Analyse de données

Description bidimensionelle et mesure de liaison entre variables

Couples de variables aléatoires

Soit 2 variables aléatoires X et Y (dans le cas discret) définies sur un espace probabilisé $(\Omega, \mathscr{C}, \mathbb{P})$

- Ω Univers (l'ensemble des éléments élémentaires)
- $\mathscr{C}$ attribu tel que $\mathscr{C}$ inclu dans $P(\Omega)$ (== “toutes les parties de Omega”)
- $\mathbb{P}$ probabilité

Exemple :

> Un évènement A = $\{(i,j)/i\} + j \geq 10$,
> $\mathscr{C} \in P$

$X(\omega) = \{X_i \ / \ i \in I\}$ l'ensemble des valeurs de X
$Y(\omega) = \{Y_j \ / \ j \in J\}$ l'ensemble des valeurs de Y

Loi conjointe (X, Y)

Définition :
On appelle loi conjointe du couple (X, Y) l'ensemble des couples $((X_i, Y_j), P_{i,j})$ avec $\forall x_i \in X_\omega, \forall y_i \in Y_\omega$

$$P_{i,j} = \mathbb{P}((X = x_i) \cap (Y = y_j))$$

$P_{i,j} \geqslant 0$ et $\sum_{i,j} P_{i,j} = 1$

Remarque :
$I = [\![1, r]\!]$, $J = [\![1, s]\!]$

$$\left[\begin{array}{ccccccc}
 & y_1 & \dots & y_j & \dots & y_s & \text{Loi de X} \\
x_1 & P_{1,1} & & \vdots & & P_{1,s} & P_{1,\cdot} \\
\vdots & & & \vdots & & & \vdots \\
x_i & \dots & \dots & \boxed{P_{i,j}} & & & P_{i,\cdot} \\
\vdots & & & & & & \vdots \\
x_r & P_{r,1} & & & & P_{r,s} & P_{r,\cdot} \\
\hline
\text{Loi de Y} & P_{\cdot,1} & \dots & P_{\cdot,j} & \dots & P_{\cdot,s} & 1
\end{array}\right]$$

Lois marginales

Les V.A. X, Y sont appelees variables marginales des couples (X, Y)

Loi marginale de X :

$$\boxed{\mathbb{P}[X = x_i] = P_{i.} = \sum_{j \in J} P_{i,j}}$$

Loi marginale de Y :

$$\boxed{\mathbb{P}[Y = y_j] = P_{.j} = \sum_{i \in I} P_{i,j}}$$

Exemple :

$$\left[\begin{array}{cccccc}
 & 1 & 2 & 3 & 4 & \text{Loi de X} \\
1 & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{4} \\
2 & 0 & \frac{2}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{4} \\
3 & 0 & 0 & \frac{3}{16} & \frac{1}{16} & \frac{1}{4} \\
4 & 0 & 0 & 0 & \frac{4}{16} & \frac{1}{4} \\
\text{Loi de Y} & \frac{1}{16} & \frac{3}{16} & \frac{5}{16} & \frac{7}{16} & 1
\end{array}\right]$$

Lois conditionnelles

On appelle loi conditionnelle de $X = x_i$ sachant que $Y = y_j$:

$$\mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}((X = x_i) \cap (Y = y_i))}{\mathbb{P}(Y = y_j)} = \frac{\mathbb{P}_{\text{ij}}}{\mathbb{P}_{.j}}$$

avec $\mathbb{P}(Y = y_j) \neq 0$

$$\mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}_{i,j}}{\mathbb{P}_{.j}}$$

$$\mathbb{P}(Y = y_j | X = x_i) = \frac{\mathbb{P}_{i,j}}{\mathbb{P}_{i.}}$$

Exemple : (matrice précédente)

$\mathbb{P}(X = x_i | Y = 3)$

X_i	**1**	**2**	**3**	**4**
$\mathbb{P}(X = x_i \| \| Y = 3)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{5}$	0

$$\mathbb{P}(X = 1 | Y = 3) = \frac{\mathbb{P}((X = 1) \cap (Y = 3))}{\mathbb{P}(Y = 3)} = \frac{\frac{1}{16}}{\frac{5}{16}} = \frac{1}{5}$$

Définition : [Indépendance]
X et Y sont indépendantes ssi

$$\mathbb{P}((X = x_i) \cap (Y = y_j)) = \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_j)$$

$\forall x_i \in X(\omega), \forall y_j \in Y(\omega)$
Equivalent à :

$$\mathbb{P}_{i,j} = \mathbb{P}_{i.} \cdot \mathbb{P}_{.j}$$

Loi d'une fonction de 2 variables

Soit g : $\mathbb{R}^2 \longmapsto \mathbb{R}$ et Z la variable aléatoire $Z = g(X, Y)$ avec (X, Y) couple de valeurs aléatoires,
g est une fonction quelconque.

Les valeurs prises par Z : $g(x_i, y_j)$, $x_i \in X(\omega)$, $y_j \in Y(\omega)$

$$Z(\omega) = \{Z_k \ / \ k \in K\}_{K \subset \mathbb{N} \text{ ou } \mathbb{Z}}$$

$$\underbrace{[Z = z_k]}_{\text{événement}} = \bigcup_{(i,j)} ([X = x_i] \cap [Y = y_j])$$

Evènements incompatibles : leur intersection est vide.
Soit
$A_{i,j} \bigcap A_{i',j'} = \emptyset$
$(i,j) \neq (i',j')$

$$\mathbb{P}(Z = z_k) = \sum_{(i,j)\ /\ g(X)_{ij} = Z_k} \mathbb{P}((X = x_i) \cap (Y = y_j))$$

Remarque :
En particulier $Z = X + Y$

$$\mathbb{P}(Z = x + y) = \sum_{x+y=Z} \mathbb{P}((X = x) \cap (Y = y))$$

Si $Z = X \cdot Y$

$$\mathbb{P}(Z = x \cdot y) = \sum_{xy=Z} \mathbb{P}((X = x) \cap (Y = y))$$

Exemple :

X\Y	1	2	3	4
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
2	0	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
3	0	0	$\frac{3}{16}$	$\frac{1}{16}$
4	0	0	0	$\frac{4}{16}$

1 Déterminer la loi de $S = X + Y$
2 Déterminer la loi de $P = X \cdot Y$

$S(\omega) = \{2, 3, 4, 5, 6, 7, 8\}$

S_i	2	3	4	5	6	7	8
$P_i = P(S = S_i)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{4}{16}$	$\frac{1}{16}$	$\frac{4}{16}$

$P = X \cdot Y$ (produits)

$p(\omega) = \{1, 2, 3, 4, 6, 8, 9, 12, 16\}$

P_i	1	2	3	4	6	8	9	12	16
$\mathbb{P}(P = P_i)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{4}{16}$

Definition :
Soit $Z = g(X, Y)$
On appelle esperance de Z

$$E(Z) = \sum_{\substack{i \in I \\ j \in J}} (g(x_i, y_j) \cdot \mathbb{P}((X = x_i) \cap (Y = y_i)))$$

Remarque :

$$E(XY) = \sum_i \sum_j x_i y_j \mathbb{P}_{i,j}$$

Si 2 V.A. X et Y sont indépendantes alors

$$E(XY) = E(X)E(Y)$$

- Reciproque est fausse en général
- Condition suffisante non necessaire

Contre-exemple :

X\Y	0	1	2	Loi de X
0	$\frac{1}{20}$	$\frac{1}{4}$	0	$\frac{3}{10}$
1	$\frac{17}{60}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{7}{10}$
Loi de Y	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$	1

$$E(X \cdot Y) = \sum_{i=0}^{1} \sum_{j=0}^{2} x_i \cdot y_j \; P_{i,j} = \frac{1}{4} \times 1 + \frac{1}{6} \times 2$$

$$E(X) = \sum_{i=0}^{1} x_i \; P_i = \frac{7}{10}$$

$$E(Y) = \sum_{j=0}^{2} j \; P_j = \frac{1}{2} + \frac{2}{6} = \frac{5}{6}$$

$E(X)E(Y) = \frac{35}{60} = \frac{7}{12} = E(XY)$
$E(X \cdot Y) = E(X)E(Y)$
Mais X et Y sont dépendantes car
$\mathbb{P}((X = 0) \cap (Y = 2)) = 0 \neq \mathbb{P}((X = 0)\mathbb{P}(Y = 2)) = \frac{1}{20}$

Définition : [Covariance]
Soit 2 V.A. X,Y discrètes
On appelle covariance de x,y l'espérance de xy - le produit des espérances respectives.

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$$V(X+Y) = V(X) + V(Y) + 2Cov(XY)$$

Si X et Y sont indépendantes $\Longrightarrow Cov(X,Y) = 0$
$\boxed{V(X+Y) = V(X) + V(Y)}$

Corrélation linéaire

Défintion :
On appelle cœfficient de corrélation entre X et Y,

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$
$$\underbrace{\sigma_X}_{\text{écart-type}} = \sqrt{V(X)}, \sigma_Y = \sqrt{V(Y)}$$

Remarque :

- $\sigma_x = ||X - E(X)||$, $\sigma_y = ||Y - E(Y)||$
- $\underbrace{\langle X - E(X), Y - E(Y) \rangle}_{\text{produit scalaire}} = Cov(X,Y)$
- $\rho(X,Y) = \frac{\langle X-E(X), Y-E(Y) \rangle}{||X-E(X)|| ||Y-E(Y)||}$
- $\cos\theta = \frac{\langle u,v \rangle}{||u|| ||v||}$
- $|\rho(X,Y)| \leqslant 1$
- $|\rho(X,Y)| = 1 \longrightarrow X, V$ colinéaire
- $|\rho(X,Y)| = 0 \longrightarrow X, V$ perpendiculaire

Exercice d'application :
"À préparer pour la prochaine fois" :

X_i	**-2**	**-1**	**0**	**1**	**2**
$\mathbb{P}(X = X_i)$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$

Soit $Y = X^2$.

1. Donner la loi conjointe de (X, Y)
2. Loi marginale de Y
3. Indépendance?
4. Calculer $Cov(X, Y)$

La loi conjointe de (X, Y)

- $P_{i,j} = \mathbb{P}((X = i) \cap (Y = j)) = 0$ si $j \neq i^2$
- $P_{i,j} = \mathbb{P}(\underbrace{(X = i) \cap (Y = i^2)}_{\{X=i\} \subset \{Y=i^2\}}) = \mathbb{P}(X = i)$ si $j = i^2$

$X\backslash Y$	**0**	**1**	**4**	**Loi de X**
-2	0	0	$\frac{1}{6}$	$\frac{1}{6}$
-1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
0	$\frac{1}{6}$	0	0	$\frac{1}{6}$
1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
2	0	0	$\frac{1}{6}$	$\frac{1}{6}$
Loi de Y	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$	1

Loi marginale de Y
D'aprés le tableau de la loi conjointe

i	**0**	**1**	**4**
$\mathbb{P}(Y = Y_i)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

Indépendance
$P_{ij} = \mathbb{P}((X = x_i) \cap (Y = y_j)) =^? \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_j) \ \forall (i, j)$
Or $P_{ij} = \mathbb{P}((X = 0) \cap (Y = 1)) = 0 \neq \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 1) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$
$\Rightarrow X$ et Y ne sont pas indépendants

La covariance

$$Cov(X,Y) = E(X \cdot Y) - E(X)E(Y)$$
$$E(X \cdot Y) = \sum_{i,j} x_i y_j \mathbb{P}_{i,j} = \sum_i x_i \sum_j y_j \mathbb{P}_{i,j}$$
$$\text{avec } P_{i,j} = \mathbb{P}((X = x_i) \cap (Y = y_j))$$
$$E(XY) = -\frac{8}{6} - \frac{1}{4} + \frac{1}{4} + \frac{8}{6} = 0$$

$$E(Y) = \sum y_i \mathbb{P}(Y = y_i) = \frac{1}{2} + \frac{4}{3} = \frac{11}{6}$$

$$E(X) = \sum x_i \mathbb{P}(X = x_i) = -\frac{1}{3} - \frac{1}{4} + \frac{1}{4} + \frac{1}{3} = 0$$

$Cov(X,Y) = 0$
On peut noter que lorsque **X** et **Y** sont indépendante la covariance est nulle. Cependant la réciproque n'est pas vrai.
La preuve dans cet exercice **X** et **Y** sont dépendante et leur covariance est nulle.

Exercice 2
On a n boîtes numérotées de 1 à n
La boîte n°k contient k boules numérotées de 1 à k.
On choisit au hasard une boîte puis une boule dans cette boîte.
X V.A le numéro de la boîte
Y V.A le numéro de la boule

1. Déterminer la loi conjointe
2. Calculer $\mathbb{P}(X = Y)$
3. Déterminer la loi marginale de Y
4. Calculer $E(Y)$

$X(\omega) = \{1, 2, \dots, n\}$
$Y(\omega) = \{1, 2, \dots, n\}$
Loi de couple (X, Y) ?
$\mathbb{P}((X = i) \cap (Y = j))$

$\forall i \in [\![1, n]\!]$
$\forall j \in [\![1, n]\!]$

Loi conjointe
Soit $j > i$
$P_{ij} = \mathbb{P}((X = x_i) \cap (Y = y_j)) = 0$

Soit $j <= i$
$P_{ij} = \mathbb{P}((X = x_i) \cap (Y = y_j))$
$= \mathbb{P}(Y = y_j | X = x_i) \cdot \mathbb{P}(X = x_i)$
$= \frac{1}{i} \cdot \frac{1}{n}$

$\mathbb{P}(X = Y)$

$$(X = Y) = \bigcup_{i=1}^{n} (X = x_i) \cap (Y = y_i)$$

évènements indépendants donc

$$\mathbb{P}(X = Y) = \sum_{i=1}^{n} P_{ii} = \sum_{i=1}^{n} \frac{1}{i} \cdot \frac{1}{n} = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{i}$$

Loi marginale de Y

$$\mathbb{P}(Y = j) = \sum_{i}^{n} ((X = x_i) \cap (Y = y_j))$$

$$= \sum_{i=j}^{n} \frac{1}{i \cdot n} = \frac{1}{n} \sum_{i=j}^{n} \frac{1}{i}$$

Espérance de Y

$$E(Y) = \sum_{j=1}^{n} j \cdot \mathbb{P}(Y = y_j)$$

$$E(Y) = \sum_{j=1}^{n} j \cdot \frac{1}{n} \sum_{i=j}^{n} \frac{1}{i}$$
$$E(Y) = \frac{1}{n} \sum_{j=1}^{n} j \sum_{i=j}^{n} \frac{1}{i}$$
$$E(Y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{i} \sum_{j=1}^{i} j$$
$$E(Y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{i} \cdot \frac{i(i+1)}{2}$$
$$E(Y) = \frac{1}{2n} \sum_{i=1}^{n} i + 1$$
$$E(Y) = \frac{1}{2n}(\frac{n(n+1)}{2} + n)$$
$$E(Y) = \frac{1}{2}(\frac{n+1}{2} + 1)$$
$$E(Y) = \frac{n+3}{4}$$

Exercice 3
On dispose d'un dé équilibré à 6 faces
et d'une pièce truquée telle que la probabilité d'obtenir pile vaut $p\in]0,1[$
Soit $N \in \mathbb{N}^*$ (fini). On effectue N lancés de dé. Si n est le nombre de six obtenu, on lance alors n fois la pièce.
Z V.A nombre de six obtenus
X V.A nombre de pile
Y V.A nombre de face
$Z = X + Y$
$Z = 0$ si $X = Y = 0$

1. déterminer la loi de Z, E(Z), V(Z)
2. $\forall k \in \mathbb{N}, \forall n \in \mathbb{N}$ déterminer $\mathbb{P}(X = k|Z = n)$
3. $\forall 0 \leq k \leq n < N$ calculer $\mathbb{P}((X = k) \cap (Z = n))$
4. calculer $P(X = 0)$
5. Montrer que $\forall 0 \leq k \leq n \leq \mathbb{N}$, $\binom{n}{k}\binom{N}{n} = \binom{N}{k}\binom{N-k}{n-k}$
   Pour cette 5è question il s'agit de prouver une égalité de deux combinaisons de type k parmi n, cette égalité servira pour la question 6. Elle est facile à prouver car il suffit de développer LHS et RHS en utilisant la formule (k parmi n) = n!/k!(n-k)! puis de simplifier des deux côtés
6. En déduire $P(X = k)$
   Reconnaître la loi de X
7. Loi de Y ?
8. Calculer $Cov(X, Y)$, X et Y sont-ils indépendants ?
9. Loi de (X, Y) ?

1. $Z \hookrightarrow B(N, \frac{1}{6})$
   $E(Z) = \frac{N}{6}, V(Z) = N \cdot \frac{1}{6}\frac{5}{6} = \frac{5N}{36}$
2.

Soit $k > n$
$\mathbb{P}(X = k|Z = n) = 0$
Soit $0 \leq k \leq n$
$\mathbb{P}(X = k|Z = n) = \binom{n}{k}p^k(1-p)^{n-k}$

Donc $X|Z = n \hookrightarrow B(n, p)$

3. $\mathbb{P}((X = k) \cap (Z = n)) = \mathbb{P}(X = k|Z = n)\mathbb{P}(Z = n)$
   $\forall 0 \leq k \leq n \leq \mathbb{N}$
   $\binom{n}{k}p^k(1-p)^{n-k}\binom{N}{n}(\frac{1}{6})^n(\frac{5}{6})^{N-n}$

4. ** $P(X = k) = 0$ **

$$\mathbb{P}(x = 0) = \bigcup_{n=0}^{N} (X = 0) \cap (Z = n)$$
évènements incompatible donc

$$\mathbb{P}(x=0) = \sum_{n=0}^{N} P((X=0)\cap(Z=n))$$
$$\mathbb{P}(x=0) = \sum_{n=0}^{N}(1-p)^n\binom{N}{n}\frac{1}{6}^n\frac{5}{6}^{N-n}$$
$$\mathbb{P}(x=0) = \sum_{n=0}^{N}(1-p)^n\binom{N}{n}\frac{5^{N-n}}{6^N}$$
$$\mathbb{P}(x=0) = \tfrac{1}{6^N}\sum_{n=0}^{N}\binom{N}{n}(1-p)^n 5^{N-n}$$
$$\mathbb{P}(x=0) = \tfrac{1}{6^N}(6-p)^N$$
$$\mathbb{P}(x=0) = (1-\tfrac{p}{6})^N$$

Formule du binôme de Newton

$$(a+b)^N = \sum_{n=0}^{N}\binom{N}{n}a^n b^{N-n}$$

5. $\binom{n}{k}\binom{N}{n} =? \binom{N}{k}\binom{N-k}{n-k}$

$$\frac{n!}{k!(n-k)!}\frac{N!}{n!(N-n)!} = \frac{N!}{k!(N-k)!}\frac{(N-k)!}{(n-k)!(N-n)!}$$
les deux expressions se simplifient en
$$\frac{N!}{k!(n-k)!(N-n)!}$$

6)$\mathbb{P}(X=k) = \binom{N}{k}\frac{p^k}{6^N}\sum_{n=k}^{N}\binom{N-k}{n-k}(1-p)^{n-k}5^{N-n}$

Posons $i = n-k$

$$\mathcal{P}(X=k) = \binom{N}{k}\frac{p^k}{6^N}\sum_{i=0}^{N-k}\binom{N-k}{i}(1-p)^i 5^{N-k-i}$$
$$= \binom{N}{k}\frac{p^k}{6^N}(1-p+5)^{N-k} = \binom{N}{k}(\tfrac{p}{6})^k(1-\tfrac{p}{6})^{N-k}$$
$X \hookrightarrow \mathcal{B}(N, p/6)$

7. Même raisonnement pour la loi de Y
   $Y \hookrightarrow \mathcal{B}(N, q/6)$ où $q = 1-p$

8. $Cov(X,Y) =?$
   Rappel : Covariance d'une somme dans le cas general
   $Var(Z) = Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

$Z \hookrightarrow \mathcal{B}(N, 1/6)$
$X \hookrightarrow \mathcal{B}(N, p/6)$
$Y \hookrightarrow \mathcal{B}(N, q/6)$

$X \hookrightarrow \mathcal{B}(n,p)$
$E(X) = np$
$V(X) = npq = np(1-p)$

$Cov(X,Y) = \frac{1}{2}(V(Z) - V(X) - V(Y))$
$Cov(X,Y) = \frac{1}{2}(\frac{5N}{36} - \frac{Np}{6}(1-\frac{p}{6}) - \frac{Nq}{6}(1-\frac{q}{6}))$
$Cov(X,Y) = \frac{1}{2} \times \frac{1}{36}(5N - NP(6-p) - Nq(6-q))$
$Cov(X,Y) = \frac{1}{2} \times \frac{1}{36}(5N - 6Np + Np^2 - 6Nq + Nq^2)$
$Cov(X,Y) = \frac{1}{2} \times \frac{1}{36}(5N - 6Np + Np^2 - 6N(1-p) + N(1-2p+p^2))$
$Cov(X,Y) = \frac{1}{2} \times \frac{1}{36}(2Np^2 - 2Np)$
$Cov(X,Y) = \frac{1}{36}(Np(p-1)) \neq 0$ car $p \neq 0, p \neq 1$
$\Rightarrow$ X et Y ne sont pas indépendants

9. Loi du couple (X,Y)
   $Z \in [\![0,N]\!]$
   $X = i$ et $Y = j$
   $Z = i + j$
   $\mathbb{P}((X=i) \cap (Y=j)) = \mathbb{P}((X=i) \cap (Z=i+j))$
   $= \binom{n}{i}\binom{N}{i+j}p^i(1-p)^{n-i}(\frac{5}{6})^{N-(i+j)}(\frac{1}{6})^{i+j}$

Exercice 4

Une urne contient une boule blanche et une boule noire. On y prélève une boule, chaque boule ayant la même probabilité d'être tirée.
On note sa couleur et on la remet dans l'urne, avec C boules de la même couleur de la boule tirée. On repète cette épreuve n fois. ($n \geq 2$)
On définit:

$$X_i = \begin{cases} 1 \text{ si on obtient une boule blanche au i}^{\text{ème}} tirage \\ 0 \text{ sinon} \end{cases}$$

$Z_p = \sum_{i=1}^{p} X_i$

1. Déterminer la loi du couple (X_1, X_2) et en déduire la loi de X_2
2. Déterminer la loi de Z_2
3. Déterminer $\mathbb{P}(X_{p+1} = 1/Z_p = k)$
4. Montrer que $\mathbb{P}(X_{p+1} = 1) = \frac{1+C\mathbb{E}(Z_p)}{2+pc}$
5. Montrer que $\forall p; 1 \leq p \leq n$
   $\mathbb{P}(X_p = 1) = \mathbb{P}(X_p = 0) = \frac{1}{2}$
   Récurrence sur p.

1. $\mathbb{P}(X_1=1)=\mathbb{P}(X_1=0)=\frac{1}{2}$
   Donc $X_1 \hookrightarrow \mathcal{B}(p\frac{1}{2})$ Bernoulli

   $\mathbb{P}((X_1=i)\cap(X_2=k))=\mathbb{P}(X_2=k|X_1=i)\overbrace{\mathbb{P}(X_1=i)}^{\frac{1}{2}}$
   1^{er} cas $i\neq k$
   $\mathbb{P}(X_2=k|X_1=i)=\frac{1}{2+C}\rightarrow\mathbb{P}((X_1=i)\cap(X_2=k))=\frac{1}{2*(2+C)}$
   2^{me} cas $i=k$
   $\mathbb{P}(X_2=i|x_1=i)=\frac{1+C}{2+C}$

$X_1\backslash X_2$	**0**	**1**	**Loi de X_1**
0	$\frac{1+C}{2(2+C)}$	$\frac{1}{2(2+C)}$	$\frac{1}{2}$
1	$\frac{1}{2(2+C)}$	$\frac{1+C}{2(2+C)}$	$\frac{1}{2}$
Loi de X_2	$\frac{1}{2}$	$\frac{1}{2}$	1

$X_2\hookrightarrow\mathcal{B}(p=\frac{1}{2})$
$2\leq p\leq n$ et $Z_p=\sum_{i=1}^{p}X_i$

2. $Z_2=X_1+X_2$
   $Z_2(\omega)=\{0,1,2\}$
   $\mathbb{P}(Z_2=0)=\mathbb{P}((X_1=0)\cap(X_2=0))=\frac{1+C}{2(2+C)}$
   $\mathbb{P}(Z_2=1)=\mathbb{P}((X_1=1)\cap(X_2=0))+\mathbb{P}((X_1=0)\cap(X_2=1))=\frac{1}{2+C}$
   $\mathbb{P}(Z_2=2)=\mathbb{P}((X_1=1)\cap(X_2=1))=\frac{1+C}{2(2+C)}$

3. $\mathbb{P}(X_{p+1}=1|Z_p=k)$
   $Z_p(\omega)=[[0,p]]$
   $(Z_p=k)\Leftrightarrow$ au cours des p tirages on a obtenu k boules blanches et $(p-k)$ boules noires
   au $(p+1)^{\text{ième}}$ tirage l'urne contient: $2+pc$ boules dont $1+kc$ boules blanches
   Donc $\mathbb{P}(X_{p+1}=1|Z_p=k)=\frac{1+kc}{2+pc}$

4. $\mathbb{P}(X_{p+1}=1)$?
   $$(X_{p+1}=1)=\bigcup_{k=0}^{p}\underbrace{[(X_{p+1}=1)\cap(Z_p=k)]}_{\text{événements incompatibles}}$$
   $$\mathbb{P}(X_{p+1}=1)=\sum_{k=0}^{p}\mathbb{P}((X_{p+1}=1)\cap(Z_p=k))$$
   $$\mathbb{P}(X_{p+1}=1)=\sum_{k=0}^{p}\mathbb{P}(\frac{X_{p+1}=1}{Z_p=k})\mathbb{P}(Z_p=k)$$
   $$=\sum_{k=0}^{p}(\frac{1+kc}{2+pc}\mathbb{P}(Z_p=k))$$

$= \frac{1}{2+pc}(\underbrace{\sum_{k=0}^{p} \mathbb{P}(Z_p = k)}_{=1} + c\underbrace{\sum_{k=0}^{p} k\mathbb{P}(Z_p = k)}_{E(Z_p)})$

5. Raisonnement par récurrence sur p
   si $p = 1, \mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 0) = \frac{1}{2} X_1 \hookrightarrow \mathcal{B}(\frac{1}{2})$
   $p = 2, \mathbb{P}(X_2 = 1) = \mathbb{P}(X_2 = 0) = \frac{1}{2}$
   $X_2 \hookrightarrow \mathcal{B}(\frac{1}{2})$
   \underline{Hypothèse de récurrence} Supposons que cette propriété reste vraie jusqu'au rang p
   $$Z_p = \sum_{i=1}^{p} X_i$$
   $$E(Z_p) = \sum_{i=1}^{p} E(X_i) = \sum_{i=1}^{p} \frac{1}{2} = \frac{p}{2}$$
   $\Rightarrow \mathbb{P}(X_{p+1} = 1) = \frac{1+cp/2}{2+pc} = \frac{1}{2}(\frac{2+pc}{2+pc}) = \frac{1}{2}$
   $\mathbb{P}(X_{p+1} = 0) = 1 - \frac{1}{2} = \frac{1}{2}$

Exercice 5

$a \in \mathbb{R}$
X, Y 2 variables aléatoires à valeurs dans $\mathbb{N}$
$\mathbb{P}((X = k) \cap (Y = j)) = \frac{a}{2^{k+1} j!}$

1. Déterminer la constante a
2. Déterminer les lois marginales de X et Y
3. Déterminer l'indépendance des variables X et Y
4. Calculer la covariance $Cov(X, Y)$

1. Donc il faut que $a \geq 0$ et $\sum_{k=0,j=0} P_{kj} = 1$

   $$\sum_{k,j} \mathbb{P}((\mathbb{X} = \mathbb{k}) \cap (\mathbb{X} = \mathbb{j})) = 1 \Rightarrow \sum_{k,j} \frac{a}{2^{k+1} j!} = 1$$

   $$\Rightarrow a \sum_{k,j=0}^{+\infty} \frac{1}{2^{k+1} j!} = 1$$

   $$\Rightarrow a \underbrace{\sum_{j=0}^{+\infty} \frac{1}{j!}}_{e} \underbrace{\sum_{k=0}^{+\infty} \frac{1}{2^{k+1}}}_{1} = 1$$

   $$\Rightarrow a \cdot e = 1 \Rightarrow a = \frac{1}{e} = e^{-1}$$

> *Rappel* :
>
> $$e^x = \sum_{0}^{+\infty} \frac{x^k}{k!} \forall x \in \mathbb{R}$$
>
> $$e = \sum_{k=0}^{+\infty} \frac{1}{k!} (x = 1)$$

$$\sum_{k=0}^{+\infty} \frac{1}{2^{k+1}} = 1/2$$

Forme géométrique $\sum_{0}^{+\infty} a^k = \frac{1}{1-a}$ avec $|a| < 1$

$$\sum_{k=0}^{+\infty} \frac{1}{2^k} = \frac{1}{2} \frac{1}{1 - \frac{1}{2}} = 1$$

2. *Loi marginal de* X

   $$\mathbb{P}(X = k) = \sum_{j=0}^{+\infty} \frac{a}{2^{k+1} j!}$$

   $\forall k \in \mathbb{N}$, a constant et 2^{k+1} aussi donc

   $$\mathbb{P}(X = k) = \frac{a}{2^{k+1}} \sum_{j=0}^{+\infty} \frac{1}{j!} = \frac{a \cdot e}{2^{k+1}} = \frac{1}{2^{k+1}}$$

Loi marginal de Y

$$\mathbb{P}(Y = j) = \sum_{k=0}^{\infty} \frac{a}{2^{k+1} j!} = \frac{a}{j!} \sum_{k=0}^{\infty} \frac{1}{2^{k+1}}$$

$$\mathbb{P}(Y = j) = \frac{a}{j!} \frac{1}{2} \sum_{k=0}^{\infty} (1/2)^k$$

$$= \frac{a}{j!} \frac{1}{2} \frac{1}{1 - \frac{1}{2}} = \frac{a}{j!} = \frac{e^{-1}}{j!}$$

$$\mathbb{P}(Y = j) = \frac{e^{-1}}{j!} \quad \forall j \in \mathbb{N}$$

3. Indépendance ?

$\forall (j,k) \in \mathbb{N}^2, \mathbb{P}((X=k) \cap (Y=j)) \overset{?}{=} \mathbb{P}(X=k)\mathbb{P}(Y=j)$
$\mathbb{P}((X=k) \cap (Y=j)) = \frac{a}{2^{k+1} j!} = \frac{1}{2^{k+1}} \frac{a}{j!}$ avec $a = e^{-1}$ donc oui

4. $Cov(X,Y) = E(XY) - E(X)E(Y) = 0$ car X et Y sont indépendants.
   Rq $\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = 0$

5) Statistique multidimensionnelle

a) Tableau des données

Les observations de p variables sur n individus sont rassemblées en une matrice X à n lignes et p colonnes

$$X = \begin{pmatrix} X_1^{(1)} & \cdots & \vdots & \cdots \\ X_2^{(1)} & \cdots & \vdots & \cdots \\ X_i^{(1)} & \cdots & X_i^{(j)} & \cdots \\ \vdots & \cdots & \vdots & \cdots \\ X_n^{(1)} & \cdots & X_n^{(j)} & \cdots \end{pmatrix}$$

$$X^{(j)} = \begin{pmatrix} X_1^{(j)} \\ X_2^{(j)} \\ \vdots \\ X_n^{(j)} \end{pmatrix}$$

$i^{\text{ème}}$ individu : ${}^t e^i = (X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(p)})$

Ligne numéro i de X :

$$l_i = \begin{pmatrix} X_i^{(1)} \\ \vdots \\ X_i^{(p)} \end{pmatrix}$$

b) Matrice des poids

On associe à chaque individu un poids $p_i \geq 0$ (p_i : de choisir l'individu numero i)
$\sum_{i=1}^n p_i = 1$

$$D = \begin{pmatrix} p_1 & 0 & \cdots & \cdots & 0 \\ 0 & P_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & p_n \end{pmatrix}$$

matrice de poids

Si $P_i = \frac{1}{n}$
$D = \frac{1}{n} I_n$ (avec I_n la matrice identite)

c) Centre de gravité

Vecteur g de moyenne arithmétique de chaque variable :
${}^t g = (\overline{X^{(1)}}, \overline{X^{(2)}}, \ldots, \overline{X^{(p)}})$

Moyenne de $X^{(j)}$:
$\overline{X^{(j)}} = \sum_{i=1}^{n} P_i X_i^{(j)}$

Le tableau des données centrées est Y:
$\forall (i,j), y_i^{(j)} = X_i^{(j)} - \overline{(X^{(j)})}$

d) Matrice variance-covariance et matrice de corrélation

Déf. : On appelle matrice de variance-covariance la matrice V tq:
$V_{ij} = Cov(X^{(i)}, X^{(j)})$
$V_{ii} = V(X^{(i)})$ variance de $X^{(i)}$
$V(X^{(i)}) = \sum_{j=1}^{n} P_j (y_j^{(i)})^2$

Une matrice de variance-covariance est toujours symétrique, ie. ${}^t V = V$.

$\sigma(X^{(i)}) = \sqrt{\sum_{j=1}^{n} P_j (y_j^{(i)})^2}$ écart-type de $X^{(i)}$

On note $D_{1/S}$ la matrice diagonale des inverses des écarts-types

$$D_{1/S} = \begin{pmatrix} 1/S_1 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 1/S_p \end{pmatrix}$$

$S_j = \sigma(X^{(j)})$
$Z = Y D_{1/S}$

Z : matrice des données **centrées** et **réduites**

$$Z_i^{(j)} = \frac{X_i^j - \bar{X}^{(j)}}{S_j} = \frac{y_i^{(j)}}{S_j}$$

La matrice de corrélation
$R = D_{1/S} V D_{1/S}$
$= \underbrace{D_{1/S}\,{}^t Y}_{Z} D \underbrace{D_{1/S} Y}_{Z}$
$R = {}^t Z D Z$ symétrique

6) Espace des individus et espace des variables

Chaque individu etant un vecteur defini par p coordonnees.
On note F l'espace des individus de dimension p. Les n individus forment un nuage de points dans F et g en est le centre de gravite.

La distance entre 2 individus l_i et l_j est définie par:
$\langle e_i, e_j\rangle = {}^t e_i M e_j$
où M : matrice symétrique definie positive.

$({}^t M = M \langle M_u, u\rangle > 0, \forall u \neq 0)$
$||u|| = \sqrt{\langle u, u\rangle} = \sqrt{{}^t_u M_u}$
$d^2(e_i, e_j) = ||e_i - e_j||^2 = {}^t(e_i - e_j)M(e_i - e_j)$
$d(e_i, e_j) = \sqrt{{}^t(e_i - e_j)M(e_i - e_j)}$
Si $M = I$ on retrouve le produit scalaire canonique
Si $M = D_{1/S^2}$ ce qui revient à diviser chaque caractère par son écart-type

Définition :
On appelle inertie totale du nuage de points la **moyenne pondérée** des **carrés** des **distances des points** au **centre de gravité**.
$I_g = \sum_{i=1}^{n} p_i \, {}^t(e_i - g)M(e_i - g)$
$I_g = \sum_{i=1}^{n} p_i ||e_i - g||^2$

7) Espace des variables

$$X^{(j)} = \begin{pmatrix} X_1^{(j)} \\ \vdots \\ X_n^{(j)} \end{pmatrix} \text{Variable numéro } j$$

Soit E l'espace des variables dim $E = n$

Pour étudier la proximité des variables il faut ?préciser? cet espace d'une métrique M
ici $M = D$ (matrice des poids)
$\langle Y^{(j)}, Y^{(k)}\rangle = {}^t Y^{(j)} D Y^{(k)}$

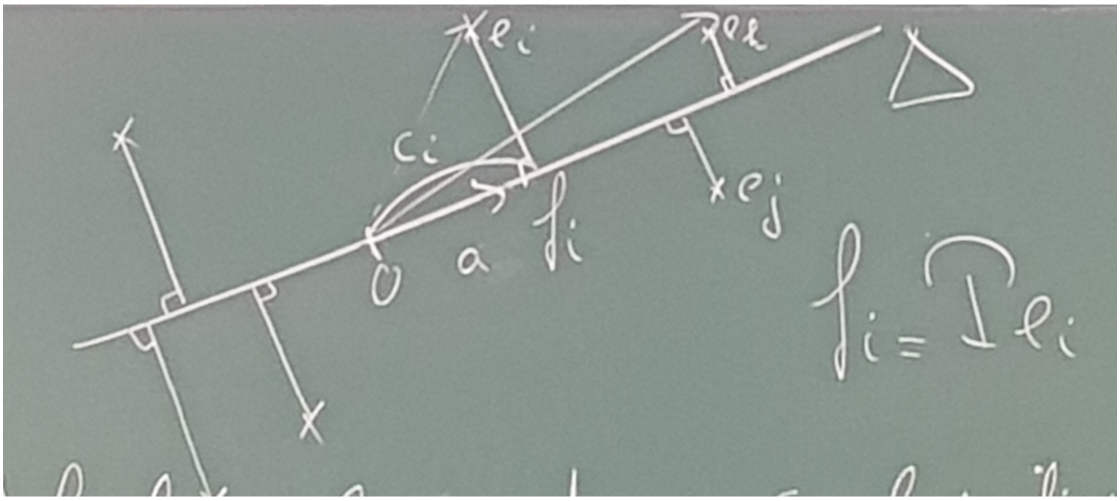
$Cov(X^{(j)}, X^{(k)}) = \langle Y^{(j)}, Y^{(k)}\rangle$
$||Y^{(j)}|| = \sqrt{{}^t Y^{(j)} D Y^{(j)}} = S_j$

$$\rho(X^{(j)}, X^{(k)}) = \frac{Cov(X^{(j)}, X^{(k)})}{S_j S_k} = \frac{\langle Y^{(j)}, Y^{(k)}\rangle}{||Y^{(j)}||||Y^{(k)}||}$$

$\rho(X^{(j)}, X^{(k)}) = \frac{\langle Y^{(j)}, Y^{(k)}\rangle}{||Y^{(j)}||||Y^{(k)}||} = \text{Cos}\Theta_{j,k}$
$|\rho(X^{(j)}, X^{(k)})| \leq 1$

A une variable $X^{(j)}$ on peut associer un axe de l'espace des individus F
et un vecteur de l'espace des variables.
On peut également déduire de $X^{(1)}, X^{(2)}, \cdots, X^{(p)}$ de nouvelles variables par combinaison linéaire,
ce qui revient à projeter les individus sur de nouveaux axes de F.
Soit Δ un axe de l'espace des individus, engendré par un vecteur unitaire a (M normé à 1; $||a|| = {}^t a M a = 1$ M métrique).

cf : projection des individus sur un axe (Projection M_orthogonale)



La liste des coordonnées C_i des individus sur Δ forme une nouvelle variable (variable artificielle)

$$C = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_i \\ \vdots \\ C_n \end{pmatrix}$$

$C_i = \langle l_i, a \rangle = {}^t e_i M a$
$C = XMa$

$$C =^t e_i \underbrace{\begin{pmatrix} X^{(1)} & \dots & X^p \\ \vdots & & \\ X^{(1)} & \dots & X^p \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} x & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & x \end{pmatrix}}_{M} \underbrace{\begin{pmatrix} a_i \\ \vdots \end{pmatrix}}_{a}$$

Mat S.D.P (on suppose que les variables $X^{(j)}$ sont centrés $Y = X$)

$$C = X\underbrace{Ma}_{u} = Xu = \sum_{j=1}^{p} u_j X^{(j)}$$

$u = Ma$

$u = Ma$ (Si $M = I \Rightarrow u = a$)
${}^t aMa = 1$
${}^t uMu = {}^t a\, {}^t MM^{-1}Ma$
$= {}^t a\underbrace{MM^{-1}}_{I}Ma$
$= {}^t aMa = 1$
$\Rightarrow$ $\boxed{{}^t uM^{-1}u = 1}$ on dit que u et M^{-1} normé à 1

u : facteur
$V(C) = {}^t CDC$
$= {}^t(Xu)D(Xu)$
$= {}^t u\, \underbrace{{}^t XDX}_{V}u$
$V(C) = {}^t uVu$

Remarque :
si u est un vecteur propre de V
$Vu = \lambda u$ (λ valeure propre de V)
$V(c) = {}^t u\lambda u = \lambda\, {}^t uu = \lambda$
si $(M = I \Rightarrow ||u|| = 1)$
$\boxed{V(c) = \lambda}$

${}^t uMu$ correspond à la norme M de U, au carré
en effet, $\langle u, v\rangle = {}^t uMv$

8) Analyse des Composantes principales

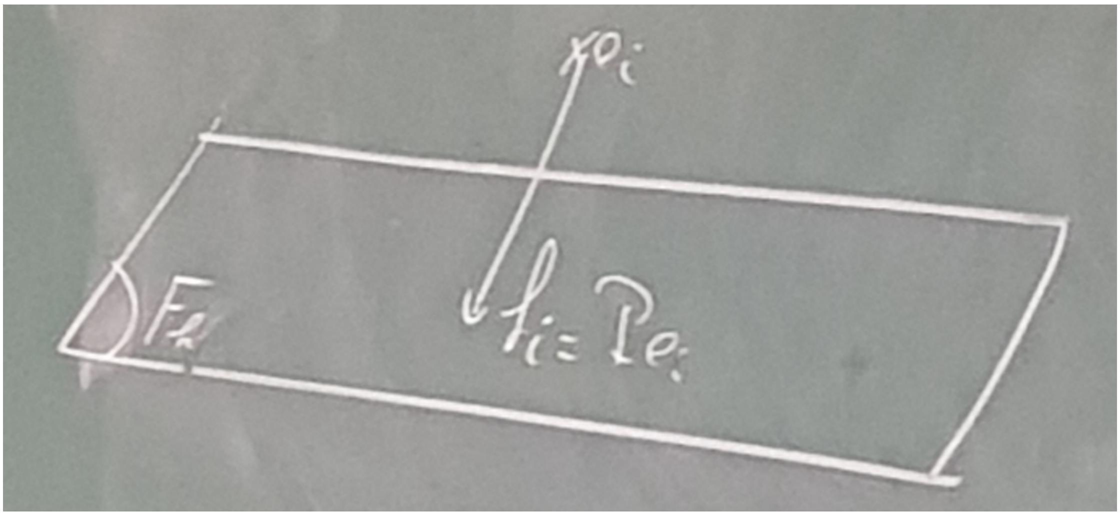
Le principe de la méthode est d'obtenir une **représentation approchée** du **nuage** des n individus dans un **S.E.V de dimension faible**. Ceci s'effectue par projection.

$\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i = \alpha$
(produit scalaire canonique)
M symétrique définie positif
$\langle u, v\rangle = {}^t uMv$

$$||u|| = \sqrt{\$\langle u, u\rangle} = \sqrt{{}^t uMu}$$

Le choix de l'espace de projection s'effectue selon le critère suivant qui revient à déformer le moins possible les distances en projections.
Il faut que l'inertie du nuage projeté sur le S.E.V F_k $\underbrace{(dimF_k = k)}_{\text{h faible par rapport à p}}$ soit maximale

P: Projection M_orthogonale sur F_k
$P^2 = P$ et ${}^tPM = MP$ (M métrique)



Le nuage projeté sur F_k est associé au tableau $X\,{}^tP$
Matrice de Var-Cov de nuage projeté

> On applique P à la transposée de X: $X\,{}^tP = {}^t(P\,{}^tX)$

${}^t(X\,{}^tP)D(X\,{}^tP)$
$P\underbrace{{}^tXdX}_{V}{}^tP = \boxed{PV\,{}^tP}$

> **Remarque**
>
> $(1)\text{Trace}(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1} \lambda_i$ (λ_i valeurs propres de A)
> $(2)\text{Trace}(AB) = \text{Trace}(BA)$
> $(3)I$(inertie) $= \text{Trace}(VM) = \text{Trace}(MV)$

L'inertie du nuage projeté d'après (3) :
$I_{NP} = Tr(PV\,{}^tPM)$
$Tr(\underbrace{P}_{A}\underbrace{VMP}_{B})$ car ${}^tPM = MP$
$Tr(XMP^2) = Tr(VMP)$

L'inertie du nuage projeté est : $\boxed{I_{NP} = Tr(VMP)}$

Le problème est donc de trouver P projection M_orhtogonale de rang k maximisant la trace de VMP
ce qui déterminera F_k

Th: Soit F_k un S.E.V portant l'inertie maximale, alors le S.E.V de dimension $k+1$ portant l'inertie maximale est la somme directe de F_k et des S.E.V de dimension 1 M orthogonal à F_k portant l'inertie maximale

$F_{k+1} = F_k \oplus \vec{b}\mathbb{R}$ (Somme directe)
$F_k \cap \vec{b}\mathbb{R} = \{\vec{0}\}$ \
et \
dim F_k+ dim $\vec{b}\mathbb{R} = k + 1$ \

$\vec{b}\mathbb{R} = \{\lambda\vec{b}, \lambda \in \mathbb{R}\}$

Rappels (Formes quadratiques et projection sur une droite)

$g : \mathbb{R}^n \longrightarrow \mathbb{R}$

$$\frac{dg}{du} = \begin{pmatrix} \frac{\partial g}{\partial u_1} \\ \vdots \\ \frac{\partial g}{\partial u_p} \end{pmatrix} \text{gradient de } g$$

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}$$

${}^t a u = \sum_{i=1}^{p} a_i u_i = g(u) (a_i \in \mathbb{R})$

$\frac{\partial g}{\partial u_i} = a_i V_i$

$$\frac{\partial ({}^t a u)}{\partial u} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

$A \in \mathcal{M}_p(\mathbb{R})$

$\boxed{\frac{d}{du}(\underbrace{{}^t u A u}_{\text{forme quadratique si A est symétrique: } {}^t A = A}) = Au + {}^t A u}$

$\frac{d}{du}({}^t u A u) = 2Au$

Projecteur M_orthogonale sur la droite Δ engendrée par a

$$\boxed{P = \frac{a\, {}^t a M}{{}^t a M a}}$$

$$a\, {}^t a = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} \begin{pmatrix} a_1 & \cdots & a_p \end{pmatrix}$$

Inertie du nuage projeté :

$I_{NP} = Tr(VMP)$

$= Tr(VM \frac{a\, {}^t a M}{{}^t a M a})$

$= \frac{1}{{}^t a M a} Tr(VMa\, {}^t a M)$ avec $Tr(AB) = Tr(BA)$

$= \frac{1}{{}^t a M a} Tr({}^t a M V M a)$

$$\boxed{I_{NP} = \frac{{}^t aMVMa}{{}^t aMa}} \frac{d}{da}(\frac{{}^t aMVMa}{{}^t aMa}) = 0$$

$\frac{({}^t aMa)2MVMa - ({}^t aMVMa)2Ma}{({}^t aMa)^2}$

$({}^t aMa)MVMa = ({}^t aMVMa)Ma$

$MVMa = (\frac{({}^t aMVMa)}{({}^t aMa)})Ma$

En multipliant par M^{-1}

$$\rightarrow \boxed{VMa = (\frac{({}^t aMVMa)}{({}^t aMa)})a}$$

$$\boxed{\begin{array}{r} VMa = \lambda a \\ \text{avec} \\ \lambda = \dfrac{({}^t aMVMa)}{({}^t aMa)} \end{array}}$$

a est un vecteur propre de VM associé à λ

Le S.E.V F_k de dimension k est engendré par les k valeurs propres de VM associés aux k plus grandes valeurs propres

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq \cdots$

$$\boxed{\text{\% d'inertie : } \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \geq 80\%}$$

Composantes principales

$$\boxed{C^{(i)} = Yu^{(i)}} i = 1 - - - k$$

$u^{(i)} = Ma^{(i)}$ facteur axes factoriels si $M = I => u^{(i)} = a^{(i)} \forall i$

Corrélations entre $C^{(i)}$ et $X^{(i)}$

$$\boxed{\rho(X^{(i)}, C^{(j)}) = \frac{Cov(X^{(i)}, C^{(j)})}{\sigma(X^{(i)})\sigma(C^{(j)})}}$$

Exercice 1 (ACP)

$$\text{Soit } X = \begin{pmatrix} 3 & 8/5 & 5 \\ 4 & 2 & 6 \\ 3 & 9/5 & 6 \end{pmatrix}$$

$p_i = \frac{1}{3} \forall i$
$M = I_3$ (Métrique de l'espace des individus)

1. Déterminer le centre de gravité
2. Calculer Y (données centrées)
3. Calculer la matrice V
4. Diagonaliser $V \; (M = I_3)$
5. Déterminer les facteurs primitives
6. calculer les cœfficients de corrélation

1. Déterminer le centre de gravité de ce nuage:

$$\bar{X}^{(j)} = \sum_{i=1}^{3} p_i x_i^{(j)} = \frac{1}{3} \sum_{i=1}^{3} x_i^{(j)}$$

$$^t g = \begin{pmatrix} \bar{X}^{(1)} & \bar{X}^{(2)} & \bar{X}^{(3)} \end{pmatrix}$$

$\bar{X}^{(1)} = \frac{10}{3}, \bar{X}^{(2)} = \frac{9}{5}, \bar{X}^{(3)} = \frac{17}{3}$
$^t g = (\frac{10}{3}, \frac{9}{5}, \frac{17}{3})$

$\bar{X}^{(j)} = \sum_{i=1}^{3} p_i x_i^j = \frac{1}{3} \sum_{i=1}^{3} X_i^j$

2) La matrice de données centrées
$Y : y_i^{(j)} = X_i^{(j)} - \bar{X}^{(j)}$

$$Y = \begin{pmatrix} -\frac{1}{3} & -\frac{1}{5} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{5} & \frac{1}{3} \\ -\frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}$$

3. Matrice de Var-Covariance
   $V = {}^tYDY$
   $D = \frac{1}{3} I_3$

$$D = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \text{matrice de poids}$$

   $V = \frac{1}{3} \, {}^tYY$
   $V_{ii} = \mathrm{Var}(X^{(i)}) => \sigma(X^{(i)}) = \sqrt{\mathrm{Var}(X^{(i)})} \forall i$

$$V = \frac{1}{3}\begin{pmatrix} -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{5} & \frac{1}{5} & 0 \\ -\frac{2}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} -\frac{1}{3} & -\frac{1}{5} & -\frac{2}{3} \\ \frac{2}{3} & \frac{1}{5} & \frac{1}{3} \\ -\frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{9} & \frac{1}{15} & \frac{1}{9} \\ \frac{1}{15} & \frac{2}{75} & \frac{1}{15} \\ \frac{1}{9} & \frac{1}{15} & \frac{2}{9} \end{pmatrix}$$

4. On doit diagonaliser $MVM = I_3$ donc on va diagonaliser V
   Le polynôme caractéristique de V
   $P_V(\lambda) = det(V - \lambda I_3)$
   $P_V(\lambda) = 0 \implies$ les racines de cette équation sont les valeurs propres de cette matrice

$$P_V(\lambda) = \begin{vmatrix} \frac{2}{9} - \lambda & \frac{1}{15} & \frac{1}{9} \\ \frac{1}{15} & \frac{2}{75} - \lambda & \frac{1}{15} \\ \frac{1}{9} & \frac{1}{15} & \frac{2}{9} - \lambda \end{vmatrix}$$

$L_1 \leftarrow L_1 - L_3$

$$P_V(\lambda) = \begin{vmatrix} \frac{1}{9} - \lambda & 0 & \lambda - \frac{1}{9} \\ \frac{1}{15} & \frac{2}{75} - \lambda & \frac{1}{15} \\ \frac{1}{9} & \frac{1}{15} & \frac{2}{9} - \lambda \end{vmatrix}$$

$C_3 \leftarrow C_3 + C_1$

$$P_V(\lambda) = \begin{vmatrix} \frac{1}{9} - \lambda & 0 & 0 \\ \frac{1}{15} & \frac{2}{75} - \lambda & \frac{2}{15} \\ \frac{1}{9} & \frac{1}{15} & \frac{1}{3} - \lambda \end{vmatrix} = (\frac{1}{9} - \lambda)((\frac{2}{75} - \lambda)(\frac{1}{3} - \lambda) - \frac{2}{15^2})$$

$= (\frac{1}{9} - \lambda)(\lambda^2 - (\frac{2}{75} + \frac{1}{3})\lambda)$
$= (\frac{1}{9} - \lambda)(\lambda^2 - \frac{27}{75}\lambda) = 0$
$= (\frac{1}{9} - \lambda)\lambda(\lambda - \frac{9}{25}) = 0$
$=> \lambda_1 = \frac{9}{25}, \lambda_2 = \frac{1}{9}, \lambda_3 = 0$

Remarque : $\boxed{\mathrm{Trace}(V) = \lambda_1 + \lambda_2 + \lambda_3}$

Valeur propre	$\lambda_1 = 0.36$	$\lambda_2 = 0.11$	$\lambda_3 = 0$
% d'inertie	$\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} = 76\%$	$\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3} = 24\%$	0

Le pourcentage d'inertie apporté par le plan factoriel $\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3} = 100\%$

> On ordonne les valeurs propres dans l'ordre décroissant, et on en prend tant que l'on a pas atteint 80%.
> par exemple, dans notre cas, 76% < 80% donc on rajoute λ_2, puis comme 100% > 80%, on s'arrete ici.

Sous-espace propre :

$E_{\lambda_1} = E_{\frac{9}{25}} = Ker(V - \frac{9}{25}I_3)$

$$\forall w = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in E_{\frac{9}{25}} \Leftrightarrow (V - \frac{9}{25}I_3)\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

$$\begin{cases} (1) : (\frac{2}{9} - \frac{9}{25})x + \frac{1}{15}y + \frac{1}{9}z = 0 \\ (2) : \frac{1}{15}x + (\frac{2}{75} - \frac{9}{25})y + \frac{1}{15}z = 0 \\ (3) : \frac{1}{9}x + \frac{1}{15}y + (\frac{2}{9} - \frac{9}{25})z = 0 \end{cases}$$

$(1) - (3) : (\frac{1}{9} - \frac{9}{25})x + (\frac{9}{25} - \frac{1}{9})z = 0$

$x = z$ et $y = \frac{2}{5}x$

$$E_{\frac{9}{25}} = \text{Vect}\begin{pmatrix} 5 \\ 2 \\ 5 \end{pmatrix} \, droite$$

$$||w_1|| = \sqrt{(5)^2 + (2)^2 + 5^2} = \sqrt{54}$$

$$\boxed{u^{(1)} = \frac{1}{\sqrt{54}}\begin{pmatrix} 5 \\ 2 \\ 5 \end{pmatrix} \text{normé } 1^{\text{er}} \text{ facteur}}$$

$$E_{1/9} = Ker(V - \frac{1}{9}I_3)$$

$$\begin{cases} (1) : \frac{1}{9}x + \frac{1}{15}y + \frac{1}{9}z = 0 \\ (2) : \frac{1}{15}x + (\frac{2}{75} - \frac{1}{9})y + \frac{1}{15}z = 0 \\ (3) : \text{identique à } (1) \end{cases}$$

(1) $-$ (2) $\Rightarrow y = 0$

(1) $\Rightarrow x = -z$

$$\boxed{E_{\frac{1}{9}} = \text{Vect}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}}$$

$$\boxed{u^{(2)} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}} 2^e \text{ facteur}$$

$(u^{(1)}, u^{(2)})$ base orthonormées

5. Composantes principales
   $\boxed{C^{(i)} = Yu^{(i)}} i = 1, 2$

$$u^{(1)} = \begin{pmatrix} 0,68 \\ 0,27 \\ 0,68 \end{pmatrix}$$

$$u^{(2)} = \begin{pmatrix} 0,71 \\ 0 \\ -0,71 \end{pmatrix}$$

$$C^{(1)} = \begin{pmatrix} -0,73 \\ 0,73 \\ 0 \end{pmatrix}$$

$$C^{(2)} = \begin{pmatrix} 0,24 \\ 0,24 \\ -0,47 \end{pmatrix}$$

6. Cœfficient de corrélation
   $\rho(X^{(i)}, C^{(j)}) = \frac{\text{Cov}(X^{(i)}, C^{(j)})}{\sigma(X^{(i)})\sigma(C^{(j)})}$
   $Cov(X^{(i)}, C^{(j)}) = \langle y^{(i)}, C^{(j)} \rangle$
   $Cov(X^{(1)}, C^{(1)}) = \langle y^{(1)}, C^{(1)} \rangle$
   $= {}^t y^{(1)} DC^{(1)}$
   $M = D = \frac{1}{3} I$
   $= \frac{1}{3}((-\frac{1}{3})(-0,73) + (\frac{2}{3})(0,73))$

$\sigma(X^{(1)}) = \sqrt{\text{Var} X^{(1)}} = 0,47$
$\sigma(C^{(1)}) = ||C^{(1)}|| = \sqrt{(\frac{1}{3})((-0,73)^2 + (0,73)^2)} = 0.59 \simeq 0.60$
$\rho(X^{(1)}, C^{(1)}) = \frac{\text{Cov}(X^{(1)}, C^{(1)})}{\sigma(X^{(1)})\sigma(C^{(1)})} = \boxed{0,87}$

	$C^{(1)}$	$C^{(2)}$
$X^{(1)}$	0,87	0,5
$X^{(2)}$	1	0
$X^{(3)}$	0,87	-0,5

Exercice 2 (ACP)
Même questions que l'exercice 1

$$X = \begin{pmatrix} 16 & 2 & 0 \\ 8 & 12 & 10 \\ 12 & 16 & 14 \\ 20 & 8 & 14 \\ 16 & 4 & 10 \\ 0 & 6 & 12 \end{pmatrix}$$

$p_i = \frac{1}{6} \forall i$

$M = I_3$ dans l'espace des individus

1. $\bar{X^{(1)}} = \frac{72}{6} = 12$
   $\bar{X^{(2)}} = \frac{48}{6} = 8$
   $\bar{X^{(3)}} = \frac{60}{6} = 10$
   ${}^t g = (12, 8, 10)$
2.

$$Y = \begin{pmatrix} 4 & -6 & -10 \\ -4 & 4 & 0 \\ 0 & 8 & 4 \\ 8 & 0 & 4 \\ 4 & -4 & 0 \\ -12 & -2 & 2 \end{pmatrix}$$

3. $V = \frac{1}{6}\,{}^tYY$

$$V = \frac{1}{6} \begin{pmatrix} 4 & -4 & 0 & 8 & 4 & -12 \\ -6 & 4 & 8 & 0 & -4 & -2 \\ -10 & 0 & 4 & 4 & 0 & 2 \end{pmatrix} \begin{pmatrix} 4 & -6 & -10 \\ -4 & 4 & 0 \\ 0 & 8 & 4 \\ 8 & 0 & 4 \\ 4 & -4 & 0 \\ -12 & -2 & 2 \end{pmatrix} = \begin{pmatrix} \frac{128}{3} & -\frac{16}{3} & \cdots \\ -\frac{16}{3} & \frac{68}{3} & \cdots \\ -\frac{16}{3} & \frac{44}{3} & \cdots \end{pmatrix}$$

4. Diago de V

$$P_V(\lambda) = \begin{vmatrix} \frac{128}{3} - \lambda & -\frac{16}{3} & -\frac{16}{3} \\ -\frac{16}{3} & \frac{68}{3} - \lambda & \frac{44}{3} \\ -\frac{16}{3} & \frac{44}{3} & \frac{68}{3} - \lambda \end{vmatrix}$$

$$C_1 \leftarrow C_1 + C_2 + C_3$$

$$P_V(\lambda) = \begin{vmatrix} 32-\lambda & -\frac{16}{3} & -\frac{16}{3} \\ 32-\lambda & \frac{68}{3}-\lambda & \frac{44}{3} \\ 32-\lambda & \frac{44}{3} & \frac{68}{3}-\lambda \end{vmatrix} = (32-\lambda)$$

$L_2 \leftarrow L_2 - L_1$
$L_3 \leftarrow L_3 - L_1$

$$= (32-\lambda)\begin{vmatrix} 28-\lambda & 20 \\ 20 & 28-\lambda \end{vmatrix}$$

$P_V(\lambda) = (32-\lambda)((28-\lambda)^2 - 20^2)$
$= (32-\lambda)(28-\lambda-20)(28-\lambda+20)$
$\boxed{P_V(\lambda) = (32-\lambda)(8-\lambda)(48-\lambda)}$
$\boxed{\lambda_1 = 48 > \lambda_2 = 32 > \lambda_3 = 8}$

Valeur propres	λ_1	λ_2	λ_3
pourcentage d'inertie	$\frac{48}{88} = 0,54$	$\frac{32}{88} = 0,36$	$\frac{8}{88} = 0,09$

$$\boxed{\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3} = \frac{80}{88} = 90\%}$$

Axes factoriels :

$$u^{(1)} = \begin{pmatrix} \frac{\sqrt{6}}{3} \\ -\frac{\sqrt{6}}{6} \\ -\frac{\sqrt{6}}{6} \end{pmatrix}, u^{(2)} = \begin{pmatrix} \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{3} \end{pmatrix} \text{base orthonormée}$$

Composantes principales :

$$C^{(1)} = Yu^{(1)} = \begin{pmatrix} 4\sqrt{6} \\ -2\sqrt{6} \\ -2\sqrt{6} \\ 2\sqrt{6} \\ 2\sqrt{6} \\ -4\sqrt{6} \end{pmatrix}$$

$$C^{(2)} = Yu^{(2)} = \begin{pmatrix} -4\sqrt{3} \\ 0 \\ 4\sqrt{3} \\ 4\sqrt{3} \\ 0 \\ -4\sqrt{3} \end{pmatrix}$$

6. cœfficient de covariance

	$C^{(1)}$	$C^{(2)}$
$X^{(1)}$	0,87	0,5
$X^{(2)}$	-0,59	0,69
$X^{(3)}$	-0,59	0,69

$\rho(X^{(i)}, C^{(j)}) = \frac{Cov(X^{(i)}, C^{(j)})}{\sigma(x^{(i)})\sigma(C^{(j)})}$

Méthode ACP

Méthode ACP = Analyse Composante Principale
On a une matrice
P variables (colonnes)
n Individus (lignes) : population de pressions, vitesses, objets, habitants…
Un élément de la matrice $i, j =$ valeur de P_j pour l'individu i

Inertie (http://jybaudot.fr/Stats/inertie.html) : moyenne des distances au carré des points au **barycentre** (https://fr.wikipedia.org/wiki/Barycentre).