

le cnam

Reconnaissance des formes et méthodes neuronales

Chapitre 2 : Analyse en composantes principales

Michel Crucianu
<http://lecnam.net>
<http://cedric.cnam.fr/~crucianm/rfmn.html>

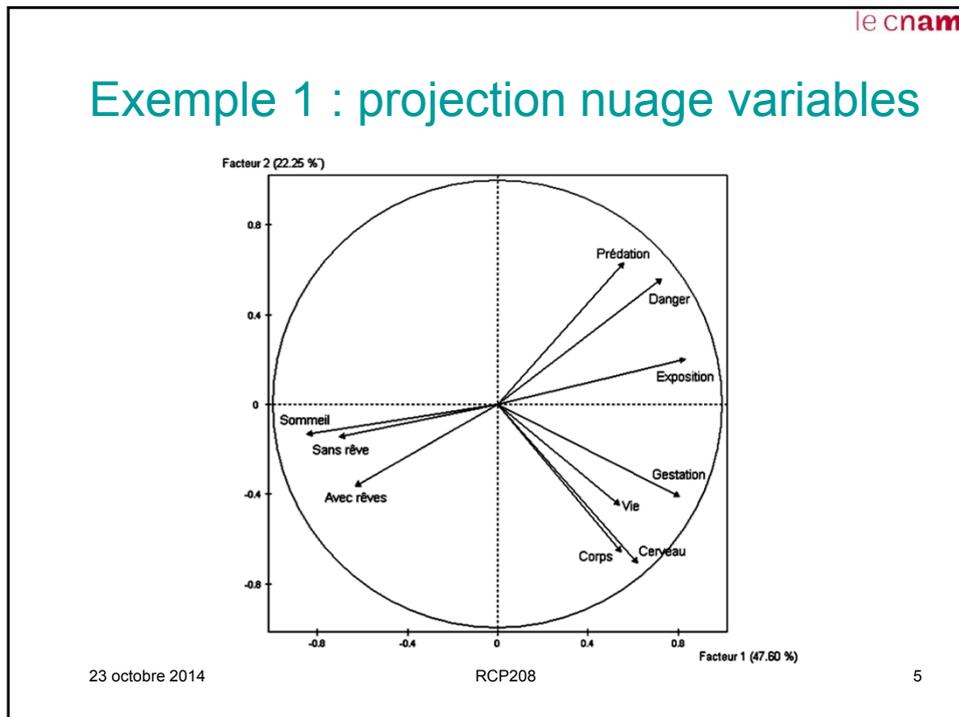
23 octobre 2014 RCP208 1

le cnam

Analyse en composantes principales

- Méthode d'**analyse multivariée exploratoire**, introduite par Hotelling en 1933 suivant des idées de Pearson (1901)
- Les données : ensemble d'individus caractérisés par un ensemble de variables quantitatives
- Objectif : résumer les variables initiales à l'aide d'un petit nombre de variables synthétiques (les **composantes principales**) obtenues comme des combinaisons linéaires des variables initiales
- Utilisations
 - ◆ **Condenser la représentation des données** en conservant au mieux leur **organisation globale**
 - ◆ Permettre une **visualisation de l'organisation prépondérante** des données grâce à une projection sur un espace de faible dimension
 - ◆ **Préparer d'autres analyses** en éliminant les variables redondantes et les directions dans lesquelles la variance des données est très faible
 - ◆ Identifier les composantes unidimensionnelles d'une loi normale multidimensionnelle

23 octobre 2014 RCP208 2



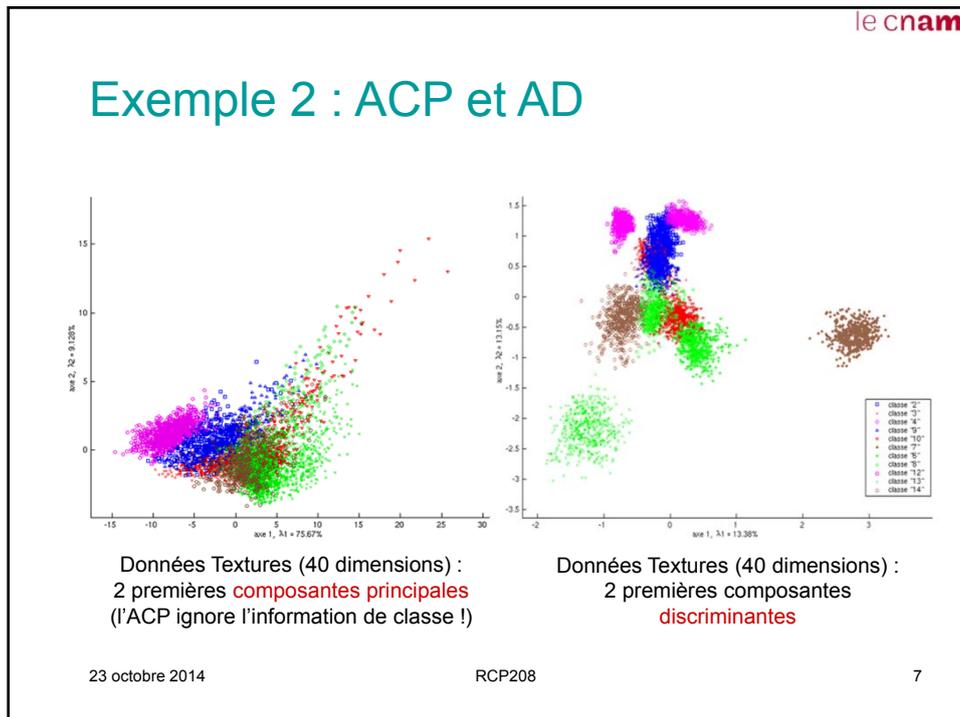
le cnam

Exemple 2 : les données

- Analyse de données caractérisant des **textures**
- Individus (**anonymes**, 5500 au total) : 500 pixels pour chacune des 11 micro-textures différentes
- Variables (40 au total) : moments statistiques modifiés d'ordre 4, déterminés pour 4 orientations différentes (0°, 45°, 90°, 135°), tenant compte des relations avec les voisins d'ordres 1 et 2
- Application d'une ACP normée pour réduire la redondance potentiellement présente dans les variables initiales et indiquer dans quelle mesure l'appartenance à une classe se reflète dans l'organisation globale des données

[Données obtenues et mises à disposition par le Laboratoire de Traitement d'Image et de Reconnaissance de Formes (LTIRF) de l'Institut National Polytechnique de Grenoble (INPG) dans le cadre du projet ESPRIT III ELENA (No. 6891) et du groupe de travail ESPRIT ATHOS (No. 6620)]

23 octobre 2014 RCP208 6



le cnam

ACP : les données

- Tableau de données \mathbf{X}
 - ◆ n lignes (les *individus*, ou observations)
 - ◆ p colonnes (les variables, à valeurs dans \mathbb{R})

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Individu (ou observation) i :

$$\mathbf{e}_i = T[x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip}]$$

Variable j :

$$\mathbf{x}_j = T[x_{1j} \quad x_{2j} \quad \cdots \quad x_{nj}]$$

23 octobre 2014
RCP208
8

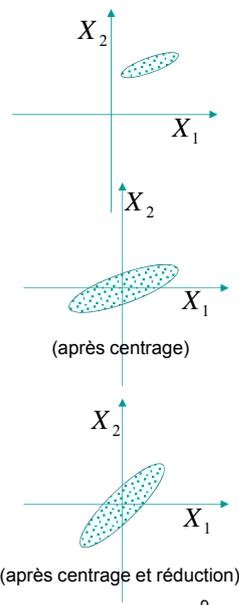
le cnam

Données et types d'analyse

- Valeurs directement recueillies → ACP **générale**
 - ◆ Interviennent dans l'analyse à la fois la position du nuage d'individus par rapport à l'origine et la forme du nuage
 - ◆ Utilisation : analyse tenant compte du zéro naturel de certaines variables

- Variables centrées (moyennes nulles) → ACP **centrée**
 - ◆ Analyse de la forme du nuage d'individus (par rapport à son centre de gravité)
 - ◆ Utilisation : variables directement comparables

- Variables centrées et réduites (moyennes nulles et écarts-types unitaires) → ACP **normée**
 - ◆ Analyse de la forme du nuage après normalisation
 - ◆ Utilisation : variables non directement comparables, ayant des unités de mesure différentes et/ou des intervalles de variation très différents



(après centrage)

(après centrage et réduction)

23 octobre 2014 RCP208 9

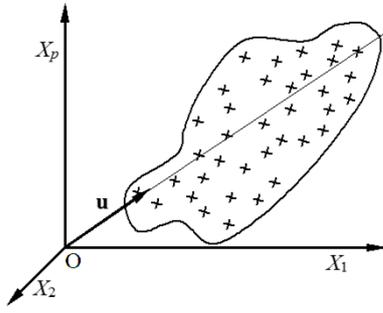
le cnam

Analyse du nuage des individus

- Chercher les sous-espaces linéaires de \mathbb{R}^p qui **maximisent la dispersion des projections** des points-individus sur ces sous-espaces
- Pour une dimension donnée, le sous-espace qui **maximise la dispersion** est également celui qui **ajuste le mieux** le nuage (qui minimise l'erreur faite lors de l'approximation du nuage initial par le nuage des projections)

- Ajustement par une droite :

- Choix pour la suite de la présentation :
 - ◆ Distance euclidienne classique
 - ◆ Pondération identique des individus



23 octobre 2014 RCP208 10

le cnam

Ajustement par une droite

- Quelle droite maximise l'inertie par rapport à l'origine des projections orthogonales des points sur la droite ?
- Direction donnée par le vecteur \mathbf{u} , unitaire ($\|\mathbf{u}\| = \sqrt{\mathbf{u}^T \mathbf{u}} = 1$)

$$d(O, P_i) = \sqrt{\mathbf{e}_i^T \mathbf{u}} = \|\mathbf{e}_i\| \|\mathbf{u}\| \cos(\mathbf{e}_i, \mathbf{u})$$
- Inertie totale des projections des points sur la droite :

$$\sum_{i=1}^n d^2(O, P_i) = \sum_{i=1}^n (\sqrt{\mathbf{e}_i^T \mathbf{u}})^2$$
- Sachant que

$$d^2(A_i, P_i) = d^2(O, A_i) - d^2(O, P_i)$$

maximiser la dispersion des projections sur la droite revient à optimiser l'ajustement du nuage

23 octobre 2014
RCP208
11

le cnam

Ajustement par une droite (2)

- Optimisation : maximiser $\sum_{i=1}^n (\sqrt{\mathbf{e}_i^T \mathbf{u}})^2$ sous la contrainte $\mathbf{u}^T \mathbf{u} = 1$ (pour éviter les solutions triviales)
- Sachant que

$$\sum_{i=1}^n (\sqrt{\mathbf{e}_i^T \mathbf{u}})^2 = \sum_{i=1}^n \mathbf{e}_i^T \mathbf{u} \mathbf{e}_i^T \mathbf{u} = \sum_{i=1}^n \mathbf{u}^T (\mathbf{e}_i \mathbf{e}_i^T) \mathbf{u} = \mathbf{u}^T \left[\sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i^T \right] \mathbf{u}$$

$$\sum_{i=1}^n \mathbf{e}_i \mathbf{e}_i^T = \mathbf{X} \mathbf{X}^T$$
- On arrive au problème : maximiser $\mathbf{u}^T \mathbf{X} \mathbf{X}^T \mathbf{u}$ sous la contrainte $\mathbf{u}^T \mathbf{u} = 1$
- Méthode des multiplicateurs de Lagrange

lagrangien : $L(\mathbf{u}) = \mathbf{u}^T \mathbf{X} \mathbf{X}^T \mathbf{u} - \lambda (\mathbf{u}^T \mathbf{u} - 1)$

dérivée partielle : $\frac{\partial L}{\partial \mathbf{u}}(\mathbf{u}) = 2 \mathbf{X} \mathbf{X}^T \mathbf{u} - 2\lambda \mathbf{u} = 0$ pour les extrêmes

solution : $\mathbf{X} \mathbf{X}^T \mathbf{u} = \lambda \mathbf{u}$ (pb. de valeurs et vecteurs propres ordinaire)

$\mathbf{u}^T \mathbf{X} \mathbf{X}^T \mathbf{u} = \lambda \mathbf{u}^T \mathbf{u} = \lambda \Rightarrow$ vecteur propre de la plus grande valeur propre

23 octobre 2014
RCP208
12

Ajustement par un sous-espace

- Identifier le sous-espace de dimension k maximisant l'inertie des projections des points du nuage des individus sur ce sous-espace
- Suivant le même procédé on obtient : une base orthonormée qui définit ce sous-espace est constituée des k vecteurs propres unitaires de ${}^T \mathbf{X} \mathbf{X}$ correspondant aux k plus grandes valeurs propres (non nulles)

$${}^T \mathbf{X} \mathbf{X} \mathbf{u}_\alpha = \lambda_\alpha \mathbf{u}_\alpha \quad \lambda_1 > \lambda_2 > \lambda_3 > \dots > 0$$

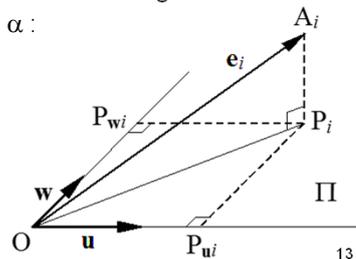
- Propriétés de ${}^T \mathbf{X} \mathbf{X}$: symétrique positive semi-définie \Rightarrow les vecteurs propres des valeurs propres différentes et $\neq 0$ sont orthogonaux
- Projection de l'individu i sur l'axe factoriel α :

$$\psi_{\alpha i} = {}^T \mathbf{e}_i \mathbf{u}_\alpha$$

- Inertie des projections sur l'axe α :

$${}^T \mathbf{u}_\alpha {}^T \mathbf{X} \mathbf{X} \mathbf{u}_\alpha = \lambda_\alpha$$

$$\text{ou encore } \sum_{i=1}^n \psi_{\alpha i}^2 = \lambda_\alpha$$



23 octobre 2014

RCP208

13

Analyse du nuage des variables

- La variable j est représentée par le vecteur $\mathbf{x}_j = {}^T [x_{1j} \ x_{2j} \ \dots \ x_{nj}]$
 - Pour chaque k , on cherche le sous-espace linéaire de \mathbb{R}^n , de dimension k , qui **maximise la dispersion des projections** des points-variables sur ce sous-espace
 - Comme pour les points-individus, on obtient : une base orthonormée qui définit ce sous-espace est constituée des k vecteurs propres unitaires de la matrice $\mathbf{X} {}^T \mathbf{X}$ correspondant aux k plus grandes valeurs propres $\neq 0$
- $$\mathbf{X} {}^T \mathbf{X} \mathbf{v}_\alpha = \mu_\alpha \mathbf{v}_\alpha \quad \mu_1 > \mu_2 > \mu_3 > \dots > 0$$
- Tout comme ${}^T \mathbf{X} \mathbf{X}$, la matrice $\mathbf{X} {}^T \mathbf{X}$ est symétrique positive semi-définie
 - Projection de la variable j sur l'axe factoriel α : $\phi_{\alpha j} = {}^T \mathbf{x}_j \mathbf{v}_\alpha$

23 octobre 2014

RCP208

14

Relation entre les 2 analyses

- Y a-t-il une **relation** entre l'analyse du nuage des points-individus et l'analyse du nuage des points-variables ?
- En pré-multipliant ${}^T \mathbf{X} \mathbf{X} \mathbf{u}_\alpha = \lambda_\alpha \mathbf{u}_\alpha$ par \mathbf{X} on obtient $\mathbf{X} {}^T \mathbf{X} (\mathbf{X} \mathbf{u}_\alpha) = \lambda_\alpha (\mathbf{X} \mathbf{u}_\alpha)$
- Donc à tout vecteur propre \mathbf{u}_α de ${}^T \mathbf{X} \mathbf{X}$ associé à une valeur propre $\lambda_\alpha \neq 0$ correspond un vecteur propre $\mathbf{X} \mathbf{u}_\alpha$ de $\mathbf{X} {}^T \mathbf{X}$ associé à la **même** valeur propre
- En pré-multipliant $\mathbf{X} {}^T \mathbf{X} \mathbf{v}_\alpha = \mu_\alpha \mathbf{v}_\alpha$ par ${}^T \mathbf{X}$ on obtient la réciproque
- Les valeurs propres non nulles de ${}^T \mathbf{X} \mathbf{X}$ et $\mathbf{X} {}^T \mathbf{X}$ sont donc toutes égales, $\lambda_\alpha = \mu_\alpha$

Relation entre les 2 analyses

- Comme $\|\mathbf{X} \mathbf{u}_\alpha\|^2 = {}^T \mathbf{u}_\alpha {}^T \mathbf{X} \mathbf{X} \mathbf{u}_\alpha = \lambda_\alpha$ et $\|{}^T \mathbf{X} \mathbf{v}_\alpha\|^2 = \lambda_\alpha$, nous obtenons les **relations entre vecteurs propres** :

$$\begin{cases} \mathbf{v}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{X} \mathbf{u}_\alpha \\ \mathbf{u}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} {}^T \mathbf{X} \mathbf{v}_\alpha \end{cases} \quad \text{qui impliquent} \quad \begin{cases} \boldsymbol{\Psi}_\alpha = \sqrt{\lambda_\alpha} \mathbf{v}_\alpha \\ \boldsymbol{\Phi}_\alpha = \sqrt{\lambda_\alpha} \mathbf{u}_\alpha \end{cases}$$

- En conséquence, pour les projections sur les axes factoriels nous avons les **relations de transition** :

$$\begin{cases} \Psi_{\alpha i} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^p x_{ij} \Phi_{\alpha j} \\ \Phi_{\alpha j} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n x_{ij} \Psi_{\alpha i} \end{cases}$$

Reconstitution du tableau de données

- Peut-on retrouver \mathbf{X} à partir des vecteurs propres \mathbf{u}_α , \mathbf{v}_α et des valeurs propres λ_α ?
- Le rang de ${}^T \mathbf{X} \mathbf{X}$ est $q \leq p$, q étant le nombre de valeurs propres $\neq 0$

- De $\mathbf{X} \mathbf{u}_\alpha {}^T \mathbf{u}_\alpha = \sqrt{\lambda_\alpha} \mathbf{v}_\alpha {}^T \mathbf{u}_\alpha$ on obtient

$$\mathbf{X} \sum_{\alpha=1}^q \mathbf{u}_\alpha {}^T \mathbf{u}_\alpha = \sum_{\alpha=1}^q \sqrt{\lambda_\alpha} \mathbf{v}_\alpha {}^T \mathbf{u}_\alpha$$

- Comme \mathbf{U} est orthogonale, ${}^T \mathbf{U} = \mathbf{U}^{-1}$, et $\sum_{\alpha=1}^p \mathbf{u}_\alpha {}^T \mathbf{u}_\alpha = \mathbf{U} {}^T \mathbf{U}$ on obtient $\mathbf{X} \sum_{\alpha=1}^p \mathbf{u}_\alpha {}^T \mathbf{u}_\alpha = \mathbf{X} \mathbf{U} {}^T \mathbf{U} = \mathbf{X} \mathbf{I}_p = \mathbf{X}$

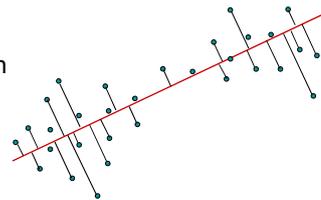
et donc $\mathbf{X} = \sum_{\alpha=1}^q \sqrt{\lambda_\alpha} \mathbf{v}_\alpha {}^T \mathbf{u}_\alpha$

qui montre comment **reconstituer de façon exacte** \mathbf{X}

Reconstitution du tableau de données

- Si on s'arrête à $k < q$, on obtient l'expression de la **reconstitution approximative de \mathbf{X}** :

$$\mathbf{X} \approx \mathbf{X}^* = \sum_{\alpha=1}^k \sqrt{\lambda_\alpha} \mathbf{v}_\alpha {}^T \mathbf{u}_\alpha$$



- **Théorème** (Eckart-Young) : \mathbf{X}^* est la meilleure approximation, au sens des moindres carrés, de la matrice \mathbf{X} par une matrice de rang k inférieur au rang de \mathbf{X} .

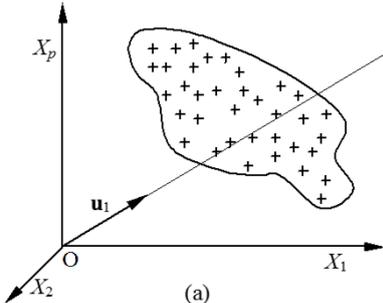
⇒ Intérêt de l'ACP pour la compression : \mathbf{X} est représentée par $n \times p$ valeurs, alors que pour stocker \mathbf{X}^* , $k(n+p)$ valeurs suffisent !

Des méthodes **non linéaires** d'analyse en composantes principales (basées, par exemple, sur des réseaux de neurones) sont également utilisées pour la compression, avec de meilleurs résultats en général

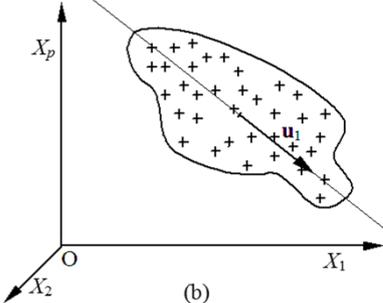
le cnam

ACP centrée, ACP normée

- Pour l'ACP générale, la **position** et la **largeur** des intervalles des valeurs des variables interviennent dans l'analyse !



(a)



(b)

Premier axe factoriel dans l'analyse du nuage des individus, (a) par rapport à l'origine (ACP générale) ou (b) par rapport au centre de gravité (ACP centrée, ACP normée)

23 octobre 2014
RCP208
19

le cnam

ACP centrée

- Opération de **centrage** : la moyenne des valeurs de chaque variable est déterminée et soustraite avant l'analyse

- Si $\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{np} \end{bmatrix}$ est le tableau des **données brutes**,

alors $\mathbf{X} = [1/\sqrt{n-1}][\mathbf{R} - \mathbf{g}^T \mathbf{1}_n]$ et on effectue l'analyse sur ${}^T \mathbf{X} \mathbf{X}$, la **matrice des covariances empiriques** entre les variables (\mathbf{g} est le centre de gravité du nuage, $\mathbf{1}_n$ est le vecteur colonne de dimension n dont toutes les composantes sont égales à 1)

23 octobre 2014
RCP208
20

ACP normée

- Opération de **réduction** : l'écart-type $\sigma_j^2 = [1/(n-1)] \sum_{i=1}^n (r_{ij} - g_j)^2$ des valeurs de chaque variable est calculé, puis les valeurs de chaque variable sont divisées par son écart-type

- Dans ce cas, $\mathbf{X} = (\mathbf{R} - \mathbf{g}^T \mathbf{1}_n) \mathbf{D}_r^{-1}$, avec

$$\mathbf{D}_r = \sqrt{n-1} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p \end{bmatrix}$$

et l'analyse est effectuée sur ${}^T \mathbf{X} \mathbf{X}$, qui est la **matrice des corrélations empiriques** entre les variables

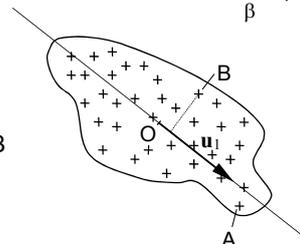
- Pour l'ACP normée, tous les points-variables sont à une même distance de l'origine des axes factoriels car $\sum_{i=1}^n x_{ij}^2 = [\sum_{i=1}^n (r_{ij} - g_j)^2] / [(n-1)\sigma_j^2] = 1$ et donc $d^2(\mathbf{x}_j, \mathbf{0}) = \sum_{i=1}^n x_{ij}^2 = 1$

Aides à l'interprétation

- Apport d'un axe factoriel
 - Contribution relative** (ou **cosinus carré**, ou **qualité de représentation**) d'un axe dans l'explication de l'inertie d'un individu ou d'une variable

$$Cr_\alpha(i) = \cos^2(\mathbf{e}_i, \mathbf{u}_\alpha) = \frac{\Psi_{\alpha i}^2}{\|\mathbf{e}_i\|^2} \quad \text{avec} \quad \sum_\alpha Cr_\alpha(i) = \sum_\alpha \frac{\Psi_{\alpha i}^2}{\sum_\beta \Psi_{\beta i}^2} = 1$$

- Le premier axe factoriel explique la majeure partie de l'inertie de A mais une faible part de l'inertie de B



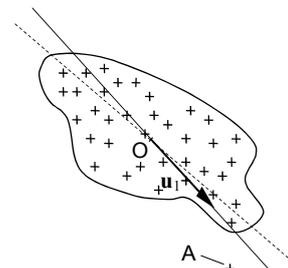
le cnam

Aides à l'interprétation

- Influence d'un individu ou d'une variable
 - ◆ Contribution absolue (ou simplement contribution) d'un individu ou d'une variable à la variance expliquée par un axe factoriel

$$Ca_{\alpha}(i) = \frac{\Psi_{\alpha i}^2}{\lambda_{\alpha}} \quad \text{avec} \quad \sum_{i=1}^n Ca_{\alpha}(i) = (1/\lambda_{\alpha}) \sum_{i=1}^n \Psi_{\alpha i}^2 = 1$$

- ◆ Un individu très excentré aura un impact très fort sur l'orientation des axes factoriels



The diagram shows a cloud of points represented by '+' signs. A solid line represents an axis labeled 'A'. A dashed line is perpendicular to 'A'. A vector labeled 'u_i' originates from the center of the cloud and points towards a specific point, illustrating its influence on the orientation of axis 'A'.

23 octobre 2014 RCP208 23

le cnam

Règles d'interprétation

- Objectifs :
 - ◆ Révéler quelles variables sont corrélées et quelles variables s'opposent
 - ◆ Mettre en évidence des regroupements d'individus et les caractériser à partir des variables
 - ◆ Si les individus ne sont pas anonymes, identifier des relations entre eux
- Associer une interprétation aux axes à partir des corrélations et des oppositions entre variables initiales
- Le regroupement de projections des individus sur les plans factoriels s'interprète en termes de similitudes de comportement par rapport aux variables
- Éviter d'interpréter les projections sur un axe des individus mal représentés par l'axe (contributions relatives faibles)
- Retirer, pour une nouvelle analyse, les individus ayant des contributions absolues excessives à l'orientation de certains axes

23 octobre 2014 RCP208 24

Variables et individus supplémentaires

- Après analyse, de nouveaux individus et/ou de nouvelles variables peuvent être projetés sur les axes factoriels afin de contribuer à l'interprétation
- Individus supplémentaires (ou illustratifs)
 1. Observations faites dans des conditions différentes, nouvelles observations
 2. Individus atypiques, éliminés de l'analyse en raison de leurs contributions absolues excessives
 3. Centre de gravité d'un groupe d'individus (par ex., individus qui possèdent une certaine modalité d'une variable nominale)
 - Projection directe sur les axes factoriels (après opérations de centrage et réduction si ACP centrée et normée)
 - Qualité de représentation : indication utile

23 octobre 2014

RCP208

25

Variables et individus supplémentaires

- Variables supplémentaires (ou illustratives)
 - ◆ Quantitatives : projection directe sur les axes factoriels de l'analyse des variables (après centrage et réduction si ACP centrée et normée)
 - ◆ Nominales (à modalités)
 - Pour chaque modalité, calcul du centre de gravité des individus qui la possèdent et projection sur les axes de l'analyse des individus
 - Interprétation : nécessité de calculer une valeur-test

$$v_{\alpha k} = \left(\frac{1}{n_k} \sum_{i \in k} \psi_{\alpha i} \right) \cdot \frac{1}{\sqrt{\frac{n - n_k}{n - 1} \cdot \frac{\lambda_{\alpha}}{n_k}}}$$

la position de la modalité (représentée par le centre de gravité des individus qui la possèdent) est interprétable si $v_{\alpha k} > 1,96$ (n_k est le nombre d'individus qui possèdent la modalité k)

23 octobre 2014

RCP208

26

Ex. 1 : corrélations entre variables

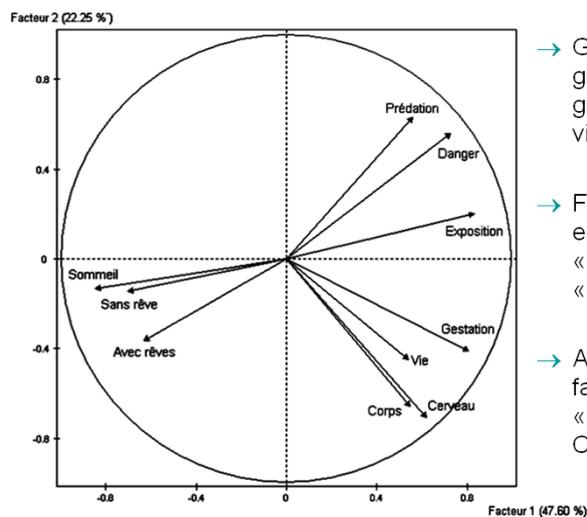
	Corps	Cerveau	Sans rêve	Avec rêves	Sommeil	Vie	Gestation	Prédation	Exposition	Danger
Corps	1,00									
Cerveau	0,93	1,00								
Sans rêve	-0,14	-0,24	1,00							
Avec rêves	-0,04	-0,07	0,50	1,00						
Sommeil	-0,31	-0,36	0,87	0,65	1,00					
Vie	0,30	0,51	-0,37	-0,28	-0,40	1,00				
Gestation	0,65	0,75	-0,44	-0,34	-0,56	0,60	1,00			
Prédation	0,06	0,03	-0,27	-0,40	-0,38	-0,10	0,19	1,00		
Exposition	0,34	0,37	-0,44	-0,47	-0,59	0,35	0,62	0,62	1,00	
Danger	0,13	0,15	-0,40	-0,51	-0,55	0,06	0,37	0,92	0,79	1,00

23 octobre 2014

RCP208

27

Exemple 1 : projection nuage variables

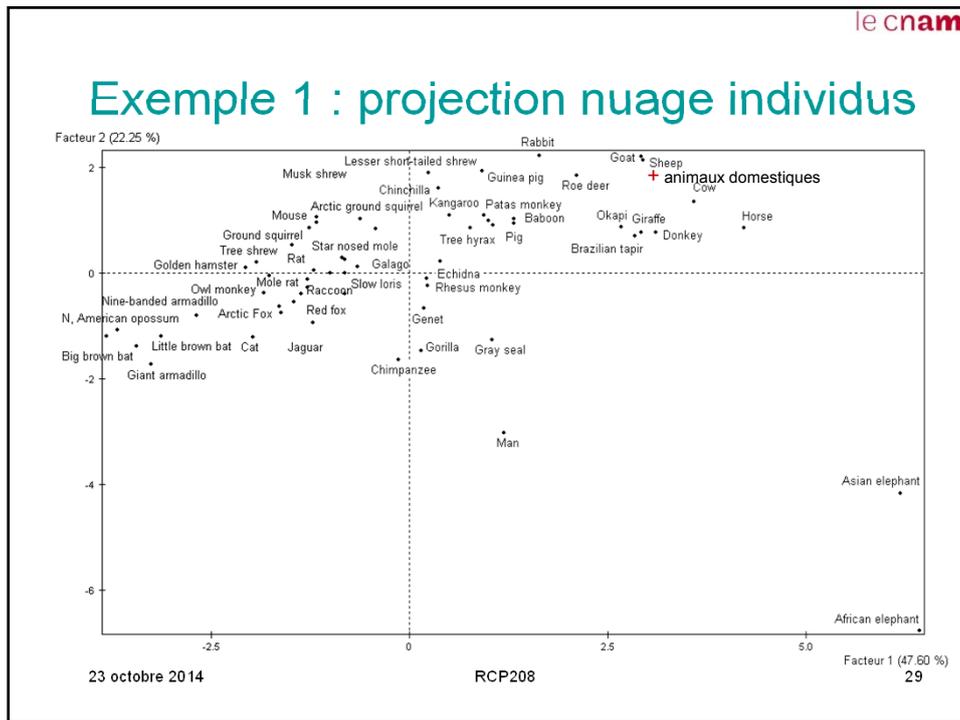


- Groupe « sommeil », groupe « danger », groupe « corps, cerveau, vie, gestation » (CCVG)
- Forte anti-corrélation entre le groupe « sommeil » et le groupe « danger »
- Anti-corrélation plus faible entre le groupe « danger » et le groupe CCVG

23 octobre 2014

RCP208

28



le cnam

Choix du nombre d'axes

- Le critère dépend de l'objectif de l'analyse !
 - ◆ Analyse descriptive avec interprétation : à partir de quel ordre les différences entre les pourcentages d'inertie expliquée ne sont plus significatives ?
 - Empiriquement : changement de pente dans la décroissance des valeurs propres
 - Hypothèse de normalité : test statistique d'égalité entre valeurs propres successives
 - ◆ Compression des données : qualité d'approximation, mesurée par le taux d'inertie expliquée
 - ◆ ACP comme pré-traitement avant méthodes décisionnelles :
 - Critère de conditionnement de la matrice des covariances empiriques ou de la matrice des corrélations empiriques
 - Introduction du nombre d'axes à retenir comme paramètre du modèle décisionnel et utilisation de méthodes de sélection de modèle
 - ◆ Visualisation des données : 2 ou 3 premiers axes factoriels

23 octobre 2014 RCP208

