CNAM MULHOUSE . Ramm Algebra Center

Formation Ingénieur Informatique

Mathématiques : PROBABILITES

Cours Michel GOZE

Chapitre 8

Vecteurs aléatoires

1. Couple de variables aléatoires

On va s'intéresser dans ce chapitre à une étude globale de plusieurs variables aléatoires sur un même espace probabilité. Nous allons commencer par le cas d'un couple de variables

1.1. **Fonction de répartition d'un couple de variables aléatoires.** Un vecteur aléatoire $\mathbf{X} = (X_1, X_2)$ de dimension 2 sur un espace probabilisé (Ω, P), appelé communément couple de variables aléatoires est défini comme une application

$$\mathbf{X} : \Omega \to \mathbb{R}^2$$

à valeurs dans le plan $\mathbb{R}^2$.

Exemple. On lance deux dés simultanément. On note par X la variable égale au plus petit des deux nombres apparus et par Y le plus grand. On définit ainsi un couple (X, Y) correspondant à la donnée du plus petit et du plus grand, dans cet ordre, des nombres apparus.

Définition 1. *On appelle fonction de répartition du couple* $\boldsymbol{X} = (X_1, X_2)$ *la fonction*

$$F_{\mathbf{X}}(x_1, x_2) = P\{X_1 \leq x_1,\ X_2 \leq x_2\}.$$

Cette fonction vérifie

(1) $F_{\mathbf{X}}(-\infty, x_2) = F_{\mathbf{X}}(x_1, -\infty) = 0$,
(2) $F_{\mathbf{X}}(+\infty, +\infty) = 0$.

Elle est également croissante en chacune de ses variables.

Cette fonction correspond à la probabilité pour que X soit inférieur à x et Y à y.

Lorsque Ω est un ensemble fini, X_1 et X_2 sont des variables aléatoires finies. Notons par $x_1^1, \cdots, x_s^1$ les valeurs prises par X_1 et par $x_1^2, \cdots, x_r^2$ celles par X_2. On suppose ces valeurs ordonnées. Pour représenter la loi du couple $\mathbf{X} = (X_1, X_2)$, on peut tracer un tableau à double entrée, les lignes correspondant aux valeurs de X_1 et les colonnes à celles de X_2. La case correspondant à la i-ème ligne et j-ième colonne contiendra la valeur

$$p_{i,j}(X_1, X_2) = P((X_1 = x_i^1) \bigcap (X_2 = x_j^2)).$$

Nous détaillerons ceci un peu plus loin.

1.2. **La loi de probabilité du couple $\mathbf{X} = (X_1, X_2)$ ou loi conjointe.** Rappelons que la loi de probabilité d'une variable aléatoire X sur Ω est la probabilité

$$P_X(]a,b]) = P\{a < X \leq b\} = P(\{\omega \in \Omega,\ a < X(\omega) \leq b\}).$$

Considérons à présent le couple $\mathbf{X} = (X_1, X_2)$. L'évènement $\{a_1 < X_1 \leq b_1, X_2 \leq x_2\}$ est formé des points (x_1, x_2) de $\mathbb{R}^2$ tels que $a_1 < X_1 \leq b_1$ et $X_2 \leq x_2$. On en déduit

$$P(\{a_1 < X_1 \leq b_1,\ X_2 \leq x_2\} = F(b_1, x_2) - F(a_1, x_2).$$

De même, on aura

$$P(\{X_1 \leq x_1,\ a_2 < X_2 \leq b_2\} = F(x_1, b_2) - F(x_1, a_2).$$

Définition 2. *La loi de probabilité du couple $\boldsymbol{X} = (X_1, X_2)$ de variable aléatoires sur Ω est donnée pour tout rectangle $R =]a_1, b_1] \times]a_2, b_2]$ du plan $\mathbb{R}^2$ par*

$$P_{\mathbf{X}} = P(\{a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2\}).$$

On a donc

$$P_{\mathbf{X}} = F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2) + F_{\mathbf{X}}(a_1, a_2).$$

C'est donc la probabilité pour que $\mathbf{X} = (X_1, X_2)$ soit dans le rectangle R.

1.3. **Lois marginales.** Soit $\mathbf{X} = (X_1, X_2)$ un couple de variable aléatoires sur Ω.

Définition 3. *On appelle loi marginale de X_i du couple $\boldsymbol{X}$ la loi suivie par la seule variable X_i.*

Connaissant la fonction de répartition de $\mathbf{X}$, il est facile de trouver la loi marginale de chacune des variables X_i. On aura par exemple

$$F_{X_1}(x_1) = F(x_1, +\infty)$$

et

$$F_{X_2}(x_2) = F(+\infty, x_2).$$

1.4. **Cas d'un couple de variables aléatoires finies.** Supposons que $\Omega = \{\omega_1, \cdots, \omega_n\}$ soit fini. Dans ce cas, la loi sera représentée par un tableau à double entrée. Les lignes correspondent aux valeurs $P(X_1(\omega_i))$ et les colonnes aux valeurs $P(X_2(\omega_i))$. On notera $p_{ij} = P(X_1 = \omega_I, X_2 = \omega_j)$. Le tableau se présentera ainsi

$X_2(\omega_j)/X_1(\omega_i)$	$X_1(\omega_1)$	$X_1(\omega_2)$	$\cdots$	$X_1(\omega_n)$
$X_2(\omega_1)$	p_{11}	p_{21}	$\cdots$	p_{n1}
$X_2(\omega_2)$	p_{12}	p_{22}	$\cdots$	p_{n2}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$X_2(\omega_n)$	p_{1n}	p_{2n}	$\cdots$	p_{nn}

Exemples. 1. Une urne contient 3 boules blanches et 4 boules noires. On tire deux boules de l'urne. Soit X_1 la variable aléatoire égale à 1 si la première boule est blanche, à 0 sinon. Soit X_2 la variable aléatoire égale à 1 si la deuxième boule est blanche, à 0 sinon. Déterminons la loi conjointe du couple (X_1, X_2). On suppose que les tirages se font sans remise. Le tableau est ainsi :

$X_2(\omega_j)/X_1(\omega_i)$	0	1
0	2/7	2/7
1	2/7	1/7

En effet, p_{11} correspond à $X_1 = 0$ et $X_2 = 0$, c'est-à-dire la première boule est noire et la deuxième est aussi noire. Comme il n'y a pas de remise, la probabilité pour que la première boule soit noire est 4/7 et la deuxième moire, sachant que la première était noire est 3/6. Ainsi $p_{11} = \frac{4}{7}\frac{3}{6} = \frac{2}{7}$. On détermine de la même façon les trois autres probabilités.

2. On lance un dé et X correspond au plus petit des deux chiffres apparus et Y le plus grand. Le tableau est le suivant

(X/Y)	1	2	3	4	5	6
1	1/36	1/18	1/18	1/18	1/18	1/18
2	0	1/36	1/18	1/18	1/18	1/18
3	0	0	1/36	1/18	1/18	1/18
4	0	0	0	1/36	1/18	1/18
5	0	0	0	0	1/36	1/18
6	0	0	0	0	0	1/36

La définition des lois marginales dans le cas fini se ramène à la définition suivante.

Notons par $x_{1,i}$ les valeurs de X_1, c'est-à-dire $x_{1,i} = X_1(\omega_i)$ et par $x_{2,i}$ celles de X_2.

Définition 4. *Soit $\boldsymbol{X} = (X_1, X_2)$ un couple de variables aléatoires sur Ω de loi $P_{\mathbf{X}}$. Alors les applications*

$$P_{X_1}(x_{1,i}) = P\{X_1 = x_{1,i}\} = p_{i,\bullet}$$

et

$$P_{X_2}(x_{2,j}) = P\{X_2 = x_{2,j}\} = p_{\bullet,j}$$

sont appelées les lois marginales du couple $\boldsymbol{X} = (X_1, X_2)$. Ce sont les lois de probabilité de chacune des variables X_1 et X_2.

Il est clair que l'on a

$$p_{i,\bullet} = \sum_{j=1}^{n} p_{ij}, \quad p_{\bullet,j} = \sum_{i=1}^{n} p_{ij}$$

et

$$\sum_{i,j} p_{ij} = 1.$$

Dans l'exemple précédent, on peut compléter le tableau en inscrivant les lois marginales. On obtient

$X_2(\omega_j)/X_1(\omega_i)$	0	1	P_{X_1}
0	2/7	2/7	4/7
1	2/7	1/7	3/7
P_{X_2}	4/7	3/7	1

Notons enfin que la seule connaissance des lois marginales ne suffit pas pour déterminer la loi du couple. Par exemple, dans le cas ci-dessus les lois marginales correspondent aux deux tableaux

X_1	0	1
	4/7	3/7

,

X_2	0	1
	4/7	3/7

et la donnée de ces deux tableaux ne permet pas de reconstruire le tableau du couple.

Soit $\mathbf{X} = (X_1, X_2)$ un couple de variable aléatoires sur (Ω, P).

Définition 5. *Pout tout $x_2 \in X_2(\Omega)$ tel que $P(X_2 = x_2) \neq 0$, la loi de $P_{X_2=x_2}$ de X_1 sachant $\{X_2 = x_2\}$ est donnée par*

$$P_{X_2=x_2}(X_1 = x_1) = P(X_1 = x_1, X_2 = x_2) = \frac{P_{\boldsymbol{X}}(x_1, x_2)}{P(X_2 = x_2)}.$$

De même, pout tout $x_1 \in X_1(\Omega)$ tel que $P(X_1 = x_1) \neq 0$, la loi de $P_{X_1=x_1}$ de X_2 sachant $\{X_1 = x_1\}$ est donnée par

$$P_{X_1=x_1}(X_2 = x_2) = P(X_1 = x_1, X_2 = x_2) = \frac{P_{\boldsymbol{X}}(x_1, x_2)}{P(X_1 = x_1)}.$$

Chacune de ces applications , $P_{X_2=x_2}$ et $P_{X_1=x_1}$ sont des probabilités sur Ω. Déterminons ces lois dans l'exemple précédent :

$X_1(\omega_i)$	0	1
$P_{X_2=0}$	1/2	1/2
$P_{X_2=1}$	2/3	1/3

,

$X_2(\omega_i)$	0	1
$P_{X_1=0}$	1/2	1/2
$P_{X_1=1}$	2/3	1/3

1.5. **Couple de variables aléatoires continues.** On généralise la notion de densité de probabilité ainsi :

Définition 6. *On appelle densité de probabilité conjointe du couple $\boldsymbol{X} = (X_1, X_2)$ la fonction*

$$f_{\boldsymbol{X}}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{\boldsymbol{X}}(x_1, x_2).$$

Evidemment on suppose que la fonction de répartition admet ces dérivées partielles. Rappelons brièvement ces notions. Soit $f(x_1, x_2)$ une fonction de deux variables. Soit x_2 fixé et considérons la fonction g_1 d'une variable réelle définie par

$$g_1(x_1) = f(x_1, x_2).$$

Dans cette définition la variable est x_1 et x_2 est considérée comme un paramètre. Supposons g_1 dérivable. Alors la dérivée de g_1 est la dérivée partielle de f par rapport à x_1. Elle est notée :

$$g_1'(x_1) = \frac{\partial}{\partial x_1} f(x_1, x_2).$$

De même, si x_1 est fixé, on considère la fonction d'une variable

$$g_2(x_2) = f(x_1, x_2).$$

Dans cette définition la variable est x_2 et x_1 est considérée comme un paramètre. Supposons g_2 dérivable. Alors la dérivée de g_2 est la dérivée partielle de f par rapport à x_2. Elle est notée :

$$g_2'(x_2) = \frac{\partial}{\partial x_2} f(x_1, x_2).$$

Prenons par exemple $f(x_1, x_2) = x_1^2 + 3x_1x_2 + 4x_2^3$. Alors

$$\frac{\partial}{\partial x_1} f(x_1, x_2) = 2x_1 + 3x_2, \quad \frac{\partial}{\partial x_2} f(x_1, x_2) = 3x_1 + 12x_2^2.$$

Considérons maintenant la fonction de deux variables $\frac{\partial}{\partial x_1} f(x_1, x_2)$. Elle admet deux dérivées partielles, l'une par rapport à x_1 l'autre par rapport à x_2. On les notes

(1) $\frac{\partial}{\partial x_1}(\frac{\partial}{\partial x_1} f(x_1, x_2)) = \frac{\partial^2}{\partial x_1 \partial x_1} f(x_1, x_2) = \frac{\partial^2}{\partial x_1^2} f(x_1, x_2),$

(2) $\frac{\partial}{\partial x_2}(\frac{\partial}{\partial x_1} f(x_1, x_2)) = \frac{\partial^2}{\partial x_1 \partial x_2} f(x_1, x_2)$

De même, en considérant la fonction de deux variables $\frac{\partial}{\partial x_2} f(x_1, x_2)$, elle admet deux dérivées partielles, l'une par rapport à x_1 l'autre par rapport à x_2. On les note

(1) $\frac{\partial}{\partial x_1}(\frac{\partial}{\partial x_2} f(x_1, x_2)) = \frac{\partial^2}{\partial x_2 \partial x_1} f(x_1, x_2)$

(2) $\frac{\partial}{\partial x_2}(\frac{\partial}{\partial x_1} f(x_1, x_2)) = \frac{\partial^2}{\partial x_2 \partial x_2} f(x_1, x_2) = \frac{\partial^2}{\partial x_2^2} f(x_1, x_2),$

On notera qu'en général

$$\frac{\partial^2}{\partial x_1 \partial x_2} f(x_1, x_2) = \frac{\partial^2}{\partial x_2 \partial x_1} f(x_1, x_2).$$

Dans l'exemple précédent, on a donc

$$\frac{\partial^2}{\partial x_1^2} f(x_1, x_2) = 2, \ \frac{\partial^2}{\partial x_1 \partial x_2} f(x_1, x_2) = 3, \ \frac{\partial^2}{\partial x_2 \partial x_1} f(x_1, x_2) = 24x_2.$$

1.6. **Variables aléatoires indépendantes.** Soit $\mathbf{X} = (X_1, X_2)$ un couple de variables aléatoires.

Définition 7. *On dit que les variables X_1 et X_2 sont indépendantes si la fonction de répartition du couple vérifie*

$$F_{\mathbf{X}}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2)$$

pour tout $x_1, x_2 \in \mathbb{R}$.

Lorsque les variables aléatoires X_1 et X_2 sont finies, elles sont indépendantes si et seulement si

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2).$$

Dans ce cas, la donnée des lois marginales permet de trouver la loi conjointe, ce qui en général faux si les variables ne sont pas indépendantes.

Si elles sont continues de densité respectives f_{X_1} et f_{X_2} alors l'indépendance se traduit par

$$f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2).$$

Exemple Deux personnes se donnent un rendez-vous. L'heure d'arrivée de ces deux personnes est une variable aléatoire uniforme répartie sur l'intervalle $(20h, 21h)$. Ces deux variables sont indépendantes. Quelle est la probabilité pour que la première personne arrivée attende plus de 10 minutes ?

Désignons par X et Y les deux variables correspondant aux dates d'arrivée en minutes de chacun des individus. Il s'agit de deux variables indépendantes de loi uniforme entre 0 et 60. On peut supposer ces variables continues de densité

$$f_X(x) = \frac{1}{60},\ f_Y(y) = \frac{1}{60}.$$

Comme ces variables sont indépendantes

$$P(X + 10 < Y) = \int\int_{x+10<y} f_{\mathbf{X}}(x, y)dxdy = \int_{10}^{60}\int_{0}^{y-10} (\frac{1}{60})^2 dxdy = \frac{25}{36}.$$

1.7. **Indépendance et corrélation.** Dans les chapitres précédents, nous avons introduit la notion de corrélation entre deux variables aléatoires. Rappelons les principales notions.

(1) Si X_1 et X_2 sont finies d'espérance $E(X_1)$ et $E(X_2)$, alors la covariance de X_1 et X_2 est le réel
$$Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$$
sachant que $E(X_1X_2) = \sum x_{1,i}x_{2,j}p_{ij}$.

(2) Si X_1 et X_2 sont continues d'espérance $E(X_1)$ et $E(X_2)$, alors la covariance de X_1 et X_2 est le réel
$$\int\int_{\mathbb{R}^2} (x_1 - E(X_1))(x_2 - E(X_2))f_{\mathbf{X}}(x_1, x_2)dx_1dx_2.$$

(3) Le coefficient de corrélation du couple $\mathbf{X} = (X_1, X_2)$ est le réel
$$r(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}}$$
où $Var(X)$ est la variance de X.

Proposition 1. *Si les variables aléatoires X_1 et X_2 sont indépendantes, alors elles sont décorrélées, c'est-à-dire vérifient*

$$r(X_1, X_2) = 0.$$

La réciproque est fausse en général.

2. CAS GÉNÉRAL : VECTEURS ALÉATOIRES

2.1. **Fonction de répartition, densité conjointe.** Nous avons défini les vecteurs aléatoires à n dimensions comme un n-uplet de variables aléatoires :

$$\mathbb{X} = (X_1, X_2, \cdots, X_n)$$

sur un même espace probabilisé Ω. Sa fonction de répartition est donnée par

$$F_{\mathbb{X}}(x_1, \cdots, x_n) = P(X_i \leq x_i,\ i = 1, \cdots, n).$$

Elle vérifie

(1) $F_{\mathbb{X}}(-\infty, x_2, \cdots, x_n) = \cdots = F_{\mathbb{X}}(x_1, x_2, \cdots, \cdots, x_n) = F_{\mathbb{X}}(x_1, x_2, \cdots, -\infty) = 0.$

(2) $F_{\mathbb{X}}(+\infty, \cdots, +\infty) = 1.$

(3) $F_{\mathbb{X}}$ est croissante en chacune de ses variables.

Si chacune des variables est continue, la densité du vecteur X sera donnée par

$$f_{\mathbf{X}}(x_1, \cdots, x_n) = \frac{\partial^n F_{\mathbf{X}}}{\partial x_1 \partial x_2 \cdots \partial x_n}.$$

2.2. **Changement de variables.** Considérons n fonctions de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^n$:

$$g_i : \mathbb{R}^n \to \mathbb{R}^n.$$

On suppose que ces fonctions admettent des dérivées partielles aux ordres souhaités et qu'elles sont bijectives ce qui signifie que le système $g_k(x_1, \cdots, x_n) = y_k,\ k = 1, \cdots, n$ admet une solution. On considère un vecteur aléatoire $\mathbf{X} = (X_1, \cdots, X_n)$ de densité $f_{\mathbf{X}}(x_1, \cdots, x_n)$. On souhaite trouver la densité du nouveau vecteur

$$\mathbf{Y} = (g_1(X_1), \cdots, g_n(X_n)).$$

Considérons la matrice

$$\begin{pmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} & \cdots & \dfrac{\partial g_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial g_n}{\partial x_1} & \dfrac{\partial g_n}{\partial x_2} & \cdots & \dfrac{\partial g_n}{\partial x_n} \end{pmatrix}$$

et notons par

$$J_{g_1, \cdots, g_n}(x_1, \cdots, x_n)$$

son déterminant. Alors la fonction densité du vecteur $\mathbf{Y}$ est donnée par

$$f_{\mathbf{Y}}(y_1, \cdots, y_n) = \frac{f_{\mathbf{X}}(x_1, \cdots, x_n)}{|J_{g_1, \cdots, g_n}(x_1, \cdots, x_n)|}. \tag{1}$$

Exemple. Considérons un vecteur $\mathbf{X} = (X_1, X_2)$ et le changement de variables

$$Y_1 = X_1 + X_2,\ Y_2 = X_1 - X_2.$$

On a donc

$$g_1(x_1, x_2) = x_1 + x_2,\ g_2(x_1 + x_2) = x_1 - x_2.$$

Ces deux fonctions sont bijectives, les équations $y_1 = x_1 + x_2,\ y_2 = x_1 - x_2$ ont pour solution

$$x_1 = \frac{y_1 + y_2}{2},\ x_2 = \frac{y_1 - y_2}{2}.$$

La fonction

$$J_{g_1,,g_2}(x_1, x_2)$$

est le déterminant de la matrice

$$\begin{pmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} \\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

et donc

$$J_{g_1,,g_2}(x_1, x_2) = -2.$$

On en déduit, si $\mathbf{Y} = (Y_1, Y_2)$:

$$f_{\mathbf{Y}}(y_1, y_2) = \frac{f_{\mathbf{X}}(x_1, x_2)}{|J_{g_1,g_2}(x_1, x_2)|} = \frac{1}{2} f_{\mathbf{X}}(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}).$$

Remarque : cas où le changement n'est pas bijectif. Nous avons supposé ci-dessus que le système d'équations $y_i = g_i(x_1, \cdots, x_n),\ i = 1, \cdots, n$ avait une et une seule solution. Supposons qu'il admette un nombre fini de solutions $(x_1^p, \cdots, x_n^p)$ pour $p = 1, \cdots, s$ (précédemment on avait $s = 1$). Dans ce cas on aura

$$(2) \qquad f_{\mathbf{Y}}(y_1, \cdots, y_n) = \frac{f_{\mathbf{X}}(x_1^1, \cdots, x_n^1)}{|J_{g_1,\cdots,g_n}(x_1^1, \cdots, x_n^1)|} + \cdots + \frac{f_{\mathbf{X}}(x_1^s, \cdots, x_n^s)}{|J_{g_1,\cdots,g_n}(x_1^s, \cdots, x_n^s)|}.$$

2.3. **Matrice de covariance.** Soit $\mathbb{X} = (X_1, X_2, \cdots, X_n)$ un vecteur aléatoire. Pour chacun des couples (X_i, X_j), $i, j = 1, \cdots, n$, notons par

$$C_{ij} = Cov(X_i, X_j)$$

la covariance des deux variables X_i et X_j.

Définition 8. *Les variables aléatoires $X_1, X_2, \cdots, X_n$ sont mutuellement non-corrélées si $C_{ij} = 0$ pour tout $i, j = 1, \cdots, n$ et $i \neq j$.i*

Par exemple, si les variables $X_1, X_2, \cdots, X_n$ sont indépendantes, elles sont non-corrélées, l'inverse n'étant pas en général vrai.

Définition 9. *On appelle matrice de covariance du vecteur aléatoire $\mathbb{X} = (X_1, X_2, \cdots, X_n)$ la matrice carrée $C_{\mathbb{X}}$ de taille n dont les coefficients sont*

$$C_{ij} = Cov(X_i, X_j)$$

pour $i, j = 1, \cdots, n$.

Rappelons que pour deux variables aléatoires X et Y,

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E((X - E(X))(Y - E(Y))).$$

Dans le cas fini, ce nombre existe toujours, dans le cas continu, on peut rencontrer un problème d'existence. Lorsqu'on parlera de covariance, on supposera que ce nombre existe. La covariance vérifie

(1) $Cov(X,Y) = Cov(Y,X)$

(2) $Cov(X,X) = Var(X)$

(3) $Cov(a_1X_1 + a_2X_2, Y) = a_1Cov(X_1,Y) + a_2Cov(X_2,Y)$.

(4) Si X et Y sont indépendantes, $Cov(X,Y) = 0$.

On en déduit que la matrice carrée $C_{\mathbb{X}}$ est symétrique, c'est-à-dire égale à sa transposée (la transposée d'une matrice est la matrice dont la ligne numéro i est la colonne numéro i de la matrice initiale). Par exemple, pour $n = 2$, on a le vecteur $\mathbb{X} = (X_1, X_2)$ et

$$C_{\mathbb{X}} = \begin{pmatrix} Var(X_1) & Cov(X_1,X_2) \\ Cov(X_1,X_2) & Var(X_2) \end{pmatrix}$$

Proposition 2. *Soit $\mathbb{X} = (X_1, X_2, \cdots, X_n)$ tel que les variables $X_1, \cdots, X_n$ soient indépendantes. Alors la matrice $C_{\mathbb{X}}$ est une matrice diagonale.*

En effet, dans ce cas $C_{ij} = 0$ et

$$C_{\mathbb{X}} = \begin{pmatrix} Var(X_1) & 0 & \cdots & 0 \\ 0 & Var(X_2) & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & Var(X_n) \end{pmatrix}$$

Remarque. La matrice de covariance est une matrice symétrique. Ces matrices jouent un rôle important en mathématiques et surtout en géométrie euclidienne ou pseudo-euclidienne. Elles sont associées à des notions de distance. Ici cette matrice donne une indication de la dispertion du vecteur $\mathbb{X} = (X_1, X_2, \cdots, X_n)$ autour de sa moyenne. On peut la considérer comme une matrice d'inertie, d'où l'utilité d'étudier cette matrice en termes de réduction de matrices en calculant les valeurs propres ou les directions invariantes.

2.4. **Vecteurs gaussiens.** On veut déterminer dans quelles conditions la réciproque de la proposition précédente est vraie.

Définition 10. *Un vecteur $\mathbb{X} = (X_1, X_2, \cdots, X_n)$ est dit gaussien lorsque toute combinaison linéaire $a_1X_1 + \cdots + a_nX_n$ de ses composantes X_i suit une même loi normale $\mathcal{N}(m, \sigma^2)$.*

On ne s'étendra pas sur les propriétés suivantes, on les donne pour illustrer la diagonalisation de la matrice de covariance. Considérons tout d'abord un vecteur $\mathbb{X} = (X_1, X_2, \cdots, X_n)$ dont les composantes sont indépendantes et suivent la loi normale $\mathcal{N}(0,1)$. Alors un tel vecteur est gaussien. Comme les composantes sont indépendantes $C_{\mathbb{X}}$ est diagonale. Inversement, soit $\mathbb{X} = (X_1, X_2, \cdots, X_n)$ un vecteur gaussien. Si la matrice de covariance est diagonale, alors les composantes de ce vecteur sont indépendantes.

3. Somme et produit de variables aléatoires

Pour bien comprendre le comportement de la somme de deux variables aléatoires, commençons par un exemple, tiré mots pour mots de l'excellent livre de Walter Appel. Supposons que je mange tous les jours une quantité de chocolat comprise entre 0 et 100 gr. On modélise cette quantité par la variable aléatoire X_1 pour le premier jour, et X_2 pour le deuxième jour. On suppose que ces deux variables suivent la même loi uniforme sur $[0, 100]$. Quelle est la loi de probabilité correspondant à la quantité de chocolat mangée en deux jours. Supposons, pour simplifier, que la quantité de chocolat mangée ne peut prendre que des valeurs entières comprise entre 0 et 100 avec une probabilité $P(k) = 1/100$. On suppose toutefois que l'on mange effectivement chaque jour une quantité non nulle. Ainsi la probabilité pour que je mange 2 gr en deux jours correspond à $X_1 = 1$ et $X_2 = 1$. Cette probabilité est donc égale à $1/100 \times 1/100 = 1/10000$. La probabilité pour que je mange $3gr$ en deux jours corresponds aux valeurs $(X_1, X_2) = (1, 2)$ ou $(2, 1)$. Cette probabilité est donc égale à $2/10000$. De même, la probabilité pour que je mange $4gr$ est donnée par $(X_1, X_2) = (1, 3), (2, 2), (3, 1)$ et donc égale à $3/10000$ soit $P(X_1 = 1)P(X_2 = 3) + P(X_1 = 2)P(X_2 = 2) + P(X_1 = 3)P(X_2 = 1)$. On se dirige donc vers la formule

$$P(n) = P(X_1 = 1)P(X_2 = n-1) + P(X_1 = 2)P(X_2 = n-2) + \cdots + P(X_1 = n-1)P(X_2 = 1).$$

Avant de s'intéresser à la loi de la somme de deux variables, faisons un petit retour sur les invariants associés.

3.1. **Somme de deux variables aléatoires.** Rappelons que si X est une variable aléatoire, son espérance est donnée par

(1) $E(X) = \sum_k P(X = x_k)x_k$ lorsque X ne prend que des valeurs finies $x_1, \cdots, x_n$,
(2) $E(X) = \int_{-\infty}^{+\infty} tf(t)dt$ pour une variable continue de densité $f(x)$.

On déduit immédiatement

Proposition 3. *Soit $\mathbb{X} = (X_1, X_2)$ un couple de variables aléatoires. Alors*

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

(en supposant, dans le cas continu ou infini, que ces valeurs existent).

Démonstration. La démonstration de cette identité dans le cas fini est assez simple. Si X_1 prend les valeurs $x_1^1, \cdots, , x_n^1$ avec les probabilités $p_1^1, \cdots, p_n^1$ et X_2 les valeurs $x_1^2, \cdots, , x_n^2$ avec les probabilités $p_1^2, \cdots, p_n^2$, alors $X_1 + X_2$. Soit p_{ij} la probabilité $P(X_1 = x_i^1, X_2 = x_j^2)$. Alors

$$E(X_1 + X_2) = (\sum (x_i^1 + x_j^2)p_{ij})/N.$$

Comme $\sum_j x_i^1 p_{ij} = x_i^1 p_i^1$, on en déduit $E(X_1 + X_2) = E(X_1) + E(X_2)$.

Dans le cas continu, si $f_{\mathbf{X}}$ est la densité du couple $\mathbf{X} = (X_1, X_2)$, alors les lois marginales de X_1 et X_2 sont données par les densités

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, x_2)dx_2, \; f_{X_2}(x_2) = \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, x_2)dx_1.$$

On en déduit

$$E(X_1 + X_2) = \int\int_{\mathbb{R}^2} (x_1 + x_2) f_{\mathbf{X}}(x_1, x_2)dx_1dx_2 = \int_{-\infty}^{+\infty} x_1 f_{\mathbf{X_1}}(x_1)dx_1 + \int_{-\infty}^{+\infty} x_2 f_{\mathbf{X_2}}(x_2)dx_2$$

et donc

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

Remarquons que cette propriété se généralise aisément : pour tout vecteur $\mathbb{X} = (X_1, \cdots, X_n)$ de variable aléatoire, et tout $(a_1, a_2, \cdots, a_n) \in \mathbb{R}^n$,

$$E(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = a_1 E(X_1) + a_2 E(X_2) + \cdots + a_n X_n.$$

La démonstration précédente montre, dans le cas continu, l'intérêt de connaître la densité de probabilité de la somme $X_1 + X_2$ et d'écrire

$$E(X_1 + X_2) = \int_{-\infty}^{+\infty} u f_{X_1+X_2}(u) du.$$

Théorème 1. *Soit* $\mathbb{X} = (X_1, X_2)$ *un couple de variables aléatoires* **indépendantes** *possédant des densités* f_{X_1} *et* f_{X_2}. *Alors la densité de probabilité de* $X_1 + X_2$ *est*

$$f_{X_1+X_2} = f_{X_1} * f_{X_2}$$

soit

$$f_{X_1+X_2}(x) = \int_{-\infty}^{+\infty} f_{\mathbf{X_1}}(t) f_{\mathbf{X_2}}(x-t) dt.$$

Ce produit $f_{X_1} * f_{X_2}$ s'appelle le produit de convolution des fonctions f_{X_1} et f_{X_2}. Rappelons que, comme nous supposons ici les variables indépendantes, la densité conjointe du couple est le produit (ordinaire) des densités :

$$f_{\mathbb{X}}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2).$$

Nous avons donc maintenant l'explication de notre histoire de chocolat.

La variance d'une variable aléatoire étant définie à partir de relations de degré 2 sur les variables, on ne peut espérer en général des relations de linéarité comme pour l'espérance qui elle est donnée par des relations de degré 1. Toutefois on a le résultat suivant dans le cas des variables indépendantes :

Théorème 2. Théorème de Bienaymé. *Soit* $\mathbb{X} = (X_1, X_2)$ *un couple de variables aléatoires* **indépendantes** *possédant chacune des variances* $Var(X_1)$ *et* $Var(X_2)$. *Alors la variance de* $X_1 + X_2$ *est*

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2).$$

Ceci se généralise à n variables aléatoires indépendantes :

$$Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n).$$

La démonstration de ce résultat peut se baser sur le comportement du produit de variables aléatoires.

3.2. **Produit de deux variables aléatoires.** Considérons un couple $\mathbb{X} = (X_1, X_2)$ de variables aléatoires. En général il est difficile de dire quelque chose de général sur $E(X_1X_2)$. Toutefois, dans le cas indépendant, on a

Théorème 3. *Soit* $\mathbb{X} = (X_1, X_2)$ *un couple de variables aléatoires* **indépendantes** *possédant des espérances* $E(X_1)$ *et* $E(X_2)$*. Alors*
$$E(X_1X_2) = E(X_1)E(X_2).$$

Démonstration. En effet, si on regarde cela dans le cas continu, on a par hypothèse $f_{\mathbb{X}}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ et
$$E(X_1X_2) = \int\int_{\mathbb{R}^2} x_1x_2f_{\mathbb{X}}(x_1, x_2)dx_1dx_2 = \int_{-\infty}^{+\infty} x_1f_{X_1}(x_1)dx_1 \cdot \int_{-\infty}^{+\infty} x_2f_{X_2}(x_2)dx_2$$
et
$$E(X_1X_2) = E(X_1)E(X_2).$$

4. Loi des grands nombres

4.1. **L'inégalité de Chebycheff.** Soit X une variable aléatoire sur un espace probabilité (Ω, P). Soit m sa moyenne et σ son écart type. Rappelons que la moyenne est une autre appellation de l'espérance :
$$m_X = E(X) = \sum x_iP(X = x_i),\ \sigma_X^2 = E((X - m_X)^2).$$

Théorème 4. Inégalité de Chebycheff *Pour tout nombre* t *positif, on a*
$$P(|X - m| > t) \leq \frac{\sigma^2}{t^2}.$$

Cette inégalité permet de montrer qu'une variable aléatoire prendra avec une grande probabilité une valeur relativement proche de sa moyenne. Elle donne un moyen d'évaluer la distance entre les valeurs prises par X et son espérance. Plus précisément elle donne une majoration de la probabilité que l'écart soit grand. Cette inégalité est la conséquentce immédiate du résultat suivant (uniquement cité ici dans le cas fini) : si X ne prend que des valeurs positives, alors pour tout $t > 0$,
$$P(X \geq t) \leq \frac{E(X)}{t}.$$
En effet si $x_1, \cdots, x_n$ sont les valeurs (positives) prises par X, alors si $x_k, x_{k+1}, \cdots, x_n$ sont les valeurs supérieures à t,
$$P(X \geq t) = P(X = x_k) + \cdots + P(X = x_n)$$
et comme $0 < t \leq x_k \leq \cdots \leq x_n$,
$$tP(X \geq t) \leq x_kP(X = x_k) + \cdots + x_nP(X = x_n)$$

et donc comme tous les x_i sont positifs

$$tP(X \geq t) \leq \sum_{i=1}^{n} x_i P(X = x_i) = E(X).$$

En appliquant cette inégalité à la variable $(X - E(X))^2$, on obtient l'inégalité de Chebycheff.

4.2. **Théoreme de Bernoulli.** Enonçons tout d'abord ce résultat.

Théorème 5. Théorème de Bernoulli. *Soient $X_1, \cdots, X_n$ des variables aléatoires indépendantes ayant toutes la même loi de probabilité, la même espérance m et le même écart type. Posons*

$$Y = \frac{1}{n}(X_1 + \cdots + X_n).$$

Alors pour tout $\varepsilon > 0$,

$$P(|Y - m| > \varepsilon) \to 0$$

quand $n \to +\infty$.

On utilise souvent ce théorème pour trouver une approximation de la probabilité d'un évènement. Avant de le démontrer, donnons une illustration "historique". Supposons que la variable $nY = X_1 + \cdots + X_n$ suive la loi binomiale $\mathcal{B}(p, q, n)$. Dans ce cas nY est considérée comme le nombre k d'occurrences d'un évènement E au bout de n expériences successives et identiques, la probabilité d'un succès étant égale à p. La variable Y est donc la fréquence K/n de E. Comme $E(nY) = np$, $E(Y) = p$ et le théorème de Bernoulli s'énonce ainsi :

"la fréquence de E tend vers sa probabilité.

Démonstration. Si $Y = \frac{1}{n}(X_1 + \cdots + X_n)$, comme $E(X_i) = m$ pour tout i et $\sigma(X_i) = \sigma$ pour tout i, alors

$$E(Y) = m, \quad \sigma(Y) = \frac{\sigma}{n}.$$

L'inégalité de Chebycheff implique

$$P(|Y - m| > \frac{t\sigma}{n}) \leq \frac{1}{t^2}.$$

Posons $\varepsilon = \frac{t\sigma}{n}$. Alors

$$P(|Y - m| > \varepsilon) \leq \frac{\sigma^2}{n^2\varepsilon^2}$$

et deuxième membre tend vers 0 quand n tend vers $+\infty$.

4.3. **Théorème central limite.** Ce résultat affirme que toute somme de variables aléatoires indépendantes ayant même moyenne et même écart type tend vers une variable aléatoire qui suit une loi normale.

Théorème 6. Théorème central limite. *Soient $X_1, \cdots, X_n$ des variables aléatoires indépendantes ayant toutes la même loi de probabilité, la même espérance m et le même écart type σ. Posons*

$$Z = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n).$$

Alors la loi de probabilité de Z tend vers la loi normale $\mathcal{N}(m, \sigma)$.

EXERCICES

Exercice 1. Une urne contient 3 boules blanches et 4 boules noires. On tire deux boules de l'urne. Soit X_1 la variable aléatoire égale à 1 si la première boule est blanche, à 0 sinon. Soit X_2 la variable aléatoire égale à 1 si la deuxième boule est blanche, à 0 sinon. On suppose que les tirages se font avec remise.

(1) Faire le tableau de la loi conjointe et déterminer les lois marginales.
(2) En comparant avec le tirage sans remise, en déduire que l'on peut avoir les mêmes lois marginales mais les lois conjointes différentes.
(3) Déterminer les lois conditionnelles.

Exercice 2. Soit $\mathbf{X} = (X_1, X_2)$ un couple de variables aléatoires. Supposons X_1 et X_2 indépendantes. Montrer que la donnée des lois marginales permet de retrouver la loi conjointe.

Exercice 3. Soit $\mathbf{X} = (X_1, X_2)$ un couple de variables aléatoires. Supposons X_1 et X_2 soient des variables de Bernoulli. Montrer qu'elles sont indépendantes si et seulement si les évènements $\{X_1 = 1\}$ et $\{X_2 = 1\}$ sont indépendants.

Exercice 4. Une urne contient r boules dont r_1 sont blanches et r_2 sont noires. On effectue n tirages successifs en replaçant à chaque fois la boule tirée dans l'urne. A l'issue des n tirages, on note X le nombre de boules blanches et Y le nombre de boules noires. Déterminer la loi conjointe du couple (X, Y) et les lois de X et Y .

Exercice 5. Soit $\mathbf{X} = (X_1, X_2)$ un couple de variables aléatoires et soit $C_{\mathbf{X}}$ sa matrice de corrélation. Calculer ses valeurs propres. On rappelle que les valeurs propres d'une matrice A sont les racines λ du polynôme caractéristique

$$det(A - \lambda Id)$$

où Id désigne la matrice identité

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Exercice 6. Soient X et Y deux variables aléatoires indépendantes sur un même espace probabilité. On suppose qu'elles suivent des lois binomiales de paramètres respectifs (m, p) et (n, p). Montrer que $X + Y$ suit une loi binomiale de paramètres $(m + n, p)$.

Exercice 6. Soient X et Y deux variables aléatoires indépendantes sur un même espace probabilité. On suppose qu'elles suivent la loi de Bernoulli de paramètre p. r. Montrer que $X + Y$ suit une loi binomiale de paramètres $(p, 2)$.

Exercice 7. Soient X et Y deux variables aléatoires sur un même espace probabilité prenant les valeurs x_1, x_2, x_3 pour X et y_1, y_2, y_3 pour Y. Soit $P_{ij} = P(X = x_i, Y = y_j)$. Pour chacune des matrices $P = (P_{ij}), i, j = 1, 2, 3$ ci-dessous, déterminer si les variables aléatoires X et Y sont indépendantes.

$$\begin{pmatrix} 1/4 & 0 & 1/12 \\ 0 & 0 & 0 \\ 1/2 & 0 & 1/6 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 3/17 & 12/12 & 2/17 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/4 & 3/32 & 1/32 \\ 2/15 & 1/20 & 1/20 \\ 1/4 & 1/10 & 1/24 \end{pmatrix}$$

Exercice 8. A-t-on plus (ou moins) d'une chance sur deux d'obtenir au moins un 6 en lançant 4 dés ?

Exercice 9. Dans une série de 400 parties de pile ou face, quelle est la probabilité pour que pile sorte k fois avec $190 < k < 210$?

Exercice 10. Un pays compte 10 millions d'électeurs dont $5,5$ millions appartiennent à un parti A et $4,5$ millions à un parti B. On désigne 20000 électeurs par tirage au sort. Quelle est la probabilité pour qu'une majorité de ces 20000 électeurs appartiennent au parti B ?