

ACP : Approfondissements :

1. Réduction
2. Dualité
3. Robustesse de la méthode
(théorique et par *bootstrap*)
4. Décomposition
5. Prévion
6. ACP et extensions

Octobre 2020

Mireille Batton-Hubert - Institut H.Fayol

UP3: 'Analyse de données et données temporelles' - Majeure Science des données

1

A. Rappel sur La réduction de dimension (*dimensionnalité*)

La réduction de dimensionnalité : pour

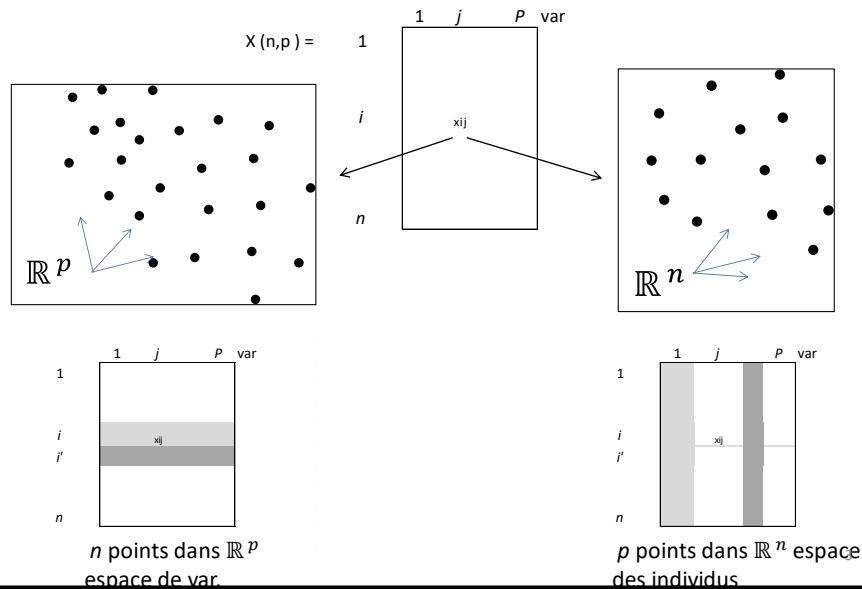
- Compression de données (avec perte)
 - Visualisation des données en 2D ou 3D
 - Extraction de caractéristiques potentiellement +fondamentales, +explicatives, +compactes
- Prétraitement => meilleure représentation de départ pour un autre algorithme (classification ou régression)

L'ACP: trouver un sous-espace linéaire qui passe proche des données: projection orthogonale de sur un sous espace linéaire de plus faible dimension M.

- Les composantes z représentent les coordonnées de la projection de x dans ce sous-espace de dimension M.
- Très utilisé comme pré-traitement (extraction de caractéristiques) ou pour la visualisation.
- 2 formulations équivalentes:
 - minimisation de l'erreur de reconstruction (Pearson 1901)
 - maximisation de la variance (Hotelling 1933).
- La réinterprétation en tant que modèle probabiliste à variables latentes : beaucoup plus récente.

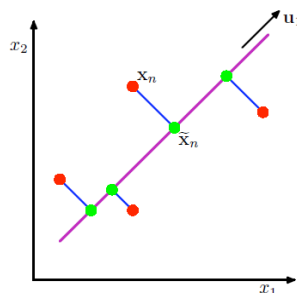
2

1. Ajustement du nuage des individus / variables dans l'espace \mathbb{R}^p et \mathbb{R}^n

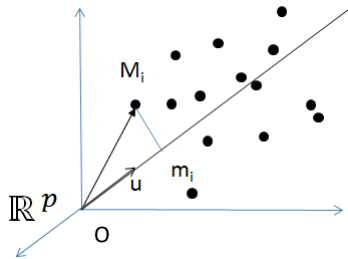


1.0 ACP : par maximisation de la variance (Hotelling 1933)

- Minimum d'erreur: minimise la moyenne des distances carrées entre les x et leur projection (lignes bleues).
- **Maximum de variance:** maximise la variance le long de la direction de projection (variance des points verts)



1.1 Ajustement du nuage des n individus dans un espace des p variables \mathbb{R}^p



Chercher le maximum de la forme quadratique $\mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u}$:

$$\begin{cases} \max_{(\mathbf{u})} \{\mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u}\} \\ \text{Sous la contrainte : } \mathbf{u}'\mathbf{u} = 1 \end{cases}$$

$$Om_i = \mathbf{x} \mathbf{u} = \sum_j^p x_{ij} u_j$$

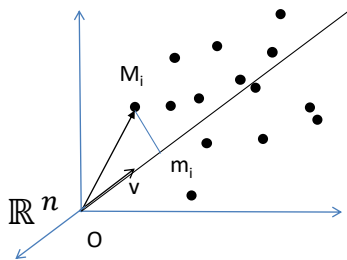
$$\mathbf{X}\mathbf{u} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix}$$

$$= \begin{bmatrix} \sum x_{1j} u_j \cdots \\ \sum x_{ij} u_j \\ \sum x_{nj} u_j \end{bmatrix}$$

Un sous-espace à q dimensions qui ajuste au mieux le nuage dans \mathbb{R}^p engendré par les q premiers vecteurs propres de la matrice $\mathbf{X}'\mathbf{X}$ (covariance)

5

1.2 Ajustement du nuage des p variables dans un espace des n individus \mathbb{R}^n



Un sous-espace à q dimensions qui ajuste au mieux le nuage dans \mathbb{R}^n engendré par les q premiers vecteurs propres de la matrice $\mathbf{X}\mathbf{X}'$

La matrice à diagonaliser est $\mathbf{X}\mathbf{X}'$
 \mathbf{v}_α vecteur propre de la matrice $\mathbf{X}\mathbf{X}'$ de valeur propre μ_α

$$Om_i = \mathbf{x}' \mathbf{v} = \sum_j^n x_{ij} v_j$$

$$\mathbf{X}'\mathbf{v} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{p1} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

On cherche à rendre maximale la somme des carrés des p projections sur \mathbf{v} (des p individus) sur les q composantes du vecteur $\mathbf{X}'\mathbf{v}$

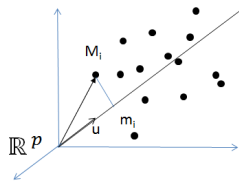
Chercher le maximum de la forme quadratique $(\mathbf{X}'\mathbf{v})'\mathbf{X}'\mathbf{v}$:

$$\begin{cases} \max_{(\mathbf{v})} \{\mathbf{v}'\mathbf{X}\mathbf{X}'\mathbf{v}\} \\ \text{Sous la contrainte : } \mathbf{v}'\mathbf{v} = 1 \end{cases}$$



6

2.1 Recherche de vecteur \mathbf{u} qui rend maximale inertie du nuage projeté de n points un espace à p variables \mathbb{R}^p



Chercher le maximum de la forme quadratique
 $\mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u}$ ($= \text{Is}(\mathbf{u})$)

$$\left\{ \begin{array}{l} \max_{(\mathbf{u})} \{\mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u}\} \\ \text{Sous la contrainte : } \mathbf{u}'\mathbf{u} = 1 \quad (\mathbf{u}'\mathbf{M}\mathbf{u} = 1) \end{array} \right.$$

Maximiser $\text{Is}(\mathbf{u}) = \mathbf{u}'\mathbf{A}\mathbf{u}$
 Sous la contrainte $\mathbf{u}'\mathbf{M}\mathbf{u} = 1$
 → Maximum lorsque annulation des dérivées du Lagrangien

Démonstration

$$\frac{\partial(\mathbf{u}'\mathbf{A}\mathbf{u})}{\partial \mathbf{u}} = 2\mathbf{A}\mathbf{u}$$

$$\frac{\partial(\mathbf{u}'\mathbf{M}\mathbf{u})}{\partial \mathbf{u}} = 2\mathbf{M}\mathbf{u}$$

$$\mathcal{L} = \mathbf{u}'\mathbf{A}\mathbf{u} - \lambda(\mathbf{u}'\mathbf{M}\mathbf{u} - 1)$$

Formulation du Lagrangien

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}} = 2\mathbf{A}\mathbf{u} - 2\lambda\mathbf{M}\mathbf{u} = 0$$

Maximum : annulation des dérivées
 du Lagrangien

7

2.2 Recherche de vecteur \mathbf{u} qui rend maximale inertie du nuage projeté

Maximiser $\text{Is}(\mathbf{u}) = \mathbf{u}'\mathbf{A}$

Sous la contrainte $\mathbf{u}'\mathbf{M}\mathbf{u} = 1$

→ Maximum lorsque annulation des dérivées du Lagrangien

Démonstration

$$\frac{\partial(\mathbf{u}'\mathbf{A}\mathbf{u})}{\partial \mathbf{u}} = 2\mathbf{A}\mathbf{u}$$

$$\frac{\partial(\mathbf{u}'\mathbf{M}\mathbf{u})}{\partial \mathbf{u}} = 2\mathbf{M}\mathbf{u}$$

$$\mathcal{L} = \mathbf{u}'\mathbf{A}\mathbf{u} - \lambda(\mathbf{u}'\mathbf{M}\mathbf{u} - 1)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}} = 2\mathbf{A}\mathbf{u} - 2\lambda\mathbf{M}\mathbf{u} = 0$$

Soit

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{M} \mathbf{u} \quad (1)$$

$$\lambda = \mathbf{u}'\mathbf{A}\mathbf{u} \text{ car } \mathbf{u}'\mathbf{M}\mathbf{u} = 1 \quad (2)$$

$$\mathbf{M}^{-1}\mathbf{A}\mathbf{u} = \lambda \mathbf{u} \text{ car } \mathbf{M} \text{ inversible } (3)$$

\mathbf{u} vecteur propre de la matrice
 $\mathbf{M}^{-1}\mathbf{A}$ correspondant à la valeur
 propre la plus grande (si unique)
 Noté \mathbf{u}_1

Recherche du vecteur propre \mathbf{u}_2 M-orthogonal à \mathbf{u}_1

8

2.3 Recherche de vecteur \mathbf{u} qui rend maximale inertie du nuage projeté

Maximiser $Is(\mathbf{u}) = \mathbf{u}'\mathbf{A}\mathbf{u}$

Sous la contrainte $\mathbf{u}'\mathbf{M}\mathbf{u} = 1$

→ Maximum lorsque annulation des dérivées du Lagrangien

Démonstration du théorème

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{M}\mathbf{u} \quad (1)$$

$$\lambda = \mathbf{u}'\mathbf{A}\mathbf{u} \text{ car } \mathbf{u}'\mathbf{M}\mathbf{u} = 1 \quad (2)$$

$$\mathbf{M}^{-1}\mathbf{A}\mathbf{u} = \lambda \mathbf{u} \text{ car } \mathbf{M} \text{ inversible} \quad (3)$$

Théorème : la direction du premier axe factoriel est définie par le vecteur propre associé à la plus grande valeur propre de la matrice variance-covariance

qui donne le vecteur propre \mathbf{u}_1 associé à la plus grande valeur propre

Recherche du vecteur propre \mathbf{u}_2 M-orthogonal à \mathbf{u}_1

$$\mathbf{u}_2'\mathbf{M}\mathbf{u}_2 = 1 \text{ et } \mathbf{u}_1'\mathbf{M}\mathbf{u}_2 = 0$$

Maximise la forme $\mathbf{u}_2'\mathbf{A}\mathbf{u}_2$

$$\mathcal{L} = \mathbf{u}_2' \mathbf{A} \mathbf{u}_2 - \lambda_2 (\mathbf{u}_2' \mathbf{M} \mathbf{u}_2 - 1) - \mu_2 (\mathbf{u}_2' \mathbf{M} \mathbf{u}_1) \quad (4)$$

Avec les contraintes (soit en multipliant les termes par \mathbf{u}_2')
alors $\mu_2 = 0$ (sur équation (5))

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}_2} = 2\mathbf{A}\mathbf{u}_2 - 2\lambda_2 \mathbf{M}\mathbf{u}_2 - \mu_2 \mathbf{M}\mathbf{u}_1 = 0 \quad (5)$$

soit

$$\mathbf{A}\mathbf{u}_2 = \lambda_2 \mathbf{M}\mathbf{u}_2$$

\mathbf{u}_2 second vecteur propre de la matrice $\mathbf{M}^{-1}\mathbf{A}$
(6) correspondant à la seconde plus grande valeur propre :
ce qui se généralise à toutes dimensions : \mathbf{u}_α équation (6)

3.1 Relation d'ajustement dans les deux espaces

Relation de transition entre les deux espaces

$$\text{dans } \mathbb{R}^p \quad \mathbf{X}'\mathbf{X}\mathbf{u}_\alpha = \lambda_\alpha \mathbf{u}_\alpha \quad (1)$$

$$\text{dans } \mathbb{R}^n \quad \mathbf{X}\mathbf{X}'\mathbf{v}_\alpha = \mu_\alpha \mathbf{v}_\alpha \quad (2)$$

Tout vecteur propre normé \mathbf{u}_α de $\mathbf{X}'\mathbf{X}$ relatif à une valeur propre λ_α non nulle correspond au vecteur propre $\mathbf{X}\mathbf{u}_\alpha$ de $\mathbf{X}\mathbf{X}'$ relatif à la même valeur propre λ_α

Car : $(\mathbf{X}\mathbf{X}')\mathbf{X}\mathbf{u}_\alpha = \lambda_\alpha \mathbf{X}\mathbf{u}_\alpha$ avec $\lambda_1 \leq \mu_1$ pour la première valeur propre (1)

Tout vecteur propre normé \mathbf{v}_α de $\mathbf{X}\mathbf{X}'$ relatif à une valeur propre μ_α non nulle correspond au vecteur propre $\mathbf{X}'\mathbf{v}_\alpha$ de $\mathbf{X}'\mathbf{X}$ relatif à la même valeur propre μ_1 (2) d'où $\mu_1 \leq \lambda_1$ pour la première valeur propre (1)

Toutes les valeurs propres non nulles de $\mathbf{X}\mathbf{X}'$ et $\mathbf{X}'\mathbf{X}$ sont égales : $\lambda_\alpha = \mu_\alpha$

Formules de transition entre les 2 espaces : (entre vecteurs propres correspondant à la même valeur propre)

$$\mathbf{v}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{X}\mathbf{u}_\alpha$$

$$\mathbf{u}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{X}'\mathbf{v}_\alpha$$

3.2 Formules de transition entre les deux espaces : coordonnées des points (individus ou var.) dans \mathbb{R}^p ou \mathbb{R}^n

- Dans \mathbb{R}^p , u_α est le α axe factoriel et le vecteur l_α des coordonnées sur cet axe :

$$l_\alpha = X u_\alpha$$

Les nouvelles coordonnées sur chacun des axes factoriels des n individus (entre 1 et n) sont :

$$C_{\text{ind}} = X U$$

- Dans \mathbb{R}^n , v_α est le α axe factoriel et le vecteur φ_α des coordonnées sur cet axe :

$$\varphi_\alpha = X' v_\alpha$$

Les nouvelles coordonnées sur chacun des axes factoriels, des p points-variables(entre 1 et p) sont :

$$C_{\text{var}} = X' V$$

- Les facteurs se calculent par :

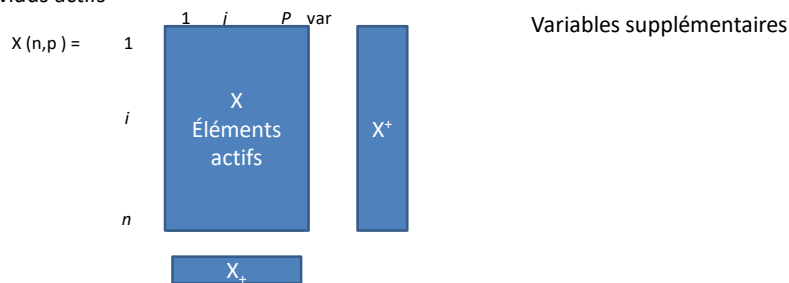
et

$$\begin{cases} l_\alpha = v_\alpha \sqrt{\lambda_\alpha} \\ \varphi_\alpha = u_\alpha \sqrt{\lambda_\alpha} \end{cases}$$

11

4. Projection d'un nouvel individu variable ou individu

Analyse factorielle permet de trouver des sous-espaces de représentation des proximités entre vecteurs de description d'observation construits à partir de var. et individus actifs



Principes des éléments complémentaires : positionner des individus ou de variables (points lignes – points colonnes) n'ayant pas participé à l'analyse → Points n'intervenant pas dans la construction de la réduction

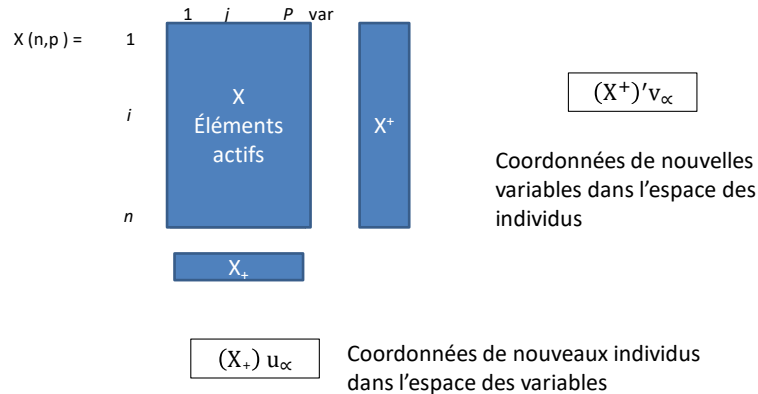
⇒ Ajouter des éléments illustratifs et enrichir

Exemple: positionnement des individus : centre de gravité d'individus, des individus complémentaires - var. écartées initialement etc...

12

4. Projection d'un nouvel individu variable ou individu : *formules de projection*

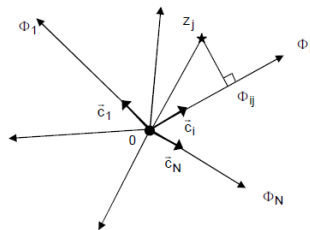
Analyse factorielle permet de trouver des sous-espaces de représentation des proximités entre vecteurs de description d'observation construits à partir de var. et individus actifs



13

5. Dans espace de points colonnes : R^n

Le changement de base orthonormée ($v_1, \dots, v_i, \dots, v_n$) sont des vecteurs unitaires déterminés par leurs coordonnées dans l'ancienne base) permet de calculer, les projections orthogonales (les coordonnées des variables sur les i axes principaux en utilisant le produit scalaire entre le vecteur unitaire v_i et le vecteur Oz où O est l'origine de n axes (n individus dans Z), il y a n axes principaux Φ_i



14

5 (a) Dans espace de points colonnes : R^n

$$\Phi(n,p) \begin{matrix} z_1 & z_j & z_p \\ \Phi_1 & \dots & \dots \\ \dots & \dots & \dots \\ \Phi_i & \dots & \dots \\ \dots & \dots & \dots \\ \Phi_p & \dots & \dots \\ \dots & \dots & \dots \\ \Phi_n & \dots & \dots \end{matrix} \Phi(p,p)$$

Pour généraliser ce résultat à une matrice qui contient toutes les coordonnées il faut tenir compte du résultat (formules de transition) qu'il y a dans cet espace $n-p$ valeurs propres de V qui sont nulles et donc $n-p$ axes principaux qui n'existent pas. La matrice est donc de dimension (p,p) et non (n,p) .

15

5.(b) Projection dans espace des individus variables sur les nouvelles composantes

Soit A la matrice qui contient les coefficients de corrélation linéaires entre les composantes principales F_j et les z_k variables avec $j = 1, \dots, p$; avec U les composantes principales (vecteurs propres dans R^p)

On a : A de dimension (p,p) et $A = 1/n F' Z$ et on a aussi : $A = \Lambda^{1/2} U'$

En retenant pour projection des variables le contenu de A , les coordonnées de ces projections sont toutes en valeur absolue inférieures ou égales à 1, c'est-à-dire sont toutes insérées dans un cercle de rayon unité appelé **cercle des corrélations**.

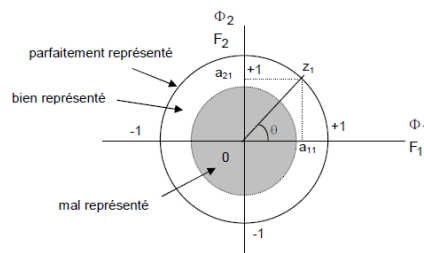
Considérons, par exemple, le cercle de corrélation dans le plan constitué par deux axes principaux et une variable z_1 sur ce cercle. Elle a pour coordonnées a_{11} et a_{21} .

diag $R = \text{diag } A/A$

1^{er} élément $\begin{bmatrix} a_{11}^2 + \dots + a_{p1}^2 \end{bmatrix} = 1 = V(z_1)$

$p^{\text{ème}}$ élément $\begin{bmatrix} a_{1p}^2 + \dots + a_{pp}^2 \end{bmatrix} = 1 = V(z_p)$

$\Sigma \quad \lambda_1 \quad \lambda_j \quad \text{Trace} = p$



5. Validation et robustesse en ACP

En analyse de données : un échantillon pour lequel, pas ou peu d'hypothèses sur la population

La réduction de dimension : Interprétation des relations entre variables , regroupement d'individus et projection

Evaluer si la projection faite (retenir $k \ll p$) est pertinente à partir de critères de qualité et de contribution

MAIS

Résultat est-il significatif (en terme en statistique) ? : des interprétations sont faites avec des méthodes d'analyse qui reposent sur un échantillon et notamment une matrice de données X :

Après réduction de la dimension, qu'est-ce qui fait que l'analyse est stable (robustesse de la méthode) et validité des résultats obtenus si on perturbe un peu les données :

→ appartient aux méthodes de validation

18

5.1 Validation et robustesse : pourquoi ?

'Out put' de l'ACP/AF :

- Matrice de corrélation/covariance est obtenue sur l'échantillon X
- Le seuil choisi pour retenir que k composante est-il judicieux (cad le choix de l'inertie expliquée retenue est il fiable ?) dépend de la matrice de covariance (cor)
- Le système de projection des individus est-elle utilisable pour de nouveaux individus et nouvelles variables ? Généralisation ou extension (prédicteur)

- Mettre en œuvre des techniques pour rendre robuste les résultats obtenus d'une méthode quelconque (ACP et AF)
 - 1) résultats théoriques liés à la méthode AFC (vrai en Analyse factorielle)
 - 2) approche empirique de validation (méthode pragmatique , systématique et universelle (*bootstrap* et *validation croisée*))

19

5.2 Stabilité en ACP (1)

Stabilité concerne :

- Les facteurs (axes factoriels)
- Les valeurs propres

Stabilité : lorsque l'on apporte des modifications aux données (matrice X)

→ Comment se fait la variation maximale des facteurs et des valeurs propres

- **Evaluer de façon empirique**

MAIS

- **Résultats théoriques** existent dans le cas de perturbations symétriques pour de matrice symétriques
→ Théorème de Wielandt-Hoffman

20

5.3 Petites variations des valeurs propres et vecteurs propres d'une matrice symétrique

Théorème de Wielandt-Hoffman

Si **A**, **B**, **C** sont des matrices symétriques (p,p) ayant respectivement pour valeurs propres classées par ordre décroissant $\alpha_i, \beta_i, \gamma_i$ (pour A, B, C),

Et que :

$$\mathbf{C} = \mathbf{A} + \mathbf{B}$$

Alors :

$$\sum_{i=1}^p (\gamma_i - \alpha_i)^2 \leq \sum_{i=1}^p \beta_i^2$$

Soit un autre théorème :

pour tout i tel que $1 \leq i \leq p$

$$\alpha_i + \beta_p \leq \gamma_i \leq \alpha_i + \beta_1$$

21

5.4 Petites variations des valeurs propres et vecteurs propres d'une matrice symétrique

Soit montrer comment des variations de la matrice symétrique à diagonaliser \mathbf{A} de terme général a_{ij} influencent les éléments propres (\mathbf{u} et λ)

Supposons que la matrice \mathbf{A} et par conséquent la matrice \mathbf{U} et Λ dépendent continument du paramètre s

Rappel : $\mathbf{A} \mathbf{u} = \lambda \mathbf{u}$,

devient : $\mathbf{A} \mathbf{U} = \Lambda \mathbf{U}$ (1)

pour l'ensemble du spectre avec comme contrainte $\mathbf{U}' \mathbf{U} = \mathbf{I}$

La dérivation de (1): $\frac{\partial \mathbf{A}}{\partial s} \mathbf{U} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial s} = \frac{\partial \mathbf{U}}{\partial s} \Lambda + \mathbf{U} \frac{\partial \Lambda}{\partial s}$

Et
$$\frac{\partial \mathbf{U}'}{\partial s} \mathbf{U} + \mathbf{U}' \frac{\partial \mathbf{U}}{\partial s} = 0$$

$$\mathbf{U}' \frac{\partial \mathbf{A}}{\partial s} \mathbf{U} = \frac{\partial \Lambda}{\partial s} + \left(\mathbf{U}' \frac{\partial \mathbf{U}}{\partial s} \Lambda - \Lambda \mathbf{U}' \frac{\partial \mathbf{U}}{\partial s} \right) \quad (2)$$

De la forme $(B\Lambda - \Lambda B)$ avec des éléments diagonaux nuls

D'où :
$$\frac{\partial \Lambda}{\partial s} = \text{diag}(\mathbf{U}' \frac{\partial \mathbf{A}}{\partial s} \mathbf{U}) \quad (3)$$

22

5.5 Petites variations des valeurs propres et vecteurs propres d'une matrice symétrique

Pour les valeurs propres la variation de la valeur propre est liée à une perturbation d'un élément de la matrice carré \mathbf{A} soit :

$$\frac{\partial \lambda_r}{\partial s} = \sum_{ij} u_{ir} u_{jr} \cdot \frac{\partial a_{ij}}{\partial s} = q_{rr}$$

A partir de

$$\mathbf{U}' \frac{\partial \mathbf{A}}{\partial s} \mathbf{U} = \frac{\partial \Lambda}{\partial s} + \left(\mathbf{U}' \frac{\partial \mathbf{U}}{\partial s} \Lambda - \Lambda \mathbf{U}' \frac{\partial \mathbf{U}}{\partial s} \right)$$

Et en posant :

$$Q = \mathbf{U}' \frac{\partial \mathbf{A}}{\partial s} \mathbf{U} \text{ d'où } q_{rr} = \sum_{i,j} u_{ir} u_{jr} \frac{\partial a_{ij}}{\partial s}$$

Pour les vecteurs, cela conduit à :

$$\frac{\partial u_{jr}}{\partial s} = \sum_{t \neq r} u_{jt} \frac{q_{rt}}{\lambda_r - \lambda_t}$$

Qui montrent le lien avec les valeurs propres et notamment que des valeurs propres voisines créent une instabilité des vecteurs propres correspondants car cet écart est au dénominateur entre les valeurs propres du terme à droite

- ✓ La variation des valeurs propres ne dépend pas des variations des vecteurs propres
- ✓ Les Variations des composantes d'un vecteur propre dépend des écarts entre la valeur propre qui lui est associée et les autres valeurs propres (incertitude sur la détermination des axes principaux)

23

Bootstrap : technique de validation

Idée de la validation : cad l'étude de la stabilité d'une réponse (*ici réduction et compressions de données*)

Méthodes de validation empirique (*par expérience numérique*)

- Méthode générale de validation d'une méthode
- Une technique de type statistique (échantillonnage)

Principe : Le calcul de la stabilité se fait par simulation de perturbations des données

Idée fondamentale de la stabilité :

Compte tenu des erreurs attendues dans les données , on génère des tableaux perturbés et on vérifie la stabilité des représentations obtenues

Bootstrap : technique de simulation particulière fondée sur la distribution empirique de l'échantillon de base :

- bootstrap non paramétrique (la première définition) : *non-parametric bootstrap*
- bootstrap paramétrique : simulation avec une distribution théorique et des paramètres calculés à partir de l'échantillon (*parametric bootstrap*)

Introduit par Efron (1979) puis codéveloppé avec Tibshirani (depuis 1993)

24

1 Bootstrap: méthode d'estimation méthode de ré échantillonnage

Principe :

Pour estimer la confiance que l'on accorde à l'estimation d'un paramètre inconnu noté θ , on simule b échantillons de même taille n de l'échantillon initial

Ces échantillons ou répliques sont obtenus par tirage au hasard **avec remise** parmi les n individus observés au départ , tous les n individus ont une même probabilité d'être choisis soit, une probabilité égale à $1/n$

Certains individus observés au départ, auront un poids plus fort s'ils sont sélectionnés plusieurs fois alors d'autres absents auront un poids nul

Chaque réplique k , donne lieu à une estimation du paramètre $\theta^*(k)$ du paramètre θ

Calcul de la distribution statistique de nouvel échantillon des b répliques θ^*

Ainsi le calcul de la variance empirique des b quantités $\theta^*(k)$ et sous conditions assez générales , on peut estimer de façon satisfaisante la variance de l'estimateur

NB : A l'origine, ou l'évaluation de la précision d'un estimateur par Efron

25

2 Bootstrap non paramétrique

- n i.i.d $x = (x_1, x_2, \dots, x_n)$

- On est intéressé par une statistique $s(x)$

- On tire B échantillons $x_b^* = (x_1^*, x_2^*, \dots, x_n^*)$

Où chaque x^* est constitué en tirant avec remise n valeurs parmi les x_i intéressé par une statistique $s(x)$

Exemple :

$$x^* = (x_7, x_1, x_2, x_7, x_3, \dots, x_4)$$

- On calcule chaque valeur de $s(x_b^*)$ pour chaque bootstrap b

- On calcule

$$s(.) = \sum_{b=1}^B s(x_b^*) / B$$

- Estimation de l'erreur-standard :

$$\hat{se}_{boot} = \left\{ \sum_{i=1}^B (s(x_i^*) - s(.))^2 / (B-1) \right\}^{1/2}$$

26

3 Cas de l'erreur standard d'une moyenne

Théorie probabiliste : (population) \rightarrow propriétés de l'échantillon

Inférence statistique : (observé) \rightarrow propriétés de l'échantillon

- ✓ Apprendre de X en observant x
- ✓ Quelle précision d'un statistique observée ?

Evaluation de l'Erreur standard d'une moyenne :

soit x une var. aléatoire de distribution F

Espérance et variance de F : $\mu_F = E_F(x)$

$$\sigma_F^2 = \text{var}_F(x) = E_F[(x - \mu_F)^2]$$

On tire au hasard un échantillon x de taille n à partir de F :

La moyenne de l'échantillon \bar{x} avec $\bar{x} = \sum_{i=1}^n x_i / n$ a pour espérance μ_F et pour variance

$$\sigma_F^2 / n, \text{ soit } \bar{x} \sim (\mu_F, \sigma_F^2 / n)$$

Erreur standard de la moyenne est : $se_F(\bar{x}) = se(\bar{x}) = \sqrt{\text{var}_F(\bar{x})} = \sigma_F / n$

Si n est assez grand, \bar{x} suit une loi normale d'espérance et de variance $\bar{x} \sim \mathcal{N}(\mu_F, \sigma^2, n)$

En autre terme : $\text{Prob}\left\{|\bar{x} - \mu_F| < \frac{\sigma}{\sqrt{n}}\right\} \doteq 0.683$

$$\text{Prob}\left\{|\bar{x} - \mu_F| < 2 \frac{\sigma}{\sqrt{n}}\right\} \doteq 0.954$$

27

4 Estimation par Bootstrap de l'erreur standard d'une moyenne (a)

- On a tiré un échantillon $x = (x_1, x_2, \dots, x_n)$ d'une fonction de distribution inconnue F

$$F \rightarrow x = (x_1, x_2, \dots, x_n)$$
- On veut estimer un paramètre $\theta = t(F)$ à partir de x
- On calcule un estimateur $\hat{\theta} = s(x)$
- \hat{F} est la fonction de distribution qui donne la probabilité $1/n$ à chaque x_i
- On tire des échantillons bootstrap à partir de

$$\hat{F} \rightarrow x_b^* = (x_1, x_2, \dots, x_n)$$
- On calcule la copie bootstrap $\hat{\theta}^* = s(x^*)$

L'estimation bootstrap idéale de l'erreur standard $se_F(\hat{\theta})$ est l'erreur standard de $\hat{\theta}$ pour des ensembles des données de taille n tirés suivant \hat{F} cad $se_{\hat{F}}(\hat{\theta}^*)$

28

5. Estimation par Bootstrap de l'erreur standard d'une moyenne (b) algorithmme

Algorithme d'estimation des erreurs standards

- 1) Tirer B échantillons bootstrap $x^{*1}, x^{*2}, \dots, x^{*B}$ à partir de x
- 2) Calculer la copie Bootstrap $\hat{\theta}^* = s(x^{*B})$
- 3) Calculer l'erreur-standard pour les B copies

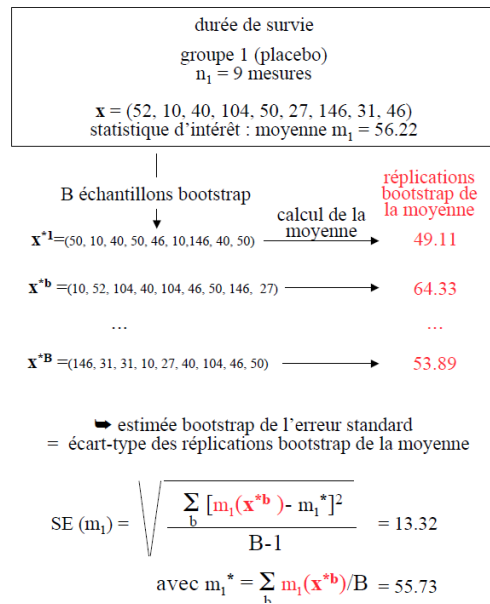
$$\hat{se}_{boot} = \left\{ \frac{\sum_{b=1}^B (\hat{\theta}^*(b) - \hat{\theta}^*(.))^2}{(B-1)} \right\}^{1/2}$$

$$\text{Avec } \hat{\theta}^*(.) = \sum_{b=1}^B \hat{\theta}^*(b) / B$$

L'estimation *bootstrap non paramétrique* de l'erreur-standard \hat{se}_B a pour limite $se_F(\hat{\theta})$ quand B tend vers l'infini

29

6. Erreur standard de la moyenne : un exemple (c)



30

7. Erreur standard de la moyenne : un exemple (d)

durée de survie	
groupe 1 (placébo) $n_1 = 9$ mesures	groupe 2 (traitement) $n_2 = 7$ mesures
52, 10, 40, 104, 50, 27, 146, 31, 46	94, 38, 23, 197, 99, 16, 141
moyenne $m_1 = 56.22$ médiane $\mu_1 = 46$	moyenne $m_2 = 86.86$ médiane $\mu_2 = 94$

erreur standard sur m_1 :

↳ classique : $se_1 = 14.14$

↳ bootstrap : $se_1^* = 13.32$

erreur standard sur m_2 :

↳ classique : $se_2 = 25.24$

↳ bootstrap : $se_2^* = 23.81$

erreur standard sur μ_1 :

↳ classique : ?

↳ bootstrap : $se_1^* = 11.54$

erreur standard sur μ_2 :

↳ classique : ?

↳ bootstrap : $se_2^* = 36.35$

In I. Buvat 2000

erreur standard sur n'importe quelle statistique

Approche classique : pas toujours possible

Par bootstrap : TOUJOURS UNE SOLUTION au prix d'un peu de calcul

31

8. Bootstrap : son usage

Deux applications fondamentales :

- 1) La réduction du biais d'un estimateur $\theta(F)$** à partir de son estimation sur un échantillon auquel est associé F_0 et l'estimateur $\theta(F_0)$ de $\widehat{\theta(F)}$
- $$\text{Biais} = E\{\theta(F) - \theta(F_0) | F\}$$

2) Calcul de l'Intervalle de confiance

Pour une loi F inconnue dont on veut estimer un paramètre $\theta(F)$ par un intervalle de confiance à 0,95 et de la loi F_0 associée à l'échantillon observé ou $\theta(F_0)$ est l'estimateur de $\theta(F)$

Pour obtenir à partir de $\theta(F_0)$, un intervalle de confiance pour $\theta(F)$, besoin de connaître la loi $\theta(F) - \theta(F_0)$ sous F ou une approximation

Alors les bornes de l'intervalle avec $t_1 = t_0.0.025$ et $t_2 = t_0.0.975$ les quantiles 0.025 et 0.975 de cette loi $[\theta(F_0) + t_1; \theta(F_0) + t_2]$:

$$P(\theta(F) - \theta(F_0) < t_1) = 0.025$$

$$P(\theta(F) - \theta(F_0) > t_2) = 0.025$$

$$P(\theta(F_0) + t_1 \leq \theta(F) \leq \theta(F_0) + t_2) = 0.925$$

→Valider la substitution par la loi $\theta(F_0) - \theta(F_1)$ sous F_0

32

9. Bootstrap en ACP

Estimation de quels paramètres en ACP ?

- ✓ Estimation du seuil de $k \ll p$ valeurs propres et taux d'inertie expliquée
- ✓ Estimation du coefficient de corrélation/covariance
- ✓ Estimation des composantes (vecteurs propres et valeurs propres) et de la projection des n individus et p variables

Définir le nombre de tirages de bootstrap :

B = 25 à 50 :début d'information

B = 200 :estimer l'erreur-standard

B = 500 estimer intervalle de confiance

→Suivre la stabilisation des résultats en fonction du la taille de B

33

10. En analyse factorielle - ACP: 1) Quantité d'inertie expliquée par un axe

Exemple

88 étudiants avec 5 notes d'examen

Matrice de données : $X(88,5)$

De moyenne empirique : $\bar{X} = (38.95, 50.59, 50.60, 46.68, 42.31)$

Matrice de covariance : A

Calcul des 5 valeurs propres de A :

$$\hat{\lambda}_1 = 679.2 \quad \hat{\lambda}_2 = 199.8 \quad \hat{\lambda}_3 = 102.6 \quad \hat{\lambda}_4 = 83.7 \quad \hat{\lambda}_5 = 31.8$$

$$QI = \hat{\theta} = \frac{\hat{\lambda}_1}{\sum_{i=1}^5 \hat{\lambda}_i} \quad \hat{\theta} = 0.619$$

Si on retient 1 composante, QI doit être proche de 1: Quelle est la précision de $\hat{\theta}$?

Chaque échantillon du Bootstrap sera une matrice X^* de 5 colonnes et 88 lignes tiré au hasard parmi les x_i

Calcul de A

Calcul des 5 $\hat{\lambda}_i^*$ puis de $\hat{\theta}^* = \frac{\hat{\lambda}_1^*}{\sum_{i=1}^5 \hat{\lambda}_i^*}$ tracer histogramme des B valeurs de QI^*

$$\bar{\theta}^* = 0.625 \quad \widehat{se}_{200} = 0.047$$

L'intervalle de confiance standard de la vraie valeur de θ de probabilité $(1-2\alpha)$ est :

$\theta \in \hat{\theta} \pm z^{(1-\alpha)} \widehat{se}$ ou $z^{(1-\alpha)}$ est les $100(1-\alpha)$ -ième percentile de la loi normale centrée réduite

$$\theta \in 0.619 \pm 0.047 = [0.572; 0.666] \text{ avec une probabilité de } 0.683$$

$$\theta \in 0.619 \pm 1.645 \times 0.047 = [0.542; 0.696] \text{ avec une probabilité de } 0.954$$

34

11. Estimation des intervalles de confiance (a)

Estimation de l'intervalle de confiance

Estimateur $\hat{\theta}$ et son erreur standard $se_F(\hat{\theta})$ son IC usuel à 90% est $\hat{\theta} \pm 1.64 se_F(\hat{\theta})$, suppose que $(\theta(F_0) - \hat{\theta}) / se_F(\hat{\theta})$ suit une loi normale centrée réduite

Deux approximations : 1^{ère}) si l'estimateur du paramètre $\hat{\theta}$ suit une loi normale et la 2^{ème}) que $Z = (\theta(F_0) - \hat{\theta}) / se_F(\hat{\theta})$ suit une loi normale centrée réduite

→ Alors que si $\hat{\theta}$ à une distribution normale, Z suit un loi de Student de degré $n-1$

OR: Avec le Bootstrap, l'hypothèse de normalité n'est pas nécessaire ce qui permet de calculer l'intervalle de confiance : généralisation de la méthode du test de Student

→ **Bootstrap-t** étudie la distribution de Z sur les données :

- pour chaque b échantillons générés par les B tirage de bootstrap

$$- \text{calculer pour chaque } b \quad Z^{*b} = \frac{\hat{\theta}^* b - \hat{\theta}}{se(\hat{\theta}^* b)} \times \frac{1}{\sqrt{n}}$$

- trier les valeurs de Z^{*b} par ordre croissant

- le α ième fractile de la loi Z est estimé par la valeur \hat{t}^α telle que $\alpha = \# \{Z^{*b} \leq \hat{t}^\alpha\} / B$ et l'intervalle bootstrap $IC(2\alpha)$

$$\text{avec} \quad IC_{inf} = \hat{\theta} - Z^{*(\frac{B(1-\alpha)}{2})} \times se_F \times \frac{1}{\sqrt{n}}$$

$$IC_{sup} = \hat{\theta} - Z^{*(B \times \alpha / 2)} \times se_F \times \frac{1}{\sqrt{n}}$$

35

12. Estimation des intervalles de confiance (b)

OR : Estimation de l'intervalle de confiance est applicable à des statistiques ponctuelles comme la moyenne d'un échantillon mais non applicable pour des statistiques plus complexes pour lesquelles on ne dispose pas d'estimateur statistique (comme la médiane , un coefficient de corrélation...)

Méthode par quantile

L'intervalle de confiance basé sur les quantiles bootstrap :

- pour chaque b échantillons générés par les B tirage de bootstrap
 - calculer pour chaque b , *repliquat bootstrap* $\hat{\theta}^{*b}$
 - trier les valeurs de $\hat{\theta}^{*b}$ par ordre croissant
 - soit le $\hat{\theta}_B^{*(\alpha)}$ est le fractile empirique d'ordre α des $\hat{\theta}^{*b}$, soit la B -ième valeurs dans la liste ordonnée des B reliquats de $\hat{\theta}^*$.
- L'intervalle approximé de niveau $(1-2\alpha)$ est alors

$$IC = \left[\hat{\theta}^{*\left(\frac{B \times (\alpha)}{2}\right)}; \hat{\theta}^{*\left(\frac{B(1-\alpha)}{2}\right)} \right]$$

36

Mise en œuvre en TP

1. Réduction en parallèle sur les 2 espaces
2. Dualité
3. Robustesse de la méthode ACP
(théorique et par *bootstrap*)
4. Projection de nouveaux individus
5. Cercle des corrélations

37

ACP : Approfondissements (2)

1. Réduction
2. Dualité
3. Robustesse de la méthode
(théorique et par *bootstrap*)
4. Décomposition
5. Prédiction
6. ACP et extensions

39

2.1 Retour sur les Relations d'ajustement dans les deux espaces

Relation de transition entre les deux espaces

$$\text{dans } \mathbb{R}^p \quad \mathbf{X}'\mathbf{X} \mathbf{u}_\alpha = \lambda_\alpha \mathbf{u}_\alpha \quad (1)$$

$$\text{dans } \mathbb{R}^n \quad \mathbf{X}\mathbf{X}' \mathbf{v}_\alpha = \mu_\alpha \mathbf{v}_\alpha \quad (2)$$

Tout vecteur propre normé \mathbf{u}_α de $\mathbf{X}'\mathbf{X}$ relatif à une valeur propre λ_α non nulle correspond au vecteur propre $\mathbf{X}\mathbf{u}_\alpha$ de $\mathbf{X}\mathbf{X}'$ relatif à la même valeur propre λ_α

Car : $(\mathbf{X}\mathbf{X}') \mathbf{X}\mathbf{u}_\alpha = \lambda_\alpha \mathbf{X}\mathbf{u}_\alpha$ avec $\lambda_1 \leq \mu_1$ pour la première valeur propre (1)

Tout vecteur propre normé \mathbf{v}_α de $\mathbf{X}\mathbf{X}'$ relatif à une valeur propre μ_α non nulle correspond au vecteur propre $\mathbf{X}' \mathbf{v}_\alpha$ de $\mathbf{X}'\mathbf{X}$ relatif à la même valeur propre μ_1 (2) d'où $\mu_1 \leq \lambda_1$ pour la première valeur propre (1)

Toutes les valeurs propres non nulles de $\mathbf{X}\mathbf{X}'$ et $\mathbf{X}'\mathbf{X}$ sont égales : $\lambda_\alpha = \mu_\alpha$

Formules de transition entre les 2 espaces : (entre vecteurs propres correspondant à la même valeur propre)

$$\mathbf{v}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{X} \mathbf{u}_\alpha$$

$$\mathbf{u}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{X}' \mathbf{v}_\alpha$$

40

2.2 Formules de transition entre les deux espaces : coordonnées des points (individus ou var.) dans \mathbb{R}^p ou \mathbb{R}^n

- Dans \mathbb{R}^p , \mathbf{u}_α est le α axe factoriel et le vecteur l_α des coordonnées sur cet axe :

$$l_\alpha = \mathbf{X} \mathbf{u}_\alpha$$

Les nouvelles coordonnées sur chacun des axes factoriels des n individus (entre 1 et n) sont :

$$\mathbf{C}_{\text{var}} = \mathbf{X} \mathbf{U}$$

- Dans \mathbb{R}^n , \mathbf{v}_α est le α axe factoriel et le vecteur φ_α des coordonnées sur cet axe :

$$\varphi_\alpha = \mathbf{X}' \mathbf{v}_\alpha$$

Les nouvelles coordonnées sur chacun des axes factoriels, des p points-variables(entre 1 et p) sont :

$$\mathbf{C}_{\text{ind}} = \mathbf{X}' \mathbf{U}$$

- Les facteurs se calculent par :

$$\begin{cases} l_\alpha = \mathbf{v}_\alpha' \sqrt{\lambda_\alpha} \\ \varphi_\alpha = \mathbf{u}_\alpha' \sqrt{\lambda_\alpha} \end{cases}$$

et

41

3. Reconstitution des données de départ (a)

Avec \mathbf{u}_α le α ième vecteur propre de norme 1 de la matrice $\mathbf{X}'\mathbf{X}$ correspondant à la valeur propre λ_α et \mathbf{v}_α le α ième vecteur propre de norme 1 de la matrice $\mathbf{X}\mathbf{X}'$:

$$l_\alpha = \mathbf{X} \mathbf{u}_\alpha = \mathbf{v}_\alpha' \sqrt{\lambda_\alpha}$$

Avec une postmultiplication de cette relation par \mathbf{u}'_α et la somme sur les p axes :

$$\mathbf{X} \left\{ \sum_{\alpha=1}^p \mathbf{u}_\alpha \mathbf{u}'_\alpha \right\} = \sum_{\alpha=1}^p \sqrt{\lambda_\alpha} \mathbf{v}_\alpha \mathbf{u}'_\alpha$$

Avec \mathbf{U} ($p \times p$) contenant les vecteurs propres de $\mathbf{X}'\mathbf{X}$; vecteurs orthogonaux et de norme 1 :

$$\mathbf{U}\mathbf{U}' = \mathbf{I}_p \text{ et } \mathbf{U}'\mathbf{U} = \mathbf{I}_p \text{ (matrice unité d'ordre } p)$$

Or :

$$\sum_{\alpha=1}^p \mathbf{u}_\alpha \mathbf{u}'_\alpha = \mathbf{U}\mathbf{U}'$$

D'où : les valeurs propres λ_α étant toujours rangées par ordre décroissante :

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

42

3. Reconstitution des données de départ (b)

les valeurs propres λ_α étant toujours rangées par ordre décroissante :

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

Est une **formule de reconstitution du tableau de données X** à partir des valeurs propres λ_α et des vecteurs propres \mathbf{u}_α et \mathbf{v}_α

Ou formule de **décomposition aux valeurs singulières** (Eckart et Young 1936)

$$\mathbf{X} = \sqrt{\lambda_1} \mathbf{v}_1 \mathbf{u}'_1 + \dots + \sqrt{\lambda_\alpha} \mathbf{v}_\alpha \mathbf{u}'_\alpha + \dots + \sqrt{\lambda_p} \mathbf{v}_p \mathbf{u}'_p$$

43

4. Valeurs propres sur matrice carrée

Pour certaines matrices carrées, A on peut faire une décomposition en valeurs propres:

$$A = X D X^{-1},$$

où D est une matrice diagonale de valeurs propres et X est une matrice inversible de vecteurs propres.

Quand on fait le produit matrice-vecteur $Ax = (XDX^{-1})x$

- On prend $x \rightarrow$ on l'exprime dans la base donnée par les vecteurs propres ($X^{-1}x$)
- On multiplie les éléments de ce vecteur par les valeurs propres D une à une, et on refait le changement de base inverse en multipliant par X.
- Si on a un système linéaire $Ax = b$, on peut effectuer les changements de base $\hat{x} = X^{-1}x$, et $\hat{b} = X^{-1}b$, on obtient le

Système $D\hat{x} = \hat{b}$,

Limitations: pour effectuer cette transformation, la matrice A doit être carrée et diagonalisable.

45

5. Décomposition en valeurs singulières (a)

La décomposition en valeurs singulières est similaire à la décomposition en valeurs propres, mais fonctionne pour n'importe quelle matrice A de taille $m \times n$: on factorise A en produit de trois matrices:

$$A = U \Sigma V^*,$$

avec U une matrice $m \times m$ unitaire, V une matrice $n \times n$ unitaire et une matrice $m \times n$ diagonale avec coefficients réels et positifs.

Interprétation géométrique

L'image d'une sphère unité dans R^n par une matrice A $m \times n$ est une hyperellipse dans R^m .

Soit les longueurs des axes principaux $\{\sigma_1, \sigma_2, \dots, \sigma_m\}$ et les directions des axes principaux $\{u_1, u_2, \dots, u_m\}$.

On a $\sigma_i \geq 0$ et $u_i \in R^m$, avec l'ensemble $\{u_i\}$ orthonormal (on peut prendre $\|u_i\| = 1$).

Les vecteurs $\sigma_i u_i$ sont alors les semi-axes principaux de l'hyperellipse. Si $\text{rang } A = r$, alors exactement r valeurs de σ_i seront non-nulles. Si $m \times n$, alors au plus n axes seront de longueurs positives.

46

5. Décomposition en valeurs singulières (b)

A de taille $m \times n$: $A = U \Sigma V^*,$

avec U une matrice $m \times m$ unitaire, V une matrice $n \times n$ unitaire et une matrice $m \times n$ diagonale avec coefficients réels et positifs.

Soit les longueurs des axes principaux $\{\sigma_1, \sigma_2, \dots, \sigma_m\}$ et les directions des axes principaux $\{u_1, u_2, \dots, u_m\}$.

On a $\sigma_i \geq 0$ et $u_i \in R^m$, avec l'ensemble $\{u_i\}$ orthonormal (on peut prendre $\|u_i\| = 1$).

Les vecteurs $\sigma_i u_i$ sont alors les semi-axes principaux de l'hyperellipse.

Si $\text{rang } A = r$, alors exactement r valeurs de σ_i seront non-nulles. Si $m \times n$, alors au plus n axes seront de longueurs positives.

les n **vecteurs singuliers à gauche** de A ou les n **vecteurs de sortie** de A , $\{u_1, u_2, \dots, u_n\}$ orientés dans les directions des semi-axes principaux, où u_i est la direction du semi-axe de longueur i .

les n **vecteurs singuliers à droite** de A ou les n **vecteurs d'entrée** de A $\{v_1, v_2, \dots, v_n\}$ qui sont les pré-images des semi-axes principaux: $Av_i = \sigma_i u_i$, pour $i = 1, \dots, n$. Sous forme matricielle:

$$A \begin{bmatrix} | & | & | & | & | \\ v_1 & v_2 & \dots & v_n \\ | & | & | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | & | & | \\ u_1 & u_2 & \dots & u_n \\ | & | & | & | & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_n & \\ & & & & \ddots \end{bmatrix}, \quad AV = \hat{U} \hat{\Sigma}.$$

47

5. Décomposition en valeurs singulières (c)

$$A = U \Sigma V^*$$

A de taille $m \times n$:

avec U une matrice $m \times m$ unitaire, V une matrice $n \times n$ unitaire et une matrice $m \times n$ diagonale avec coefficients réels et positifs.

$$A \begin{bmatrix} | & | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & | & \cdots & | \end{bmatrix} = \begin{bmatrix} | & | & | & \cdots & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & | & \cdots & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_n & \\ & & & & 0 \end{bmatrix}, \quad \begin{aligned} AV &= \hat{U} \hat{\Sigma}. \\ A &= \hat{U} \hat{\Sigma} V^*. \end{aligned}$$

Factorisation appelée décomposition en valeurs singulières réduite (les colonnes de U ne forment pas une base ($m > n$)) ; si la matrice diagonale et la matrice de U est augmentée des $m-n$ lignes de zéros, on obtient la **décomposition en valeurs singulières complète**

$$A = \hat{U} \hat{\Sigma} V^*.$$

Théorème 3.1. existence et unicité de la SVD. Toute matrice $A \in \mathbb{C}^{m \times n}$ possède une SVD. Les valeurs singulières $\{\sigma_i\}$ sont déterminées de façon unique. Si A est carrée ($m = n$) et les valeurs singulières σ_j sont distinctes, les vecteurs d'entrée et de sortie v_j et u_j sont déterminés de façon unique à un facteur complexe unité près.

Théorème 3.2. Les valeurs singulières d'une matrice A sont les racines carrées des valeurs propres non-nulles de A^*A et AA^* .

Preuve. $A^*A = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma^*U^*U\Sigma V^* = V\Sigma^*\Sigma V^*$. La matrice A^*A est donc semblable à $\Sigma^*\Sigma$, ce qui implique qu'elles ont les mêmes valeurs propres. Les valeurs propres de $\Sigma^*\Sigma$ sont $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$.

Théorème 3.3. Les colonnes de U sont les vecteurs propres orthogonaux de AA^* et les colonnes de V sont les vecteurs propres orthogonaux de A^*A à unité complexe près.

$$A = U \Sigma V^*.$$

48

ACP comme décomposition du signal

les valeurs propres λ_α étant toujours rangées par ordre décroissante :

$$X = \sum_{\alpha=1}^p \sqrt{\lambda_\alpha} v_\alpha u'_\alpha$$

Est une **formule de reconstitution du tableau de données X** à partir des valeurs propres λ_α et des vecteurs propres u_α et v_α

Ou formule de **décomposition aux valeurs singulières (Eckart et Young 1936)**

$$X = \sqrt{\lambda_1} v_1 u'_1 + \dots + \sqrt{\lambda_\alpha} v_\alpha u'_\alpha + \dots + \sqrt{\lambda_p} v_p u'_p$$

$\begin{matrix} X & & v_1 & u'_1 & & v_2 & u'_2 & & v_p & u'_p \end{matrix}$

Ou encore en SVD : $X = U \Sigma V^*$

Avec Σ est la matrice diagonale de racines des valeurs propres, U la matrice des vecteurs singuliers à gauche de X (vecteurs propres XX') et V les vecteurs singuliers à droite de X (vecteurs propres de la matrice $X'X$)

Application sur Donnée temporelles sur le dernier TP conjoint ACP et séries temporelles

49