Statistique
Cours 3

Vecteurs aléatoires

$X \in [10°, 800°]$

$Y \in [20\,cm, 30\,cm]$

$P(X = x, Y = y)$

~~proba~~ probabilité (con)jointe

Deux dés

X = max de faces

Y = ~~somme~~ somme de faces

$$P(X=3, Y=4) = P(3,1) + P(1,3)$$
$$= \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$$

$X = (x_1 = 1, x_2 = 2, \dots x_6 = 6)$

Probabilités marginales

$X \backslash Y$	$y_1 \dots$	$\dots y_j \dots$	y_x
x_1			
$\vdots$			
x_i		$p(x_i, y_j)$	$\sum_{j=1}^{x} p(x_i, y_j)$
x_m			
		$\sum_{i=1}^{m} p(x_i, y_j)$	1

$$W = \begin{bmatrix} X \\ Y \end{bmatrix}$$

$$W = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix} \qquad E(W) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_m) \end{pmatrix}$$

Matrice de variance - covariance

$$C(W) = E\left[(W - E(W))(W - E(W))^T\right]$$
$$= E(WW^T) - E(W)\,E^T(W)$$

X	Y
0	1
1	3
2	4
5	4

$E(X) = 2 \quad E(Y) = 3$

$$C(X,Y) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} X_i Y_j}{} - E(X)E(Y)$$

$$\frac{\sum x_i^2}{4} = \frac{30}{4} = 7,5 \qquad \frac{\sum y_j^2}{4} = 10,5$$

$$\frac{\sum\sum x_i y_j}{4} = \frac{31}{4} = 7,75$$

$$\begin{array}{cc} C(X,X) & C(X,Y) \\ \end{array}$$

$$C(X,Y) = \begin{bmatrix} 7,5-4 & 7,75-6 \\ 7,75-6 & 10,5-9 \end{bmatrix} = \begin{bmatrix} 2,5 \ {\color{red}= V(X)} & 1,75 \\ 1,75 & 1,5 \ {\color{red}= V(Y)} \end{bmatrix}$$

$C(Y,Y)$

$$\rho(X,Y) = \frac{C(X,Y)}{\sigma_X \sigma_Y} = \frac{1,75}{\sqrt{2,5}\sqrt{1,5}} = 0,76$$ corrélation des variables X et Y

Echantillonnage

population

échantillon

⇓

statistique descriptive

- sondage
- tirage aléatoire avec remise
  - simple
  - par strates homogènes
  - par grappes (pour augmenter la précision des calculs)

⇒ moyenne empirique (expérimentale) $\bar{X}$, m_X

⇒ variance empirique S^2_X, $V(X)$

$\Rightarrow$ inférence statistique : μ_x de la population
σ^2_x

fondé $\Downarrow$ sur l'aléa $\rightarrow$ associé à un coût de se tromper.

Echantillonnage en vue d'étudier une statistique
$S = f(X_1, \dots, X_n)$

Une réalisation de la statistique est :
$s = f(X_1 = x_1, \dots, X_n = x_n)$

Exemples d'étude
X va
$E(X) = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} (X_i = x_i)$

$\bar{X}$ estimateur de $\mu \Rightarrow \bar{X} = $ va.

- $\bar{X}$ estimateur sans biais : $E(\bar{X}) = \mu$
- asymptotiquement efficace $V(\bar{X}) = \frac{\sigma^2_x}{n}$
- fortement convergent $E(\bar{X}) \xrightarrow{ps} \mu$
- si n grand alors $E(\bar{X})$ est approximativement gaussienne. $E(\bar{X}) = E\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n} \cdot n\, E(\bar{X})$
- Si n grand alors $E(\bar{X})$ est approximativement gaussienne $\Rightarrow \sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma_x} \sim \mathcal{N}(0, 1)$

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ — S^2 estimateur sans biais $E(S^2) = \sigma_x^2$

$S_1^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ — S_1^2 — avec biais si $n > 1000$

$V(S^2) = \frac{\mu^4 - \sigma_x^4}{n}$