Lycée Henri IV — **Cours**

MPSI — Durée : 4 h 15

VECTEURS ALÉATOIRES FINIS

Les énoncés et parties suivis du symbole [✠] ne seront pas traités en cours.

Prérequis

Revoir le chapitre sur :

- les sommes ;
- les probabilités ;
- les variables aléatoires.

Dans tout ce chapitre, on considère un espace probabilisé fini $(\Omega, \mathscr{T}, P)$.

Les variables aléatoires considérées dans ce chapitre sont à valeurs dans un ensemble quelconque (sauf si celui-ci est précisé). En particulier, lorsqu'on considère plusieurs variables aléatoires, celles-ci peuvent prendre leurs valeurs dans des ensembles différents.

De même, lorsque l'ensemble d'arrivée d'une fonction n'est pas précisé, c'est que celui-ci est quelconque.

La lettre n désigne un entier naturel non nul. La lettre p désigne un nombre réel de l'intervalle $[0; 1]$.

Les trois lettres AQT signifient « Âne Qui Trotte » et sont utilisées pour désigner une démonstration facile laissée au lecteur.

a. Vecteurs aléatoires finis

Définition 1

On appelle vecteur aléatoire (ou $\overrightarrow{\text{v.a.}}$ en abrégé) tout n-uplet $\overrightarrow{U} = (X_1, X_2, \ldots, X_n)$ tel que, pour tout $k \in [\![1;n]\!]$, X_k soit une variable aléatoire sur $(\Omega, \mathscr{T})$.

Dans le cas où $n = 2$, on dit que $\overrightarrow{U} = (X, Y)$ est un couple de variables aléatoires.

Se donner un vecteur aléatoire revient à considérer simultanément plusieurs variables aléatoires attachées à la même expérience aléatoire. Un vecteur aléatoire n'est donc rien d'autre qu'une variable aléatoire à valeurs dans un produit cartésien.

Dans la suite de ce cours, nous donnerons la plupart des énoncés et exemples dans le cadre des couples de variables aléatoires, laissant au lecteur le soin de généraliser au cas des dimensions supérieures ou égales à 3.

Comme dans le cas des variables aléatoires, l'univers image $(X, Y)(\Omega)$ des valeurs prises par le couple aléatoire (X, Y) a plus d'importance que l'univers Ω.
Dans la pratique, on préfère souvent déterminer $X(\Omega) \times Y(\Omega)$ (parce que c'est plus simple) mais il faut garder en tête que $(X, Y)(\Omega)$ n'est qu'une partie de $X(\Omega) \times Y(\Omega)$.

En première année, l'univers Ω étant toujours fini, l'univers image l'est également. On parle alors de vecteur aléatoire fini.

Nous considérerons souvent des événements définis par une conjonction de conditions portant sur X et Y. Pour abréger la notation $(X \in A) \cap (X \in B)$, nous noterons $(X \in A\,, X \in B)$ l'événement conjoint « $X \in A$ et $X \in B$ » ; la virgule désignant donc le $\boxed{\text{et}}$ logique.

Exemples :

- Une variable aléatoire est un vecteur aléatoire avec un seul élément.
- Un vecteur aléatoire dont toutes les coordonnées sont des variables aléatoires certaines est dit certain. C'est bien-sûr un exemple de vecteur aléatoire fini.
- On tire au hasard simultanément deux numéros dans $[\![1;n]\!]$ et l'on note X le plus petit et Y le plus grand. On a $X(\Omega) = [\![1;n-1]\!]$ et $Y(\Omega) = [\![2;n]\!]$ donc $X(\Omega) \times Y(\Omega)$ est le carré $[\![1;n-1]\!] \times [\![2;n]\!]$. L'univers image est plus petit puisqu'il faut tenir compte du fait que $X < Y$, ce qui donne $(X, Y)(\Omega) = \{(i, j) \in [\![1;n]\!]^2 : i < j\}$ (c'est un triangle).
- Exemple fil rouge : Une urne contient 3 boules blanches et 4 boules noires. On tire successivement deux boules de cette urne. On peut alors considérer la variable aléatoire X prenant la valeur 1 si la première boule tirée est blanche et 0 sinon ainsi que la variable aléatoire Y qui prend la valeur 1 si la seconde boule tirée est blanche et 0 sinon. Alors (X, Y) est un couple de variable aléatoire qui prend ses valeurs dans $\{0; 1\}^2$.

À tout couple de variables aléatoires est associé naturellement le système complet d'événements qui prend en compte une par une les valeurs conjointes prises par le couple.

Proposition 1

Soit (X, Y) un couple de variables aléatoires finies. On appelle système complet d'événements associé à (X, Y) la famille d'événements $\big((X = x\,, Y = y)\big)_{(x,y) \in (X,Y)(\Omega)}$.

■ Les événements $(X = x\,, Y = y)$ sont clairement incompatibles deux à deux et l'on peut écrire $\bigsqcup_{(x,y)\in(X,Y)(\Omega)} (X = x\,, Y = y) = \big(X \in X(\Omega)\,, Y \in Y(\Omega)\big) = \Omega$, ce qui établit le résultat. ■

B. Loi d'un vecteur aléatoire fini

B.1. Densité conjointe

Définition 2

Soit (X, Y) un couple de variables aléatoires finies. On appelle loi conjointe du couple (X, Y) ou densité (singulière) conjointe du couple (X, Y) l'application $f_{(X,Y)}$ définie par

$$f_{(X,Y)} \left\{ \begin{array}{ccc} (X,Y)(\Omega) & \longrightarrow & [0;1] \\ (x,y) & \longmapsto & P(X = x\,,\, Y = y) \end{array} \right.$$

Pour définir la loi conjointe d'un couple (X, Y), on utilise souvent un tableau à double entrée (lignes et colonnes) donnant les probabilités des différents éléments de l'ensemble $X(\Omega) \times Y(\Omega)$.

Lorsque $(X, Y)(\Omega)$ est strictement plus petit que $X(\Omega) \times Y(\Omega)$, on attribue évidemment une probabilité nulle aux événements $(X = x\,,\, Y = y)$ pour les (x, y) qui sont dans $X(\Omega) \times Y(\Omega)$ et pas dans $(X, Y)(\Omega)$.

Comme $\big((X = x\,,\, Y = y)\big)_{(x,y)\in(X,Y)(\Omega)}$ est un système complet d'événements, on a

$$\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} P(X = x\,,\, Y = y) \quad = \sum_{(x,y)\in(X,Y)(\Omega)} P(X = x\,,\, Y = y) \quad = \quad 1,$$

ce qui signifie que la somme de tous les nombres du tableau est égale à 1.

Exemples :

- Exemple fil rouge : Reprenons l'urne qui contient 3 boules blanches et 4 noires. On tire successivement deux boules. On note X la variable aléatoire prenant la valeur 1 si la première boule tirée est blanche et 0 sinon et l'on note Y la variable aléatoire qui vaut 1 si la seconde boule tirée est blanche et 0 sinon. Alors (X, Y) est un couple de variables aléatoires qui prend ses valeurs dans $\{0; 1\}^2$ et l'on obtient les lois conjointes suivantes, suivant que l'on considère le tirage avec ou sans remise :

$Y \backslash X$	0	1
0	$\frac{16}{49}$	$\frac{12}{49}$
1	$\frac{12}{49}$	$\frac{9}{49}$

avec remise

$Y \backslash X$	0	1
0	$\frac{2}{7}$	$\frac{2}{7}$
1	$\frac{2}{7}$	$\frac{1}{7}$

sans remise

La proposition suivante donne une condition nécessaire et suffisante pour qu'une fonction soit une densité conjointe.

Proposition 2

Soient A un ensemble fini constitué de couples et $f : A \longrightarrow \mathbb{R}$ une application. Si f est à valeurs positives et si $\sum_{(x,y)\in A} f(x, y) = 1$, alors f est une loi de probabilité, c'est-à-dire qu'il existe un espace probabilisé $(\Omega, \mathscr{T}, P)$ et un couple de variables aléatoires (X, Y) sur $(\Omega, \mathscr{T})$ tels que l'on ait $X(\Omega) \times Y(\Omega) = A$ et $\forall (x, y) \in A,\ P(X = x\,,\, Y = y) = f(x, y)$.

■ Comme $\sum_{(x,y)\in A} f(x, y) = 1$ par hypothèse, on sait, d'après le théorème de construction des probabilités finies, qu'il existe une probabilité P_f sur $\big(A; \mathscr{P}(A)\big)$ telle que $\forall (x, y) \in A,\ P_f\big((x, y)\big) = f(x, y)$. Il suffit alors de prendre $\Omega = A$, $\mathscr{T} = \mathscr{P}(A)$, $P = P_f$ et $(X, Y) = \mathrm{Id}_A$ pour avoir gain de cause. ■

B.2. Lois marginales

Proposition 3

Soit (X, Y) un couple de variables aléatoires finies. Les lois de probabilités f_X et f_Y de X et Y sont appelées lois marginales de X et de Y et sont données par

$$f_X \left\{ \begin{array}{ccc} X(\Omega) & \longrightarrow & [0;1] \\ x & \longmapsto & \displaystyle\sum_{y \in Y(\Omega)} P(X = x\,,\, Y = y) \end{array} \right. \quad \text{et} \quad f_Y \left\{ \begin{array}{ccc} Y(\Omega) & \longrightarrow & [0;1] \\ y & \longmapsto & \displaystyle\sum_{x \in X(\Omega)} P(X = x\,,\, Y = y). \end{array} \right.$$

■ La famille $\big((Y = y)\big)_{y \in Y(\Omega)}$ est un système complet d'événements, donc, pour tout $x \in X(\Omega)$, on a $P(X = x) = \sum_{y \in Y(\Omega)} P(X = x\,,\, Y = y)$, d'après la formule de filtration. De même pour $P(Y = y)$. ■

Lorsque X et Y suivent la même loi de probabilité $\mathscr{L}$, on dit que X et Y sont des variables aléatoires identiquement distribuées (en abrégé : des v.a.i.d.) sur la loi $\mathscr{L}$.

Sommations dans les marges

Dans le tableau de la loi conjointe, $P(X = x)$ est la somme des probabilités de la colonne de x et $P(Y = y)$ est la somme des probabilités de la ligne de y. On reporte alors les résultats dans une colonne et une ligne supplémentaires, appelées les « marges », d'où le nom de lois marginales :

$Y \backslash X$	x_1	$\cdots$	x_n	loi de Y
y_1				
$\vdots$				
y_m				
loi de X				1

$$P(X = x) = \sum_{y \in Y(\Omega)} P(X = x\,,\, Y = y)$$

$$P(Y = y) = \sum_{x \in X(\Omega)} P(X = x\,,\, Y = y).$$

Le 1 dans la case en bas à droite est la somme de toutes les probabilités de la loi conjointe mais aussi la somme de chacune des marges.

On constate donc que la loi conjointe permet de déterminer les lois marginales. En revanche, les lois marginales ne suffisent pas en général pour reconstituer la loi du couple ! C'est ce qu'illustre l'exemple fil-rouge ci-dessous.

Exemples :

- Exemple fil rouge : Avec ou sans remise, X et Y sont identiquement distribuées sur $\mathscr{B}(3/7)$:

avec remise

$Y \backslash X$	0	1	loi de Y
0	$\frac{16}{49}$	$\frac{12}{49}$	$\frac{4}{7}$
1	$\frac{12}{49}$	$\frac{9}{49}$	$\frac{3}{7}$
loi de X	$\frac{4}{7}$	$\frac{3}{7}$	1

sans remise

$Y \backslash X$	0	1	loi de Y
0	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{4}{7}$
1	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{3}{7}$
loi de X	$\frac{4}{7}$	$\frac{3}{7}$	1

B.3. Loi conditionnelle

Définition 3

Soit (X,Y) un couple de variables aléatoires finies. Pour tout $y \in Y(\Omega)$, l'application

$$f_{X\,|\,Y=y} \left\{ \begin{array}{ccc} X(\Omega) & \longrightarrow & [0;1] \\ x & \longmapsto & P(X=x\,|\,Y=y) = \dfrac{P(X=x\,,\,Y=y)}{P(Y=y)} \end{array} \right.$$

est appelée loi conditionnelle de X sachant $(Y=y)$.

■ Il faut vérifier que l'on a bien affaire à une loi de probabilité. Il est clair que $f_{X\,|\,Y=y}$ est bien à valeurs positives. De plus,

$$\begin{aligned} \sum_{x\in X(\Omega)} P(X=x\,|\,Y=y) &= \sum_{x\in X(\Omega)} \frac{P(X=x\,,\,Y=y)}{P(Y=y)} \\ &= \frac{1}{P(Y=y)} \sum_{x\in X(\Omega)} P(X=x\,,\,Y=y) \\ &= \frac{1}{P(Y=y)} P(Y=y) \\ &= 1, \end{aligned}$$

où l'avant-dernière égalité découle de la filtration de l'événement $(Y=y)$ à travers le système complet d'événements associé à X. ■

Il est bien évidemment possible de « renverser » la définition précédente en considérant la loi de probabilité conditionnelle de Y sachant $(X=x)$, où $x \in X(\Omega)$.

Lorsqu'on considère une loi conditionnelle, on ne modifie pas la variable aléatoire elle-même mais seulement sa loi de probabilité.

La connaissance de la loi marginale et des lois conditionnelles permet de déterminer la loi conjointe grâce à la formule de conditionnement :

$$P(X=x\,,\,Y=y) = P(X=x)P(Y=y\,|\,X=x).$$

Ainsi, dans la pratique, pour trouver la loi conjointe de (X,Y), il est classique de déterminer la loi de X et les lois conditionnelles de Y sachant $(X=x)$ pour tout $x \in X(\Omega)$.

Exemples :

- Exemple fil rouge : Dans le cas du tirage sans remise, la loi conditionnelle de X sachant $(Y=0)$ est donnée par

i	0	1
$P(X=i\,\|\,Y=0)$	$\frac{1}{2}$	$\frac{1}{2}$

et la loi conditionnelle de X sachant $(Y=1)$ est donnée par

i	0	1
$P(X=i\,\|\,Y=1)$	$\frac{2}{3}$	$\frac{1}{3}$

B.4. Loi d'une composée

Définition 4

Soient (X, Y) un couple de variables aléatoires finies et g une fonction dont l'ensemble de définition contient $(X, Y)(\Omega)$. La variable aléatoire $g \circ (X, Y)$, obtenue en composant les applications (X, Y) et g, est notée $g(X, Y)$.

L'encadré suivant explique comment trouver la loi de $g(X, Y)$.

Loi d'une composée

Soit (X, Y) un couple de variables aléatoires finies dont on connait l'univers image $(X, Y)(\Omega)$ et la loi conjointe. Pour déterminer la loi de probabilité d'une variable composée $Z = g(X, Y)$, on effectue les étapes suivantes :

- on détermine l'univers image de $Z = g(X, Y)$ en calculant les images par g des couples de $(X, Y)(\Omega)$ (certaines valeurs peuvent être égales) ;
- pour chaque $z \in Z(\Omega)$,
  - on recherche tous les couples $(x_1, y_1), \dots, (x_k, y_k)$ de $(X, Y)(\Omega)$ tels que $g(x_1, y_1) = \cdots = g(x_k, y_k) = z$ de sorte que l'événement $(Z = z)$ puisse se décomposer sous la forme $(Z = z) = (X = x_1\,,\, Y = y_1) \sqcup \cdots \sqcup (X = x_k\,,\, Y = y_k)$;
  - on calcule $P(Z = z) = P(X = x_1\,,\, Y = y_1) + \cdots + P(X = x_k\,,\, Y = y_k)$.

On a donc

$$f_Z \left\{ \begin{array}{ccc} Z(\Omega) & \longrightarrow & \mathbb{R} \\ z & \longmapsto & \displaystyle\sum_{(x,y) \in g^{-1}(\{z\})} P(X = x\,,\, Y = y) \end{array} \right.$$

Exemples :

- Exemple fil rouge : Dans le cas du tirage sans remise, recherchons la loi de probabilité de la somme $S = X + Y$. On a $S(\Omega) = \{0; 1; 2\}$ et pour $s \in \{0; 1; 2\}$, on a

$$\begin{aligned} P(S = s) &= P(X + Y = s) \\ &= P\Big(\bigsqcup_{x \in X(\Omega)} (X = x\,,\, Y = s - x) \Big) \\ &= \sum_{x \in X(\Omega)} P(X = x\,,\, Y = s - x) \\ &= \sum_{x \in X(\Omega)} f_{(X,Y)}(x, s - x) \qquad \longleftarrow \text{ convolution} \end{aligned}$$

donc

$$P(S = 0) = P(X = 0\,,\, Y = 0) = \frac{2}{7}$$
$$P(S = 1) = P(X = 0\,,\, Y = 1) + P(X = 1\,,\, Y = 0) = \frac{2}{7} + \frac{2}{7} = \frac{4}{7}$$
$$P(S = 2) = P(X = 1\,,\, Y = 1) = \frac{1}{7},$$

c'est-à-dire

i	0	1	2
$P(S = i)$	$\frac{2}{7}$	$\frac{4}{7}$	$\frac{1}{7}$

C. Indépendance de variables aléatoires

C.1. Indépendance de deux variables aléatoires

Définition 5

Soit (X, Y) un couple de variables aléatoires finies. On dit que X et Y sont indépendantes (et on note $X \perp\!\!\!\perp Y$) lorsque, pour toutes parties $A \subset X(\Omega)$ et $B \subset Y(\Omega)$, on a

$$P(X \in A\,,\, Y \in B) = P(X \in A)P(Y \in B).$$

Notons que, puisque les vecteurs aléatoires sont des variables aléatoires, on peut très bien parler de vecteurs aléatoires indépendants.

On peut caractériser l'indépendance à l'aide des densités singulières.

Proposition 4

Soit (X, Y) un couple de variables aléatoires finies. Les variables X et Y sont indépendantes si, et seulement si, pour tout $x \in X(\Omega)$ et tout $y \in Y(\Omega)$, on a

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

■ $\Rightarrow$ Supposons que X et Y sont indépendantes. Soient $x \in X(\Omega)$ et $y \in Y(\Omega)$. En prenant $A = \{x\}$ et $B = \{y\}$ dans la définition de l'indépendance, on obtient $P(X = x\,,\, Y = y) = P(X = x)P(Y = y)$.

$\Leftarrow$ Supposons que, pour tous $x \in X(\Omega)$ et $y \in Y(\Omega)$, on a $P(X = x\,,\, Y = y) = P(X = x)P(Y = y)$. Soient $A \subset X(\Omega)$ et $B \subset Y(\Omega)$. On a

$$\begin{aligned}
P(X \in A\,,\, Y \in B) &= P\Big(\bigsqcup_{(a,b)\in A\times B} (X = a\,,\, Y = b)\Big) = \sum_{(a,b)\in A\times B} P(X = a\,,\, Y = b)\\
&= \sum_{(a,b)\in A\times B} P(X = a)P(Y = b) = \sum_{a\in A} P(X = a) \sum_{b\in B} P(Y = b)\\
&= P\Big(\bigsqcup_{a\in A} (X = a)\Big)P\Big(\bigsqcup_{b\in B} (Y = b)\Big) = P(X \in A)P(Y \in B),
\end{aligned}$$

ce qui démontre que X et Y sont indépendantes. ■

On peut reformuler cet énoncé à l'aide des densités singulières : X et Y sont indépendantes si, et seulement si, les lois marginales permettent de reconstituer la loi conjointe du couple en effectuant le produit tensoriel : $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$ pour tout $x \in X(\Omega)$ et tout $y \in Y(\Omega)$.

Pour résumer, on retient que, dans le tableau des lois, il y a indépendance lorsque la loi conjointe est le produit des lois marginales.

On notera aussi que X et Y sont indépendantes lorsque, pour tous $x \in X(\Omega)$, la loi conditionnelle de Y sachant $(X = x)$ est égale à celle de Y. Autrement dit, il y a indépendance lorsque toute information apportée par l'une des variables n'influence pas la loi de l'autre.

Exemples :

- Une variable aléatoire certaine est indépendante de toute autre variable.
- Exemple fil rouge : La lecture des tableaux des lois conjointes nous dit que, lorsque les tirages s'effectuent avec remise, X et Y sont indépendantes et qu'au contraire, lorsque lorsque les tirages s'effectuent sans remise, X et Y ne sont pas indépendantes.

C.2. Indépendance mutuelle

Définition 6

Soit $(X_1, X_2, \dots, X_n)$ un vecteur aléatoire fini. On dit que $X_1, X_2, \dots, X_n$ sont mutuellement indépendantes lorsque, pour toutes parties $A_1 \subset X_1(\Omega)$, $A_2 \subset X_2(\Omega)$, $\dots$, $A_n \subset X_n(\Omega)$, on a

$$P(X_1 \in A_1\,,\ \dots\ ,\ X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n)$$

Lorsque $X_1, \dots, X_n$ sont des variables aléatoires mutuellement indépendantes identiquement distribuées (v.a.i.i.d.) sur la loi $\mathscr{L}$, on dit que $(X_1, \dots, X_n)$ est un échantillon de la loi $\mathscr{L}$.

À la différence de l'indépendance mutuelle d'une famille d'événements, il n'est pas nécessaire, pour définir l'indépendance mutuelle d'une famille $(X_1, X_2, \dots, X_n)$ de variables aléatoires, de faire intervenir les sous-familles de $(X_1, X_2, \dots, X_n)$. L'énoncé suivant explique pourquoi.

Proposition 5

Toute sous-famille d'une famille de variables aléatoires finies mutuellement indépendantes est constituée de variables mutuellement indépendantes.

■ Soient $X_1, X_2, \dots, X_n$ des variables aléatoires finies mutuellement indépendantes et $(X_{i_1}, \dots, X_{i_k})$ une sous-famille de $(X_1, \dots, X_n)$, où $k \in [\![1;n]\!]$. Pour tout $i \in [\![1;n]\!] \setminus \{i_1, \dots, i_k\}$, on choisit $A_i = X_i(\Omega)$ dans la définition de l'indépendance de $X_1, \dots, X_n$. Comme $\big(X_i = X_i(\Omega)\big) = \Omega$ pour tout $i \in [\![1;n]\!] \setminus \{i_1, \dots, i_k\}$, on obtient l'égalité $P(X_{i_1} \in A_{i_1}\,,\ \dots\ ,\ X_{i_k} \in A_{i_k}) = P(X_{i_1} \in A_{i_1}) \cdots P(X_{i_k} \in A_{i_k})$ pour toutes parties $A_{i_1} \subset X_{i_1}(\Omega), \dots, A_{i_k} \subset X_{i_k}(\Omega)$. Donc $X_{i_1}, \dots, X_{i_k}$ sont mutuellement indépendantes. ■

Cette proposition dit en particulier que des variables mutuellement indépendantes sont deux à deux indépendantes. En revanche, il faut être convaincu que la réciproque est fausse (on donne ci-dessous un contre-exemple). Par conséquent, comme pour les événements, l'indépendance mutuelle est une notion plus forte que l'indépendance deux à deux. C'est pourquoi, dans la pratique, on utilise quasi-exclusivement l'indépendance mutuelle (et l'on omet en général l'adjectif « mutuelle »).

Exemples :

- Soient X et Y deux variables aléatoires indépendantes de même loi $\mathscr{B}(1/2)$. On pose $Z = |X - Y|$ et l'on constate que Z suit également la loi $\mathscr{B}(1/2)$ puisque $Z(\Omega) = \{0;1\}$ et $P(Z = 1) = P(X = 0\,,\ Y = 0) + P(X = 1\,,\ Y = 1) = (1/2)^2 + (1/2)^2 = 1/2$.

  Alors $P(X = 0\,,\ Z = 0) = P(X = 0\,,\ Y = 0) = P(X = 0)P(Y = 0) = P(X = 0)P(Z = 0)$ et de même pour les autres événements conjoints, donc Z est indépendante de X.
  De la même façon, Z est indépendante de Y.
  Autrement dit, les variables X, Y, Z sont indépendantes deux à deux.

  En revanche, les variables X, Y, Z ne sont pas mutuellement indépendantes. En effet, on a $P(X = 0\,,\ Y = 0\,,\ Z = 1) = 0$ alors que $P(X = 0)P(Y = 0)P(Z = 1) = 1/8$.

Comme dans le cas de l'indépendance de deux variables, on peut caractériser l'indépendance mutuelle à l'aide des densités singulières.

Proposition 6

Soit $(X_1, X_2, \dots, X_n)$ un vecteur aléatoire fini. Les variables $X_1, X_2, \dots, X_n$ sont mutuellement indépendantes si, et seulement si, pour tous $x_1 \in X_1(\Omega)$, $x_2 \in X_2(\Omega), \dots, x_n \in X_n(\Omega)$, on a

$$P(X_1 = x_1\,,\ X_2 = x_2\,,\ \dots\ ,\ X_n = x_n) = P(X_1 = x_1)P(X_2 = x_2) \cdots P(X_n = x_n).$$

■ On adapte la démonstration de la proposition 4. ■

C.3. Lemme des coalitions

Voici le lemme des coalitions qui donne les propriétés de stabilité de l'indépendance lorsqu'on compose des variables aléatoires.

Proposition 7

Soient $(X_1, X_2, \dots, X_n)$ un vecteur aléatoire fini et $k \in [\![2;n]\!]$. Si $X_1, X_2, \dots, X_n$ sont mutuellement indépendantes, alors, pour tous entiers $1 \leqslant p_1 < p_2 < \cdots < p_{k-1} \leqslant n$ et toutes fonctions g_1, g_2, $\dots$, g_k définies respectivement sur les ensembles $(X_1, \dots, X_{p_1})(\Omega)$, $(X_{p_1+1}, \dots, X_{p_2})(\Omega)$, $\dots$, $(X_{p_{k-1}+1}, \dots, X_n)(\Omega)$, les variables aléatoires $g_1(X_1, \dots, X_{p_1})$, $g_2(X_{p_1+1}, \dots, X_{p_2})$, $\dots$, $g_k(X_{p_{k-1}+1}, \dots, X_n)$ sont mutuellement indépendantes.

■ On traite le cas $k = 2$ (le cas général est similaire mais lourd à rédiger). On pose $p = p_1$, $g = g_1$ et $h = g_2$. Pour $y \in g(X_1, \dots, X_p)(\Omega)$ et $z \in h(X_{p+1}, \dots, X_n)(\Omega)$, on a

$$\begin{aligned}
& P\big(g(X_1, \dots, X_p) = y \,,\, h(X_{p+1}, \dots, X_n) = z\big) \\
= \ & P\Bigg(\bigsqcup_{\substack{(x_1, \dots, x_p) \in g^{-1}(\{y\}) \\ (x_{p+1}, \dots, x_n) \in h^{-1}(\{z\})}} (X_1 = x_1 \,,\, \dots \,,\, X_p = x_p \,,\, X_{p+1} = x_{p+1} \,,\, \dots \,,\, X_n = x_n) \Bigg) \\
= \ & \sum_{\substack{(x_1, \dots, x_p) \in g^{-1}(\{y\}) \\ (x_{p+1}, \dots, x_n) \in h^{-1}(\{z\})}} P(X_1 = x_1 \,,\, \dots \,,\, X_p = x_p \,,\, X_{p+1} = x_{p+1} \,,\, \dots \,,\, X_n = x_n) \\
= \ & \sum_{\substack{(x_1, \dots, x_p) \in g^{-1}(\{y\}) \\ (x_{p+1}, \dots, x_n) \in h^{-1}(\{z\})}} P(X_1 = x_1 \,,\, \dots \,,\, X_p = x_p) P(X_{p+1} = x_{p+1} \,,\, \dots \,,\, X_n = x_n) \quad \text{par indépendance} \\
= \ & \sum_{(x_1, \dots, x_p) \in g^{-1}(\{y\})} P(X_1 = x_1 \,,\, \dots \,,\, X_p = x_p) \sum_{(x_{p+1}, \dots, x_n) \in h^{-1}(\{z\})} P(X_{p+1} = x_{p+1} \,,\, \dots \,,\, X_n = x_n) \\
= \ & P\Bigg(\bigsqcup_{(x_1, \dots, x_p) \in g^{-1}(\{y\})} (X_1 = x_1 \,,\, \dots \,,\, X_p = x_p) \Bigg) P\Bigg(\bigsqcup_{(x_{p+1}, \dots, x_n) \in h^{-1}(\{z\})} (X_{p+1} = x_{p+1} \,,\, \dots \,,\, X_n = x_n) \Bigg) \\
= \ & P\big(g(X_1, \dots, X_p) = y\big) P\big(h(X_{p+1}, \dots, X_n) = z\big),
\end{aligned}$$

ce qui prouve l'indépendance de $g(X_1, \dots, X_p)$ et $h(X_{p+1}, \dots, X_n)$. ■

Cette proposition nous dit en particulier que, si $X_1, X_2, \dots, X_n$ sont mutuellement indépendantes, alors :

- pour toutes fonctions g_1, g_2, $\dots$, g_n définies respectivement sur $X_1(\Omega)$, $X_2(\Omega)$, $X_n(\Omega)$, les variables aléatoires $g_1(X_1)$, $g_2(X_2)$, $\dots$, $g_n(X_n)$ sont mutuellement indépendantes ;
- pour tous entiers $1 \leqslant p_1 < p_2 < \cdots < p_{k-1} \leqslant n$ (où $k \in [\![2;n]\!]$), les vecteurs aléatoires $(X_1, \dots, X_{p_1})$, $(X_{p_1+1}, \dots, X_{p_2})$, $\dots$, $(X_{p_{k-1}+1}, \dots, X_n)$ sont mutuellement indépendants.

En substance, on retient que l'indépendance se conserve tant que l'on considère des variables composées à l'aide de coalitions disjointes. Il est interdit d'appartenir à deux camps opposés !

Ainsi, si l'on considère une variable X et l'une de ses composées, disons $Z = g(X)$, alors X et Z ne sont jamais indépendantes (sauf dans le cas très particulier où X ou Z est une variable certaine). En effet, si X et $g(X)$ sont indépendantes, alors pour $x_0 \in X(\Omega)$ tel que $P(X = x_0) > 0$, on a $P(X = x_0)P\big(g(X) = g(x_0)\big) = P\big(X = x_0 \,,\, g(X) = g(x_0)\big) = P(X = x_0)$, c'est-à-dire $P\big(g(X) = g(x_0)\big) = 1$. Cela n'est possible que lorsque $g(X)$ suit une loi certaine.

La réciproque du lemme de coalition est fausse ! En effet, si $h : Y(\Omega) \longrightarrow \mathbb{R}$ est une fonction constante, alors $h(Y)$ est une variable aléatoire certaine qui est nécessairement indépendante de $g(X)$ même si X et Y ne sont pas indépendantes.

C.4. Sommes de variables de Bernoulli

Proposition 8

Soit $(X_1, \dots, X_n)$ un échantillon de la loi $\mathscr{B}(p)$, c'est-à-dire que $X_1, \dots, X_n$ sont des variables aléatoires indépendantes identiquement distribuées sur la loi $\mathscr{B}(p)$. Alors $S_n = X_1 + \cdots + X_n$ suit la loi $\mathscr{B}(n;p)$.

■ On procède par récurrence sur $n \in \mathbb{N}^*$.

Initialisation : Lorsque $n = 1$, $S_1 = X_1$ suit la loi $\mathscr{B}(p)$, qui est bien la loi $\mathscr{B}(1;p)$.

Hérédité : Fixons $n \in \mathbb{N}^*$ et supposons que $S_n \hookrightarrow \mathscr{B}(n;p)$. Démontrons que $S_{n+1} \hookrightarrow \mathscr{B}(n+1;p)$. Soit $k \in [\![0; n+1]\!]$. D'après la formule des probabilités totales appliquées à $(S_{n+1} = k)$ à travers le système complet d'événément $\big((X_{n+1} = 0), (X_{n+1} = 1)\big)$, on a

$$\begin{aligned}
P(S_{n+1} = k) &= P(X_{n+1} = 0)P(S_{n+1} = k \,|\, X_{n+1} = 0) + P(X_{n+1} = 1)P(S_{n+1} = k \,|\, X_{n+1} = 1) \\
&= P(X_{n+1} = 0)P(S_n = k \,|\, X_{n+1} = 0) + P(X_{n+1} = 1)P(S_n = k-1 \,|\, X_{n+1} = 1) \\
&= P(X_{n+1} = 0)P(S_n = k) + P(X_{n+1} = 1)P(S_n = k-1) \qquad \text{par indépendance} \\
&= q \times \binom{n}{k} p^k q^{n-k} + p \times \binom{n}{k-1} p^{k-1} q^{n-k+1} \\
&= \left[\binom{n}{k} + \binom{n}{k-1}\right] p^k q^{n+1-k} \\
&= \binom{n+1}{k} p^k q^{n+1-k} \qquad \text{d'après la relation de Pascal,}
\end{aligned}$$

donc $S_{n+1} \hookrightarrow \mathscr{B}(n+1;p)$.

Conclusion : D'après le principe de récurrence, $S_n \hookrightarrow \mathscr{B}(n;p)$ pour tout $n \geqslant 1$. ■

Nous retiendrons donc le résultat suivant : une variable aléatoire suivant une loi $\mathscr{B}(n;p)$ a la même loi qu'une somme de n variables aléatoires indépendantes de loi $\mathscr{B}(p)$.

Ce résultat peut nous servir à retrouver l'espérance d'une loi binomiale puisque l'espérance est linéaire (et donc $E(S_n) = E(X_1) + \cdots + E(X_n) = np$) mais aussi la variance puisque nous verrons, au paragraphe D. 2., que la variance d'une somme de variables indépendantes est la somme des variances (et donc $V(S_n) = V(X_1) + \cdots + V(X_n) = npq$).

Corollaire 1

Considérons $S_1, S_2, \dots, S_m$ des variables aléatoires mutuellement indépendantes qui suivent respectivement les lois $\mathscr{B}(n_1;p)$, $\mathscr{B}(n_2;p)$, …, $\mathscr{B}(n_m;p)$. Alors la somme $S_1 + S_2 + \cdots + S_m$ suit la loi $\mathscr{B}(n_1 + \cdots + n_m;p)$.

■ Soient $X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_{n_2}, \dots, X_{n_{m-1}+1}, \dots, X_{n_m}$ des variables indépendantes identiquement distribuées sur la loi $\mathscr{B}(p)$. Alors $\Sigma_1 = X_1 + \cdots + X_{n_1}$ suit $\mathscr{B}(n_1, p)$, …, $\Sigma_m = X_{n_{m-1}+1} + \cdots + X_{n_m}$ suit $\mathscr{B}(n_m, p)$ et $\Sigma_1, \dots, \Sigma_m$ sont indépendantes. Par suite, d'une part, $\Sigma_1 + \cdots + \Sigma_m$ suit la même loi que $S_1 + \cdots + S_m$ et, d'autre part, $\Sigma_1 + \cdots + \Sigma_m = X_1 + \cdots + X_{n_m}$ suit la loi $\mathscr{B}(n_1 + \cdots + n_m;p)$. Le résultat en découle. ■

Pour la loi hypergéométrique, on a le résultat suivant.

Proposition 9

Soit S une variable aléatoire suivant la loi $\mathscr{H}(N;n;p)$. Alors il existe des variables aléatoires $X_1, \dots, X_n$ identiquement distribuées sur la loi $\mathscr{B}(p)$ et non indépendantes telles que S a la même loi que $X_1 + \cdots + X_n$.

■ Nous le démontrerons en exercice. ■

Ce résultat peut nous servir à retrouver l'espérance d'une loi hypergéométrique (puisque l'espérance est linéaire et donc $E(S) = E(X_1) + \cdots + E(X_n) = np$) mais pas la variance. 2 h 40

D. Paramètres de position et de dispersion

D.1. Espérance d'une composée

Proposition 10

Soit (X, Y) un couple de variables aléatoires finies. Pour toute fonction de deux variables $g : (X, Y)(\Omega) \longrightarrow \mathbb{R}$, l'espérance de la variable aléatoire $g(X, Y)$ est donnée par la formule de transfert :

$$E\big(g(X,Y)\big) = \sum_{(x,y)\in(X,Y)(\Omega)} g(x,y)P(X = x\,,\, Y = y).$$

■ On a

$$\begin{aligned} E\big(g(X,Y)\big) &= \sum_{z\in g(X,Y)(\Omega)} zP\big(g(X,Y) = z\big) = \sum_{z\in g(X,Y)(\Omega)} z \sum_{(x,y)\in g^{-1}(\{z\})} P(X = x\,,\, Y = y) \\ &= \sum_{z\in g(X,Y)(\Omega)} \sum_{(x,y)\in g^{-1}(\{z\})} zP(X = x\,,\, Y = y) = \sum_{(x,y)\in(X,Y)(\Omega)} g(x,y)P(X = x\,,\, Y = y), \end{aligned}$$

ce qui démontre le résultat. ■

Pour calculer l'espérance de $Z = g(X, Y)$, on additionne donc les valeurs $g(x, y)$ prises par $g(X, Y)$ en pondérant chacune d'elles par la probabilité conjointe d'avoir $X = x$ et $Y = y$.

Exemples :

- Pour une somme de deux variables aléatoires, on obtient

$$\begin{aligned} E(X+Y) &= \sum_{(x,y)\in(X,Y)(\Omega)} (x+y)P(X = x\,,\, Y = y) \\ &= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} (x+y)P(X = x\,,\, Y = y) \qquad \text{quitte à rajouter des termes nuls} \\ &= \sum_{x\in X(\Omega)} x \sum_{y\in Y(\Omega)} P(X = x\,,\, Y = y) + \sum_{y\in Y(\Omega)} y \sum_{x\in X(\Omega)} P(X = x\,,\, Y = y) \\ &= \sum_{x\in X(\Omega)} xP(X = x) + \sum_{y\in Y(\Omega)} yP(Y = y) \\ &= E(X) + E(Y), \end{aligned}$$

  ce qui fournit une nouvelle démonstration de la linéarité de l'espérance.

- Exemple fil rouge : Dans le cas du tirage avec remise, on a

$$E(XY) = \sum_{k,\ell\in\{0;1\}} k\ell\, P(X = k\,,\, Y = \ell) = (0\times0)\frac{16}{49} + (0\times1)\frac{12}{49} + (1\times0)\frac{12}{49} + (1\times1)\frac{9}{49} = \frac{9}{49}$$

  et dans le cas du tirage sans remise, on a

$$E(XY) = (0\times0)\frac{2}{7} + (0\times1)\frac{2}{7} + (1\times0)\frac{2}{7} + (1\times1)\frac{1}{7} = \frac{1}{7}.$$

- Soient X et Y deux variables aléatoires indépendantes de même loi de Bernoulli de paramètre p. Alors

$$\begin{aligned} E\big(\min(X;Y)\big) &= \sum_{k,\ell\in\{0;1\}} \min\{k;\ell\}\, P(X = k)P(Y = \ell) \\ &= \min\{0;0\}q^2 + \min\{0;1\}qp + \min\{1;0\}pq + \min\{1;1\}p^2 \\ &= p^2. \end{aligned}$$

D.2. Covariance

a) Définition

Définition 7

Soit (X,Y) un couple de variables aléatoires réelles finies. On appelle covariance de X et Y le nombre réel, noté $\mathrm{Cov}(X,Y)$, défini par

$$\mathrm{Cov}(X,Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big].$$

Elle est également définie par la formule de Kœnig–Huygens

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

■ On a

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] \\ &= E\big[XY - E(Y)X - E(X)Y + E(X)E(Y)\big] \\ &= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(Y)E(X). \end{aligned}$$ ■

En pratique, on utilise la formule de Kœnig–Huygens pour calculer la covariance.

Heuristiquement, la covariance mesure l'influence qui s'exerce entre deux variables aléatoires. Ainsi, si l'on note X_j le nombre de loups garous le jour j, la covariance de X_j et X_{j+28} est supérieure aux covariances de X_j et X_{j+k} pour k différent de 28.

Exemples :

- Si X est une variable aléatoire certaine, alors $\mathrm{Cov}(X,Y) = 0$.
- Exemple fil rouge : Pour le tirage avec remise, on a $E(XY) = 9/49$ et $E(X) = E(Y) = 3/7$, donc
$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{9}{49} - \left(\frac{3}{7}\right)^2 = 0$$
et dans le cas du tirage sans remise, on a $E(XY) = 1/7$ et $E(X) = E(Y) = 3/7$, d'où
$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{7} - \left(\frac{3}{7}\right)^2 = -\frac{2}{49}.$$

La proposition suivante dresse la liste des propriétés de la covariance.

Proposition 11

Soient (X,Y) un couple de variables aléatoires réelles finies et $a, b \in \mathbb{R}$. Alors

(i) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ (symétrie) ;

(ii) $\mathrm{Cov}(aX_1 + bX_2, Y) = a\,\mathrm{Cov}(X_1,Y) + b\,\mathrm{Cov}(X_2,Y)$ (linéarité à gauche) ;
$\mathrm{Cov}(X, aY_1 + bY_2) = a\,\mathrm{Cov}(X,Y_1) + b\,\mathrm{Cov}(X,Y_2)$ (linéarité à droite) ;

(iii) $\mathrm{Cov}(X,X) = V(X) \geqslant 0$ (positivité).

■ AQT ■

La covariance est donc une forme bilinéaire symétrique positive sur l'espace vectoriel des variables aléatoires réelles finies. La covariance est donc presque un produit scalaire. Avec cette analogie, l'indépendance s'interprète comme une sorte d'orthogonalité des variables aléatoires (d'où la notation $\perp\!\!\!\perp$). Cela dit, nous allons voir que la covariance ne caractérise pas l'indépendance aussi clairement que le produit scalaire caractérise l'orthogonalité.

b) Variance d'une somme

Proposition 12

Soit (X, Y) un couple de variables aléatoires réelles finies. Alors
$$V(X+Y) = V(X) + 2\,\mathrm{Cov}(X,Y) + V(Y).$$

■ On a

$$\begin{aligned} V(X+Y) &= \mathrm{Cov}(X+Y, X+Y) \\ &= \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,X) + \mathrm{Cov}(Y,Y) \\ &= V(X) + 2\,\mathrm{Cov}(X,Y) + V(Y) \end{aligned}$$

où l'on utilisé la bilinéarité de la covariance. ■

On notera la ressemblance avec le développement du carré scalaire d'une somme.

Exemples :

- Exemple fil rouge : Pour le tirage sans remise, on vient de voir que $\mathrm{Cov}(X,Y) = -2/49$. Par ailleurs, $V(X) = V(Y) = 12/49$, donc
$$V(X+Y) = \frac{12}{49} + 2 \times \frac{-2}{49} + \frac{12}{49} = \frac{20}{49}.$$

Le corollaire suivant traite le cas où l'on a plus de deux variables aléatoires.

Corollaire 2

Soit $(X_1, \ldots, X_n)$ un vecteur aléatoire réel. Alors
$$\begin{aligned} V(X_1 + \cdots + X_n) &= \underbrace{\sum_{1 \leqslant i,j \leqslant n} \mathrm{Cov}(X_i, X_j)}_{n^2 \text{ termes}} \\ &= \sum_{k=1}^{n} V(X_k) + 2 \underbrace{\sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}(X_i, X_j)}_{\binom{n}{2} \text{ termes}}. \end{aligned}$$

■ AQT (par récurrence). ■

Pour retenir ces formules, on peut utiliser la matrice de covariance, c'est-à-dire la matrice carrée $C = (c_{i,j})_{1 \leqslant i,j \leqslant n}$ dont les coefficients sont donnés par $c_{i,j} = \mathrm{Cov}(X_i, X_j)$. La variance de la somme $X_1 + \cdots + X_n$ est alors tout simplement la somme de tous les coefficients de cette matrice (c'est la première formule du corollaire). En tenant compte du fait que cette matrice est symétrique et du fait que ses coefficients diagonaux sont les variances des X_k, on peut regrouper la somme de tous les termes de la matrice en la somme de ses termes diagonaux à laquelle on ajoute le double de la somme des termes situés strictement au dessus de la diagonale (c'est la seconde formule du corollaire).

c) *Covariance et indépendance*

Proposition 13

Soit (X, Y) un couple de variables aléatoires réelles finies. Si X et Y sont indépendantes, alors

$$\operatorname{Cov}(X, Y) = 0,$$

c'est-à-dire

$$E(XY) = E(X)E(Y) \qquad \text{et} \qquad V(X+Y) = V(X) + V(Y).$$

■ On a

$$\begin{aligned}
E(XY) &= \sum_{(x,y)\in(X,Y)(\Omega)} xyP(X = x\,,\, Y = y) \\
&= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xyP(X = x\,,\, Y = y) && \text{quitte à rajouter des termes nuls} \\
&= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xyP(X = x)P(Y = y) && \text{par indépendance} \\
&= \sum_{x\in X(\Omega)} xP(X = x) \sum_{y\in Y(\Omega)} yP(Y = y) \\
&= E(X)E(Y),
\end{aligned}$$

ce qui implique que $\operatorname{Cov}(X, Y) = 0$, d'après la formule de Kœnig–Huygens. ■

Attention, la réciproque est fausse ! Pour le voir, considérons une variable aléatoire X de loi uniforme sur $\{-1; 0; 1\}$ et Y l'indicatrice de l'événement $(X = 0)$. Alors $XY = 0$ (car Y est nulle lorsque X ne l'est pas), d'où $E(XY) = 0$. Comme $E(X) = 0$ puisque la densité de X est paire, on en déduit que $\operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$. Mais X et Y ne sont pas indépendantes puisque $P(X = 1\,,\, Y = 1) = 0$ et $P(X = 1)P(Y = 1) = (1/3)(1/3) \neq 0$.

La covariance ne caractérise donc pas l'indépendance comme le produit scalaire caractérise l'orthogonalité !

Exemples :

- Exemple fil rouge : Pour le cas où les tirages se font avec remise, on a vu que $\operatorname{Cov}(X, Y) = 0$ et $V(X) = V(Y) = 12/49$, donc

$$V(X+Y) = \frac{12}{49} + \frac{12}{49} = \frac{24}{49}.$$

Le corollaire suivant traite le cas où l'on a plus de deux variables aléatoires.

Corollaire 3

Soit $(X_1, \ldots, X_n)$ un vecteur aléatoire réel. Si les variables $X_1, \ldots, X_n$ sont mutuellement indépendantes (deux à deux indépendantes suffit ici), alors

$$V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n).$$

■ Cela découle du corollaire 2. ■

C'est ce résultat qui permet de dire que la variance d'une loi binomiale $\mathscr{B}(n; p)$ est n fois la variance d'une loi de Bernoulli $\mathscr{B}(p)$, puisque l'on sait qu'une variable suivant la loi $\mathscr{B}(n; p)$ a la même loi qu'une somme de n variables indépendantes identiquement distribuées sur la loi $\mathscr{B}(p)$.

d) Corrélation

Définition 8

Soit (X, Y) un couple de variables aléatoires réelles finies dont les écart-types sont non nuls. On appelle coefficient de corrélation linéaire de X et Y le nombre réel, noté $\rho(X, Y)$, défini par

$$\rho(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sigma(X)\sigma(Y)}\cdot$$

On dit que X et Y sont corrélées lorsque $\rho(X, Y) \neq 0$.

Le coefficient de corrélation est indépendant des unités choisies (en physique, on dirait que c'est une grandeur sans unité).

Exemples :

- Exemple fil rouge : Pour le tirage avec remise, on a $\rho(X, Y) = 0$. Pour le tirage sans remise, on a $\operatorname{Cov}(X, Y) = -2/49$ et $V(X) = V(Y) = 12/49$, donc

$$\rho(X, Y) = \frac{-2/49}{12/49} = -\frac{1}{6}\cdot$$

La proposition suivante rasssemble les propriétés du coefficient de corrélation.

Proposition 14

Soit (X, Y) un couple de variables aléatoires réelles finies dont les écart-types sont non nuls. Alors

(i) $\rho(X, Y) \in [-1; 1]$;

(ii) $\big(\rho(X, Y) = \pm 1\big) \Longleftrightarrow (\exists \lambda, \mu \in \mathbb{R},\ X = \lambda Y + \mu \text{ p.s.})$;

(iii) $(X \perp\!\!\!\perp Y) \Longrightarrow \big(\rho(X, Y) = 0\big)$.

■ (i) Posons $Q(\lambda) = V(X + \lambda Y)$. La fonction $\lambda \longmapsto Q(\lambda)$ est positive puisque c'est une variance. Par ailleurs, $\forall \lambda \in \mathbb{R},\ Q(\lambda) = a\lambda^2 + b\lambda + c$ où $a = V(Y)$, $b = 2\operatorname{Cov}(X, Y)$ et $c = V(X)$, donc Q est un trinôme du second degré puisque $a = V(Y) \neq 0$. Comme il est positif, son discriminant est nécessairement négatif ou nul, c'est-à-dire $b^2 \leqslant 4ac$, ce qui donne $4\big(\operatorname{Cov}(X, Y)\big)^2 \leqslant 4V(Y)V(X)$. Cela implique immédiatement que $\rho(X, Y) \in [-1; 1]$.

(ii) Avec les notations du (i), on voit que la relation $\rho(X, Y) = \pm 1$ signifie que le discriminant de Q est nul ou, ce qui revient au même, que Q admet une racine double, c'est-à-dire $\exists \lambda_0 \in \mathbb{R},\ Q(\lambda_0) = 0$ ou encore $\exists \lambda_0 \in \mathbb{R},\ V(X + \lambda_0 Y) = 0$. Autrement dit, $\rho(X, Y) = \pm 1$ si, et seulement si, $X + \lambda_0 Y$ suit une loi certaine, ce qui correspond au résultat.

(iii) découle de la proposition 13. ■

La propriété (i) est équivalente à l'inégalité $|\operatorname{Cov}(X, Y)| \leqslant \sigma(X)\sigma(Y)$. C'est une inégalité de Cauchy–Schwarz.

La proposition (iii) nous dit que

$$\boxed{\text{indépendance} \quad \Longrightarrow \quad \text{non corrélation}}$$

Comme pour la covariance, la réciproque est fausse !

Lorsque la covariance est non nulle, les variables sont nécessairement non indépendantes. Cela ne signifie pas qu'il existe une relation de cause à effet entre les deux phénomènes mesurés par les variables (par exemple, le nombre de coups de soleil observés dans une station balnéaire est certainement corrélé au nombre de lunettes de soleil vendues mais aucun des deux phénomènes n'est la cause de l'autre) mais cela donne une indication statistique de la corrélation des deux phénomènes, c'est-à-dire une mesure de l'intensité de la liaison mathématique qui peut exister entre les variables.

En particulier, on peut interpréter le signe du coefficient de corrélation (c'est-à-dire de la covariance) de la façon suivante :

- si $\rho(X, Y) > 0$, alors X et Y ont tendance à varier « dans le même sens » (c'est ce que l'on peut attendre de variables qui mesurent le poids et la taille des individus : « plus la taille augmente, plus le poids augmente ») ;
- si $\rho(X, Y) < 0$, alors X et Y ont tendance à varier « dans des sens contraires » (c'est ce que l'on peut attendre de variables qui mesurent la quantité de testostérone et le nombre de cheveux des individus d'une population).

Il faut toutefois rester très prudent sur l'interprétation que l'on fait de la non nullité du coefficient de corrélation. En effet, la covariance est extrêmement sensible à la présence de valeurs aberrantes ou extrêmes (ces valeurs sont appelées des « déviants ») dans notre ensemble de données (valeurs très éloignées de la majorité des autres, pouvant être considérées comme des exceptions).

D.3. Loi faible des grands nombres

Le théorème suivant, communément appelé la loi faible des grands nombres, vient confirmer l'intuition selon laquelle plus on effectue de répétitions indépendantes d'une même expérience aléatoire, plus la moyenne des valeurs observées s'approche de l'espérance.

Théorème 1

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles finies indépendantes et de même loi (on parle de v.a.i.i.d. pour « variables aléatoires indépendantes identiquement distribuées »). On note m l'espérance commune des X_i. Pour tout $\varepsilon > 0$, on a

$$\lim_{n\to+\infty} P\Big(\Big|\frac{X_1+\cdots+X_n}{n} - m\Big| \geqslant \varepsilon\Big) = 0.$$

■ Commençons par dire que l'existence d'un espace probabilisé $(\Omega, \mathscr{T}, P)$ sur lequel existe une telle suite de variables aléatoires indépendantes identiquement distribuées est admise !

Notons σ^2 la variance commune des X_i. Alors, par indépendance, on a

$$V\Big(\frac{X_1+\cdots+X_n}{n}\Big) = \frac{V(X_1)+\cdots+V(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}\cdot$$

Pour tout $\varepsilon > 0$, l'inégalité de Bienaymé–Tchebychev appliquée à $(X_1+\cdots+X_n)/n$ donne

$$P\Big(\Big|\frac{X_1+\cdots+X_n}{n} - m\Big| \geqslant \varepsilon\Big) \leqslant \frac{\sigma^2}{n\varepsilon^2},$$

ce qui permet de conclure en vertu du théorème des gendarmes. ■

On dit que la moyenne empirique $(X_1+\cdots+X_n)/n$ converge en probablité vers m.

Exemples :

- On considère une suite d'épreuves indépendantes ainsi qu'un événement A fixé. Pour tout $n \in \mathbb{N}^*$, on note X_n l'indicatrice de l'événement A lors de la n-ème épreuve (X_n vaut 1 si l'événement A est réalisé lors de la n-ème épreuve et vaut 0 dans le cas contraire). La suite $(X_n)_{n\geqslant 1}$ est ainsi constituée de variables aléatoires indépendantes de même loi $\mathscr{B}\big(P(A)\big)$.

  On note alors M_n la fréquence de réalisation de A lors des n premières épreuves, c'est-à-dire

  $$M_n = \frac{X_1+\cdots+X_n}{n}\cdot$$

  La loi faible des grands nombres nous dit alors que, pour tout $\varepsilon > 0$, on a

  $$\lim_{n\to+\infty} P(|M_n - P(A)| \geqslant \varepsilon) = 0.$$

  C'est ce résultat qui justifie l'idée intuitive que la probabilité d'un événement est la limite de sa fréquence de réalisation lorsqu'on répète indépendamment l'expérience un grand nombre de fois.

4 h 15