Chp 3 : Analyse en composantes principales
A.C.P

I / Introduction

En vue d'éliminer la redondance d'information due aux éventuelles liaisons pouvant exister entre les p variables initiales, on se propose de remplacer le tableau X de départ par un tableau Y, où le nombre de variables est plus petit.

$$(X^1, \ldots, X^p) = \underset{(n,p)}{X} \longrightarrow \underset{(n,q)}{Y} = (Y^1, \ldots, Y^q) \text{ avec } q << p.$$

Rq : les nouvelles variables sont non corrélées

Généralement, en pratique, les variables initiales ne sont pas toutes de la même grandeur, et comme les nouvelles variables sont des combinaisons linéaires de variables initiales, les variables prenant les petites valeurs seront ~~sans~~ negligées devant les variables à grandes valeurs. A.C.P répond à ce problème, en utilisant l'A.F.G du tableau centré X, et munir chaque individu d'un poids p_i $\left(\sum_{i=1}^{n} p_i = 1\right)$

II / Position du problème

Soit X un tableau croisant n individus et p variables

$$\underset{(n,p)}{X} = (x_i^j)_{\substack{1 \le i \le n \\ 1 \le j \le p}}$$

On note par $\underset{(n,p)}{\tilde{X}}$ le tableau centré, $\underset{(n,p)}{\tilde{X}} = (\tilde{x}_i^j)_{\substack{1 \le i \le n \\ 1 \le j \le p}}$

où $\tilde{x}_i^j = x_i^j - \overline{x}^j$

But de l'A.C.P.

L'oeil humain ne sachant regarder dans un espace de dimension supérieure à 3, le but de l'A.C.P est de décrire et visualiser le nuage $N(I) = \{ \underline{X}_i \,/\, {}^t\underline{X}_i \in \mathbb{R}^p \}$ et ce en cherchant un sous espace vectoriel de dimension q $(q < p)$ le plus proche de $N(I)$

Methode

1) Chercher la droite Δu_1, le plus proche de $N(I)$ (Δu_1 est un s.e.v de $\mathbb{R}^p$ de dimension 1, engendré par $\{u_1\}$)

2) Recherche de la droite Δu_2, orthogonale à Δu_1, la plus proche de $N(I)$

Proposition : Le plan le plus proche de $N(I)$ est le plan engendré par les vecteurs $\{u_1, u_2\}$.

3) Recherche de la droite Δu_3, orthogonale à Δu_1 et Δu_2 la plus proche du nuage.

⋮

Ainsi de suite jusqu'à.

⋮

q) Recherche de la droite Δu_q, orthogonale à $\Delta u_1, \dots, \Delta u_{q-1}$, la + proche du nuage.

Déf : Les droites $\Delta u_1, \dots, \Delta u_q$ qui ajuste au mieux le nuage sont dites axes factoriels ou axes principaux.

Proposition : le s.e.v de dimension q qui ajuste au mieux les p^ts du nuage $N(I)$ est le s.e.v engendré par $\{u_1, \dots, u_q\}$

$$A.C.P \text{ du tableau } X \iff A.F.G \text{ de } \tilde{X} \text{ où } \begin{cases} (\mathbb{R}^p, I_p) \\ (\mathbb{R}^n, D_p) \end{cases}$$

$$\underset{(p,p)}{D_p} = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix} \text{ metrique des poinds}$$

II/ On se place sur $\mathbb{R}^p$.

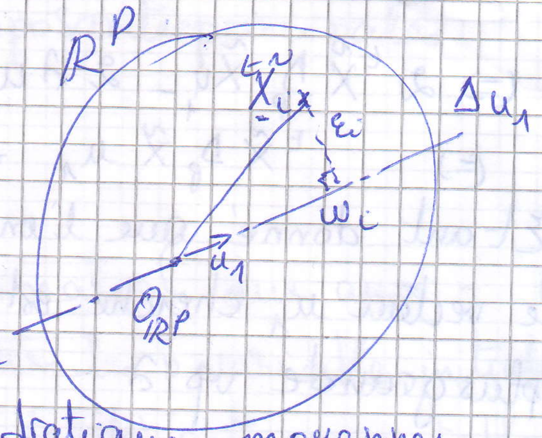
On considère le nuage des n points de $\mathbb{R}^p$, $N(I) = \{(\tilde{X}_i, p_i), \tilde{X}_i \in \mathbb{R}^p\}$

$N(I)$ nuage des n individus

1/ Recherche du s.e.v de dim 1

On note par u_1 le vecteur de la base de ce s.e.v de dim 1 noté Δu_1, on supose que u_1 est normé.

w_i est la projection I_p-orthogonale de ${}^t\tilde{X}_i$ sur Δu_1



$$w_i = \langle {}^t\tilde{X}_i , u_1 \rangle_{I_p} = \tilde{X}_i I_p u_1$$

ε_i erreur dûe à la projection.

On utilise le critère des M.C qui consiste à minimiser la somme des erreurs quadratique moyennes

Problème d'optimisation :

$$\min_{\|u_1\|^2_{I_p}} \sum_{i=1}^{n} p_i \varepsilon_i^2$$

$$\varepsilon_i^2 = \|{}^t\tilde{X}_i\|^2_{I_p} - w_i^2$$

Le pb revient à $\max\limits_{\|u_1\|^2_{I_p}} \sum_{i=1}^{n} p_i w_i^2$

$$\iff \max_{\|u_1\|^2_{I_p}} \|W\|^2_{D_p} \quad \text{avec } W = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \text{ vecteur des n projections}$$

En effet $\|W\|^2_{D_p} = {}^tW D_p W = (w_1, \ldots, w_n) \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$

$$= (p_1 w_1, \ldots, p_n w_n) \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \sum_{i=1}^{n} p_i w_i^2$$

$$\text{Pb} \iff \max_{\|u_1\|^2_{I_p}} {}^tW D_p W \quad \text{avec } W = \begin{pmatrix} \tilde{X}_1 \\ \vdots \\ \tilde{X}_n \end{pmatrix} D_p u_1 = \tilde{X} D_p u_1$$

$$\iff \max_{\|u_1\|^2_{I_p}} {}^t(\tilde{X} I_p u_1) D_p (\tilde{X} I_p u_1)$$

$$\iff \max_{\|u_1\|^2_{I_p}} {}^tu_1 I_p {}^t\tilde{X} D_p \tilde{X} I_p u_1$$

En définitive, il s'agit de résoudre le problème à 1 contrainte suivant :

$$\begin{cases} \max\limits_{u_1 \in \mathbb{R}^p} {}^tu_1 I_p {}^t\tilde{X} D_p \tilde{X} I_p u_1 \\ {}^tu_1 . u_1 = \|u_1\|^2_{I_p} = 1 \end{cases} \iff \begin{cases} \max\limits_{u_1 \in \mathbb{R}^p} {}^tu_1 {}^t\tilde{X} D_p \tilde{X} u_1 \\ {}^tu_1 . u_1 = 1 \end{cases}$$

Le pb de multiplicateur de Lagrange donne comme solution un vecteur u_1 tq

$$\frac{\partial}{\partial u_1} \left[{}^tu_1 \, I_p \, {}^t\tilde{X} D_p \tilde{X} u_1 - \lambda({}^tu_1 . u - 1) \right] = 0$$

$\Leftrightarrow 2\, {}^t\tilde{X} D_p \tilde{X} u_1 - 2\lambda u_1 = 0$

$\Leftrightarrow \quad {}^t\tilde{X} D_p \tilde{X} u_1 = \lambda u_1$ $\quad u_1$ est alors $\vec{vp}$ de ${}^t\tilde{X} D_p \tilde{X}$ associé à λ

Étant donné que l'on cherche à maximiser ${}^tu_1 {}^t\tilde{X} D_p \tilde{X} u_1$

le vecteur u_1 cherché est le $\vec{vp}$ normé de ${}^t\tilde{X} D_p \tilde{X}$ associé à la plus grande vp λ

Rq. ${}^t\tilde{X} D_p \tilde{X}$ est la matrice de dimension (p,p) des variances-covariance.

2) La recherche du $1^{ère}, \dots, q^{ème}$ axes principaux

Nous allons généraliser (1)

Soient $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ les p valeurs propres (vp) de V

Soient $u_1, \dots, u_p$ les p vecteurs propres ($\vec{vp}$) de V (orthogonaux) associés aux vps $\lambda_1, \dots, \lambda_p$ resp.

1^{er} axe ~~factoriel~~ principal Δu_1

$2^{ème}$ axe principale est Δu_2

$\vdots$

$q^{ème}$ axe principal est Δu_q

Déf : On appelle pourcentage d'inertie expliquée par l'axe principale Δu_l la quantité

$$I_l = \frac{\lambda_l}{\lambda_1 + \dots + \lambda_p}$$

. Plan principal : Déf

Le plan principale est le plan le plus proche du nuage N(I)

$P = (\Delta u_1, \Delta u_2)$

Inertie expliquée par le plan $I_{1,2} = \dfrac{\lambda_1 + \lambda_2}{\lambda_1 + \dots + \lambda_p}$

. Sous espace principal de dim q.

Le s.e principal de dimension q, est le s.e.v de $\mathbb{R}^p$ le plus proche du nuage.

$$E_p = (\Delta u_1, \Delta u_2, \dots, \Delta u_q)$$

Le pourcentage d'inertie expliqué par E_p est

$$\frac{\lambda_1 + \dots + \lambda_q}{\lambda_1 + \dots + \lambda_p}$$

Composante principale.

Soient $u_1, \dots, u_q$ les q vp normés orthogonaux de V, associés au q premières vps $\lambda_1 \geq \dots \geq \lambda_q$ resp.

La composante $C^l = \tilde{X} u$ est appelée la $l^{ième}$ composante principal

$$C^l = W^{(l)} = \tilde{X} u_l.$$

Proposition

- Les composantes principales $C^1, \dots, C^q$ sont centrées
- $\mathrm{Var}(C^l) = \lambda_l$
- $\mathrm{cov}(C^l, C^{l'}) = 0$

Preuve.

- $\bar{C}^l = \sum_{i=1}^n p_i C_i^l = \sum_{i=1}^n p_i \tilde{X}_i u_l = \left(\sum_{i=1}^n p_i \tilde{X}_i\right) u_l = (\bar{\tilde{X}}^1, \dots, \bar{\tilde{X}}^p) u_l$
$= (0, \dots, 0)\, u_l = 0$

- $\mathrm{Var}(C^l) = \mathrm{Var}(W^{(l)}) = {}^t W^{(l)} D_p W^{(l)}$
$= {}^t(\tilde{X} u_l) D_p (\tilde{X} u_l) = {}^t u_l\, {}^t\tilde{X} D_p \tilde{X} u_l = {}^t u_l (V u_l)$
$= {}^t u_l (\lambda u_l) = \lambda_l \|u_l\|^2_{I_p} = \lambda_l$

- $\mathrm{cov}(C^l, C^{l'}) = \langle C^l, C^{l'} \rangle_{D_p} = {}^t C^l D_p C^{l'}$
$= {}^t W^{(l)} D_p\, {}^t W^{(l')}$
$= {}^t(\tilde{X} u_l) D_p (\tilde{X} u_{l'}) = {}^t u_l V u_{l'} = \langle u_l, u_{l'} \rangle_V = 0$

IV/ Cas où on se place sur $(\mathbb{R}^n, D_p)$

Soit le nuage de p points de $\mathbb{R}^n$ $N(J) = \{ \tilde{X}^j \,/\, \tilde{X}^j \in \mathbb{R}^n, j = 1, \dots, p \}$

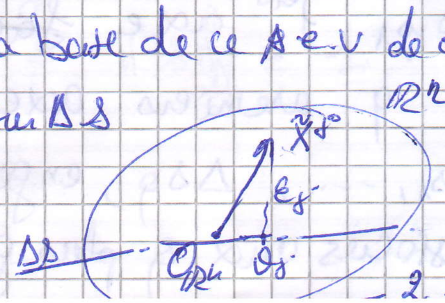
De la même manière que dans le cas où on se place sur $\mathbb{R}^p$

On cherche le s.e.v de $\mathbb{R}^n$ de dimension 1 qui ajuste au mieux les p points de $\mathbb{R}^n$

On noté par s le vecteur normé de la base de ce s.e.v de dim 1

O_j est la projection D_p-orthogonale de $\tilde{X}^j$ sur Δs

$$O_j = \langle \tilde{X}^j, s \rangle_{D_p} = {}^t\tilde{X}^j D_p s.$$



E_j erreur de projection

En utilise le critere M-C qui consiste à minimiser la somme des erreurs quadratique

$$pb \iff \min_{\|s\|^2_{D_p}=1} \sum_{j=1}^{p} (\varepsilon^j)^2$$

$$(\varepsilon^j)^2 = \|\tilde{x}^j\|^2_{D_p} - \vartheta^{j\,2} \quad (\text{pytagore})$$

$$pb \iff \max_{\|s\|^2_{D_p}=1} \sum_{j=1}^{p} (\vartheta^j)^2$$

$$\iff \max_{\|s\|^2_{D_p}=1} \sum_{j=1}^{p} {}^t\tilde{X}^j s \qquad \text{on pose } \vartheta = \begin{pmatrix} \vartheta^1 \\ \vdots \\ \vartheta^p \end{pmatrix} = \begin{pmatrix} {}^t\tilde{X}^1 D_p s \\ \vdots \\ {}^t\tilde{X}^p D_p s \end{pmatrix} = {}^t\tilde{X} D_p s$$

$$\iff \max_{\|s\|^2_{D_p}=1} {}^t\vartheta \cdot \vartheta \qquad {}^t\vartheta \cdot \vartheta = (\vartheta^1, \dots, \vartheta^p) \begin{pmatrix} \vartheta^1 \\ \vdots \\ \vartheta^p \end{pmatrix} = \sum_{j=1}^{p} (\vartheta^j)^2$$

$$\iff \max_{\|s\|^2_{D_p}} {}^t({}^t\tilde{X} D_p s)({}^t\tilde{X} D_p s) \iff \max_{\|s\|^2_{D_p}} {}^ts D_p \tilde{X}\,{}^t\tilde{X} D_p \cdot s$$

le problème d'optimisation à une contrainte revient à resoudre

$$\begin{cases} \max\limits_{s \in \mathbb{R}^n} {}^ts D_p \tilde{X}\,{}^t\tilde{X} \cdot D_p \cdot s \\ \|s\|^2_{D_p} = {}^ts D_p s = 1 \end{cases}$$

le Lagrangien

$$L(s) = {}^ts D_p \tilde{X}\,{}^t\tilde{X} D_p s - \lambda({}^ts D_p s - 1)$$

$$\frac{\partial L(s)}{\partial s} = 0 \iff 2 D_p \tilde{X}\,{}^t\tilde{X} D_p s - 2\lambda D_p s = 0$$

$$\iff \tilde{X}\,{}^t\tilde{X} D_p = \lambda s$$

s est un vecteur propre de la matrice $\tilde{X}\,{}^t\tilde{X} D_p$ associé a la vp λ

Comme le pb d'optimisation consiste a maximiser ${}^ts D_p \tilde{X}\,{}^t\tilde{X} D_p \cdot s$ ce qui revient à, le s.e.v de dim 1 le plus proche du $N(J)$ est le s.e.v engendre par le $\vec{vp}$ normé s_1 de $\tilde{X}\,{}^t\tilde{X} D$ associé à la plus grande vp λ_1.

Δs_1 1er axe ~~factori~~ principale

les q premiers axes principaux sont les s.e.v de dim 1 $\Delta s_1, \dots, \Delta s_q$, engendré par les $\vec{vp}$ normés $s_1, \dots, s_q$ de $\tilde{X}\,{}^t\tilde{X} D_p$ associés aux q plus grande vp $\lambda_1 \geq \dots \geq \lambda_q$