Probabilités et statistique

Distribution conjointe

Module 3

Plan

- **Variables aléatoires bidimensionnelles**

  Fonction de masse conjointe

  Fonction de masse marginale

  Fonction de masse conditionnelle

  Indépendance de variables

- **Espérance bidimensionnelle**

  Espérance conditionnelle

  Espérance d'une fonction de variable bidimensionnelle

- **Corrélation de variables**

  Covariance

  Interprétation de la corrélation

1. Variables aléatoires bidimensionnelles

A. Fonction de masse conjointe

➢ ***Définition***

Soit (X, Y) un couple de variables aléatoires discrètes. La fonction de masse est définie par

$$\boxed{\begin{array}{l} p(x, y) = \mathrm{P}(\mathrm{X} = x\, ,\, \mathrm{Y} = y) \\ x \in R_{\mathrm{X}} \;\;,\;\; y \in R_{\mathrm{Y}} \end{array}}$$

et

➢ ***Propriétés***

- $0 \leq p(x\, , y) \leq 1$
- $\sum\limits_{x \in R_{\mathrm{X}}} \sum\limits_{y \in R_{\mathrm{Y}}} p(x\, , y) = 1$

Exemple

Les accidents provoqués par les jeunes conducteurs.

X : « nombre d'années d'expérience de conduite »

Y : « nombre d'accidents dans la dernière année »

X \ Y	0	1	Total
1	0.05	0.25	0.30
2	0.15	0.20	0.35
3	0.25	0.1	0.35
Total	0.45	0.55	1

$$R_{\mathrm{X}} = \{1\,,2\,,3\}$$

$$R_{\mathrm{Y}} = \{0\,,1\}$$

a) Quelle est la probabilité qu'un conducteur ait 2 ans d'expérience et ait eu 1 accident ?

$$P(\mathrm{X}=2\,,\mathrm{Y}=1)=p(2\,,1)=0.20$$

b) Quelle est la probabilité qu'un conducteur de plus d'un an d'expérience ait eu aucun accident ?

$$P(X>1\,,Y=0)=\mathrm{P}(\mathrm{X}=2\,,\mathrm{Y}=0)+\mathrm{P}(\mathrm{X}=3\,,\mathrm{Y}=0)$$

$$=p(2\,,0)+p(3\,,0)=0.15+0.25=0.40$$

B. Fonction de masse marginale

- Fonction de masse marginale de X .

$$p_{\mathrm{X}}(x) = \sum_{y \in R_{\mathrm{Y}}} p(x\,,\,y)$$

Exemple: Les jeunes conducteurs

c) Quelle est la probabilité qu'un jeune conducteur ait 1 an d'expérience ?

$$p_{\mathrm{X}}(1) = p(1\,,0) + p(1\,,1) = 0.05 + 0.25 = 0.3$$

d) Quelle est la probabilité qu'un jeune conducteur ait 2 ans d'expérience ?

$$p_{\mathrm{X}}(2) = p(2\,,0) + p(2\,,1) = 0.15 + 0.20 = 0.35$$

e) Quelle est la probabilité qu'un jeune conducteur ait 3 ans d'expérience ?

$$p_{\mathrm{X}}(3) = p(3\,,0) + p(3\,,1) = 0.25 + 0.10 = 0.35$$

- Fonction de masse marginale de Y .

$$p_{\mathrm{Y}}(y) = \sum_{x \in R_{\mathrm{X}}} p(x\,,\,y)$$

Exemple : Les jeunes conducteurs

f) Quelle est la probabilité qu'un jeune conducteur n'ait pas eu d'accident ?

$$p_{\mathrm{Y}}(0) = p(1\,,0) + p(2\,,0) + p(3\,,0) = 0.05 + 0.15 + 0.25 = 0.45$$

g) Quelle est la probabilité qu'un jeune conducteur ait 1 accident ?

$$p_{Y}(1) = p(1\,,1) + p(2\,,1) + p(3\,,1) = 0.25 + 0.20 + 0.10 = 0.55$$

Remarques

- $\sum_{x \in R_{\mathrm{X}}} p_{\mathrm{X}}(x) = 1$ et $\sum_{y \in R_{\mathrm{Y}}} p_{\mathrm{Y}}(y) = 1$

c. **Fonction de masse conditionnelle**

- Fonction de masse conditionnelle de X sachant $\mathrm{Y} = y$.

$$p_{\mathrm{X}/\mathrm{Y}=y}(x) = \mathrm{P}\left(\mathrm{X} = x / \mathrm{Y} = y\right) = \frac{\mathrm{P}\left(\mathrm{X} = x\,,\,\mathrm{Y} = y\right)}{\mathrm{P}\left(\mathrm{Y} = y\right)} = \frac{p(x\,,\,y)}{p_{Y}(y)}$$

Exemple : Les jeunes conducteurs

h) Quelle est la probabilité qu'un jeune conducteur n'ayant pas eu d'accident, ait 1 an d'expérience ?

$$p_{\mathrm{X}/\mathrm{Y}=0}(1) = \frac{p(1\,,0)}{p_{\mathrm{Y}}(0)} = \frac{0.05}{0.45} = 0.11$$

i) Quelle est la probabilité qu'un jeune conducteur n'ayant pas eu d'accident, ait 2 ans d'expérience ?

$$p_{\mathrm{X}/\mathrm{Y}=0}(2) = \frac{p(2\,,0)}{p_{\mathrm{Y}}(0)} = \frac{0.15}{0.45} = 0.33$$

j) Quelle est la probabilité qu'un jeune conducteur n'ayant pas eu d'accident, ait 2 ans d'expérience ?

$$p_{\mathrm{X}/\mathrm{Y}=0}(3) = \frac{p(3\,,0)}{p_{\mathrm{Y}}(0)} = \frac{0.25}{0.45} = 0.56$$

➢ Fonction de masse conditionnelle de Y sachant $\mathrm{X}=x$.

$$p_{\mathrm{Y}/\mathrm{X}=x}(y)=\mathrm{P}\left(\mathrm{Y}=y/\mathrm{X}=x\right)=\frac{\mathrm{P}\left(\mathrm{X}=x\,,\,\mathrm{Y}=y\right)}{\mathrm{P}\left(\mathrm{X}=x\right)}=\frac{p(x\,,\,y)}{p_{\mathrm{X}}(x)}$$

Exemple : Les jeunes conducteurs

k) Quelle est la probabilité qu'un jeune conducteur ayant 2 ans d'expérience, n'ait pas eu d'accident ?

$$p_{\mathrm{Y}/\mathrm{X}=2}(0)=\frac{p(2\,,0)}{p_{\mathrm{X}}(2)}=\frac{0.15}{0.35}=0.43$$

l) Quelle est la probabilité qu'un jeune conducteur ayant 2 ans d'expérience, ait eu 1 accident ?

$$p_{\mathrm{Y}/\mathrm{X}=2}(1)=\frac{p(2\,,1)}{p_{\mathrm{X}}(2)}=\frac{0.20}{0.35}=0.57$$

Remarques

- $\sum_{x\in R_{\mathrm{X}}} p_{\mathrm{X}/\mathrm{Y}=y}(x)=1$ et $\sum_{y\in R_{\mathrm{Y}}} p_{\mathrm{Y}/\mathrm{X}=x}(y)=1$

D. Indépendance de variables

- Les variables X et Y sont indépendantes si

$$p(x,y) = p_X(x) \times p_Y(y)$$
$$\text{pour tout } x \in R_X \text{ et } y \in R_Y$$

Exemple : Les jeunes conducteurs

X et Y sont-ils indépendants ?

$$p(1,0) = 0.05 \;;\; p_X(1) = 0.3 \;;\; p_Y(0) = 0.45$$
$$p_X(1) \times p_Y(0) = 0.3 \times 0.45 = 0.135 \neq 0.05 = p(1,0)$$
$$p_X(1) \times p_Y(0) \neq p(1,0)$$

Donc les variables X et Y ne sont pas indépendants.

- Les variables X et Y sont indépendantes si

  - $p_{X/Y=y}(x) = p_X(x)$ pour tout $x \in R_X$
  - $p_{Y/X=x}(y) = p_Y(y)$ pour tout $y \in R_Y$

2. Espérance bidimensionnelle

A. Espérance conditionnelle

$$E(\mathrm{X}/\mathrm{Y}=y)=\sum_{x\in R_{\mathrm{X}}} x p_{\mathrm{X}/\mathrm{Y}=y}(x) \quad \text{et} \quad E(\mathrm{Y}/\mathrm{X}=x)=\sum_{y\in R_{\mathrm{Y}}} y p_{\mathrm{Y}/\mathrm{X}=x}(y)$$

Exemple : Les jeunes conducteurs

- La moyenne en années d'expérience des conducteurs qui n'ont pas eu d'accidents

$$E(\mathrm{X}/\mathrm{Y}=0)=1\times\tfrac{0.05}{0.45}+2\times\tfrac{0.15}{0.45}+3\times\tfrac{0.25}{0.45}=2.44$$

- La moyenne en années d'expérience des conducteurs ayant eu 1 accident.

$$E(\mathrm{X}/\mathrm{Y}=1)=1\times\tfrac{0.25}{0.55}+2\times\tfrac{0.20}{0.55}+3\times\tfrac{0.10}{0.55}=1.73$$

- La moyenne en nombre d'accidents des conducteurs qui ont eu 1 an d'expérience

$$E(\mathrm{Y}/\mathrm{X}=1)=0\times\tfrac{0.05}{0.30}+1\times\tfrac{0.25}{0.30}=0.83$$

- La moyenne en nombre d'accidents des conducteurs qui ont eu 2 ans d'expérience

$$E(\mathrm{Y}/\mathrm{X}=2)=0\times\tfrac{0.15}{0.35}+1\times\tfrac{0.20}{0.35}=0.57$$

- La moyenne en nombre d'accidents des conducteurs qui ont eu 3 ans d'expérience

$$E(\mathrm{Y}/\mathrm{X}=3)=0\times\tfrac{0.25}{0.35}+1\times\tfrac{0.1}{0.35}=0.28$$

B. Espérance d'une fonction de variable bidimensionnelle

$$\boxed{E[g(\mathrm{X},\mathrm{Y})] = \sum_{x \in R_\mathrm{X}} \sum_{y \in R_\mathrm{Y}} g(x,y)\, p(x,y)}$$

- $g(\mathrm{X},\mathrm{Y}) = \mathrm{XY}$ $\quad E(\mathrm{XY}) = \sum_{x \in R_\mathrm{X}} \sum_{y \in R_\mathrm{Y}} xyp(x,y)$
- $g(\mathrm{X},\mathrm{Y}) = \mathrm{X}$ $\quad E(\mathrm{X}) = \sum_{x \in R_\mathrm{X}} \sum_{y \in R_\mathrm{Y}} xp(x,y) = \sum_{x \in R_\mathrm{X}} x\left(\sum_{y \in R_\mathrm{Y}} p(x,y)\right) = \sum_{x \in R_\mathrm{X}} xp_\mathrm{X}(x)$
- $g(\mathrm{X},\mathrm{Y}) = \mathrm{Y}$ $\quad E(\mathrm{Y}) = \sum_{y \in R_\mathrm{Y}} \sum_{x \in R_\mathrm{X}} yp(x,y) = \sum_{y \in R_\mathrm{Y}} y\left(\sum_{x \in R_\mathrm{X}} p(x,y)\right) = \sum_{y \in R_\mathrm{Y}} yp_\mathrm{Y}(y)$
- $g(\mathrm{X},\mathrm{Y}) = a\mathrm{X} + b\mathrm{Y}$ $\quad E(a\mathrm{X} + b\mathrm{Y}) = aE(\mathrm{X}) + bE(\mathrm{Y})$
  avec *a et b* des constantes réelles.

Exemple : Les jeunes conducteurs

- $E(\mathrm{XY}) = 1\times 0\times 0.05 + 1\times 1\times 0.25 + 2\times 0\times 0.15 + 2\times 1\times 0.20 + 3\times 0\times 0.25 + 3\times 1\times 0.1 = 0.95$
- $E(\mathrm{X}) = 1\times 0.3 + 2\times 0.35 + 3\times 0.35 = 2.05$ et $E(\mathrm{Y}) = 0\times 0.45 + 1\times 0.55 = 0.55$
- $E(2\mathrm{X} - 3\mathrm{Y}) = 2\times E(\mathrm{X}) - 3\times E(\mathrm{Y}) = 2\times 2.05 - 3\times 0.55 = 2.45$

3. Corrélation de variables

A. Covariance

La covariance de deux variables aléatoires X et Y est définie par :

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

Interprétation

- $Cov(X,Y) = 0 \Rightarrow$ pas de corrélation.
- $Cov(X,Y) > 0 \Rightarrow$ Corrélation positive : X et Y évoluent de la même manière.
- $Cov(X,Y) < 0 \Rightarrow$ Corrélation négative : X et Y évoluent en sens opposé

Exemple : Les jeunes conducteurs

$E(XY) = 0.95$; $E(X) = 2.05$; $E(Y) = 0.55$

$Cov(X,Y) = 0.95 - 2.05 \times 0.55 = -0.1775 \Rightarrow$ Corrélation négative

Propriété

$Cov(aX, bY) = abCov(X,Y)$; a et b des constantes réelles.

B. Coefficient de corrélation

Ce coefficient mesure le sens et le niveau de corrélation entre les variables.

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{V(X) \times V(Y)}}$$

Exemple : Les jeunes conducteurs

$\bullet\, Cov(X, Y) = -0.1775$

$E(X) = 1 \times 0.3 + 2 \times 0.35 + 3 \times 0.35 = 2.05$

$E(X^2) = 1^2 \times 0.3 + 2^2 \times 0.35 + 3^2 \times 0.35 = 4.6$

$\bullet\, V(X) = 4.6 - (2.05)^2 = 0.397$

$E(Y) = 0 \times 0.45 + 1 \times 0.55 = 0.55$

$E(Y^2) = 0^2 \times 0.45 + 1^2 \times 0.55 = 0.55$

$\bullet\, V(Y) = 0.55 - (0.55)^2 = 0.247$

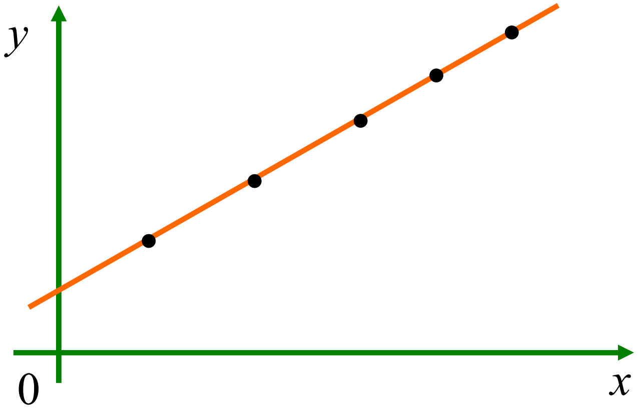
$$\bullet\, \rho_{XY} = \frac{-0.1775}{\sqrt{0.397 \times 0.247}} = -0.5668 \Rightarrow \text{Corrélation négative et moyenne.}$$

Remarques

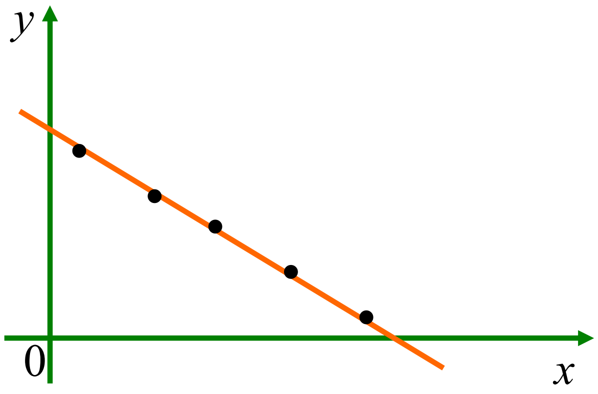
$\bullet\, -1 \leq \rho_{XY} \leq 1$

$\bullet\, \rho_{XY}$ est sans unité.

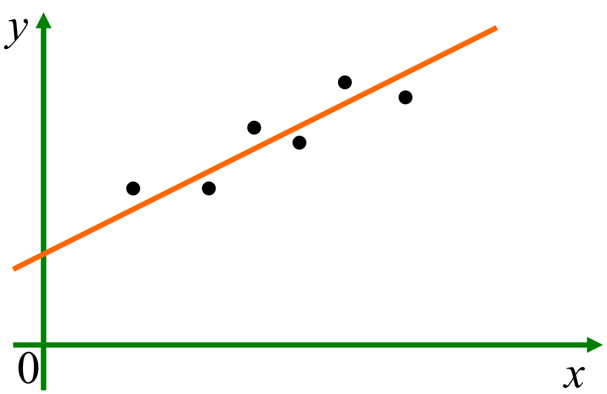
c. Interprétation de la corrélation



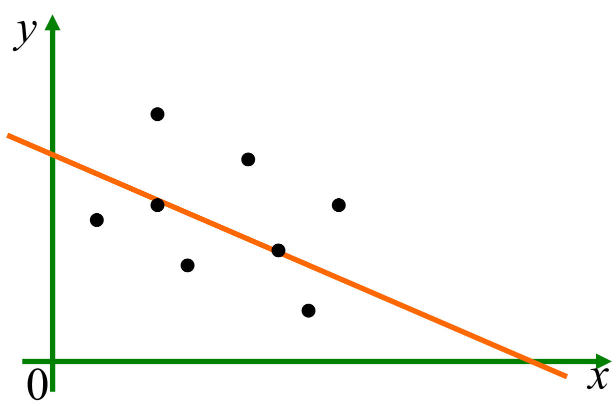
$\rho_{XY} = 1 \Rightarrow$ Corrélation parfaitement positive



$\rho_{XY} = -1 \Rightarrow$ Corrélation parfaitement négative



$\rho_{XY} = 0.8 \Rightarrow$ Corrélation fortement positive



$\rho_{XY} = -0.5 \Rightarrow$ Corrélation moyennement négative

- Cas où $\rho_{XY} = 0$.

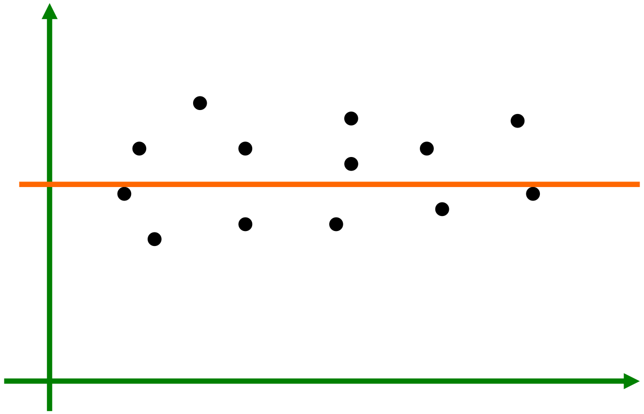
$$\text{Si les variables X et Y sont indépendantes} \Rightarrow \rho_{XY} = 0$$
$$\text{mais l'inverse n'est pas vrai.}$$

Exemple

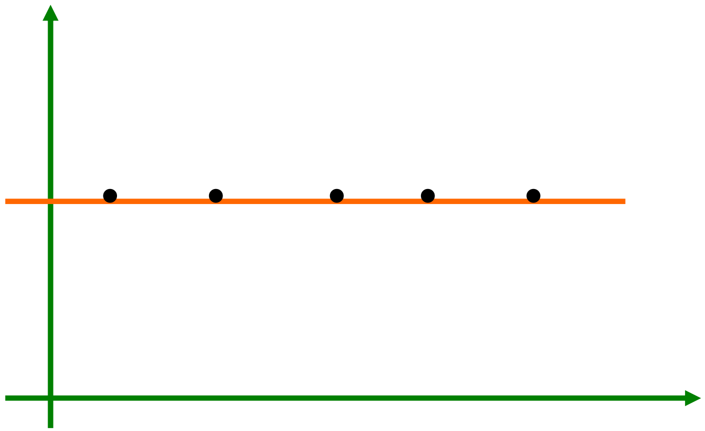
X \ Y	0	1	Total
-1	0	1/3	1/3
0	1/3	0	1/3
1	0	1/3	1/3
Total	1/3	2/3	1

- $\rho_{XY} = 0.$
- Les variables X et Y ne sont pas indépendantes.

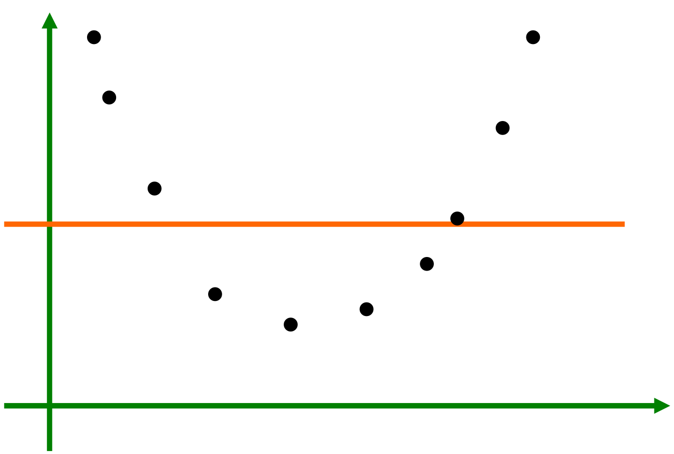
- Cas où $\rho_{XY} = 0$ et les variables X et Y ne sont pas indépendantes.



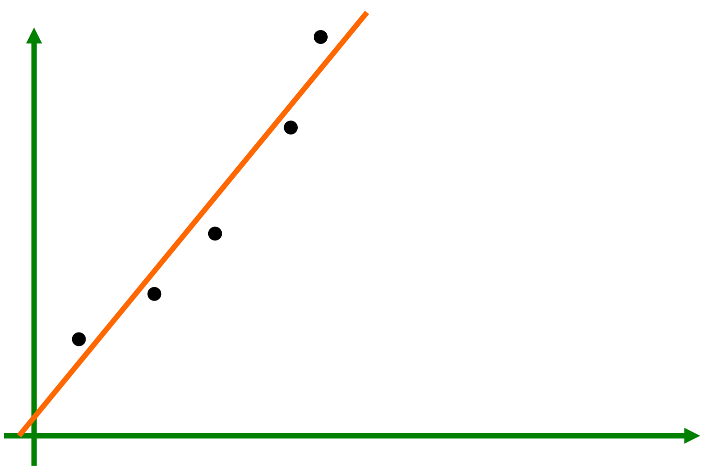
$\rho_{XY} = 0 \quad ; \quad \text{pente} = 0$



$\rho_{XY} = 0 \quad ; \quad \text{pente} = 0$



$\rho_{XY} = 0 \quad ; \quad \text{pente} = 0$



$\rho_{XY} = 0.8$

Variance

$$V(aX+bY)=a^2V(X)+b^2V(Y)+2abCov(X,Y)$$

avec a et b des constantes réelles.

Exemple : Les jeunes conducteurs

- $V(X)=4.6-(2.05)^2=0.397 \quad ; \qquad V(Y)=0.55-(0.55)^2=0.247$
- $Cov(X,Y)=-0.1775$
- $V(2X-3Y)=2^2V(X)+(-3)^2V(Y)+2(2)(-3)Cov(X,Y)$

$$=4\times 0.397+9\times 0.247-12\times(-0.1775)$$

$$=5.941$$

Remarque

Si les variables X et Y sont indépendantes alors

$V(aX+bY)=a^2V(X)+b^2V(Y)$