Analyse en Composantes Principales (ACP)

L3$\{MA, Info_SI, Info_Reseaux\}$ - UFR S.A.T

Pr. Ousmane THIARE

othiare@ugb.edu.sn
[www.ousmanethiare.com]

16 avril 2020

Analyse en Composantes Principales (ACP)

- L'étude d'une population statistique de taille n passe le plus souvent par le recueil d'un nombre élevé p de données quantitatives par élément observé. L'analyse de ces données doit tenir compte de leur caractère multidimensionnel et révéler les liaisons existantes entre les comoposantes.

- L'étude d'une population statistique de taille n passe le plus souvent par le recueil d'un nombre élevé p de données quantitatives par élément observé. L'analyse de ces données doit tenir compte de leur caractère multidimensionnel et révéler les liaisons existantes entre les comoposantes.
- Méthode très puissante pour explorer la structure de telles données, introduite en 1901 par K. Pearson et développée par H. Hotelling en 1933

- L'étude d'une population statistique de taille n passe le plus souvent par le recueil d'un nombre élevé p de données quantitatives par élément observé. L'analyse de ces données doit tenir compte de leur caractère multidimensionnel et révéler les liaisons existantes entre les comoposantes.
- Méthode très puissante pour explorer la structure de telles données, introduite en 1901 par K. Pearson et développée par H. Hotelling en 1933
- Chaque donnée étant représentée dans un espase à p dimensions, l'ensemble des donées forment un "nuage de n points" dans $\mathbb{R}^p$.

- Principe $\Longrightarrow$ obtenir un représentation approchée du nuage dans un sous-espace de dimension faible k par projections sur des axes bien choisis.

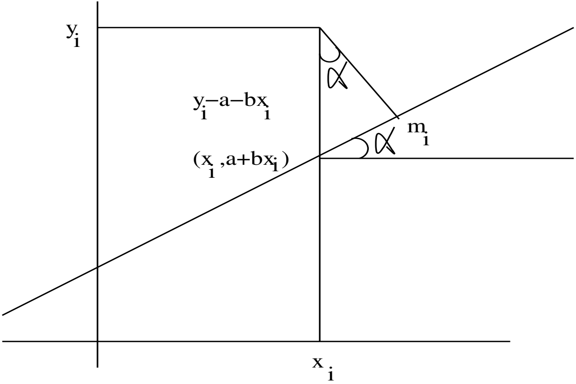
- Principe $\Longrightarrow$ obtenir un représentation approchée du nuage dans un sous-espace de dimension faible k par projections sur des axes bien choisis.
- Les composantes principales sont les n vecteurs ayant pour coordonnées celles des projections orthogonales des n éléments du nuage sur les k axes principaux.

- Principe $\Longrightarrow$ obtenir un représentation approchée du nuage dans un sous-espace de dimension faible k par projections sur des axes bien choisis.
- Les composantes principales sont les n vecteurs ayant pour coordonnées celles des projections orthogonales des n éléments du nuage sur les k axes principaux.
- L'ACP construit ainsi de nouvelles variables, artificielles, et des représenatations graphiques permettant de visualiser les relations entre variables, ainsi que l'existence éventuelle de groupes d'éléments et de groupes de variables.

Soit $\mathbb{R}^2$ l'espace des individus, muni du produit scalaire canonique et de la base canonique $\{e_1, e_2\}$ qui, on l'a vu, est orthonormée pour ce produit scalaire.
Si aucune des variables statistiques, X ou Y ne peut s'interpréter par rapport à l'autre, il n'y a pas de raison de privilégier la régression linéaire de Y par rapport à X ou la régression linéaire de X par rapport à Y.
Nous sommes alors conduits à un autre point de vue, celui de la **réduction des données**.

Nous cherchons alors dans $\mathbb{R}^2$ une droite (D) qui minimise la somme S^2 des carrés des distances des points du nuage de points à la droite.
La solution est donnée par la **droite de régression orthogonale**.

L'équation de la droite de régression orthogonale est de la forme : $y = a + bx$.



b est la tangente de l'angle de la droite avec l'axe des abscisses : $b = \tan\alpha$.

$$\|\overrightarrow{M_i m_i}\|^2 = \cos^2\alpha (y_i - a - bx_i)^2 = \frac{1}{1+b^2}(y_i - a - bx_i)^2$$

En introduisant le point moyen $(\overline{X}, \overline{Y})$, on peut écrire :

$$\frac{1}{n}\sum_{i=1}^{n}\|M_i m_i\|^2 =$$

$$\frac{1}{1+b^2} \times \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y} - b(x_i - \overline{X}) + (\overline{Y} - a - b\overline{X}))^2$$

$$= \frac{1}{1+b^2} \times \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y} - b(x_i - \overline{X}))^2 + \frac{1}{1+b^2}(\overline{Y} - a - b\overline{X})^2$$

$$+2\frac{1}{1+b^2} \times (\overline{Y} - a - b\overline{X})\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y} - b(x_i - \overline{X}))$$

Régression orthogonale : Axe principal

Calcul du terme constant a (suite)

Les relations $\overline{Y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$ et $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ entraînent que le dernier terme de la somme est nul. Il reste :

$$\frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{M_i m_i}\|^2 =$$

$$\frac{1}{1+b^2} \times \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{Y} - b(x_i - \overline{X}))^2 + \frac{1}{1+b^2}(\overline{Y} - a - b\overline{X})^2$$

Quelle que soit la valeur de b, cette somme sera la plus petite possible lorsque le deuxième terme est nul : $\overline{Y} = a + b\overline{X}$.

Ce résultat signifie que le **point moyen est sur la droite de régression orthogonale** et que, lorsque b est connu, le terme constant a est donné par :

$$a = \overline{Y} - b\overline{X}$$

Puisque le point moyen $G = (\overline{X}, \overline{Y})$ est sur la droite de régression orthogonale, nous le prendrons comme **origine** dans $\mathbb{R}^2$.

La droite de régression orthogonale a une équation de la forme :

$$y_0 = bx_0$$

avec $y_0 = y - \overline{Y}$ et $x_0 = x - \overline{X}$

Régression orthogonale : Axe principal

Analyse en composantes principales (ACP)

En fait, la forme de la relation précédente fait disparaître la symétrie initiale entre les rÙles de X et Y : ce n'est pas sous cette forme que nous exprimerons l'équation de la droite (D) de régression orthogonale.
Etant donnée une droite (D) passant par l'origine G, on considère plutÙt le vecteur unitaire de $\mathbb{R}^2$ orthogonal à la droite (D) :

$$u_1 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \text{ avec } \alpha^2 + \beta^2 = 1$$

Le vecteur unitaire u porté par la droite (D) est :

$$u = \begin{pmatrix} \beta \\ -\alpha \end{pmatrix}$$

La droite (D) est l'ensemble des points $M = (x, y)$ vérifiant $\langle u_1 | \overrightarrow{GM} \rangle = 0$,
soit $\alpha x_0 + \beta y_0 = 0$.
Etant donné un point M_i du nuage de point et sa projection orthogonale m_i sur la droite (D), le vecteur $\overrightarrow{Gm_i}$ est le projeté orthogonal de $\overrightarrow{GM_i}$ sur le vecteur u :

$$\overrightarrow{Gm_i} = \langle \overrightarrow{GM_i} | u \rangle u = (\beta x_{i0} - \alpha y_{i0}) \begin{pmatrix} \beta \\ -\alpha \end{pmatrix}$$

$$\overrightarrow{m_i M_i} = \overrightarrow{GM_i} - \overrightarrow{Gm_i} = \begin{pmatrix} x_{i0} \\ y_{i0} \end{pmatrix} - (\beta x_{i0} - \alpha y_{i0}) \begin{pmatrix} \beta \\ -\alpha \end{pmatrix} =$$

$$\begin{pmatrix} (1-\beta^2) x_{i0} + \alpha\beta y_{i0} \\ \alpha\beta x_{i0} + (1-\alpha^2) y_{i0} \end{pmatrix} = \begin{pmatrix} \alpha^2 x_{i0} + \alpha\beta y_{i0} \\ \alpha\beta x_{i0} + \beta^2 y_{i0} \end{pmatrix} =$$

$$(\alpha x_{i0} + \beta y_{i0}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$\|\overrightarrow{m_iM_i}\|^2 = (\alpha x_{i0} + \beta y_{i0})^2 (\alpha \quad \beta) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} =$$

$$(\alpha x_{i0} + \beta y_{i0})^2(\alpha^2 + \beta^2) = (\alpha x_{i0} + \beta y_{i0})^2$$

$$\frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{m_iM_i}\|^2 = \frac{1}{n}\sum_{i=1}^{n}(\alpha x_{i0} + \beta y_{i0})^2 =$$

$$\langle \alpha X_0 + \beta Y_0 | D_{\frac{1}{n}} | \alpha X_0 + \beta Y_0 \rangle = \|\alpha X_0 + \beta Y_0\|^2_{D_{\frac{1}{n}}}$$

La recherche de la droite de régression orthogonale se ramène donc à une question que l'on peut envisager d'un double point de vue :

- soit rechercher, dans l'espace des individus $\mathbb{R}^2$, un vecteur unitaire $u_1 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ avec $\alpha^2 + \beta^2 = 1$ qui minimise la somme :

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{m_i M_i}\|^2 = \frac{1}{n}\sum_{i=1}^{n} (\alpha x_{i0} + \beta y_{i0})^2,$$

- soit rechercher, dans l'espace des variables $\mathbb{R}^n$, un vecteur $\alpha X_0 + \beta Y_0$, combinaison linéaire fictive des deux variables centrées X_0 et Y_0, avec $\alpha^2 + \beta^2 = 1$, qui minimise $\|\alpha X_0 + \beta Y_0\|^2_{D_{\frac{1}{n}}}$, c'est-à-dire un vecteur de l'hyperplan défini par X_0 et Y_0, de norme minimum pour le produit scalaire défini par la matrice diagonale $D_{\frac{1}{n}}$, sous la contrainte $\alpha^2 + \beta^2 = 1$.

Sous la deuxième forme, la résolution du problème est appelée l'**analyse en composantes principales**.

Appelons Z la matrice $\begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix}$ des variables centrées

Inertie totale

On appelle inertie totale du nuage de points de $\mathbb{R}^2$ par rapport à l'origine G des axes, la quantité :

$$I_T = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{GM_i}\|^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{i0}^2 + y_{i0}^2) = s^2(X) + s^2(Y).$$

On appelle inertie statistique du nuage de points de $\mathbb{R}^2$ par rapport à une direction Δ de $\mathbb{R}^2$ définie par un vecteur unitaire u, la quantité :

$$I_S(u) = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{Gm_i}\|^2$$

où $\overrightarrow{Gm_i}$ est le projeté orthogonal de $\overrightarrow{GM_i}$ sur u.

Le rapport $\dfrac{I_S(u)}{I_T}$ est le **taux d'inertie totale expliquée par la direction u**.

Par exemple, l'inertie statistique du nuage de points par rapport à l'axe des x est la variance de X et l'inertie statistique du nuage de points par rapport à l'axe des y est la variance de Y.

On appelle inertie mécanique du nuage de points de $\mathbb{R}^2$ par rapport à une direction Δ définie par un vecteur unitaire u, la quantité :

$$I_M(u) = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{M_i m_i}\|^2$$

où $\overrightarrow{Gm_i}$ est le projeté orthogonal de $\overrightarrow{GM_i}$ sur u.
Par exemple, l'inertie mécanique du nuage de points par rapport à l'axe des x est la variance de Y et l'inertie mécanique du nuage de points par rapport à l'axe des y est la variance de X.
Le théorème de Pythagore $\|\overrightarrow{GM_i}\|^2 = \|\overrightarrow{Gm_i}\|^2 + \|\overrightarrow{M_i m_i}\|^2$ entraîne :

$$I_M(u) = I_T - I_S(u)$$

On appelle **premer axe factoriel** du nuage de points de $\mathbb{R}^2$, l'axe dont la direction définie par un vecteur unitaire u maximise l'inertie statistique $I_S(u)$.
La direction définie par le vecteur u est appelée la **direction principale** ou **direction factorielle**.
On remarquera que, comme le premier axe factoriel maximise $I_S(u)$, il minimise $I_M(u)$: il donne donc la solution de notre problème, c'est-à-dire la droite de régression orthogonale.

Pour u=$\begin{pmatrix} \beta \\ -\alpha \end{pmatrix}$, l'inertie statistique $I_S(u) = \frac{1}{n}\sum_{i=1}^{n} \|\overrightarrow{Gm_i}\|^2$ s'écrit, avec $\overrightarrow{Gm_i} = \langle \overrightarrow{GM_i} | u \rangle u = (\beta x_{i0} - \alpha y_{i0}) \begin{pmatrix} \beta \\ -\alpha \end{pmatrix}$, sous la forme :

$$I_S(u) = \frac{1}{n}\sum_{i=1}^{n} (\beta x_{i0} - \alpha y_{i0})^2 =$$

$$\beta^2 \times \frac{1}{n}\sum_{i=1}^{n} x_{i0}^2 + \alpha^2 \times \frac{1}{n}\sum_{i=1}^{n} y_{i0}^2 - 2\alpha\beta \times \frac{1}{n}\sum_{i=1}^{n} x_{i0} y_{i0}$$

Et comme on sait que :

$$\frac{1}{n}\sum_{i=1}^{n} x_{i0}^2 = s^2(X), \frac{1}{n}\sum_{i=1}^{n} y_{i0}^2 = s^2(Y), \frac{1}{n}\sum_{i=1}^{n} x_{i0} y_{i0} = Cov(X, Y)$$

L'inertie statistique devient :

$I_S(u) = \beta^2 s^2(X) + \alpha^2 s^2(Y) - 2\alpha\beta Cov(X,Y) = \begin{pmatrix} \beta & -\alpha \end{pmatrix} \begin{pmatrix} s^2(X) & Cov(X,Y) \\ Cov(X,Y) & s^2(Y) \end{pmatrix} \begin{pmatrix} \beta \\ -\alpha \end{pmatrix} = u^t A u$

La matrice A= $\begin{pmatrix} s^2(X) & Cov(X,Y) \\ Cov(X,Y) & s^2(Y) \end{pmatrix} = \frac{1}{n} \begin{pmatrix} x_{10} & \cdots & x_{n0} \\ y_{10} & \cdots & y_{n0} \end{pmatrix} \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix}$ s'appelle la **matrice des variances covariances.**

En introduisant la matrice

$Z=\begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix}$ des variables centrées, la matrice des variances covariances s'écrit sous les formes :

$A = \begin{pmatrix} s^2(X) & Cov(X,Y) \\ Cov(X,Y) & s^2(Y) \end{pmatrix} =$

$\frac{1}{n}\begin{pmatrix} x_{10} & \cdots & x_{n0} \\ y_{10} & \cdots & y_{n0} \end{pmatrix} \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} =$

$\frac{1}{n}Z^tZ = Z^tD_{\frac{1}{n}}Z$ et l'inertie totale est la trace de cette matrice, somme des éléments diagonaux $s^2(X)$ et $s^2(Y)$:

$$I_T = Tr(A)$$

1re remarque : **valeurs propres**

La matrice des variances-covariances A est, comme on le voit, symétrique réelle.

Une valeur propre de A est un nombre réel λ tel qu'il existe un vecteur v non nul vérifiant $Av = \lambda v$.

Les valeurs propres de A sont donc les nombres réels λ tels que le noyau de l'endomorphisme (application linéaire de $\mathbb{R}^2$ dans $\mathbb{R}^2$) défini par la matrice $A - \lambda I_2$ ne soit pas réduit à 0.

Dire que le noyau n'est pas réduit à 0, c'est-à-dire que l'application lin'éaire n'est pas injective, donc qu'elle n'est pas bijective (puisque dans $\mathbb{R}^2$, injective=bijective) : pour cela, il faut et il suffit que son déterminant soit nul.

Les valeurs propres sont donc les solutions de l'équation :

$$Det(A - \lambda I_2) = \lambda^2 - (s^2(X) + s^2(Y))\lambda + s^2(X)s^2(Y) - (Cov(X, Y))^2 = 0$$

Le discriminant de cette équation du second degré est :

$$(s^2(X) + s^2(Y))^2 - 4(s^2(X)s^2(Y) - (Cov(X, Y))^2) = (s^2(X) - s^2(Y))^2 + 4(Cov(X, Y))^2 \geq 0$$

La matrice A possède donc, ainsi qu'on l'avait déjà dit pour toute matrice symétrique réelle, deux valeurs propres réelles λ_1 et λ_2 :

- la somme de ces valeurs propres est la **trace** de la matrice, somme des éléments diagonaux de la première diagonale :
  $\lambda_1 + \lambda_2 = s^2(X) + s^2(Y) \geq 0$

La matrice A possède donc, ainsi qu'on l'avait déjà dit pour toute matrice symétrique réelle, deux valeurs propres réelles λ_1 et λ_2 :

- la somme de ces valeurs propres est la **trace** de la matrice, somme des éléments diagonaux de la première diagonale :
  $\lambda_1 + \lambda_2 = s^2(X) + s^2(Y) \geq 0$
- le produit de ces valeurs propres est le **déterminant** de la matrice :
  $\lambda_1 \lambda_2 = s^2(X) s^2(Y) - (Cov(X, Y))^2 \geq 0$ (d'après l'inégalité de Schwarz)

Les deux valeurs propres de la matrice des variances-covariances sont donc des nombres réels positifs : il est très improbable que l'une soit nulle (il faudrait, pour cela, que le coefficient de corrélation linéaire soit rigoureusement égal à 1, en valeur absolue, ce qui ne saurait se produire que si X et Y sont déduits l'un de l'autre par une relation linéaire, ou si X et Y sont constantes. Il est très improbabable aussi aussi que les deux valeurs propres soient égales : il faudrait pour cela que la covariance de X et Y soit strictement égale à 0 et que les variances de X et Y soient strictement égales, ce qui ne se produit jamais en pratique.)

Dans le cas général, on peut donc appeler λ_1 et λ_2 les **les valeurs propres de la matrice des variances-covariances**, rangés par ordre décroissant.

- $\lambda_1 > \lambda_2 > 0$

Dans le cas général, on peut donc appeler λ_1 et λ_2 les **les valeurs propres de la matrice des variances-covariances**, rangés par ordre décroissant.

- $\lambda_1 > \lambda_2 > 0$
- $\lambda_1 = \frac{1}{2}\left(s^2(X) + s^2(Y) + \sqrt{(s^2(X) - s^2(Y))^2 + 4(Cov(X,Y))^2}\right)$

Régression orthogonale : Axe principal

Matrices des variances-covariances (suite)

Dans le cas général, on peut donc appeler λ_1 et λ_2 les **les valeurs propres de la matrice des variances-covariances**, rangés par ordre décroissant.

- $\lambda_1 > \lambda_2 > 0$
- $\lambda_1 = \frac{1}{2}\left(s^2(X) + s^2(Y) + \sqrt{(s^2(X) - s^2(Y))^2 + 4(Cov(X, Y))^2}\right)$
- $\lambda_2 = \frac{1}{2}\left(s^2(X) + s^2(Y) - \sqrt{(s^2(X) - s^2(Y))^2 + 4(Cov(X, Y))^2}\right)$

2^me^ remarque : **vecteurs propres**

On démontre aussi, en algèbre, que $\mathbb{R}^2$ possède une **base propre orthonormée**, c'est-à-dire une base $\{u_1, u_2\}$, orthonormée pour le produit scalaire canonique, formée de vecteurs propres de la matrice A :

$$Au_1 = \lambda_1 u_1 \text{ et } Au_2 = \lambda_2 u_2,$$

avec

$$\|u_1\|^2 = 1, \|u_2\|^2 = 1, \langle u_1 | u_2 \rangle = 0$$

Ces vecteurs propres peuvent se calculer.
Soit λ une valeur propre. On a :

$$\begin{pmatrix} s^2(X)-\lambda & Cov(X,Y) \\ Cov(X,Y) & s^2(Y)-\lambda \end{pmatrix}\begin{pmatrix} s^2(Y)-\lambda \\ -Cov(X,Y) \end{pmatrix} =$$

$$\begin{pmatrix} (s^2(X)-\lambda)(s^2(Y)-\lambda)-(Cov(X,Y))^2 \\ (s^2(Y)-\lambda)Cov(X,Y)-(s^2(Y)-\lambda)Cov(X,Y) \end{pmatrix} =$$

$\begin{pmatrix} Det(A-\lambda I_2) \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$ donc le vecteur $\begin{pmatrix} s^2(Y)-\lambda \\ -Cov(X,Y) \end{pmatrix}$ est un **vecteur propre pour la valeur propre λ.**

Le carré de la norme de ce vecteur pour le produit scalaire est donné par :
$(s^2(Y) - \lambda - Cov(X,Y)) \begin{pmatrix} s^2(Y) - \lambda \\ -Cov(X,Y) \end{pmatrix} =$
$(s^2(Y) - \lambda)^2 + (Cov(X,Y))^2$ On peut donc prendre pour vecteur propre relatif à la valeur propre λ, le vecteur :
$u = \frac{1}{\sqrt{(s^2(Y) - \lambda)^2 + (Cov(X,Y))^2}} \begin{pmatrix} s^2(Y) - \lambda \\ -Cov(X,Y) \end{pmatrix}$ Le produit scalaire des deux vecteurs propres ainsi obtenu est nul, parce que la relation $\lambda_1 + \lambda_2 = s^2(X) + s^2(Y)$ entraîne :

$$\left(s^2(Y)-\lambda_1 \quad -Cov(X,Y)\right)\begin{pmatrix} s^2(Y)-\lambda_2 \\ -Cov(X,Y)\end{pmatrix} =$$

$$\left(\lambda_2 - s^2(X) \quad -Cov(X,Y)\right)\begin{pmatrix} s^2(Y)-\lambda_2 \\ -Cov(X,Y)\end{pmatrix} =$$

$Det(A-\lambda_2 I_2)=0$

Les deux vecteurs $\begin{pmatrix} s^2(Y)-\lambda_1 \\ -Cov(X,Y)\end{pmatrix}$ et $\begin{pmatrix} s^2(Y)-\lambda_2 \\ -Cov(X,Y)\end{pmatrix}$ forment une base de $\mathbb{R}^2$ parce que le déterminant de leurs coordonnées n'est pas nul :
$Cov(X,Y)\times(s^2(Y)-\lambda_1)+Cov(X,Y)\times(s^2(Y)-\lambda_2)=$
$Cov(X,Y)\times(\lambda_1-\lambda_2)\neq 0$ de sorte que les deux vecteurs ne sont pas proportionnels.

Les deux vecteurs :

- $$u_1 = \frac{1}{\sqrt{(s^2(Y)-\lambda_1)^2 + (Cov(X,Y))^2}} \begin{pmatrix} s^2(Y)-\lambda_1 \\ -Cov(X,Y) \end{pmatrix}$$

forment donc une **base orthonormée** de $\mathbb{R}^2$.
Remarquons que, au lieu de prendre pour vecteur propre pour le valeur propre λ, le vecteur $\begin{pmatrix} s^2(Y)-\lambda \\ -Cov(X,Y) \end{pmatrix}$, on aurait pu prendre aussi le vecteur $\begin{pmatrix} -Cov(X,Y) \\ s^2(X)-\lambda \end{pmatrix}$ qui lui est proportionnel (le déterminant de la matrice de ces vecteurs est le déterminant de la matrice $A-\lambda I_2$).

Régression orthogonale : Axe principal
Matrices des variances-covariances (suite et fin)

Les deux vecteurs :

- $u_1 = \dfrac{1}{\sqrt{(s^2(Y) - \lambda_1)^2 + (Cov(X,Y))^2}} \begin{pmatrix} s^2(Y) - \lambda_1 \\ -Cov(X,Y) \end{pmatrix}$
- $u_2 = \dfrac{1}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X,Y))^2}} \begin{pmatrix} s^2(Y) - \lambda_2 \\ -Cov(X,Y) \end{pmatrix}$

forment donc une **base orthonormée** de $\mathbb{R}^2$.
Remarquons que, au lieu de prendre pour vecteur propre pour le valeur propre λ, le vecteur $\begin{pmatrix} s^2(Y) - \lambda \\ -Cov(X,Y) \end{pmatrix}$, on aurait pu prendre aussi le vecteur $\begin{pmatrix} -Cov(X,Y) \\ s^2(X) - \lambda \end{pmatrix}$ qui lui est proportionnel (le déterminant de la matrice de ces vecteurs est le déterminant de la matrice $A - \lambda I_2$).

Diagonalisation de la matrice des variances-covariances

Soit V=

$$\begin{pmatrix} \dfrac{s^2(Y)-\lambda_1}{\sqrt{(s^2(Y)-\lambda_1)^2+(Cov(X,Y))^2}} & \dfrac{s^2(Y)-\lambda_2}{\sqrt{(s^2(Y)-\lambda_2)^2+(Cov(} \\ \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y)-\lambda_1)^2+(Cov(X,Y))^2}} & \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y)-\lambda_2)^2+(Cov(} \end{pmatrix}$$

la matrice des coordonnées des vecteurs propres u_1 et u_2.

$$Ve_1 = u_1, Ve_2 = u_2.$$

la matrice des coordonnées des vecteurs propres u_1 et u_2.

$$Ve_1 = u_1, Ve_2 = u_2.$$

Diagonalisation de la matrice des variances-covariances (suite)

V donne, par produits, pour image d'une base orthonormée, une base orthonormée : c'est ce qu'on appelle une matrice "orthogonale", ce qui veut dire que son inverse est égale à sa transposée :

$$V^{-1} = V^t$$

Pour le vérifier, remarquons que ouisque les bases $\{e_1, e_2\}$ et $\{u_1, u_2\}$ sont orthonormées, les coordonnées des vecteurs s'obtiennent par produits scalaires : :

$$u_1 = \langle u_1 | e_1 \rangle e_1 + \langle u_1 | e_2 \rangle e_2$$

$$u_2 = \langle u_2 | e_1 \rangle e_1 + \langle u_2 | e_2 \rangle e_2$$

Diagonalisation de la matrice des variances-covariances (suite)

de sorte que la matrice V, qui a, pour colonnes les vecteurs u_1 et u_2 dans la base $\{e_1, e_2\}$, est :

$$V = \begin{pmatrix} \langle u_1 | e_1 \rangle & \langle u_2 | e_1 \rangle \\ \langle u_1 | e_2 \rangle & \langle u_2 | e_2 \rangle \end{pmatrix}$$

et les relations inverses

$$e_1 = \langle e_1 | u_1 \rangle u_1 + \langle e_1 | u_2 \rangle u_2$$

$$e_2 = \langle e_2 | u_1 \rangle u_1 + \langle e_2 | u_2 \rangle u_2$$

montrent que la matrice inverse de V est la matrice :

$$V^{-1} = \begin{pmatrix} \langle e_1 | u_1 \rangle & \langle e_2 | u_1 \rangle \\ \langle e_1 | u_2 \rangle & \langle e_2 | u_2 \rangle \end{pmatrix}$$

Diagonalisation de la matrice des variances-covariances (suite)

qui, compte tenu de la symétrie du produit scalaire, est la transposée de V.

$$V^{-1} = \begin{pmatrix} \langle u_1|e_1\rangle & \langle u_1|e_2\rangle \\ \langle u_2|e_1\rangle & \langle u_2|e_2\rangle \end{pmatrix} = V^t$$

Il résulte alors des relations $Ve_1 = u_1$ et $Ve_2 = u_2$, que l'on a :

$$V^t u_1 = V^{-1} u_1 = e_1; \quad V^t u_2 = V^{-1} u_2 = e_2$$

Considérons maintenant la matrice $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, matrice diagonale des valeurs propres de A.

A est la matrice, dans la base canonique $\{e_1, e_2\}$, d'un endomorphisme f.

Cet endomorphisme f se réduit à deux homothéties, de rapport λ_1 selon le vecteur u_1, et de rapport λ_2 selon le vecteur u_2.

Λ est donc la matrice, dans la base propre $\{u_1, u_2\}$, de

Diagonalisation de la matrice des variances-covariances (suite)

La matrice de l'application identique de $\mathbb{R}^2$ muni de la base $\{u_1, u_2\}$ dans $\mathbb{R}^2$ muni de la base $\{e_1, e_2\}$ donne, par produits, pour image du vecteur $u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ le vecteur

$$u_1 = \frac{1}{\sqrt{(s^2(Y) - \lambda_1)^2 + (Cov(X,Y))^2}} \begin{pmatrix} s^2(Y) - \lambda_1 \\ -Cov(X,Y) \end{pmatrix}$$

et pour image du vecteur $u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ le vecteur

$$u_2 = \frac{1}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X,Y))^2}} \begin{pmatrix} s^2(Y) - \lambda_2 \\ -Cov(X,Y) \end{pmatrix}$$

Diagonalisation de la matrice des variances-covariances (suite)

C'est donc la matrice V des vecteurs propres.

$$V = [Id_{\mathbb{R}^2}, \{u_1, u_2\}, \{e_1, e_2\}]$$

Réciproquement, la matrice de l'application identique de $\mathbb{R}^2$ muni de la base $\{e_1, e_2\}$ dans $\mathbb{R}^2$ muni de la base $\{u_1, u_2\}$ donne, par produits, pour image du vecteur $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ le vecteur

$$e_1 = \begin{pmatrix} \dfrac{s^2(Y) - \lambda_1}{\sqrt{(s^2(Y) - \lambda_1)^2 + (Cov(X, Y))^2}} \\ \dfrac{s^2(Y) - \lambda_2}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X, Y))^2}} \end{pmatrix}$$

Diagonalisation de la matrice des variances-covariances (suite)

et pour image du vecteur $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

le vecteur $e_2 = \begin{pmatrix} \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y)-\lambda_1)^2 + (Cov(X,Y))^2}} \\ \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y)-\lambda_2)^2 + (Cov(X,Y))^2}} \end{pmatrix}$.

C'est donc la matrice V^t transposée et inverse de la matrice V des vecteurs propres.

$$V^t = [Id_{\mathbb{R}^2}, \{e_1, e_2\}, \{u_1, u_2\}]$$

Diagonalisation de la matrice des variances-covariances (suite)

Le diagramme commutatif suivant :

$$\begin{tikzcd}
\mathbb{R}^2, \{e_1, e_2\} \arrow[r, "f"', "A"] & \mathbb{R}^2, \{e_1, e_2\} \arrow[d, "Id"', "V"] \\
\mathbb{R}^2, \{u_1, u_2\} \arrow[u, "Id", "V^t"'] \arrow[r, "f", "A"'] & \mathbb{R}^2, \{u_1, u_2\}
\end{tikzcd}$$

met en évidence la relation $f = Id \circ f \circ Id$.

Diagonalisation de la matrice des variances-covariances (suite et fin)

En termes de produit de matrices, cette relation s'écrit :

$$\Lambda = VA^tV$$

d'où l'on déduit aussitôt

$$V = V^t \Lambda V.$$

On dit qu'on a **diagonalisé le matrice** A.

Pour un vecteur normé u, posons $v = Vu$. On a $v^t = u^t V^t$.

$$\|v\|^2 = v^t v = u^t V^t V u = u^t u = \|u\|^2 = 1$$

Le vecteur v est normé lui aussi. L'inertie statistique par rapport à u s'écrit :

$$I_S(u) = u^t A u = u^t V^t \Lambda V u = v^t \Lambda v$$

Dans $\mathbb{R}^2$ rapporté à la base $\{u_1, u_2\}$, notons $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$

$I_S(u) = v^t \Lambda v = \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda_1 v_1^2 + \lambda_2 v_2^2,$

avec $v_1^2 + v_2^2 = 1$

Recherche des axes principaux (suite)

Le problème de la recherche de la droite de régression orthogonale se ramène maintenant à la résolution du problème suivant :
maximiser $\lambda_1 \mathbf{v_1^2} + \lambda_2 \mathbf{v_2^2}$ sous la contrainte $v_1^2 + v_2^2 = 1$, avec $\lambda_1 > \lambda_2 > 0$
C'est maintenant un problème facile à résoudre :

$$I_S(u) = \lambda_1 v_1^2 + \lambda_2 v_2^2 = \lambda_1(1 - v_2^2) + \lambda_2 v_2^2 = \lambda_1 - (\lambda_1 - \lambda_2) v_2^2$$

La quantité $\lambda_1 - (\lambda_1 - \lambda_2) v_2^2,$ avec $\lambda_1 > \lambda_2$ atteint sa valeur maximum λ_1 lorsqu'on prend $v_2 = 0,$ donc $|v_1| = 1$.
La direction du premier axe factoriel est donc définie par le vecteur v de coordonnées $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ dans la base $\{u_1, u_2\}$: $v = u_1$.

$$I_S(u_1) = \lambda_1$$

Recherche des axes principaux (suite)

D'où le résultat, qu'on peut énoncer sous forme de **thèorème.**

Théorème

La direction du premier axe factoriel est définie par le vecteur propre associé à la plus grande valeur propre de la matrice des variances-covariances.

Le premier axe factoriel est la droite de régression orthogonale.
Comme corollaire, la direction perpendiculaire au premier axe factoriel définit le **deuxième axe factoriel** : elle est définie par le vecteur propre associé à la plus petite valeur propre de la matrice des variances-covariances.

Le deuxième axe factoriel minimise l'inertie statistique $I_S(u$:

$I_S(u) = \lambda_2$ lorsque $|v_2| = 1$, donc $v_1 = 0$ et $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = u_2$ par exemple (on pourrait prendre aussi, bien sûr, $v = -u_2$, la direction serait la même).

$$I_S(u_2) = \lambda_2$$

Le taux d'inertie totale expliquée par le premier axe factoriel est le rapport

$$\frac{I_S(u_1)}{I_T} = \frac{\lambda_1}{s^2(X) + s^2(Y)} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Recherche des axes principaux (suite et fin)

Le taux d'inertie totale expliquée par le deuxième axe factoriel est le rapport

$$\frac{I_S(u_2)}{I_T} = \frac{\lambda_2}{s^2(X) + s^2(Y)} = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

La relation $\lambda_1 + \lambda_2 = s^2(X) + s^2(Y)$ (la somme des valeurs propres est la trace de la matrice des variances-covariances) s'écrit :

$$I_S(u_1) + I_S(u_2) = I_T$$

La somme des inerties statistiques par rapport aux deux axes factoriels est l'inertie totale du nuage de points. Chaque valeur propre de la matrice des variances-covariances correspond à l'inertie expliquée par l'axe factoriel correspondant.

Coordonnées factorielles et composantes principales

Dans $\mathbb{R}^2$ rapporté à la base propre orthonormée $\{u_1, u_2\}$ les coordonnées des vecteurs $\overrightarrow{GM_i}$ s'appellent les **coordonnées factorielles.**

Comme la base $\{u_1, u_2\}$ est orthonormée, les coordonnées factorielles s'obtiennent par produit scalaire :

$$\overrightarrow{GM_i} = \langle \overrightarrow{GM_i} | u_1 \rangle u_1 + \langle \overrightarrow{GM_i} | u_2 \rangle u_2$$

Or la base canonique $\{e_1, e_2\}$ est, elle même, orthonormée

$$\overrightarrow{GM_i} = \langle \overrightarrow{GM_i} | e_1 \rangle e_1 + \langle \overrightarrow{GM_i} | e_2 \rangle e_2 = x_{i0} e_1 + y_{i0} e_2$$

d'où :

$$\langle \overrightarrow{GM_i} | u_1 \rangle = x_{i0} \langle e_1 | u_1 \rangle + y_{i0} \langle e_2 | u_1 \rangle$$
$$\langle \overrightarrow{GM_i} | u_2 \rangle = x_{i0} \langle e_1 | u_2 \rangle + y_{i0} \langle e_2 | u_2 \rangle$$

Coordonnées factorielles et composantes principales (suite)

Les coordonnées factorielles s'obtiennent donc par la formule matricielle :

$$\begin{pmatrix} \langle \overrightarrow{GM_i} | u_1 \rangle \\ \langle \overrightarrow{GM_i} | u_2 \rangle \end{pmatrix} = \begin{pmatrix} \langle e_1 | u_1 \rangle & \langle e_2 | u_1 \rangle \\ \langle e_1 | u_2 \rangle & \langle e_2 | u_2 \rangle \end{pmatrix} \begin{pmatrix} x_{i0} \\ y_{i0} \end{pmatrix} = V^t \begin{pmatrix} \langle \overrightarrow{GM_i} | e_1 \rangle \\ \langle \overrightarrow{GM_i} | e_2 \rangle \end{pmatrix}$$

$$\begin{pmatrix} \langle \overrightarrow{GM_i} | u_1 \rangle \\ \langle \overrightarrow{GM_i} | u_2 \rangle \end{pmatrix} = V^t \begin{pmatrix} \langle \overrightarrow{GM_i} | e_1 \rangle \\ \langle \overrightarrow{GM_i} | e_2 \rangle \end{pmatrix} = V^t \begin{pmatrix} x_{i0} \\ y_{i0} \end{pmatrix}$$

La matrice V^t est ce qu'on appelle la **matrice du changement de base.**
Elle donne les nouvelles coordonnées (sur la base $\{u_1, u_2\}$) en fonction des anciennes (sur la base $\{e_1, e_2\}$).

Coordonnées factorielles et composantes principales (suite)

Les relations

$$\left(\langle \overrightarrow{GM_i} | u_1 \rangle \quad \langle \overrightarrow{GM_i} | u_2 \rangle\right) = \begin{pmatrix} \langle \overrightarrow{GM_i} | u_1 \rangle \\ \langle \overrightarrow{GM_i} | u_2 \rangle \end{pmatrix}^t = \left(V^t \begin{pmatrix} x_{i0} \\ y_{i0} \end{pmatrix} \right)^t = \left(x_{i0}\right.$$

peuvent se condenser en une seule formule matricielle :

$$L = ZV$$

formule dans laquelle :

$$L = \begin{pmatrix} \langle \overrightarrow{GM_1} | u_1 \langle & \langle \overrightarrow{GM_1} | u_2 \rangle \\ \vdots & \vdots \\ \langle \overrightarrow{GM_n} | u_1 \langle & \langle \overrightarrow{GM_n} | u_2 \rangle \end{pmatrix}$$

Coordonnées factorielles et composantes principales (suite)

est la matrice, à n ligne et 2 colonnes, dont les lignes sont les cordonnées factorielles du nuage de points dans $\mathbb{R}^2$ muni de la base $\{u_1, u_2\}$,

$$Z = \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix}$$

est la matrice, à n lignes et 2 colonnes, dont les colonnes sont les variables centrées
$X - \overline{X}$ et $Y - \overline{Y}$,

Coordonnées factorielles et composantes principales (suite)

$$V = \begin{pmatrix} \dfrac{s^2(Y)-\lambda_1}{\sqrt{(s^2(Y)-\lambda_1)^2+(Cov(X,Y))^2}} & \dfrac{s^2(Y)-\lambda_2}{\sqrt{(s^2(Y)-\lambda_2)^2+(Cov(} \\ \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y)-\lambda_1)^2+(Cov(X,Y))^2}} & \dfrac{-Cov(X,Y)}{\sqrt{(s^2(Y)-\lambda_2)^2+(Cov(} \end{pmatrix}$$

est la matrice des coordonnées des vecteurs propres orthonormés $\{u_1, u_2\}$ de la matrice des variances-covariances, dans la base canonique $\{e_1, e_2\}$.

Les deux colonnes de la matrice L sont des éléments de l'espace des variables $\mathbb{R}^n$: on les appelle les **composantes principales** de la variable statistique (X,Y).

Coordonnées factorielles et composantes principales (suite)

La première colonne de la matrice V est le vecteur propre u_1.
La première colonne de la matrice L=ZV est donc le vecteur propre $L_1 = Zu_1$.
De même, la deuxième colonne de la matrice L est le vecteur $L_2 = Zu_2$.
Les deux composantes principales L_1 et L_2 de la variable statistique (X,Y) s'obtiennent ainsi par les formules :

Coordonnées factorielles et composantes principales (suite)

$$L_1 = \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} u_1 =$$

$$\frac{1}{\sqrt{(s^2(Y)-\lambda_1)^2 + (Cov(X,Y))^2}} \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} \begin{pmatrix} s^2(Y)-\lambda_1 \\ -Cov(X,Y) \end{pmatrix}$$

$$L_2 = \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} u_2 =$$

$$\frac{1}{\sqrt{(s^2(Y) - \lambda_2)^2 + (Cov(X, Y))^2}} \begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} \begin{pmatrix} s^2(Y) - \lambda_2 \\ -Cov(X, Y) \end{pmatrix}$$

Coordonnées factorielles et composantes principales (suite)

avec les valeurs propres λ_1 et λ_2 de la matrice

$$A = \frac{1}{n}\begin{pmatrix} x_{10} & \cdots & y_{10} \\ x_{n0} & \cdots & y_{n0} \end{pmatrix}\begin{pmatrix} x_{10} & y_{10} \\ \vdots & \vdots \\ x_{n0} & y_{n0} \end{pmatrix} =$$

$$\frac{1}{n}Z^tZ = Z^tD_{\frac{1}{n}}Z = \begin{pmatrix} s^2(X) & Cov(X,Y) \\ Cov(X,Y) & s^2(Y) \end{pmatrix}$$ des variances-covariances :

$$\lambda_1 = \frac{1}{2}\left(s^2(X) + s^2(Y) + \sqrt{(s^2(X) - s^2(Y))^2 + 4(Cov(X,Y)$$

$$\lambda_2 = \frac{1}{2}\left(s^2(X) + s^2(Y) - \sqrt{(s^2(X) - s^2(Y))^2 + 4(Cov(X,Y)$$

$$\overline{L_1} = \langle L_1 | D_{\frac{1}{n}} | 1_n \rangle = \frac{1}{n} \langle Zu_1 | 1_n \rangle = \frac{1}{n} (Zu_1)^t 1_n = \frac{1}{n} u_1^t Z^t 1_n$$

$$Z^t 1_n = \begin{pmatrix} x_{10} & \cdots & y_{10} \\ x_{n0} & \cdots & y_{n0} \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_{i0} \\ \sum_{i=1}^{n} y_{i0} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

puisque les variables X_0 et Y_0 sont centrées.

Il reste donc :

$$\overline{L_1} = \frac{1}{n} u_1^t \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

De même

$$\overline{L_2} = \langle L_2 | D_{\frac{1}{n}} | 1_n \rangle = \frac{1}{n} \langle Zu_2 | 1_n \rangle = \frac{1}{n} (Zu_2)^t 1_n = \frac{1}{n} u_2^t Z^t 1_n = \frac{1}{n} u_2^t \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

Propriétés des composantes principales

La variance d'une composante principale est la valeur propre correspondante

Comme les composantes sont centrées, leur variance est le carré de leur norme pour le produit scalaire défini par $D_{\frac{1}{n}}$:

$$s^2(L_1) = \|L_1\|^2_{D_{\frac{1}{n}}} = \langle L_1 | D_{\frac{1}{n}} | L_1 \rangle = \frac{1}{n} L_1^t L_1 = \frac{1}{n} u_1^t Z^t Z u_1$$

$$\frac{1}{n} Z^t Z = A$$

$$s^2(L_1) = u_1^t A u_1 = u_1^t \lambda_1 u_1 = \lambda_1 \|u_1\|^2 = \lambda_1$$

De même $s^2(L_2) = \|L_2\|^2_{D_{\frac{1}{n}}} = \langle L_2 | D_{\frac{1}{n}} | L_2 \rangle = \frac{1}{n} L_2^t L_2 = \frac{1}{n} u_2^t Z^t Z u_2 = u_2^t A u_2 = u_2^t \lambda_2 u_2 = \lambda_2 \|u_2\|^2 = \lambda_2$

$$Cov(L_1, L_2) = \langle L_1 | D_{\frac{1}{n}} | L_2 \rangle = \frac{1}{n} L_1^t L_2 = \frac{1}{n} u_1^t Z^t Z u_2 = \frac{1}{n} u_1^t A u_2 = \frac{\lambda_2}{n} \langle u_1 | u_2 \rangle = 0$$

puisque les vecteurs u_1 et u_2 sont orthogonaux pour le produit scalaire canonique.