

Pattern recognition/reconnaissance de formes

V. Vigneron
vvigne@iup.univ-evry.fr
Univ. Evry Val d'Essonne

Quelques exemples d'entropie différentielle (cas continu)

a) X a une distribution uniforme $[0, X_{\max}]$

$$H(X) = - \int_0^{X_{\max}} \frac{1}{X_{\max}} \cdot \log_2 \left(\frac{1}{X_{\max}} \right) dx = \log_2 X_{\max}$$

b) X est a une distribution gaussienne $N(0, G^2)$

$$H(X) = \int_{-\infty}^{\infty} \varphi(x) \left[\frac{\log_2(2\pi G^2)}{2} + \frac{x^2 \log_2 e}{2G^2} \right] dx$$

$$H(X) = \frac{\log_2(2\pi e G^2)}{2} \quad / \quad \text{où } \varphi(x) \text{ est la densité de probabilité gaussienne (ou normale)}$$

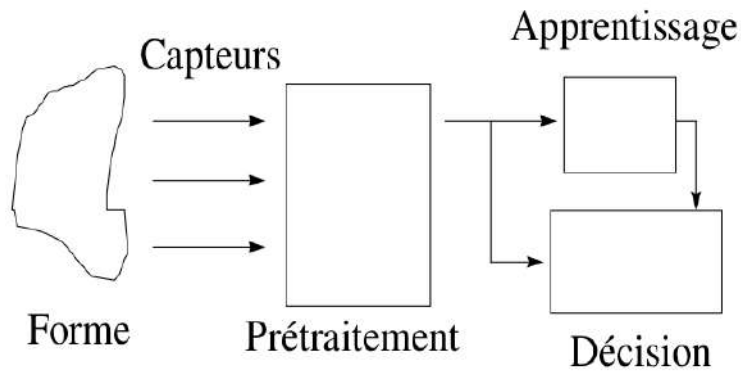
c) distribution gaussienne mais \underline{X} est une multi variable à n dimensions

$$H(\underline{X}) = \int_{-\infty}^{\infty} \varphi(\underline{x}) \left[\frac{\log_2((2\pi)^n \det C)}{2} + \frac{(\underline{x}^T C^{-1} \underline{x}) \log_2 e}{2} \right] d\underline{x}$$

Jan. 25, 2013

ETUDES DE CAS

1. diagnostic précoce de la maladie d'Alzheimer
2. Tri cellulaire pour le diagnostic cancérologique du col de l'utérus
3. Contrôle neuronal pour la suppléance fonctionnelle (neural-shunt)
4. prédiction du geste chirurgical optimal pour la correction postural d'enfants handicapés cérébraux
5. détection de défauts dans des wafers (SIEMENS/INFINEON)
6. robotique mobile COOPERCOM / carte de perception étendue à partir de communications intervéhiculaires



1. Traitement d'images PET SCAN (Univ. Sherbrooke, Quebec)

Démences gériatriques touchent ≈ 24 millions de pers., la maladie d'Alzheimer = la plus courante. Ce nombre devrait tripler dans les 50 ans à venir.

Objectif : Aide à la détection précoce de la maladie d'Alzheimer à partir de l'analyse de FDG PET Scan qui correspondent à des images du métabolisme du cerveau (imagerie fonctionnelle).

Altérations liées à la maladie d'Alzheimer visibles dans les PET scans

avant les IRM et avant que les symptômes soient identifiables.

3 classes définies :

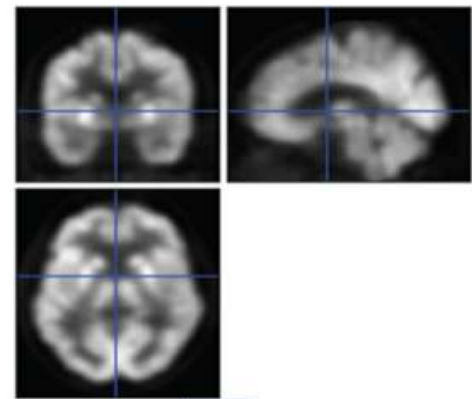
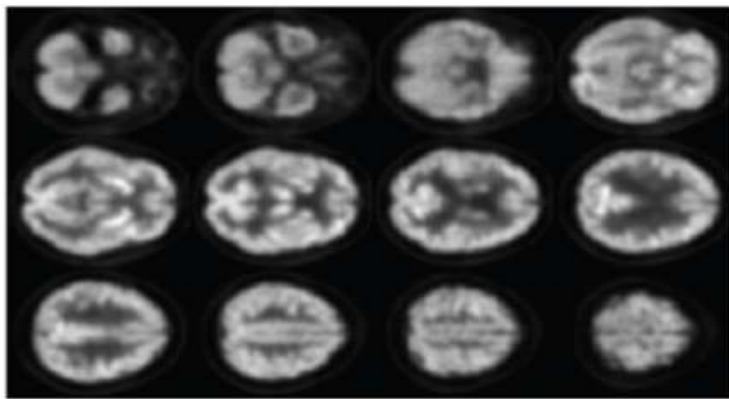


Figure: Normal aged brain in PET scan - **NC**

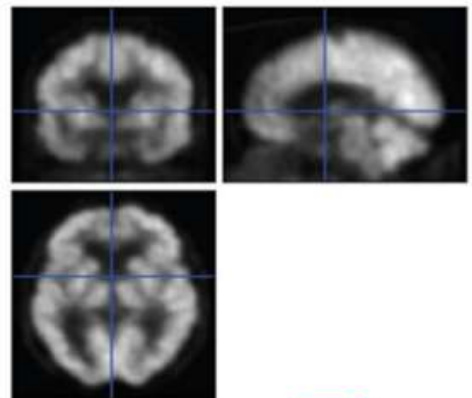
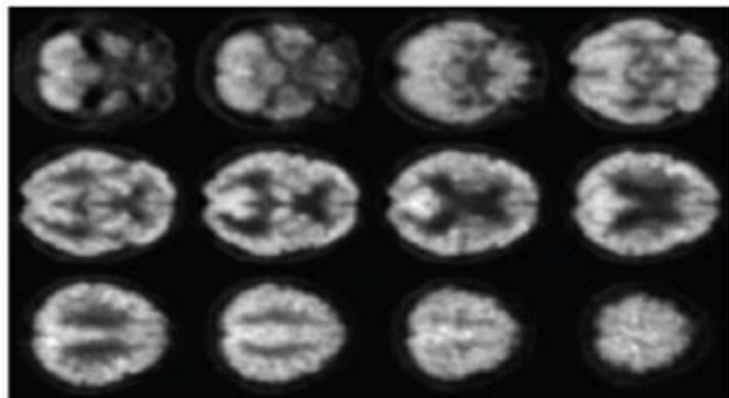
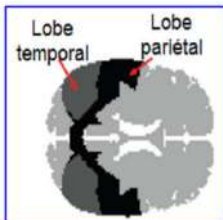


Figure: Alzheimer's disease affected brain in PET scan - **AD**



	NC	MCI	dMCI	AD	Σ
patients baseline	84	124	25	82	335

2. Tri cellulaire (Novacyt)

Cytologie= étude des caractéristiques morphologiques de cellules isolées
Cellules issues du col de l'utérus



Détection du cancer du col de l'utérus



PROBLÈME : actuellement, détection à "l'œil nu"



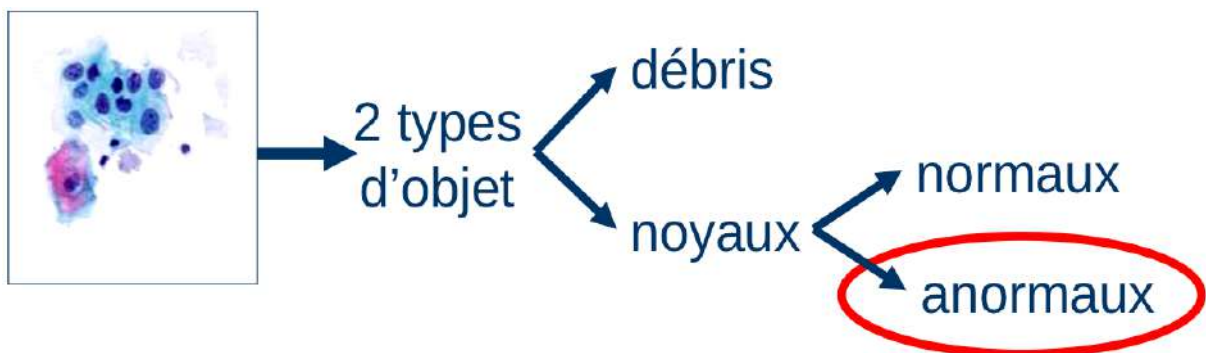
RISQUE DE FAUX DIAGNOSTIC !

D'où outil qui aide à détecter les cellules anormales

Cet outil va s'inscrire dans tout un processus industriel intégré depuis le traitement de la lame jusqu'au diagnostic des cellules.

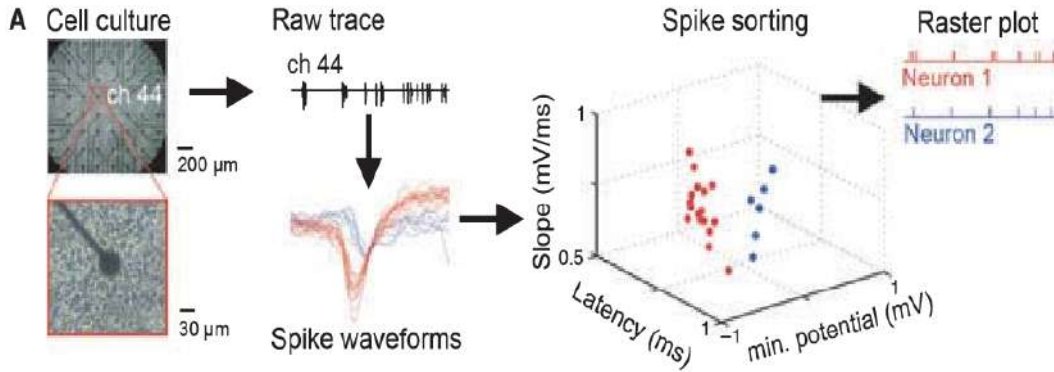


Phase de classification



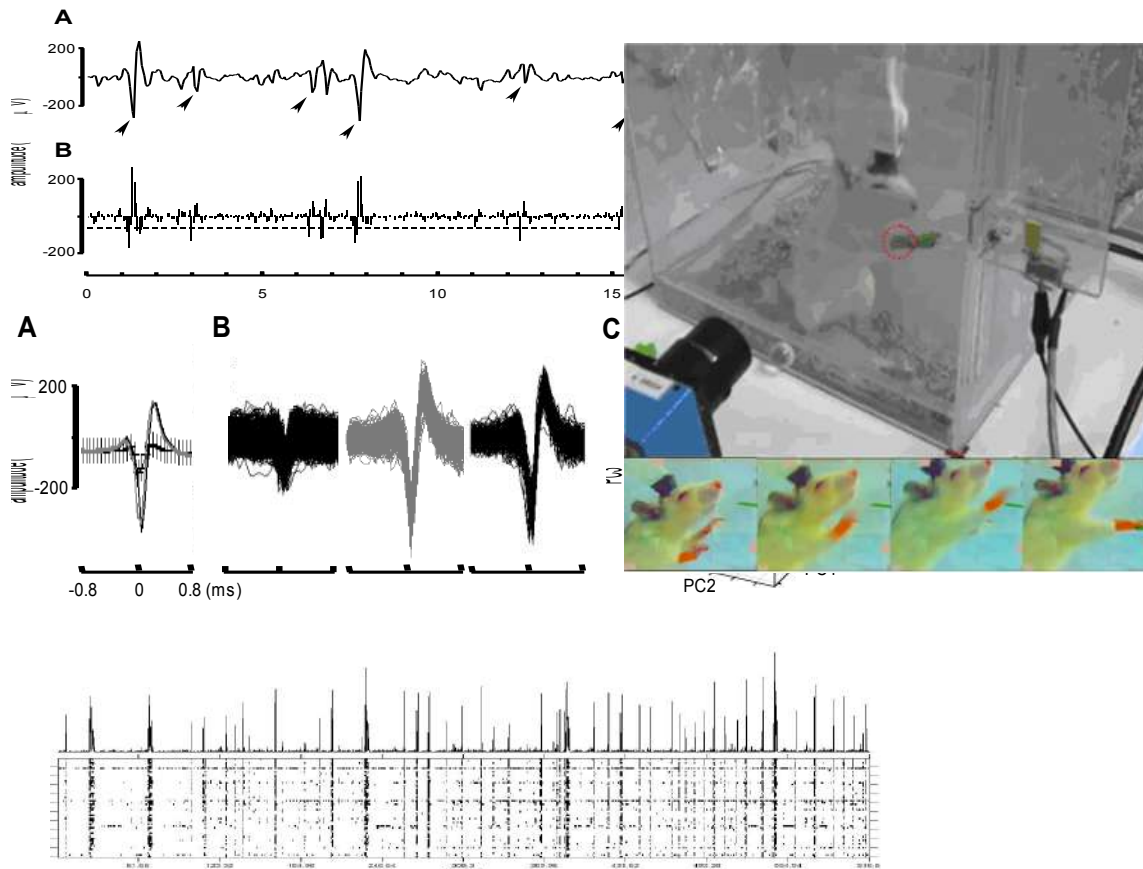
3. Neural Shunt (NTHU/Taiwan, Physio Lab/Mainz, Allemagne)

But : évaluer la faisabilité d'une commande moto-corticale pour la réparation fonctionnelle

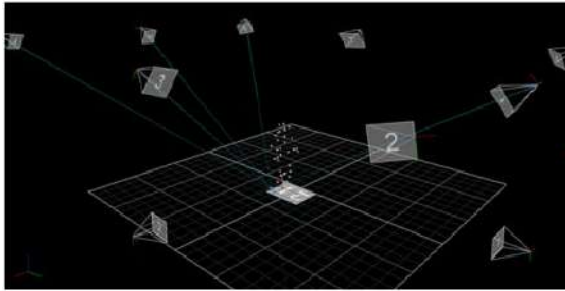


Champ disciplinaire : classification, filtrage de Kalman

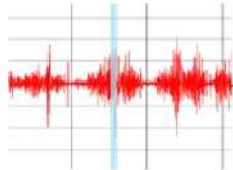
Moyens : expérimentation animale et interface Cerveau-Machine (stimulation de l'aire somato-sensorielle → cortex moteur)



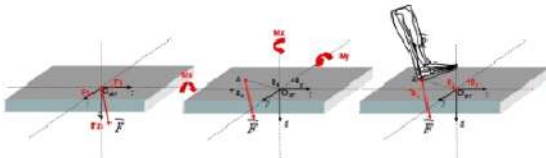
4. Correction de posture (Institut POIDATZ)



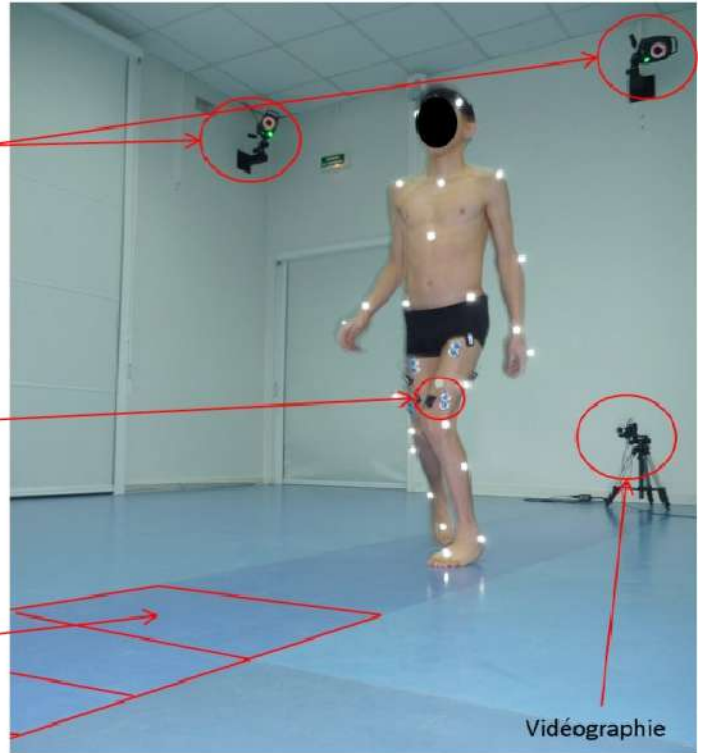
Capture des coordonnées 3D des marqueurs



Enregistrement de l'activité électrique musculaire

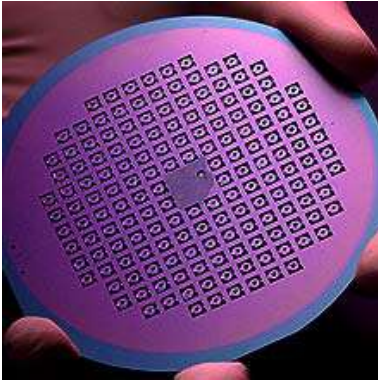


Mesure des efforts d'interaction pied-sol

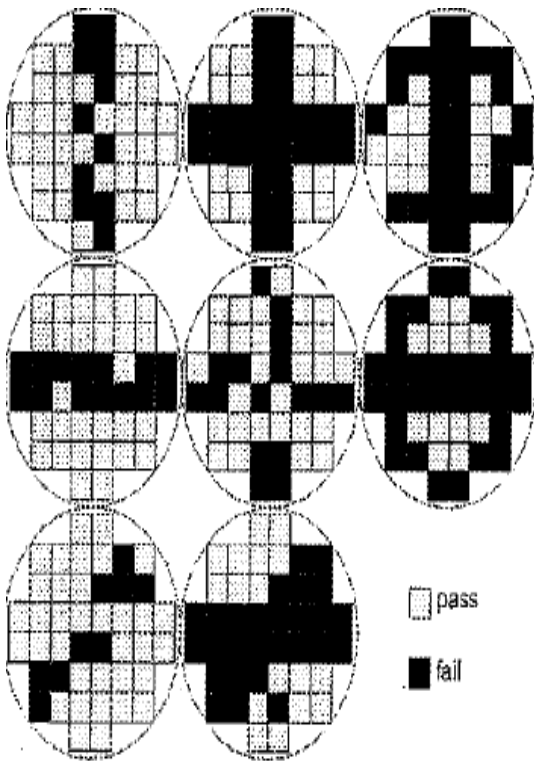


Vidéographie

5. détection de défauts dans des wafers (U. Regebsburg)



Binary Wafer Test data. N wafers, M processeurs



6. COOPERCOM (Avec IFFSTAR/LIVIC)

But : Fusionner les capacités perceptives et communicantes d'un cluster de véhicules en mouvement pour la construction d'une carte de perception étendue.

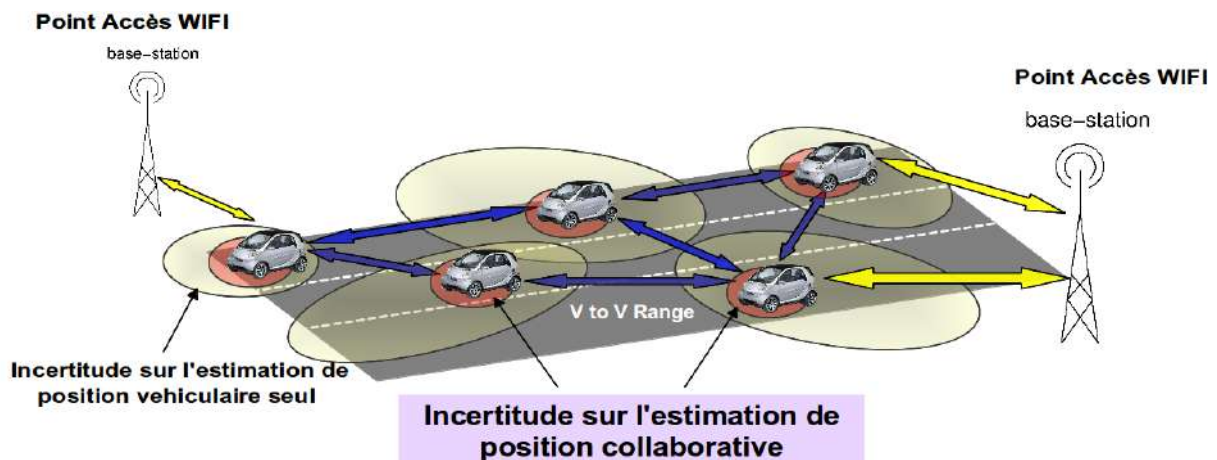
Champ disciplinaire : extraction de caractéristiques, réduction de dimensions.

Communications inter-véhiculaires vues comme un **filtrage fédératif**

- cluster de véhicules en mouvement
- réseau de capteurs (caméras, centrale inertielle, GPS)
- architecture décentralisée

Avantages :

- ↗ précision et fiabilité de la position
- ↗ étendue de la carte perceptive
- Évitement des collisions, gestion des changements de voies = système d'assistance à la conduite
- Robustesse aux pannes et incertitudes
- ↗ rapidité de localisation par rapport à un seul véhicule



Ce cours sera illustré en utilisant le logiciel libre **Matlab** : <http://www.scilab.org/>. Matlab est un bon choix pour le traitement des données neurophysiologiques : il est libre (sous Unix, Linux, MSDOS, ...), il contient le code source (en Fortran, C et en sa propre langue), il est puissant et très bien documenté, et il est raisonnablement facile à employer.

Les structures d'algorithmes et de données constituent des programmes. Dans Matlab, la matrice est probablement la structure principale. Un vecteur est une matrice de dimension $1 \times n$ ou $n \times 1$. Une valeur réelle ou complexe peut être considérée comme une matrice 1×1 .

Un **vecteur** est une **matrice** de dimension $1 \times n$ ou $n \times 1$. Une valeur réelle ou complexe peut être considérée comme une matrice 1×1 .

Matlab a beaucoup de fonctions qui permettent le traitement des matrices comme une entité. Dans la plupart des cas il est inutile d'accéder aux valeurs qui les composent.

Exemple: *Soit A une matrice ; $\max(A)$ - $\min(A)$ produit la gamme (maximum - minimum) des valeurs intégrées dans la matrice A.*

Exemple:

1. *Création d'une matrice T rangées et M colonnes composées par des nombres aléatoires entre 0 et 1*

`mat = rand(T,M)` ; Si nous voulons savoir plus au sujet de la fonction « `rand` » (ou autre fonction) nous pourrions écrire

`help(rand)`

et nous obtiendrions une liste de fonctions disponibles contenant `random` et sa définition.

Nombre de rangées `l` et colonnes `c` :

- `[l,c]=size(A)`
2. Extraction d'une composant d'une matrice `mat`
`mat (l,c); // l=numéro de la ligne, c=numéro de la colonne`
 3. Longueur d'un vector
`length (vec);`
 4. Extraction de plusieurs composants d'une matrice sur la ligne `l` ou d'un vecteur
`mat (l,c1:c2);`
 5. Extraction de plusieurs composants d'un vecteur `vec`
`length (vec);`
 6. Éliminer certains composants: `mat (l,c1:c2) = [];`
 7. Insérer certains composants:
 8. Vecteurs ou matrices peuvent contenir des variable booléennes ou des chaîne de caractères
`vec = [1 2.3 4 0 -1];vec > 0 ; vec = [1 1 1 0 0];`
 9. Somme des composantes d'un vecteur
`sum (vec);`
 10. Produit des composantes d'un vecteur
`prod (vec);`
 11. Moyenne
`mean (vec);`
 12. Médiane
`median (vec);`
 13. Écart-type
`std(vec);`
 14. Combiner plusieurs commandes
`sum (abs(vec));`

15. Minimum ou Maximum d'un vecteur

```
[minvec, ind] = min (vec); //minvec : valeur minimale, ind : son indice dans vec
```

```
[maxvec, ind] = max (vec); //maxvec : valeur maximale, ind : son indice dans vec
```

16. tri par ordre décroissant

```
[ind, d] = sort (v);
```

17. Tracer un vecteur ou une partie d'un vecteur

```
k=[0:0.02:5];
```

```
signal = sin(2 * pi * 9.7 * k) + sin(2 * pi * 10.3 * k);
```

```
plot2d(k,signal);
```

```
xtitle('sin(2*pi*9.7*k)+ sin(2 * pi * 10.3 * k)', 'k');
```

Identification d'un système

= déterminer le modèle mathématique d'un système inconnu en observant i/o.

Buts :

- prédiction d'un comportement
- explique interactions/rerelations entre i/o.
- concevoir un contrôleur d'avion, de bateau,...

Principe (processus itératif) :

- identifier une classe de modèles (paramétriques)

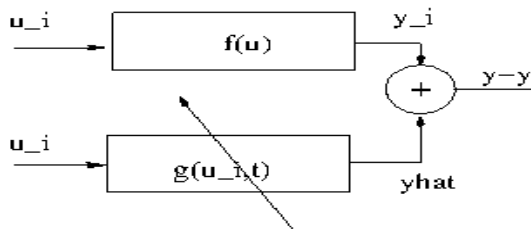
$$y = f(u, \theta)$$

où y est la sortie, u l'entrée et θ un vecteur de paramètres.

- déterminer les paramètres à partir de techniques d'optimisation
- test de validation /données non-observées

sans connaissance a priori → c'est un problème difficile.

Données d'apprentissage (échantillons) : $(u_i, y_i)_{i=1}^m$.



BASES DU CALCUL MATRICIEL

Notations :

- matrices A, B, X
- vecteurs $\vec{a}, \vec{b}, \vec{x}$ ou plus simplement $\mathbf{a}, \mathbf{b}, \mathbf{x}$
- scalaires a, b, x .

Lemmes :

$$(AB)^T = B^T A^T \quad (1)$$

$$(AB)^{-1} = B^{-1} A^{-1} \quad (2)$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, \quad \text{avec } A_1 = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}, A_2 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, A_3 = [7 \ 8], A_4 = 9$$

$$A = [\mathbf{a}_1 \quad \dots \quad \mathbf{a}_n] = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix} \quad (3)$$

si $\mathbf{x} = [x_1 \quad \dots \quad x_n]^T, f(\mathbf{x})$ une fonction scalaire, alors

où \mathbf{a}_i est un vecteur colonne.

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \quad (4)$$

$$\text{si } \mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \quad \dots \quad f_m(\mathbf{x})] \quad (5)$$

$$J_f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_n} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \frac{\partial \mathbf{f}(\mathbf{x})}{\mathbf{x}^T} = [\nabla f_1(x) \quad \dots \quad \nabla f_m(x)] \quad (6)$$

Le hessien d'une fonction scalaire=jacobien de la fonction gradient

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \vdots & \frac{\partial^2 f}{\partial x_i x_j} & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} = \begin{bmatrix} \nabla^T \frac{\partial f}{\partial x_1} \\ \vdots \\ \nabla^T \frac{\partial f}{\partial x_n} \end{bmatrix} \quad (7)$$

- Gradient d'une fonction linéaire :

$$\mathbf{c} = [c_1 \ \dots \ c_n]^T, \mathbf{x} = [x_1 \ \dots \ x_n]^T, \quad f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} = \mathbf{x}^T \mathbf{c} \Rightarrow \nabla f(\mathbf{x}) = \mathbf{c}$$

- Forme quadratique :

$$A = (a_{ij})_{n \times n}, \mathbf{x} = [x_1 \ \dots \ x_n]^T \quad (8)$$

$$\mathbf{x}^T A \mathbf{x} = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \mathbf{x}^T \left(\frac{A + A^T}{2} \right) \mathbf{x} \quad (9)$$

$$\nabla f(\mathbf{x}) = \begin{cases} (A + A^T) \mathbf{x}, & \text{si } A \text{ n'est pas symétrique,} \\ 2A \mathbf{x}, & \text{si } A \text{ est symétrique,} \end{cases} \quad (10)$$

$$H_f(\mathbf{x}) = \begin{cases} (A + A^T), & \text{si } A \text{ n'est pas symétrique,} \\ 2A, & \text{si } A \text{ est symétrique,} \end{cases} \quad (11)$$

- Matrice définie positive (négative): $A > 0$ ($A < 0$) si $\mathbf{x}^T A \mathbf{x} > 0$ (< 0), $\forall \mathbf{x} \neq \mathbf{0}$ ou si toutes les valeurs propres de A sont > 0 (resp. < 0).

- optimum d'une fonction quadratique : $f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$, A symétrique.

$$\text{si } \begin{cases} A > 0 & \min f(\mathbf{x}) \\ A < 0, & \max f(\mathbf{x}) \end{cases} \text{ est en } \mathbf{x}^* = -A^{-1} \mathbf{b}$$

- Si \mathbf{x} et \mathbf{y} sont 2 vecteurs

$$\nabla(\mathbf{x}^T \mathbf{y}) = \nabla(\mathbf{y}^T \mathbf{x}) = \mathbf{y} \quad (12)$$

$$\nabla(\mathbf{x}^T \mathbf{x}) = 2\mathbf{x} \quad (13)$$

$$\nabla_{\mathbf{x}^T} A \mathbf{y} = A \mathbf{y} \quad (14)$$

$$\nabla_{\mathbf{y}^T} A \mathbf{x} = A^T \mathbf{x} \quad (15)$$

$$\nabla_{\mathbf{x}^T} A \mathbf{x} = (A + A^T) \mathbf{x} \quad (16)$$

$$\nabla \mathbf{f}^T(\mathbf{x}) \mathbf{g}(\mathbf{x}) = J_{fg}^T + J_{gf}^T \quad (17)$$

- Formule d'inversion de matrice : si A et $(I + CA^{-1}B)$ sont non singulières et carrées : $(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}$

Exercice 1 (Preuve ?)

- développement en série de Taylor d'une fonction scalaire $f(\cdot)$ de \mathbf{x} en fonction de $\mathbf{d} = [d_1 \ \dots \ d_n]^T$

$$f(\mathbf{x} + \mathbf{d}) = f(\mathbf{x}) + \sum_i^n \frac{\partial f(\mathbf{x})}{\partial x_i} d_i + \frac{1}{2} \sum_i^n \sum_j^n \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} d_i d_j + \mathcal{O}(n) \quad (18)$$

$$\approx f(\mathbf{x}) + \mathbf{g}^T \mathbf{d} + \frac{1}{2} \mathbf{d}^T H \mathbf{d} \quad (19)$$

Estimation de paramètres

- méthode du maximum de vraisemblance,
- méthode des moments.
- méthode des moindres carrés.

1. **Idée de base pour le maximum de vraisemblance** : si un événement a lieu, sa probabilité d'occurrence (sa vraisemblance) est grande.

Technique d'estimation des paramètres d'une distribution.

Soit $x \sim \mathcal{L}$ de densité de probabilité $f(x; \theta)$. Si je dispose de n observations x_1, \dots, x_n , la fonction de vraisemblance L est :

$$L = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta).$$

Sans information *a priori* sur la vraie valeur de θ , on choisira θ qui maximise la probabilité d'observer x_1, \dots, x_n . C'est l'estimateur de *maximum de vraisemblance*, *i.e.* :

$$\left. \frac{\partial L}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0 \quad \text{ou} \quad \left. \frac{\partial \ln L}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0. \quad (20)$$

Exemple 1 (Estimation du maximum de vraisemblance d'une panne) La durée X de fonctionnement d'un système avant panne est dist. exponentiellement selon $f(x) = \lambda e^{-\lambda x} u(x)$. λ la fréquence des pannes est inconnue. Pour estimer λ , si X_1, \dots, X_n sont indépendantes,

$$L(x_1, \dots, x_n; \lambda) = f_{X_1, \dots, X_n}(x_1, \dots, x_n | \lambda), \quad (21)$$

$$= \prod_{i=1}^n f_{X_i}(x_i) = \lambda^n e^{-\lambda x_1} \dots e^{-\lambda x_n} u(x_1) \dots u(x_n) \quad (22)$$

$$= \lambda^n e^{-\lambda(x_1 + \dots + x_n)} u(x_1) \dots u(x_n) = \lambda^n e^{-\lambda n \bar{x}} \prod_{i=1}^n u(x_i) \quad (23)$$

$$\ln L = n(\ln \lambda - \bar{x} \lambda) + \sum_{i=1}^n \ln u(x_i) \quad (24)$$

avec $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. L et $\ln L$ ont le même max. pour λ t.q. $\frac{d}{d\lambda} \ln L = 0$, soit $\hat{\lambda} = \frac{1}{\bar{x}}$.

2. **Idée de base pour la méthode des moments** : le moment d'un échantillon est proche du vrai moment (premier ou deuxième).

Exemple 2 (Maximum de vraisemblance et méthode des moments) Soit (X_1, \dots, X_n) un échantillon indépendant de DP $f(X) = (\theta + 1)x^\theta$, $0 < x < 1, \theta > -1$, θ un paramètre inconnu à estimer. Fonction de vraisemblance :

$$L(x_1, \dots, x_n; \lambda) = (\theta + 1)x_1^\theta \dots (\theta + 1)x_n^\theta = (\theta + 1)^n (x_1 \dots x_n)^\theta \quad (25)$$

$$\ln L = n \ln(\theta + 1) + \theta(\ln x_1 + \dots + \ln x_n) \quad (26)$$

$$\frac{d}{d\theta} \ln L = 0 \quad \Rightarrow \frac{n}{\theta + 1} + \sum_i^n \ln x_i = 0 \quad (27)$$

Estimateur du maximum de vraisemblance :

$$\hat{\theta} = - \left[\frac{n}{\sum_i^n \ln x_i} + 1 \right]. \quad (28)$$

$$\text{Notons que : } \bar{x} = \int_0^1 x(\theta + 1)x^\theta = \frac{\theta + 1}{\theta + 2} x^{\theta+2} \Big|_0^1 = \frac{\theta + 1}{\theta + 2}. \quad (29)$$

$\bar{x} = \hat{x}$. Donc, l'estimateur par la méthode des moments $\hat{\theta} = \frac{2\bar{X}-1}{1-\bar{X}}$ est différent de l'estimateur du maximum de vraisemblance.

Exercice 2 (MV des paramètres d'une distribution uniforme) (X_1, \dots, X_n) un échantillon indépendant de DP Uniforme $f(X) = \begin{cases} \frac{1}{(\theta_2 - \theta_1)} & \theta_1 < x \leq \theta_2 \\ 0 & \text{ailleurs.} \end{cases}$, où $\theta_2 > \theta_1$ sont inconnus. Trouver les par une méthode MV et des moments.

Réponse :

Les moyenne et variance d'une dist. uniforme sont $E[X] = \frac{\theta_1 + \theta_2}{2}$ et $\text{var}[X] = \frac{(\theta_2 - \theta_1)^2}{12}$. Posons $E[X] = \hat{X}$ et $\text{var}[X] = S^2$, où \hat{X} est la moyenne empirique et $S^2 = \frac{1}{n} \sum_i^n (X_i - \hat{X})^2$ est la variance empirique (biaisée). En résolvant ces 2 équations, on trouve l'estimateur de la méthode des moments :

$$\theta_1 = \hat{X} - \sqrt{3}S \quad (30)$$

$$\theta_2 = \hat{X} + \sqrt{3}S \quad (31)$$

La fonction de vraisemblance est donné par :

$$L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n f_{X_i}(x_i) = \begin{cases} \frac{1}{(\theta_2 - \theta_1)^n} & \theta_1 \leq x_i \leq \theta_2, i = 1, \dots, n \\ 0 & \text{ailleurs.} \end{cases} \quad (32)$$

Les équations de vraisemblance :

$$\frac{d \ln L}{d\theta_1} = \frac{n}{\theta_2 - \theta_1} = 0 \quad (33)$$

$$\frac{d \ln L}{d\theta_2} = \frac{n}{\theta_2 - \theta_1} = 0 \quad (34)$$

ne permettent pas de trouver les estimateurs MV. Mais, on sait que $X_i \sim \mathcal{U}_{[\theta_1, \theta_2]}$, et θ_1 plus petit que n'importe quel X_i (et symétriquement pour θ_2). donc les estimateur du maximum de vraisemblance sont $\hat{\theta}_1 = \min_{1 \leq i \leq n} X_i$ et $\hat{\theta}_2 = \max_{1 \leq i \leq n} X_i$.

3. Estimateur des Moindres Carrés

Soit

$$\boxed{y = \theta_1 f_1(\mathbf{u}) + \dots + \theta_n f_n(\mathbf{u})} \quad (35)$$

avec $\mathbf{u} = [u_1 \dots u_n]^T$ and f_1, \dots, f_n n fonctions connues.

(35) est une **équation de régression**, θ_i les coefficients de régression.

Si on dispose de données d'apprentissage $\{(u_i, y_i), i = 1, \dots, m\}$, on doit résoudre le système :

$$= \begin{cases} f_1(u_1)\theta_1 + \dots + f_n(u_1)\theta_n = y_1 \\ \vdots \\ f_1(u_m)\theta_1 + \dots + f_n(u_m)\theta_n = y_m \end{cases} \quad (36)$$

$$\Leftrightarrow \boxed{A\theta = \mathbf{y}}, \quad (37)$$

$$\text{avec } A = \begin{bmatrix} f_1(u_1) & \dots & f_n(u_1) \\ \vdots & & \vdots \\ f_1(u_m) & \dots & f_n(u_m) \end{bmatrix}, \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \quad (38)$$

Adapté à l'estimation de fonctions paramétriques non-linéaires ?

Pour identifier de façon unique θ , il faut que $m \geq n$.

$$\begin{cases} \text{si } m = n \text{ et } A \text{ carrée non-singulière} & \Rightarrow \theta = A^{-1}\mathbf{y} \\ \text{si } m > n & \Rightarrow A\theta = \mathbf{y} \rightarrow A\theta + \mathbf{e} = \mathbf{y} \end{cases}$$

où \mathbf{e} est le vecteur d'erreur de modélisation.

⇓

On cherche $\theta = \hat{\theta}$ qui minimise

$$C(\theta) = \sum_i^n (y_i - \mathbf{a}_i^T \theta)^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - A\theta)^T (\mathbf{y} - A\theta) \quad (39)$$

$$= \mathbf{y}^T \mathbf{y} + \theta A^T A \theta - 2\mathbf{y} A^T \theta \quad (40)$$

en notant que $\theta^T A^T \mathbf{y} = \mathbf{y}^T A \theta$

$$\nabla C(\theta) = 2A^T A \theta - 2A^T \mathbf{y} = 0 \quad (41)$$

$$\theta = \boxed{(A^T A)^{-1} A^T \mathbf{y}} \quad (42)$$

- Si $A^T A$ est singulière \Rightarrow le LSE (Least Square Estimator) n'est pas unique
- Si chaque terme d'erreur à un poids particulier, alors on peut modifier la fonction coût $C_W(\theta) = (\mathbf{y} - A\theta)^T W (\mathbf{y} - A\theta)$ et

$$\boxed{\hat{\theta}_W = (A^T W A)^{-1} A^T W \mathbf{y}.}$$

Exemple 3 (loi de Hook)

$$\ell = k_0 + k_1 f \rightarrow \begin{bmatrix} 1 & 1,1 \\ 1 & 1,9 \\ 1 & 3,2 \\ 1 & 3,3 \\ 1 & 4,4 \\ 1 & 5,9 \end{bmatrix} \begin{bmatrix} k_0 \\ k_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} = \begin{bmatrix} 1,5 \\ 2,1 \\ 2,5 \\ 3,3 \\ 4,1 \end{bmatrix} \quad (43)$$

$$\begin{bmatrix} \hat{k}_0 \\ \hat{k}_1 \end{bmatrix} = (A^T A)^{-1} A^T \ell = \begin{bmatrix} 1,2 \\ 0,44 \end{bmatrix} \quad (44)$$

$\Rightarrow \ell = k_0 + k_1 f + k_2 f^2 + \dots + k_n f^n$ (on peut toujours prendre un modèle plus complexe...)

Qualité d'un estimateur θ (le meilleur ?)

- mesure de qualité =

erreur des moindres carrés $MSE(\hat{\theta}) = E[(\theta - \hat{\theta})^2]$ (on pourrait utiliser $E[|\theta - \hat{\theta}|]$ mais pas $E[(\theta - \hat{\theta})]$)

ou bien la **variance de l'estimateur** $MSE(\hat{\theta}) = E[(\theta - E[\hat{\theta}])^2]$

- le **biais** : erreur d'estimation est nulle $E[\theta - \hat{\theta}] = 0$.
- l'**efficacité** : sa MSE est la plus petite parmi tous les estimateurs possibles du même paramètre.
Borne de Cramer-Rao : $MSE(\hat{\theta}) \geq I^{-1}$, où $I = -E[\frac{\partial^2}{\partial \theta^2} \ln L(x_1, \dots, x_n; \lambda)]$.
- la **consistance** : l'erreur d'estimation $\rightarrow 0$ quand $n \rightarrow \infty$.

Exercice 3 (X_1, X_2) un échantillon d'une population $X \sim \mathcal{N}(\mu, \sigma^2)$, μ inconnu mais σ^2 connu. Étant donné trois estimateurs $\hat{\mu}_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2$, $\hat{\mu}_2 = \frac{2}{5}X_1 + \frac{3}{5}X_2$, $\hat{\mu}_3 = \frac{1}{3}X_1 + \frac{2}{3}X_2$:

1. Prouver que $\hat{\mu}_1$, $\hat{\mu}_2$ et $\hat{\mu}_3$ sont non-biaisés.
2. Trouver celui avec la plus petite variance.

Estimateur des Moindres Carrés Multivariés

$$\text{si } \mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_q]^T, q > 1 \quad (45)$$

$$A\Theta + E = Y, A_{(n,m)} = \begin{bmatrix} f_1(u_1) & \dots & f_n(u_1) \\ \vdots & & \vdots \\ f_1(u_m) & \dots & f_n(u_m) \end{bmatrix} \quad (46)$$

$$\Theta_{(n,m)} = \begin{bmatrix} \theta_{11} & \dots & \theta_{1q} \\ \vdots & & \vdots \\ \theta_{n1} & \dots & \theta_{nq} \end{bmatrix}, \quad Y_{(m,q)} = \begin{bmatrix} y_{11} & \dots & y_{1q} \\ \vdots & & \vdots \\ y_{m1} & \dots & y_{mq} \end{bmatrix} \quad (47)$$

On doit minimiser

$$E[\Theta] = \sum_{i=1}^m \sum_{j=1}^q e_{ij}^2 = \sum_{j=1}^q \left(\sum_{i=1}^m e_{ij}^2 \right) \quad (48)$$

et

$$\min_{\Theta} \sum_{j=1}^q \left(\sum_{i=1}^m e_{ij}^2 \right) = \sum_{j=1}^q \left(\min_{\theta_j} \sum_{i=1}^m e_{ij}^2 \right) \leftrightarrow A\theta_j + \mathbf{e}_j = \mathbf{y}_j \quad (49)$$

$$\hat{\theta}_j = (A^T A)^{-1} A^T \mathbf{y}_j \quad (50)$$

$$\boxed{\hat{\Theta} = (A^T A)^{-1} A^T Y} \quad (51)$$

Interprétation géométrique

Si $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$, alors $A\theta = [\mathbf{a}_1, \dots, \mathbf{a}_n] \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$. Pour que $A\theta$ approx-

ime \mathbf{y} au sens des moindres carrés, $A\theta$ doit être égal à la projection de \mathbf{y} sur le sous-espace $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$.

Si $A = [\mathbf{a}_1, \mathbf{a}_n]$, $A\theta = \mathbf{a}_1\theta_1 + \mathbf{a}_n\theta_n \rightarrow A\theta \in \text{plan } \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$.

Pour que $\mathbf{e} = \mathbf{y} - A\theta$ soit minimum, il faut que $A\theta$ soit égal à la projection de \mathbf{y} sur $[\mathbf{a}_1, \mathbf{a}_n] \Leftrightarrow \mathbf{e} \perp [\mathbf{a}_1, \mathbf{a}_n]$.

$$\Leftrightarrow \mathbf{a}_1^T (\mathbf{y} - A\hat{\theta}) = 0 \quad (52)$$

$$\Leftrightarrow \mathbf{a}_2^T (\mathbf{y} - A\hat{\theta}) = 0 \quad (53)$$

$$\Leftrightarrow A^T (\mathbf{y} - A\hat{\theta}) = 0 \quad (54)$$

ou

$$\text{proj}_A(\mathbf{y}) = A\hat{\theta} = A(A^T A)^{-1} A^T \mathbf{y} \quad (55)$$

où $H = A(A^T A)^{-1} A^T$ est l'**opérateur de projection**.

$M = I - A(A^T A)^{-1} A^T$ est l'**opérateur orthogonal**.

↓

$$\hat{\mathbf{y}} = H\mathbf{y} \text{ et } \hat{\mathbf{e}} = M\mathbf{y}$$

Propriétés des estimateurs des moindres carrés

- l'approximation optimale de \mathbf{y} ($\hat{\mathbf{y}} = A\hat{\theta}$) est \perp à $\hat{\mathbf{e}}$: $\hat{\mathbf{y}}\hat{\mathbf{e}} = 0$.

- Remarquez que :

$$M\mathbf{e} = (I - A(A^T A)^{-1} A^T)(\mathbf{y} - A\hat{\theta}) \quad (56)$$

$$= \mathbf{y} - A\hat{\theta} - A(A^T A)^{-1} A^T (\mathbf{y} - A\hat{\theta}) \quad (57)$$

$$= \mathbf{y} - A(A^T A)^{-1} A^T \mathbf{y} = \mathbf{y} - H\mathbf{y} = M\mathbf{y} \quad (58)$$

$$(59)$$

- $HA = A$ ou $MA = 0$

- H est *idempotente* $\Leftrightarrow \forall k > 1, H^k = H$

\Downarrow

$$\forall \mathbf{y}, \underbrace{HHH \dots H}_{k \text{ fois}} \mathbf{y} = H\mathbf{y}.$$

Estimateurs des moindres carrés récursifs

Supposons que k désigne le nombre de lignes de A et de \mathbf{y} , alors

$$\theta_k = (A^T A)^{-1} A^T \mathbf{y} \quad (60)$$

et supposons que les données arrivent dans un ordre séquentiel.

J'observe une nouvelle paire (\mathbf{a}^T, y) , au lieu de recalculer l'estimateur des MC θ_{k+1} à partir des $k + 1$ données disponibles, on a intérêt à utiliser l'estimateur θ_k déjà obtenu pour mettre à jour θ_{k+1} .

Peut-on Trouver la formule générale de mise à jour des θ_k ?

$$\theta_k = \left(\begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} \begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{y} \\ y \end{bmatrix} \quad (61)$$

$$\text{Posons } P_k = (A^T A)^{-1} \quad (62)$$

$$P_{k+1} = \left(\begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} = \left([A^T \mathbf{a}] \begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} \quad (63)$$

$$\boxed{P_{k+1} = (A^T A + \mathbf{a}\mathbf{a}^T)^{-1}} \quad (64)$$

$$\text{On sait que : } P_k^{-1} = A^T A \quad (65)$$

$$P_{k+1}^{-1} = (A^T A + \mathbf{a}\mathbf{a}^T) = P_k^{-1} + \mathbf{a}\mathbf{a}^T \quad (66)$$

$$\text{De } \theta_k = P_k A^T \mathbf{y}, \text{ il vient } A^T \mathbf{y} = P_k^{-1} \theta_k \quad (67)$$

$$\theta_{k+1} = P_{k+1} ((P_{k+1}^{-1} - \mathbf{a}\mathbf{a}^T) \theta_k + \mathbf{a}y) \quad (68)$$

$$= P_{k+1} P_{k+1}^{-1} \theta_k - P_{k+1} \mathbf{a}\mathbf{a}^T \theta_k + P_{k+1} \mathbf{a}y \quad (69)$$

$$= \theta_k - P_{k+1} (\mathbf{a}\mathbf{a}^T \theta_k - \mathbf{a}y) \quad (70)$$

$$\boxed{\theta_{k+1} = \theta_k - \underbrace{P_{k+1} \mathbf{a}(\mathbf{a}^T \theta_k - y)}_{\text{terme correcteur} \times \text{gain}}} \quad (71)$$

Terme correcteur \times gain = (erreur de prédiction) \times P_{k+1}

erreur de prédiction $\boxed{= (\mathbf{a}^T \theta_k - y)}$.

$$P_{k+1} = (P_k^{-1} + \mathbf{a}\mathbf{a}^T)^{-1} \quad (72)$$

En appliquant le lemme d'inversion avec $A = P_k^{-1}$, $B = \mathbf{a}$ et $C = \mathbf{a}^T$, on trouve

$$P_{k+1} = P_k - P_k \mathbf{a} (I + \mathbf{a}^T P_k \mathbf{a})^{-1} \mathbf{a}^T P_k \quad (73)$$

$$\Rightarrow P_{k+1} = P_k - \frac{P_k \mathbf{a} \mathbf{a}^T P_k}{I + \mathbf{a}^T P_k \mathbf{a}} \quad (74)$$

En résumé, l'estimateur récursif des MC pour $A\theta = \mathbf{y}$ est

$$\boxed{\begin{cases} P_{k+1} &= P_k - \frac{P_k \mathbf{a} \mathbf{a}^T P_k}{I + \mathbf{a}^T P_k \mathbf{a}} \\ \theta_{k+1} &= \theta_k - P_{k+1} \mathbf{a}_{k+1} (\mathbf{a}_{k+1}^T \theta_k - y_{k+1}) \end{cases}, k = 0, \dots, m-1.} \quad (75)$$

Pour commencer cet algorithme, on a besoin de θ_0 et P_0 . On choisit par exemple θ_0 et P_0 calculés à partir de n points : $\theta_n = (A_n^T A_n)^{-1} A_n^T \mathbf{y}_n$. Or, on remarque que $P_k = (P_0 + \sum_{i=1}^k \mathbf{a}_i^T \mathbf{a}_i)^{-1} = (P_0 + A_k^T A_k)^{-1}$. En prenant $P_0 = \alpha I \Rightarrow \lim_{\alpha \rightarrow \infty} P_0^{-1} = \lim_{\alpha \rightarrow \infty} \frac{1}{\alpha} I = 0$.

Remarque :

- $P_k \propto$ covariance des estimateurs (théorème de Gauss-Markov)
- l'estimateur des MC peut-être interprété come un filtre de Kalman

$$\begin{cases} \theta(k+1) &= \theta(k) \\ \mathbf{y}(k) &= \mathbf{a}^T(k) \theta(k) + \mathbf{e}(k) \end{cases} \quad (76)$$

- Pour des sorties multiples :

$$\begin{cases} P_{k+1} &= P_k - \frac{P_k \mathbf{a} \mathbf{a}^T P_k}{I + \mathbf{a}^T P_k \mathbf{a}} \\ \Theta_{k+1} &= \Theta_k - P_{k+1} \mathbf{a}_{k+1} (\mathbf{a}_{k+1}^T \Theta - y_{k+1}) \end{cases} \quad (77)$$

MC Récursifs pour séries temporelles

On a $P_k = (A^T A)^{-1}$, où k est le temps (*i.e.* le nombre de rangées de mon tableau). Si $k >$ le nombre de paramètres, alors $A^T A > 0$. Si $k \rightarrow \infty$, $\frac{1}{k} A^T A \rightarrow$ matrice constante non-singulière.

$$\lim_{k \rightarrow \infty} P_k = \lim_{k \rightarrow \infty} \frac{1}{k} \left(\frac{1}{k} A^T A \right)^{-1} = 0 \quad (78)$$

donc le gain d'adaptation $P_{k+1}a_{k+1}$ diminue.

+ chaque composante d'erreur est traitée de la même manière

+ ça marche bien pour les systèmes invariants dans le temps

– marche pas dans le cas des modèles temporelles

⇓

2 solutions

- ré-initialisé régulièrement P_k à $P_0 \Rightarrow$ convergence rapide
- introduire un facteur “d’oubli” λ qui pondère les données les plus récentes.

Par exemple,

$$\boxed{C(\theta) = \sum_i^n \lambda^{n-i} (y_i - \mathbf{a}_i^T \theta)^2 = (\mathbf{y} - A\theta)^T W (\mathbf{y} - A\theta)} \quad (79)$$

$$\text{avec } W = \begin{bmatrix} \lambda^{m-i} & & & 0 \\ O & \lambda^{m-2} & & 0 \\ & & \ddots & \\ & & & 1 \end{bmatrix}, 0 < \lambda < 1, \quad (80)$$

\Rightarrow estimateur des MC : $\boxed{\theta = (A^T W A)^{-1} W A^T \mathbf{y}}$

Exercice 4 (Forme récursive) *Développer les nouvelles équations pour les MC pondérés à partir de*

$$\theta_{k+1} = \left(\begin{bmatrix} [A]^T & [\lambda W & 0] \\ [\mathbf{a}^T] & [0 & 1] \end{bmatrix} \begin{bmatrix} [A] \\ [\mathbf{a}^T] \end{bmatrix} \right)^{-1} \begin{bmatrix} [A]^T & [\lambda W & 0] \\ [\mathbf{a}^T] & [0 & 1] \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ y \end{bmatrix} \quad (81)$$

Réponse :

$$\theta_{k+1} = \left(\begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \lambda W & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} \begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \lambda W & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ y \end{bmatrix} \quad (82)$$

$$= (\lambda A^T W A + \mathbf{a} \mathbf{a}^T)^{-1} (\lambda A^T W \mathbf{y} + \mathbf{a} y) \quad (83)$$

Posons $P_k = (A^T W A)^{-1}$ (84)

$$P_{k+1} = \left(\begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \lambda W & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} = (\lambda A^T W A + \mathbf{a} \mathbf{a}^T)^{-1} \quad (85)$$

$$\Rightarrow P_{k+1}^{-1} = P_k^{-1} + \mathbf{a} \mathbf{a}^T \quad (86)$$

Or :

$$\theta_k = P_k A^T W \mathbf{y} \quad (87)$$

$$\Rightarrow \theta_{k+1} = P_{k+1} (\lambda A^T W \mathbf{y} + \mathbf{a} y) \quad (88)$$

$$= \theta_k + P_{k+1} \mathbf{a} (y - \mathbf{a}^T \theta_k) \quad (89)$$

$$P_{k+1} = \frac{1}{\lambda} \frac{P_k \mathbf{a} \mathbf{a}^T P_k}{I + \mathbf{a}^T P_k \mathbf{a}} \quad (90)$$

$$(91)$$

Si $\lambda = 1$, on retrouve la formule précédente.

Propriétés statistiques de ces estimateurs

Examinons $y = A\theta + e$ d'un point de vue statistique, où $e \sim \mathcal{L}$ (loi aléatoire quelconque), θ la vraie valeur du paramètre et y une v.a. dépendent de e .

Definition 1 (estimateur biaisé) $\hat{\theta}$ est une estimateur non-biaisé si $E[\hat{\theta}] = \theta$.

Definition 2 (estimateur consistant) $\hat{\theta}_k$ est une estimateur consistant si $\lim_k P(\|\hat{\theta}_k - \theta\| \geq \epsilon) = 0, \forall \epsilon > 0$.

Definition 3 (estimateur minimum de variance) $\hat{\theta}$ est une estimateur à minimum de variance si $\forall \theta^*, cov(\hat{\theta}) \leq cov(\hat{\theta}^*)$.

Conditions de Gauss-Markov sur le vecteur-erreur \mathbf{e} qui rendent compte du bruit de mesure et de modélisation :

$$E[\mathbf{e}] = 0 \quad (92)$$

$$E[\mathbf{e}\mathbf{e}^T] = 0 \quad (93)$$

\mathbf{e} est un vecteur de dimension m et un vecteur aléatoire $\sim \mathcal{N}(0, \sigma^2 I_n)$.

Théoreme 1 (Gauss-Markov) *Si les conditions de Gauss-Markov sont vérifiées, alors l'estimateur $\hat{\theta}$ est non-biaisé et de variance minimale.*

Preuve:

$$E[\mathbf{y}] = E[A\theta] + E[\mathbf{e}] = A\theta \quad (94)$$

$$E[\hat{\theta}] = E[(A^T A)^{-1} A^T \mathbf{y}] = (A^T A)^{-1} A^T E[\mathbf{y}] = (A^T A)^{-1} A^T A\theta = \theta \quad (95)$$

Théoreme 2 (Gauss-Markov consistance) *Sous les conditions de GM, $\hat{\theta}$ est consistant si $(A^T A)^{-1} \rightarrow 0$ quand $m \rightarrow \infty$.*

Preuve :

$$\hat{\theta} = (A^T A)^{-1} A^T \mathbf{y} \quad (96)$$

$$= (A^T A)^{-1} A^T (A\theta + \mathbf{e}) = \theta + (A^T A)^{-1} A^T \mathbf{e} \quad (97)$$

$$\Rightarrow \text{cov}(\hat{\theta}) = E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \quad (98)$$

$$= E[(A^T A)^{-1} A^T \mathbf{e} \mathbf{e}^T A (A^T A)^{-T}] \quad (99)$$

$$= (A^T A)^{-1} A^T E[\mathbf{e} \mathbf{e}^T] A (A^T A)^{-T} \quad (100)$$

$$\boxed{\text{cov}(\hat{\theta}) = (A^T A)^{-1} A^T \sigma^2 I_n A (A^T A)^{-T} = \sigma^2 (A^T A)^{-1}} \quad (101)$$

$\Rightarrow \text{cov}(\hat{\theta}) \rightarrow 0$ quand $m \rightarrow \infty$ si $(A^T A)^{-1} \rightarrow 0$.

RAPPELS sur les TESTS DE CONFORMITÉ

Estimation d'une proportion

p : fréquence d'apparition du caractère C dans une population.

f : idem dans un échantillon de taille n . f est une estimation biaisée de p .

Quelle confiance accorder à cette estimation ?

On répond à cette question en déterminant un $]a, b[$ tel que l'on ait la probabilité α de se tromper en affirmant que p appartient à cet intervalle. L'intervalle obtenu est un *intervalle de confiance* de p au risque α .

nf : nombre individus ayant le caractère C dans un échantillon de taille n , suit la loi binomiale $\mathcal{B}(n, p)$.

Si on peut l'approximer par une loi normale, alors

$$U = \frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1).$$

On peut donc au risque α écrire :

$$f - u_\alpha \sqrt{\frac{p(1-p)}{n}} \leq p \leq f + u_\alpha \sqrt{\frac{p(1-p)}{n}}.$$

Pour trouver les bornes, on remplace p par f :

$$\left] f - u_\alpha \sqrt{\frac{f(1-f)}{n}}, f + u_\alpha \sqrt{\frac{f(1-f)}{n}} \right[.$$

Exercice 5 on a prélevé au hasard 100 individus. Sur ces 100 personnes, 20 sont atteints de myopie. Que peut on dire du pourcentage de myopes au niveau de la population ?

Estimation d'une moyenne

1. *cas d'une population gaussienne (μ, σ connus)*

si $X \sim \mathcal{N}(\mu, \sigma)$, $\bar{X} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$. Pour un risque α donné, on lit l'écart-réduit u_α dans la table 2 et :

$$-u_\alpha \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

soit :

$$\mu \in]\bar{x} - u_\alpha \frac{\sigma}{\sqrt{n}}, \bar{x} + u_\alpha \frac{\sigma}{\sqrt{n}}[.$$

2. *cas d'une population gaussienne (σ inconnu)*

$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$ suit la loi de Student à $\nu = n - 1$ ddl. Pour un risque α . Pour un risque α donné, on lit t_α dans la table 3 et :

$$\mu \in]\bar{x} - t_\alpha \frac{s}{\sqrt{n}}, \bar{x} + t_\alpha \frac{s}{\sqrt{n}}[.$$

3. *loi quelconque et grand échantillon*

Si $n \geq 30$, la va $U = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$. L'intervalle de confiance de μ au risque α s'écrit :

$$]\bar{x} - u_\alpha \frac{s}{\sqrt{n}}, \bar{x} + u_\alpha \frac{s}{\sqrt{n}}[.$$

Exercice 6 On suppose que la taille d'un enfant de 13 à 14a dans une population donnée est une va gaussienne de moyenne μ et d'écart-type σ . Pour estimer μ dans une ville, un chercheur commence par mesurer la taille (en cm) de 25 enfants choisis au hasard et on obtient les résultats suivants :

Tailles]130;135[]135;140[]140;145[]145;150[]150;155[]155;160[]160;165[
Effectifs	1	3	5	8	5	1	2

1. Calculer une estimation ponctuelle non-biaisée de μ et de σ^2 .
2. Donner au niveau de confiance 0.95 un intervalle de la moyenne de la population.
3. Le chercheur désire améliorer sa connaissance de μ en augmentant la taille de l'échantillon. En supposant que s reste la même que celle obtenue, quelle taille minimum doit-il choisir pour obtenir un intervalle de la moyenne au seuil de 5% d'amplitude inférieure à 1cm ?

Exercice 7 *On mesure la force à la traction d'une prothèse en mesurant la force à partir de laquelle elle se casse. Pour 10 prothèses utilisées, on note :*

19,6; 19,9; 20,4; 19,8; 20,5; 21,0; 18,5; 19,7; 18,4; 19,4

On suppose que X suit une loi gaussienne $\mathcal{N}(\mu, \sigma)$.

a. calculer une estimation non-biaisée de μ et de σ .

b. donner pour $\alpha = 0.05$ un intervalle de confiance I_1 de la moyenne de la population

Comparaison de 2 moyennes observées (échantillons indépendants)

Dans 2 populations P_1 et P_2 , on étudie la va X .

population	P_1	P_2
moyenne	μ_1	μ_2
écart-type	σ_1	σ_2

 \rightarrow

échantillon	1	2
taille	n_1	n_2
moyenne	\bar{x}_1	\bar{x}_2
variance estimée	s_1^2	s_2^2

La différence entre les moyennes expérimentales \bar{x}_1 et \bar{x}_2 est significative ou au contraire explicable par les fluctuations de l'échantillonnage ?

$H_0 : \mu_1 = \mu_2$ ou encore P_1 et P_2 sont homogènes.

$H_1 : \mu_1 \neq \mu_2$ (test bilatéral) ou $\mu_1 > \mu_2$ (test unilatéral) si le signe de $\mu_1 - \mu_2$ est connu.

Cas des grands échantillons ($n_{1,2} > 30$)

- Sous H_0 , \forall la loi de X ,
$$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1).$$
- on calcule $u = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$.
- Pour α fixé, on lit 1 page 44 la borne u_α .

$$\begin{cases} \text{Si } u \in] -u_\alpha, u_\alpha[, & H_0 \text{ ne peut être rejetée} \\ \text{Si } u \notin] -u_\alpha, u_\alpha[, & \text{on rejette } H_0 \text{ au risque } \alpha \end{cases}$$

Cas des petits échantillons de populations Gaussiennes

- Sous H_0 , si $X \sim \mathcal{N}$ dans P_1 et P_2 et si $\sigma_1^2 = \sigma_2^2 = \sigma^2$, alors la va $T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})}} \sim T$ (Student) à $n_1 + n_2 - 2$ ddl.
- Il faut d'abord tester l'égalité des 2 variances. Sous l'hypothèse $\sigma_1^2 = \sigma_2^2$, alors $\sigma^2 \approx \hat{\sigma}^2 = \frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}$.

et

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\hat{\sigma}^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

- Pour α fixé, on lit Table 2 page 45 le nombre t_α pour ν ddl t.q. $P(|T| \geq t_\alpha) = \alpha$

$$\begin{cases} \text{Si } t \in] -t_\alpha, t_\alpha[, & H_0 \text{ ne peut être rejetée} \\ \text{Si } t \notin] -t_\alpha, t_\alpha[, & \text{on rejette } H_0. \end{cases}$$

Exemple 4 (Comparaison de 2 moyennes observées) *On compare la consommation de caféine chez 112 cancéreux : moyenne 147,2 mg/jour, écart type estimé 101,8 mg/jour à celle de 185 non cancéreux : moyenne 132,9 mg/jour - écart type 115,7 mg/jour. On prend un risque à 5%.*

Réponse: Test bilatéral, grands échantillons : $u = \frac{147,2 - 132,9}{\sqrt{\frac{101,8^2}{112} + \frac{115,7^2}{185}}} = 1,11 < 1,96 \Rightarrow$

Différence non significative.

On aurait pu utiliser un t (Student) mais ce serait plus long car il faut alors calculer la variance commune.

Exercice 8 *La comparaison dynamométrique de la force ventilatoire d'un groupe A d'adultes en souffrance respiratoire à celle d'une groupe B d'adultes de référence a donné les résultats suivants (en kg.m^{-1})*

groupe A : 57; 52; 49; 52; 53; 56

groupe B : 45; 51; 37; 52; 40; 42; 50; 35; 52

Que peut on conclure en ce qui concerne la différence de force ventilatoire entre les 2 groupe ? On admettra que la va étudiée suit une loi normale dans chacune des 2 populations et que les variances sont égales.

Analyse de la variance=Test d'homogénéité

Tester l'effet d'un facteur contrôlé A ayant p modalités sur les moyennes d'une variable aléatoire Y

Facteur A					
modalité	$i = 1$...	i	...	$i = p$
résultats expérimentaux	y_{11}		y_{i1}		y_{p1}
	\vdots		\vdots		\vdots
	y_{1j}		y_{ij}		y_{pj}
	\vdots		\vdots		\vdots
	y_{1n_1}		y_{in_i}		y_{pn_p}
Sommes	$\bar{y}_1 = \sum_j y_{1j}$...	$\bar{y}_j = \sum_j y_{ij}$...	$\bar{y}_p = \sum_j y_{pj}$

Modèle de l'ANOVA :

$$\begin{aligned}
 H_0 : \mu_i = \mu_j = \mu &\Leftrightarrow \boxed{y_{ij} = \mu + e_{ij} \sim \mathcal{N}(\mu, \sigma)} && \text{pas d'effet facteur } A \\
 &\hat{\mu} = \bar{y} = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij}, n = \sum_{i=1}^p n_i \\
 H_1 : \exists \mu_i \neq \mu_j &\Leftrightarrow \boxed{y_{ij} = \mu + a_i + e_{ij} \sim \mathcal{N}(\mu + a_i, \sigma^2)} && \text{effet du facteur } A \\
 &\hat{\mu} + \hat{a}_i = \bar{y}_i, \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}.
 \end{aligned}$$

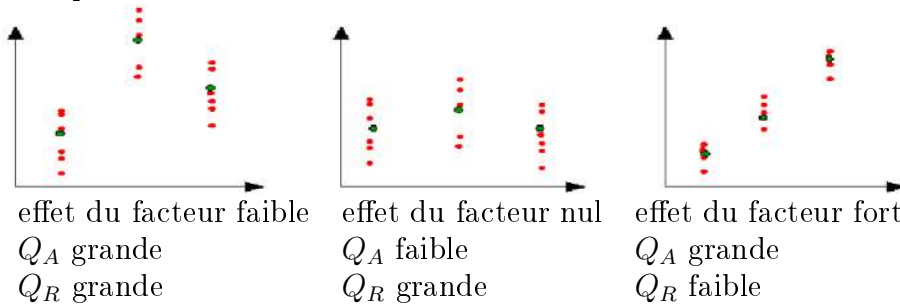
Décomposition de la variation totale (Équation d'analyse de variance)

$$\underbrace{\sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2}_{Q_T} = \underbrace{\sum_{i=1}^p n_i (\bar{y}_i - \bar{y})^2}_{Q_A} + \underbrace{\sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}_{Q_R} \quad (102)$$

Q_T : variation totale,

Q_A : variation des écarts liés au effets du facteur A (**variation inter**),

Q_R : variation des écarts résiduelles (**variation intra**) interne à chaque modalité. Par exemple :



$F_A = \frac{q_A}{q_R}$ suit une loi de Fisher-Snedecor, comparée à F_{seuil} pour un risque d'erreur α fixé et $(p - 1, N - p)$ ddl

- si $F_A > F_{\text{seuil}}$, H_0 est rejetée au risque α : le facteur contrôlé A a un effet significatif en moyenne sur les valeurs de la variable étudiée.
- si $F_A > F_{\text{seuil}}$, H_0 est acceptée : le facteur A n'a pas d'effet significatif.

Exemple 5 *Un manufacturier souhaite voir si trois machines à remplissage ont des rendements différents. Il assigne aléatoirement 15 ouvriers entraînés de la même manière aux 3 machines (5 par machine). Avec un risque de 5%, les 3 machines ont elles un rendement*

	M_1	M_2	M_3
différents.	25,40	23,40	20,00
	26,31	21,80	22,20
	24,10	23,50	19,75
	23,74	22,75	20,60
	25,10	21,60	20,40

$H_0 : \mu_1 = \mu_2 = \mu_3$

$H_1 : \text{pas toutes égales.}$

Source de variation	Somme des carrés	ν (ddl)	$\frac{Q}{\nu}$	F_A
Facteur	$Q_A = 47,1640$	$3 - 1 = 2$	$q_A = 23,5820$	25,60
Erreur	$Q_R = 11,0532$	$15 - 3 = 12$	$q_R = 0,9211$	
Total	$Q_T = 58,2172$	$15 - 1 = 14$		

Statistique de test : $F_A = \frac{23,58}{0,92} = 25,60 \gg F_{\alpha=0,05\%}(2, 12) = 3,89 \Rightarrow$ Rejet à $\alpha = 0,05\%$, au moins une des μ_i diffère des autres.

Exemple 6 *Pour vérifier si la présence de molybdène influe sur le dosage de l'aluminium, on fait $p = 4$ prélèvements d'une même solution d'aluminium. A ces solutions sont ajoutées des quantités variables de molybdène de sorte que les rapports Mo/Al obtenus se situent dans un certain domaine de variation.*

$i = 1$	$i = 2$	$i = 3$	$i = 4$
157	140	162	122
148	129	159	134
145	135	155	135

Réponse:

Source de variation	Somme des carrés	ν (ddl)	$\frac{Q}{\nu}$	F_A
Facteur	$Q_A = 1571$	3	$q_A = 523,6$	15,63
Erreur	$Q_R = 268$	8	$q_R = 33,5$	
Total	$Q_T = 1839$	11		

Statistique de test : $F_A = 15,63 \gg F_{0,05\%}(3,8) = 7,59 \Rightarrow$ Rejet à $\alpha = 0,05\%$, de l'hypothèse d'absence d'influence du facteur contrôlé : le dosage de l'aluminium est perturbé par la présence de molybdène.

Les tests statistiques. Pourquoi faire ?

Un certain nombre d'**hypothèses** doivent en principe être vérifiées avant une analyse statistique :

- **Les mesures reflètent les vraies valeurs** : Cette hypothèse n'est malheureusement jamais réalisée en pratique, du fait des erreurs systématiques ou aléatoires.
- **Les données sont consistantes** : Aucune modification dans les conditions internes du système n'intervient durant la période d'observation (position du pluviomètre, procédures d'observation, observateur unique).
- **La série des données est stationnaire** : Les propriétés de la loi statistique qui régit le phénomène (moyenne, variance ou moments d'ordre supérieur) sont **invariantes au cours du temps**.
- **Les données sont homogènes** : Durant la période de mesure, les **caractéristiques** du phénomène étudié **n'évoluent pas** (*i.e.* pas de variations climatiques, ni de variations du régime des débits dues à une déforestation ou un reboisement). Les données reflètent un seul et unique phénomène.
- **La série d'observation possède un caractère aléatoire et simple** : cette hypothèse de base sert de **préalable** à la mise en œuvre de nombreuses méthodes statistiques. Un échantillon aléatoire signifie que tous les individus de la population ont la même probabilité d'être prélevés. Un échantillon simple signifie que le prélèvement d'un individu n'influe pas sur la probabilité de tirage des individus suivants.
- **La série doit être suffisamment longue** : La **longueur de la série** influe sur les erreurs d'échantillonnage, notamment sur le calcul des **moments d'ordre supérieurs**.

Un test statistique = méthode permettant de prendre une décision à partir d'informations fournies par un échantillon.

Tester une hypothèse H_1 contre une hypothèse H_0

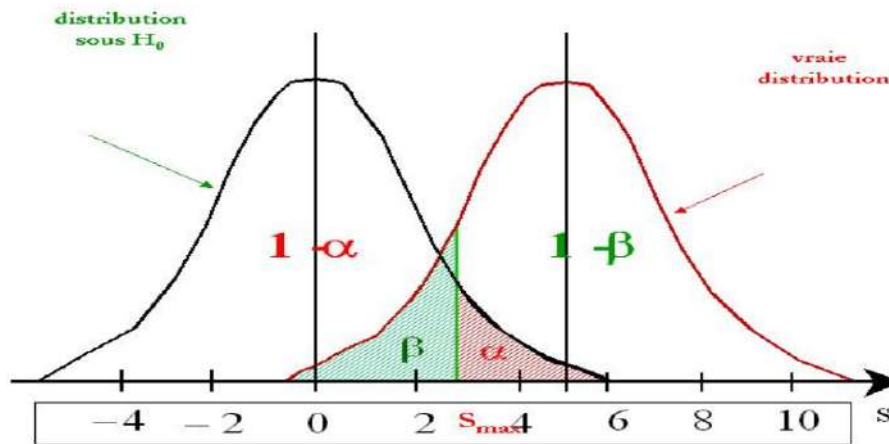
	H_0 vraie	H_1 vraie
H_0 décidée	$1 - \alpha$	β
H_0 rejetée	α	$1 - \beta$

- Le comportement de X doit être connu sous H_0 ;
- α = **erreur de première espèce** (erreur "**type I**" en anglais), $(1 - \alpha)$ est le niveau de confiance
- β **erreur de seconde espèce** (erreur "**type II**")

Un test est bon si on arrive à minimiser α et β .

La seule façon d'y parvenir est d'augmenter n .

- L'hypothèse nulle, H_0 est l'hypothèse que l'on désire contrôler \Leftrightarrow "il n'existe pas de différence entre les paramètres comparés ou que la différence observée n'est pas significative" (due aux fluctuations d'échantillonnage). H_0 est formulée dans le but d'être rejetée.
- L'hypothèse alternative, H_1 équivaut à dire " H_0 est fausse".
- La puissance d'un test augmente avec taille n de l'échantillon.



$$\beta = P(\text{accepter } H_0 | H_0 \text{ fausse}) \quad \alpha = P(\text{rejeter } H_0 | H_0 \text{ vraie})$$

Exemple 7 (Calibration d'un instrument) L'erreur de mesure d'un instrument suit une loi Gaussienne $X \sim \mathcal{N}(0, 0.01)$ quand il est bien calibré. On souhaite vérifier sa calibration en testant si la moyenne de son erreur de mesure peut être acceptée comme étant nulle. Donc notre hypothèse ici est $H_0 : \bar{x} = 0$.

On fait les mesures suivantes :

0.1294, -0.0336, 0.1714, 0.2624, 0.030, 0.1858, 0.2254, -0.0594, -0.044, 0.157

Comme la moyenne empirique $\hat{X} \sim \mathcal{N}(\bar{x}, \sigma^2/n) = \mathcal{N}(0, 0.01/10)$, alors $\hat{X} \in [-1.96 \frac{\sigma}{\sqrt{10}}, 1.96 \frac{\sigma}{\sqrt{10}}] = [-0.062, 0.062]$ avec une probabilité de 95% si H_0 est vraie. Or, $\hat{X} = 0.1025$ est dehors de l'intervalle. Donc on rejette H_0 . Après calibration, on mesure :

-0.040, 0.069, 0.0816, 0.0712, 0.129, 0.1191, -0.1202, -0.002, -0.0157

Donc la moyenne empirique $\hat{X} = 0.0359$ est dans l'intervalle, donc H_0 est acceptée avec un niveau de confiance de 95%.

Test d'hypothèse de Neyman-Pearson

Les erreurs de type I et II peuvent être réduites simultanément pendant le même test (elles sont liées : réduire l'une augmente l'autre) en augmentant la **taille de l'échantillon**.

On peut donc contrôler l'erreur de type I.

⇓

Affecter l'erreur de décision qui aura le plus de conséquences à l'erreur de type I.

Soit le test d'hypothèse : $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ et X_1, \dots, X_n un échantillon.

Le **test du taux de vraisemblance de Neyman-Pearson** s'écrit :

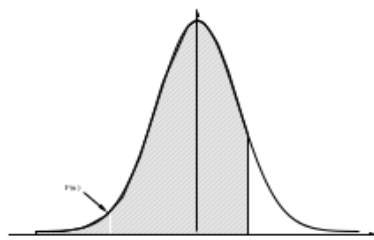
$$\begin{cases} \text{rejeter } H_0 & \text{si } \frac{f(X_1, \dots, X_n | \theta = \theta_1)}{f(X_1, \dots, X_n | \theta = \theta_0)} < \lambda, \\ \text{accepter } H_0 & \text{si } \frac{f(X_1, \dots, X_n | \theta = \theta_1)}{f(X_1, \dots, X_n | \theta = \theta_0)} > \lambda, \end{cases} \quad (103)$$

Ce test est optimal.

Dans (103), $\frac{f(X_1, \dots, X_n | \theta = \theta_1)}{f(X_1, \dots, X_n | \theta = \theta_0)}$ est un scalaire. λ est choisi de sorte que la probabilité d'erreur de type I $= \alpha$.

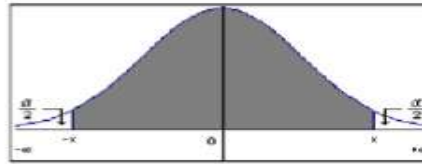
Exercice 9 On a mesuré chez 50 patients monitorés choisis au hasard et de façon indépendante en clinique la production d'adrénaline après un choc opératoire. On a obtenu : $\sum_{i=1}^{50} (x_i - \bar{x})^2 = 3,713 \text{ g}^2$.

On suppose que la production individuelle peut être considéré comme une va approximativement gaussienne, d'écart-type 0,1 g. Au risque 5%, les résultats expérimentaux sont ils conformes à cette valeur théorique.



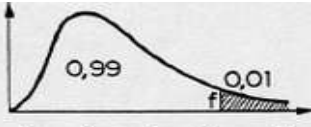
u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Table 1: Table loi normale.



α	1	0,8	0,6	0,4	0,2	0,1	0,05	0,02	0,01	0,002	0,001
$1 - \alpha$	0	0,2	0,4	0,6	0,8	0,9	0,95	0,98	0,99	0,998	0,999
$v = ddl$											
1	0,0000	0,3249	0,7265	1,3764	3,0777	6,3137	12,706	31,821	63,656	318,29	636,58
2	0,0000	0,2887	0,6172	1,0607	1,8856	2,9200	4,3027	6,9645	9,9250	22,328	31,600
3	0,0000	0,2767	0,5844	0,9785	1,6377	2,3534	3,1824	4,5407	5,8408	10,214	12,924
4	0,0000	0,2707	0,5686	0,9410	1,5332	2,1318	2,7765	3,7469	4,6041	7,1729	8,6101
5	0,0000	0,2672	0,5594	0,9195	1,4759	2,0150	2,5706	3,3649	4,0321	5,8935	6,8685
6	0,0000	0,2648	0,5534	0,9057	1,4398	1,9432	2,4469	3,1427	3,7074	5,2075	5,9587
7	0,0000	0,2632	0,5491	0,8960	1,4149	1,8946	2,3646	2,9979	3,4995	4,7853	5,4081
8	0,0000	0,2619	0,5459	0,8889	1,3968	1,8595	2,3060	2,8965	3,3554	4,5008	5,0414
9	0,0000	0,2610	0,5435	0,8834	1,3830	1,8331	2,2622	2,8214	3,2498	4,2969	4,7809
10	0,0000	0,2602	0,5415	0,8791	1,3722	1,8125	2,2281	2,7638	3,1693	4,1437	4,5868
11	0,0000	0,2596	0,5399	0,8755	1,3634	1,7959	2,2010	2,7181	3,1058	4,0248	4,4369
12	0,0000	0,2590	0,5386	0,8726	1,3562	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178
13	0,0000	0,2586	0,5375	0,8702	1,3502	1,7709	2,1604	2,6503	3,0123	3,8520	4,2209
14	0,0000	0,2582	0,5366	0,8681	1,3450	1,7613	2,1448	2,6245	2,9768	3,7874	4,1403
15	0,0000	0,2579	0,5357	0,8662	1,3406	1,7531	2,1315	2,6025	2,9467	3,7329	4,0728
16	0,0000	0,2576	0,5350	0,8647	1,3368	1,7459	2,1199	2,5835	2,9208	3,6861	4,0149
17	0,0000	0,2573	0,5344	0,8633	1,3334	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651
18	0,0000	0,2571	0,5338	0,8620	1,3304	1,7341	2,1009	2,5524	2,8784	3,6105	3,9217
19	0,0000	0,2569	0,5333	0,8610	1,3277	1,7291	2,0930	2,5395	2,8609	3,5793	3,8833
20	0,0000	0,2567	0,5329	0,8600	1,3253	1,7247	2,0860	2,5280	2,8453	3,5518	3,8496
21	0,0000	0,2566	0,5325	0,8591	1,3232	1,7207	2,0796	2,5176	2,8314	3,5271	3,8193
22	0,0000	0,2564	0,5321	0,8583	1,3212	1,7171	2,0739	2,5083	2,8188	3,5050	3,7922
23	0,0000	0,2563	0,5317	0,8575	1,3195	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676
24	0,0000	0,2562	0,5314	0,8569	1,3178	1,7109	2,0639	2,4922	2,7970	3,4668	3,7454
25	0,0000	0,2561	0,5312	0,8562	1,3163	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251
26	0,0000	0,2560	0,5309	0,8557	1,3150	1,7056	2,0555	2,4786	2,7787	3,4350	3,7067
27	0,0000	0,2559	0,5306	0,8551	1,3137	1,7033	2,0518	2,4727	2,7707	3,4210	3,6895
28	0,0000	0,2558	0,5304	0,8546	1,3125	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739
29	0,0000	0,2557	0,5302	0,8542	1,3114	1,6991	2,0452	2,4620	2,7564	3,3963	3,6595
30	0,0000	0,2556	0,5300	0,8538	1,3104	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460
40	0,0000	0,2550	0,5286	0,8507	1,3031	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510
50	0,0000	0,2547	0,5278	0,8489	1,2987	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960
60	0,0000	0,2545	0,5272	0,8477	1,2958	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602
70	0,0000	0,2543	0,5268	0,8468	1,2938	1,6669	1,9944	2,3808	2,6479	3,2108	3,4350
80	0,0000	0,2542	0,5265	0,8461	1,2922	1,6641	1,9901	2,3739	2,6387	3,1952	3,4164
90	0,0000	0,2541	0,5263	0,8456	1,2910	1,6620	1,9867	2,3685	2,6316	3,1832	3,4019
100	0,0000	0,2540	0,5261	0,8452	1,2901	1,6602	1,9840	2,3642	2,6259	3,1738	3,3905
200	0,0000	0,2537	0,5252	0,8434	1,2858	1,6525	1,9719	2,3451	2,6006	3,1315	3,3398
∞	0,0000	0,2533	0,5244	0,8416	1,2816	1,6449	1,9600	2,3263	2,5758	3,0903	3,2906

Table 2: Table loi de Student.



(Les valeurs de la première ligne doivent être multipliées par 10)

$\frac{p_1}{p_2}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	405	500	540	563	576	586	593	598	602	606	608	611	613	614	616	617	618	619
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1	27.0	26.9	26.9	26.8	26.8	26.8
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.4	14.3	14.2	14.2	14.2	14.1	14.1
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.80	9.82	9.77	9.72	9.68	9.64	9.61
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60	7.56	7.52	7.48	7.45
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36	6.31	6.27	6.24	6.21
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56	5.52	5.48	5.44	5.41
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.00	4.96	4.92	4.89	4.86
10	10.0	7.50	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56	4.52	4.49	4.46
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29	4.25	4.21	4.18	4.15
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01	3.97	3.94	3.91
13	9.17	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.91	3.86	3.82	3.78	3.75	3.72
14	8.96	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70	3.66	3.62	3.59	3.56
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52	3.49	3.45	3.42
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50	3.45	3.41	3.37	3.34	3.31
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40	3.35	3.31	3.27	3.24	3.21
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.32	3.27	3.23	3.19	3.16	3.13
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19	3.15	3.12	3.08	3.05
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09	3.05	3.02	2.99
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.12	3.07	3.03	2.99	2.96	2.93
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.07	3.02	2.98	2.94	2.91	2.88
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	3.02	2.97	2.93	2.89	2.86	2.83
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.98	2.93	2.89	2.85	2.82	2.79
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.94	2.89	2.85	2.81	2.78	2.75
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96	2.90	2.86	2.82	2.78	2.74	2.72
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.99	2.93	2.87	2.82	2.78	2.75	2.71	2.68
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90	2.84	2.79	2.75	2.72	2.68	2.65
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.93	2.87	2.81	2.77	2.73	2.69	2.66	2.63
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70	2.66	2.63	2.60
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93	2.86	2.80	2.74	2.70	2.66	2.62	2.58	2.55
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89	2.82	2.76	2.70	2.66	2.62	2.58	2.55	2.51
36	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86	2.79	2.72	2.67	2.62	2.58	2.54	2.51	2.48
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83	2.75	2.69	2.64	2.59	2.55	2.51	2.48	2.45
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52	2.48	2.45	2.42
42	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78	2.70	2.64	2.59	2.54	2.50	2.46	2.43	2.40
44	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75	2.68	2.62	2.56	2.52	2.47	2.44	2.40	2.37
46	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73	2.66	2.60	2.54	2.50	2.45	2.42	2.38	2.35
48	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.72	2.64	2.58	2.53	2.48	2.44	2.40	2.37	2.33
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79	2.70	2.63	2.56	2.51	2.46	2.42	2.38	2.35	2.32
55	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53	2.47	2.42	2.38	2.34	2.31	2.28
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39	2.35	2.31	2.28	2.25
65	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61	2.53	2.47	2.42	2.37	2.33	2.29	2.26	2.23
70	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45	2.40	2.35	2.31	2.27	2.23	2.20
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.36	2.31	2.27	2.23	2.20	2.17
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52	2.45	2.39	2.33	2.29	2.24	2.21	2.17	2.14
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.31	2.26	2.22	2.19	2.15	2.12
125	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.66	2.55	2.47	2.39	2.33	2.28	2.23	2.19	2.15	2.11	2.08
150	6.81	4.75	3.92	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.37	2.31	2.25	2.20	2.16	2.12	2.09	2.06
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27	2.22	2.17	2.13	2.09	2.06	2.02
300	6.72	4.68	3.85	3.38	3.08	2.86	2.70	2.57	2.47	2.38	2.31	2.24	2.19	2.14	2.10	2.06	2.03	1.99
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.28	2.22	2.17	2.12	2.07	2.04	2.00	1.97
1000	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.27	2.20	2.15	2.10	2.06	2.02	1.98	1.95
∞	6.62	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.13	2.08	2.04	2.00	1.97	1.93

Table 3: Table loi de Fisher-Snedecor.

Les 4 Tests les plus utiles en recherche

1. Test de Dixon

Au cours d'une expérimentation, un des résultats semble s'écarter notablement des autres. Trop souvent, cette valeur est éliminée en la considérant comme aberrante.

Dangereux sans vérification préalable.

La bonne attitude : Si aucune cause accidentelle (erreur de lecture, faute de calcul, etc) n'a pu être détectée, il faut avoir recours à un test statistique justifiant l'élimination de la valeur aberrante avec une probabilité p , choisie à l'avance, de se tromper.

Le test de Dixon, que nous allons exposer, permet de réaliser cela.

Une expérimentation a conduit à n résultats numériques ($n \geq 3$).

On classe ces résultats par ordre **croissant**

$$x_1 < x_2 < \dots < x_{n-1} < x_n \quad (104)$$

Le test permet alors de tester si la première valeur x_1 ou la dernière valeur x_n est aberrante.

On calcule les rapports suivants :

$3 \leq n \leq 7$	$r_{10} = \frac{y_2 - y_1}{y_n - y_1}$	$r_{10} = \frac{y_n - y_{n-1}}{y_n - y_1}$
$8 \leq n \leq 10$	$r_{10} = \frac{y_2 - y_1}{y_{n-1} - y_1}$	$r_{10} = \frac{y_n - y_{n-1}}{y_n - y_2}$
$11 \leq n \leq 13$	$r_{10} = \frac{y_3 - y_1}{y_{n-1} - y_1}$	$r_{10} = \frac{y_n - y_{n-2}}{y_n - y_2}$
$14 \leq n \leq 30$	$r_{10} = \frac{y_3 - y_1}{y_{n-2} - y_1}$	$r_{10} = \frac{y_n - y_{n-2}}{y_n - y_3}$

La table de Dixon donne les **valeurs critiques** de ces rapports au niveau de risque 10%, 5% et 1%.

La règle de décision : si la valeur du rapport est inférieure à la valeur critique, on est pas justifié, au risque donné, d'éliminer l'observation.

Exemple 8 *Valeurs observées* : 148, 151, 152, 153, 160

Nous avons 5 valeurs et la dernière semble anormalement élevée.

Or, $r_{10} = \frac{160-153}{160-148} = 0,583$.

Pour $n = 5$ observations, la valeur critique lue dans la table est, au risque de 5%, 0,642. $0,583 < 0,642$. Donc, il n'est pas justifié, au risque de 5%, d'éliminer la valeur 160.

2. Test de normalité, test de Shapiro et Wilk

(valable pour des tailles d'échantillons $5 < n < 38$).

Observations rangées par ordre croissant :

$$y_1 \leq y_2 \leq \dots \leq y_{n-1} \leq y_n. \quad (105)$$

calculs :

$$y = \sum_{i=1}^n y_i \quad T_n = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (106)$$

$$\begin{aligned} d_1 &= y_n - y_1 \\ d_2 &= y_{n-1} - y_{n-2} \\ &\vdots \end{aligned} \quad (107)$$

$$W = \frac{\sum_{j=1}^p a_j d_j}{T_n} \quad \text{où les } a_j \text{ sont des coefficients tabulés} \quad (108)$$

Pour un risque à α ,

$$\text{Si } \begin{cases} W > W_{\text{crit}} & \text{accepter l'hypothèse de normalité de la série} \\ W < W_{\text{crit}} & \text{rejeter l'hypothèse de normalité} \end{cases} \quad (109)$$

à partir de la table 1 page 50 de Shapiro et Wilk.

Exercice 10 On a titré une série de 10 lots de Streptomycine par dosage au maltol. On a obtenu les résultats suivants diminués de 700 unités : 60 – 80 – 55 – 45 – 60 – 65 – 65 – 60 – 70 – 40. Tester la normalité de cette série.

3. Test sur un coefficient de corrélation linéaire

On considère X et Y deux variables aléatoires quantitatives. Le coefficient de corrélation linéaire est le nombre

$$\rho = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2} \sqrt{\sum_i^n (y_i - \bar{y})^2}}, \quad (110)$$

Tester l'hypothèse H_0 contre l'alternative H_1 :

H_0 : $\rho = 0$, i.e. X et Y sont indépendantes

H_1 : X et Y sont significativement corrélés

Si H_0 est vraie, alors $T = \sqrt{n-2} \frac{\rho}{\sqrt{1-\rho^2}} \sim \mathcal{T}(n-2)$

n = nombre de couples de valeurs expérimentales (au minimum 3)
 $\nu = n - 2$ est le nombre de ddl.

(cf table loi de Student page 45 pour α et ν fixés la valeur de $r_{\mathcal{T}(n-2)}$). Donc,

$$\begin{cases} \text{si } -r_{\mathcal{T}(n-2)} < \rho < +r_{\mathcal{T}(n-2)}, & \text{accepter } H_0 \text{ au risque } \alpha, \\ \text{sinon,} & \text{rejeter } H_0. \end{cases} \quad (111)$$

Exercice 11 *On étudie la vitesse de dissolution d'un type de streptomycine en poudre. Pour cela, on utilise des lots de streptomycine et, à chaque lot, on associe la densité X et la vitesse de dissolution Y . Les résultats sur 13 lots sont les suivants :*

1140	1092	1127	1175	1162	1105	1160	1143	1170	1105	1150	1145	1120
95	35	15	110	105	20	70	90	100	45	45	55	45

On cherche à savoir s'il y a corrélation entre ces deux variables.

i \ n	2	3	4	5	6	7	8	9	10
1	0,7071	0,7071	0,6872	0,6646	0,6431	0,6233	0,6052	0,5888	0,5739
2			0,1677	0,2413	0,2806	0,3031	0,3164	0,3244	0,3291
3					0,0875	0,1401	0,1743	0,1976	0,2141
4							0,0561	0,0947	0,1224
5									0,0399

i \ n	11	12	13	14	15	16	17	18	19	20
1	0,5601	0,5475	0,5359	0,5251	0,5150	0,5056	0,4963	0,4886	0,4808	0,4734
2	0,3315	0,3325	0,3325	0,3318	0,3306	0,3290	0,3273	0,3253	0,3232	0,3211
3	0,2260	0,2347	0,2412	0,2460	0,2495	0,2521	0,2540	0,2553	0,2561	0,2565
4	0,1429	0,1586	0,1707	0,1802	0,1878	0,1939	0,1988	0,2027	0,2059	0,2085
5	0,0695	0,0922	0,1099	0,1240	0,1353	0,1447	0,1524	0,1587	0,1641	0,1686
6		0,0303	0,0539	0,0727	0,0880	0,1005	0,1109	0,1197	0,1271	0,1334
7				0,0240	0,0433	0,0593	0,0725	0,0837	0,0932	0,1013
8						0,0196	0,0359	0,0496	0,0612	0,0711
9								0,0163	0,0303	0,0422
10										0,0140

i \ n	21	22	23	24	25	26	27	28	29	30
1	0,4643	0,4590	0,4542	0,4493	0,4450	0,4407	0,4366	0,4328	0,4291	0,4254
2	0,3185	0,3156	0,3126	0,3098	0,3069	0,3043	0,3018	0,2992	0,2968	0,2944
3	0,2578	0,2571	0,2563	0,2554	0,2543	0,2533	0,2522	0,2510	0,2499	0,2487
4	0,2119	0,2131	0,2139	0,2145	0,2148	0,2151	0,2152	0,2151	0,2150	0,2148
5	0,1736	0,1764	0,1787	0,1807	0,1822	0,1836	0,1848	0,1857	0,1064	0,1870
6	0,1399	0,1443	0,1480	0,1512	0,1539	0,1563	0,1584	0,1601	0,1616	0,1630
7	0,1092	0,1150	0,1201	0,1245	0,1283	0,1316	0,1346	0,1372	0,1395	0,1415
8	0,0804	0,0878	0,0941	0,0997	0,1046	0,1089	0,1128	0,1162	0,1192	0,1219
9	0,0530	0,0618	0,0696	0,0764	0,0823	0,0876	0,0923	0,0965	0,1002	0,1036
10	0,0263	0,0368	0,0459	0,0539	0,0610	0,0672	0,0728	0,0778	0,0822	0,0862
11		0,0122	0,0228	0,0321	0,0403	0,0476	0,0540	0,0598	0,0650	0,0697
12				0,0107	0,0200	0,0284	0,0358	0,0424	0,0483	0,0537
13						0,0094	0,0178	0,0253	0,0320	0,0381
14								0,0084	0,0159	0,0227
15										0,0076

Figure 1: Coefficients (a_1, \dots, a_p) de Shapiro-Wilk.

Réponse :

$\rho = 0,777$ ce qui laisse penser corrélation est mauvaise.

Effectuons alors un test sur ce coefficient de corrélation. Il y a 13 observations, donc $\nu = 13 - 2 = 11$. Au risque de 5 % on trouve dans la table ci-dessous $r = 0,5529$.

ν	1	2	3	4	5	6	7	8	9	10
5%	0.9969	0.9500	0.8783	0.8114	0.7545	0.7067	0.6664	0.6319	0.6021	0.5760
10%	0.9999	0.9900	0.9587	0.9172	0.8745	0.8343	0.7977	0.7646	0.7348	0.7079
ν	11	12	13	14	15	16	17	18	19	20
5%	0.5529	0.5324	0.5139	0.4973	0.4821	0.4683	0.4555	0.4438	0.4329	0.4227
10%	0.6835	0.6614	0.6411	0.6226	0.6055	0.5897	0.5751	0.5614	0.5487	0.5368

$\rho \notin [-0.5529; 0.5529]$, donc on accepte H_0 (on a moins de 5 chances sur 100 de se tromper en affirmant que X et Y sont indépendantes).

Remarquons que même au risque de 1%, $r = 0.6835$: là encore, on accepterait H_0 .

4. Comparaison de 2 écarts-types. Test de Fisher

Soit deux séries de mesures de taille n_1 et n_2 .

échantillon	1	2
taille	n_1	n_2
moyenne	\bar{x}_1	\bar{x}_2
variance estimée	s_1^2	s_2^2

Le plus grand écart-type sera noté σ_1 et le plus petit σ_2 , on aura donc $\sigma_1 > \sigma_2$. σ_1 est connu avec $\nu_1 = n_1 - 1$ ddl et l'écart-type σ_2 avec $\nu_2 = n_2 - 1$ ddl.

- **Hypothèses du test**

$$H_0 \quad \sigma_1 \approx \sigma_2$$

$$H_1 \quad \sigma_1 > \sigma_2 \text{ significativement}$$

- statistique à calculer : $F = \frac{\sigma_1}{\sigma_2}$. Sous H_0 , $F \sim$ loi de Fisher.

- On lit dans la table de Fisher-Snedecor (au risque) $\alpha F_{\text{critique}}(\nu_1, \nu_2)$

$F < F_{\text{critique}}(\nu_1, \nu_2)$	accepter H_0
$F > F_{\text{critique}}(\nu_1, \nu_2)$	rejeter H_0

- Si le test de Fisher rejette H_0 , on peut fusionner les deux écarts-types en un nouvel écart-type s^* :

$$s^* = \sqrt{\frac{\nu_1 \sigma_1^2 + \nu_2 \sigma_2^2}{\nu_1 + \nu_2}} \quad (112)$$

L'écart-type fusionné est alors connu avec $\nu = n_1 + n_2 - 2$ ddl.

Exemple 9 Deux élèves ont titré le même échantillon de sel. L'un A a réalisé 6 mesures et l'autre B a réalisé 8 mesures. On désire tester la précision de ces deux élèves lors de ce TP.

A	161	156	170	154	155	152		
B	163	169	155	158	159	151	157	163

Réponse :

Écart-type de A : $\sigma_A = 6,6030$ avec $\nu_A = 5$

Écart-type de B : $\sigma_B = 5,5533$ avec $\nu_B = 7$.

On a donc $\sigma_A > \sigma_B \Rightarrow F = \frac{6,6030}{5,5533} = 1,4138 > F_{\text{crit}}(5, 7) = 4,88$.

On accepte donc H_0 et l'on conclut qu'il n'y a pas de différence significative de précision entre ces deux élèves.

CLASSIFICATION

Déterminer le partitionnement en classes, ainsi que les critères d'appartenance d'un échantillon à la classe la plus probable, en partant uniquement d'une probabilité d'appartenance à une même classe (sans toutefois préjuger de laquelle), c'est-à-dire à partir d'une mesure de ressemblance entre échantillons.

On désigne aussi par regroupement ou clustering les méthodes d'inférence d'une classification.



Plus deux échantillons sont proches dans \mathcal{M} , plus leur probabilité d'appartenir à la même classe est grande.

Definition 4 Les classes/catégories obtenues sont appelées groupes/agrégats.

Definition 5 Une forme est représentée par un n -vecteur correspondant à la mesure de n caractéristiques.

On exige 2 propriétés pour la classification :

- **compacité** : les points représentant une classe donnée sont plus proche entre eux que des points de toutes les autres classes
- **séparabilité** : les classes sont bornées et il n'y a pas de recouvrement entre elles

Étant donné un n -échantillon $\{x_1, \dots, x_n\}$ dans un espace \mathcal{M} quelconque, ainsi qu'une **mesure de "dissemblance"** (*dissimilarity measure*)

$$\boxed{d(x, y) \geq 0, \quad \forall (x, y) \in \mathcal{M} \times \mathcal{M}} \quad (113)$$

On cherche un regroupement des échantillons en c classes t.q. $d(x, y)$ soit plus petite lorsque $K(x) = K(y)$ que lorsque $K(x) \neq K(y)$,

Les objets à classer sont des variables ou des individus

1. Données: tableau $n \times p$ individus \times variables
2. Objectif = recherche d'une typologie ou segmentation, c'est à dire d'une partition ou répartition des n individus dans des classes, sur la base de l'observation de p descripteurs
3. Moyen = chaque classe doit être la plus homogène possible et, entre elles, les plus distinctes possibles, au sens d'un critère à définir

La **classification** regroupe des objets en groupes/classes/familles/segments/clusters de sorte que

- 2 objets d'un même groupe se ressemblent le plus possible
- 2 objets de groupes distincts diffèrent le plus possible
- le nombre de groupe est parfois fixé

Mise en œuvre d'une classification :

- Choix de la **mesure d'éloignement** (dissimilarité, distance) entre individus (généralement distance euclidienne)
- Choix du **critère d'homogénéité** des classes à optimiser (généralement inertie).
- Choix de la **méthode utilisée** : la Classification Ascendante Hiérarchique (CAH) ou celle par ré-allocation dynamique sont les plus utilisées
- Mesure de la **qualité de la classification**
- Choix du nombre de classes et leur **interprétation**

Notion de distances

Definition 6 La distance est une fonction $d : E \times E \rightarrow \mathbb{R}$ définie sur un espace métrique E vérifiant les 4 propriétés :

- séparabilité : $\forall (a, b) \in E^2, a \neq b \rightarrow d(a, b) > 0$
- réflexivité : $\forall a \in E, d(a, a) = 0$
- symétrie : $\forall (a, b) \in E^2, d(a, b) = d(b, a)$
- inégalité triangulaire : $\forall (a, b, c) \in E^2, d(a, c) \leq d(a, b) + d(b, c).$

Exercice 12 Soit la fonction $d : F \times F \rightarrow \mathbb{R}$ tq $d(f, g) = \max_{x \in \Delta} |f(x) - g(x)|$, où F est l'ensemble des fonctions définies bornées sur Δ . Démontrer que d est une distance.

Exemple 10 (Exemples de distance entre vecteurs) $X = \{x_i\}$ et $Y = \{y_i\}, i = 1, \dots, N$

- Hamming : $d_1(X, Y) = \sum_{i=1}^N |x_i - y_i|$;
- Euclidienne : $d_2(X, Y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$;
- infini : $d_\infty(X, Y) = \max_i |x_i - y_i|$;
- Minkowski (q) : $d_q(X, Y) = (\sum_{i=1}^N |x_i - y_i|^q)^{\frac{1}{q}}$; ($q = 2$ est Euclidien) • le cosinus de l'angle entre 2 vecteurs: $s(X, Y) = \frac{X^T Y}{\|X\| \|Y\|}$ est invariant en rotation et en dilatation (pas pour une transformation linéaire quelconque).

Pour permettre d'attribuer un point X à une classe C_k selon,

$$\boxed{X \in C_k \Leftrightarrow C_k = \arg \min_i d(X, C_i)} \quad (114)$$

Plus généralement, comment définir le cluster le plus proche ?

$$d_{\min}(X, C) = \min_{Y \in C} \|X - Y\|, \quad (115)$$

$$d_{\max}(C_1, C_2) = \min_{X \in C_1, Y \in C_2} \|X - Y\| \quad (116)$$

$$d_{\text{avg}}(C_1, C_2) = \frac{1}{n_1 n_2} \sum_{X \in C_1} \sum_{Y \in C_2} \|X - Y\| \quad (117)$$

$$d_{\text{moy}}(C_1, C_2) = \|M_1 - M_2\|, \quad M_1 = \frac{1}{n_1} \sum_{X \in C_1}, M_2 = \frac{1}{n_2} \sum_{X \in C_2} \quad (118)$$

Mesure d'éloignement

Le choix d'une mesure dépend de la nature des descripteurs. Le plus souvent, le choix se porte sur une distance euclidienne (i.e. engendrée par un produit scalaire) appropriée :

$$d_1(\mathbf{x}, \mathbf{y}) = \|x - y\|_M^2 = \mathbf{x}^T M \mathbf{y}$$

Pour des Variables quantitatives : M de produit scalaire sur l'espace des variables:

(1) $M = I_p$

(2) $M = D_\sigma^{-1}$

(3) $M = V^{-1}$

(1) variables homogènes, (2) hétérogènes, (3) corrélées

Critère d'homogénéité Si l'ensemble des points est regroupé en K classes

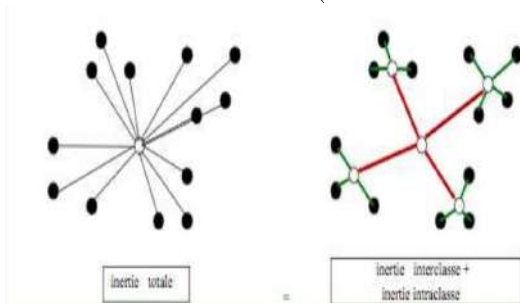
$$f_k \sim \mathcal{N}(\mu_k, P_k)$$

1. Inertie intra-classes : $I_W = \sum_{k=1}^K I_k = \sum_{k=1}^K \sum_{e_i \in f_k} p_i \|x_i - \mu_k\|^2$

2. Inertie inter-classes : $I_B = \sum_{k=1}^K P_k \|\mu_k - \mu\|^2$

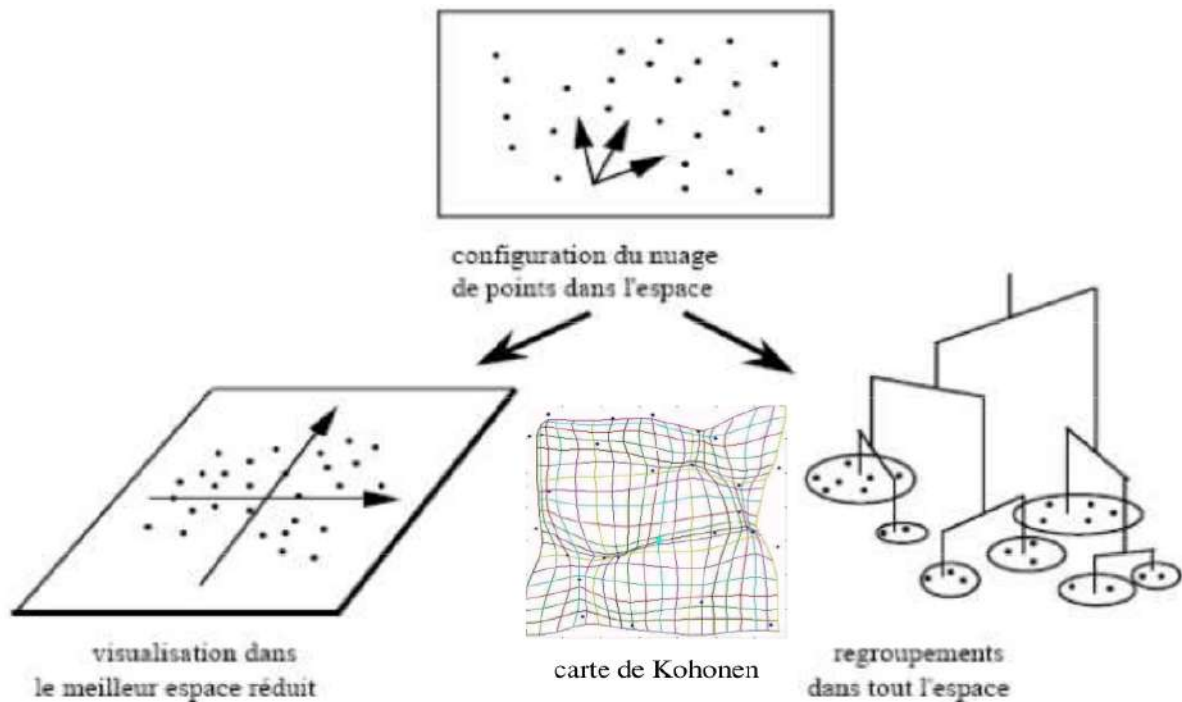
et $I = I_W + I_B$ (formule de Huygens).

Critère le plus utilisé (variables quantitatives) = $\min_{N_k} I_W$ pour K fixé.



- une classe est homogène \Leftrightarrow son inertie est faible
- 2 critères de bonne classification : grande I_B , petite I_W

LES MÉTHODES



Source : Lebart-Morineau-Piron, *Statistique exploratoire multidimensionnelle*, page 10

- Méthodes **hiérarchiques**

1. ascendantes (agglomératives) (basées sur une notion de distance ou de densité)
2. descendantes (divisives)

- Méthodes de **partitionnement**

1. centres mobiles, k-means et nuées dynamiques
2. k-modes, k-prototypes, k-représentants (k-medoids)
3. réseaux de Kohonen
4. méthodes basées sur une notion de densité

- Méthodes mixtes

- Analyse floue (fuzzy clustering)

Classification \Rightarrow Homogénéité des groupes \Rightarrow – de variables discriminantes
 \Rightarrow + grande fiabilité des prédictions

Critères de bon algorithme

1. Détecter les structures présentes dans les données
2. Permettre de déterminer le nombre optimal de classes
3. Fournir des classes bien différenciées
4. Fournir des classes stables vis-à-vis de légères modifications des données
5. Traiter efficacement les grands volumes de données
6. Traiter tous les types de variables (quantitatives et qualitatives) → rarement obtenu sans transformation
7. Mais pas de critère universel de qualité semblable à l'aire sous la courbe ROC des méthodes de scoring

Quand on a des variables qualitatives → se ramener à une classification de variables continues par une ACM

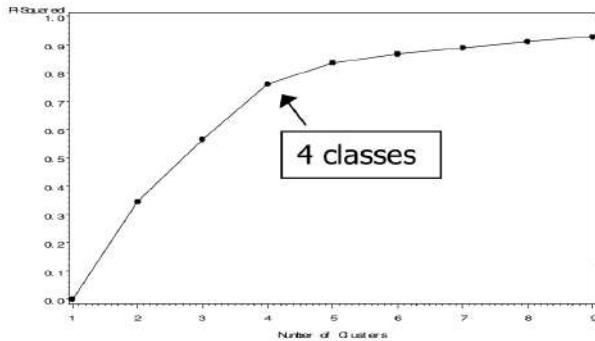
Performances d'un classifieur

- il faut des données (en qualité et en quantité) pour apprendre
 - il faut des données "inconnues" pour évaluer le classifieur
 - mesures globales: taux de reconnaissance, de confusion, de rejet
 - mesures locales: matrices de confusion
- bon classifieur = bonne généralisation
- il faudrait théoriquement 3 jeux de données différents
- apprentissage, test, évaluation
- rapidité dans la prise de décision
- mieux vaut un algo d'apprentissage lent et un algo de décision rapide que l'inverse!!!

Mesures de qualité d'une classification

$$R^2 = \frac{I_B}{I_B + I_W} \text{ proportion de la variance expliquée par les classes}$$

- être le plus proche possible de 1 sans avoir trop de classes
- s'arrêter après le dernier saut important



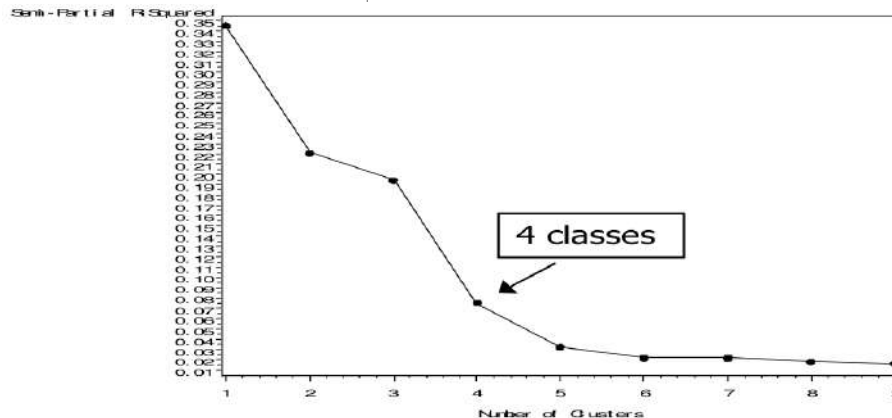
$$F = \frac{\frac{R^2}{K-1}}{\frac{1-R^2}{n-K}} \text{ mesure la séparation entre toutes les classes}$$

(statistique de Fisher)

- rechercher une grande valeur
- n nb d'observations, K nb de classes

R^2 semi-partiel (SPRSQ) = mesure la perte d'inertie interclasse provoquée en regroupant 2 classes

But = avoir une inertie interclasse maximum, on recherche un faible SPRSQ suivi d'un fort SPRSQ à l'agrégation suivante un pic pour k classes et un creux pour $k+1$ classes indique une bonne classification en $k+1$ classes



Distances/variables binaires

Souvent, les traits caractéristiques des formes ne sont pas mesurables \Rightarrow codage qui consiste à coder la composante x_{ik} de la représentation X_i par 1 si la forme possède l'attribut k et 0 sinon.

On calcule le nombre de fois où :

- X_i et X_j possèdent le même caractère (couples de 11) $a = \sum_{k=1}^N x_{ik}x_{jk}$;
- X_i et X_j ne possèdent aucun caractère commun (couples de 00)
 $b = \sum_{k=1}^N (1 - x_{ik})(1 - x_{jk})$;
- X_i ne possède pas le caractère possédé par X_j (couples de 01)
 $h = \sum_{k=1}^N (1 - x_{ik})x_{jk}$;
- X_i possède un caractère non possédé par X_j (couples de 10)
 $g = \sum_{k=1}^N x_{ik}(1 - x_{jk})$;

Plus a est grand plus la ressemblance est grande.

Exemple 11 Soit une classification de légumes (z'avez rien contre ?!)

forme	mesures	rond	allongé	rouge	vert
chou	X_1	1	0	0	1
concombre	X_2	0	1	0	1
tomate	X_3	1	0	1	0

On peut construire une **fonction de similarité** \nearrow avec a , symétrique en g et h , \searrow avec b appelée distance binaire. Par exemple :

$$\begin{aligned}
 S_1(X_i, X_j) &= \frac{a}{a + b + g + h} && \text{(Russel et Rao)} \\
 S_2(X_i, X_j) &= \frac{a}{n - b} && \text{(Joccard et Needham)} \\
 S_3(X_i, X_j) &= \frac{a}{2a + g + h} && \text{(Dice)} \\
 S_4(X_i, X_j) &= \frac{a}{a + 2(g + h)} && \text{(Sokal et Sneath)} \\
 S_5(X_i, X_j) &= \frac{a + b}{n} && \text{(Sokal et Michenon)}
 \end{aligned}$$

Exercice 13 Comparer X_1 et X_3 . Peut on dire que X_1 et X_2 se ressemblent plus que X_1 et X_3 ?

Distances entre listes (Kendall)

Supposons que la forme à reconnaître soit représentée par une **liste** de mesures. Comment mesurer la ressemblance entre 2 listes $X = \{x_i\}$ et $Y = \{y_i\}$?

Le facteur de comparaison pour la liste X est défini par

$$\Delta_{rk}^X = \begin{cases} +1 & \text{si } x_r > x_k \\ -1 & \text{si } x_r < x_k \\ 0 & \text{si } x_r = x_k \text{ avec } r < k \leq N \end{cases} \quad (119)$$

et permet de calculer la distance de Kendall:

$$d(X, Y) = 1 - \frac{2}{N(N-1)} \sum_{r < k} \Delta_{rk}^X \Delta_{rk}^Y.$$

Si les 2 listes sont toutes 2 strictement \nearrow ou \searrow , alors la distance de Kendall entre elles est nulle.

Exercice 14 1. Soient $X = \{-1, 2, 3\}$ et $Y = \{3, 5, 8\}$. Calculer les Δ_{ij}^X et Δ_{ij}^Y , puis $d(X, Y)$.

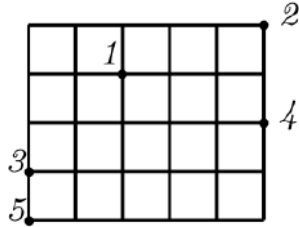
2. Soient maintenant $X = \{7, 4, 0\}$ et $Y = \{15, 20, 21\}$. Calculer $d(X, Y)$.

Remarque : Pour le code des variables binaires. Coder 0/1 une variable symétrique, c'est pareil (Exemple: le sexe d'une personne), mais si la variable est asymétrique on code par 1 la modalité la moins fréquente (Test HIV. Le test peut être positif ou négatif (0 ou 1) mais il y a une valeur qui sera plus présente que l'autre.)

Groupements hiérarchiques

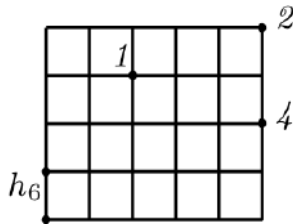
- Méthodes **divisives** : partent du n -échantillon et le subdivisent récursivement en (sous-)groupes plus spécifiques.
- Méthodes **agglomératives** procèdent dans le sens ascendant, par le regroupement des échantillons individuels en petits agrégats d'abord, fusionnés à leur tour dans les niveaux, etc.

Exemple 12 Soit un ensemble de formes représentées par des points (numérotés de 1 à 5) dans un repère euclidien.



d	1	2	3	4	5
1	0	$\sqrt{10}$	$\sqrt{8}$	$\sqrt{10}$	$\sqrt{13}$
2		0	$\sqrt{34}$	2	$\sqrt{41}$
3			0	$\sqrt{26}$	$\sqrt{41}$
4				0	1
5					0

\Rightarrow on regroupe 3 et 5 : $h_6 = \{3, 5\}$, d'indice d'agrégation $f(h_6) = 1$.



\Rightarrow Problème : déterminer la distance de h_6 aux formes restantes. 2 solutions :

saut minimal (single linkage)

diamètre maximal (complete linkage)

d	1	2	4	h_6
1	0	$\sqrt{10}$	$\sqrt{10}$	$\sqrt{8}$
2		0	2	$\sqrt{34}$
4			0	$\sqrt{26}$
h_6				0

d	1	2	4	h_6
1	0	$\sqrt{10}$	$\sqrt{10}$	$\sqrt{13}$
2		0	2	$\sqrt{41}$
4			0	$\sqrt{29}$
h_6				0

saut minimal : affecte à la distance entre 2 groupes la distance entre leurs formes les plus proches.

diamètre maximal : distance entre les formes les plus éloignées.

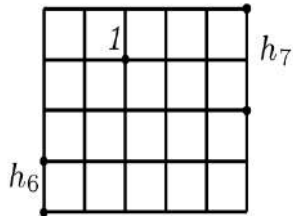
A l'étape suivante, on a la même agrégation pour les 2 distances, $h_7 = \{2, 4\}$, $f(h_7) = 2$.

saut minimal

d	1	h_6	h_7
1	0	$\sqrt{8}$	$\sqrt{10}$
h_6		0	$\sqrt{26}$
h_7			0

diamètre maximal

d	1	h_6	h_7
1	0	$\sqrt{13}$	$\sqrt{10}$
h_6		0	$\sqrt{41}$
h_7			0



A la troisième étape, les hiérarchies diffèrent. A la dernière étape, toutes les formes sont regroupées.

saut minimal

$$h_9 = h_7 \cup h_8 = \{1, 2, 3, 4, 5\}$$

$$f(h_9) = \sqrt{10}$$

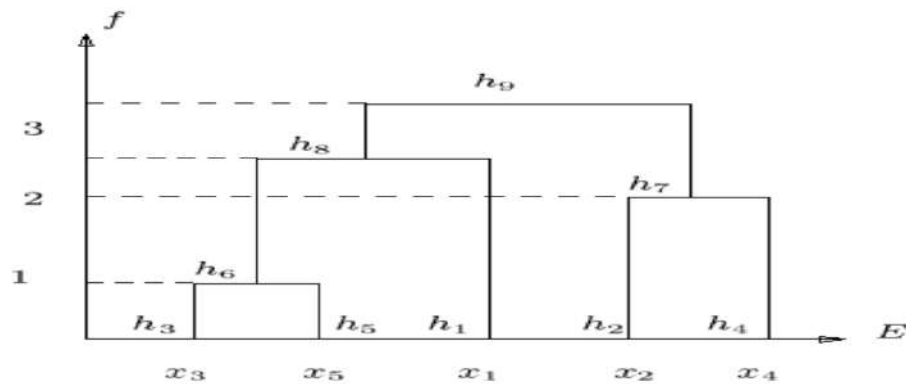
diamètre maximal

$$h_{9'} = h_6 \cup h_{8'} = \{1, 2, 3, 4, 5\}$$

$$f(h_{9'}) = \sqrt{41}$$

Le choix de la distance influe sur les regroupements

Figure 2: Représentation arborescente d'une hiérarchie (obtenue par le saut minimal)



Critère de Ward (baisse d'inertie interclasse résultant de la fusion des 2 classes)

- la distance de 2 classes A et B , de barycentres \mathbf{a} et \mathbf{b} , et d'effectifs n_A et n_B , vaut :

$$d(A, B) = \frac{d(\mathbf{a}, \mathbf{b})}{\frac{1}{n_A} + \frac{1}{n_B}}$$

- fonction de la distance des barycentres
- produire des classes sphériques de mêmes effectifs — peu efficace sur classes allongées
- très sensible aux outliers

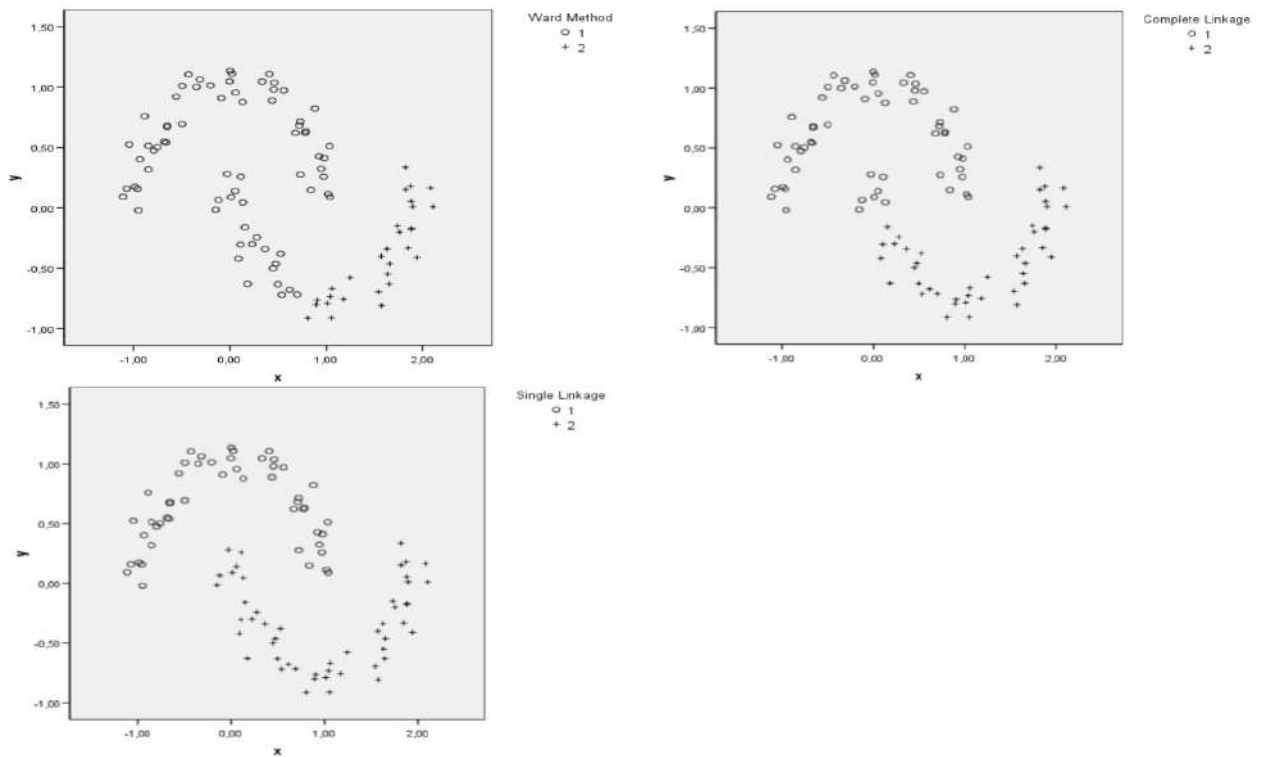
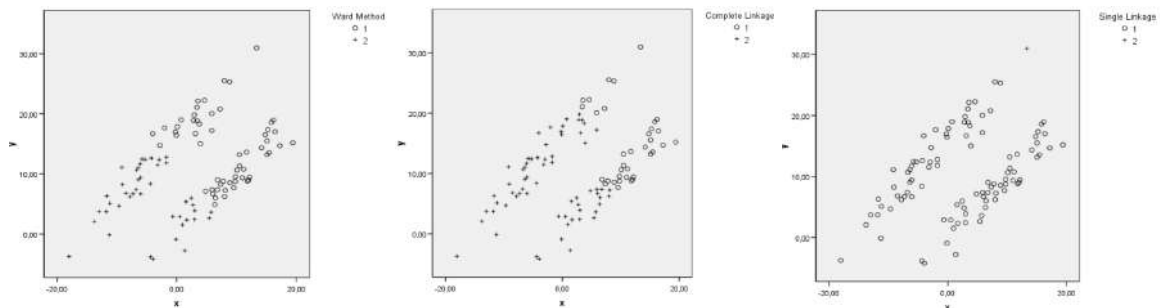

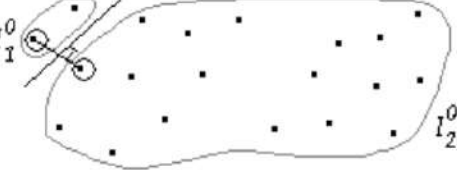
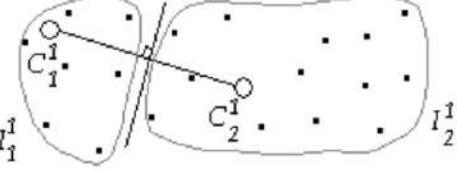
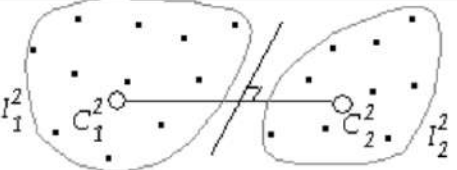


Figure 3: Illustration de la CAH - Données allongées.



Méthode des centres mobiles

- On choisit k individus comme centres initiaux des classes
- On calcule les distances entre chaque individu et chaque centre c_i de l'étape précédente, et on affecte chaque individu au centre le plus proche, ce qui définit k classes
- On remplace les k centres c_i par les barycentres des k classes définies à l'étape 2
- On regarde si les centres sont restés suffisamment stables ou si un nombre fixé d'itérations a été atteint :
 - si oui, on arrête (en général, après au moins une dizaine d'itérations)
 - si non, on revient à l'étape 2

	<p>Tirage au hasard des centres C_1^0 et C_2^0</p>
	<p>Constitution des classes I_1^0 et I_2^0</p>
	<p>Nouveaux centres C_1^1 et C_2^1 et nouvelles classes I_1^1 et I_2^1</p>
	<p>Nouveaux centres C_1^2 et C_2^2 et nouvelles classes I_1^2 et I_2^2</p>

Source : Ludovic Lebart – *Analyse des données appliquée* - 2002

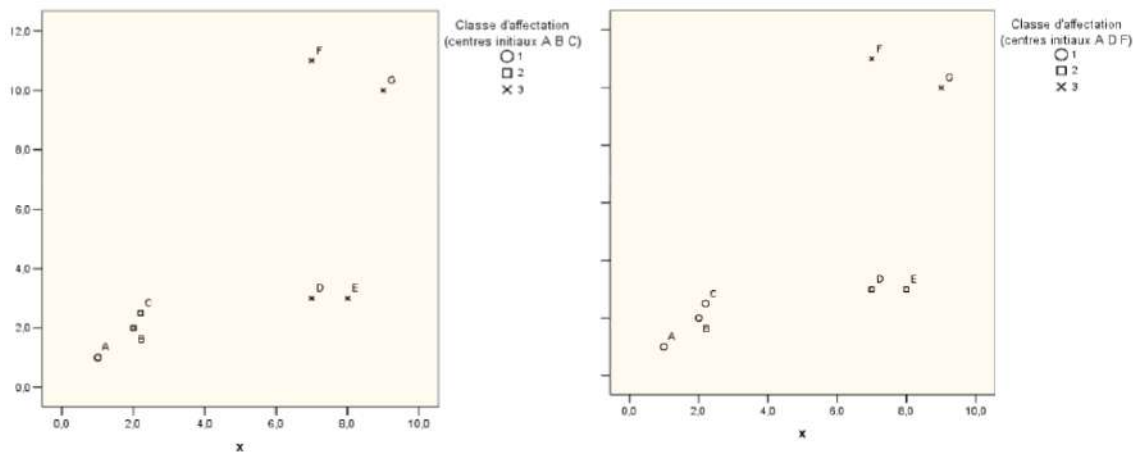
- **k -means** (Mc Queen): les barycentres des classes ne sont pas recalculés à la fin des affectations, mais à la fin de chaque allocation d'un individu à une classe. L'algorithme est ainsi plus rapide, mais l'ordre d'apparition des individus dans le fichier n'est pas neutre.
- **nuées dynamiques** (Diday): ce n'est plus un seul point qui représente une classe mais un noyau de points constitués d'éléments représentatifs de la classe. Cela permet de corriger l'influence d'éventuelles valeurs extrêmes sur le calcul du barycentre.

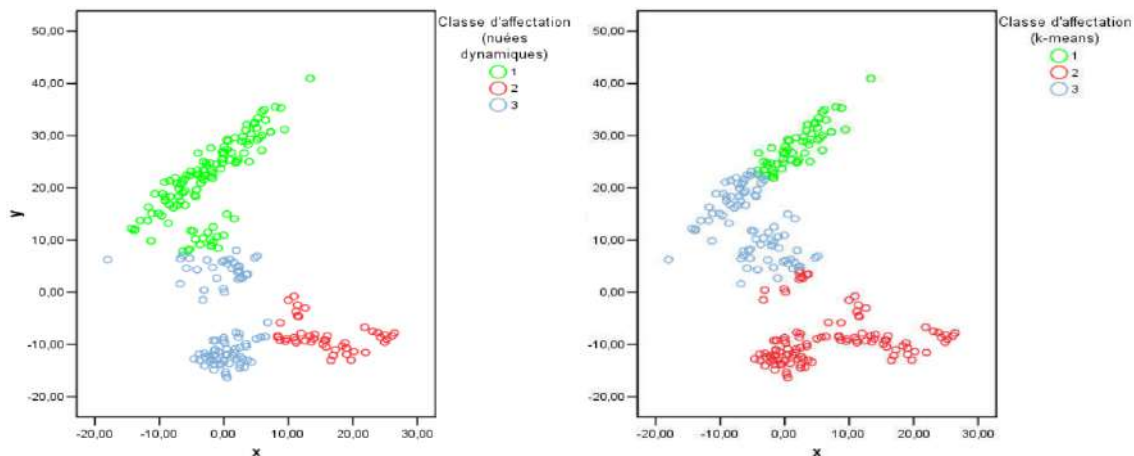
PROS :

- le nombre d'itérations nécessaires est très faible
- on ne compare pas toutes les observations entre elles
- l'algorithme converge

CONS

- Instabilité : Le minimum obtenu est un minimum local (dépend du choix initial des centres)
- Le nombre de classes est fixé par avance





Méthodes mixtes Combiner les avantages de CAH et des k-means

- Combiner les avantages de CAH et des k-means
- effectuer une 1ère classification sur les n observations par les centres mobiles ou les k-means, en fixant le nombre de classes entre 10 et 100
- valeur assez grande pour limiter le risque de fusion de classes naturelles
- puis effectuer la CAH sur les centres de ces pré-classes
- On peut aussi quantifier la **force discriminante** de chaque variable sur la classification dans son ensemble, afin de déterminer quelles sont les caractéristiques expliquant le plus les différenciations construites.
- En fin de course, il est fréquent de **nommer** chacune des classes obtenues par un qualificatif résumant la caractérisation.
- Pour illustrer ce qui caractérise une classe, on peut aussi rechercher **l'individu le plus typique** (ou central) de la classe, ou bien encore un noyau d'individus la représentant bien.

Exercice 15 (*k-means*) $A = \{1, 2, 3, 6, 7, 8, 13, 15, 17\}$. Créer 3 clusters à partir de A .

Solution:

On prend 3 objets au hasard. Supposons que c'est 1, 2 et 3. Ça donne $C_1 = \{1\}$, $M_1 = 1$, $C_2 = \{2\}$, $M_2 = 2$, $C_3 = \{3\}$ et $M_3 = 3$.

Chaque objet O est affecté au cluster au milieu duquel, O est le plus proche. 6 est affecté à C_3 car $d(M_3, 6) < d(M_2, 6)$ et $d(M_3, 6) < d(M_1, 6)$. On a :

$$\begin{array}{ll} C_1 = \{1\} & M_1 = 1 \\ C_2 = \{2\} & M_2 = 2 \\ C_3 = \{3, 6, 7, 8, 13, 15, 17\} & M_3 = 69/7 = 9.86 \end{array}$$

- $d(3, M_2) < d(3, M_3) \rightarrow 3$ passe dans C_2 . Tous les autres objets ne bougent pas. $C_1 = \{1\}$, $M_1 = 1$, $C_2 = \{2, 3\}$, $M_2 = 2.5$, $C_3 = \{6, 7, 8, 13, 15, 17\}$ et $M_3 = 66/6 = 11$
- $d(6, M_2) < d(6, M_3) \rightarrow 6$ passe dans C_2 . Tous les autres objets ne bougent pas. $C_1 = \{1\}$, $M_1 = 1$, $C_2 = \{2, 3, 6\}$, $M_2 = 11/3 = 3.67$, $C_3 = \{7, 8, 13, 15, 17\}$, $M_3 = 12$
- $d(2, M_1) < d(2, M_2) \rightarrow 2$ passe en C_1 . $d(7, M_2) < d(7, M_3) \rightarrow 7$ passe en C_2 . Les autres ne bougent pas. $C_1 = \{1, 2\}$, $M_1 = 1.5$, $C_2 = \{3, 6, 7\}$, $M_2 = 5.34$, $C_3 = \{8, 13, 15, 17\}$, $M_3 = 13.25$
- $d(3, M_1) < d(3, M_2) \rightarrow 3$ passe en C_1 . $d(8, M_2) < d(8, M_3) \rightarrow 8$ passe en C_2 . $C_1 = \{1, 2, 3\}$, $M_1 = 2$, $C_2 = \{6, 7, 8\}$, $M_2 = 7$, $C_3 = \{13, 15, 17\}$, $M_3 = 15$

Méthode des K-Medoids

- Trouver des objets représentatifs (medoïdes) dans les clusters (au lieu de la moyenne)
- Principe : Commencer avec un ensemble de medoïdes puis itérativement remplacer un par un autre si ça permet de réduire la distance globale (Efficace pour des données de petite taille)

Algo. des K-Medoids

Choisir arbitrairement k objets représentatifs

$\forall (h, j)$ d'objets t.q h est choisi et j non,

calculer le coût TC_{jh} du remplacement de j par h

Si $TC_{jh} < 0$, j est remplacé par h

avec

assigner chaque objet non sélectionné au medoïde qui lui est le plus similaire

jusqu'à convergence

$$TC_{jh} = \sum_i d(j, h) - d(j, i).$$

Exercice 16 (*k-means*) $A = \{1, 2, 3, 6, 7, 8, 13, 15, 17\}$. Créer 3 clusters à partir de A .

Méthode des K-Medoids

- Trouver des objets représentatifs (medoïdes) dans les clusters (au lieu de la moyenne)
- Principe : Commencer avec un ensemble de medoïdes puis itérativement remplacer un par un autre si ça permet de réduire la distance globale (Efficace pour des données de petite taille)

Algo. des K-Medoids

Choisir arbitrairement k objets représentatifs

$\forall (h, j)$ d'objets t.q h est choisi et j non,

calculer le coût TC_{jh} du remplacement de j par h

Si $TC_{jh} < 0$, j est remplacé par h

avec

affecter chaque objet non sélectionné au medoïde qui lui est le plus similaire

jusqu'à convergence

$$TC_{jh} = \sum_i d(j, h) - d(j, i).$$

Exercice 17 (*k-means*) $A = \{1, 2, 3, 6, 7, 8, 13, 15, 17\}$. Créer 3 clusters à partir de A .

Solution:

- Soit $A = \{1, 3, 4, 5, 8, 9\}$, $k = 2$ et $M = \{1, 8\}$ ensemble des medoïdes
→ $C_1 = \{1, 3, 4\}$ et $C_2 = \{5, 8, 9\}$
 $TC_{1,8} = d(3, 1)^2 + d(4, 1)^2 + d(5, 8)^2 + d(5, 9)^2 + d(9, 8)^2 = 39$
- Comparons 1 et 3 → $M = \{3, 8\}$ → $C_1 = \{1, 3, 4, 5\}$ et $C_2 = \{8, 9\}$
 $E_{3,8} = d(1, 3)^2 + d(4, 3)^2 + d(5, 3)^2 + d(9, 8)^2 = 10$
 $E_{3,8} - E_{1,8} = -29 < 0$ donc le remplacement est fait.
- Comparons 3 et 4 → $M = \{4, 8\}$ → C_1 et C_2 inchangés et $E_{4,8} = d(1, 4)^2 + d(3, 4)^2 + d(5, 4)^2 + d(8, 9)^2 = 12$ → 3 n'est pas remplacé par 4
- Comparons 3 et 5 → $M = \{5, 8\}$ → C_1 et C_2 inchangés et $E_{5,8} > E_{3,8}$

Le réseau de Kohonen

La réduction de la complexité est obtenue en diminuant le nombre n d'exemples ou le nombre P de caractéristiques

$$D^{n \times P} \longrightarrow D^{n' \times P'} \begin{cases} n' < n \\ P' < P \end{cases}$$

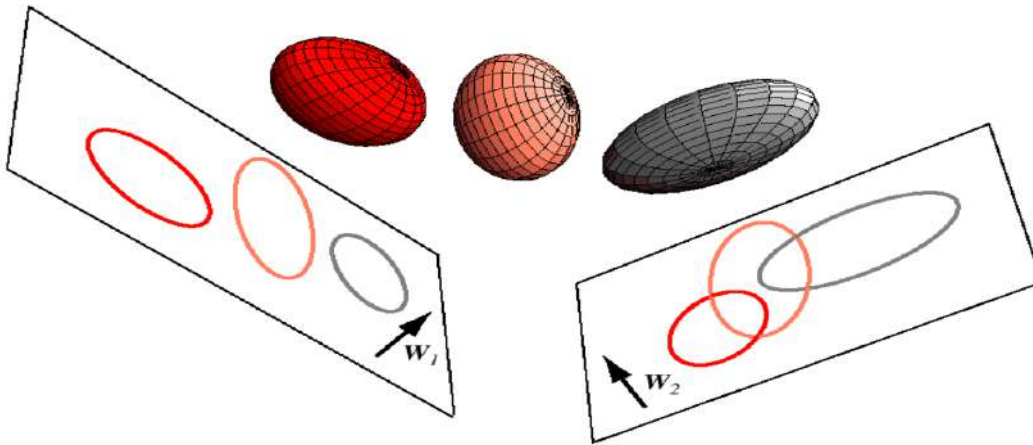
- par **quantification vectorielle** passer de n exemples à n' représentant, comme la méthode de l'apprentissage compétitif (Martinetz & Schulten, 1991).
- par **projection** de $R^P \rightarrow \mathbb{R}^{P'}$, comme la méthode d'Analyse en Composantes Principales (ACP), mais aussi l'Analyse en Composantes Indépendantes (ACI).

But : Minimiser l'inertie intraclasse

$$\hat{D}_0(\mathbf{f}; C_1, C_2, \dots, C_K) = \frac{1}{K} \sum_{i=1}^K \sum_{j \in A_i} \|\mathbf{x}_j - C_i\|^2. \quad (120)$$

Sur des données quantitatives/**qualitatives** :

- représentation/visualisation (vecteurs codes, voisinages de classes, grille, ficelle, cylindre, etc.)
- représentation des contenus des classes, listes, courbes,
- classes et super-classes, visualisation des séparation entre les classes
- analyse des tableaux de contingence
- tables de Burt, croisement de plusieurs variables qualitatives

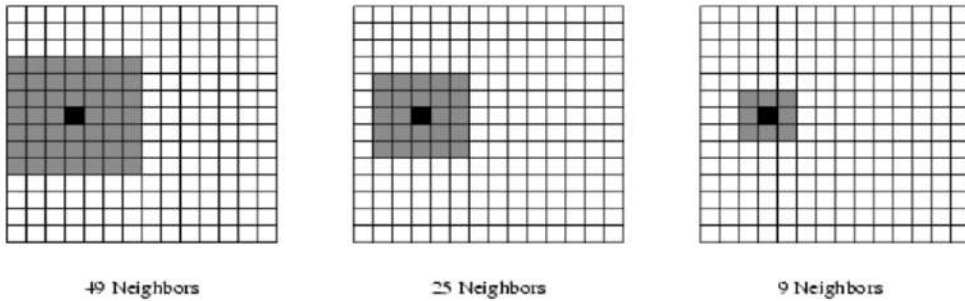


L'algorithme

Algorithme de classification (T. Kohonen, 80) qui regroupe les observations en classes en respectant la topologie de l'espace des observations.

- Des variables (dans \mathbb{R}^p) appartiennent (après classement) à la même classe ou à des classes voisines.

Les classes sont disposées sur une grille rectangulaire qui définit naturellement les voisins de chaque classe. Mais on peut choisir une autre topologie...



- algo. itératif.

L'initialisation consiste à associer à chaque classe un vecteur code de P dimensions choisi de manière aléatoire. Ensuite, à chaque étape, on choisit une observation au hasard (loi $\mathcal{U}_{[1,n]}$), on la compare à tous les vecteurs codes, et on détermine la classe gagnante, c'est-à-dire celle dont le vecteur code est le plus proche au sens d'une distance donnée *a priori*.

On rapproche alors de l'observation les codes de la classe gagnante et des classes voisines.

$$\mathbf{c}_i(t+1) = \begin{cases} \mathbf{c}_i(t) + \epsilon(t+1)\sigma(i_0(t+1), i)(\mathbf{x}_{t+1} - \mathbf{c}_i(t)) & \text{si } i \in \mathcal{V}(i, \ell(t)), \\ \mathbf{c}_i(t) & \text{sinon.} \end{cases} \quad (121)$$

où¹

- \mathbf{x}_{t+1} est l'observation présentée à l'étape $(t+1)$,
- \mathbf{c}_i est le vecteur code associé à la classe i ,
- ϵ un paramètre d'adaptation > 0 et < 1 .
- analogue à l'algorithme des nuées dynamiques
- En résumé,

Algo. Kohonen

Pour tout le corpus d'apprentissage

présenter d'une entrée \mathbf{x} au hasard

chercher l'unité gagnante $i_0(\mathbf{x})$

rapprocher C_{i_0} et les C_i voisins de l'entrée \mathbf{x}

jusqu'à convergence

¹ $\mathcal{V}(i, \ell(t))$ est le voisinage de l'unité i de rayon $\ell(t)$.

- sélection unité gagnante \mathbf{c}_{i_0} la plus proche de \mathbf{x}_k

$$\boxed{\exists i_0, \|\mathbf{c}_{i_0} - \mathbf{x}_k\| \leq \|\mathbf{c}_i - \mathbf{x}_k\|, \forall i}$$

- et des unités voisines \mathbf{c}_v de l'unité gagnante

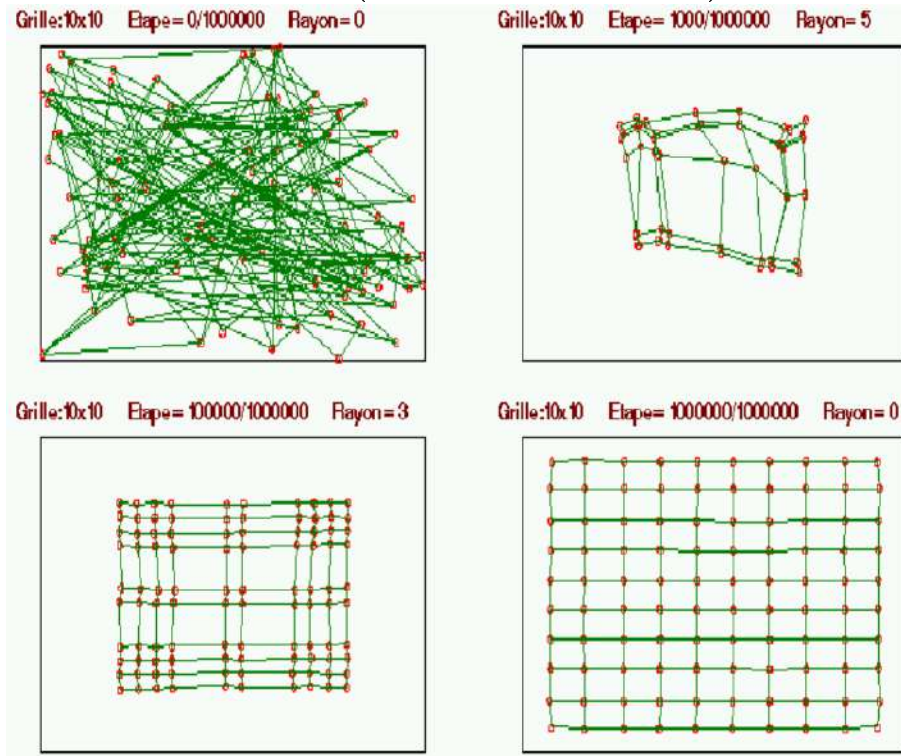
$$\boxed{\mathbf{c}_i = \mathbf{c}_i + \epsilon(\mathbf{x}_k - \mathbf{c}_i)}$$

- ϵ suit une loi décroissante

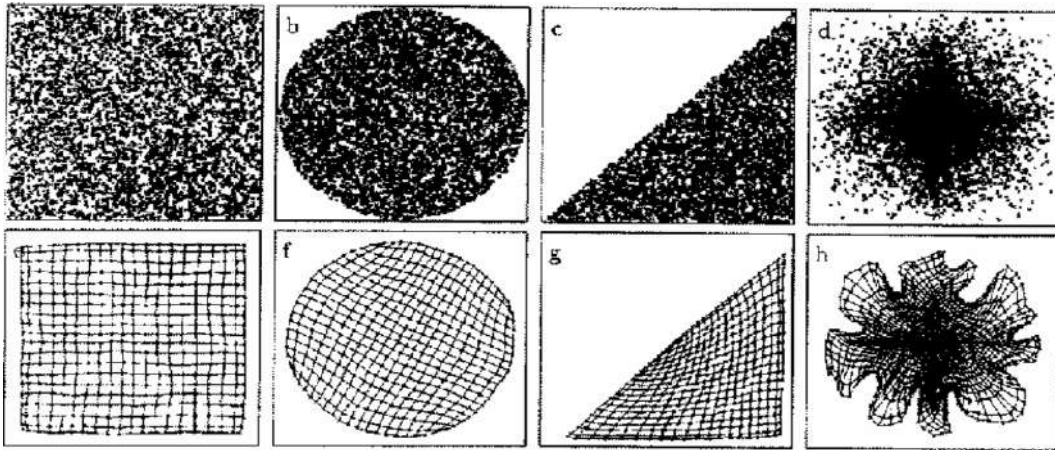
$$\epsilon(t) = \epsilon_i \left(\frac{\epsilon_f}{\epsilon_i} \right)^{t/t_{\max}} \quad (122)$$

- $\sigma(\cdot, \cdot)$: fonction de voisinage $\sigma(i, j) = \begin{cases} 1 & \text{ssi } i \text{ et } j \text{ sont voisins,} \\ 0 & \text{sinon} \end{cases}$
- des voisinages Gaussiens adoucissent l'auto-organisation.

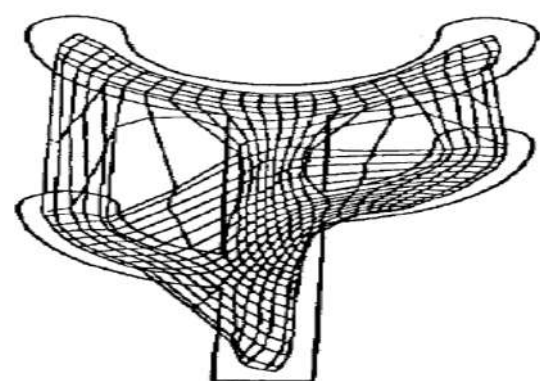
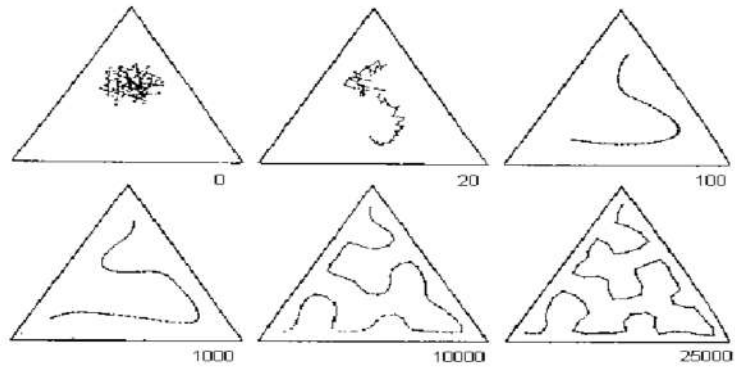
Exemple 13 (Démonstration en dimension 2)



Exemple 14 (Quantification de distributions par l'algorithme de Kohonen)



Exemple 15 (Exemples de simulations (Kohonen) [?])



Exemple 16 (En 3D (Kohonen) [?])

Auto-organisation

- **Point de vue biologique** : à toute zone du corps humain (la main par exemple) correspond une zone dans le cortex sensoriel. On peut ainsi dresser une carte du cortex sensoriel dans laquelle les “voisinages” sont préservés.

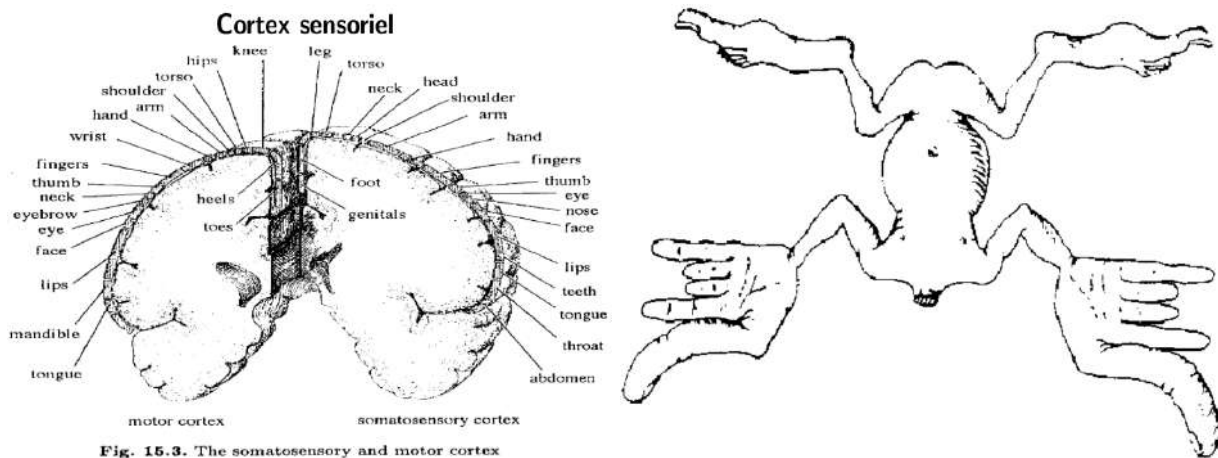
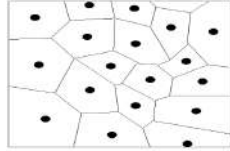


Figure 4: Cortex sensoriel et Homunculus (Anderson, Penfield and Boldrey)

Algorithmes de quantification vectorielle

Definition 7 Tout algorithme ayant pour objectif de réduire le nombre d'exemples ou de quantifier la distribution en construisant une famille réduite de représentants.



Pavage de Voronoï

Dans l'espace des entrées, les classes forment une partition, ou mosaïque de Voronoï, dépendant des \mathbf{c}_i .

2 exemples d'algorithmes :

	Winner take all	Winner take Most
1	initialisation des I représentants $\mathbf{c}_i, 1 \leq i \leq I$	1 choix d'une distance $g(a, b)$ dans \mathbb{R}^p
2	choix d'un \mathbf{x}_k parmi n exemples	2 initialisation des I représentants $\mathbf{c}_i, 1 \leq i \leq I$
3	sélection de l'unité gagnante \mathbf{c}_{i_0}	3 choix d'un \mathbf{x}_k parmi n exemples
4	loi d'adaptation pour la seule unité gagnante $\Delta \mathbf{c}_{i_0} = \alpha(\mathbf{x}_k - \mathbf{c}_{i_0})$	4 sélection de l'unité gagnante \mathbf{c}_{i_0}
5	test d'arrêt sinon retour en 2.	5 loi d'adaptation de toutes les unités \mathbf{c}_i en fonction du gain $g(\mathbf{c}_{i_0}, \mathbf{x}_k)$ de l'unité gagnante $\Delta \mathbf{c}_i = \alpha g(-\mathbf{c}_i, \mathbf{x}_k)(\mathbf{x}_k - \mathbf{c}_i), \forall i, 1 \leq i \leq I$
		6 test d'arrêt sinon retour en 3.

Classification par carte de Kohonen

- Pour représenter des données au moyen de l'algorithme de Kohonen, on prend comme entrées les lignes de la matrice des données
- Après apprentissage, chaque individu (ligne) correspond à une unité du réseau (celle qui gagne quand on présente cet individu)
- On classe une observation dans la classe A_i définie par l'unité gagnante qui lui correspond ($i = i_0(\mathbf{x})$)
- On obtient donc une classification des individus, avec respect des voisinages
- La carte ainsi obtenue fournit une représentation plane
- les proximités entre classes qui se ressemblent est essentielle

Pour chaque classe, on peut

- représenter le vecteur code à p composantes en dessinant une courbe à p points
- faire la liste des observations de cette classe
- représenter en superposition les observations de la classe
- ceci fournit une représentation plane, analogue à l'analyse en composantes principales (mais une seule carte et pas de projection orthogonale)

Exemples d'application

1. Représentation des pays, (Blayo et Letremy)
2. Communes d'Ile-de France, (Ibbou, Tutin)
3. Courbes de consommation, prévision, (Rousset)
4. Consommation au Canada, (Gaubert, Gardes, Rousset)
5. Classification de dessins d'enfants (Vigneron et Streri)
6. Dépenses de formations en entreprise (Vigneron et Petit)

Exemple 17 (Recensements de 1783 communes de la vallée du Rhône)

- Chiffres de population aux recensements de 1936, 1954, 1962, 1968, 1975, 1982, 1990.
- Communes de la vallée du Rhône (dans le cadre d'une étude du rapport entre l'évolution démographique et la répartition de la population entre différentes catégories socio-professionnelles).
- Ardèche, Bouches-du-Rhône, Drôme, Gard, Hérault, Isère, Haute-Loire, Vaucluse
- Les chiffres sont transformés en pourcentages en divisant par la somme de tous les recensements (pour supprimer l'effet taille)
- La distance utilisée est la distance du χ^2
- 10.000 itérations, une grille 8×8 , 5 super-classes (82% d'inertie)

CODE	NOM	POUR36	POUR54	POUR62	POUR68	POUR75	POUR82	POUR90
07001	ACCONS	0.178	0.147	0.137	0.113	0.118	0.154	0.153
07002	ALHON	0.196	0.146	0.117	0.103	0.098	0.127	0.213
07003	ALZAC	0.210	0.170	0.150	0.135	0.105	0.110	0.120
07004	AJOUX	0.304	0.173	0.143	0.123	0.097	0.085	0.076
07005	ALBA-LA-ROMAINE	0.158	0.122	0.131	0.145	0.144	0.137	0.164
07006	ALBON	0.257	0.190	0.157	0.127	0.107	0.083	0.079
07007	ALBOUSIERE	0.164	0.150	0.144	0.139	0.138	0.120	0.144
07008	ALISSAS	0.125	0.121	0.147	0.133	0.125	0.161	0.188
07009	ANDANCE	0.137	0.145	0.135	0.139	0.155	0.138	0.151
07010	ANNONAY	0.121	0.125	0.142	0.160	0.160	0.150	0.143
07011	ANTRAGUES	0.202	0.149	0.134	0.133	0.125	0.131	0.127
07012	ARCENS	0.187	0.146	0.152	0.145	0.128	0.122	0.120
07013	ARDOIX	0.151	0.139	0.122	0.130	0.141	0.152	0.166
07014	ARLEBOSC	0.199	0.159	0.149	0.143	0.129	0.113	0.108
07015	ARRAS-SUR-RHONE	0.147	0.158	0.155	0.155	0.128	0.122	0.135
07016	ASPERJOC	0.197	0.162	0.145	0.130	0.130	0.117	0.119
07017	ASSIONS	0.175	0.136	0.152	0.140	0.123	0.135	0.138
07018	ASTET	0.277	0.215	0.152	0.127	0.100	0.067	0.062
07019	AUBENAS	0.112	0.121	0.129	0.151	0.169	0.162	0.156
07020	AUBIGNAS	0.209	0.114	0.140	0.113	0.125	0.133	0.167
07022	BAIX	0.135	0.128	0.135	0.128	0.115	0.202	0.157
07023	BALAZUC	0.215	0.154	0.127	0.112	0.109	0.141	0.142
07024	BANNE	0.194	0.157	0.134	0.118	0.122	0.134	0.140
07025	BARNAS	0.262	0.175	0.153	0.133	0.095	0.089	0.093
07026	BEAGE	0.234	0.183	0.149	0.142	0.118	0.097	0.078
07027	BEAUCHASTEL	0.100	0.103	0.120	0.127	0.184	0.192	0.174
07028	BEAULIEU	0.175	0.153	0.147	0.136	0.136	0.132	0.122
07029	BEAUMONT	0.308	0.166	0.127	0.112	0.094	0.100	0.093
07030	BEAUVENE	0.227	0.167	0.166	0.142	0.105	0.099	0.094
07031	BERRIAS-ET-CASTELJAU	0.183	0.148	0.147	0.140	0.139	0.121	0.123
07032	BERZEME	0.221	0.184	0.141	0.147	0.105	0.111	0.091
07033	BESSAS	0.180	0.134	0.133	0.133	0.134	0.154	0.134
07034	BIDON	0.203	0.117	0.079	0.092	0.102	0.187	0.219
07035	BOFFRES	0.224	0.175	0.146	0.135	0.105	0.110	0.105
07036	BOGY	0.164	0.137	0.133	0.119	0.128	0.144	0.175
07037	BOREE	0.269	0.200	0.157	0.135	0.100	0.074	0.064
07038	BORNE	0.311	0.188	0.167	0.124	0.073	0.058	0.079
07039	BOZAS	0.208	0.177	0.164	0.144	0.114	0.106	0.088
07040	BOUCIEU-LE-ROI	0.212	0.163	0.145	0.147	0.116	0.104	0.113
07041	BOULIEU-LES-ANNONAY	0.115	0.116	0.126	0.137	0.157	0.166	0.183
07042	BOURG-SAINT-ANDEOL	0.091	0.090	0.107	0.173	0.168	0.181	0.190
07044	BROSSAINC	0.173	0.160	0.157	0.144	0.124	0.117	0.124
07045	BURZET	0.243	0.202	0.154	0.129	0.095	0.095	0.082
07047	CELLER-DU-LUC	0.234	0.166	0.149	0.128	0.113	0.098	0.111
07048	CHALENCON	0.223	0.174	0.164	0.147	0.097	0.099	0.095
07049	CHAMBON	0.331	0.205	0.154	0.110	0.071	0.069	0.061
07050	CHAMBONAS	0.159	0.145	0.148	0.148	0.138	0.127	0.135
07051	CHAMPAGNE	0.131	0.129	0.129	0.132	0.143	0.155	0.182
07052	CHAMPS	0.197	0.169	0.159	0.141	0.121	0.096	0.118
07053	CHANDOLAS	0.181	0.141	0.141	0.142	0.136	0.132	0.127

Figure 5: Les données.

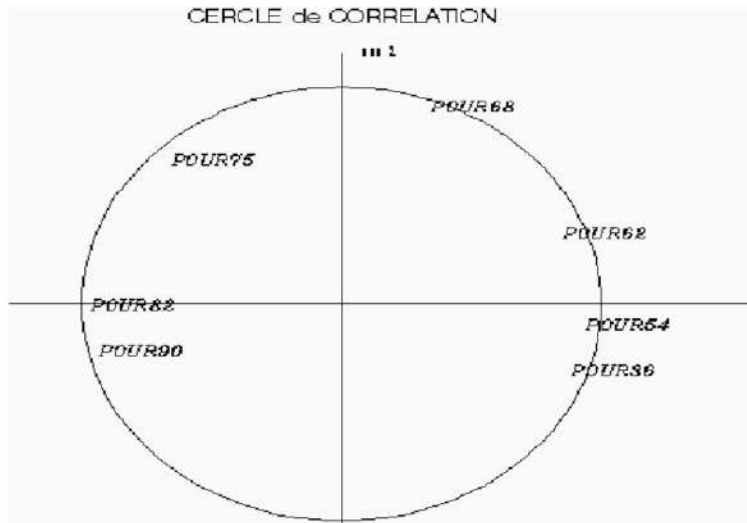
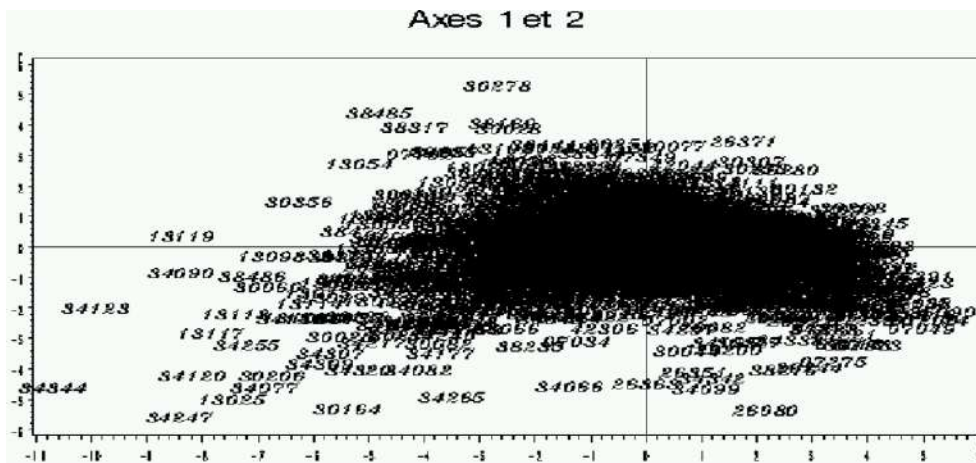


Figure 6: ACP sur les communes (88% sur les axes 1 et 2)



7004	07 162	07098	07114	07006	07271	07009	70164	07016	70126	07060	70226	07008	70077	07118	70046
7005	07 164	07045	07 150	07022	07272	07008	70204	07002	70116	07034	70222	07 077	70007	07 010	70071
71 52	07 165	07015	07 151	07126	07273	07007	70205	07001	70117	07035	70223	07 078	70008	07 011	70072
7077	07 166	07002	07 152	07 116	70016	07006	70206	07 010	70002	70043	70224	07 079	70009	07 012	70073
7082	07 167	07068	07 153	07 120	70017	07005	70207	07000	70118	70044	70225	07 080	70010	07 013	70074
7048	07 163	07003	07 154	07 124	70018	07004	70208	07003	70119	70045	70226	07 081	70011	07 014	70075
7020	07 170	07004	07 155	07 128	70019	07003	70209	07002	70120	70046	70227	07 082	70012	07 015	70076
7022	07 172	07005	07 156	07 132	70020	07002	70210	07001	70121	70047	70228	07 083	70013	07 016	70077
7066	07 173	07006	07 157	07 136	70021	07001	70211	07000	70122	70048	70229	07 084	70014	07 017	70078
7055	07 178	07007	07 158	07 140	70022	07000	70212	07000	70123	70049	70230	07 085	70015	07 018	70079
7051	07 180	07008	07 159	07 144	70023	07000	70213	07000	70124	70050	70231	07 086	70016	07 019	70080
7074	07 182	07009	07 160	07 148	70024	07000	70214	07000	70125	70051	70232	07 087	70017	07 020	70081
7084	07 183	07010	07 161	07 152	70025	07000	70215	07000	70126	70052	70233	07 088	70018	07 021	70082
71 70	07 184	07011	07 162	07 156	70026	07000	70216	07000	70127	70053	70234	07 089	70019	07 022	70083
7 02	07 185	07012	07 163	07 160	70027	07000	70217	07000	70128	70054	70235	07 090	70020	07 023	70084
7 02	07 186	07013	07 164	07 164	70028	07000	70218	07000	70129	70055	70236	07 091	70021	07 024	70085
7 02	07 187	07014	07 165	07 168	70029	07000	70219	07000	70130	70056	70237	07 092	70022	07 025	70086
7 02	07 188	07015	07 166	07 172	70030	07000	70220	07000	70131	70057	70238	07 093	70023	07 026	70087
7 02	07 189	07016	07 167	07 176	70031	07000	70221	07000	70132	70058	70239	07 094	70024	07 027	70088
7 02	07 190	07017	07 168	07 180	70032	07000	70222	07000	70133	70059	70240	07 095	70025	07 028	70089
7 02	07 191	07018	07 169	07 184	70033	07000	70223	07000	70134	70060	70241	07 096	70026	07 029	70090
7 02	07 192	07019	07 170	07 188	70034	07000	70224	07000	70135	70061	70242	07 097	70027	07 030	70091
7 02	07 193	07020	07 174	07 192	70035	07000	70225	07000	70136	70062	70243	07 098	70028	07 031	70092
7 02	07 194	07021	07 178	07 196	70036	07000	70226	07000	70137	70063	70244	07 099	70029	07 032	70093
7 02	07 195	07022	07 182	07 200	70037	07000	70227	07000	70138	70064	70245	07 100	70030	07 033	70094
7 02	07 196	07023	07 186	07 204	70038	07000	70228	07000	70139	70065	70246	07 101	70031	07 034	70095
7 02	07 197	07024	07 190	07 208	70039	07000	70229	07000	70140	70066	70247	07 102	70032	07 035	70096
7 02	07 198	07025	07 194	07 212	70040	07000	70230	07000	70141	70067	70248	07 103	70033	07 036	70097
7 02	07 199	07026	07 198	07 216	70041	07000	70231	07000	70142	70068	70249	07 104	70034	07 037	70098
7 02	07 200	07027	07 202	07 220	70042	07000	70232	07000	70143	70069	70250	07 105	70035	07 038	70099
7 02	07 201	07028	07 206	07 224	70043	07000	70233	07000	70144	70070	70251	07 106	70036	07 039	70100
7 02	07 202	07029	07 210	07 228	70044	07000	70234	07000	70145	70071	70252	07 107	70037	07 040	70101
7 02	07 203	07030	07 214	07 232	70045	07000	70235	07000	70146	70072	70253	07 108	70038	07 041	70102
7 02	07 204	07031	07 218	07 236	70046	07000	70236	07000	70147	70073	70254	07 109	70039	07 042	70103
7 02	07 205	07032	07 222	07 240	70047	07000	70237	07000	70148	70074	70255	07 110	70040	07 043	70104
7 02	07 206	07033	07 226	07 244	70048	07000	70238	07000	70149	70075	70256	07 111	70041	07 044	70105
7 02	07 207	07034	07 230	07 248	70049	07000	70239	07000	70150	70076	70257	07 112	70042	07 045	70106
7 02	07 208	07035	07 234	07 252	70050	07000	70240	07000	70151	70077	70258	07 113	70043	07 046	70107
7 02	07 209	07036	07 238	07 256	70051	07000	70241	07000	70152	70078	70259	07 114	70044	07 047	70108
7 02	07 210	07037	07 242	07 260	70052	07000	70242	07000	70153	70079	70260	07 115	70045	07 048	70109
7 02	07 211	07038	07 246	07 264	70053	07000	70243	07000	70154	70080	70261	07 116	70046	07 049	70110
7 02	07 212	07039	07 250	07 268	70054	07000	70244	07000	70155	70081	70262	07 117	70047	07 050	70111
7 02	07 213	07040	07 254	07 272	70055	07000	70245	07000	70156	70082	70263	07 118	70048	07 051	70112
7 02	07 214	07041	07 258	07 276	70056	07000	70246	07000	70157	70083	70264	07 119	70049	07 052	70113
7 02	07 215	07042	07 262	07 280	70057	07000	70247	07000	70158	70084	70265	07 120	70050	07 053	70114
7 02	07 216	07043	07 266	07 284	70058	07000	70248	07000	70159	70085	70266	07 121	70051	07 054	70115
7 02	07 217	07044	07 270	07 288	70059	07000	70249	07000	70160	70086	70267	07 122	70052	07 055	70116
7 02	07 218	07045	07 274	07 292	70060	07000	70250	07000	70161	70087	70268	07 123	70053	07 056	70117
7 02	07 219	07046	07 278	07 296	70061	07000	70251	07000	70162	70088	70269	07 124	70054	07 057	70118
7 02	07 220	07047	07 282	07 300	70062	07000	70252	07000	70163	70089	70270	07 125	70055	07 058	70119
7 02	07 221	07048	07 286	07 304	70063	07000	70253	07000	70164	70090	70271	07 126	70056	07 059	70120
7 02	07 222	07049	07 290	07 308	70064	07000	70254	07000	70165	70091	70272	07 127	70057	07 060	70121
7 02	07 223	07050	07 294	07 312	70065	07000	70255	07000	70166	70092	70273	07 128	70058	07 061	70122
7 02	07 224	07051	07 298	07 316	70066	07000	70256	07000	70167	70093	70274	07 129	70059	07 062	70123
7 02	07 225	07052	07 302	07 320	70067	07000	70257	07000	70168	70094	70275	07 130	70060	07 063	70124
7 02	07 226	07053	07 306	07 324	70068	07000	70258	07000	70169	70095	70276	07 131	70061	07 064	70125
7 02	07 227	07054	07 310	07 328	70069	07000	70259	07000	70170	70096	70277	07 132	70062	07 065	70126
7 02	07 228	07055	07 314	07 332	70070	07000	70260	07000	70171	70097	70278	07 133	70063	07 066	70127
7 02	07 229	07056	07 318	07 336	70071	07000	70261	07000	70172	70098	70279	07 134	70064	07 067	70128
7 02	07 230	07057	07 322	07 340	70072	07000	70262	07000	70173	70099	70280	07 135	70065	07 068	70129
7 02	07 231	07058	07 326	07 344	70073	07000	70263	07000	70174	70100	70281	07 136	70066	07 069	70130
7 02	07 232	07059	07 330	07 348	70074	07000	70264	07000	70175	70101	70282	07 137	70067	07 070	70131
7 02	07 233	07060	07 334	07 352	70075	07000	70265	07000	70176	70102	70283	07 138	70068	07 071	70132
7 02	07 234	07061	07 338	07 356	70076	07000	70266	07000	70177	70103	70284	07 139	70069	07 072	70133
7 02	07 235	07062	07 342	07 360	70077	07000	70267	07000	70178	70104	70285	07 140	70070	07 073	70134
7 02	07 236	07063	07 346	07 364	70078	07000	70268	07000	70179	70105	70286	07 141	70071	07 074	70135
7 02	07 237	07064	07 350	07 368	70079	07000	70269	07000	70180	70106	70287	07 142	70072	07 075	70136
7 02	07 238	07065	07 354	07 372	70080	07000	70270	07000	70181	70107	70288	07 143	70073	07 076	70137
7 02	07 239	07066	07 358	07 376	70081	07000	70271	07000	70182	70108	70289	07 144	70074	07 077	70138
7 02	07 240	07067	07 362	07 380	70082	07000	70272	07000	70183	70109	70290	07 145	70075	07 078	70139
7 02	07 241	07068	07 366	07 384	70083	07000	70273	07000	70184	70110	70291	07 146	70076	07 079	70140
7 02	07 242	07069	07 370	07 388	70084	07000	70274	07000	70185	70111	70292	07 147	70077	07 080	70141
7 02	07 243	07070	07 374	07 392	70085	07000	70275	07000	70186	70112	70293	07 148	70078	07 081	70142
7 02															

Contenus des super-classes

Classe 1 : forte décroissance, 323 communes, aucune des Bouches du Rhône, par exemple Issarlès (07), St Maurice Navacelles (34), St Léger-du-Ventoux (84)

Classe 2 : décroissance moyenne, 320 communes, aucune des Bouches-du-Rhône, par exemple Lamastre (07), Antraigues (07), Génolhac (30), Fontaine-de-Vaucluse (84)

Classe 3 : stabilité, 726 communes, la classe la plus nombreuse, par exemple Aubenas (07), Marseille (13), Crozes-Hermitage (26), Nîmes (30), Clermont-l'Hérault (34)

Classe 4 : croissance modérée, 322 communes, par exemple Bourg-St Andéol (07), Aix-en-Provence (13), Baux-de-Provence (13), Montélimar (26), Pont-St Esprit (30)

Classe 5 : croissance forte, 92 communes, aucune d'Ardèche, beaucoup des Bouches-du-Rhône, par exemple Fos-sur-mer (13), Vitrolles (13), La-Grande-Motte (34), Seyssins (38), Puget (84)

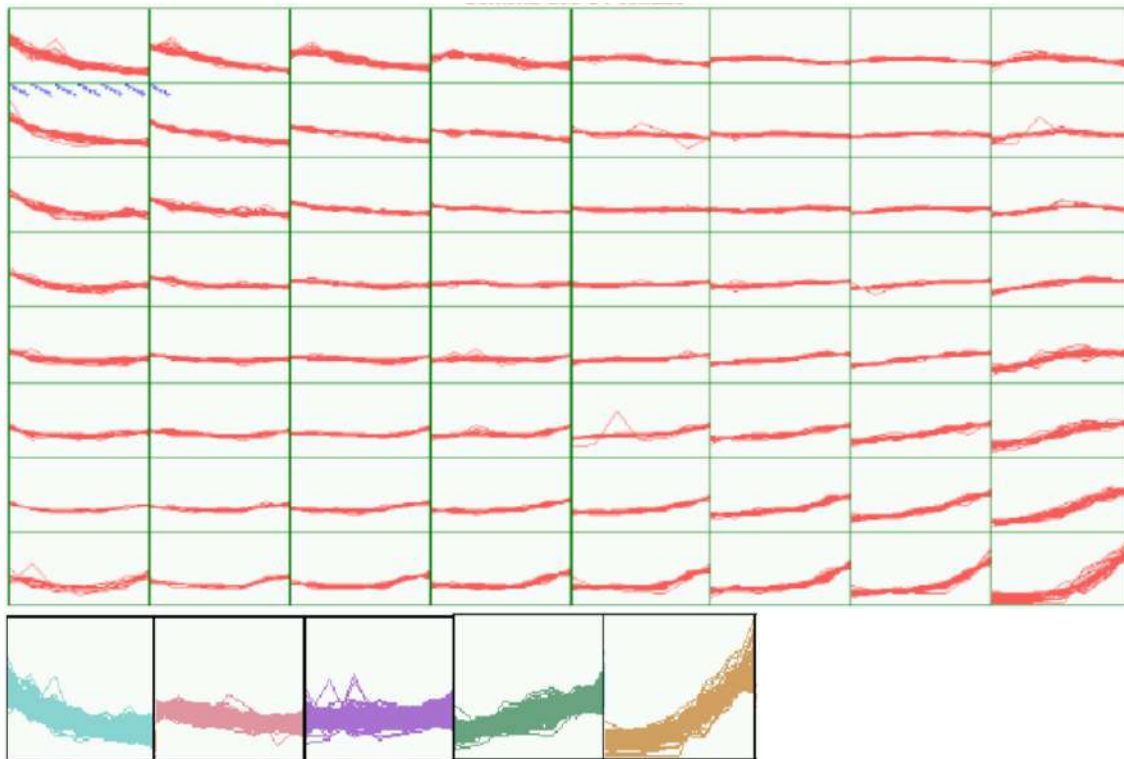


Figure 9: Contenu des classes et des super-classes

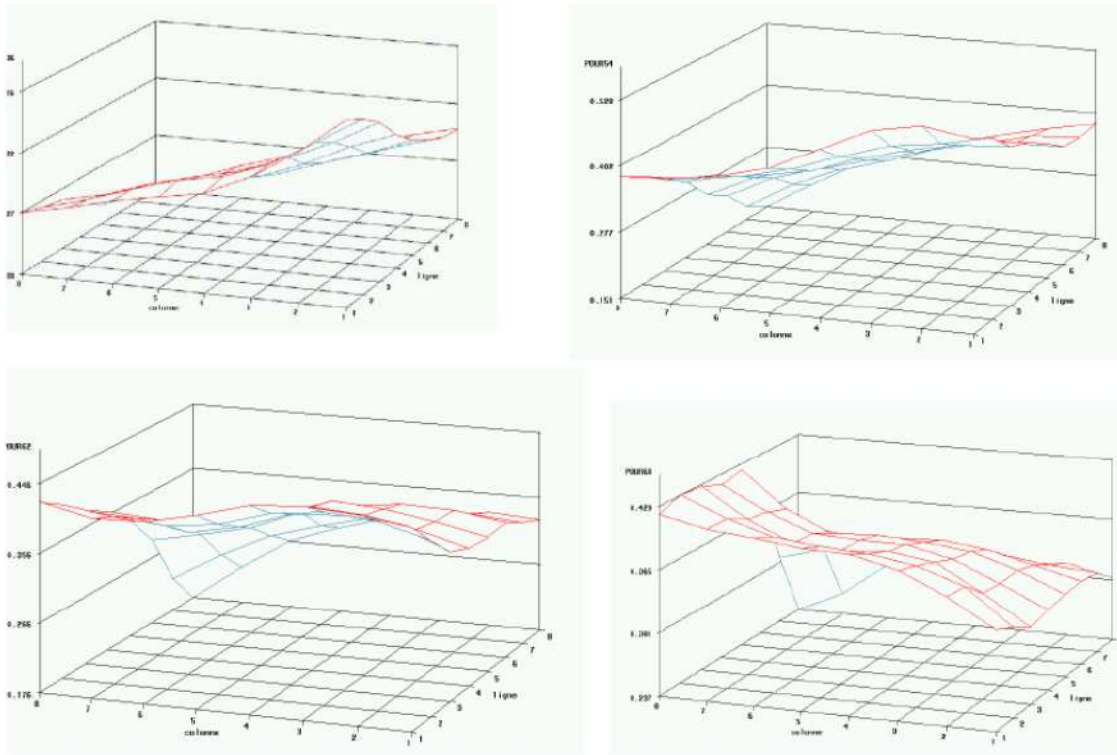


Figure 10: Les quatre premiers recensements

Qualités de la classification

	<i>DL</i>	<i>POUR36</i>	<i>POUR54</i>	<i>POUR62</i>	<i>POUR68</i>	<i>POUR75</i>	<i>POUR82</i>	<i>POUR90</i>
<i>SC_INTER</i>	4	4.560	1.668	0.656	0.153	0.361	2.029	5.745
<i>SC_INTRA</i>	1 778	0.980	0.425	0.443	0.484	0.515	0.522	1.365
<i>SC_TOTAL</i>	1 782	5.540	2.093	1.100	0.637	0.876	2.551	7.110
<i>FISHER</i>		2068.472	1744.982	657.897	140.003	311.868	1727.231	1870.749
<i>P_VALUE</i>			0	0	0	0	0	

Les recensements des années 62, 68 et 75 sont moins discriminants.

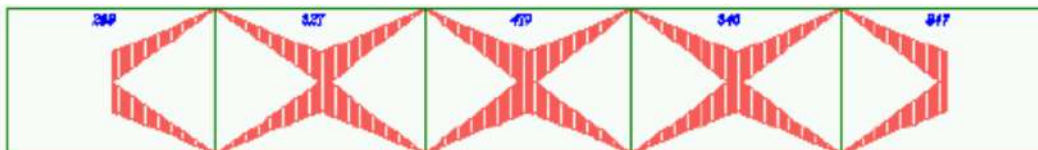


Figure 11: Ficelle à 5 classes

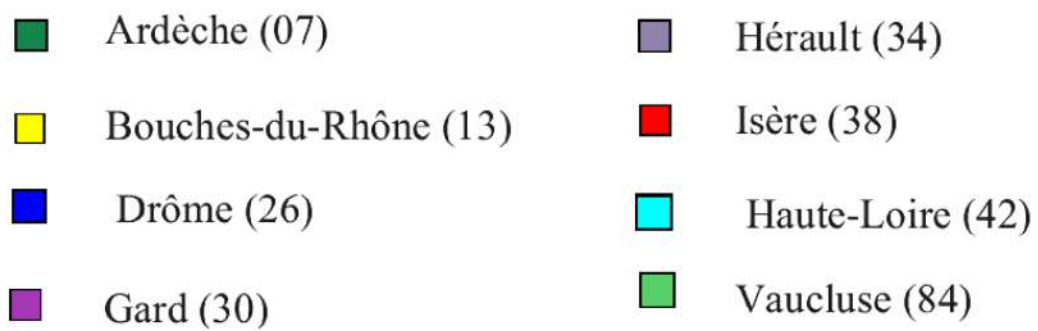
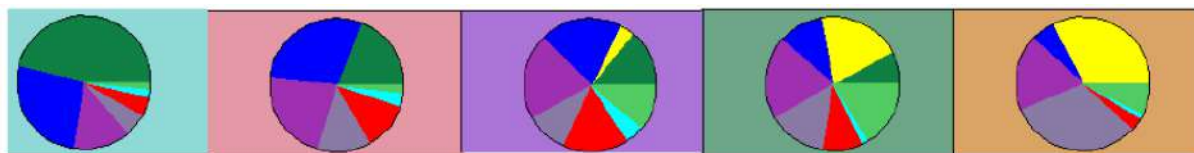


Figure 12: Département par super-classes

Exemple 18 (Un exemple : 96 pays en 1996) ANCRX Croissance annuelle de la population en %

TXMORT Taux de mortalité infantile (en pour mille)

TXANAL Taux d'illettrisme en %

SCOL2 Indice de fréquentation scolaire au second degré

PNBH PNB par habitant exprimé en dollars

CHOMAG Taux de chômage en %

INFLAT Taux d'inflation en %

NIVIDH Niveau de l'Indice de Développement Humain (6 niveaux codés par libellés : faible1, faible2, moyen1, moyen2, fort1, fort2) Niveau d'Indice de Développement Humain

CATIDH (6 niveaux codés de 1 à 6)

PAYS	ANCRX	TXMORT	TXANAL	SCOL2	PNBH	CHOMAG	INFLAT	NIVIDH	CATIDH
Afghanistan	6	159	70,9	15	276	19	17	faible1	1
Afrique du sud	2,6	46,9	23,5	77	2873	33	10	moyen2	4
Albanie	1,1	33,1	8	29,2	828	17,9	16,1	moyen2	4
Algerie	2,2	42,1	42	61	1570	27	31	moyen2	4
Allemagne	0,2	5,8	1	101,2	24993	9,4	3,1	fort2	6
Angola	3,6	126,8	58	14	575	25	951	faible1	1
Arabie Saoudite	3	68,8	39,5	49	7081	6,6	0,7	moyen2	4
Argentine	1,1	33,8	4,4	72,3	7827	11,3	4	fort1	5
Australie	1,3	5,9	0,1	84	17688	9,7	2,5	fort2	6
Bahrein	2,5	24,2	17	99	7500	15	2	fort1	5
Belgique	0,1	7,8	0,3	103,2	22225	12,6	2,6	fort2	6
Bolivie	2,2	74,9	20	37	733	6,2	8,5	moyen1	3
Bresil	1,6	59,8	18	43	3073	5,5	1094	fort1	5
Bulgarie	-0,2	15,3	2,1	68,2	1058	17	33	moyen2	4
Cameroun	2,9	85,8	36,5	32	733	25,1	12,8	moyen1	3
Canada	1	6,7	3,1	104,2	18286	10,4	0,3	fort2	6
Chili	1,4	14,4	5,7	67	3643	6,1	11,2	fort1	5
Chine	1	25,8	22,4	55	418	2,5	22	moyen1	3
Chypre	1	9,9	4,5	95	9459	2	4,8	fort2	6
Colombie	1,7	36,8	8,5	62	1379	8	22,9	fort1	5
Comores	3,5	81,7	42,5	19	317	16	24,8	faible2	2
Coree du Sud	1	14,9	3,7	96	7572	2,3	6	fort1	5
Costa Rica	2,2	13,5	5,2	47	1896	5	15	fort1	5
Cote d'Ivoire	3,3	90,9	46,8	25	587	17	25,6	faible1	1
Croatie	0,1	11,5	3,2	83,2	2755	13,1	97,6	moyen2	4
Danemark	0,2	5,6	1	114,2	28346	12,1	2,1	fort2	6
Egypte	1,9	58,8	50,5	76	632	20,2	8,3	moyen1	3
Emirats arabes uni	2,2	23,2	20,9	89	23809	0,2	5	fort1	5
Equateur	2,1	36,8	12,8	55	1205	7,2	26	moyen2	4
Espagne	0,2	7,3	7,1	110,2	12283	24,4	4,8	fort2	6
Etats Unis	1	8,2	3	97,2	25219	5,6	2,8	fort2	6

Figure 13: Les données.

Analyse des correspondances

Établir les liaisons (dangereuses?) entre deux variables qualitatives :

- tableau de contingence, marge, profils

$X \setminus Y$	y_1	\dots	y_m	marg.
x_1	n_{11}	\dots	n_{m1}	$n_{1.}$
\vdots				
x_r	n_{1r}	\dots	n_{mr}	$n_{r.}$
marg.	$n_{.1}$	\dots	$n_{.m}$	$n_{..}$

- le χ^2 d'écart à l'indépendance (K. Pearson) :

$$\boxed{\sum_i^r \sum_j^m \frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n_{..}})^2}{\frac{n_{i.}n_{.j}}{n_{..}}}} = n \left(\sum_{i,j} \frac{n_{ij}^2}{n_{i.}n_{.j}} - 1 \right) \sim \chi_{(r-1)(m-1)}^2, \quad (123)$$

- on a les relations suivantes :

$$n_{i.} = \sum_j^p n_{ij}, n_{.j} = \sum_i^p n_{ij}, n_{..} = \sum_{i,j} n_{ij} \text{ soit en terme de fréquences :}$$

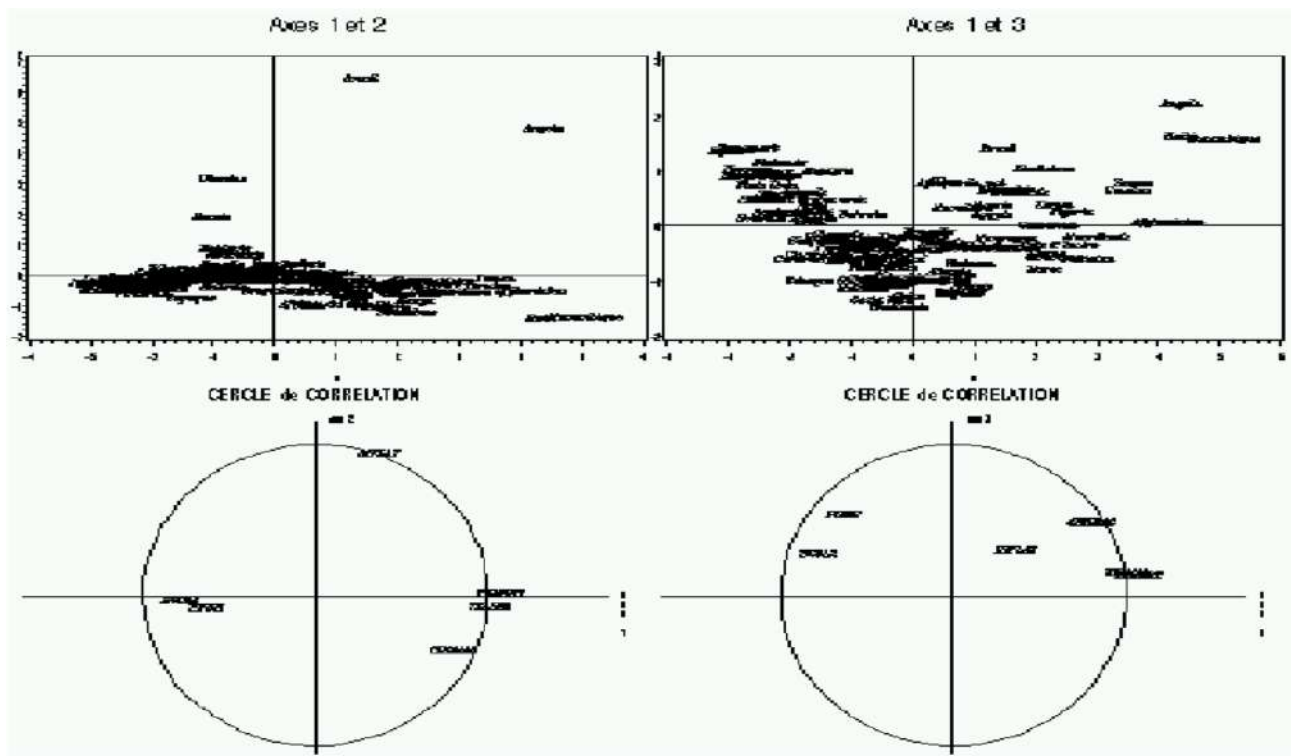
$$f_{ij} = \frac{n_{ij}}{n_{..}}, f_{i.} = \sum_j^p f_{ij}, f_{.j} = \sum_i^p f_{ij}, f_{..} = \sum_{i,j} f_{ij} \quad (124)$$

- transformation du tableau de contingence : ce n'est pas le tableau d'effectifs brutes qui nous intéresse mais les tableaux des profils lignes et colonnes. on note les profils lignes : $\frac{f_{ij}}{f_{i.}} = \frac{n_{ij}}{n_{i.}}$ et colonnes par : $\frac{f_{ij}}{f_{.j}} = \frac{n_{ij}}{n_{.j}}$
- hypothèse d'indépendance (empirique) : $f_{ij} = f_{i.}f_{.j}$
- choix des distances : la distance euclidienne entre 2 profils-lignes traduit la ressemblance ou la différence entre des modalités sans tenir compte des effectifs de ces modalités :

$$d^2(i, i') = \sum_j^p \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2, \quad (125)$$

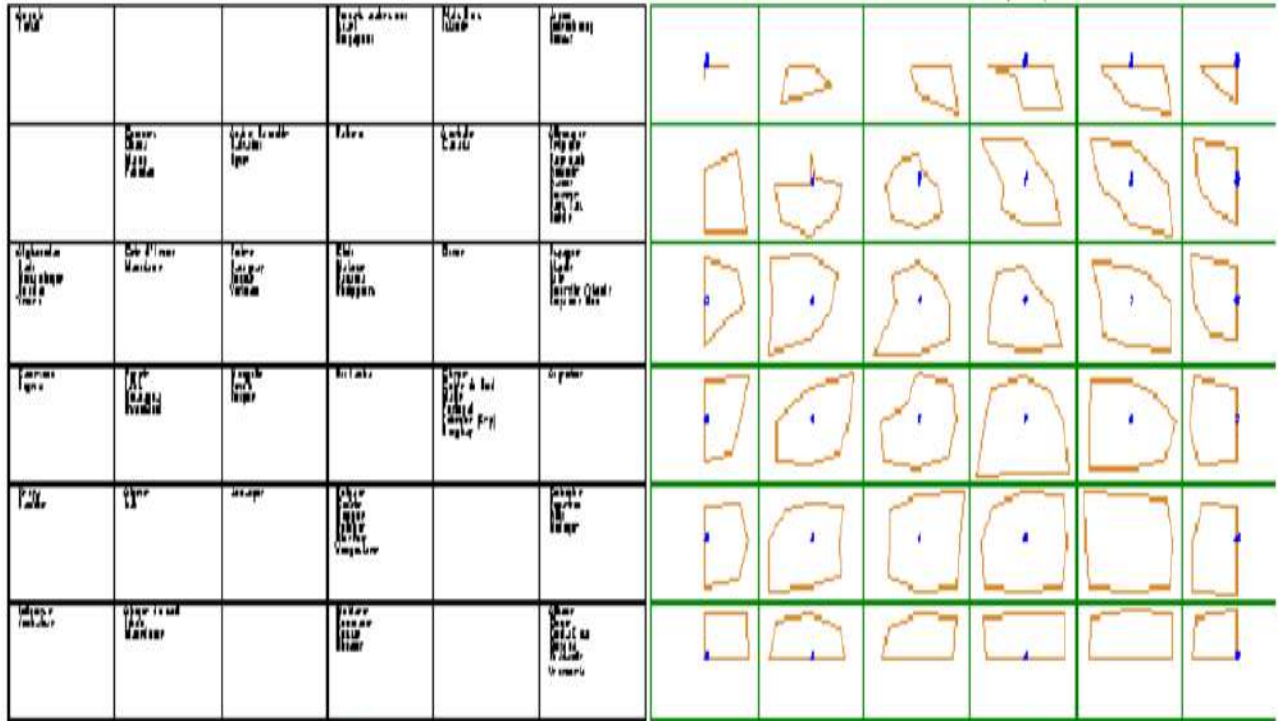
mais cette distance favorise les colonnes avec une masse importante \Rightarrow distance entre les lignes :

$$\boxed{d^2(i, i') = \sum_j^p \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2}, \quad (126)$$



Angola Bresil			Emirats arabes uni Israël Singapour	Etats Unis Irlande	Japon Luxembourg Suisse
	Comores Ghana Maroc Pakistan	Arabie Saoudite Salvador Syrie	Bahrein	Australie Canada	Allemagne Belgique Danemark Finlande France Norvege Pays Bas Suede
Afghanistan Bahr Mozambique Soudan Yemen	Cote d'Ivoire Mauritanie	Bolnie Paraguay Jordnie Vietnam	Chine Malaisie Panama Philippines	Grèce	Espagne Irlande Italie Nouvelle Zelande Royaume Uni
Cameroun Nigeria	Egypte Cuba Nicaragua Swaziland	Mongolie Perou Turquie	Sri Lanka	Chypre Corée du Sud Malte Portugal Tchéquie (Rep) Uruguay	Argentine
Kenya Namibie	Algérie Irak	Jamaïque	Bulgarie Croatie Hongrie Pologne Slovenie Yougoslavie		Colombie Equateur Inde Mexique
Indonésie Zimbabwe	Afrique du sud Liban Macedoine		Moldavie Roumanie Russie Ukraine		Albanie Chine Costa Rica Guyana Thailande Venezuela

Figure 14: Analyse en Composantes Principales et Carte de Kohonen



Libelles des 7 clusters

7 Super-classes avec les Proches Voisins

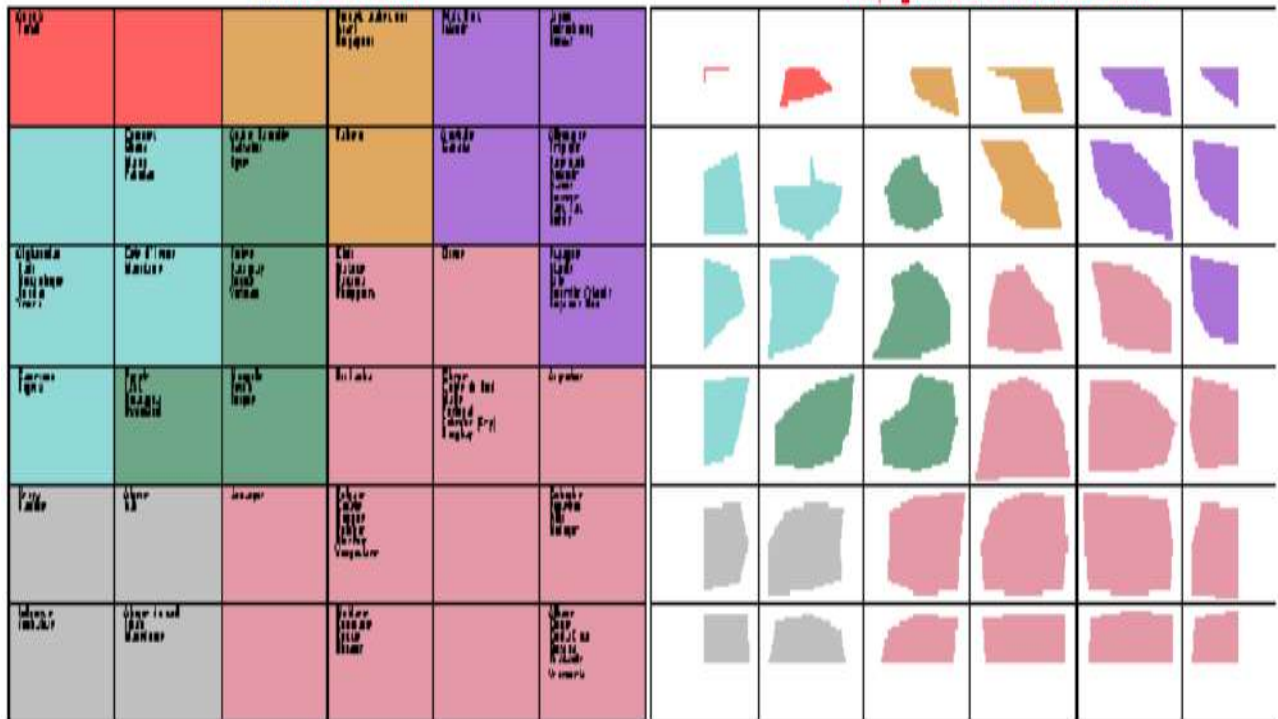


Figure 15: Classes, super-classes, distances

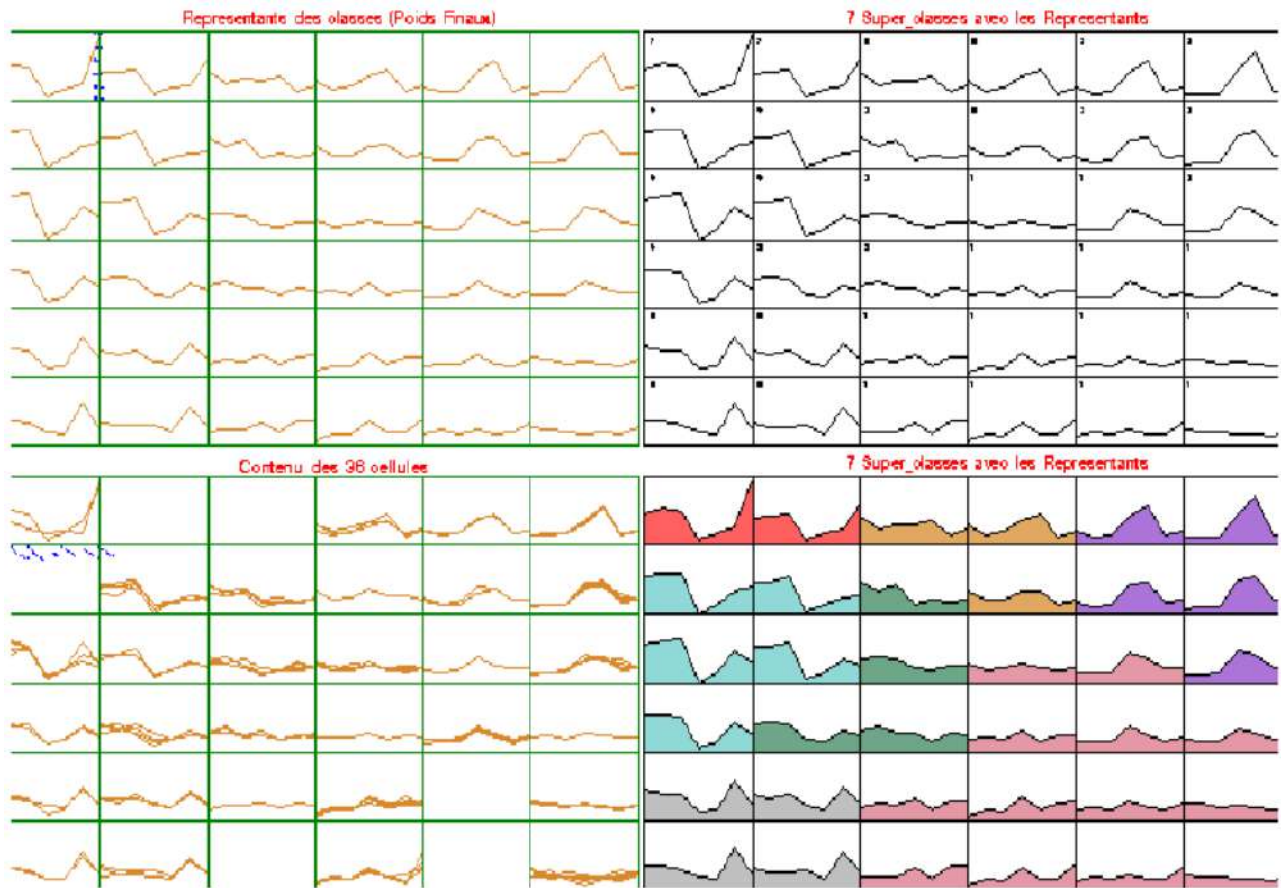


Figure 16: Vecteurs-codes, contenus des classes, et super-classes

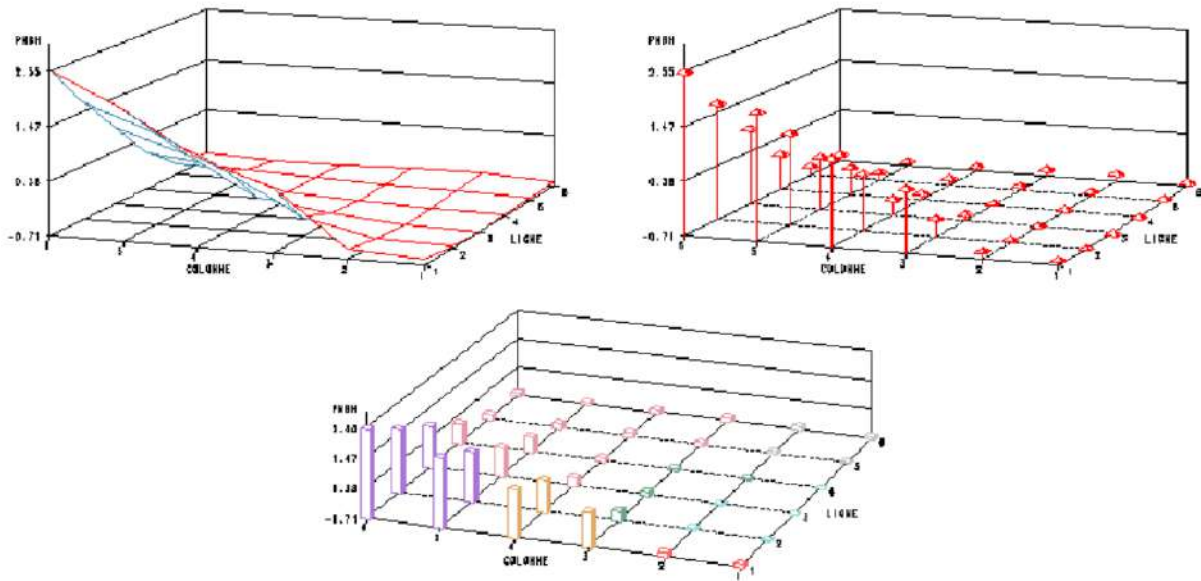


Figure 17: PNB par habitant le long de la grille

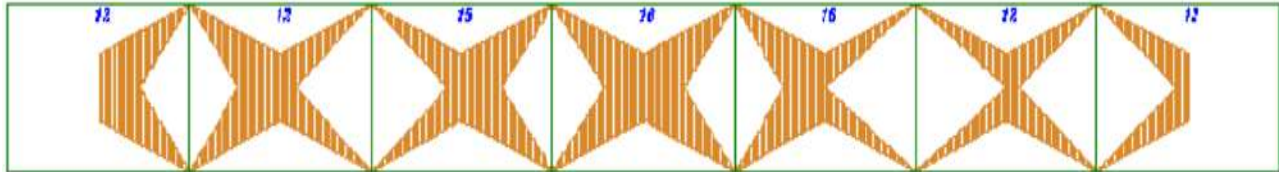
Positions des 24 Données Supplémentaires

Turkmenistan			Koweït		Liechtenstein
	Vanuatu	Irak Jordanie Libye			
Benin Guinée Liberia	Bangladesh				
	Inde	Tuvalu	Vatican		
Birmanie Guatemala Lesotho	Congo		Cuba Dominique Grenade		
				Kirghizstan	Andorre Corée du Nord

Figure 18: Classement des observations avec données manquantes

Albanie Angola Comores Cote d'Ivoire Haïti Mauritanie Mozambique Nigeria	Pakistan Soudan Yemen	Algérie Cameroun Égypte Indonésie Iran Kenya Laos Maroc Namibie	Nicaragua Swaziland Zimbabwe	Afrique Albanie Arabie S. Bolivie Guyane Liban Macedoine Mongolie Paraguay	Pérou Salvador Syrie Turquie Vietnam	Bahrein Bresil Chine Colombie Costa R. Equateur Fiji Jamaïque Malaisie	Mexique Panama Philippines Sri Lanka Thaïlande Venezuela Yougoslavie	Argentine Bulgarie Chili Corée du Sud Croatie Hongrie Malte Moldavie Pologne	Portugal Roumanie Russie Slovenie Tchéquie Ukraine Uruguay	Australie Canada Chypre Espagne Irlande Grèce Irlande Israël Italie	Nouvelle Royaume Singapour	Allemagne Belgique Danemark Émirats Etats Un. France Islande Japon Luxembourg	Norvège Pays Bas Suède Suisse
---	-----------------------------	---	------------------------------------	--	--	--	--	--	--	---	----------------------------------	---	--

Distances (M) avec les plus proches voisins



Contenu des 7 cellules

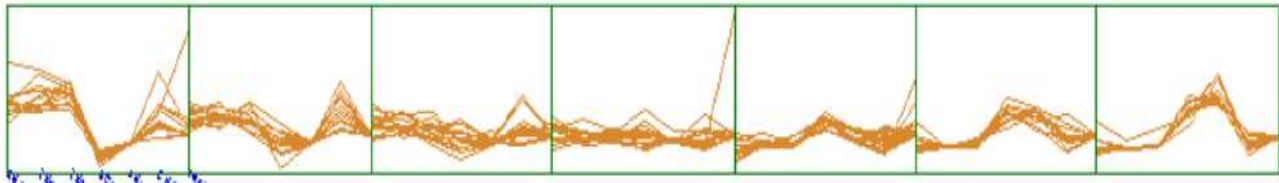


Figure 19: KACP sur une ficelle 7

et de la même manière la distance entre les profils-colonnes :

$$d^2(j, j') = \sum_i^n \frac{1}{f_i} \left(\frac{f_{ij}}{f_j} - \frac{f_{i'j'}}{f_{j'}} \right)^2, \quad (127)$$

Schéma général de l'analyse des correspondances

nuage de n points-lignes dans \mathbb{R}^p $X = D_n^{-1}F$	analyse du tableau X	nuage de p points-lignes dans \mathbb{R}^n $X = D_p^{-1}F^T$
p -coord. (point-ligne i) $x_{ij} = \frac{f_{ij}}{f_i}, j = 1, \dots, p$		n -coord. (point-colonne j) $x_{ij} = \frac{f_{ij}}{f_j}, i = 1, \dots, n$
$M = D_p^{-1}$	avec la métrique M	$M = D_n^{-1}$
$d^2(i, i') = \sum_j^p \frac{1}{f_j} \left(\frac{f_{ij}}{f_i} - \frac{f_{i'j}}{f_i'} \right)^2$		$d^2(j, j') = \sum_i^n \frac{1}{f_i} \left(\frac{f_{ij}}{f_j} - \frac{f_{ij'}}{f_{j'}} \right)^2$
$N = D_n$ masse du point i : f_i .	et le critère N	$N = D_p$ masse du point j : f_j .

Critère à maximiser et matrice à diagonaliser : rendre maximal la quantité

$$\sum_i^n f_i \cdot d^2(i, O) \qquad u^T D_p^{-1} F^T D_n^{-1} F D_p^{-1} u \qquad (128)$$

sous la contrainte $U^T D_p^{-1} u = 1$. u est le vecteur propre de la matrice $S = F^T D_n^{-1} F D_p^{-1}$,
soit $s_{jj'} = \sum_i^n \frac{f_{ij} f_{ij'}}{f_i \cdot f_{j'}}$. Idem pour \mathbb{R}^n .

Principe de l'analyse factorielle

“Mise en facteur” de $a^2 - b^2$ sous la forme de 2 *facteurs* : $(a + b)(a - b)$.

idée : Peut-on décomposer un tableau qui n'est pas simple en une somme de plusieurs tableaux plus simples à lire car pouvant être résumés par un produit de facteurs ?

Quel est le nombre de tableaux nécessaires pour représenter un tableau quelconque ?

Exemple 19 (Orientation des bacheliers 1975) *Tableau issu des statistiques scolaires.*

	Destination			
	Université	Prépa	Autres	Total
$T =$ séries A	13	2	5	20
B,D,D'	20	2	8	30
C,E	10	5	5	20
Technique	7	1	22	30
Total	50	10	40	100

Question : cette répartition moyenne varie t'elle selon les séries du bac ?

L'effectif théorique d'une case est obtenu par le produit des marginales ligne et colonne car le respect de la proportion moyenne pour toutes les lignes du tableau correspond à ce qu'on nomme la situation d'indépendance.

	Université	Prépa	Autres	
$T_0 =$ séries A	10	2	8	20
B,D,D'	15	3	12	30
C,E	10	2	8	20
Technique	15	3	12	30
	50/100	10/100	40/100	

$$\begin{array}{|c|c|c|} \hline 13 & 2 & 5 \\ \hline 20 & 2 & 8 \\ \hline 10 & 5 & 5 \\ \hline 7 & 1 & 22 \\ \hline \end{array} - \begin{array}{|c|c|c|} \hline 10 & 2 & 8 \\ \hline 15 & 3 & 12 \\ \hline 10 & 2 & 8 \\ \hline 15 & 3 & 12 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 3 & 0 & -3 \\ \hline 5 & -1 & -4 \\ \hline 0 & 3 & -3 \\ \hline -8 & -2 & 10 \\ \hline \end{array}$$

T

T_0

R_1

R_1 : tableau des écarts à l'indépendance information la plus importante pour l'interprétation). Permet de repérer s'il y a

- attraction entre une ligne et une colonne (écart positif),
- indépendance, écart nul

- répulsion, écart négatif.

A titre d'exemple voici une décomposition de ce tableau R_1 :

$$\begin{array}{c} R_1 \\ \hline \begin{array}{ccc} 3 & 0 & -3 \\ 5 & -1 & -4 \\ 0 & 3 & -3 \\ -8 & -2 & 10 \end{array} \\ \hline \end{array} = \begin{array}{c} T_1 \\ \hline \begin{array}{ccc} 1 & 1 & -2 \\ 1 & 1 & -2 \\ 2 & 2 & -4 \\ -4 & -4 & 8 \end{array} \\ \hline \begin{array}{ccc} 1 & 1 & -2 \end{array} \end{array} + \begin{array}{c} T_2 \\ \hline \begin{array}{ccc} 2 & -1 & -1 \\ 4 & -2 & -2 \\ -2 & 1 & 1 \\ -4 & 2 & 2 \end{array} \\ \hline \begin{array}{ccc} 2 & -1 & -1 \end{array} \end{array} + \begin{array}{c} \begin{array}{c} 1 \\ 1 \\ 2 \\ -4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ -1 \\ -2 \end{array} \end{array}$$

Donc, $T = T_0 + T_1 + T_2$. Les choses sont plus simples car les 3 tableaux T_0, T_1 et T_2 peuvent être obtenus à partir des vecteurs marginaux.

Comment fait on une analyse factorielle ?

Principe de base : remplacer le tableau initial par une approximation de rang 1 (donc facile à lire) qui soit *une bonne approximation de celui-ci* au nom d'un certain critère.

- Contribution au χ^2 ou écart pondéré² : on passe de 100 à 1000.

	Univ.	Prépa	Autres	Total
séries A	1 300	200	500	2 000
B,D,D'	2 000	200	800	3 000
C,E	1 000	500	500	2 000
Technique	700	100	2 200	3 000
Total	5 000	1 000	4 000	10 000

$$\text{eff. pondéré} = \text{obs.} \times \frac{\text{observé}}{\text{eff. théorique}}$$

Exemple : $426,7 = -800 \frac{-800}{1500}$

Ecart observé			Effectifs théoriques		
300	0	-300	1 000	200	800
500	-100	-400	1 500	300	1 200
0	300	-300	1 000	200	800
-800	-200	1 000	1 500	300	1 200

↓

	Univ.	Prépa	Autres	Total	%
A	90	0	112.5	202.5	8.1
B,D,D'	166.7	33.3	133.3	333.3	13.4
C,E	0	450	112.5	562.5	22.6
Technique	426.7	133.3	833.3	1 393.3	55.9
Total	683.3	616.7	1 191.7	2 491.7	100
%	27.4	24.7	47.8	100	

- première approximation : l'indépendance. Elle traduit le niveau zéro de la relation entre les lignes et les colonnes (tableau T_0).
- Dans $R_1 = T - T_0$, les écarts > 0 correspondent à des attractions entre les lignes et les colonnes ; les écarts < 0 à des répulsions.

²Nous avons vu l'écart à l'indépendance.

- la décomposition des écarts (le cœur de l'analyse factorielle). T_0 étant une approximation de T , de combien se trompe t-on en prenant T_0 à la place de T ? de $T - T_0 = R_1$. On soustrait T_0 de *rang 1* à T de *rang 3* $\Rightarrow R_1$ est de *rang 2*.

T_1 : première approximation des écart (premier facteur) n'est pas R_1 . On commet une erreur $R_1 - T_1 = T_2$ de rang 1. C'est le *dernier facteur* ou *correction de deuxième espèce*.

- décomposition dans le cas général : elle peut se résumer par :

$$\begin{array}{ll} T & \text{est approximé par } T_0, \quad T - T_0 = R_1 \text{ (de rang 2);} \\ R_1 & \text{---} \quad T_0, \quad R_1 - T_1 = R_2 \text{ (de rang 1);} \\ R_2, & \text{dernier facteur = } T_2, \quad R_2 - T_2 = 0. \end{array}$$

ou encore $T = T_0 + T_1 + T_2$ d'importance décroissante.

\rightarrow algorithmes d'extraction des vecteurs propres. T_1

Choix de l'indicateur d'importance d'un tableau calculée de façon différente selon le type d'analyse :

- **analyse en composantes principales** : somme des carrés,
- **analyse des correspondances** : somme des carrés pondérés.

Extraction des vecteurs propres

0 1
1 2.
3 3

Considérons le tableau 3×2 suivant :

But : avoir un couple de vecteurs qui résume la structure des lignes et des colonnes.

Principe : Partir d'une approximation grossière du 1er vecteur propre (par ex. $(1, 2, 4)^T$) et l'affiner par des itérations successives. On multiplie le vecteur arbitraire par chacune des colonnes du tableau : indicateur de la structure des colonnes...

Exemple 20 (algorithme d'extraction sur tableau 3×2)

		0	2	4	6
0	1	1	17	392	9 085
1	2	2	48	1 111	25 572
2	3	4	93	2 157	50 001

1	14	17
3	327	392
5	7 582	9 085
7	175	210
7	755	592

Y a t'il une régularité dans la répartition des éléments ? Regardez les pourcentages...

0	2	4	6	8
1	17	392	9 085	210 592
14,286	10,759	10,710	10,709	10,709
2	48	1 111	25 752	596 939
28,571	30,380	30,355	30,354	30,354
4	93	2 157	50 001	1 159 641
57,143	58,861	58,934	58,937	58,937
7	158	3 660	84 838	1 966 572
100	100	100	100	100

Remarques :

- l'algorithme stabilisent les pourcentages *en colonne* au bout de 6 itérations (règle d'arrêt).
- idem pour les résumés *en ligne*
- égalité du coefficient d'accroissement entre itérations 6 et 8 : $9085 \xrightarrow{\times 23,180} 210592$ et $50001 \xrightarrow{\times 23,180} 1159041$

⇓

coefficient d'accroissement \Leftrightarrow valeur propre
--

- cette première valeur propre est tjrs $<$ la somme des carrés des nombres du tableau. Ici : $32^2 + 3^2 + 1^2 + 2^2 + 1^2 = 24 > 23,180 !$

Algorithme en analyse en composantes principales

Idée (par rapport au paragraphe précédent) : neutraliser l'accroissement des vecteurs propres.

Considérons le tableau 3×3 suivant :

1	2	3
4	5	6
7	8	9

Vecteur de départ : $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

<table style="border-collapse: collapse; width: 60px;"> <tr><td style="padding: 2px 5px;">1</td><td style="padding: 2px 5px;">2</td><td style="padding: 2px 5px;">3</td></tr> <tr><td style="padding: 2px 5px;">4</td><td style="padding: 2px 5px;">5</td><td style="padding: 2px 5px;">6</td></tr> <tr><td style="padding: 2px 5px;">7</td><td style="padding: 2px 5px;">8</td><td style="padding: 2px 5px;">9</td></tr> </table>	1	2	3	4	5	6	7	8	9	0	2	4	6
1	2	3											
4	5	6											
7	8	9											
	1	y_i x_i	y_i x_i	y_i x_i									
	1	3,6467 0,2165	3,6197 0,2148	3,6196 0,2148									
	1	8,7750 0,5211	8,7709 0,5206	8,7709 0,5206									
	1	13,9032 0,8256	13,9222 0,8263	13,9222 0,8263									
		$\lambda = 283,5974$	$\lambda = 283,8586$	$\lambda = 283,8586$									
		$\xi = 16,8404$	$\xi = 16,8481$	$\xi = 16,8481$									
	12 15 18	y_j											
1	$\lambda = 693$ $\xi = 26,3249$	x_j											
	<u>0,4558 0,5698 0,6838</u>	y_j											
	12 15 18	y_j											
3	$\lambda = 283,8575$ $\xi = 16,8481$	x_j											
	<u>0,4796 0,5724 0,6651</u>	y_j											
	8,0815 9,6433 11,2051	y_j											
5	$\lambda = 283,8586$ $\xi = 16,8481$	x_j											
	<u>0,4797 0,5724 0,6651</u>	y_j											
	8,0815 9,6433 11,2051	y_j											

Exemple 21 (illustration du processus de réduction) Cas du vecteur $\begin{pmatrix} 12 \\ 15 \\ 18 \end{pmatrix}$. Posons :

$(12, 15, 18)^2 = 12^2 + 15^2 + 18^2 = 693 \rightarrow \text{norme} = \sqrt{693} = 26,3249$. On aura $\text{norme}^2 = \lambda$ et $\text{norme} = \xi$. Et,

$$\left(\frac{12}{\sqrt{693}}, \frac{15}{\sqrt{693}}, \frac{18}{\sqrt{693}} \right) = (0,4558; 0,5698; 0,6838).$$

Puis, en multipliant le vecteur réduit par chaque ligne du tableau, par exemple, la première ligne : $0,4558 \times 1 + 0,5698 \times 2 + 0,6838 \times 3 = 3,6467$. On obtient le vecteur colonne

$$\begin{pmatrix} 3,6467 \\ 8,7750 \\ 13,9032 \end{pmatrix}, \text{ de norme } \lambda = 283,5974.$$

Doit-on utiliser pour la reconstitution les vecteurs propres calibrés de type y ou ceux de type x ? Comme le tableau reconstitué doit être *d'importance* $\lambda = 283,8586$.

le plus simple est de multiplier un vecteur calibré par un vecteur réduit.

	1,7362	2,0717	2,4073	y_i	3,6196
Exemple 22 (suite exemple)	4,2072	5,0202	5,8332		8,7709
	6,6781	7,9686	9,2592		13,9222
	y_j	8,0815	9,6433	11,2051	$\xi = 16,8481$

et, la somme des carrés d'une ligne (colonne) est égale au carré de la composante du vecteur propre, par exemple $1,7362^2 + 2,0717^2 + 2,4073^2 = 3,6196^2$. Tableau à comparer

avec le tableau initial

1	2	3
4	5	6.
7	8	9

La contribution absolue est le carré de l'élément du vecteur propre (calibré).

Exemple, la contribution absolue de la première ligne est $\frac{3,6196^2}{283,8586} = 4,6\%$.

La contribution relative au facteur de la ligne, sachant que la somme des carrés de la ligne initiale est $1^2 + 2^2 + 3^2 = 14$, est de $\frac{13,1015}{14} = 93,6\%$.

La reconstitution pour l'ensemble du tableau apporte $\frac{283,8586}{285} = 99,6\%$.

Exercice 18 (suite exemple 21 : Poursuivre le reste de la reconstitution)

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} - \begin{pmatrix} 1,7362 & 2,0717 & 2,4073 \\ 4,2072 & 5,0202 & 5,8332 \\ 6,6781 & 7,9686 & 9,2592 \end{pmatrix} = \begin{pmatrix} -0,7362 & -0,0717 & 0,5927 \\ -0,2072 & -0,0202 & 0,1668 \\ 0,3219 & 0,0314 & -0,2592 \end{pmatrix} \quad (129)$$

On poursuit les calculs comme à la première décomposition.

Solution :

	0	2	4	
	-1	y_i	x_i	y_i
	-1	-0,9479	-0,8872	-0,9479
	1	-0,02667	-0,2496	-0,2667
		0,4145	0,3879	0,4145
		$\lambda = 1,1414$		stop
		$\xi = 1,0684$		

	1,2653	0,1233	-1,0187	y_j
1	$\lambda = \mathbf{2,6538}$	$\xi = \mathbf{1,6291}$		
	0,7767	0,0757	-0,6253	x_j
	0,8298	0,0809	-0,6681	y_j
3	$\lambda = 1,1414$	$\xi = 1,0684$		
	0,7767	0,0757	-1,0684	x_j

On a la relation : somme des carrés initiale = $\lambda_1 + \lambda_2 \Leftrightarrow 285 = 283,8586 + 1,1414$. Idem pour les contribution absolues des lignes et des colonnes : par ex., pour la première ligne $14 = 3,6196^2 + (-0,9479)^2$.

Algorithme en analyse des correspondances

Considérons T le tableau 4×3 suivant :

40	28	32
4	124	72
28	132	40
128	116	256

Idée (par rapport au l'ACP) : introduire une pondération à l'intérieur de chaque itération, consistant à diviser chaque élément du vecteur par *la marge du tableau d'origine*.

Tableau reconstitué à partir des marginales $T_0 =$

20	40	40	20	-12	-8
40	80	80	-36	44	-8
40	80	80	-12	52	-40
100	200	200	28	-84	56

et $R_1 =$

le total faisant 1000.

Exemple 23 (reconstitution du tableau d'indépendance (premier facteur)) *Les totaux marginaux (=pondération) sont $(100, 200, 200, 500)^T$ pour les lignes et $(200, 400, 400)^T$ pour les colonnes (resp.). Partant d'un vecteur arbitraire $(1, 1, 1, 1)^T$, multiplié avec chaque colonne du tableau divisé par le total, on obtient $(200, 200, 400)^T$. En divisant chaque élément du vecteur ligne par le total marginal correspondant \Rightarrow le vecteur ligne pondéré vaut $(1, 1, 1)^T$. etc.*

<table style="border-collapse: collapse; width: 100%;"> <tr><td>40</td><td>28</td><td>32</td></tr> <tr><td>4</td><td>124</td><td>72</td></tr> <tr><td>28</td><td>132</td><td>40</td></tr> <tr><td>128</td><td>116</td><td>256</td></tr> </table>	40	28	32	4	124	72	28	132	40	128	116	256	<table style="border-collapse: collapse; width: 100%;"> <tr><td><i>pond.</i></td><td><i>0</i></td><td><i>2</i></td></tr> <tr><td>100</td><td>1</td><td>100</td></tr> <tr><td>200</td><td>1</td><td>200</td></tr> <tr><td>200</td><td>1</td><td>200</td></tr> <tr><td>500</td><td>1</td><td>500</td></tr> </table>	<i>pond.</i>	<i>0</i>	<i>2</i>	100	1	100	200	1	200	200	1	200	500	1	500
40	28	32																										
4	124	72																										
28	132	40																										
128	116	256																										
<i>pond.</i>	<i>0</i>	<i>2</i>																										
100	1	100																										
200	1	200																										
200	1	200																										
500	1	500																										
<i>Pond</i>	200	400	400																									
	200	200	400																									
	1	1	1																									

et reconstitution du tableau des écarts à l'indépendance ($R_1 = T - T_0$) :

20	40	40	100
40	80	80	200
40	80	80	200
100	200	200	500
200	400	400	1000

La pondération, à chaque itération, consiste à diviser chaque élément du vecteur par la marge du tableau d'origine.

Exemple 24 (Constitution de l'approximation de rang 2 du tableau de départ)

<table style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 10px;">20</td><td style="padding: 2px 10px;">-12</td><td style="padding: 2px 10px;">-8</td></tr> <tr><td style="padding: 2px 10px;">-36</td><td style="padding: 2px 10px;">44</td><td style="padding: 2px 10px;">-8</td></tr> <tr><td style="padding: 2px 10px;">-12</td><td style="padding: 2px 10px;">52</td><td style="padding: 2px 10px;">-40</td></tr> <tr><td style="padding: 2px 10px;">28</td><td style="padding: 2px 10px;">-84</td><td style="padding: 2px 10px;">56</td></tr> </table>	20	-12	-8	-36	44	-8	-12	52	-40	28	-84	56	<table style="border-collapse: collapse; text-align: center;"> <tr><td style="padding: 0 5px;">coef.</td><td style="padding: 0 5px;">0</td><td style="padding: 0 5px;">2</td></tr> <tr><td style="padding: 0 5px;">pond.</td><td style="padding: 0 5px;">1</td><td style="padding: 0 5px;">2</td></tr> <tr><td style="padding: 0 5px;"></td><td style="padding: 0 5px;">1</td><td style="padding: 0 5px;">-1</td></tr> <tr><td style="padding: 0 5px;"></td><td style="padding: 0 5px;">2</td><td style="padding: 0 5px;">-1</td></tr> <tr><td style="padding: 0 5px;"></td><td style="padding: 0 5px;">5</td><td style="padding: 0 5px;">1</td></tr> </table>	coef.	0	2	pond.	1	2		1	-1		2	-1		5	1	<table style="border-collapse: collapse; text-align: center;"> <tr><td style="padding: 0 5px;">y_i</td><td style="padding: 0 5px;">y^i</td><td style="padding: 0 5px;">z_i</td><td style="padding: 0 5px;">z^i</td></tr> <tr><td style="padding: 0 5px;">-10,5830</td><td style="padding: 0 5px;">10,5830</td><td style="padding: 0 5px;">0,2582</td><td style="padding: 0 5px;">0,2582</td></tr> <tr><td style="padding: 0 5px;">-31,7490</td><td style="padding: 0 5px;">-15,8745</td><td style="padding: 0 5px;">-0,7746</td><td style="padding: 0 5px;">-0,3873</td></tr> <tr><td style="padding: 0 5px;">-31,7490</td><td style="padding: 0 5px;">-15,8745</td><td style="padding: 0 5px;">-0,7746</td><td style="padding: 0 5px;">-0,3873</td></tr> <tr><td style="padding: 0 5px;">52,9150</td><td style="padding: 0 5px;">10,5830</td><td style="padding: 0 5px;">1,2910</td><td style="padding: 0 5px;">0,2582</td></tr> <tr><td colspan="4" style="padding: 0 5px;">$\lambda = 1680,00 \quad \xi = 40,9878$</td></tr> </table>	y_i	y^i	z_i	z^i	-10,5830	10,5830	0,2582	0,2582	-31,7490	-15,8745	-0,7746	-0,3873	-31,7490	-15,8745	-0,7746	-0,3873	52,9150	10,5830	1,2910	0,2582	$\lambda = 1680,00 \quad \xi = 40,9878$			
20	-12	-8																																																			
-36	44	-8																																																			
-12	52	-40																																																			
28	-84	56																																																			
coef.	0	2																																																			
pond.	1	2																																																			
	1	-1																																																			
	2	-1																																																			
	5	1																																																			
y_i	y^i	z_i	z^i																																																		
-10,5830	10,5830	0,2582	0,2582																																																		
-31,7490	-15,8745	-0,7746	-0,3873																																																		
-31,7490	-15,8745	-0,7746	-0,3873																																																		
52,9150	10,5830	1,2910	0,2582																																																		
$\lambda = 1680,00 \quad \xi = 40,9878$																																																					

	2	4	4	coef.	pond.	
	96	-192	96	y_j		
	48	-48	24	y^j		
1	$\lambda = 16128 \quad \xi = 126,9961$					
	0,7559	-1,5119	0,7559	z_j		
	0,3780	-0,3780	0,1890	z^j		
	30,9839	-61,9677	30,9839	y_j		
	15,4919	-15,4919	7,746	y^j		
3	$\lambda = 1680 \quad \xi = 40,9878$					
	0,7559	-1,5119	0,7559	z_j		

Le vecteur ligne pondéré y_j s'obtient à la première itération par

$$\left(\frac{96}{2}, \frac{-192}{4}, \frac{96}{4}\right) = (48, -48, 24) = y^j.$$

Ce vecteur est réduit ensuite en divisant par la racine de sa norme $\xi = \sqrt{\lambda}$ et

$$\lambda = \frac{96^2}{2} + \frac{-192^2}{4} + \frac{96^2}{4} = 16128 \rightarrow \xi = 126,9961 :$$

$$z_j = \left(\frac{96}{126,9961}, \frac{-192}{126,9961}, \frac{96}{126,9961}\right) = (0,7559; -1,5119, 0,7559) \quad (130)$$

$$z^j = \left(\frac{48}{126,9961}, \frac{-48}{126,9961}, \frac{24}{126,9961}\right) = (0,3780, -0,3780, 0,1890) \quad (131)$$

Tableau reconstitué :

	8	-16	8	y_i	
	-24	48	-24	-10,5830	
	-24	48	-24	-31,7490	
	40	-80	40	-31,7490	
				52,9150	
	y_j	30,9839	-61,9677	30,9839	$\xi = 40,9878$

On vérifiera que $z_j z^j = 1$.

A chaque itération, on trouvera 4 vecteurs :

- réduit calibré non-pondéré (y_j),
- puis après-pondération (y^j),
- réduit non-pondéré (z_j)
- puis réduit pondéré (z^j).

Le critère d'évaluation est une somme de carrés pondérés.

Analyse de tableaux de contingence

- on considère un tableau de contingence croisant 2 variables qualitatives
- on calcule les profils-lignes et les profils-colonnes
- on fabrique une matrice de données dont les observations sont
 - un profil-ligne suivi du profil-colonne le plus probable correspondant
 - ou d'un profil-colonne suivi du profil-ligne le plus probable correspondant
- on prend la distance du χ^2 .
- on applique l'algorithme de Kohonen à cette matrice de données et on représente le classement obtenu.
- les modalités les plus ressemblantes se retrouvent associées dans la même classe ou dans des classes voisines.
- analogue à l'analyse des correspondances en statistique.

Analyse des relations entre modalités de variables qualitatives / Analyse de tables de Burt

- Classiquement, l'analyse des correspondances des modalités de plus de 2 variables qualitatives se fait par **l'analyse des correspondances multiples**, qui est une **ACP pondérée sur la table de Burt associée**. La distance considérée est la distance du χ^2 . La table de Burt est un tableau de contingence généralisé, qui croise toutes les variables qualitatives 2 à 2.
- on applique ici un algorithme de Kohonen sur cette table de Burt, avec la même pondération et la distance du χ^2 .
- L'interprétation se fait de la même manière que pour KORRESP.
- Les modalités associées se retrouvent dans la même classe ou dans des classes voisines.

La table de Burt ci-dessous est obtenue en prenant $s = 3, m_1 = 3, m_2 = 2$ et $m_3 = 4$.

M : nombre total de modalités ($M = \sum_{q=1}^s m_q$)

\mathbf{z}_i , ($1 \leq i \leq M$) : le vecteur-colonne construit par les n réponses à la i -ième modalité ($z_{ij} = 1$ ou 0 selon que le sujet i a choisi la modalité j à la question q).

$Z = \{\mathbf{z}_i\}$: **tableau disjonctif complet**,

$$z_{i\cdot} = \sum_{j=1}^M z_{ij} = s \text{ marge en ligne,}$$

$$z_{\cdot j} = \sum_{i=1}^n z_{ij},$$

$$\text{effectif total : } z = \sum_{j=1}^n \sum_{i=1}^M z_{ij} = ns.$$

<table border="1" style="border-collapse: collapse; width: 100px; height: 150px;"> <thead> <tr><th>m_1</th><th>m_2</th><th>m_3</th></tr> </thead> <tbody> <tr><td>2</td><td>2</td><td>4</td></tr> <tr><td>2</td><td>1</td><td>3</td></tr> <tr><td>3</td><td>1</td><td>2</td></tr> <tr><td>1</td><td>2</td><td>4</td></tr> <tr><td>1</td><td>2</td><td>3</td></tr> <tr><td>2</td><td>2</td><td>3</td></tr> <tr><td>3</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>1</td><td>2</td></tr> <tr><td>2</td><td>2</td><td>3</td></tr> <tr><td>3</td><td>2</td><td>2</td></tr> <tr><td>1</td><td>1</td><td>4</td></tr> </tbody> </table>	m_1	m_2	m_3	2	2	4	2	1	3	3	1	2	1	2	4	1	2	3	2	2	3	3	1	1	1	1	1	2	1	2	2	2	3	3	2	2	1	1	4	→	<table border="1" style="border-collapse: collapse; width: 150px; height: 150px;"> <thead> <tr><th>m_1</th><th>m_2</th><th colspan="3">m_3</th></tr> </thead> <tbody> <tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr> </tbody> </table>	m_1	m_2	m_3			0	1	0	0	1	0	1	0	1	0	0	0	1	1	0	1	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	0	1
m_1	m_2	m_3																																																																																																																		
2	2	4																																																																																																																		
2	1	3																																																																																																																		
3	1	2																																																																																																																		
1	2	4																																																																																																																		
1	2	3																																																																																																																		
2	2	3																																																																																																																		
3	1	1																																																																																																																		
1	1	1																																																																																																																		
2	1	2																																																																																																																		
2	2	3																																																																																																																		
3	2	2																																																																																																																		
1	1	4																																																																																																																		
m_1	m_2	m_3																																																																																																																		
0	1	0	0	1																																																																																																																
0	1	0	1	0																																																																																																																
0	0	1	1	0																																																																																																																
1	0	0	0	1																																																																																																																
1	0	0	1	0																																																																																																																
0	1	0	0	1																																																																																																																
0	0	1	1	0																																																																																																																
1	0	0	1	0																																																																																																																
0	1	0	1	0																																																																																																																
0	1	0	0	1																																																																																																																
0	0	1	0	1																																																																																																																
0	0	1	0	0																																																																																																																
1	0	0	1	0																																																																																																																
1	0	0	0	1																																																																																																																
$R_{n,s}$		$Z_{n,M}, M=9, n=12$																																																																																																																		

$B = Z^T Z$: **tableau de contingence de Burt** ($M \times M$) qui rassemble les croisements 2 à 2 de toutes les variables ($b_{ij} = \sum_{i=1}^n z_{ij} z_{ij'}$). Le tableau est formé de $s \times s$ blocs.

$$b = \sum_{i=1}^M b_{ij} = s^2 n.$$

m_1			m_2			m_3		
0	1	0	0	1	0	0	0	1
0	1	0	1	0	0	0	1	0
0	0	1	1	0	0	1	0	0
1	0	0	0	1	0	0	0	1
1	0	0	0	1	0	0	1	0
0	1	0	0	1	0	0	1	0
0	0	1	1	0	1	0	0	0
1	0	0	1	0	1	0	0	0
0	1	0	1	0	0	1	0	0
0	1	0	0	1	0	0	1	0
0	0	1	0	1	0	1	0	0
1	0	0	1	0	0	0	0	1

$Z_{n,M}, M=9, n=12$

→

4	0	1	2	2	1	0	1	2
0	5	0	2	3	0	1	3	1
0	0	3	2	1	1	2	0	0
2	2	2	6	0	1	2	1	1
2	3	1	0	6	0	1	3	2
1	0	1	2	0	2	0	1	0
0	1	2	2	1	0	3	0	0
1	3	0	1	3	0	0	4	0
2	1	0	1	2	0	0	0	3

Table (symétrique) de Burt B

Le tableau B de correspondance multiple est un **assemblage de tableaux de contingence** (croisant 2 variables). L'analyse des correspondances (AC), appliquée au tableau disjonctif complet Z , est équivalente à l'analyse du tableau de contingence tableau de Burt B et produit les mêmes facteurs.

- **ACM sur Z**

Soit D la matrice diagonale $p \times p$ ayant les même éléments diagonaux que B , *i.e.* $d_{jj} = z_{.j}$. Chaque individu i est affecté d'une masse égale $\frac{1}{n}$ et chaque colonne j (modalité) est pondéré par $\frac{z_{.j}}{ns}$. La distance entre modalités dans \mathbb{R}^n et entre individus dans \mathbb{R}^p s'écrit :

$$d^2(j, j') = \sum_{i=1}^n n \left(\frac{z_{ij}}{z_{.j}} - \frac{z_{ij'}}{z_{.j'}} \right)^2 \quad (132)$$

$$d^2(i, i') = \sum_{j=1}^p \frac{ns}{z_{.j}} (z_{ij} - z_{i'j})^2. \quad (133)$$

F : **tableau des fréquences relatives** $F = \frac{1}{ns}Z$ de terme général $f_{ij} = \frac{z_{ij}}{ns}$,

les marges : $D_n = \frac{1}{n}I$ de terme général $f_{.i} = \sum_j f_{ij} = \frac{\delta_{ij}}{n}$ et $D_p = \frac{1}{ns}D$ de terme général

$$f_{.j} = \sum_i f_{ij} = \frac{\delta_{ij} z_{.j}}{ns},$$

Pour trouver les axes factoriels, on diagonalise la matrice³ :

$$S = F^T D_n^{-1} F D_p^{-1} = \frac{1}{s} Z^T Z D^{-1} \quad (134)$$

de terme général $s_{jj'} = \frac{1}{sz_{.j'}} \sum_i^n z_{ij} z_{ij'}$. Dans \mathbb{R}^p , l'équation du α -ième axe factoriel u_α est :

$$\frac{1}{s} Z^T Z D^{-1} u_\alpha = \lambda u_\alpha. \quad (135)$$

En vrac :

³La distance du χ^2 est la métrique D_n^{-1} .

– la distance entre la modalité j et le centre de gravité du nuage G , dont toutes les n coordonnées valent $\frac{1}{n}$ s'écrit :

$$d^2(j, G) = n \sum_{i=1}^n \left(\frac{z_{ij}}{z_{.j}} - \frac{1}{n} \right)^2 = \frac{n}{z_{.j}} - 1, \quad (136)$$

– inertie d'une modalité j : $I(j) = m_j d^2(j, G) = \frac{z_{.j}}{ns} d^2(j, G) = \frac{1}{s} (1 - \frac{z_{.j}}{ns})$ et l'inertie d'une question q est simplement : $I(q) = \sum_j^{m_q} I(j) = \frac{1}{s} (m_q - 1)$. Inertie totale $I = \sum_q I(q) = \frac{p}{s} - 1$.

• **ACM sur B**

L'analyse des correspondances appliquée au tableau Z est équivalente à l'analyse du tableau B . et produit les mêmes facteurs.

tableau des fréquences relatives $F = \frac{1}{ns^2} B$,

$$D_p = \frac{1}{ns} D,$$

Pour trouver les axes factoriels, la matrice à diagonaliser est :

$$S = \frac{1}{s} D^{-1} B. \quad (137)$$

on diagonalise la matrice $S^* = \frac{1}{s^2} D^{-1} B D^{-1} B$,

on obtient

$$\frac{1}{s^2} D^{-1} B D^{-1} B \phi_\alpha = \lambda_\alpha^2 \phi_\alpha. \quad (138)$$

les facteurs des 2 analyses sont donc colinéaires dans \mathbb{R}^p mais les valeurs propres diffèrent ($\lambda_B = \lambda^2$).

Exemple 25 (La famille, exemple tiré de “Statistique exploratoire multidimensionnelle” de Le
105 ménages, 8 questions, 20 modalités :

- La famille est l'endroit où on se sent bien : oui, non
- Les dépenses de logement sont une charge : négligeable, sans gros problème, une lourde charge, une très lourde charge
- Avez-vous eu récemment mal au dos : oui, non
- Vous imposez-vous des restrictions : oui, non
- Sexe de l'enquêté : masculin, féminin
- avez-vous un magnétoscope : oui, non
- Avez-vous eu récemment des maux de tête : oui, non
- Regardez-vous la télévision : tous les jours, assez souvent, pas très souvent, jamais

MAUX1	FAMILLE1 TV1	MAUX2	SEXE1	
MDOS1	SEXE2	DEPLOG2	MDOS2	MAGNETO1 RESTRIC2
MAGNETO2 RESTRIC1				DEPLOG1
		DEPLOG3		DEPLOG4
TV2	FAMILLE2	TV3		TV4

Figure 20: Carte des modalités

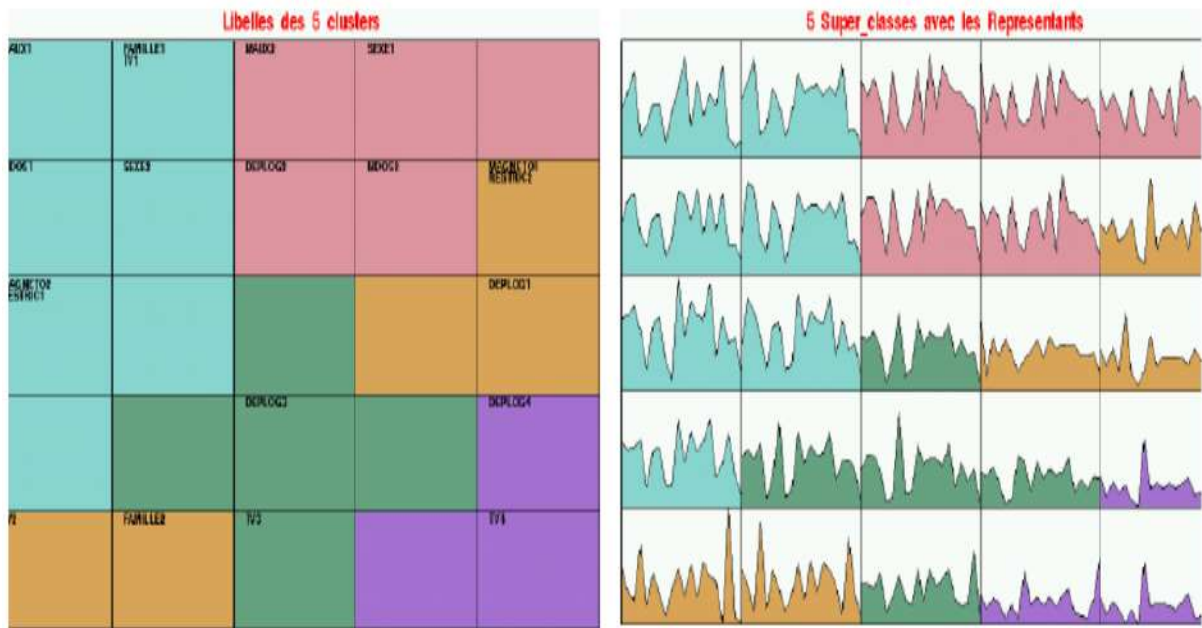


Figure 21: Super classes pour les modalités

Analyse du tableau disjonctif complet : modalités et individus

- Si l'on souhaite représenter aussi les individus (et pas seulement les modalités), on travaille sur le tableau disjonctif complet
- Classiquement, on fait alors une ACP sur le tableau disjonctif complet, correctement normalisé et en utilisant la distance du χ^2 .

	Q_{11}	Q_{12}	Q_{13}	Q_{21}	Q_{22}	Q_{31}	Q_{32}	Q_{33}	Q_{34}
1	0	1	0	0	1	0	0	0	1
2	0	1	0	1	0	0	0	1	0
3	0	0	1	1	0	0	1	0	0
4	1	0	0	0	1	0	0	0	1
5	1	0	0	0	1	0	0	1	0
6	0	1	0	0	1	0	0	1	0
7	0	0	1	1	0	1	0	0	0
8	1	0	0	1	0	1	0	0	0
9	0	1	0	1	0	0	1	0	0
10	0	1	0	0	1	0	0	1	0
11	0	0	1	0	1	0	1	0	0
12	1	0	0	1	0	0	0	0	1

- La méthode KACM1 : un algorithme de Kohonen sur le tableau disjonctif complet, avec la même normalisation et la distance du χ^2 .
 - un individu i est défini par $(\frac{z_{ij}}{\sqrt{z_{i.}\sqrt{z_{.j}}}})$, $i = 1, \dots, n$ avec $d_{i.} = Q$
 - On classe ainsi les individus, puis les modalités normalisées pour représenter des individus types . La modalité j est définie par $(\frac{b_{j\ell}}{z_{.j}\sqrt{z_{.l}\sqrt{Q}}})$.
 - La représentation graphique est malaisée (trop grand nombre de points), mais la classification obtenue est très utile.
- La méthode KACM2
 - consiste à pratiquer un algorithme de Kohonen sur la table de Burt, corrigée par la normalisation usuelle et la distance du χ^2 .
 - On classe ainsi les modalités (comme avec KACM), puis les individus correctement normalisées pour être comparables aux vecteurs qui représentent les modalités.
 - un individu i est défini par $(\frac{z_{ij}}{Q})$, une modalité par $(\frac{b_{ij}}{\sqrt{b_{i.}b_{.j}}})$
 - Avec KACM2, l'apprentissage est rapide puisqu'il ne porte que sur les modalités, mais il faut prolonger le nombre d'itérations pour calculer avec précision les vecteurs codes qui servent à classer ensuite les individus.

Libelles des 5 clusters

DEPLOG3	MADDC1 38	TV3 69	98	DEPLOG1 TV3
DEPLOG3	SEXE2 MADDC1 TV1 69 77	TV3 69	71 73 69	MAGNETO1 RESTRIC2
DEPLOG3	FAMILLE1 MAGNETO2 MADDC2 21	98 80 86 87 84	DEPLOG2 81	DEPLOG2 81
RESTRI1	98 80	TV3 69	MADDC2 TV3 69 67	FAMILLE2 TV3 69 67
TV3	98	DEPLOG1 81 80 86 87 84	SEXE1 RESTRIC1 80	DEPLOG1 81
MAGNETO2 RESTRIC1		DEPLOG2 MADDC2		FAMILLE2 TV3
	MADDC2			
SEXE2 FAMILLE1 TV1		SEXE1	RESTRIC2	DEPLOG1 MAGNETO1
MADDC1 MADDC1	DEPLOG3	TV3		DEPLOG4 TV4

Figure 22: Clusters KACM et modalmités seules

Exemple 26 (Questionnaire sur la famille : suite)

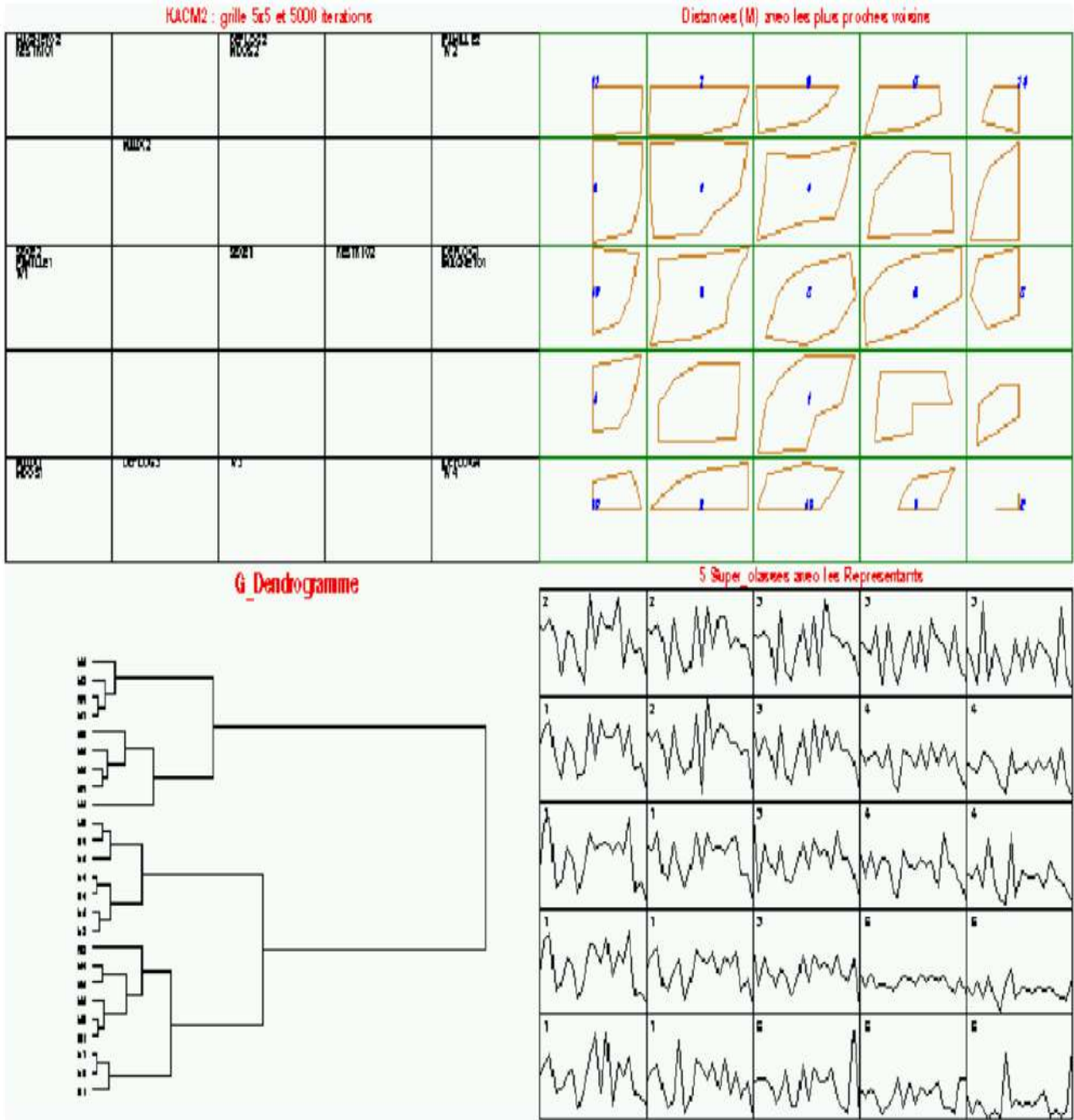


Figure 23: KACM2 : modalités, distances dendrogramme et représentants

Exemple 27 (**Demandeurs d'emplois inscrits à l'ANPE**) *Pour lesquels on a observé au moins deux périodes de chômages*⁴

- 204 personnes, extrait de l'enquête complète
8 variables qualitatives, 32 modalités
AGEC, classe d'âge, <25, 25-35, 35-45, 45-55, >55
CTINDMO, indemnisation journalière, <60F, 60-100, 100-150, >150
DIPL3, niveau de formation, >bac, niveau bac, <bac
DURC, durée totale du chômage, <12 mois, 12-24, >24
HAR, horaire mensuel en activité réduite, 0, 0-39, 39-78, 78-117, >117
PPARC, % d'AR dans la durée totale de chômage, 0, 0-0.1, 0.1-0.3, >0.3
RMOTIFA, type de sortie de chômage, (4 motifs)
RMOTIFI, type d'inscription au chômage (4 motifs)
- Motifs de sortie du chômage :
 - 1 : sortie vers l'emploi (que ce soit par l'ANPE ou par ses propres moyens)
 - 2 : sortie vers un stage ou un CES
 - 3 : retrait par maladie, retraite, service national, etc.
 - 4 : radiation, sanction, découragement, décès, etc.
- Motifs d'inscription regroupés en 4 modalités :
 - 1 : licenciement économique, fin d'intérim, autres
 - 2 : autre licenciement, fin de contrat à durée déterminée
 - 3 : démission, fin de conversion
 - 4 : recherche du premier emploi

⁴(entre le 1er juillet 1993 et le 31 août 1996)

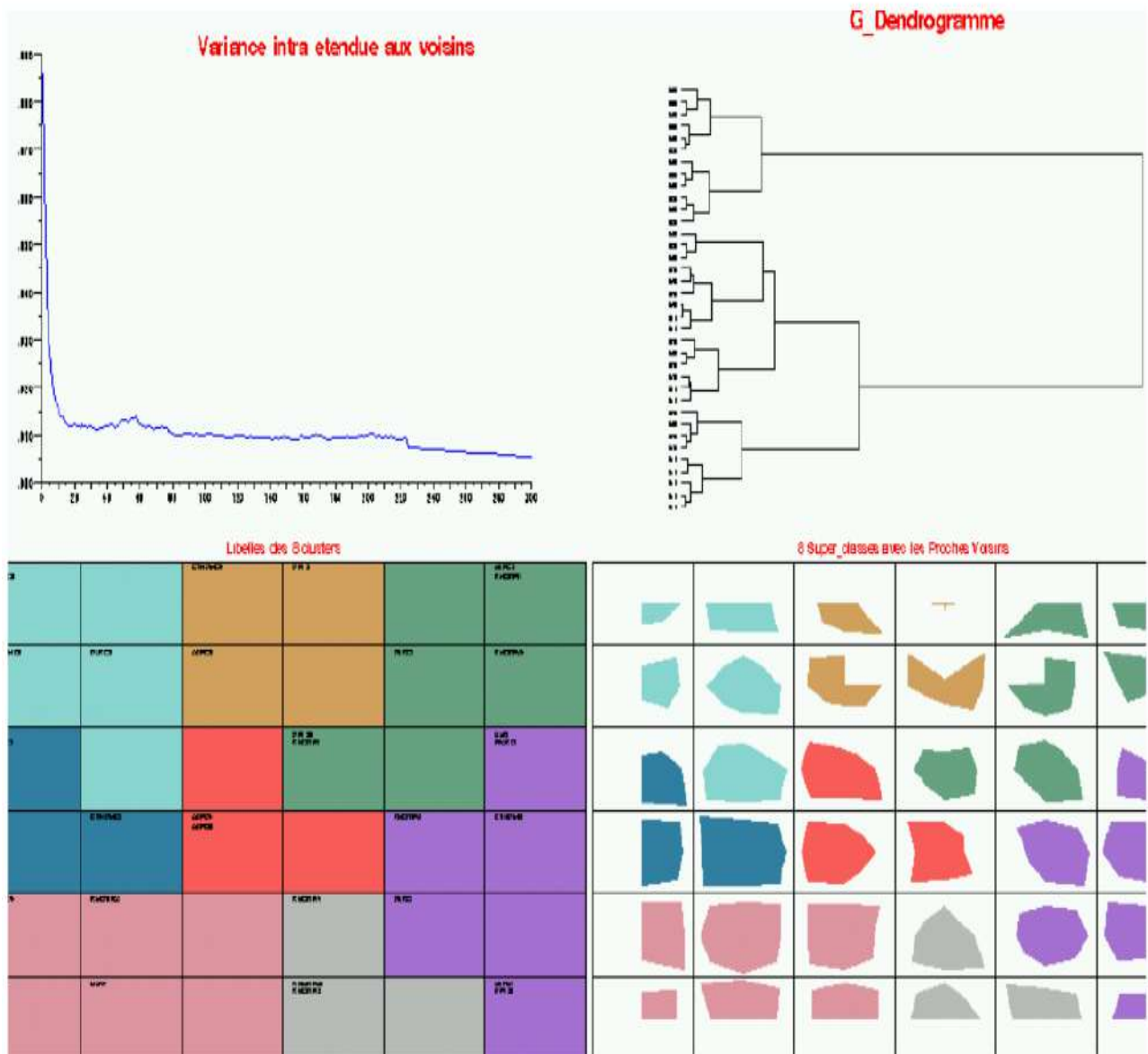


Figure 24: KACM : les modalités

EXTRACTION DE CARACTÈRES et PERCEPTION

RF=reproduire les capacités de l'homme à reconnaître des caractères, des objets, des sons, des signaux temporels, ...

Au départ, la RF est surtout du traitement du signal

- test de la présence d'un signal
- identification de sources multiples
- traitement de la parole

Progressivement, on a envisagé des tâches plus complexes. L'humain fait beaucoup de choses, reconnaître :

- des visages, des sons, des formes, et ceci indépendamment
- du point de vue sous lesquels on les observe
- des conditions d'observation
- de leur variabilité

The collage illustrates various aspects of character and object recognition. On the left, a grid shows the letters A through Z in multiple different fonts and styles. In the center, there is a scanned check from 'LA POSTE' with handwritten text, and below it, a form with the name 'David Pety' and address '3816 E. 100th Street Tulsa, OK 74137'. To the right of these documents are several handwritten labels, mostly 'titulaire', some in different colors and orientations. Below the documents is a 4x4 grid of face images, with the image in the second row, third column highlighted with a red border. To the left of the face grid are a fingerprint, a person wearing sunglasses, and a noisy, textured image.

Caractériser les données

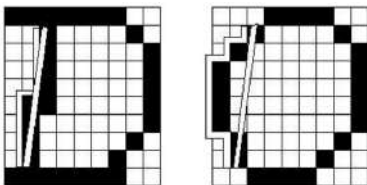
Objectif : Extraire les caractéristiques de ces données pour réduire la dimension du problème.

Une caractéristique est un trait distinctif de l'objet. Il peut être symbolique (couleur par ex.) ou numérique (ex : taille). La combinaison de ces caractéristiques forme un vecteur de dimension d (d étant le nombre de caractéristiques). L'espace de dimension d contenant les vecteurs de caractéristiques est appelé **espace de représentation**.

Une forme est un ensemble de traits de caractéristiques d'une observation donnée. Elle est composée d'un vecteur de caractéristiques et d'un label.

Deux objectifs:

- **réduire la taille du problème** en décrivant la donnée par quelques caractéristiques
- que ces caractéristiques soient discriminantes le plus possible



Quelles caractéristiques ?

caractéristiques possibles: aire, périmètre, compacité, histogramme

Décrire la forme (bi/tri-dimensionnelle) au moyen d'un ensemble de paramètres adéquats
= attributs

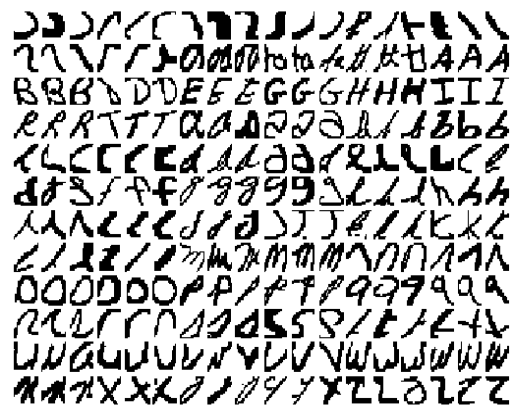
Large variété de représentations suivant les caractéristiques utilisées:

- Caractéristiques globales:
 - Surface, périmètre, largeur, hauteur
 - Élongation, circularité
 - Moments statistiques
- Caractéristiques locales:
 - Coins ou sommets (nombre, positions relatives ou absolues, angles, etc.)
 - Segments (nombre, positions relatives ou absolues, longueur, etc.) Les caractéristiques peuvent être extraites sur:
 - la forme elle-même
 - son squelette
 - ses contours

Les caractéristiques peuvent être extraites sur:

- la forme elle-même
- son squelette
- ses contours

0	0	0	0	0	C	σ	α	ω	0
1	1	1	1	1	z				
2	2	2	2	2					
3	3	3	3	3					
4	4	4	4	4					
5	5	5	5	5					
6	6	6	6	6					
7	7	7	7	7					
8	8	8	8	8	8	8			
9	9	9	9	9					
F	F	F	F	F					



Extraction et caractérisation de primitives

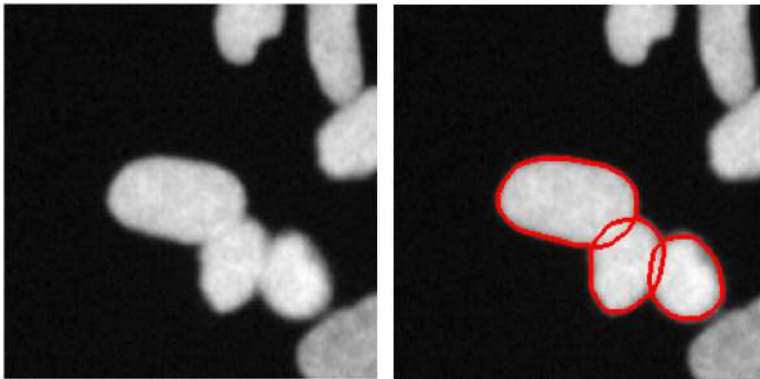
Qu'est ce qu'une primitive dans l'image ? pixel (approche dite dense), contour, segment issu de la détection de contour Les primitives doivent

- porter de l'information (Description exploitable)
- leur extraction doit être "répétable"
- précise, bien localisée



La primitive contour est très utilisée en image → Région d'intérêt (RoI)

Definition 8 (**Contour**) *discontinuité dans une direction de la fonction d'intensité*



2 étapes clefs :

- Détection de la primitive (feature)
- Description, caractérisation de la primitive

Images binaires

- Codage de contours
- Détection de lignes par la Transformée de Hough

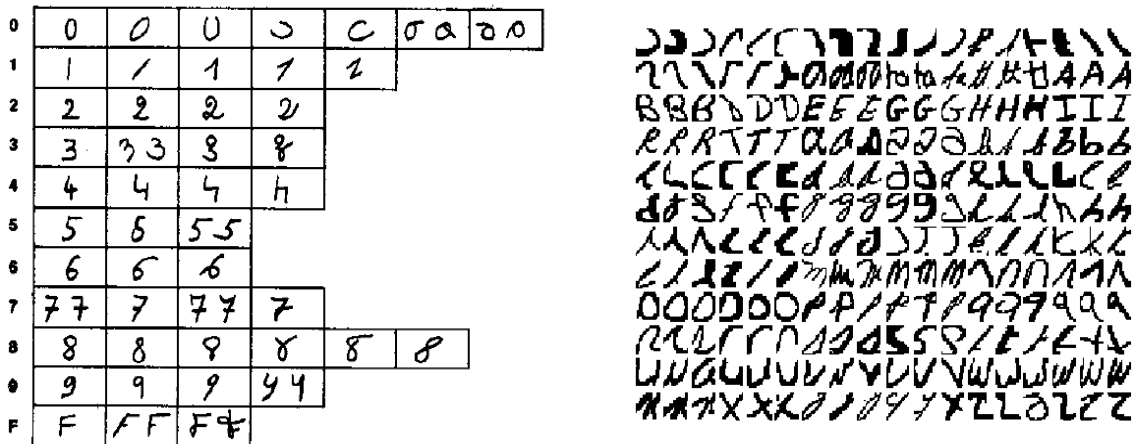
Images NdG : caractérisation de texture

- Matrice de cooccurrences
- Local binary patterns
- Filtres de Gabor

Contraintes:

- faible variance intra-classe
- grande variance inter-classe
- indépendance en translation, rotation et facteur d'échelle
- faible nombre d'attributs

Variabilité inter-classe vs variabilité intra-classe?



Quelques exemples :

- Rapport isopérimétrique : $p = \frac{\text{périmètre au carré}}{4\pi \text{surface}}$ ($\geq 1, = 1$ pour un disque)

- Concavité: $c = \frac{\text{périmètre enveloppe convexe}}{\text{périmètre objet}}$

- Moments: $M_{m,n} = \int \int x^m y^n f(x,y) dx dy$, $M_{m,n} = \sum_{(x,y \in \mathcal{S})} x^m y^n$

ordre 0: $M_{0,0}$ = surface de l'objet

ordre 1: centre de gravité de l'objet $\bar{x} = \frac{M_{1,0}}{M_{0,0}}$, $\bar{y} = \frac{M_{0,1}}{M_{0,0}}$

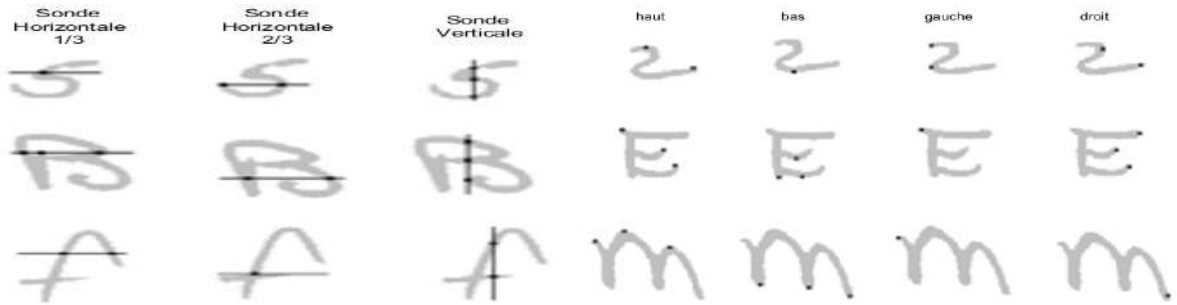
ordre 2: paramètres de l'ellipsoïde d'inertie

$$\theta = \frac{1}{2} \text{atan} \frac{2M_{1,1}}{M_{2,0} - M_{0,2}}$$

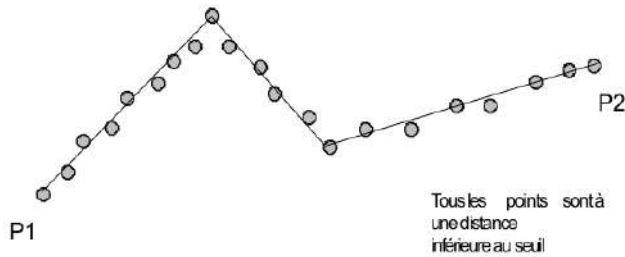
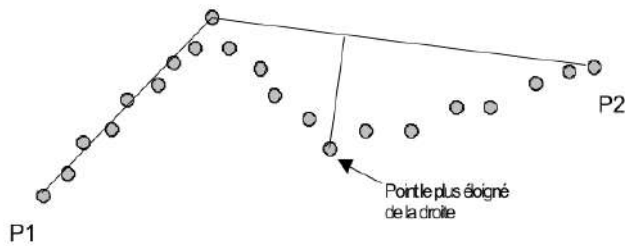
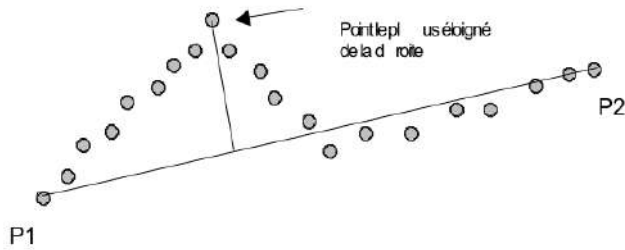
Faire preuve d'imagination mais faire aussi un peu de biblio!!!



Exemples en reconnaissance de caractères:



Tout se joue dans l'espace de caractéristiques!!!
 Le processus de RdF est un processus de réduction progressive et sélective de l'information



$C = \{c_i, i = 0, 1, \dots, n\}$ l'ensemble des points d'une chaîne

c_i a 2 voisins qui sont c_{i-1} et c_{i+1}

S une partition de C et M un sous-ensemble de C contenant les points de cassure. On a

$$S = \{s_0, s_1, \dots, s_{N-1}\} \text{ et } M = \{m_0, m_1, \dots, m_N\}$$

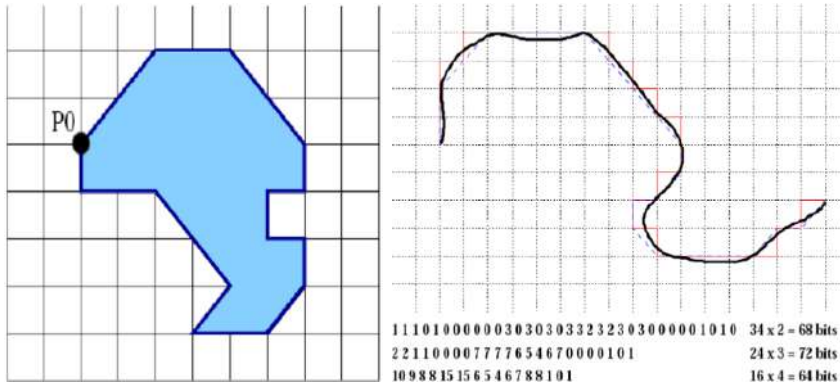
Un segment est ici défini comme

$$S_k = \{c_j | i_k < j < i_{k+1}, c_{i_k} = m_k, c_{i_{k+1}} = m_{k+1}\}$$

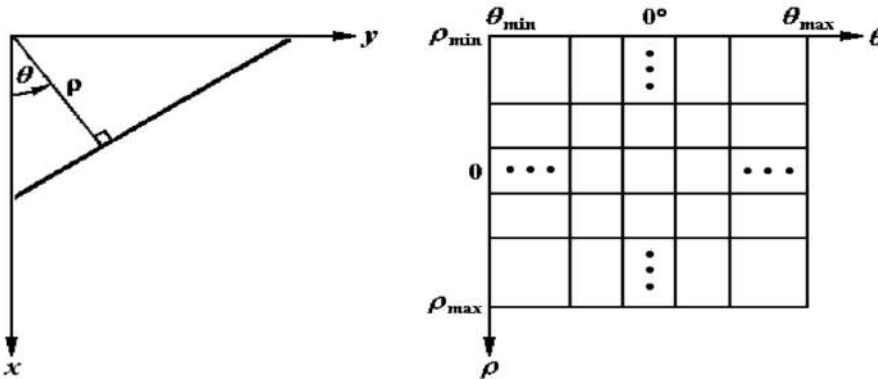
Figure 25: chaînage des contours : Illustration du découpage récursif et calcul des distances

Comment coder le contour ?

- A partir du codage de Freeman (Histogramme des orientations)



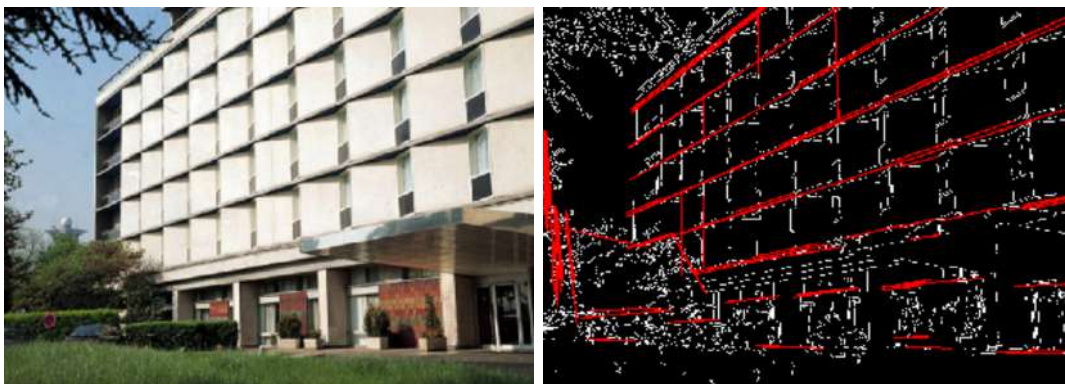
- Transformée de Hough : Détection de lignes dans une image, à partir d'une image binarisée.



$$\rho = x \cos \theta + y \sin \theta$$

$\rho \in [-(H^2 + L^2)^{1/2}, (H^2 + L^2)^{1/2}] \Leftrightarrow$ L'espace paramétré est borné par la hauteur de l'image H , la largeur de l'image L , $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

Parcourir l'espace paramétré et déterminer si le nombre de points colinéaires dépasse un seuil fixé (compteurs)

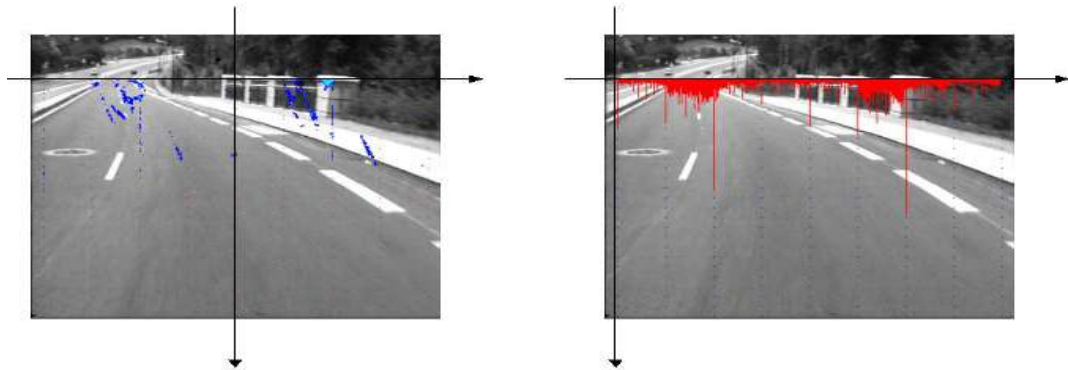


Exemple 28 (Détection et Caractérisation de marquage routier orientée par les données)
(IFFSTAR/Livic)

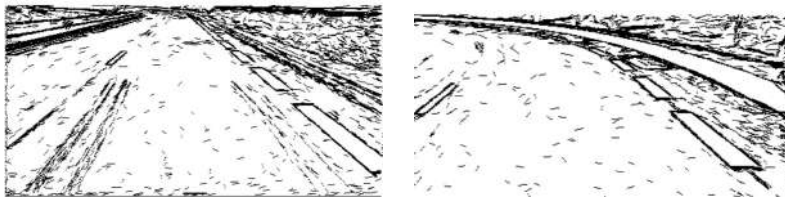
Module de Décision : Où chercher l'information la plus pertinente ?

Module de Reconnaissance : Définir la forme la plus probable de la route

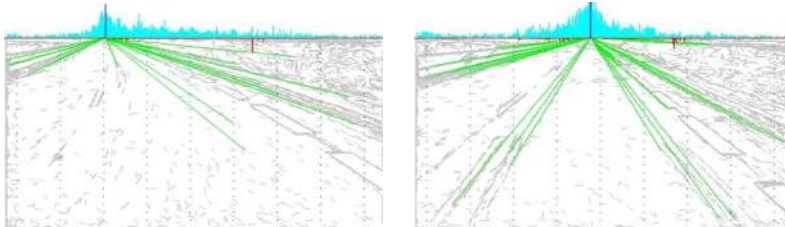
Hypothèse : faible variation angulaire \Rightarrow voies peuvent être décrites par des histogrammes et/ou des nuages de points



Extraction de contour du profil routier par filtre de type "gradient"



construction des histogrammes en ρ, θ (bleu) et reconstruction des droites extraites de l'image (vert)



- Échantillonnage spatial



256×256



128×128



64×64



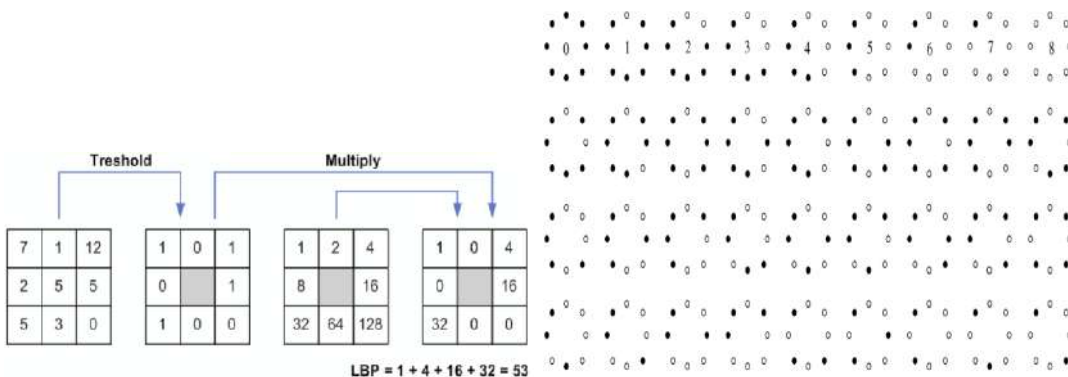
32×32

Échantillonnage et information : tenir compte du contenu informationnel pertinent de l'image à analyser.



La quantification peut faire apparaître (comme ici) des distortions dans les images. Comme pour l'échantillonnage (Shannon), il existe des règles pour déterminer la bonne quantification (le bon nombre de bits) pour coder les images numériques.

- Texture par Local Binary Patterns (LBP)

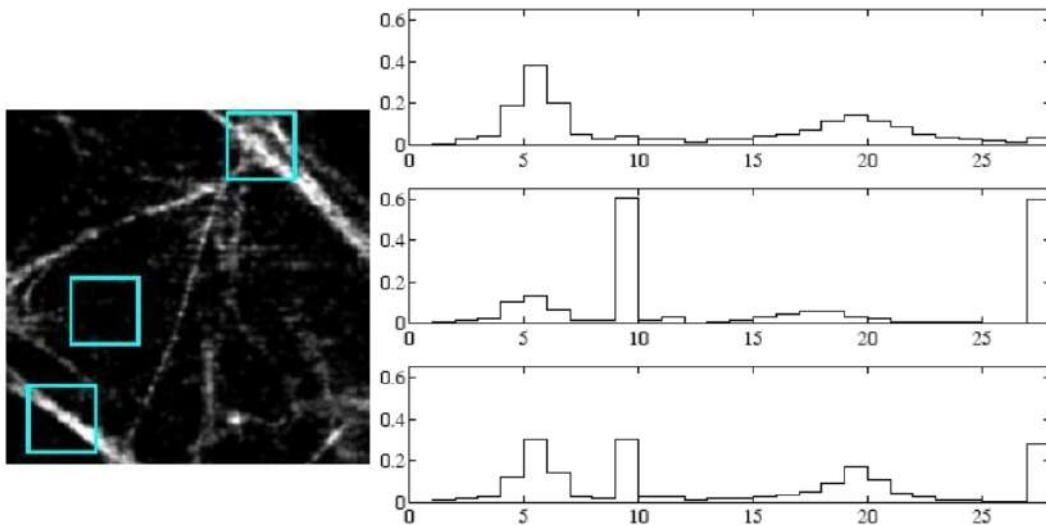


$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p g_c) 2^p$$

Gray-scale and rotation invariant

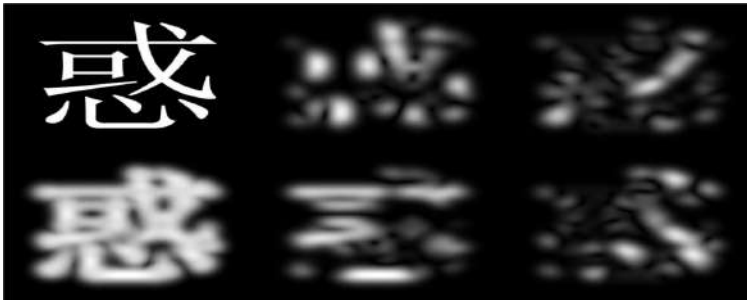
- Pour chaque pixel, on calcule son “pattern” correspondant.
- Pour la résolution P , il y a $P + 2$ patterns possibles.
- Pour chacun des $P + 2$ patterns, on compte le nombre de pixels ayant ce pattern (distribution).
- L'image est caractérisée par l'histogramme des patterns.

Exemple 29 (*Système basé-texture pour la discrimination de cas pathologiques*)

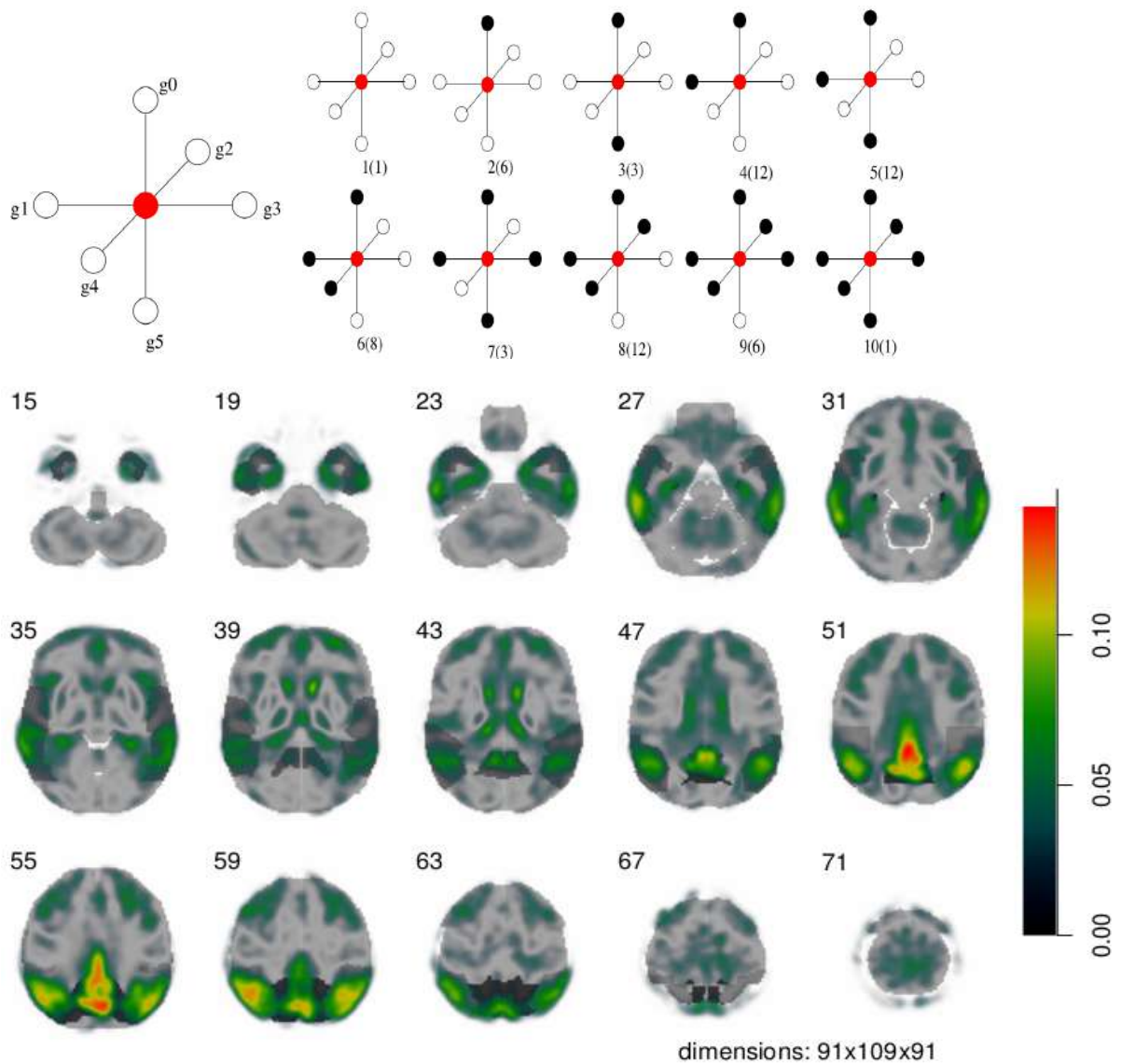


De nombreux outils pour les caractériser : Stat du 1er ordre, matrice de cooccurrence (Haralick), Filtres de Gabor

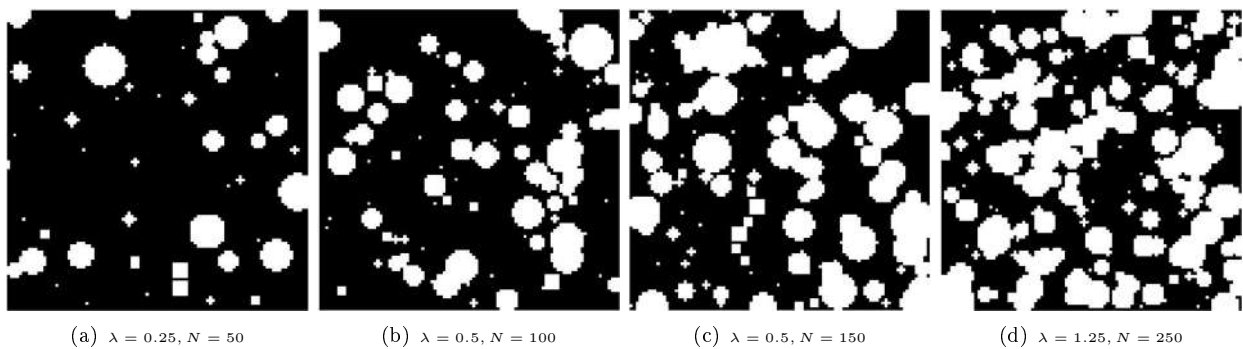
- Filtres de Gabor

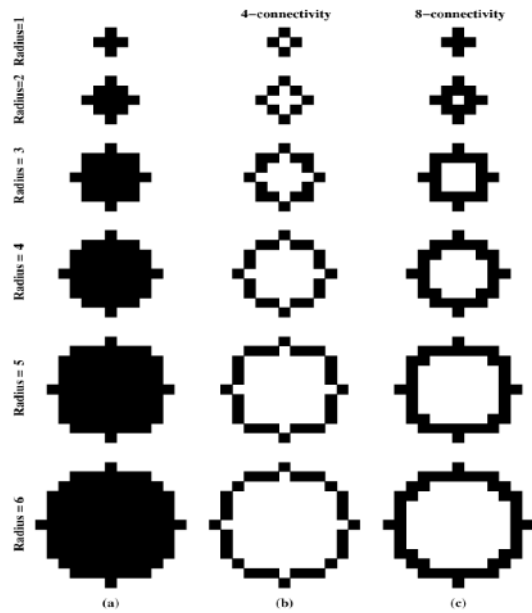


- LBP 3D : regroupement des 2^6 motifs en 10 groupes; ce (chiffre) indique le nombre de motifs différents dans un groupe donné.



- 4 et 8-connectivité. Processus Booléen 2D ? un processus de position de la graine ($\mathcal{P}(\lambda)$) et un processus de forme créant la forme ($\mathcal{N}(r_0, \sigma)$)

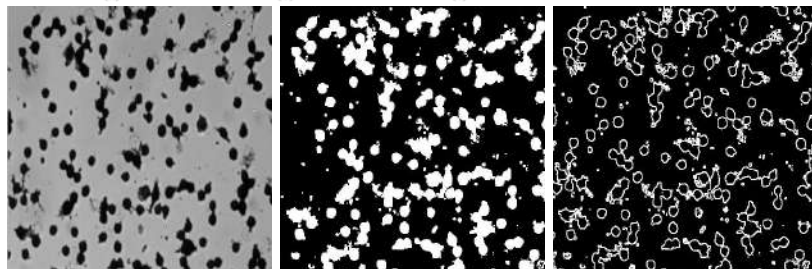




$$\begin{aligned}
 f &= 1 - e^{-\lambda A} \\
 L &= \lambda(1 - f)\Pi \\
 C &= \lambda(1 - f)
 \end{aligned}$$

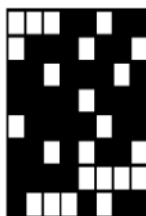
L : estimée par le nombre de pixels de bord.

paramètres		exemple	formule
fraction d'aire	f	0.38	$\#b/n$
longueur tour/unité de surf.	con- L	0.36	$\frac{\pi}{2n} \sum_i^n P(\beta_i)$
convexité spécif.	C	0.06	$\#$ hit and miss



(e) Original cell image (f) The boundary pixels identified in the image (g) realisation of the 2D Boolean model

- Comment décrire une matrice de 1/0 par un modèle booléen 1D ?



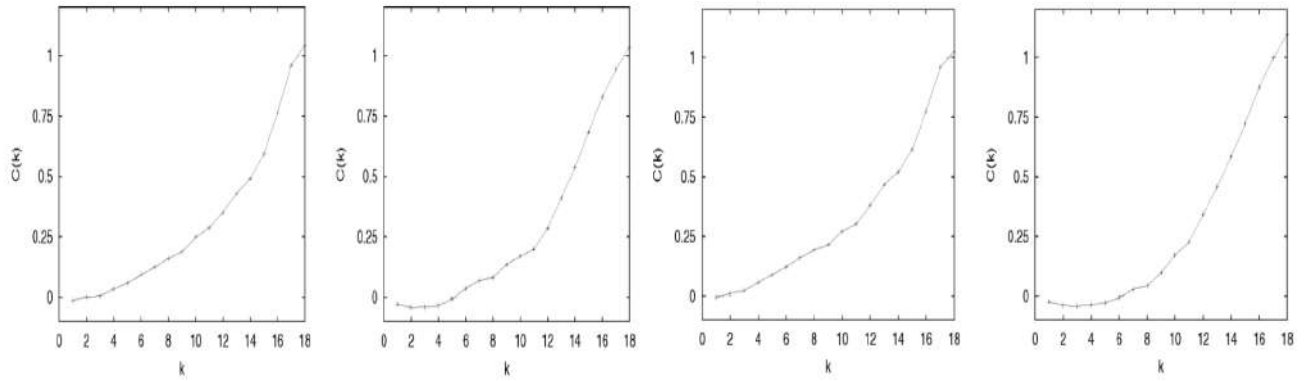
N_p : the number of white points in the sequence

N_{ps} : the number of white segments in the sequence;

$S(k)$: the number of pairs of white pixels that are k pixels

$$p = \frac{N_{ps}}{N_p}$$

$$C(k) = 1 - \frac{1}{p} \left(1 - \frac{S(k+1)}{S(k)} \right)$$



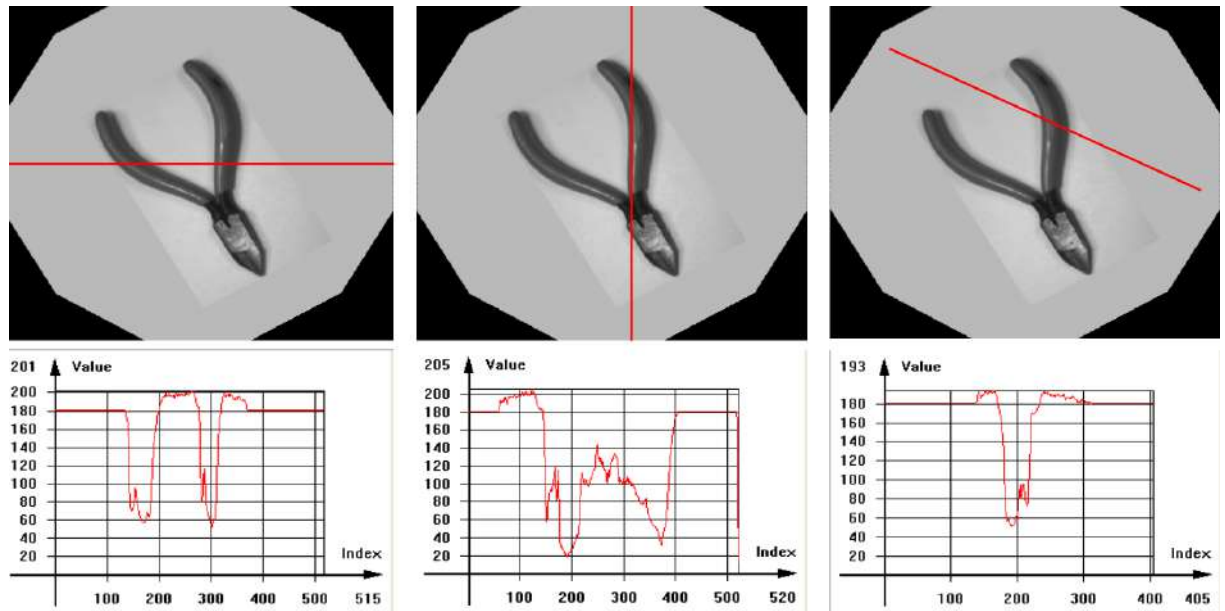
Analyse statistique des images

Structure de données pour une image $I(M, N)$, M lignes, N colonnes

- Exemple 1: $M = N = 512$ pixels \rightarrow image carrée de 262.144 pixels
 $k=8 \rightarrow K = 256$ niveaux de gris $\in [0; 255]$ (0 = noir, 255 = blanc)
 taille de fichier informatique non compressé ≈ 2 Mb
- Exemple 2 : $M = N = 512$ pixels \rightarrow image carrée de 262.144 pixels
 $k = 3 \times 8 = 24 \rightarrow K = 224 = 16.777.216$ couleurs
 un pixel est codé sur 3 valeurs : $R \in [0; 255]$ $V \in [0; 255]$ $B \in [0; 255]$
 taille de fichier informatique non compressé ≈ 6 Mb



- Moyenne des niveaux de gris : $m = \frac{1}{P} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i, j)$
- Variance des niveaux de gris : $\sigma^2 = \frac{1}{P} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i, j) - m)^2$
- Rapport signal sur bruit (SNR), traduit la dispersion des niveaux de gris dans l'image :
 $SNR = \frac{m}{\sigma}$
- Extremas (niveaux min et max) et étendue (max-min)
- Normalisation des données :
 - Modification des données brutes permettant de comparer des images entre elles
 - En les centrant (mêmes moyennes ≈ 0) : $I'(i, j) = I(i, j) - m$
 - En les centrant et en les réduisant : $I'(i, j) = \frac{I(i, j) - m}{\sigma}$



- les images suivantes ont le même nombre de pixels noirs et blancs?

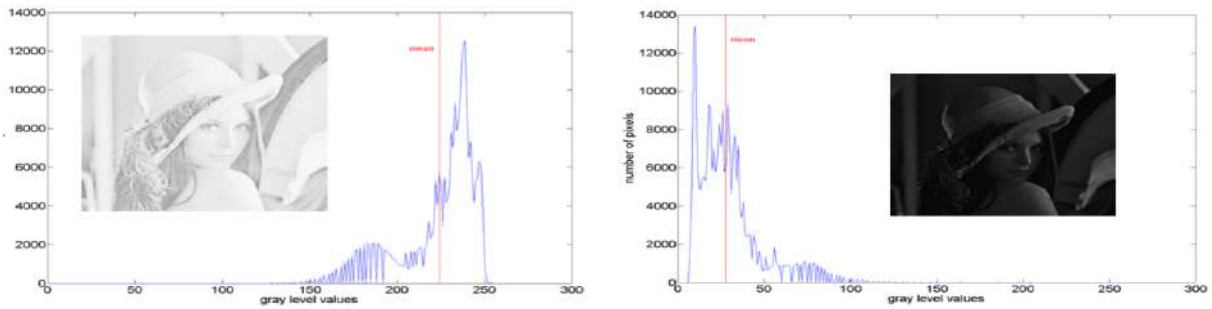


Les images précédentes ont le même histogramme

- Histogramme des niveaux $H(n)$ = Répartition de la population des pixels appartenant à chacune des composante ou NdG
- Représentation de l'image en termes d'intensité et de contraste $\sum_{n=0}^{2^k-1} H(n) = P$
- Probabilité d'un NdG $p(n) = H(n)/p$ et $\sum_{n=0} p(n) = 1$.
- Quantité d'information : La quantité d'information apportée par un niveau n est inversement proportionnelle à la probabilité de ce niveau

$$I(n) = -\log_2(p(n))$$

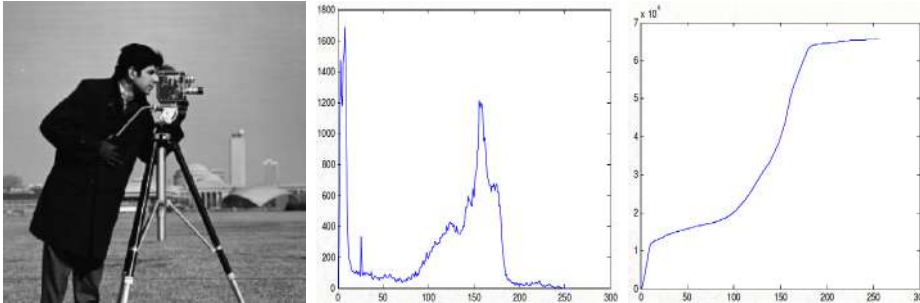
- moyenne = tendance générale des valeurs des pixels



- Entropie : Mesure de la quantité moyenne d'information apportée par chaque niveau de l'image

$$E(I) = - \sum_{n=0}^{2^k-1} p(n) \log_2(p(n))$$

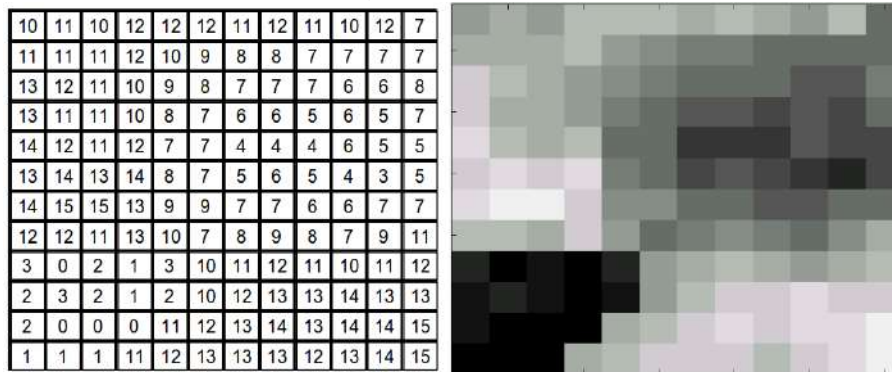
- Histogramme cumulé



- Histogramme en x et en y : sommer tous les niveaux de gris des pixels présents sur une même ligne ou une même colonne et ceci pour toutes les lignes ou colonnes de l'image.

$$H_{\text{ligne}}(i) = \sum_{j=0}^{N-1} I(i, j) \quad H_{\text{col}}(j) = \sum_{i=0}^{M-1} I(i, j), \forall i \in [0, M-1], \forall j \in [0, N-1],$$

Exercice 19 Soit : $M = 12, N = 12 \rightarrow P = 144, k = 4$

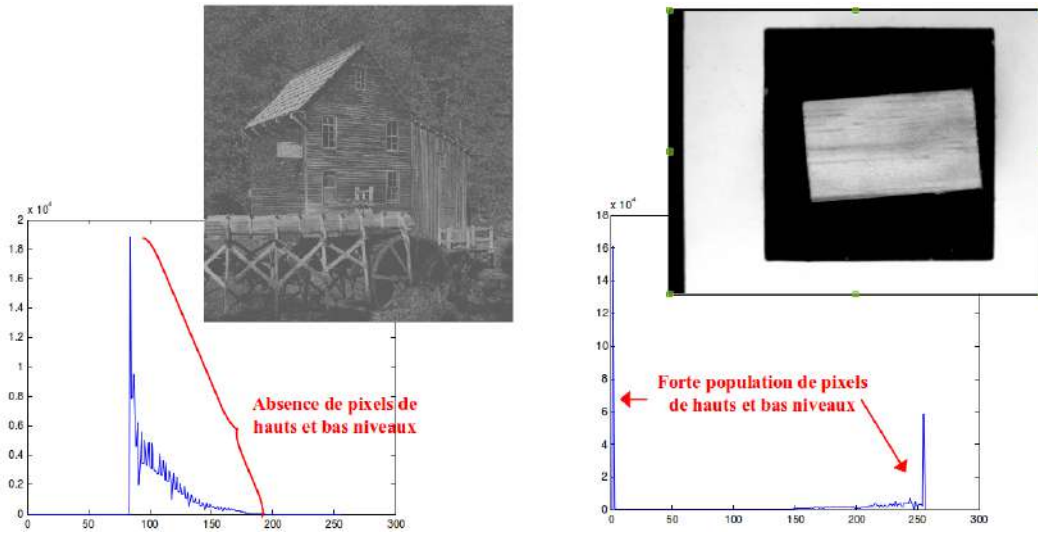


Calculer et proposer un code/algorithmme pour

- Valeur moyenne, mode, médiane, et écart-type

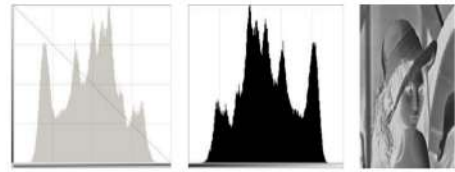
- Valeur min et max
- Histogramme en ligne et colonne
- Histogramme et histogramme cumulé

- Exemples d'image à "faible" et "fort" contraste :

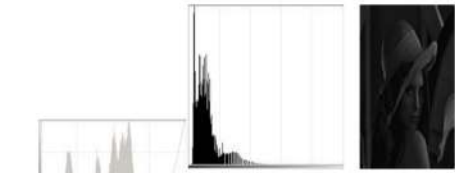


solution : Modification de contraste= "contrast stretching"

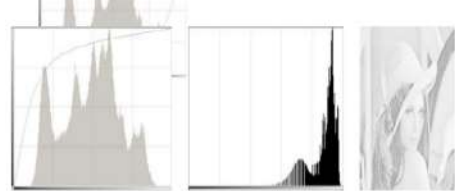
– Négation



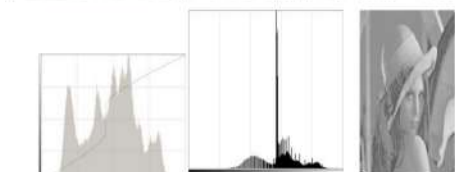
– Éclaircissement



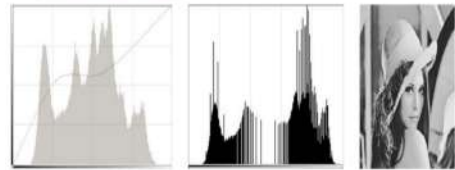
– Assombrissement



– Diminution du contraste :



– Augmentation du contraste



- Seuillage = Diminution du nombre de NdG d'une image pour mettre en évidence des régions et simplifier l'information
binarisation (seuillage simple donnant une image binaire)

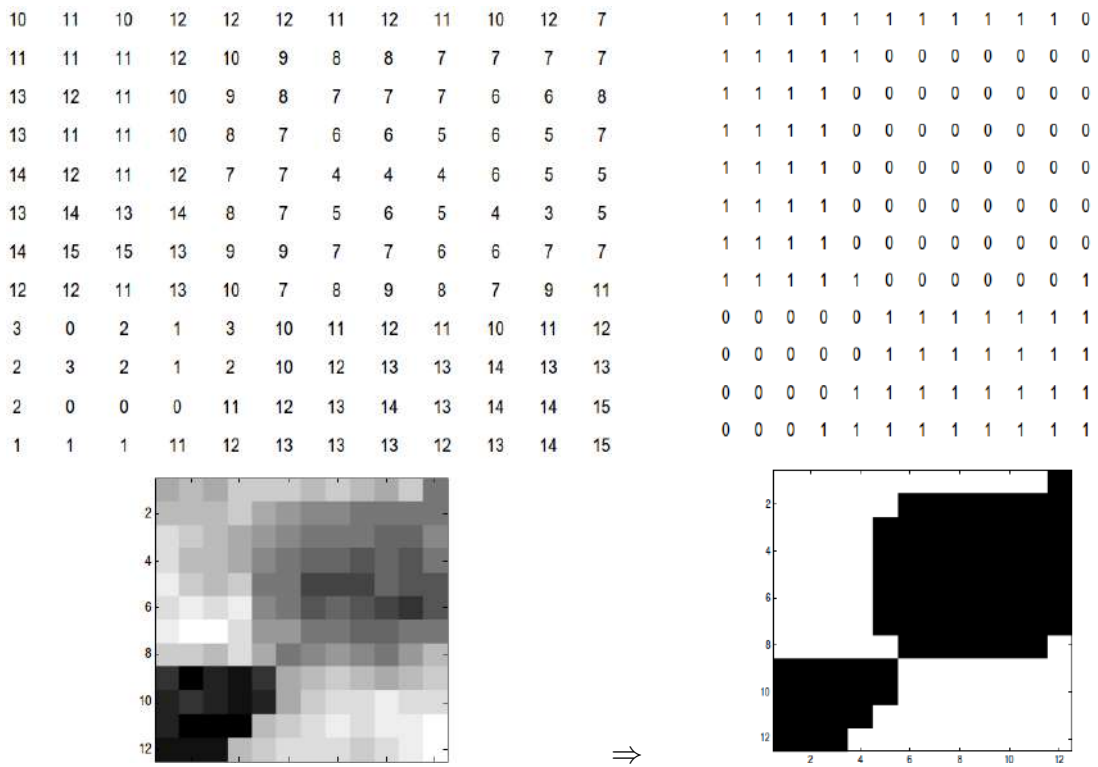
Exemple de seuillage automatique : maximisation de la variance interclasses

Soit une image de moyenne m et de population P , s_0 le seuil optimal recherché, $s \in [0; K - 1]$ un seuil courant.

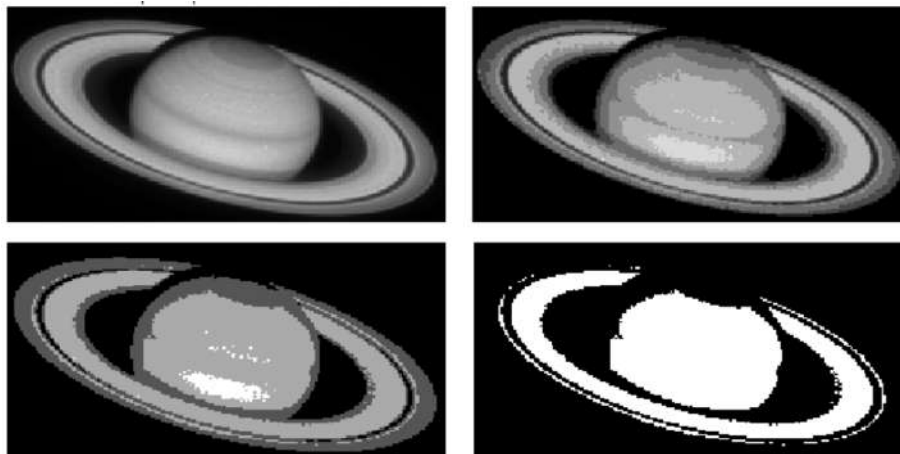
A chaque valeur de s , on associe deux classes C_0^s et C_1^s de moyenne et de population $\{m_0^s; \sigma_0^{2s}\}$ et $\{m_1^s; \sigma_1^{2s}\}$ resp. A chaque seuil s , on calcule la variance interclasses V^s définie par $V(x) = p(C_0^s)(m_0^s - m)^2 + p(C_1^s)(m_1^s - m)^2$ avec $p(C_0^s) = \sigma_0^{2s}/\sigma^2$ et $p(C_1^s) = \sigma_1^{2s}/\sigma^2$.

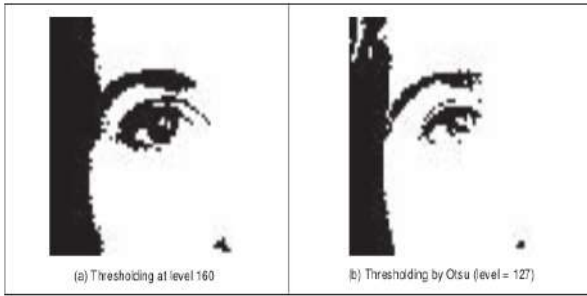
On sélectionne s_0 tel que $V(s_0) = \max_s \{V(s)\}$ pour $s \in [0; K - 1]$.

Pas très bien adaptée pour des classes à écarts-types très différents, mais généralisable à N classes



Exemple 30 (Image à 16, 8, 4 et 2 niveaux) La réduction du nombre de niveaux est obtenue par arrondis successifs ; l'image à 2 niveaux correspond à un seuillage $\theta = \frac{1}{2}$.



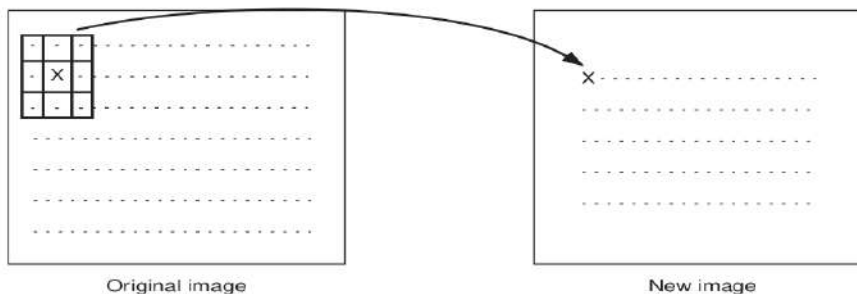


- L'atténuation de bruit → superpose N images de la même source, *i.e.* addition d'images discrètes, point par point
- Opérateur de lissage / Filtrage spatial

$$M = \frac{1}{4} \sum_{|i-k|+|j-\ell| \leq 1} I(i, j), \quad (139)$$

⇒ *effet de flou* qui en résulte peut être limité, et les contrastes conservés, par d'autres variantes, par exemple $I(i, j) = \frac{1}{2}(I(i, j) + M)$ ou par : si $|I(i, j) - M| \geq \epsilon$ alors M sinon $I(i, j)$. Les nouvelles valeurs doivent être arrondies pour entrer dans la même gamme que les valeurs de l'image originale. Par exemple, pour des images en noir (1) et blanc (0), on choisira comme nouvelle valeur de $I(k, \ell)$: si $\alpha I(k, \ell) + \beta M \geq 1$ alors 1 sinon 0.

Exemple 31 (Lissage avec une moyenne sur un voisinage carré de côté 3, 5, 9)
Un effet similaire peut être obtenu en itérant l'opérateur de lissage de taille 3×3 .



1. Filtrage médian

$$\begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array} \rightarrow 10 \ 10 \ 20 \ 20 \ \boxed{25} \ 25 \ 30 \ 30 \ 250$$

2. Masques du **gradient** $G_x = \frac{1}{2} \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$ $G_y = \frac{1}{2} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

3. Masques de **Prewitt et Sobel** : dérivatif du 1 ordre

$$P_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad P_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

4. Masques de **Freeman**

$$H_0 = \begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix} \quad H_1 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix} \quad H_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix} \quad H_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix}$$

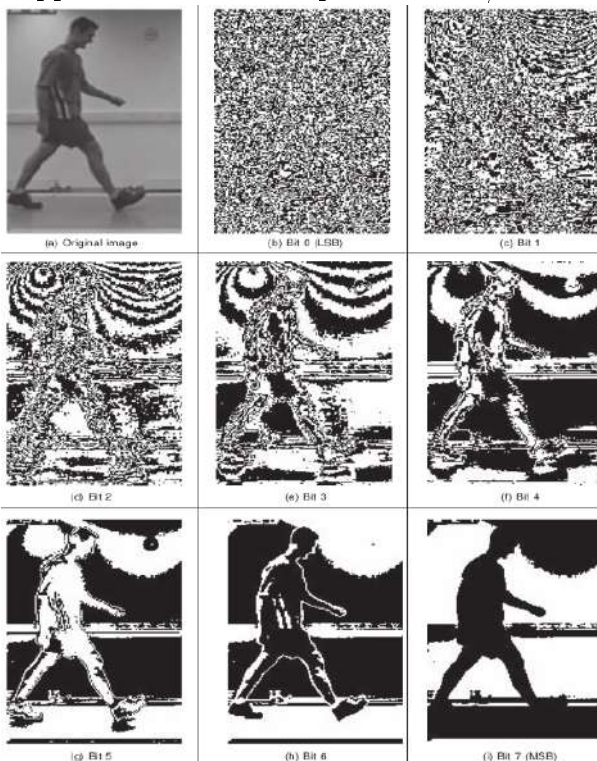
$$H_4 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{bmatrix} \quad H_5 = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{bmatrix} \quad H_6 = \begin{bmatrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad H_7 = \begin{bmatrix} -1 & -1 & 1 \\ -1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

De même, les 8 masques de Freeman sont normalement utilisés pour extraire la direction maximale de contraste au voisinage du point considéré (codée à l'aide d'une valeur comprise entre 0 et 7).

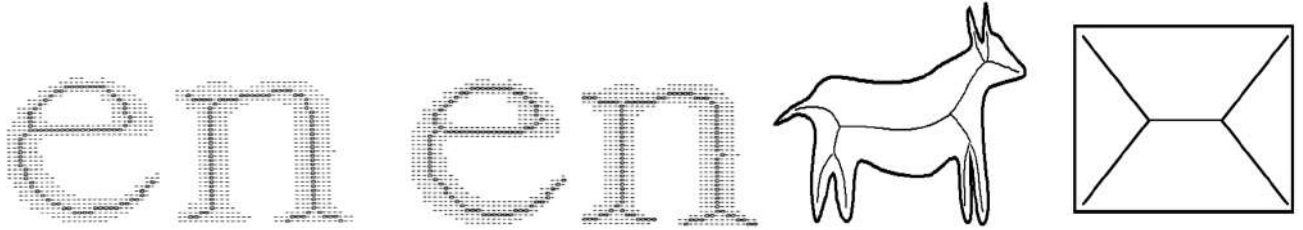
5. Masques du **Laplacien et de la moyenne** (dérivatif de 2nd ordre)

$$L = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad L = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad L = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \quad M = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

- Suppression de bits de poids faibles/forts



- **Squelettisation** = amincissement itéré, aboutissant à un domaine mince de même topologie, le squelette (morpho-mathématique)



En pratique, c'est le lieu des collisions de deux fronts d'onde = érosion extrême!

- Couleur = élément fondamental de la perception visuelle
Stimulus d'entrée : conversion de la lumière en un signal électrochimique. Elle est représentée comme la composition de différentes couleurs de base qui sont appelées couleurs primaires

Avantages :

- Amélioration visuelle
- Réduction de l'information inutile (bruit)
- Renforcement de l'information utile

Représentation des couleurs fréquentielle et **spatiale**

Espace à p dimensions (le plus souvent, $p = 3$)

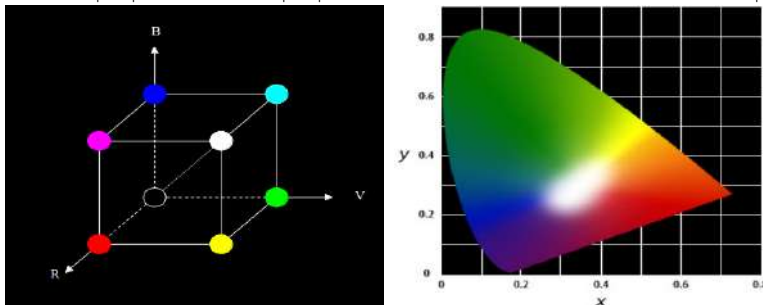
Dimension = composante couleur = axe couleur

Nombreux espaces correspondants à des préoccupations différentes

Passage d'un espace à un autre : transformation linéaire ou non

- Systèmes XYZ (espace normalisé xyz)
$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 2.769 & 1.751 & 1.130 \\ 1.0 & 4.59 & 0.06 \\ 0.0 & 0.056 & 5.594 \end{pmatrix} \begin{pmatrix} R \\ V \\ B \end{pmatrix} \text{ et}$$

$$x = \frac{X}{X+Y+Z} \quad y = \frac{Y}{X+Y+Z} \quad (\text{comp. chromatiques}) \quad z = \frac{Z}{X+Y+Z} \quad (\text{comp. de luminance})$$



- Systèmes HSI-HSL = Hue(teinte)/Saturation/Luminance
Hue = longueur d'onde dominante
Saturation = Distance au niveau de gris ayant la même intensité
Luminance = intensité de lumière

– Système $I_1I_2I_3$ calculé pour représenter au mieux le système visuel humain

$$\begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ -1/4 & 1/2 & -1/4 \end{pmatrix} \begin{pmatrix} R \\ V \\ B \end{pmatrix}$$



original

HSL

$I_1I_2I_3$

DÉCISION STATISTIQUE

Représenter les classes sous forme de densités de probabilités.

Chaque groupe (classe) est associée à un nuage de points.

La classification permet d'assigner un objet à l'une des K classes possibles $w_i, i = 1, \dots, K$ en se basant sur la mesure \mathbf{x} avec $\sum_{i=1}^K w_i = 1$.

Comment maximiser la probabilité d'appartenance à une classe ?

Probabilité d'appartenance à un des deux groupes

$$f_1(x) = p(x|w_1) = \frac{1}{(2\pi)^{0.5}\sigma_1} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}} \quad \text{et} \quad f_2(x) = p(x|w_2) = \frac{1}{(2\pi)^{0.5}\sigma_2} e^{-\frac{(x-m_2)^2}{2\sigma_2^2}}$$

Probabilité d'erreur

Pour une caractéristique observée x , la probabilité d'erreur d'une décision $d(x) = w_i$ est égale à

$$P(\text{erreur}|x) = \sum_{j \neq i} p(w_j|x) = 1 - p(w_i|x)$$

La probabilité d'erreur globale pour le système est

$$P(\text{erreur}) = \int_{-\infty}^{\infty} P(\text{erreur}|x)p(x)dx$$

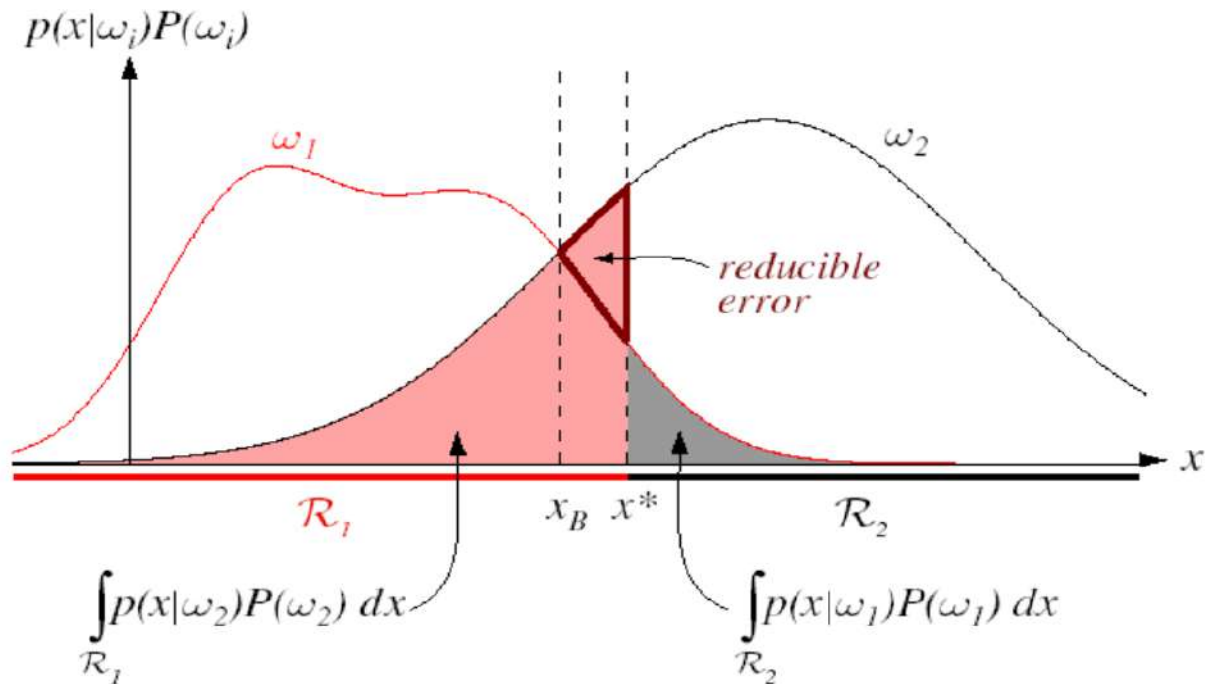
La décision optimale (au sens de la probabilité d'erreur) consiste donc à choisir $d(x) = w_i$ telle que $p(w_i|x)$ soit maximale

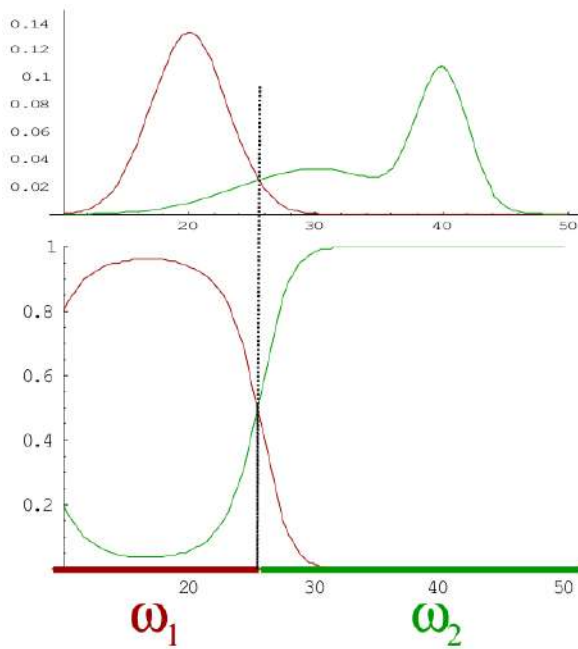
$$p(w_i|x) \geq p(w_j|x) \forall j \quad (140)$$

$$p(x|w_i)p(w_i) \geq p(x|w_j)p(w_j) \forall j \quad (141)$$

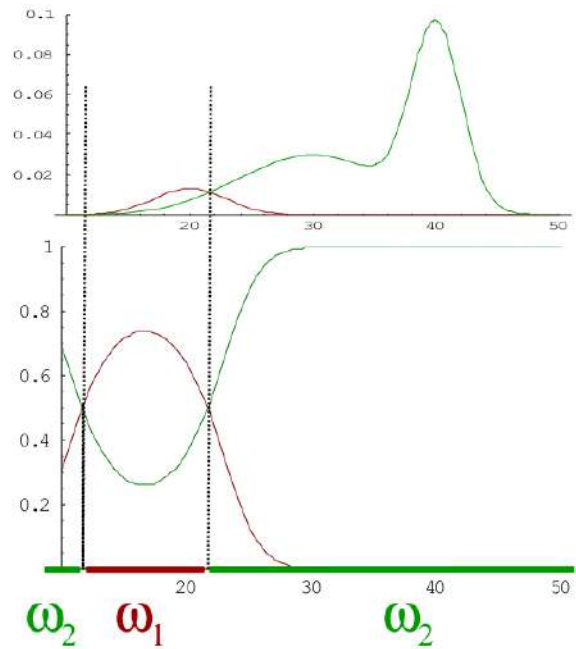
Si les probabilités *a priori* des classes sont équivalentes

$$p(x|w_i) \geq p(x|w_j) \forall j$$



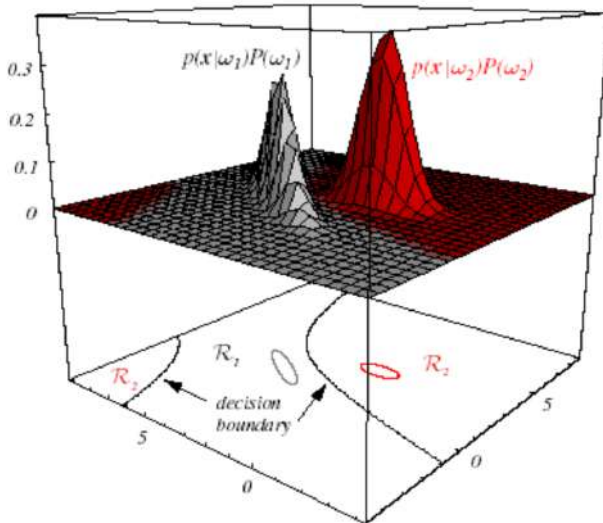


$$P(\omega_1)=0.5, P(\omega_2)=0.5$$



$$P(\omega_1)=0.1, P(\omega_2)=0.9$$

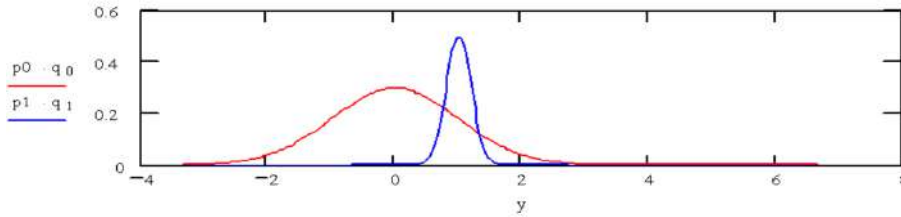
La règle de Bayes se généralise au cas où les caractéristiques sont représentées sous forme de vecteur \mathbf{x} :



$$P(w_i|\mathbf{x}) = \frac{P(\mathbf{x}|w_i)P(w_i)}{p(\mathbf{x})}$$

avec $p(\mathbf{x}) = \sum_i P(\mathbf{x}|w_i)P(w_i)$.

Exemple 32 (Deux lois de probabilités conditionnelles Gaussiennes) $m_0 = 0, \sigma_0 = 1, m_1 = 1, \sigma_1 = 0.2, q_0 = 3/4, q_1 = 1/4$



choix des valeurs des pénalités : $\frac{c(1|0)-c(0|0)}{c(0|1)-c(1|1)} = 2$. Pour minimiser le critère, on choisira l'hypothèse $u = 1$ si y est dans l'intervalle $[0.87; 1.21]$, et $u = 0$ si $y \notin [0.87; 1.21]$.

probabilité de détection correcte

$$p(d = 1|u = 1) = \int_{0.87}^{-1.21} \frac{1}{\sigma_1(2\pi)^{0.5}} e^{-\frac{(y-m_1)^2}{2\sigma_1^2}} dy = .595$$

probabilité d'erreur (fausse alarme)

$$p(d = 1|u = 0) = \int_{0.87}^{-1.21} \frac{1}{\sigma_0(2\pi)^{0.5}} e^{-\frac{(y-m_0)^2}{2\sigma_0^2}} dy = .079$$

Exercice 20 Soit $\Omega = \{w_1, w_2\}$. $p(x|w_1) \sim \mathcal{N}(2; 0.5)$ et $p(x|w_2) \sim \mathcal{N}(1.5; 0.2)$.
 $p(w_1) = 2/3$ et $p(w_2) = 1/3$ et $c = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

Minimisation du risque pour deux classes

Soit $c(d_i|w_j)$ le coût engendré par la décision d_i lorsque l'objet appartient effectivement à la classe w_j . La décision d_i correspond à choisir la classe w_i . la probabilité d'erreur correspond au cas particulier où

$$c(d_i|w_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

Soit $c(i|j) = c(d_i|w_j)$.

Le problème de la décision à 2 classes :

- on a mesuré y ; choisir parmi les deux propositions ($d = 0$ et $d = 1$):
 $\boxed{y \text{ a été causé par } u = 0 \text{ ou } y \text{ a été causé par } u = 1}$;
- il faut se donner une fonction de pénalité (4 coûts associés aux cas possibles) :
 $c(0|0)$ quand on choisit $d = 0$ et que la vraie valeur est $u = 0$
 $c(0|1)$ quand on choisit $d = 0$ et que la vraie valeur est $u = 1$
 $c(1|0)$ quand on choisit $d = 1$ et que la vraie valeur est $u = 0$
 $c(1|1)$ quand on choisit $d = 1$ et que la vraie valeur est $u = 1$
- Pour une valeur de y mesurée, on choisira $d = 0$ si le coût associé à ce choix est moins élevé que le coût associé au choix $d = 1$

- coûts moyens associés aux décisions

$d = 0$	$d = 1$
$c(0 0)p(y u = 0).q_0 + c(0 1)p(y u = 1).q_1$	$c(1 0)p(y u = 0).q_0 + c(1 1)p(y u = 1).q_1$

avec $q_0 = p(u = 0)$ et $q_1 = p(u = 1)$.

- **On choisit $d = 0$ si, en moyenne, cela coûte moins que de choisir $d = 1$**

$$c(0|0)p(y|u = 0).q_0 + c(0|1)p(y|u = 1).q_1 < c(1|0)p(y|u = 0).q_0 + c(1|1)p(y|u = 1).q_1$$

hypothèse : coûts des mauvaises décisions plus élevés que coûts des décisions correctes
on choisira $d = 0$ lorsque

$$\boxed{\frac{p(y|u = 1) q_1}{p(y|u = 0) q_0} < \frac{c(1|0) - c(0|0)}{c(0|1) - c(1|1)}} \quad (\text{taux de vraisemblance})$$

Autres critères de décision

- La minimisation du risque n'est pas le seul critère de décision possible
- On peut vouloir minimiser l'erreur maximale (minimax) quelque soient les probabilités $P(w_i)$

⇓

Risque minimax

coût est linéaire (tangente pointillée) et risque minimax est atteint avec la tangente horizontale

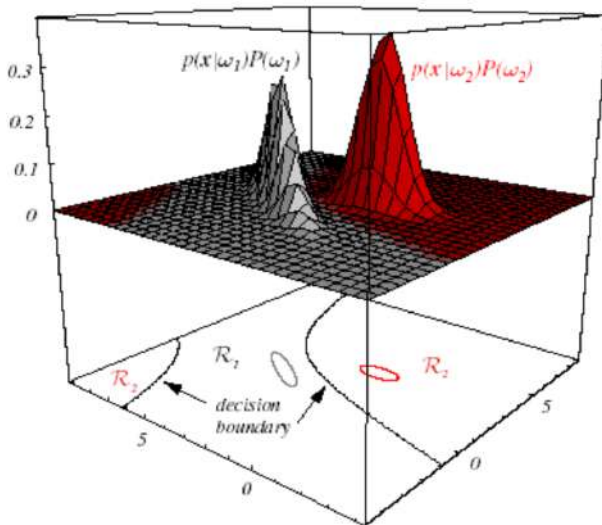


Figure 26: Erreur minimale en fonction de la probabilité a priori $P(w_i)$.

Le risque global

$$R = \int_{R_1} [c_{11}p(x|w_1)p(w_1) + c_{12}p(x|w_2)p(w_2)] dx + \int_{R_2} [c_{21}p(x|w_1)p(w_1) + c_{22}p(x|w_2)p(w_2)] dx$$

Posons $P(w_1) = 1 - P(w_2)$, on peut formuler le risque selon une fonction linéaire

$$R = c_{22} + (c_{12} - c_{22}) \int_{R_1} p(x|w_2) dx + \alpha P(w_1) \quad (142)$$

$$\alpha = (c_{11} - c_{22}) + (c_{21} - c_{11}) \int_{R_2} p(x|w_1) dx - (c_{12} - c_{22}) \int_{R_1} p(x|w_2) dx \quad (143)$$

La solution au problème correspond à $\alpha = 0$

$$R = c_{22} + (c_{12} - c_{22}) \int_{R_1} p(x|w_2) dx = c_{11} + (c_{21} - c_{11}) \int_{R_2} p(x|w_1) dx$$

Fonctions discriminantes

Dans le cas de plusieurs classes un classifieur peut utiliser des fonctions discriminantes $g_i(x)$ et décider

$$d_i \Leftrightarrow g_i(x) > g_j(x), \forall i \neq j$$

Un classifieur Bayésien utilise les fonctions discriminantes

$$g_i(x) = -R(d_i|x)$$

ou toute fonction $f(g_i(x))$ où f est monotone croissante, en particulier

$$g_i(x) = -p(x|w_i)p(w_i) \tag{144}$$

$$g_i(x) = -\ln p(x|w_i) + \ln p(w_i) \tag{145}$$

Principe de la décision Bayésienne

Trouver la classe w_k qui maximise la probabilité a posteriori $P(w_i|x)$ On considère

- un ensemble de K classes notée $\{w_1, w_2, \dots, w_K\}$
- une caractéristique à valeur réelle notée x
- Pour chaque classe w_i on suppose connaître la probabilité a priori de cette classe $p(w_i)$ avec $p(w_i) \geq 0, \forall i$ et $\sum p(w_i) = 1$.

La règle de Bayes permet de calculer la probabilité a posteriori de chaque classe, c'est-à-dire :

$$\boxed{P(w_i|x) = \frac{p(x|w_i)p(w_i)}{p(x)}} \quad \text{où} \quad p(x) = \sum_i p(x|w_i)p(w_i)$$

est appelée **évidence**.

Critère de Neyman-Pearson

probabilités a priori des causes $q_0 = p(u = 0)$ et $q_1 = p(u = 1)$ inconnues.

Décider si une mesure x correspond à l'émission d'une donnée $u = 0$, et dans ce cas la densité de probabilité de x est $p(x|w_1)$ ou si elle correspond à l'émission $u = 1$, et dans ce cas la densité de probabilité de x est $p_1(x|w_2)$;

$$\text{On décidera } w_1 \text{ si } r(x) = \frac{p(x|w_2)}{p(x|w_1)} \underset{d_1}{\overset{d_2}{\gtrless}} \theta$$

Maximiser la probabilité PDC = p_{dc} de détection correcte ($d = 1$ quand $u = 1$) ;

La probabilité de fausse alarme PFA = p_{fa} : probabilité que $u = 0$ alors que $r(x) = \frac{p(x|w_2)}{p(x|w_1)} > \theta$.

Donc PFA = ensemble des valeurs (domaine \mathcal{D}) de x pour lequel ce seuil est dépassé

$$PFA = \int_{\mathcal{D}} p(x|w_1) dx$$

Exemple 33 les probabilités a priori des causes ne sont pas prises en compte ;

si le seuil θ est choisi égal à 2, on décide $d = 1$ lorsque x est dans l'interv. (0.7, 1.38).

probabilité de fausse alarme

$$p_{fa} = \int_{0.7}^{1.38} \frac{1}{\sigma_0(2\pi)^{0.5}} e^{-\frac{x^2}{2\sigma_0^2}} dy = .158$$

probabilité de détection correcte

$$\int_{0.7}^{1.38} \frac{1}{\sigma_1(2\pi)^{0.5}} e^{-\frac{(y-m_1)^2}{2\sigma_1^2}} dy = .904$$

Neyman Pearson : on se donne un seuil α que cette probabilité de fausse alarme p_{fa} ne doit pas dépasser et on en déduit le seuil θ utilisé dans la décision

Dans l'intervalle $[x_{\min}, x_{\max}]$: $r(x) > \theta$

Comment ajuster θ et par conséquent les bornes x_{\min} et x_{\max} pour maximiser la probabilité de décision correcte tout en assurant que la probabilité de fausse alarme :

$$p_{fa} = \int_{x_{\min}}^{x_{\max}} p(x|w_1) dx$$

ne dépasse pas le seuil α

Exemple 34 $m_0 = 0, \sigma_0 = 1, m_1 = 1, \sigma_1 = 0.2$.

si θ est fixé : calculer les valeurs x_{\min} et x_{\max} entre lesquelles on décidera $d = 1$.

Le dépassement du seuil :

$$\frac{\sigma_0}{\sigma_1} \exp\left(\frac{x^2}{2\sigma_0^2} - \frac{(x - m_1)^2}{2\sigma_1^2}\right) > \theta$$

soit en logarithme:

$$\log\left(\frac{\sigma_0}{\sigma_1\theta}\right) + \left(\frac{x^2}{2\sigma_0^2} - \frac{(x - m_1)^2}{2\sigma_1^2}\right) > 0$$

en fonction des puissances de x :

$$\frac{\sigma_0^2 - \sigma_1^2}{\sigma_0^2\sigma_1^2}x^2 - 2\frac{m_1}{\sigma_1^2}x + \frac{m_1^2}{\sigma_1^2} - 2\log\left(\frac{\sigma_0}{\sigma_1\theta}\right) > 0$$

Les deux valeurs du dépassement du seuil sont racines d'une équation du deuxième degré

$$\Delta = \frac{m_1^2}{\sigma_1^4} - \frac{\sigma_0^2 - \sigma_1^2}{\sigma_0^2\sigma_1^2} \left(\frac{m_1^2}{\sigma_1^2} - 2\log\left(\frac{\sigma_0}{\sigma_1\theta}\right) \right) \quad (146)$$

$$x_{\min} = \frac{\sigma_0^2\sigma_1^2}{\sigma_0^2 - \sigma_1^2} \left\{ \frac{m_1}{\sigma_1^2} - \Delta^{1/2} \right\} \quad x_{\max} = \frac{\sigma_0^2\sigma_1^2}{\sigma_0^2 - \sigma_1^2} \left\{ \frac{m_1}{\sigma_1^2} + \Delta^{1/2} \right\} \quad (147)$$

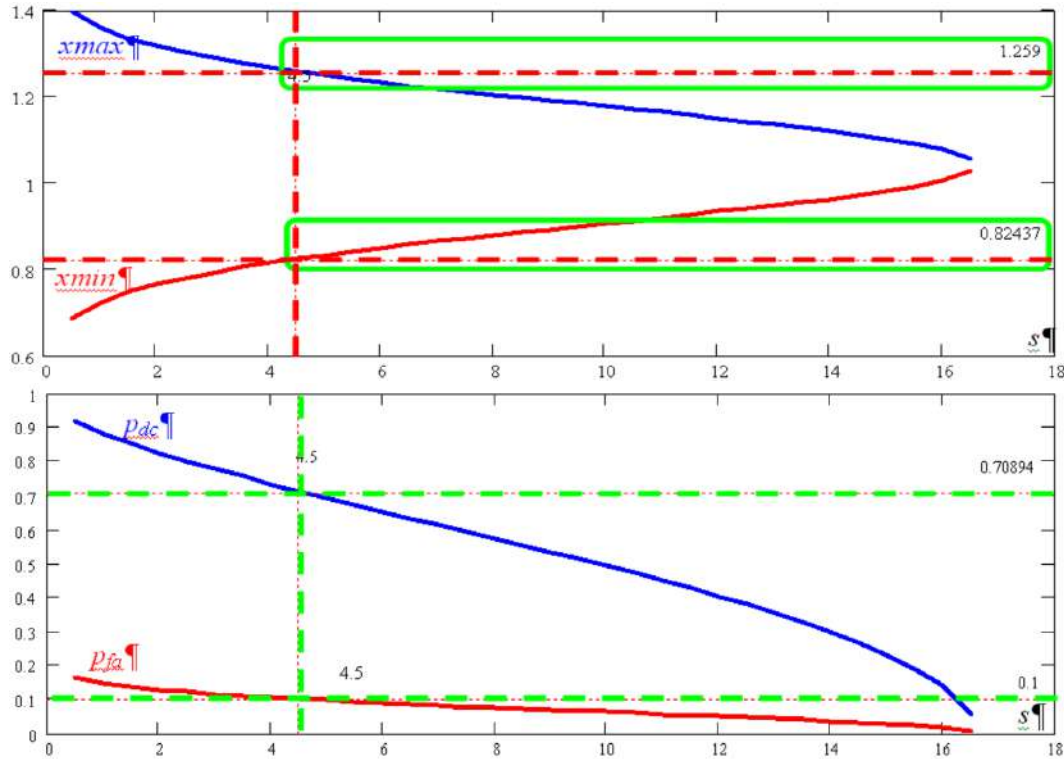


Figure 27: a) Valeur des limites x_{\min} et x_{\max} du domaine de décision $d = 1$, en fonction du seuil θ . b) probabilité de fausse alarme et probabilité de détection correcte en fonction du seuil θ

Si la probabilité de fausse alarme est de 0.1, on choisira un seuil de décision à 4.5, ce qui correspondra aux bornes $x_{\min} = 0.824$ et $x_{\max} = 1.259$ et une probabilité de décision correcte de 0.709.

Exercice 21 On veut reconnaître des signaux Morse, c'est-à-dire des impulsions de durée t variable qui sont soit des points (classe 1) soit des traits (classe 2). On admet que les densités de probabilité conditionnelles associées aux deux classes sont:

$$f(t|w_i) = \frac{\frac{1}{\pi b}}{1 + \left(\frac{t-a_i}{b}\right)^2}$$

i étant l'indice de la classe, t étant la durée de l'impulsion mesurée par le chronomètre numérique, a_i la valeur probable de la classe i

1. Déterminer la règle de décision Bayésienne permettant de reconnaître les points et les traits
2. supposons $a_1 = 3, a_2 = b = 5$, quel est le comportement de $P(w_1|t)$ quand $t \rightarrow \infty$?
3. calculer la probabilité d'erreur de la règle de Bayes.

Exercice 22 On considère dans le plan muni de la distance euclidienne, les 5 points suivants :

$$x_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, x_2 = \begin{pmatrix} 2 \\ -2 \end{pmatrix}, x_3 = \begin{pmatrix} -2 \\ -2 \end{pmatrix}, x_4 = \begin{pmatrix} -2 \\ 2 \end{pmatrix}, x_5 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Les 3 premiers points sont éléments de la classe w_1 et les 2 derniers de la classe w_2 .

1. Tracer les frontières de décision obtenues en utilisant la règle du plus proche voisin.
2. Préciser les zones de décision obtenues avec la règle de distance au barycentre

Exercice 23 On dispose d'une base de données expertisée qui servira à l'apprentissage.

Classe 1:

$$x_1 = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, x_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, x_4 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, x_5 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

Classe 2:

$$x_6 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, x_7 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}, x_8 = \begin{pmatrix} 3.5 \\ 3.5 \end{pmatrix}, x_9 = \begin{pmatrix} 4 \\ 4 \end{pmatrix}, x_{10} = \begin{pmatrix} 4 \\ 4.5 \end{pmatrix}$$

On désire utiliser l'algorithme des k -ppv avec $k = 3$ pour les points suivants:

$$A = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}, B = \begin{pmatrix} 3.5 \\ 2.5 \end{pmatrix}, C = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, D = \begin{pmatrix} 4.5 \\ 4 \end{pmatrix}, E = \begin{pmatrix} 3/2 \\ 1/2 \end{pmatrix}, F = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, G = \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix}, H = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

1. Donner les résultats intermédiaires permettant la classification de ces points puis déterminer leur classe. Peut il y avoir ambiguïté ? Qu'auraient été les conclusions avec la règle du 1ppv ?
 2. Déterminer le meilleur axe discriminant correspondant à l'analyse factorielle et les projections des points A à H sur ces axe ? Donner un représentation graphique.
1. Une analyse détaillée de la loi de probabilité de 3 classes w_1, w_2, w_3 permet de supposer qu'elles sont définies par la densité :
 - 2.
 - 3.

Exercice 24

- [1] BELAID Abdel, BeLAID Yolande, *Reconnaissance des formes : Méthodes et applications*. InterEditions, 1992.
- [2] DEVIJVER Pierre, KITTLER J., *Pattern Recognition: a statistical approach*. Prentice Hall, 1982.
- [3] DUBUISSON Bernard, *Diagnostic et reconnaissance des formes*. Hermes, 1990.
- [4] FU King-Sun, *Syntactic Methods in Pattern Recognition*. Academic Press, 1974.
- [5] GAILLAT Gérard, *Méthodes statistiques de reconnaissance des formes*. Publication ENSTA, 1983.
- [6] MICLET Laurent, *Méthodes structurales pour la reconnaissance des formes*. Eyrolles et CNET-ENST, 1984.
- [7] MILGRAM Maurice, *Reconnaissance des formes : Méthodes numériques et connexionnistes*. Armand Colin, 1993.
- [8] PAVLIDIS T., *Structural Pattern Recognition*. Springer Verlag, 1982.
- [9] SIMON Jean-Claude, *La reconnaissance des formes par algorithmes*. Masson, 1984.
- [10] WATANABE Satosi, *Knowing and Guessing*. John Wiley, 1969.
- [11] M. Kunt, G. Granlund, M. Kocher, *Traitement numérique des signaux*, Presses Polytechniques et Universitaires Romandes, Lausanne, Switzerland, 1993.
- [12] J.S. Lim, *Two-dimensional signal and image processing*, Prentice Hall, New Jersey, 1990
- [13] A.K. Jain, "Fundamentals of digital image processing", Prentice Hall, New Jersey, 1989
- [14] D.E. Dudgeon, R.M. Mersereau, "Multidimensional signal processing", Prentice Hall, New Jersey, 1984
- [15] DUDA Richard, HART Peter, *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.