Analyse des données

I) Description bidimensuelles et mesure des liaisons entre X et Y

- Loi conjointe (Loi du couple (X, Y))
- Lois marginales
- Lois conditionnelles
- Indépendance
-

II Description multidimensionnelle

- Tableau des données

$$X = \begin{array}{cccc} X^{(1)} & & \cdots & X^{(p)} \\ \left(\begin{matrix} \vert \\ \vert \\ \vert \end{matrix}\right. & \begin{matrix} \vert \\ \vert \\ \vert \end{matrix} & \begin{matrix} \vert \\ \vert \\ \vert \end{matrix} & \left.\begin{matrix} \vert \\ \vert \\ \vert \end{matrix}\right) \end{array}$$

- Matrice des poids
- Centre de gravité du nuage des pts formé par les individus
- Matrice de Var-Covariance
- Matrice de Corrélation
- Projection des individus

Analyse des données

I) Description bidimensionnelle

1) Couple de variables aléatoires (X, Y)

Soient X et Y 2 variables aléatoires discrètes définies sur un même espace probabilisé $(\Omega, \mathcal{C}, P)$

Univers — Tribu — Prob

$X(\Omega) = \{ x_i \mid i \in I \}$ Valeurs de X

$Y(\Omega) = \{ y_j \mid j \in J \}$ —— de Y

<u>Déf</u> on appelle la loi de (X,Y) (Loi conjointe) l'ensemble des couples $((x_i, y_j), p_{ij})$ où $x_i \in X(\Omega)$, $y_j \in Y(\Omega)$

$$\boxed{p_{ij} = P\left((X = x_i) \cap (Y = y_j)\right)}$$

<u>Rq</u> Si $I = [\![1, r]\!]$ et $J = [\![1, s]\!]$

$$
\begin{array}{c|ccccc}
X \backslash Y & y_1 & \cdots & y_j & \cdots & y_s \\
\hline
x_1 & p_{11} & \cdots & p_{1j} & \cdots & p_{1s} \\
\vdots & \vdots & & \vdots & & \vdots \\
x_i & p_{i1} & \cdots & p_{ij} & \cdots & \vdots \\
\vdots & \vdots & & \vdots & & \\
x_r & p_{r1} & \cdots & p_{rj} & \cdots & p_{rs}
\end{array}
\qquad
\boxed{\begin{array}{l} p_{ij} \geq 0 \\ \displaystyle\sum_{i,j} p_{ij} = 1 \end{array}}
$$

$$
p_{ij} = \mathbb{P}\big((X = x_i) \cap (Y = y_j)\big)
$$

1) Lois marginales

Def Les variables X et Y sont appelées variables marginales du couple (X,Y)

$$P_{i.} = \mathbb{P}(X = x_i) = \sum_{j \in J} P_{ij}$$

$$P_{.j} = \mathbb{P}(Y = y_j) = \sum_{i \in I} P_{ij}$$

Exemple

$X \backslash Y$	1	2	3	4	$P_{i.}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
2	0	$\frac{2}{16}$	$1/16$	$\frac{1}{16}$	$1/4$
3	0	0	$3/16$	$1/16$	$1/4$
4	0	0	0	$4/16$	$1/4$
$P_{.j}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$	1.

(Loi marginale de X)

Loi marginale de Y

3) Loi conditionnelle

La loi conditionnelle de $X = x_i$ sachant que $Y = y_j$

$$\boxed{\mathbb{P}\left(X = x_i / Y = y_j\right) = \frac{P_{ij}}{P_{\cdot j}}}$$

de même

$$\boxed{\mathbb{P}\left(Y = y_j / X = x_i\right) = \frac{P_{ij}}{P_{i \cdot}}}$$

Exemple (précédent)

$$P\left(X=1 / Y=3\right) = \frac{P_{13}}{P_{\cdot 3}} = \frac{1/16}{5/16} = \frac{1}{5}$$

x_i	1	2	3	4
$P\left(X=x_i / Y=3\right)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{5}$	0

Déf X et Y sont 2 variables indépendants ssi :

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

$$\boxed{P_{ij} = P_{i\cdot}\, P_{\cdot j}} \quad \forall i \in I,\ \forall j \in J$$

$$\Longleftrightarrow \boxed{\mathbb{P}(X = x_i / Y = y_j) = \mathbb{P}(X = x_i)}$$

3) Fonction de 2 variables X et Y

$g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ fonction définie sur l'ensemble des valeurs prises par (X, Y)

$$Z = g(X, Y)$$

$$g(x_i, y_j) = z_k \quad \forall x_i \in X(\Omega) \quad \forall y_j \in Y(\Omega)$$

$$Z(\Omega) = \{ z_k \;/\; k \in K \} \qquad K \subset \mathbb{N}$$

$$(Z = z_k) = \bigcup_{\substack{(i,j) \\ g(x_i, y_j) = z_k}} \big((X = x_i) \cap (Y = y_j) \big)$$

L'événement

$$\boxed{P(Z = z_k) = \sum_{\substack{(i,j) \\ g(x_i, y_j) = z_k}} P\big((X = x_i) \cap (Y = y_j) \big)}$$

En particulier

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

$$\mathbb{P}(X.Y=z) = \sum_{\substack{(x,y)/ \\ xy=z}} \mathbb{P}((X=x)\cap(Y=y))$$

$g(x,Y) = X+Y$

$g(x,Y) = X.Y$

Exemple

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

Déterminer la loi de la somme $S = X + Y$

$$P(S=5) = P((X=1)\cap(Y=4)) + P((X=2)\cap(Y=3)) + P((X=3)\cap(Y=2)) + P((X=4)\cap(Y=1)) = 1/8$$

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

Loi du produit

$$\Pi = X.Y$$

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

Espérance de $Z = g(X,Y)$

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

$$\boxed{E\big(g(X,Y)\big) = \sum_{i,j} g(x_i, y_j) \, P_{ij}}$$

Rq $$\boxed{E(X.Y) = \sum_{i,j} x_i y_j \, P_{ij}}$$

__Rq__ Si X et Y 2 v.a. indépendantes

alors $\boxed{E(X.Y) = E(X)E(Y)}$

__Exemple__

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

$$E(X.Y) = \sum_{i=0}^{1} \sum_{j=0}^{2} (i.j) P_{ij} = \frac{1}{4} + \frac{2}{6}$$

$$= \frac{7}{12}$$

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

$$E(X) = \sum_i x_i P(X = x_i)$$

$$= \boxed{\frac{7}{10}}$$

$$E(Y) = \sum y_j P(Y = y_j) = \frac{1}{2} + \frac{1}{3}$$

$$= \boxed{\frac{5}{6}}$$

on a bien $E(X.Y) = E(X)E(Y)$

mais $P((X=0)\cap(Y=2)) = 0$

$$\neq P(X=0).P(Y=2) = \frac{3}{60}$$

$$= \frac{1}{20}$$

$\Rightarrow$ X et Y ne sont pas indépendantes

Covariance de (X,Y)

Déf On appelle covariance de (X,Y)

$$\boxed{\mathrm{Cov}(X,Y) = E(X.Y) - E(X)E(Y)}$$

Déf On appelle coefficient de corrélation linéaire :

$$\boxed{\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}}$$

Rq

$$\rho(X,Y) = \frac{\langle X - E(X), Y - E(Y) \rangle}{\sigma_X \cdot \sigma_Y}$$

$$\sigma_X = \sqrt{V(X)} = \|X - E(X)\|$$

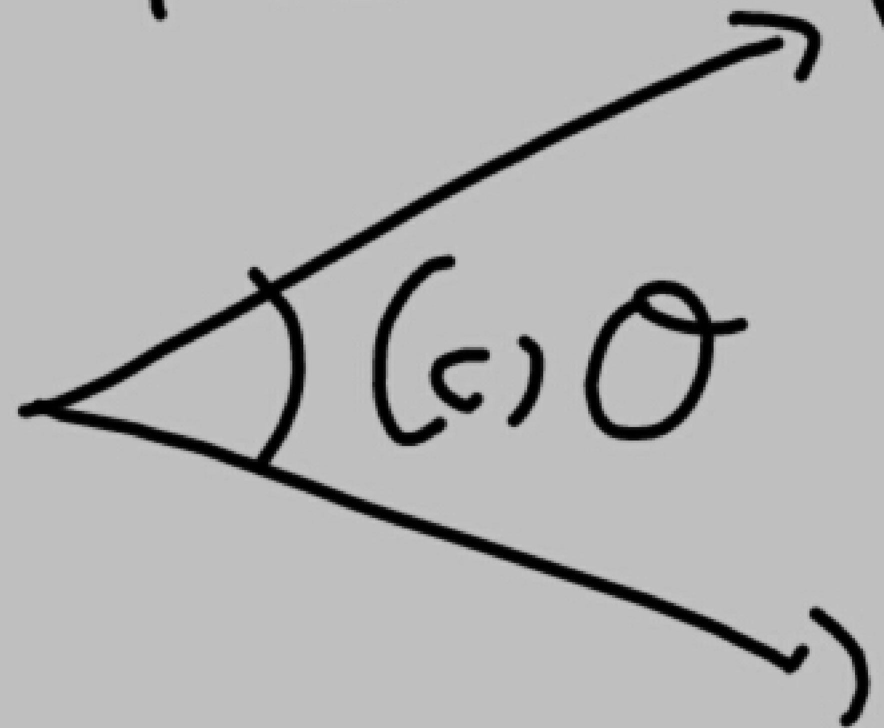
$$\sigma_Y = \sqrt{V(Y)} = \|Y - E(Y)\|$$

$\langle X - E(X), Y - E(Y) \rangle$

produit scalaire

$$\rho(X,Y) = \cos\underbrace{\widehat{\left(X - E(X), Y - E(Y)\right)}}_{\hat{\theta}}$$

$$|\rho| < 1$$



$|\rho| \leq 1$

si $\rho = 1$ forte corrélation linéaire entre X et Y

si $\rho = 0 \Rightarrow$ pas de corrélation

Ex1 Soit X = V.a

X_i	-2	-1	0	1	2
P_i	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$

1) Donner la loi conjointe

$\mathbb{P}((X = x_i) \cap (Y = y_j)) = 0$ si $y_j \neq x_i^2$

$\mathbb{P}((X = x_i) \cap (Y = x_i^2)) =$

2) Loi marginale de Y ?

3) Indépendance ?

4) $\mathrm{Cov}(X, Y)$

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

$X \backslash Y$	0	1	4	$p_{i\cdot}$
-2	0	0	1/6	
-1	0	1/4	0	
0	1/6	0	0	
1	0	1/4	0	
2	0	0	1/6	
$p_{\cdot j}$	$\frac{1}{6}$	1/2	1/3	

Analyse des données

I) Description bidimensionnelle et mesure de Corrélation

- Loi Conjointe de (X, Y)
- Lois marginales
- Lois Conditionnelles
- Covariance $Cov(X, Y)$ et Corrélation

II) Description multidimensionnelle

- Tableau des données

$$X = \begin{array}{c} \begin{array}{cccc} X^{(1)} & \dots & \dots & X^{(p)} \end{array} \\ \left(\begin{array}{cccc} | & | & | & | \end{array} \right) \end{array}$$

- Matrice des poids
- Matrice des données centrées
- Matrice de Var-Cov
- Matrice des corrélations
- Algo A.C.P (Analyse composantes principales)
- Projection

Analyse des données

I) Description bidimensionnelle

1) Loi du couple (X, Y)

Soient X et Y 2 variables aléatoires discrètes définies sur le même espace probabilisé : $(\Omega, \mathcal{C}, P)$

Univers — Tribu — Prob

$X(\Omega) = \{ x_i \,/\, i \in I \}$ valeurs de X

$Y(\Omega) = \{ y_j \,/\, j \in J \}$ ——— Y

<u>Déf</u> On appelle loi <u>conjointe</u> du couple (X, Y) l'ensemble des couples $((x_i, y_j), P_{ij})$ où $x_i \in X(\Omega)$, $y_j \in Y(\Omega)$ $\boxed{P_{ij} = P((X = x_i) \cap (Y = y_j))}$

Rq Si $I = [\![1, r]\!]$ et $J = [\![1, s]\!]$

$X \backslash Y$	y_1	$\cdots$	y_j	$\cdots$	y_s
x_1					
$\vdots$					
x_i			p_{ij}		
$\vdots$					
x_r					

$$p_{ij} = P\big((X = x_i) \cap (Y = y_j)\big)$$

$$\boxed{\begin{array}{l} p_{ij} \geqslant 0 \\ \displaystyle\sum_{i,j} p_{ij} = 1 \end{array}}$$

2) Lois marginales

<u>Déf</u> Les variables X et Y sont appelées variables marginales

$$P(X=x_i) = \sum_{j\in J} p_{ij} = \sum_{j\in J} P\big((X=x_i)\cap(Y=y_j)\big)$$

$$\boxed{P(X=x_i) = p_{i.} \quad \forall i}$$

$$\boxed{P(Y=y_j) = \sum_{i\in I} p_{ij} = p_{.j}}$$

Exemple

$X \backslash Y$	1	2	3	4	$P_{i\cdot}$ (Loi marginale X)
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
2	0	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$1/4$
3	0	0	$\frac{3}{16}$	$\frac{1}{16}$	$1/4$
4	0	0	0	$\frac{4}{16}$	$1/4$
Loi de Y $P_{\cdot j}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$	1

$$P(X=4) = \frac{1}{4}$$

$$P(Y=3) = \frac{5}{16}$$

3) Loi conditionnelles

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

$$P\left(X = x_i / Y = y_j\right) = \frac{P\left((X = x_i) \cap (Y = y_j)\right)}{P(Y = y_j)} = \boxed{\frac{p_{ij}}{p_{\cdot j}}}$$

$$\boxed{P\left(Y = y_j / X = x_i\right) = \frac{p_{ij}}{p_{i\cdot}}}$$

$$P\left(X=1 / Y=3\right) = \frac{1/16}{5/16} = \frac{1}{5}$$

Exemple

$$\mathbb{P}\left((X=1)/_{Y=3}\right) = \frac{\mathbb{P}\left((X=1)\cap(Y=3)\right)}{\mathbb{P}(Y=3)}$$

$$= \frac{1/16}{\frac{5}{16}} = \frac{1}{5}$$

x_i	1	2	3	4
$\mathbb{P}\left(X/_{Y=3}\right)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{5}$	0

Indépendance

X et Y sont indépendants, ssi $\mathbb{P}\left((X=x)\cap(Y=y)\right)$

$$= \mathbb{P}(X=x) \cdot \mathbb{P}(Y=y)$$

$$\Longleftrightarrow \boxed{P_{ij} = P_{i.} \times P_{.j}} \quad \forall\, (i,j) \in I \times J$$

$$\Longleftrightarrow \mathbb{P}\left(X=x \,/\, Y=y\right) = \mathbb{P}(X=x)$$

4) Loi d'une fonction de 2 variables

Soit $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie sur l'ensemble des valeurs prises par X et Y

$$Z = g(X, Y)$$

$$\underbrace{(Z = z_k)}_{\text{événement}} = \bigcup_{\substack{(i,j) \\ g(x_i, y_j) = z_k}} \left((X = x_i) \cap (Y = y_j) \right)$$

$$\boxed{P(Z = z_k) = \sum_{\substack{(i,j)/ \\ g(x_i, y_j) = z_k}} P\big((X = x_i) \cap (Y = y_j)\big)}$$

En particulier $g(X, Y) = X + Y = S$

$$P(S = s_k) = \sum_{\substack{(i,j)/ \\ x_i + y_j = s_k}} P\big((X = x_i) \cap (Y = y_j)\big)$$

Exemple précédent

$S = X + Y \qquad \Pi = X \cdot Y$

Loi de S

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

$$P(S = 5) = P_{1,4} + P_{23} + P_{32} + P_{41}$$

$$= \frac{2}{16} = \frac{1}{8}$$

Loi du produit $\mathbb{P} = X.Y$

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

Espérance de $Z = g(X,Y)$

Déf $$\boxed{E(Z) = \sum_{i,j} g(x_i, y_j)\, P_{ij}}$$

__Rq__ Si X et Y sont 2 variables indépendantes alors $\boxed{E(X.Y) = E(X)E(Y)}$

__Exemple__

$X \backslash Y$	0	1	2	$P_{i.}$ (Loi de X)
0	$\frac{1}{20}$ $\boxed{0}$	$1/4$ $\boxed{0}$	0 $\boxed{0}$	$\frac{3}{10}$
1	$\frac{17}{60}$ $\boxed{0}$	$1/4$ $\boxed{1}$	$1/6$ $\boxed{2}$	$\frac{7}{10}$
$P_{.j}$	$1/3$	$1/2$	$1/6$	1

$$E(X.Y) = \sum_{i=0}^{1} \sum_{j=0}^{2} i.j.P_{ij}$$

$$= \frac{1}{4} + \frac{2}{6} = \frac{1}{4} + \frac{1}{3} = \boxed{\frac{7}{12}}$$

$\boxed{E(X) = \frac{7}{10}}$, $E(Y) = 1/2 + \frac{1}{3} = \boxed{5/6}$

$$E(X.Y) = \frac{7}{12} = E(X)E(Y)$$

Mais $\mathbb{P}\big((X=0)\cap(Y=2)\big) = 0 \neq \mathbb{P}(X=0).\mathbb{P}(Y=2)$

$$\frac{3}{10}\cdot\frac{1}{6} = \frac{1}{20}$$

$\Longrightarrow$ X et Y ne sont pas indépendants

$$E(X) = \sum_{i\in I} x_i \mathbb{P}(X=x_i) = 0\times\frac{3}{10} + 1\times\frac{7}{10} = \frac{7}{10}$$

Covariance et coefficient de corrélation linéaire

Déf On appelle covariance du couple (X,Y)

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

et le coefficient de corrélation :

$$\boxed{\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}}$$

$\sigma_X = \sqrt{V(X)}$

$\sigma_Y = \sqrt{V(Y)}$

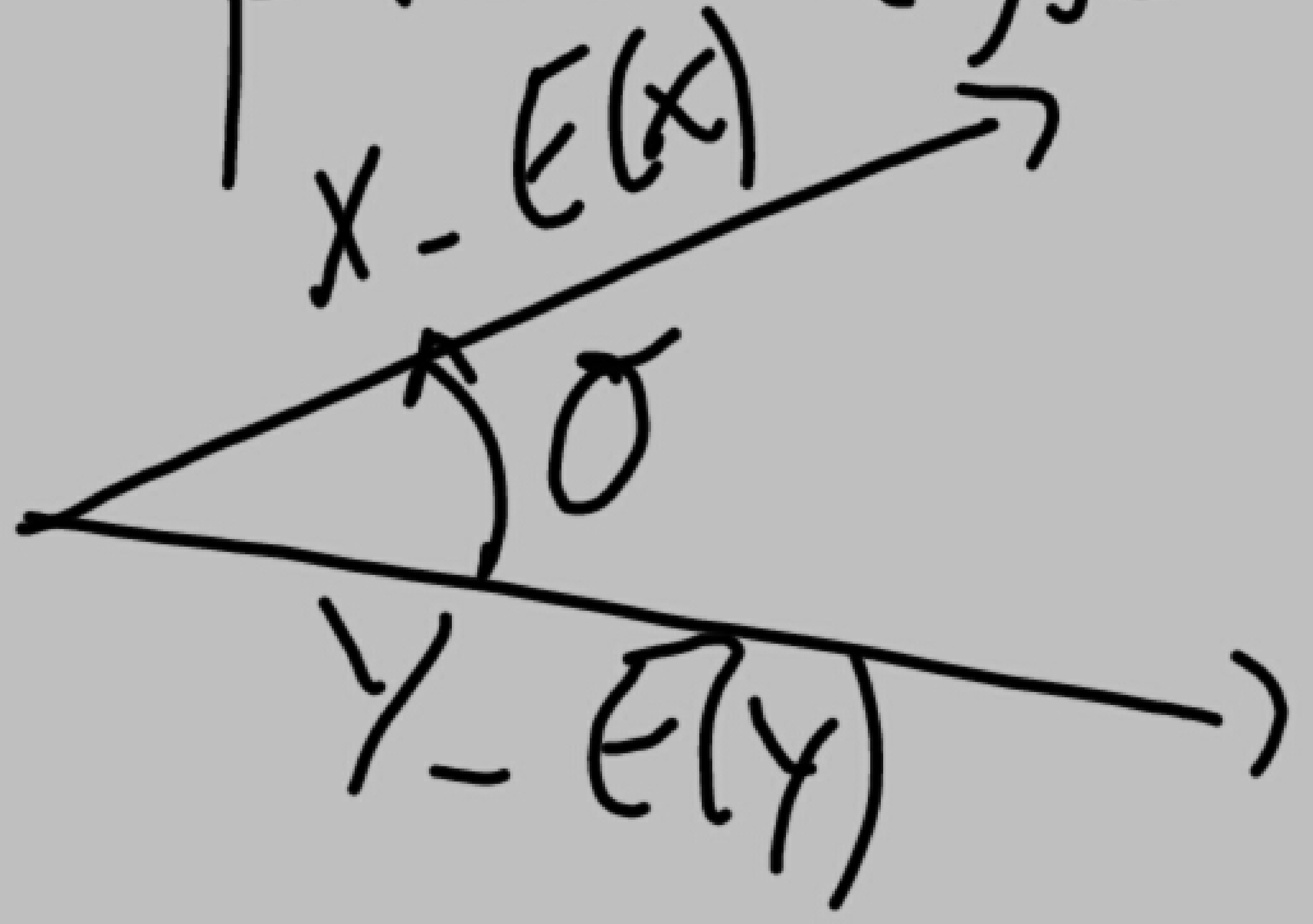
<u>Rq</u> E : espace des V.A

$$\boxed{\cos\theta = \rho(X,Y) = \frac{\langle X-E(X),\ Y-E(Y)\rangle}{\|X-E(X)\|\ \|Y-E(Y)\|}}$$

$$Cov(X,Y) = \langle X-E(X),\ Y-E(Y)\rangle$$ produit scalaire

$$\|X-E(X)\| = \sigma(X)$$

$$\|Y-E(Y)\| = \sigma(Y)$$



$$|\rho| \leq 1$$

si $\rho = 1$ $\left(\theta = 0\ [2\pi]\right) \Rightarrow$ forte corrélation lin

si $\rho = 0$ $\left(\theta = \frac{\pi}{2}\ [\pi]\right)$

$\Rightarrow$ pas de corrélation lin

Ex1

X v.a. de loi :

x_i	-2	-1	0	1	2
P_i	$\frac{1}{6}$	$\frac{1}{4}$	$1/6$	$1/4$	$\frac{1}{6}$

Soit $Y = X^2$

1) Donner la loi du couple (X, Y)

2) En déduire la loi marginale de Y

3) Indépendance et calculer $\text{Cov}(X, Y)$

1) $P((X=i)\cap(Y=j)) = 0$ si $j \neq i^2$

$P((X=i)\cap(Y=i^2)) = P(X=i)$

X \ Y	0	1	4	Loi de X
−2	0	0	1/6	1/6
−1	0	1/4	0	1/4
0	1/6	0	0	1/6
1	0	1/4	0	1/4
2	0	0	1/6	1/6
Loi Y	1/6	1/2	1/3	1

2) Loi de Y
dernière ligne
du tableau

3) $\mathbb{P}((X=0)\cap(Y=1)) = 0 \neq \mathbb{P}(X=0).\mathbb{P}(Y=1)$

$\frac{1}{6}\cdot\frac{1}{2}=\frac{1}{12}$

$\Rightarrow$ X et Y ne sont pas independants

4) $\mathrm{cov}(X,Y) = E(X.Y) - E(X)E(Y)$

$E(X.Y) = \sum_{i,j} x_i y_j P_{ij} = -\frac{1}{4}+\frac{1}{4}-\frac{8}{6}+\frac{8}{6}$

$= 0$

$E(X) = -\frac{2}{6}+\frac{2}{6} = 0$