Devoir Stat

Exercice 1 : ACP sur la matrice des distances

1. Puisque x_i est centré pour $i = 1, \dots, n$, on a :

$$x_i = \tilde{x}_i - \bar{\tilde{x}},$$ où $\bar{\tilde{x}}$ désigne la moyenne empirique des $\tilde{x}_i$, et où $\tilde{x}_i$ est le vecteur initial avant qu'il ne soit centré.

Dès lors, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_i - \bar{\tilde{x}} = \bar{\tilde{x}} - \bar{\tilde{x}} = 0$

De ceci, on a :

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})' = \frac{1}{n} \sum_{i=1}^{n} (x_i - 0)(x_i - 0)'$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i x_i'$$

$$= \frac{1}{n} \mathbf{X}'\mathbf{X}$$

$$= X' \Delta X$$

où nous avons utilisé :

- $\mathbf{X}'\mathbf{X} = [x_1, \dots, x_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 x_1' + \dots + x_n x_n'$ où chaque x_i est un vecteur de p composantes pour $i = 1, \dots, n$.

- La multiplication par $\Delta = \begin{bmatrix} 1/n & & 0 \\ & \ddots & \\ 0 & & 1/n \end{bmatrix}$ d'une matrice de dimension adaptée équivaut à multiplier les composantes de la matrice par $1/n$.

2. Montrer que $\forall i = 1, \dots, n.\quad d_{i.}^2 = \|x_i\|_M^2 + I_g$.

Nous avons que :

- $I_g = \sum_{i=1}^{n} p_i \|x_i\|_M^2 = \frac{1}{n} \sum_{i=1}^{n} x_i' M x_i$

- $d_{i.}^2 = \sum_{j=1}^{n} p_j d_{ij}^2$

Développons :

$$d_{i.}^2 = \frac{1}{n} \sum_{j=1}^{n} \|x_i - x_j\|_M^2$$

$$= \frac{1}{n} \sum_{j=1}^{n} (x_i - x_j)' M (x_i - x_j)$$

$$= \frac{1}{n} \sum_{j=1}^{n} x_i' M x_i - x_i' M x_j - x_j' M x_i + x_j' M x_j$$

$$= x_i' M x_i + \underbrace{\sum_{j=1}^{n} \frac{1}{n} x_j' M x_j}_{= I_g} - x_i' M \sum_{j=1}^{n} \frac{1}{n} x_j - \left(\sum_{j=1}^{n} \frac{1}{n} x_j' \right) M x_i$$

$$= \|x_i\|_M^2 + I_g - \underbrace{x_i' M \bar{x}}_{=0} - \underbrace{\bar{x}' M x_i}_{=0}$$

ce qui livre le résultat.

3. En déduire que $d_{..}^2 = 2 I_g$.

On développe : $d_{..}^2 = \sum_{i=1}^{n} p_i d_{i.}^2 = \sum_{i=1}^{n} p_i \left[\|x_i\|_M^2 + I_g \right]$

$$= \sum_{i=1}^{n} p_i \|x_i\|_M^2 + \underbrace{\sum_{i=1}^{n} p_i I_g}_{= I_g}$$

$$= 2 I_g$$

où nous avons utilisé la linéarité de la somme.

4. Posons $W = \left(w_{ij} = \langle x_i, x_j \rangle_M = x_i' M x_j \right)_{ij}$

la matrice des produits scalaires.

Montrer que : $w_{ij} = -\frac{1}{2}\left(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2 \right)$.

Nous avons précédemment obtenu les résultats suivants :

$$\begin{cases} d_{..}^2 = 2 I_g \\ d_{i.}^2 = \|x_i\|_M^2 + I_g \ (i) \end{cases}$$

ainsi $d_{.j}^2$ se déduit de (i) par symétrie de la matrice de Variance-Covariance i.e. $d_{.j}^2 = \|x_j\|_M^2 + I_g$.

et donc par symétrie de $\mathcal{D}$ (D ronde).

On remplace et on développe :

$$-\frac{\|x_i - x_j\|_M^2 - \left(\|x_i\|_M^2 + \cancel{I_g} \right) - \left(\|x_j\|_M^2 + \cancel{I_g} \right) + \cancel{2 I_g}}{2}$$

$$= -\left[\frac{\|x_i - x_j\|_M^2 - \|x_i\|_M^2 - \|x_j\|_M^2}{2} \right]$$

$$= -\left[\frac{(x_i - x_j)' M (x_i - x_j) - x_i' M x_i - x_j' M x_j}{2} \right]$$

$$= -\left[\frac{\cancel{x_i' M x_i} - x_i' M x_j - x_j' M x_i + \cancel{x_j' M x_j} - \cancel{x_i' M x_i} - \cancel{x_j' M x_j}}{2} \right]$$

$$= \frac{x_i' M x_j + x_j' M x_i}{2} = \frac{2 \langle x_i, x_j \rangle_M}{2} = w_{ij}$$

où nous à nouveau utilisé l'argument de symétrie.

~~5.~~ Exprimer W en fonction de D. → Pour cette question Voir annexe. "Devoir_maison Q5"

Il faut remarquer les faits suivants :

$$\begin{cases} d_{i\cdot}^2 \text{ est constant sur la ligne } i \\ d_{\cdot j}^2 \text{ est constante sur la colonne } j \\ d_{\cdot\cdot}^2 \text{ est constant.} \end{cases}$$

Ainsi, en réutilisant l'égalité prouvée au point 4.

$$W = -\frac{1}{2}\left[\Delta - M - N + P\right]$$

où M est la matrice $M = \begin{bmatrix} d_{1\cdot}^2 & d_{1\cdot}^2 & \cdots & d_{1\cdot}^2 \\ d_{2\cdot}^2 & d_{2\cdot}^2 & \cdots & d_{2\cdot}^2 \\ \vdots & & & \vdots \\ d_{n\cdot}^2 & d_{n\cdot}^2 & \cdots & d_{n\cdot}^2 \end{bmatrix}$

$$N = \begin{bmatrix} d_{\cdot 1}^2 & d_{\cdot 2}^2 & \cdots & d_{\cdot n}^2 \\ d_{\cdot 1}^2 & d_{\cdot 2}^2 & & \vdots \\ \vdots & \vdots & & d_{\cdot n}^2 \\ d_{\cdot 1}^2 & d_{\cdot 2}^2 & & \end{bmatrix}$$

$$P = \begin{bmatrix} d_{\cdot\cdot}^2 & \cdots & d_{\cdot\cdot}^2 \\ \vdots & \ddots & \vdots \\ d_{\cdot\cdot}^2 & \cdots & d_{\cdot\cdot}^2 \end{bmatrix} = d_{\cdot\cdot}^2 \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

Nous avons reproduit la définition de w_{ij} et l'avons exprimée de manière matricielle, en répliquant les constants sur les lignes, ou sur les colonnes ou les deux à la fois.

Cependant, nous mobilisons trop d'expressions... Pour simplifier on retourne à la ligne 2 du développement calculatoire de la question 4 où l'on a :

~~$w_{ij} = -\frac{1}{2}\left[\|x_i - x_j\|^2 - \|x_i\|_M^2 - \|x_j\|_M^2\right]$, dès lors il suffit alors de poser~~

~~$Q = \begin{bmatrix} 2\|X_1\|_M^2 & \cdots & \|X_1\|_M^2 + \|X_n\|_M^2 \\ \vdots & & \vdots \\ \|X_n\|_M^2 + \|X_1\|_M^2 & \cdots & 2\|X_n\|_M^2 \end{bmatrix}$ i.e. la matrice où chaque élément est la somme des 2 normes indicées par la ligne et la colonne correspondante.~~

$$\boxed{W = -\frac{1}{2}\Delta + Q.}$$

6. Puisque les axes principaux normés ont valeurs propres associés λ_k nous obtenons que :

$X S u_k = X \lambda_k u_k$, puisque les vecteurs propres de la matrice de covariance S sont des directions spéciales (axes principaux) le long desquelles les données varient, et les valeurs propres sont les mesures de l'importance de ces directions.

Ensuite, nous utilisons la propriété fondamentale de l'ACP donnée par : $X u_k = v_k$ (projection).

Or $X S u_k = X \lambda_k u_k = \lambda_k X u_k = \lambda_k v_k$. $M = I_p$

① On remarque que les composants de v_k sont : $\langle x_i, u_k \rangle_M$

7. Montrer que, toujours si $M = I_p$, v_k est également vecteur propre de WD.

↳ tel est le cas ($M = I_p$) alors : $w_{ij} = \langle x_i, x_j \rangle = x_i' x_j$

ainsi chaque élément de WD est donné par : $\frac{1}{n} x_i' x_j$. et $W = XX'$

Or, $XS = XX'DX = WDX$. Donc on a :

$XS = WDX$

#on remarque que on remarque également que $WD = \frac{1}{n} XX' D$

Or, $X S u_k = \lambda_k v_k$, ce qui implique que :

$WD \underbrace{X u_k}_{= v_k} = \lambda_k v_k$, ce qui montre que v_k vecteur propre de WD.

8. Soit le vecteur $f_k \in \mathbb{R}^n$ dont la composante numéro i est $f_{ik} = \sqrt{p_i}\, v_{ik}$. En déduire que la matrice $W\Delta$ admet pour vecteur propre f_k avec valeur propre associée λ_k.

Ecrivons $f_k \in R^n$:

$$\begin{bmatrix} f_{1,k} \\ \vdots \\ f_{n,k} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{n}\; v_{1k} \\ \vdots \\ 1/\sqrt{n}\; v_{nk} \end{bmatrix} = \underbrace{\begin{bmatrix} 1/\sqrt{n} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{n} \end{bmatrix}}_{=\sqrt{P}} \begin{bmatrix} v_{1k} \\ \vdots \\ v_{nk} \end{bmatrix}$$

Or $W\Delta v_k = \lambda_k v_k$ et donc :

$$W\Delta \sqrt{P} v_k = \sqrt{P}\, W\Delta v_k = \sqrt{P}\, \lambda_k v_k$$
$$= \lambda_k \sqrt{P} v_k$$
$$= \lambda_k f_k.$$

ce qui livre le résultat.

9. Montrer que le vecteur $(\sqrt{p_i})_{i=1,\dots,n}$ est vecteur propre de $W\Delta$ associé à la valeur propre 0.

Il suffit de montrer que le système : $W\Delta \begin{bmatrix} 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$, admet une solution non triviale

Pour rappel $\bar{X} = 0$ car les x_i sont centrés.

$$W\Delta = \frac{1}{n} \begin{bmatrix} x_1' x_1 & \cdots & x_1' x_n \\ \vdots & & \\ x_n' x_1 & \cdots & x_n' x_n \end{bmatrix}.$$

On calcule :

$$W\Delta \begin{bmatrix} 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}\, n} (x_1' x_1 + \cdots + x_1' x_n) \\ \vdots \\ \frac{1}{\sqrt{n}\, n} (x_n' x_1 + \cdots + x_n' x_n) \end{bmatrix} = \begin{bmatrix} 1/\sqrt{n} \, (x_1' \underbrace{\tfrac{1}{n} \sum_{i=1}^n x_i}_{=0}) \\ \vdots \\ 1/\sqrt{n} \, (x_n' \underbrace{\tfrac{1}{n} \sum_{i=1}^n x_i}_{=0}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

10. Montrer que $\sum_{i=1}^{n} f_{ik}^2 = \lambda_k$ et pour tout $k \neq l$. $\sum_{i=1}^{n} f_{ik} f_{il} = 0$

Ana :

$$\sum_{i=1}^{n} \left(1/\sqrt{n}\, v_{ik} \right)^2 = \sum_{i=1}^{n} \frac{1}{n} v_{ik}^2$$

On :

$$\sum_{i=1}^{n} f_{ik}^2 = \frac{1}{n} \left[v_{1k}, \dots, v_{nk} \right] \begin{bmatrix} v_{1k} \\ \vdots \\ v_{nk} \end{bmatrix}$$

$= (X u_k)' \Delta X u$ par la question 6.

$= u_k' X' \Delta X u_k$

$= u_k' S u_k$

$= u_k' \lambda_k u_k$ à nouveau par 6.

$= \lambda_k u_k' u_k$

$= \lambda_k$ car u_k est de norme 1.

Ensuite, $\sum_{i=1}^{n} f_{ik} f_{il} = \frac{1}{n} \sum v_{ik} v_{il} = 0$

puisque $u_k \perp u_l$ pour $k \neq l$. En effet :

$$\frac{1}{n} \sum_{i=1}^{n} v_{ik} v_{il} = \frac{1}{n} \left[v_{1k} \dots v_{nk} \right] \begin{bmatrix} v_{1l} \\ \vdots \\ v_{nl} \end{bmatrix}$$

$= \frac{1}{n} u_k' X' X u_l$

$= u_k' S u_l = u_k' \lambda_k u_l = 0$

Exercice 3

1- Montrer que $\sqrt{n}(\hat{\Lambda} - I_p) = O_p(1)$ pour $n \to \infty$.

Tout d'abord on remarque que :

$$\sqrt{n}(\hat{S} - I_p) = \sqrt{n}(\hat{\beta}\hat{\Lambda}\hat{\beta}' - I_p) = \hat{\beta}\sqrt{n}(\hat{\Lambda} - I_n)\hat{\beta}'$$

La norme de $\sqrt{n}(\hat{S} - I_p)$ est donnée par :

$$\|\sqrt{n}(\hat{S} - I_p)\| = \|\sqrt{n}(\hat{\beta}\hat{\Lambda}\hat{\beta}' - I_p)\| = \|\hat{\beta}\sqrt{n}(\hat{\Lambda} - I_n)\hat{\beta}'\|$$

$$\leq \underbrace{\|\hat{\beta}\|}_{=1} \|\sqrt{n}(\hat{\Lambda} - I_n)\| \underbrace{\|\hat{\beta}'\|}_{=1}.$$

ainsi, $\|\sqrt{n}(\hat{S} - I_p)\| \leq \|\sqrt{n}(\hat{\Lambda} - I_n)\|$

Puis, on remarque que :

$$\hat{\beta}'\sqrt{n}(\hat{S} - I_p)\hat{\beta} = \sqrt{n}(\hat{\Lambda} - I_n)$$

et que :

$$\|\underbrace{\hat{\beta}'\sqrt{n}(\hat{S} - I_n)\hat{\beta}}_{=\sqrt{n}(\hat{\Lambda} - I_n)}\| \leq \underbrace{\|\hat{\beta}'\|}_{=1} \|\sqrt{n}(\hat{S} - I_n)\| \underbrace{\|\hat{\beta}'\|}_{=1}$$

De ceci, on en déduit que $\|\sqrt{n}(\hat{\Lambda} - I_n)\| = \|\sqrt{n}(\hat{S} - I_n)\|$

Or, puisque $\sqrt{n}(\hat{S} - I_n) = O_p(1)$, ceci implique que

$$\sqrt{n}(\hat{\Lambda} - I_n) = O_p(1).$$

2) Montrer que pour tout $\theta \in SO_p$
$$\theta \sqrt{n}(\hat{S} - I_p)\theta' \overset{\mathcal{L}}{=} \sqrt{n}(\hat{S} - I_p).$$

Soit $\theta \in SO_p$. On remarque que :
$$\theta \sqrt{n}(\hat{S} - I_p)\theta' = \sqrt{n}(\theta \hat{S}\theta' - I_p)$$
dès lors, il suffit de montrer que $\theta \hat{S}\theta' \overset{\mathcal{L}}{=} \hat{S}$.
Nous utilisons la caractérisation des lois de Wishart.
$$\hat{S} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$$
qui se simplifie comme suit :
$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right) - \bar{x}\bar{x}'.$$ On en déduit du lemme de Fisher
que $n\hat{S} \sim W(n-1, I_p)$.
Or, $\theta n \hat{S}\theta \sim W(n-1, \underbrace{\theta I_p \theta'}_{= I_p})$

ce qui livre que : $\theta n\hat{S}\theta' \sim W(n-1, I_p)$

ainsi, $n\hat{S} \overset{\mathcal{L}}{=} \theta n \hat{S}\theta'$ implique que $\hat{S} \overset{\mathcal{L}}{=} \theta\hat{S}\theta'$

et la conclusion suit. Nous aurions aussi pu développer le terme : $\theta\sqrt{n}(\hat{S} - I_p)\theta'$ et remarquer que les x_i sont sphériques et donc invariants par l'action de θ, ce qui aurait aussi permis de conclure.

3. Montrer que $\exists\ \tilde{O} \in SO_p$ telle que
$\tilde{O}\sqrt{n}(\hat{\Lambda} - I_p)\tilde{O}'$ et $\sqrt{n}(\hat{\Lambda} - I_p)$ ne converge pas vers la même loi.

On va utiliser l'astuce. On fixe $\tilde{O} \in SO_p$ telle que celle-ci permute deux éléments de la diagonale. Par exemple pour $n = 3$ et $\tilde{O} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

On a :

$$\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & & 0 \\ & 2 & \\ 0 & & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & & 0 \\ & 1 & \\ 0 & & 3 \end{bmatrix}.$$

Supposons sans perte de généralité que $\tilde{O}$ permute les deux dernières valeurs sur la diagonale, i.e $\tilde{O} = \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 0 & 1 \\ 0 & & 1 & 0 \end{bmatrix}$

Il suffit dès lors de considérer l'évènement $E =$ " l'avant dernière valeur sur la diagonale est plus petite que la dernière."

ainsi, on a : $\mathbb{P}[\hat{\Lambda} \in E] = 1.$ et $\mathbb{P}[\tilde{O}\hat{\Lambda}\tilde{O}' \in E] = 0.$

dès lors,
$$\lim_{n\to\infty} \mathbb{P}\left[\sqrt{n}(\hat{\Lambda} - I_p) \in E\right] = 1$$

$$\text{et } \lim_{n\to\infty} \mathbb{P}\left[\sqrt{n}(\tilde{O}\hat{\Lambda}\tilde{O}' - I_p) \in E\right] = 0.$$

Et puisque l'évènement E est un Borélien sur lequel les distributions sont continues ceci suffit pour achever la preuve.

4) Montrer que les 3 questions précédentes impliquent qu'il n'existe pas de matrice β telle que $\hat{\beta} \xrightarrow{P} \beta$.

Supposons qu'une telle matrice existe.

On commence par remarquer que :

$$\theta \sqrt{n}(\hat{S} - I_p)\theta' \overset{\mathcal{D}}{=} \sqrt{n}(\hat{S} - I_p) \quad \underline{\text{exo 2.}}$$

$$\Leftrightarrow \theta \hat{\beta} \sqrt{n}(\hat{\Lambda} - I_p)(\theta\hat{\beta})' \overset{\mathcal{D}}{=} \hat{\beta}\sqrt{n}(\hat{\Lambda} - I_p)\hat{\beta}'.$$

$$\Leftrightarrow \hat{\beta}'\theta\hat{\beta}\sqrt{n}(\hat{\Lambda} - I_p)(\hat{\beta}'\theta\hat{\beta})' \overset{\mathcal{D}}{=} \sqrt{n}(\hat{\Lambda} - I_p)$$

Si on pose $\tilde{\theta} = \beta'\theta\beta$, la matrice de l'exercice 3, par choix malin de θ, on trouve par un résultat sur les convergences en probabilité que :

$$\hat{\beta}'\theta\hat{\beta} \xrightarrow{P} \beta'\theta\beta.$$

Notons maintenant X la distribution limite de $\sqrt{n}(\hat{\Lambda} - I_p)$.

Par Slutsky :

$$\begin{pmatrix} \sqrt{n}(\hat{\Lambda} - I_p) \\ \hat{\beta}'\theta\hat{\beta} \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ \beta'\theta\beta \end{pmatrix}$$

ainsi, $\hat{\beta}'\theta\hat{\beta}\sqrt{n}(\hat{\Lambda} - I_p)(\hat{\beta}'\theta\hat{\beta})' \xrightarrow{\mathcal{L}} \underbrace{\beta'\theta\beta X (\beta'\theta\beta)'}_{= \tilde{\theta} \times \tilde{\theta}'}$

On voit contredit l'exo 3 : puisque $\tilde{\theta} \times \tilde{\theta} \overset{\mathcal{D}}{\neq} X$ $\sqrt{n}(\hat{\Lambda} - I_p)$.

car ~~et~~ $\hat{\beta}'\theta\hat{\beta}\sqrt{n}(\hat{\Lambda} - I_p)(\hat{\beta}'\theta\hat{\beta})' \overset{\mathcal{D}}{=}$

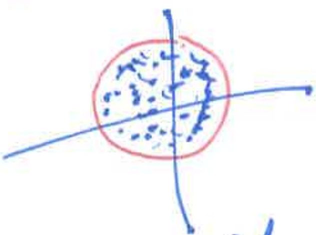
5. En considérant le résultat obtenu au point 4 est-il pertinent de donner une interprétation du type donnée à l'exo 2.3 aux différents vecteurs propres dans $\hat{\beta}$?

On peut remarquer que puisque il n'existe pas de β tel que $\hat{\beta} \xrightarrow{P} \beta$, on en déduit que $\hat{\beta}$ n'est pas un estimateur (faiblement) convergent.

Or, par un résultat du cours de STAT 1., un estimateur non convergent est à écarter.

En effet, il devient difficile de tirer des informations supplémentaires quant à la contribution des vecteurs propres à la variance

De plus, le fait que $\hat{\beta} \hat{\Lambda} \hat{\beta}' \xrightarrow{P} I_p$ et que nos $X_i \overset{iid}{\sim} \mathcal{N}(0, I_p)$ nos composantes principales auraient pour distribution $\sim \mathcal{N}_p(0, I_p)$ et donc nos courbes de niveaux forment une sphère



Ainsi, retenir un axe ne serait pas d'une grande utilité.