

Statistique Cours 3

Vecteurs aléatoires

$$X \in [10^\circ, 800^\circ]$$

$$Y \in [20 \text{ cm}, 30 \text{ cm}]$$

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

probabilité (con)jointe

Deux dés

X = max de faces

Y = 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

$X \setminus Y$	y_1	\dots	y_j	\dots	y_n
x_1					
\vdots					
x_i			$P(x_i, y_j)$		$\sum_{j=1}^n P(x_i, y_j)$
\vdots					
x_m					$\sum_{j=1}^n P(x_i, y_j)$
			$\sum_{i=1}^m P(x_i, y_j)$		1

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

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

Matrice de variance-covariance

$$C(W) = E[(W - E(W))(W - E(W))^T]$$

$$= E(WW^T) - E(W)E^T(W)$$

X	Y
0	1
1	3
2	4
5	4

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

$$C(X, Y) = \frac{\sum_{i=0}^m \sum_{j=0}^m X_i Y_j}{n} - E(X)E(Y)$$

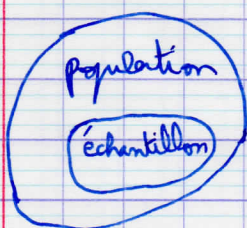
$$\frac{\sum x_i^2}{4} = \frac{30}{4} = 7,5 \quad \frac{\sum y_j^2}{4} = 10,5$$

$$\frac{\sum \sum x_i y_j}{4} = \frac{32}{4} = 7,75$$

$$C(X, Y) = \begin{bmatrix} C(X, X) & C(X, Y) \\ C(X, Y) & C(Y, Y) \end{bmatrix} = \begin{bmatrix} 7,5 - 4 & 7,75 - 6 \\ 7,75 - 6 & 10,5 - 9 \end{bmatrix} = \begin{bmatrix} 3,5 & 1,75 \\ 1,75 & 1,5 \end{bmatrix}$$

$$r(X, Y) = \frac{C(X, Y)}{\sqrt{\sigma_x^2 \sigma_y^2}} = \frac{1,75}{\sqrt{2,5 \cdot 1,5}} = 0,76 \quad \text{corrélation des variables X et Y}$$

Echantillonnage



statistique descriptive

- ⇒ moyenne empirique (expérimentale) \bar{X} , m_x
- ⇒ variance empirique S_x^2 , $V(X)$

- sondage

- tirage aléatoire avec remise

- simple

- par strates homogènes

- par grappes (pour augmenter la précision des calculs)

⇒ inférence statistique : μ_x de la population

fondé sur l'aléa → associé à un coût de σ_x^2 & tromper

Echantillonnage en vue d'étudier une statistique

$$S = f(X_1, \dots, X_m)$$

Une réalisation de la statistique est :

$$S = f(X_1 = x_1, \dots, X_m = x_m)$$

Exemples d'étude

X_{nra}

$$E(X) = \bar{X} = \frac{1}{m} \sum_{i=1}^m (X_i = x_i)$$

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

- \bar{X} estimateur sans biais : $E(\bar{X}) = \mu$
- asymptotiquement efficace $V(\bar{X}) = \frac{\sigma_x^2}{m}$
- fortement convergent $E(\bar{X}) \xrightarrow{p_0} \mu$
- si m grand alors $E(\bar{X})$ est approximativement gaussienne. $E(\bar{X}) = E\left(\frac{1}{m} \sum X_i\right) = \frac{1}{m} \cdot m E(X)$
- Si m grand alors $E(\bar{X})$ est approximativement gaussienne $\Rightarrow \sqrt{m} \cdot \frac{\bar{X} - \mu}{\sigma_x} \sim d'(0, 1)$

$$S^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$$

$$S_{\Delta}^2 = \frac{1}{m} \sum_{i=1}^m (X_i - \bar{X})^2$$

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

S^2 estimateur sans biais $E(S^2) = \sigma_x^2$

S_{Δ}^2 — avec biais

si $m > 1000$