck	base enceinte est haut-château	107		1
Geroldseck (Grand-)	donjon	12	2	3
Geroldseck (Grand-)	enceinte logis-nord	51		1
Geroldseck (Grand-)	enceinte est	108		3
Geroldseck (Petit-)	donjon	52		4
Geroldseck (Petit-)	enceinte ouest logis-sud	53		4
Geroldseck (Petit-)	enceinte est logis-sud	109		4
Geroldseck (Petit-)	enceinte logis-nord	110		1
Girbaden (Nouveau-)	donjon	18	3	3
Girbaden (Nouveau-)	enceinte	111	3	3
Girbaden (Vieux-)	enceinte sud	4	1	1
Girbaden (Vieux-)	enceinte nord	54		1
Girbaden (Vieux-)	fausses braies	33		4
Girbaden (Vieux-)	mur transversal	55		4
Greifenstein (Grand-)	donjon	56	1	1
Greifenstein (Petit-)	donjon	34		4
Gutenberg	donjon	113		4

Nom	Élément	Numéro <i>t</i>	Période	Affectation
Hagelschloss	enceinte dans fossé	16		3
Hagelschloss	enceinte basse-cour	57		4
Hagelschloss	donjon	114		1
Hagelschloss	mur renversé et enceinte ouest haut-château	115		1
Haut-Barr	tour rocher nord	5	1	1
Haut-Barr	tour du puits	58		4
Haut-Barr	donjon du rocher nord	117		3
Haut-Koenigsburg	bouclier roman	<u>6</u>	1	1
Haut-Koenigsburg	donjon	62	1	1
Haut-Ribeaupierre	donjon	25	4	3
Heidenschlossfels	pierres au sol	112		4
Helfenstein	enceinte extérieure sud	<u>35</u>	5	1
Herrenstein	bouclier	13	2	3
Herrenstein	bastion 16è	<u>99</u>	6	1
Herrlisheim		36	5	5
Hohenburg	enceinte ouest	26	4	4
Hohenburg	mur adossé au rocher au-dessus escalier	97		4
Hohenburg	enceinte sud	98	4	4
Hohenfels	logis	<u>37</u>	5	4
Hohenfels	enceinte du ressaut rocheux	96		1
Hohenstein	donjon	59		1
Hohenstein	enceinte est	94		4
Hohenstein	enceinte nord	95		3
Hohnack (Petit-)	donjon	60	1	1
Ingwiller	tour ronde	93		1
Kronenburg		<u>92</u>	3	4
Landsberg	donjon	14	2	2
Landsberg	tour d'angle nord-ouest	19		4
Lemberg	enceinte nord	91		1
Lichtenberg	donjon	90	2	2
Loewenstein	pierres au sol	63	4	4
Lutzelburg	donjon carré au centre façade sud	7	1	1
Lutzelburg	donjon carré à côté porte	79		2
Lutzelburg	donjon pentagonal	118	1	1
Lutzelburg (Hinter-)	tour habitat	15	2	2
Lutzelburg (Hinter-)	agrandissement noyau primitif	64		4
Lutzelburg (Hinter-)	donjon	65		3
Lutzelburg (Vorder-)	enceinte ouest	66		3
Lutzelburg (Vorder-)	tour primitive	88		2
Lutzelburg (Vorder-)	donjon	89		1
Meywahr		<u>67</u>	4	3
Mittelburg / Eguisheim	donjon	2	1	1
Nideck inférieur	donjon	38	5	2
Nideck supérieur	enceinte	27	4	4

Nom	Élément	Numéro <i>t</i>	Période	Affectation
Ochsenstein	enceinte	87		5
Oedenburg	enceinte nord logis	28		4
Oedenburg	donjon	68		1
Ottrott	tour ronde du château primitif	8	1	2
Ottrott-bas	chaînage d'angle	<u>17</u>	2	1
Pfalz / Eguisheim	enceinte	61		1
Pierre-Percée	donjon	69		2
Ramstein-Baerenthal	enceinte logis	71	4	4
Reichenstein	donjon	<u>29</u>	4	4
Ringelsberg (Petit-)	pierres dans carrière	86		2
Ringelstein	enceinte est haut-château	72		4
Ringelstein	mur contre-bas rocher est	73		1
Ringelstein	enceinte ouest	74		4
Rothenburg	enceinte haut-château	39	5	5
Saint Rémy	tour d'angle sud-est	<u>40</u>	5	1
Saint Ulrich	donjon	9	1	1
Saint Ulrich	tour-habitat	85	2	1
Salm	enceinte est	20	3	4
Scharrachbergheim	socle enceinte	84		4
Schoeneck	donjon	75		1
Waldeck	donjon	41	4	3
Wangenburg	enceinte nord	76		3
Wangenburg	donjon	83		4
Warthenberg	donjon et bouclier	10	1	1
Warthenberg	mur transversal	100		1
Wasenburg	logis	30	4	1
Wasenburg	bouclier	82	4	1
Wasigenstein (Grand-)	donjon	31	4	2
Wasigenstein (Petit-)	tour-habitat	<u>42</u>	5	1
Weckmund / Eguisheim	donjon	47		1
Wegelnburg	enceinte haut-château	81	4	4
Windstein (Nouveau-)	tour-habitat	21	3	3
Windstein (Vieux-)	tour-habitat château occidental	77		1
Windstein (Vieux-)	enceinte sud	78		4
Wineck	donjon (chaînage d'angle)	<u>22</u>	2	4

Conclusion

Dans ce travail, nous avons introduit et explicité le concept d'analyses "factorielles" de distributions de probabilité de vecteurs aléatoires que nous avons trouvé à l'état latent dans la méthode STATIS sur matrices de variance, élargissant ainsi les techniques d'investigation des tableaux de données à trois indices. Nous avons illustré ce concept par des exemples pour en voir les possibilités d'application.

Ayant défini une distance entre distributions de probabilité, des techniques de classification de telles distributions sont possibles. Nous n'avons pas développé cette piste mais nous en entrapercavons déjà l'intérêt par exemple pour obtenir une partition ou une hiérarchie des malades opérés du cœur pouvant mener à la définition d'un score de souffrance cellulaire (paragraphe 5.4.2).

Outre cette question, de nombreuses autres questions ont émergé tout au long de ce parcours. Certaines ont trouvé des réponses mais bien d'autres, faute de temps ou de méthodologies de traitement, sont restées en suspens. Nous en citons les principales :

- Calcul de la mesure d'affinité entre deux densités de familles de lois de probabilité autres que celles traitées.
- Calcul de la mesure d'affinité entre deux densités appartenant à deux familles de lois différentes.
- Distribution asymptotique de la mesure d'affinité entre deux densités non gaussiennes.
- Calcul des régions critiques et puissances des tests d'égalité de deux densités, tests que l'on peut obtenir à partir de la distribution asymptotique de

la mesure d'affinité.

- Etude de la convergence des analyses factorielles de densités lorsque les densités sont estimées par la méthode des noyaux.
- Dans le cadre de l'analyse en composantes principales, les représentations sur le premier plan principal sont-elles suffisantes ou doit-on aller au-delà? Quelle explication donnée aux formes de nuages typiques d'un "effet Guttman"? Quelle pourrait être l'utilité du compromis obtenu par cette analyse? Les représentations obtenues pour des processus gaussiens seraient-elles aussi régulières pour d'autres types de processus?
- Pour les processus étudiés (mouvement brownien, Ornstein-Uhlenbeck,...), on n'a considéré que l'analyse en composantes principales d'un nombre fini de densités. Que se passe-t-il si on a un continuum de densités? Les équations intégrales que l'on obtiendrait admettent-elles des solutions? Lesquelles?
- Dans le cadre de l'analyse discriminante, quand a-t-on équivalence entre deux règles parmi les quatre règles d'affectation proposées?
- Pour des exemples de processus à évolution temporelle régulière quels seraient les pourcentages de bon classement? Seraient-ils meilleurs que dans l'exemple des châteaux? La hiérarchie obtenue entre règles d'affectation dans l'exemple des châteaux serait-elle identique?

Nous avons systématiquement traité l'ensemble des indices des densités de probabilités comme un ensemble d'instantants. Ceci est principalement dû au fait que notre travail a toujours été sous-tendu par l'application "datation des châteaux". Cependant rien ne s'oppose à l'extension des analyses proposées à des processus spatiaux, car comme nous l'avons indiqué en introduction le temps n'intervient que comme élément d'interprétation par la contiguïté de son caractère.

Enfin un dernier mot sur le concept d'analyses factorielles de distributions de probabilité que nous avons mis en forme avec B. Ycart. Nous avons le sentiment

qu'au terme de cette étape, nous n'avons pas encore entrevu toute la richesse et les limites de ce concept dont "le champ potentiel s'élargit sans cesse".
C'est là la conclusion principale que nous en tirons.

Bibliographie

- [1] Bar-Hen, A. and Daudin, J.J. (1998). Asymptotic distribution of Matusita's distance: Application to the location model. *Biometrika*, 85 (2), 477–481.
- [2] Barndorff-Nielsen, O. (1978). *Information and exponential families is statistical theory*. Wiley, New York.
- [3] Boumaza, R. (1998). Analyse en composantes principales de distributions gaussiennes multidimensionnelles. *Revue de statistique appliquée*, XLVI (2), 5–20.
- [4] Carlier, A. et Gueguen, A. (1994). *Etude de cas*. In : Celeux, G. et Nakache, J.P. *Analyse discriminante sur variables qualitatives*. Polytechnica, Paris, 227–253.
- [5] Celeux, G., Editeur (1991). *Analyse discriminante sur variables continues*. Collection didactique INRIA, Paris.
- [6] Celeux, G. et Nakache, J.P. (1994). *Analyse discriminante sur variables qualitatives*. Polytechnica, Paris, 1994.
- [7] Coppi, R. and Bolasco, S. (1989). *Multiway data analysis*. North-Holland, Amsterdam. Proceedings of the International Meeting on the Analysis of Multiway Data Matrices, Rome, March 28-30, 1988.
- [8] Dauxois, J. et Pousse, A. (1976). *Les analyses factorielles en calcul des probabilités et en statistique: essai d'étude synthétique*. Thèse d'état, Université Paul Sabatier, Toulouse, France.

- [9] Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function : some applications to statistical inference. *Journal of multivariate analysis*, 12, 136–154.
- [10] Diday, E., Lemaire, J., Pouget, J. et Testu, F. (1982). *Éléments d'analyse des données*. Dunod, Paris.
- [11] Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, 751–760.
- [12] Escoufier, Y. (1985). Objectifs et procédures de l'analyse conjointe de plusieurs tableaux de données. *Statistique et analyse de données*, 10, 1–10.
- [13] Fang, K.T. and Zhang, Y.T. (1990). *Generalized multivariate analysis*. Science Press, Beijing.
- [14] Friedman, J.H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84, 165–175.
- [15] Gard, T.C. (1988). *Introduction to stochastic differential equations*. Dekker, New York.
- [16] Girard, S. (1996). *Construction et apprentissage statistique de modèles auto-associatifs non-linéaires*. Thèse de doctorat, Université de Cergy-Pontoise, France.
- [17] Glaçon, F. (1981). *Analyse conjointe de plusieurs matrices de données*. Thèse de 3ème cycle, Université Joseph Fourier, Grenoble, France.
- [18] Gourieroux, C. et Monfort, A. (1990). *Séries temporelles et modèles dynamiques*. Economica, Paris.
- [19] Kailath, T. (1967). The divergence and Bhattacharyya distance measure in signal selection. *IEEE Transactions on Communication Technology*, COM-15(1), 52–60.

- [20] Kiers, H. A.L. (1991). Hierarchical relations among three-way methods. *Psychometrika*, 56(3), 449–470.
- [21] Kroonenberg, P.M. (1983). *Three-mode principal component analysis. Theory and applications*. DSWO Press, Leiden. Reprint 1989.
- [22] Lavit, C. (1988). *Analyse conjointe de tableaux quantitatifs*. Masson, Paris.
- [23] Le Breton, A. and Musiela, M. (1983). A look at a bilinear model for multidimensional stochastic systems in continuous time. *Statistics and Decisions*, 1, 285–303.
- [24] L'Hermier des Plantes, H. (1976). *Structuration des tableaux à trois indices de la statistique*. Thèse de 3ème cycle, Université Montpellier II, Montpellier, France.
- [25] Manly, B.F.J. (1986). *Multivariate statistical methods*. Chapman and Hall, London.
- [26] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1992). *Multivariate analysis*. Academic Press, London.
- [27] Matusita, K. (1967). Classification based on distance in multivariate Gaussian case. In *Proc. 5th Berkeley Symposium*, University of California Press, vol.1, 299–304.
- [28] McLachlan, G.J. (1992). *Discriminant analysis and statistical pattern recognition*. Wiley, New York.
- [29] Mkhadri, A., Celeux, G. and Nasroallah, A. (1997). Regularization in discriminant analysis: an overview. *Computational statistics and Data analysis*, 23, 403–423.
- [30] Qannari, E.M. (1983). *Analyses factorielles de mesures. Applications*. Thèse de 3ème cycle, Université Paul Sabatier, Toulouse, France.

- [31] Rao, C.R. (1973). *Linear statistical inference and its applications*. Wiley, New York.
- [32] Romain, Y. (1979). *Etude asymptotique des approximations par échantillonnage de l'analyse en composantes principales d'une fonction aléatoire. Quelques applications*. Thèse de 3ème cycle, Université Paul Sabatier, Toulouse, France.
- [33] Romain, Y. (1997). Une introduction à l'approche fonctionnelle stochastique de la statistique multidimensionnelle. *Publications du laboratoire de statistique et probabilités, Université Paul Sabatier, Toulouse*, 9-97.
- [34] Rudrauf, J.M. (1987). Petit-Geroldseck. Mise au point sur son origine. *Etudes Médiévales, Centre de recherches archéologiques médiévales de Saverne, Saverne*, 89–120.
- [35] Saporta, G. (1990). *Probabilités, analyse des données et statistique*. Technip, Paris.
- [36] Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, London.
- [37] Tomassone, R., Danzart, M., Daudin, J.J. et Masson, J.P. (1988). *Discrimination et classement*. Masson, Paris.
- [38] Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31, 279–311.
- [39] Vinograd, I.M., Editor (1987). *Encyclopaedia of mathematics*, Kluwer academic publishers, Dordrecht, vol.7, 163–164.
- [40] Volle, M. (1981). *Analyse des données*. Economica, Paris.
- [41] Ycart, B. (1989). Markov processes and exponential families on a finite set. *Statistics and Probability Letters*, 8, 371–376.

- [42] Ycart, B. (1992a). Integer valued markov processes and exponential families. *Statistics and probability letters*, 14, 71–78.
- [43] Ycart, B. (1992b). Markov processes and exponential families. *Stochastic processes and their applications*, 41, 203–214.