CNAM MULHOUSE . Ramm Algebra Center

Technicien Développeur. Année 2.

Mathématiques

Cours Michel GOZE

Chapitre 4

Couple de variables aléatoires. Corrélation

Le but de ce chapitre est de présenter une étude simultanée de variables aléatoires. Le cas le plus simple est celui d'un couple de variables aléatoires que l'on va considérer comme une variable aléatoire mais à valeurs dans $\mathbb{R}^2$. Dans ce cas, on n'étudie pas simplement chacune des deux variables, mais on présente une étude globale, analogue à celle que l'on fait en analyse avec des fonctions de plusieurs variables à valeurs dans $\mathbb{R}^n$.

1. Couple de variables aléatoires

1.1. **Définition : Fonction de répartition d'un couple de variables aléatoires.** Soient X et Y deux variables aléatoires sur l'espace Ω. On s'intéresse dans ce paragraphe au vecteur aléatoire $\mathbf{X} = (X, Y)$ de dimension 2, appelé communément couple de variables aléatoires. Ainsi

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

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

Définition 1. *La fonction de répartition du vecteur $\boldsymbol{X} = (X, Y)$ la fonction de deux variables x, y donnée par*

$$F_{\boldsymbol{X}}(x, y) = P\{X \leq x, \text{ et } Y \leq y\}.$$

Exemple. On dispose d'une urne contenant quatre jetons numérotés de 1 à 4. On tire au sort successivement deux jetons sans remise. On note (X, Y) les résultats des deux tirages. Considérons le couple (X, Y). Sa fonction de répartition est donnée par

$$F(i, j) = P(X = i, Y = j), \quad i, j = 1, \cdots, 4.$$

On a

$$\begin{cases} P(X = i, Y = i) = 0 & i = 1, \cdots, 4 \\ P(X = i, Y = j) = 1/12 & 1 \leq i, j \leq 4, \ i \neq j. \end{cases}$$

On en déduit la fonction de répartition de ce couple. On peut présenter ces valeurs dans un tableau :

X/Y	1	2	3	4
1	0	1/12	1/2	1/12
2	1/12	0	1/12	1/12
3	1/12	1/12	0	1/12
4	1/12	1/12	1/12	0

1.2. **Propriétés.** La fonction de répartition $F_{\mathbf{X}}$ vérifie

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

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

Cette fonction est également croissante en chacune de ses variables.

1.3. **La loi de probabilité du couple $\mathbf{X} = (X, Y)$ 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 $\mathbf{X} = (X, Y)$. L'évènement $\{a_1 < X \leq b_1,\ a_2 < Y \leq x_2\}$ est le rectangle formé des points (x, y) de $\mathbb{R}^2$ tels que $a_1 < X \leq b_1$ et $a_2 < Y \leq b_2$.

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

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

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.

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

Définition 3. *On appelle loi marginale de X du couple $\boldsymbol{X} = (X, Y)$ la loi suivie par la seule variable aléatoire X.*

Connaissant la fonction de répartition de $\mathbf{X}$, il est facile de trouver la fonction de répartition de la variable X. On aura par exemple

$$F_X(x_1) = F(x, +\infty).$$

On définit de même la loi marginale de Y. Dans ce cas on aura

$$F_Y(y) = F(+\infty, y).$$

1.5. **Lois marginales.** Supposons que $\Omega = \{\omega_1, \cdots, \omega_n\}$. Dans ce cas, la loi du couple $\mathbf{X} = (X, Y)$ sera représentée par un tableau à double entrée. Les lignes correspondent aux valeurs $P(X(\omega_i))$ et les colonnes aux valeurs $P(Y(\omega_i))$. On notera

$$p_{i,j} = P(X = \omega_I, Y = \omega_j).$$

Le tableau se présentera ainsi

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

Exemple. 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

Posons

$$X(\omega_i) = x_i,\ Y(\omega_i) = y_i.$$

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

$$P_X(x_i) = P\{X = x_i\} = p_{i,\bullet} = p_{i,1} + p_{i,2} + \cdots + p_{i,n}$$

et

$$P_Y(y_j) = P\{Y = y_j\} = p_{\bullet,j} = p_{1,j} + p_{2,y} + \cdots + p_{n,j}$$

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

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.

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

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

$$P_{Y=y_j}(X = x_i) = P(X = x_i, Y = y_j) = \frac{P_{\boldsymbol{X}}(x_i, y_j)}{P(Y = y_j)}.$$

De même, pout tout $x_i \in X(\Omega)$ *tel que* $P(X = x_i) \neq 0$, *la loi* $P_{X=x_i}$ *de* Y *sachant* $\{X = x_i\}$ *est donnée par*

$$P_{X=x_i}(Y = y_j) = P(X = x_i, Y = y_j) = \frac{P_{\boldsymbol{X}}(x_i, y_j)}{P(X = x_i)}.$$

Chacune de ces applications , $P_{X=x_i}$ et $P_{Y=y_j}$ sont des probabilités sur Ω. Déterminons ces lois dans l'exemple précédent :

$X(\omega_i)$	0	1
$P_{Y=0}$	1/2	1/2
$P_{Y=1}$	2/3	1/3

,

$Y(\omega_i)$	0	1
$P_{X=0}$	1/2	1/2
$P_{X=1}$	2/3	1/3

2. Corrélation

2.1. **Variance de la loi du couple $\mathbf{X} = (X, Y)$.** Rappelons que l'espérance de la variable X est

$$E(X) = x_1 P(X = x_1) + x_2 P(X = x_2) + \cdots + x_n P(X = x_n).$$

Nous avons vu, qu'étant donné un couple $\mathbf{X} = (X, Y)$ de variables aléatoires la loi de X et Y correspondait aux lois marginales. On suppose donc donnée la loi du couple $\mathbf{X} = (X, Y)$, c'est-à-dire

$$p_{i,j} = P(X = x_i, Y = y_j).$$

L'espérance de X sera donc donnée par

$$E(X) = x_1 p_{1,\bullet} + x_2 p_{2,\bullet} + \cdots + x_n p_{n,\bullet}$$

et celle de Y par

$$E(Y) = y_1 p_{\bullet,1} + y_2 p_{\bullet,2} + \cdots + y_n p_{\bullet,n}.$$

Définition 6. *On appelle variance de la loi du couple* $\boldsymbol{X} = (X, Y)$, *la donnée des variance* $V(X)$ *et* $V(Y)$ *avec*

$$V(X) = (x_1 - E(X))^2 p_{1,\bullet} + (x_2 - E(X))^2 p_{2,\bullet} + \cdots + (x_n - E(X))^2 p_{n,\bullet}$$

et

$$V(Y) = (y_1 - E(Y))^2 p_{\bullet,1} + (y_2 - E(Y))^2 p_{\bullet,2} + \cdots + (y_n - E(Y))^2 p_{\bullet,n}.$$

2.2. **Corrélation.** Commençons par définir la covariance du couple $\mathbf{X} = (X, Y)$.

Définition 7. *On appelle covariance de la loi du couple* $\boldsymbol{X} = (X, Y)$*,*

$$Cov(X, Y) = \sum_{i,j} x_i y_j p_{i,j}.$$

Remarquons que la covariance du couple (X, X) est égale à $V(X)^2$.

2.3. **Coefficient de corrélation.** Introduisons enfin le coefficient de corrélation du couple $\mathbf{X} = (X, Y)$.

Définition 8. *Le coefficient de corrélation du couple* $\boldsymbol{X} = (X, Y)$ *est le scalaire*

$$r = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}}.$$

Ce coefficient est un réel compris entre -1 et 1. Lorsque $r = 1$ on a

$$Cov(X, Y) = \sqrt{V(X)V(Y)}.$$

On montrera en exercice que dans ce cas les variables X et Y sont reliées par une relation linéaire.

2.4. **Variables aléatoires décorrélées.** Le coefficient de corrélation est surtout utilisé pour établir un rapport entre deux variables prenant un nombre assez conséquent de valeurs. Il faudra toutefois se méfier de son interprétation, surtout lorsqu'il est proche de 1. Voir exercice

Toutefois

Définition 9. *Soit* r *le coefficient de corrélation du couple* $\boldsymbol{X} = (X, Y)$*. On dira que les variables aléatoires sont décorrélées si*

$$r = 0.$$

3. Variables aléatoires indépendantes

3.1. **Définition.** Soit $\mathbf{X} = (X, Y)$ un couple de variables aléatoires.

Définition 10. *On dit que les variables* X *et* X *sont indépendantes si la fonction de répartition du couple vérifie*

$$F_{\boldsymbol{X}}(x, y) = F_X(x) F_Y(y)$$

pour tout $x, y \in \mathbb{R}$*.*

Comme les variables aléatoires X et Y sont supposées finies, elles sont indépendantes si et seulement si

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

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.

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

$$P(X < Y + 10) + P(Y < X + 10) = 2P(X + 10 < Y).$$

On en déduit

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

3.2. **Indépendance et corrélation.** Nous avons donc deux notions, la corrélation et l'indépendance, permettant de comparer deux variables aléatoires. Il ne faut surtout pas confondre ces deux notions. Toutefois, on a

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

$$r(X, Y) = 0.$$

La réciproque est fausse en général.

3.3. **Application : La variance de la somme de deux variables aléatoires.** Nous avons énoncé ce résultat dans le chapitre précédent, mais sans aucun commentaire. Nous pouvons donc maintenant apporter des précisions.

Théorème 1. Egalité de Bienaymé *Si les variables aléatoires X et Y sont indépendantes, alors la variance de la somme et la somme des variances :*

$$V(X + Y) = V(X) + V(Y).$$

Démonstration. En effet, en utilisant les mêmes notations que ci-dessus

$$V(X) = (x_1 - E(X))^2 p_{1,\bullet} + (x_2 - E(X))^2 p_{2,\bullet} + \cdots + (x_n - E(X))^2 p_{n,\bullet}$$
$$V(Y) = (y_1 - E(Y))^2 p_{\bullet,1} + (y_2 - E(Y))^2 p_{\bullet,2} + \cdots + (y_n - E(Y))^2 p_{\bullet,n}.$$

On a aussi

$$V(X+Y) = (x_1+y_1-E(X+Y))^2 p_{1,1}+(x_2+y_2-E(X+Y))^2 p_{2,2}+\cdots+(x_n+y_n-E(X+Y))^2 p_{n,n}.$$

Comme les variables aléatoires sont indépendantes,

$$p_{i,j} = p_{i,\bullet} p_{\bullet,j}.$$

On a aussi
$$E(X+Y) = (x_1+y_1)p_{1,1} + \cdots + (x_n+y_n)p_{n,n}.$$
En reportant ceci dans l'identité précédente, on aboutit à la relation voulue.