

ANDO - Analyse de données

Professeur : Regragui Mohamed

Cours de : Sébastien Goubeau

8 juin 2020

Table des matières

1	Description bidimensionnelle et mesure de corrélation	3
1.1	Loi conjointe de \mathcal{X}, \mathcal{Y}	3
1.2	Loi marginale	4
1.3	Loi conditionnelle	4
1.4	Loi d'une fonction de 2 variables ($\mathcal{Z} = g(\mathcal{X}, \mathcal{Y})$) (Loi composé)	5
1.5	Espérance de $\mathcal{Z} = g(\mathcal{X}, \mathcal{Y})$	6
1.6	Covariance et coefficient de corrélation linéaire	8
1.7	Exercice 1	8
1.7.1	Donner la loi du couple \mathcal{X}, \mathcal{Y}	9
1.7.2	En déduire la loi marginale de \mathcal{Y}	9
1.7.3	Indépendance et calculer $COV(\mathcal{X}, \mathcal{Y})$	9
1.7.4	$COV(\mathcal{X}, \mathcal{Y}) = E(\mathcal{X}, \mathcal{Y}) - E(\mathcal{X})E(\mathcal{Y})$	9
1.8	Exercice 2	10
1.8.1	Déterminer a	10
1.8.2	Déterminer les lois marginales de \mathcal{X} et \mathcal{Y}	11
1.8.3	Indépendance?	11
1.8.4	Calculer $COV(\mathcal{X}, \mathcal{Y})$	11
1.9	Exercice 3	12
1.9.1	Déterminer la loi du couple $(\mathcal{X}, \mathcal{Y})$	12
1.9.2	Calculer $P(\mathcal{X} = \mathcal{Y})$	12
1.9.3	Donner la loi marginale de \mathcal{Y} et $E(\mathcal{Y})$	13
1.10	Exercice 4	14
1.10.1	La loi du couple $(\mathcal{X}_1, \mathcal{X}_2)$ et en déterminer la loi de \mathcal{X}_2	14
1.10.2	Déterminer la loi de \mathcal{Z}_2	15
1.10.3	Déterminer $\mathcal{Z}_p(\Omega)$ et calculer $P(\mathcal{X}_{p+1} = 1 / \mathcal{Z}_p = k)$	15
1.10.4	Montrer que $P(\mathcal{X}_{p+1} = 1) = \frac{1+cE(\mathcal{Z}_p)}{2+pc}$	16
1.10.5	Montrer que $P(\mathcal{X}_p = 1) = P(\mathcal{X}_p = 0) = \frac{1}{2} \quad (\forall p \in \llbracket 1, n \rrbracket)$	16
1.11	Exercice 5	17
1.11.1	Question 1 : loi marginale et indépendance	17
1.11.2	Question 2 : loi conjointe	18
1.12	Exercice 6	22
1.12.1	Déterminer les loi marginales ainsi que $E(\mathcal{X}), V(\mathcal{X}), E(\mathcal{Y}), V(\mathcal{Y})$	22
1.12.2	\mathcal{X} et \mathcal{Y} sont-elles indépendantes?	24
1.12.3	Déterminer la loi de $\mathcal{Z} = \mathcal{X} - \mathcal{Y}$	25
1.12.4	\mathcal{Y} et \mathcal{Z} sont-elle indépendantes?	25

2	Analyse en composante principale	26
2.1	Les données et leurs caractéristiques	26
2.1.1	Tableau de donnée	26
2.1.2	Matrice de poids	26
2.1.3	Centre de gravité	26
2.1.4	Matrice de Variance - Covariance et Matrice de Corrélation	27
2.2	Espace des individus	28
2.3	Espace des variables	29
2.4	Variables engendrées par un tableau de données	30
2.5	Algorithme A.C.P. (Analyse Composante Principale)	30
2.6	Projection	30

1 Description bidimensionnelle et mesure de corrélation

1.1 Loi conjointe de \mathcal{X}, \mathcal{Y}

Soient \mathcal{X} et \mathcal{Y} deux variables aléatoires discrètes définies sur le même espace probabilisé : (Ω, φ, P) .

Avec :

- Ω : Univers
- φ : Tribu (classe des évènements)
- P : Loi de définition restreinte

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

$\mathcal{Y}(\Omega) = \{y_j / j \in \mathcal{J}\}$ valeurs de \mathcal{Y}

Définition - Loi conjointe

On appelle loi conjointe du couple $(\mathcal{X}, \mathcal{Y})$ l'ensemble des couples (x_i, y_j, P_{ij}) où $x_i \in \mathcal{X}(\Omega), y_j \in \mathcal{Y}(\Omega)$

$$P_{ij} = P((\mathcal{X} = x_i) \cap (\mathcal{Y} = y_j))$$

Remarque

Si $\mathcal{I} = \llbracket 1, r \rrbracket$ et $\mathcal{J} = \llbracket 1, s \rrbracket$

\mathcal{X}/\mathcal{Y}	y_1	...	y_j	...	y_s
x_1			:		
...			:		
x_i	P_{ij}
...			:		
x_r			:		

$$P_{ij} = P((\mathcal{X} = x_i) \cap (\mathcal{Y} = y_j))$$

$$P_{ij} \geq 0$$

$$\sum_{i,j} P_{ij} = 1$$

1.2 Loi marginale

Définition - Loi marginale

Les variables \mathcal{X} et \mathcal{Y} sont appelées variables marginales.

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

$$P_{\forall j}(\mathcal{X} = x_i) = \sum_{j \in \mathcal{J}} P_{ij} = P_{i\bullet} \text{ (Notation statistique de la loi marginale)}$$

$$P_{\forall i}(\mathcal{Y} = y_j) = \sum_{i \in \mathcal{I}} P_{ij} = P_{\bullet j} \text{ (Notation statistique de la loi marginale)}$$

Exemple

\mathcal{X}/\mathcal{Y}	1	2	3	4	$P_{i\bullet}$
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	$\frac{4}{16}$	$\frac{1}{4}$
$P_{\bullet j}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$	1

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

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

Le 1 en bas à droite du tableau permet de vérifier que les calculs sont bons (la valeur doit toujours être à 1).

1.3 Loi conditionnelle

Définition - Loi Conditionnelle

On appelle loi conditionnelle de $\mathcal{X} = x_i$ sachant $\mathcal{Y} = y_j$:

$$P(\mathcal{X} = x_i / \mathcal{Y} = y_j) = \frac{P((\mathcal{X}=x_i) \cap (\mathcal{Y}=y_j))}{P(\mathcal{Y}=y_j)} = \frac{P_{ij}}{P_{\bullet j}}$$

$$P(\mathcal{Y} = y_j / \mathcal{X} = x_i) = \frac{P_{ij}}{P_{i\bullet}}$$

Exemple

On reprend le tableau de l'exemple précédent.

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

On peut s'aider d'un tableau :

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

Définition - Indépendance

\mathcal{X} et \mathcal{Y} sont indépendants si et seulement si :

$$P((\mathcal{X} = x) \cap (\mathcal{Y} = y)) = P(\mathcal{X} = x)P(\mathcal{Y} = y)$$

$$\Leftrightarrow P_{ij} = P_{i\bullet} \times P_{\bullet j} \quad (\forall (ij) \in \mathcal{I} \times \mathcal{J})$$

$$\Leftrightarrow P(\mathcal{X} = x/\mathcal{Y} = y) = P(\mathcal{X} = x)$$

1.4 Loi d'une fonction de 2 variables ($\mathcal{Z} = g(\mathcal{X}, \mathcal{Y})$) (Loi composée)

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

$$\mathcal{Z} = g(\mathcal{X}, \mathcal{Y})$$

$$(\mathcal{Z} = zk) = \bigcup_{(i,j) | g(x_i, y_j) = zk} ((\mathcal{X} = x_i) \cap (\mathcal{Y} = y_j))$$

Remarque

L'expression ci-dessus signifie en français :

L'événement $(\mathcal{Z} = zk)$ (symbolisé par des parenthèses) est égale à l'union des $(\mathcal{X} = x_i) \cap (\mathcal{Y} = y_j)$ qui vérifie le couple (i, j) tel que $g(x_i, y_j) = zk$.

$$P(\mathcal{Z} = zk) = \sum_{(i,j) | g(x_i, y_j) = zk} P((\mathcal{X} = x_i) \cap (\mathcal{Y} = y_j))$$

En particulier $g : (\mathcal{X}, \mathcal{Y}) = \mathcal{X} + \mathcal{Y} = \mathcal{S}$

$$P(\mathcal{S} = s) = \sum_{(i,j) | x_i + y_j = s_k} P((\mathcal{X} = x_i) \cap (\mathcal{Y} = y_j))$$

Exemple

Reprise de l'exemple précédent.

$$\mathcal{S} = \mathcal{X} + \mathcal{Y}$$

$$\mathbb{P} = \mathcal{X} \cdot \mathcal{Y}$$

Loi de \mathcal{S}

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

Pour remplir le tableau : $P(\mathcal{S} = 5) = P_{1,4} + P_{2,3} + P_{3,2} + P_{4,1} = \frac{2}{16} = \frac{1}{8}$

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

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

La ligne doit être égale à 1.

1.5 Espérance de $\mathcal{Z} = g(\mathcal{X}, \mathcal{Y})$

Définition - Espérance d'une Loi composé

Soit \mathcal{Z} une loi composé, l'espérance de cette loi est :

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

Remarque

Si \mathcal{X} et \mathcal{Y} sont 2 variables indépendantes alors : $E(\mathcal{X}, \mathcal{Y}) = E(\mathcal{X})E(\mathcal{Y})$

Exemple

\mathcal{X}/\mathcal{Y}	0	1	2	$P_{i\bullet}$
0	$\frac{1}{20}^{[0]}$	$\frac{1}{4}^{[0]}$	$0^{[0]}$	$\frac{3}{10}^{[0]}$
1	$\frac{17}{16}^{[0]}$	$\frac{1}{4}^{[1]}$	$\frac{1}{6}^{[2]}$	$\frac{7}{10}^{[1]}$
$P_{\bullet j}$	$\frac{1}{3}^{[0]}$	$\frac{1}{2}^{[1]}$	$\frac{1}{6}^{[2]}$	1

Information

Les indices entre crochet (" $[0]$ ") dans le tableau représente $i \times j$ (le nombre de la colonne fois celui de la ligne).

Exemple :

- colonne 0 ligne 0 (case $\frac{1}{20}$) : $0 \times 0 = 0$
- colonne 1 ligne 2 (case $\frac{1}{6}$) : $1 \times 2 = 2$

$$\begin{aligned}
 E(\mathcal{X}.\mathcal{Y}) &= \sum_{i=0}^1 \sum_{j=0}^2 i \times j \times P_{ij} \\
 &= \frac{1}{4} \times 1 + \frac{1}{6} \times 2 \\
 &= \frac{7}{12}
 \end{aligned}
 \quad \left| \quad
 \begin{aligned}
 E(\mathcal{X}) &= \frac{7}{10} \times 1 \\
 &= \frac{7}{10} \\
 E(\mathcal{Y}) &= \frac{1}{2} \times 1 + \frac{1}{6} \times 2 \\
 &= \frac{5}{6}
 \end{aligned}$$

Vérification : $E(\mathcal{X}.\mathcal{Y}) = E(\mathcal{X})E(\mathcal{Y})$

$$\frac{7}{12} = \frac{7}{10} \times \frac{5}{6} \quad \text{OK} \checkmark$$

Mais :

$$\begin{aligned}
 P((\mathcal{X} = 0) \cap (\mathcal{Y} = 2)) &= 0 \neq P(\mathcal{X} = 0).P(\mathcal{Y} = 2) = \frac{3}{10} \times \frac{1}{6} \\
 &= \frac{1}{20}
 \end{aligned}$$

$\Rightarrow \mathcal{X}$ et \mathcal{Y} ne sont pas indépendants.

Rappel

$$\begin{aligned}
 E(\mathcal{X}) &= \sum_{i \in \mathcal{I}} x_i P(\mathcal{X} = x_i) \\
 &= 0 \times \frac{3}{10} + 1 \times \frac{7}{10} \\
 &= \frac{7}{10}
 \end{aligned}$$

1.6 Covariance et coefficient de corrélation linéaire

Définition - Covariance d'un couple $(\mathcal{X}, \mathcal{Y})$

On appelle covariance du couple $(\mathcal{X}, \mathcal{Y})$:

$$COV(\mathcal{X}, \mathcal{Y}) = E(\mathcal{X} \cdot \mathcal{Y}) - E(\mathcal{X})E(\mathcal{Y})$$

et le coefficient de corrélation :

$$\rho(\mathcal{X}, \mathcal{Y}) = \frac{COV(\mathcal{X}, \mathcal{Y})}{\sigma_x \sigma_y}$$

$$(\sigma_x = \sqrt{V(\mathcal{X})}, \sigma_y = \sqrt{V(\mathcal{Y})})$$

Remarque

E : espace des v.a. (variables aléatoires)

ρ : rôle

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

$$COV(\mathcal{X}, \mathcal{Y}) = \langle \mathcal{X} - E(\mathcal{X}), \mathcal{Y} - E(\mathcal{Y}) \rangle \text{ (produit scalaire)}$$

$$\|\mathcal{X} - E(\mathcal{X})\| = \sigma(\mathcal{X})$$

$$\|\mathcal{Y} - E(\mathcal{Y})\| = \sigma(\mathcal{Y})$$

"schmas/" "angle-theta".png

$$|\rho| \leq 1$$

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

si $\rho = 0 (\theta = \frac{\pi}{2}[\pi]) \Rightarrow$ forte corrélation linéaire

1.7 Exercice 1

Soit \mathcal{X} v.a. de la loi :

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

Soit $\mathcal{Y} = \mathcal{X}^2$

1.7.1 Donner la loi du couple \mathcal{X}, \mathcal{Y}

$$P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) = 0 \text{ si } j \neq i^2$$

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

\mathcal{X}/\mathcal{Y}	0	1	4	$P_{i\bullet}$
-2	$0_{[0]}$	$0_{[-2]}$	$\frac{1}{6}_{[-8]}$	$\frac{1}{6}_{[-2]}$
-1	$0_{[0]}$	$\frac{1}{4}_{[-1]}$	$0_{[-4]}$	$\frac{1}{4}_{[-1]}$
0	$\frac{1}{6}_{[0]}$	$0_{[0]}$	$0_{[0]}$	$\frac{1}{6}_{[0]}$
1	$0_{[0]}$	$\frac{1}{4}_{[1]}$	$0_{[4]}$	$\frac{1}{4}_{[1]}$
2	$0_{[0]}$	$0_{[2]}$	$\frac{1}{6}_{[8]}$	$\frac{1}{6}_{[2]}$
$P_{\bullet j}$	$\frac{1}{6}_{[0]}$	$\frac{1}{2}_{[1]}$	$\frac{1}{3}_{[4]}$	1

1.7.2 En déduire la loi marginale de \mathcal{Y}

Loi de $\mathcal{Y} \rightarrow$ dernière ligne du tableau

1.7.3 Indépendance et calculer $COV(\mathcal{X}, \mathcal{Y})$

Contre exemple pour $\mathcal{X} = 0$ et $\mathcal{Y} = 1$.

$$P((\mathcal{X} = 0)(\mathcal{Y} = 1)) = 0 \neq P(\mathcal{X} = 0).P(\mathcal{Y} = 1)$$

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

$\Rightarrow \mathcal{X}$ et \mathcal{Y} ne sont pas indépendants.

1.7.4 $COV(\mathcal{X}, \mathcal{Y}) = E(\mathcal{X}, \mathcal{Y}) - E(\mathcal{X})E(\mathcal{Y})$

$$COV(\mathcal{X}, \mathcal{Y}) = E(\mathcal{X}, \mathcal{Y}) - E(\mathcal{X})E(\mathcal{Y})$$

$$E(\mathcal{X}, \mathcal{Y}) = \sum_{i,j} x_i \times y_j \times P_{ij}$$

$$= -\frac{1}{4} + \frac{1}{4} - \frac{8}{6} + \frac{8}{6}$$

$$= 0$$

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

1.8 Exercice 2

Soit $a \in \mathbb{R}_+^x$, \mathcal{X}, \mathcal{Y} 2 v.a. à valeurs dans \mathbb{N} .
 $P((\mathcal{X} = k) \cap (\mathcal{Y} = j)) = \frac{a}{2^{k+1}(j!)} \forall (k, j) \in \mathbb{N}^2$

1.8.1 Déterminer a

$$\sum_{k,j} P((\mathcal{X} = 4) \cap (\mathcal{Y} = j)) = 1$$

Rappel - e^x

$$e^x = \sum_{j=0}^{+\infty} \frac{x^j}{j!} (\forall x \in \mathbb{R})$$

$$\text{si on remplace } x \text{ par } 1 \Rightarrow e = \sum_{j=0}^{+\infty} \frac{1}{j!}$$

Rappel - Série Géométrique

$$\sum_{n=0}^{+\infty} a^n = \frac{1}{1-a} \text{ si } |a| < 1$$

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

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

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

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

$$\Rightarrow a \frac{e}{2} \left(\frac{1}{1 - \frac{1}{2}} \right) = 1$$

$$\Rightarrow a e = 1$$

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

1.8.2 Déterminer les lois marginales de \mathcal{X} et \mathcal{Y}

Loi marginale de \mathcal{X} :

$$\begin{aligned}
 P(\mathcal{X} = k) &= \sum_{j=0}^{+\infty} P((\mathcal{X} = k) \cap (\mathcal{Y} = j)) \\
 &= \frac{e^{-1}}{2^{k+1}} \sum_{j=0}^{+\infty} \frac{1}{j!} \\
 &= \frac{e^{-1} \times e}{2^{k+1}} \\
 &= \frac{1}{2^{k+1}} (\forall k \in \mathbb{N})
 \end{aligned}$$

Loi marginale de \mathcal{Y} :

$$\begin{aligned}
 P(\mathcal{Y} = j) &= \sum_{k=0}^{+\infty} P((\mathcal{X} = k) \cap (\mathcal{Y} = j)) \\
 &= \frac{e^{-1}}{j!} \sum_{k=0}^{+\infty} \frac{1}{2^{k+1}} \\
 &= \frac{e^{-1}}{j! \times 2} \left(\frac{1}{1 - \frac{1}{2}} \right) \\
 &= \frac{e^{-1}}{j!}
 \end{aligned}$$

1.8.3 Indépendance ?

Il faut montrer que le produit des lois marginales est égale à la loi conjointe. En d'autres termes : $P(\mathcal{X} = k) \times P(\mathcal{Y} = j) = P((\mathcal{X} = k) \cap (\mathcal{Y} = j))$.

$$\frac{1}{2^{k+1}} \times \frac{e^{-1}}{j!} = \frac{e^{-1}}{2^{k+1} \times j!} \Rightarrow \text{OK } \checkmark$$

Les deux variables sont indépendantes.

Remarque

\mathcal{X}, \mathcal{Y} sont indépendantes $\Rightarrow E(\mathcal{X}, \mathcal{Y}) = E(\mathcal{X})E(\mathcal{Y})$

1.8.4 Calculer $COV(\mathcal{X}, \mathcal{Y})$

$$COV(\mathcal{X}, \mathcal{Y}) = E(\mathcal{X}, \mathcal{Y}) - E(\mathcal{X}) \times E(\mathcal{Y}) = 0$$

1.9 Exercice 3

On considère n boîtes numérotées de 1 à n . La boîte numéro k contient k boules numérotées de 1 à k . On choisit une boîte au hasard, puis on choisie une boule dans cette boîte.

\mathcal{X} v.a. : numéro de la boîte

\mathcal{Y} v.a. : numéro de la boule

1.9.1 Déterminer la loi du couple $(\mathcal{X}, \mathcal{Y})$

Valeurs possible de \mathcal{X} et \mathcal{Y} :

\mathcal{X}	\mathcal{Y}
1	1
2	1
2	2
3	1
3	2
3	3

$$\mathcal{X}(\Omega) = \mathcal{Y}(\Omega) = \llbracket 1, n \rrbracket$$

$$P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) = P(\mathcal{Y} = j | \mathcal{X} = i) \times P(\mathcal{X} = i) (\forall (i, j) \in \llbracket 1, n \rrbracket^2)$$

$$\text{1^{er} cas : Quand } j > i \Rightarrow P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) = 0$$

$$\text{2^{ème} cas : Quand } j \leq i \Rightarrow P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) = \frac{1}{i} \times \frac{1}{n}$$

1.9.2 Calculer $P(\mathcal{X} = \mathcal{Y})$

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

Remarque

$(\mathcal{X} = i) \cap (\mathcal{Y} = i)$ est un évènement incompatible.

$$\begin{aligned}\mathcal{P}(\mathcal{X} = \mathcal{Y}) &= \sum_{i=1}^n P((\mathcal{X} = i) \cap (\mathcal{Y} = i)) \\ &= \sum_{i=1}^n \frac{1}{i \times n} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{i}\end{aligned}$$

1.9.3 Donner la loi marginale de \mathcal{Y} et $E(\mathcal{Y})$

$$\begin{aligned}\mathcal{P}(\mathcal{Y} = j) &= \sum_{i=1}^n P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) \\ &= \sum_{i \geq j}^n \frac{1}{i \times n} (\forall j \in \llbracket 1, n \rrbracket) \\ &= \frac{1}{n} \sum_{i \geq j}^n \frac{1}{i}\end{aligned}$$

Remarque - Suite Arithmétique

$$\sum_{j=1}^i j = \frac{i(i+1)}{2}$$

$$\begin{aligned}
 E(\mathcal{Y}) &= \sum_{j=1}^n j P(\mathcal{Y} = j) \\
 &= \sum_{j=1}^n \sum_{i \geq j}^n \frac{1}{i \times n} \\
 &= \frac{1}{n} \sum_{j=1}^n \left(j \sum_{i \geq j}^n \frac{1}{i} \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{i} \sum_{j=1}^i j \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{i} \times \frac{i(i+1)}{2} \right) \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{i+1}{2} \\
 &= \frac{1}{2n} \left(\frac{n(n+1)}{2} + n \right) \\
 &= \frac{1}{2} \left(\frac{n+1}{2} + 1 \right) \\
 &= \boxed{\frac{n+3}{4}}
 \end{aligned}$$

1.10 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. On répète cette expérience n fois ($n \geq 2$). c est une constante fixée qui ne change pas.

$$\mathcal{X}_i \begin{cases} 1 & \text{si on obtiens une boule blanche au } i^{\text{ème}} \text{ tirage} \\ 0 & \text{sinon} \end{cases}$$

1.10.1 La loi du couple $(\mathcal{X}_1, \mathcal{X}_2)$ et en déterminer la loi de \mathcal{X}_2

$$\begin{aligned}
 P(\mathcal{X}_1 = 1) &= P(\mathcal{X}_1 = 0) = \frac{1}{2} \\
 \mathcal{X}_i &\text{ suis la variable de Bernoulli } \mathcal{B}\left(\frac{1}{2}\right) \\
 P((\mathcal{X}_1 = i) \cap (\mathcal{X}_2 = j)) &= P(\mathcal{X}_2 = j / \mathcal{X}_1 = i) \times P(\mathcal{X}_1 = i)
 \end{aligned}$$

$$\begin{aligned}
 &\text{1er cas } i \neq j : P((\mathcal{X}_1 = i) \cap (\mathcal{X}_2 = j)) = \frac{1}{c+2} \times \frac{1}{2} \\
 &\underline{\text{2ème cas } i = j :}
 \end{aligned}$$

$$P((\mathcal{X}_1 = i) \cap (\mathcal{X}_2 = i)) = P(\mathcal{X}_2 = i / \mathcal{X}_1 = i) \times P(\mathcal{X}_1 = i) \\ = \frac{1+c}{2+c} \times \frac{1}{2}$$

$\mathcal{X}_1 / \mathcal{X}_2$	0	1	loi \mathcal{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 \mathcal{X}_2	$\frac{1}{2}$	$\frac{1}{2}$	1

D'après le tableau : \mathcal{X}_2 suis la variable $\mathcal{B}(\frac{1}{2})$ (Bernoulli).

1.10.2 Déterminer la loi de \mathcal{Z}_2

$$\mathcal{Z}_p = \sum_{i=1}^p \mathcal{X}_i \quad (2 \leq p \leq n)$$

\mathcal{Z}_p : nombre de boules blanches obtenue lors des premiers tirages.

$$\mathcal{Z}_2 = \mathcal{X}_1 + \mathcal{X}_2$$

$$\mathcal{Z}_2(\Omega) = \{0, 1, 2\}$$

$$P(\mathcal{Z}_2 = 0) = P((\mathcal{X}_1 = 0) \cap (\mathcal{X}_2 = 0)) = \frac{1+c}{2(2+c)}$$

$$P(\mathcal{Z}_2 = 1) = P((\mathcal{X}_1 = 0) \cap (\mathcal{X}_2 = 1)) + P((\mathcal{X}_1 = 1) \cap (\mathcal{X}_2 = 0)) \\ = \frac{1}{2+c}$$

$$P(\mathcal{Z}_2 = 2) = P((\mathcal{X}_1 = 1) \cap (\mathcal{X}_2 = 1)) \\ = \frac{1+c}{2(2+c)}$$

1.10.3 Déterminer $\mathcal{Z}_p(\Omega)$ et calculer $P(\mathcal{X}_{p+1} = 1 / \mathcal{Z}_p = k)$

$$P(\mathcal{X}_{p+1} = 1 / \mathcal{Z}_p = k) \text{ et } \mathcal{Z}_p = \sum_{i=1}^p \mathcal{X}_i$$

$(\mathcal{Z}_p = k) \Leftrightarrow$ Evenement « Au cours des p premier tirage on a obtenue k boules blanches et $(p - k)$ boules noires. »

Avant de passer au $(p + 1)^{\text{ième}}$ tirage l'urne contiens : $kc + (p - k)c + 2 = 2 + pc$ dont $1 + kc$ boules blanches.

$$\text{Ainsi : } P(\mathcal{X}_{p+1} = 1 / \mathcal{Z}_p = k) = \frac{1+kc}{2+pc}$$

1.10.4 Montrer que $P(\mathcal{X}_{p+1} = 1) = \frac{1+cE(\mathcal{Z}_p)}{2+pc}$

$$(\mathcal{X}_{p+1} = 1) = \bigcup_{k=0}^p ((\mathcal{X}_{p+1} = 1) \cap (\mathcal{Z}_p = k))$$

Remarque

$(\mathcal{X}_{p+1} = 1) \cap (\mathcal{Z}_p = k)$ est un évènement incompatible.

$$\begin{aligned} P(\mathcal{X}_{p+1} = 1) &= \sum_{k=0}^p P((\mathcal{X}_{p+1} = 1) \cap (\mathcal{Z}_p = k)) \\ &= \sum_{k=0}^p P(\mathcal{X}_{p+1} = 1 / \mathcal{Z}_p = k) \times P(\mathcal{Z}_p = k) \\ &= \sum_{k=0}^p \left(\frac{1 + kc}{2 + pc} \right) \times P(\mathcal{Z}_p = k) \\ &= \frac{1}{2 + pc} \times \left(\sum_{k=0}^p P(\mathcal{Z}_p = k) + c \sum_{k=0}^p k P(\mathcal{Z}_p = k) \right) \\ &= \frac{1 + cE(\mathcal{Z}_p)}{2 + pc} \end{aligned}$$

1.10.5 Montrer que $P(\mathcal{X}_p = 1) = P(\mathcal{X}_p = 0) = \frac{1}{2} \quad (\forall p \in \llbracket 1, n \rrbracket)$

Rappel

$$E(\mathcal{X}) = \sum_k k P(\mathcal{X} = k)$$

Soit R_p la propriété : $P(\mathcal{X}_p = 1) = P(\mathcal{X}_p = 0) = \frac{1}{2}$.
 R_1 et R_2 sont vérifiées (1^{ère} question).

Rappel - Bernoulli

Soit \mathcal{X}_i suis $\mathcal{B}(q)$
 Alors : $E(\mathcal{X}_i) = q, V(\mathcal{X}_i) = q(1 - q)$

Hypothèse : Supposons R_k vraie $\forall 1 \leq k \leq p$

$$\begin{aligned} P(\mathcal{X}_{p+1} = 1) &= \frac{1+cE(\mathcal{Z}_p)}{2+pc} \\ \mathcal{Z}_p = \sum_{i=1}^p \mathcal{X}_i &\Rightarrow E(\mathcal{Z}_p) = \sum_{i=1}^p E(\mathcal{X}_i) \end{aligned}$$

$$P(\mathcal{X}_{p+1} = 1) = \frac{1 + \frac{cp}{2}}{2 + pc} = \frac{1}{2}$$

$$P(\mathcal{X}_{p+1} = 0) = 1 - \frac{1}{2} = \boxed{\frac{1}{2}}$$

Conclusion : R_p est vraie $\forall 1 \leq p \leq n$

1.11 Exercice 5

Une urne contient des boules noires en proportion p ($1 \leq p \leq 1$) et des boules blanches en proportion $q = 1 - p$. On effectue une suite de tirage d'une boule avec remise.

1.11.1 Question 1 : loi marginale et indépendance

On note N le rang aléatoire d'apparition de la première boule noire et B celui de la première boule blanche.

a) Déterminer les lois de N et B , $E(N)$, $V(N)$, $E(B)$, $V(B)$

- $\left\{ \begin{array}{l} N : \text{temps d'attente de la réalisation de la 1}^{\text{ère}} \text{ boule noire} \\ B : \text{temps d'attente de la réalisation de la 1}^{\text{ère}} \text{ boule blanche} \end{array} \right.$

N suit la loi géométrique de paramètre p (autrement dit : $N \nearrow \mathcal{G}(p)$)

$$P(N = k) = (1 - p)^{k-1} \times p$$

$$N(\Omega) = \llbracket 1, +\infty \rrbracket$$

$$B(\Omega) = \llbracket 1, +\infty \rrbracket$$

$$P(B = k) = (1 - q)^{k-1} \times q$$

$$B \nearrow \mathcal{G}(q) \text{ pour } q = 1 - p$$

$$E(N) = \frac{1}{p} \quad V(N) = \frac{q}{p^2}$$

$$E(B) = \frac{1}{q} \quad V(B) = \frac{p}{q^2}$$

$$P(N = k) = q^{k-1} \times p \quad (\forall k \in \mathbb{N}^*)$$

$$P(B = k) = p^{k-1} \times q \quad (\forall k \in \mathbb{N}^*)$$

b) N et B sont-elles indépendantes ?

$$P((N = 1) \cap (B = 1)) = 0 \quad (N = 1) \cap (B = 1) = \emptyset$$

$$P(N = 1) \times P(B = 1) = pq \neq 0$$

Donc N et B ne sont pas indépendantes.

1.11.2 Question 2 : loi conjointe

On note \mathcal{X} la longueur de la 1^{ère} suite de boule de a même couleur et \mathcal{Y} celle de la 2^{ème} suite de boule de la même couleur.

Exemple : $(\mathcal{X} = 1) \cap (\mathcal{Y} = 2) : (B_1 \cap N_2 \cap N_3 \cap B_4) \cup (N_1 \cap B_2 \cap B_3 \cap N_4)$

a) Déterminer la loi conjointe de $(\mathcal{X}, \mathcal{Y})$

$$\mathcal{X}(\Omega) = \mathcal{Y}(\Omega) = \mathbb{N}^*$$

$$(\mathcal{X} = i) \cap (\mathcal{Y} = j) :$$

$$(N_1 \cap \dots \cap N_i \cap B_{i+1} \cap \dots \cap B_{i+j} \cap N_{i+j+1}) \cup (B_1 \cap \dots \cap B_i \cap N_{i+1} \cap \dots \cap N_{i+j} \cap B_{i+j+1})$$

$$P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) = p^{i+1} \times q^j + q^{i+1} \times p^j$$

b) Loi de \mathcal{X} ? $E(\mathcal{X})$ et $E(\mathcal{X}) \geq 2$

Rappel

$$\sum_{j=0}^{+\infty} \mathcal{X}^j = \frac{1}{1-\mathcal{X}} \quad |\mathcal{X}| < 1$$

$$\begin{aligned} \forall i \in \mathbb{N}^* \quad P(\mathcal{X} = i) &= \sum_{j=1}^{+\infty} (p^{i+1} \times q^j + q^{i+1} \times p^j) \\ &= p^{i+1} \sum_{j=1}^{+\infty} q^j + q^{i+1} \sum_{j=1}^{+\infty} p^j \\ &= \frac{p^{i+1}q}{1-q} + \frac{q^{i+1}p}{1-p} \\ &= p^i q + q^i p \end{aligned}$$

Rappel

$$\begin{aligned} f(p) &= \sum_{n=0}^{+\infty} p^n = \frac{1}{1-p} \quad (|p| < 1) \\ f'(p) &= \sum_{n=1}^{+\infty} n p^{n-1} = \frac{1}{(1-p)^2} \end{aligned}$$

$$P(\mathcal{X} = i) = p^i q + q^i p$$

$$E(\mathcal{X}) = \sum_{n \geq 1}^{+\infty} n \times P(\mathcal{X} = n)$$

$$= \sum_{n \geq 1}^{+\infty} (np^n q + nq^n p)$$

$$= q \sum_{n \geq 1}^{+\infty} np^n + p \sum_{n \geq 1}^{+\infty} nq^n$$

$$= \frac{qp}{(1-p)^2} + \frac{pq}{(1-q)^2}$$

$$= \frac{p}{q} + \frac{q}{p}$$

$$= \frac{p^2 + q^2}{pq}$$

$$\text{or } (p - q)^2 \geq 0 \Rightarrow p^2 + q^2 - 2pq \geq 0$$

$$\Rightarrow \frac{p^2 + q^2}{pq} \geq 2$$

$$\Rightarrow E(\mathcal{X}) \geq 2$$

c) Loi de \mathcal{Y} ? $E(\mathcal{Y})$ et $E(\mathcal{Y})$

$$\forall j \in \mathbb{N}^* \quad P(\mathcal{Y} = j) = \sum_{i=1}^{+\infty} (p^{i+1} q^j + q^{i+1} p^j)$$

$$= q^j p \sum_{i=1}^{+\infty} p^i + qp^j \sum_{i=1}^{+\infty} q^i$$

$$= q^j p^2 \frac{1}{1-p} + \frac{q^2 p^j}{1-q}$$

$$= p^2 q^{j-1} + q^2 p^{j-1}$$

$$= p^2 q^{j-1} + q^2 p^{j-1}$$

$$\begin{aligned}
 E(\mathcal{Y}) &= \sum_{n \geq 1} np^2 q^{n-1} + \sum_{n \geq 1} nq^2 p^{n-1} \\
 &= p^2 \sum_{n \geq 1} nq^{n-1} + q^2 \sum_{n \geq 1} np^{n-1} \\
 &= \frac{p^2}{(1-q)^2} + \frac{q^2}{(1-p)^2} \\
 &= 1 + 1 \\
 &= 2
 \end{aligned}$$

$$V(\mathcal{Y}) = E(\mathcal{Y}^2) - E^2(\mathcal{Y})$$

Rappel

$$\sum_{n \geq 1} np^n = \frac{p}{(1-p)^2} \quad (\text{en dérivant cette fonction en fonction de } p).$$

$$\begin{aligned}
 \sum_{n \geq 1} n^2 p^{n-1} &= \frac{(1-p)^2 + 2p(1-p)}{(1-p)^4} \\
 &= \frac{1-p+2p}{(1-p)^3} \\
 &= \frac{1+p}{(1-p)^3}
 \end{aligned}$$

$$\begin{aligned}
 E(\mathcal{Y}^2) &= \sum_{n \geq 1} n^2 P(\mathcal{Y} = n) \\
 &= \sum_{n \geq 1} n^2 p^2 q^{n-1} + \sum_{n \geq 1} n^2 q^2 p^{n-1} \\
 &= p^2 \sum_{n \geq 1} n^2 q^{n-1} + q^2 \sum_{n \geq 1} n^2 p^{n-1} \\
 &= p^2 \left(\frac{1+q}{(1-q)^3} \right) + q^2 \left(\frac{1+p}{(1-p)^3} \right) \\
 &= \frac{1+q}{p} + \frac{1+p}{q} \\
 &= \frac{2q}{p} + \frac{2p}{q} + 2
 \end{aligned}$$

$$\begin{aligned}
 V(\mathcal{Y}) &= \frac{2q}{p} + \frac{2p}{q} + 2 - 4 \\
 &= 2\left(\frac{q}{p} + \frac{p}{q} - 1\right)
 \end{aligned}$$

d) Calculer $P(\mathcal{X} = \mathcal{Y})$

$$P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) = p^{i+1}q^j + q^{i+1}p^j$$

$$(\mathcal{X} = \mathcal{Y}) = \bigcup_{n=1}^{+\infty} ((\mathcal{X} = n) \cap (\mathcal{Y} = n))$$

Note

$(\mathcal{X} = n) \cap (\mathcal{Y} = n)$ est in évènement indépendant.

$$\begin{aligned} P(\mathcal{X} = \mathcal{Y}) &= \sum_{n=1}^{+\infty} P((\mathcal{X} = n) \cap (\mathcal{Y} = n)) \\ &= \sum_{n=1}^{+\infty} (p^{n+1}q^n + q^{n+1}p^n) \\ &= p \sum_{n=1}^{+\infty} (pq)^n + q \sum_{n=1}^{+\infty} (pq)^n \\ &= p^2q \left(\frac{1}{1-pq} \right) + pq^2 \left(\frac{1}{1-pq} \right) \\ &= \frac{p^2q + pq^2}{1-pq} \\ &= \frac{pq(p+q)}{1-pq} \\ &= \boxed{\frac{pq}{1-pq}} \end{aligned}$$

Remarque

Dans le cas ou on a 2 variables souvent on obtiens une série.

e) Calculer la loi de $\mathcal{X} + \mathcal{Y}$ ($p = \frac{1}{2}$)

$$\mathcal{X}(\Omega) = \mathcal{Y}(\Omega) = \mathbb{N}^*$$

Note

$$\sum_{n=0}^{+\infty} p^n = \frac{1}{1-p}$$

$$(\mathcal{X} + \mathcal{Y})(\Omega) = \llbracket 2, +\infty \llbracket$$

On cherche à prouver : $\forall k \in \llbracket 2, +\infty \rrbracket \quad P(\mathcal{X} + \mathcal{Y} = k)$

$$\begin{aligned}
 P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) &= \frac{1}{2^{i+1}} \times \frac{1}{2^j} + \frac{1}{2^{i+1}} \times \frac{1}{2^j} \\
 &= \left(\frac{1}{2}\right)^{i+j} \\
 P(\mathcal{X} + \mathcal{Y} = k) &= \sum_{(i,j)/i+j=k} P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) \\
 &= \sum_{i=1}^{k-1} P((\mathcal{X} = i) \cap (\mathcal{Y} = k-i)) \\
 &= \sum_{i=1}^{k-1} \left(\frac{1}{2}\right)^k \\
 &= (k-1) \left(\frac{1}{2}\right)^k
 \end{aligned}$$

1.12 Exercice 6

Soit $a \in]0, 1[$ et $b \in]0, +\infty[$. \mathcal{X} et \mathcal{Y} deux v.a. dont la loi conjointe est :

$$P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) \begin{cases} 0 & i < j \\ \frac{b^i e^{-b} a^j (1-a)^{i-j}}{j!(i-j)!} & i \geq j \end{cases}$$

$\mathcal{X}(\Omega) = \mathcal{Y}(\Omega) = \mathbb{N}$

1.12.1 Déterminer les lois marginales ainsi que $E(\mathcal{X})$, $V(\mathcal{X})$, $E(\mathcal{Y})$, $V(\mathcal{Y})$

Note

$$\binom{i}{j} = \frac{i!}{j!(i-j)!}$$

Rappel - Formule du Binôme de Newton

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

$$\begin{aligned}
 P(\mathcal{X} = i) &= \sum_{j \in \mathbb{N}} P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) \\
 &= \sum_{j=0}^i \frac{b^i e^{-b} a^j (1-a)^{i-j}}{j!(i-j)!} \\
 &= b^i e^{-b} \sum_{j=0}^i \frac{a^j (1-a)^{i-j}}{j!(i-j)!} \\
 &= \frac{b^i e^{-b}}{i!} \sum_{j=0}^i \binom{i}{j} a^j (1-a)^{i-j} \\
 &= \frac{b^i e^{-b}}{i!} (a + 1 - a)^i \quad (\text{Formule du Binôme de Newton}) \\
 &= e^{-b} \frac{b^i}{i!} \quad \forall i \in \mathbb{N}
 \end{aligned}$$

Rappel - Loi de Poisson

On dit que \mathcal{X} suit $\mathcal{P}(\lambda)$ la loi de Poisson :

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}$$

$$E(\mathcal{X}) = V(\mathcal{X}) = \lambda$$

$$\mathcal{X} \nearrow \mathcal{P}(b) \text{ ainsi } E(\mathcal{X}) = V(\mathcal{X}) = \lambda \quad \forall j \in \mathcal{N}$$

Rappel

$$e^n = \sum_{k=0}^{+\infty} \frac{n^k}{k!} \quad \forall n \in \mathbb{R}$$

$$\begin{aligned}
 P(\mathcal{Y} = j) &= \sum_{i \in \mathbb{N}} P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) \\
 &= \sum_{i=j}^{+\infty} \frac{b^i e^{-b} a^j (1-a)^{i-j}}{j!(i-j)!} \\
 &= \frac{e^{-b} a^j}{j!} \sum_{i=j}^{+\infty} \frac{b^i (1-a)^{i-j}}{(i-j)!} \\
 &= \frac{e^{-b} (ab)^j}{j!} \sum_{i=j}^{+\infty} \frac{(b(1-a))^{i-j}}{(i-j)!} \\
 &= \frac{e^{-b} (ab)^j}{j!} \times e^{b(1-a)} \\
 &= \boxed{\frac{(ab)^j}{j!} \times e^{-ab}}
 \end{aligned}$$

Donc $\mathcal{Y} \nearrow \mathcal{P}(ab)$.

1.12.2 \mathcal{X} et \mathcal{Y} sont-elles indépendantes ?

$\mathcal{X} \nearrow \mathcal{P}(b) \quad P((\mathcal{X} = 0) \cap (\mathcal{Y} = 1)) = 0$
 $P(\mathcal{X} = 0) \times P(\mathcal{Y} = 1) = e^{-b} a b e^{-ab} \neq 0$
 \mathcal{X} et \mathcal{Y} ne sont pas indépendantes.

1.12.3 Déterminer la loi de $\mathcal{Z} = \mathcal{X} - \mathcal{Y}$

$$\begin{aligned}
 P(\mathcal{Z} = k) &= \sum_{(i,j)/i-j=k} P((\mathcal{X} = i) \cap (\mathcal{Y} = j)) \\
 &= \sum_{i \geq k}^{+\infty} P((\mathcal{X} = i) \cap (\mathcal{Y} = i - k)) \\
 &= \sum_{i=k}^{+\infty} \frac{b^i e^{-b} a^{i-k} (1-a)^k}{(i-k)! k!} \\
 &= \frac{e^{-b} (1-a)^k}{k!} \sum_{i=k}^{+\infty} \frac{b^i a^{i-k}}{(i-k)!} \\
 &= \frac{e^{-b} ((1-a)b)^k}{k!} \sum_{i=k}^{+\infty} \frac{(ab)^{i-k}}{(i-k)!} \\
 &= \frac{e^{-b} ((1-a)b)^k}{k!} \times e^{ab} \\
 &= e^{-b(1-a)} \frac{((1-a)b)^k}{k!}
 \end{aligned}$$

$$\mathcal{Z} \nearrow \mathcal{P}((1-a)b).$$

1.12.4 \mathcal{Y} et \mathcal{Z} sont-elle indépendantes ?

$$\mathcal{Y} \nearrow \mathcal{P}(ab), \mathcal{Z} \nearrow \mathcal{P}((1-a)b) \quad \forall (j, k) \in \mathbb{N}$$

$$\begin{aligned}
 P((\mathcal{Y} = j) \cap (\mathcal{Z} = k)) &= P((\mathcal{Y} = j) \cap (\mathcal{X} - \mathcal{Y} = k)) \\
 &= P((\mathcal{Y} = j) \cap (\mathcal{X} = j + k)) \\
 &= \frac{b^{j+k} e^{-b} a^j (1-a)^k}{j! k!}
 \end{aligned}$$

$$\begin{aligned}
 P(\mathcal{Y} = j) \times P(\mathcal{Z} = k) &= e^{-ab} \frac{(ab)^j}{j!} \times e^{(1-a)b} \frac{((1-a)b)^k}{k!} \\
 &= \frac{(ab)^j e^{-b} ((1-a)b)^k}{j! k!}
 \end{aligned}$$

2 Analyse en composante principale

2.1 Les données et leurs caractéristiques

2.1.1 Tableau de donnée

Les observations de p variables sur n individus sont rassemblé en une matrice X à n ligne et p colonne :

$$X = \begin{pmatrix} X_1^{(1)} & \dots & X_1^{(j)} & \dots & X_1^{(p)} \\ \vdots & & \vdots & & \vdots \\ x_i^{(1)} & \dots & X_i^{(j)} & \dots & X_i^{(p)} \\ \vdots & & \vdots & & \vdots \\ X_n^{(1)} & \dots & X_n^{(j)} & \dots & X_n^{(p)} \end{pmatrix}$$

$X_i^{(j)}$: valeurs prises par la valeur $X^{(j)}$ sur le $i^{\text{ème}}$ individu.

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

$t_{e_i} = (X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(p)})$ $i^{\text{ème}}$ individu.

2.1.2 Matrice de poids

On associe à chaque individu un poids p_i ($p_i \geq 0$) (probabilité de choisir l'individu i).

$$\sum_{i=1}^n p_i = 1 \quad D = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix}$$

Si $p_i = \frac{1}{n} \quad \forall i \quad D = \frac{1}{n} I_n$ (I_n : Matrice Identité)

2.1.3 Centre de gravité

Le vecteur $g : t_g = (\bar{X}^{(1)}, \bar{X}^{(2)}, \dots, \bar{X}^{(p)})$

$$\bar{X}^{(j)} = \sum_{i=1}^n p_i X_i^{(j)} \quad \text{Moyenne arithmétique de } X^{(j)}.$$

Le tableau des données centré : Y

$$Y_i^{(j)} = X_i^{(j)} - \bar{X}^{(j)}$$

2.1.4 Matrice de Variance - Covariance et Matrice de Corrélation

Définition - Matrice de Variance-Covariance

On appelle matrice de variance-covariance : $V = {}^t Y D Y$

Si on note $D_{\frac{1}{S}}$ la matrice diagonale des inverses des écarts-types :

$$D_{\frac{1}{S}} = \begin{pmatrix} \frac{1}{S_1} & & 0 \\ & \frac{1}{S_2} & \\ 0 & & \ddots \\ & & & \frac{1}{S_p} \end{pmatrix}$$

où $S_j = \sqrt{V(X^{(j)})}$ écart type

$$= \sum_{i=1}^n p_i \left(Y_i^{(j)} \right)^2$$

$V(X^{(j)})$: variance de $X^{(j)}$

S_j : écart-type de $X^{(j)}$

On appelle la matrice des données centrées et réduites Z telle que : $Z_i^{(j)} = \frac{Y_i^{(j)}}{S_j}$

Matriciellement : $Z = Y D_{\frac{1}{S}}$

La matrice regroupant tous les coefficients de corrélation linéaire entre les p variables est R :

$$R = \begin{pmatrix} 1 & & l_{ij} \\ & \ddots & \\ l_{ij} & & 1 \end{pmatrix} \text{ symétriques}$$

$$r_{ij} = l_{ij} = \frac{COV(X^{(i)}, X^{(j)})}{S_i S_j} \quad (l_{ij} : \text{coefficient de corrélation})$$

où $COV(X^{(i)}, X^{(j)})$: covariance

$$COV(X^{(i)}, X^{(j)}) = \sum_{k=1}^n l_k \cdot Y_k^{(i)} \cdot Y_k^{(j)} \quad (\text{produit scalaire des variables centrées})$$

Remarque

$$\begin{aligned} R &= D_{\frac{1}{S}} V D_{\frac{1}{S}} \\ &= D_{\frac{1}{S}} {}^t Y D Y D_{\frac{1}{S}} \\ \Rightarrow \boxed{R = {}^t Z D Z} \end{aligned}$$

2.2 Espace des individus

Chaque individus étant un vecteur défini par p coordonnées est considéré comme un élément d'un espace vectoriel F appelé l'espace des individus.

Les n individus forment alors un nuage de points dans F et g en est le barycentre (ou centre de gravité).

On munit l'espace F d'une métrique (distance) :

$\langle e_i, e_j \rangle = {}^t e_i M l_j$ ($\langle x, y \rangle$: produit scalaire) où M est une matrice S.D.P (Symétrique et Définie Positive).

Remarque

Si $M = I$ (I : Identité), on retrouve le produit scalaire usuel.

Si $M = D_{\frac{1}{S^2}} = \begin{pmatrix} \frac{1}{S_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{S_p^2} \end{pmatrix}$ cela revient à diviser chaque caractère par son écart-type.

Définition - Inertie

On appelle inertie totale du nuage de points la moyenne pondérée des carrés des distances des points au centre de gravité :

$$\begin{aligned} I_g &= \sum_{i=1}^n p_i {}^t(e_i - g)M(e_i - g) \\ &= \sum_{i=1}^n p_i \|e_i - g\|^2 \end{aligned}$$

Propriété - Inertie

On peut montrer que l'inertie du nuage est égale à le trace de la matrice MV :

$$I_g = \text{Trace}(MV) = \text{Trace}(VM)$$

2.3 Espace des variables

On note E : l'espace des variables $X^{(j)} = \begin{pmatrix} X_1^{(j)} \\ \vdots \\ X_n^{(j)} \end{pmatrix}$

On munit E de la métrique $M = D$ (D : matrice de poids)
 $\langle X^{(j)}, X^{(k)} \rangle = {}^t X^{(j)} D X^{(k)}$

Si les variables sont centrées on a :

$$\begin{aligned} {}^t X^{(j)} D X^{(k)} &= \sum_{i=1}^n p_i X_i^{(j)} X_i^{(k)} \\ &= \text{COV}(X^{(j)}, X^{(k)}) \end{aligned}$$

La norme de $X^{(j)}$ (variable centrée)

$$\|X^{(j)}\|^2 = \langle X^{(j)}, X^{(j)} \rangle$$

$$= \sum_{i=1}^n p_i (X_i^{(j)})^2$$

$$= S_j^2$$

$$\Rightarrow \|X^{(j)}\| = S_j \quad (\text{écart-type})$$

On mesure l'angle entre 2 variables $X^{(j)}$ et $X^{(k)}$ (centrées) :

$$\begin{aligned}\cos(\theta_{jk}) &= \frac{\langle X^{(j)}, X^{(k)} \rangle}{\|X^{(j)}\| \|X^{(k)}\|} \\ &= \frac{COV(X^{(j)}, X^{(k)})}{S_j S_k} \\ &= l_{jk}\end{aligned}$$

On retrouve le coefficient de corrélation linéaire.

2.4 Variables engendrées par un tableau de données

(image : Numériser 55.jpeg)

2.5 Algorithme A.C.P. (Analyse Composante Principale)

2.6 Projection