Variable aléatoire a plusieurs dimensions

1- Définition

Soit $(\Omega, \Lambda, \mathrm{P})$ un espace de probabilité et $X = (X_1, X_1, \dots \dots X_n)$ une application de Ω dans $\mathbb{R}^n$, qui à tout élément w de Ω fait correspondre une suite $X(w) = (X_1(w), X_2(w), \dots \dots X_n(w))$. On dit que X est un vecteur aléatoire si pour tout $i = 1,2, \dots, n$, on dit couple de variables aléatoire, si $n = 2$, on dit couple de variable aléatoires ou variable aléatoire à deux dimensions.

2- variable aléatoire à deux dimensions cas discret

on appelle fonction de répartition d'une variable aléatoire à deux dimensions (X, Y) la fonction définie par :

$$\boldsymbol{F(x,y) = P[(X \leq X)(Y \leq y)]}$$

on suppose que la variable aléatoire X prend l'une quelconque des m valeur $x_1, x_2, \dots \dots \dots . x_n$ et que la variable aléatoire Y prend quelconque des n valeurs $y_1, y_2, \dots \dots \dots . y_n$. La loi de répartition d'une variable aléatoire à deux dimensions (X, Y) est définie par :

$\boldsymbol{P(X = x_i, Y = y_i) = P_{ij} \; i = 1,2, \dots . . m. \; j = 1,2, \dots . . n.}$

avec $\boldsymbol{P_{ij} \geq 0}$ et $\boldsymbol{\sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} = 1}$

Cette loi de répartition peut être donnée par le tableau suivant :

	$\boldsymbol{y_1}$	$\boldsymbol{y_2}$	**…**	$\boldsymbol{y_i}$	**…**	$\boldsymbol{y_n}$	**Total**
$\boldsymbol{x_1}$	P_{11}	P_{12}	…	P_{1j}	…	P_{1n}	P_1
$\boldsymbol{x_2}$	P_{21}	P_{22}	…	P_{2j}	…	P_{2n}	P_2
			…		…		
$\boldsymbol{x_i}$	P_{i1}	P_{i2}	…	P_{ij}	…	P_{in}	P_i
			…		…		
$\boldsymbol{x_n}$	P_{mj}	P_{m2}	…	P_{mj}	…	P_{mn}	P_m
Total	P_1	P_2	…	P_j	…	P_n	P_1

La probabilité que $\boldsymbol{X = x_i; \; i = 1,2, \dots . . m}$ est donnée par $\boldsymbol{P(X = x_i) = \sum_{j=1}^{n} P_{ij} = P_i}$.

La probabilité que $\boldsymbol{Y = y_j; \; i = 1,2, \dots . . n}$ est donnée par $\boldsymbol{P(Y = y_i) = \sum_{j=1}^{n} P_{ij} = P_j}$.

Les $\boldsymbol{P_i}$ **et** $\boldsymbol{P_j}$ **;** $\boldsymbol{i = 1,2, \dots . . m. \; j = 1,2, \dots . . n.}$ Constituent les lois de probabilité pour variables aléatoires X et Y.Elles sont dites **lois marginales des variables aléatoires X et Y** respectivement. On a:

$$\boldsymbol{\sum_{j=1}^{n} P_i = \sum_{j=1}^{n} P_j = \sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} = 1}$$

La loi conditionnelle de X si $Y = y_i$ est définie par les probabilités.

$$\boldsymbol{P_i^j = P\big(X = x_i / Y = y_j\big) = \frac{P\big(X = x_i\,,\ Y = y_j\big)}{P\big(\ Y = y_j\big)} = \frac{P_{ij}}{P_j}}$$

$$P_j \neq 0$$

La loi conditionnelle de Y si $X = x_i$ est définie par les probabilités.

$$\boldsymbol{P_j^i = P\big(Y = y_j / X = x_i\big) = \frac{P\big(\ Y = y_j, X = x_i\ \big)}{P(\ X = x_i\ )} = \frac{P_{ij}}{P_i}}$$

$$P_i \neq 0$$

On a

$$\boldsymbol{\sum_{j=1}^{m} \frac{P_{ij}}{P_j} = \sum_{j=1}^{n} \frac{P_{ij}}{P_i} = 1}$$

Les deux variables aléatoires **X et Y sont dites indépendantes** si et seulement si; pour tout couple (i, j) on a:

$P_{ij} = P_i\, P_j$; $i = 1,2, \ldots \ldots m.\ j = 1,2, \ldots \ldots n.$

L'espérance mathématique

$$E(XY) = \sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} x_i \quad y_j$$

En cas **d'indépendance des deux variables aléatoires X et Y**, c'est à dire si $P_{ij} = P_i P_j$ on obtient

$$\boldsymbol{E(XY) = \sum_{i=1}^{m} P_i\ x_i \sum_{j=1}^{n} P_j\ y_j = E(X)E(Y)}$$

La réciproque n'est pas toujours vraie; on peut avoir l'égalité : $\boldsymbol{E(XY) = E(X)E(Y)}$ sans que les variables aléatoires X et Y soient indépendantes.

3- Variable aléatoire à deux dimensions cas des VA absolument continues

Une variable aléatoire $\boldsymbol{Z = (X, Y)}$ est dite absolument continue s'il existe une application $\boldsymbol{f(x, y)}$; $\boldsymbol{f(x, y)}$ est dite densité de probabilité du couple $\boldsymbol{(X, Y)}$ vérifiant les propriétés :

a- $f(x, y) \geq 0$; $\forall (x, y) \in \mathbb{R}^2$

b- $\iint_{-\infty}^{+\infty} \boldsymbol{f(x, y) dxdy = 1}$

La probabilité que le point (x, y) se trouve dans le domaine Δ est

$$\boldsymbol{P}[(\boldsymbol{X}, \boldsymbol{Y}) \in \boldsymbol{\Delta}] = \iint_{\boldsymbol{\Delta}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{dxdy}$$

En particulier si $\Delta = [\mathrm{a}, \mathrm{b}] \times [\mathrm{c}, \mathrm{d}]$

$$\boldsymbol{P}[(\boldsymbol{a} \leq \boldsymbol{X} \leq \boldsymbol{b})(\boldsymbol{c} \leq \boldsymbol{X} \leq \boldsymbol{d})] = \int_{\boldsymbol{a}}^{\boldsymbol{b}} \int_{\boldsymbol{c}}^{\boldsymbol{d}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{dxdy}$$

La fonction de répartition du couple de variables aléatoires (X, Y) est définie par :

$$\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{P}[\boldsymbol{X} \leq \boldsymbol{x}, \boldsymbol{Y} \leq \boldsymbol{y}] = \int_{-\infty}^{\boldsymbol{x}} \int_{-\infty}^{\boldsymbol{y}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{dxdy}$$

on a:

$$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{\partial^2 F}(\boldsymbol{x}, \boldsymbol{y})}{\boldsymbol{\partial x \partial y}}$$

ou $f(x, y)$ est la densité de probabilité et $F(x, y)$ est **la fonction de répartition** du couple (X, Y).

Les fonctions :

$$\boldsymbol{F_1}(\boldsymbol{x}) = \boldsymbol{P}[\boldsymbol{X} \leq \boldsymbol{x}] = \int_{-\infty}^{\boldsymbol{x}} \int_{-\infty}^{+\infty} \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{v}) \boldsymbol{dvdu}$$

$$\boldsymbol{F_2}(\boldsymbol{y}) = \boldsymbol{P}[\boldsymbol{Y} \leq \boldsymbol{y}] = \int_{-\infty}^{\boldsymbol{y}} \int_{-\infty}^{+\infty} \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{v}) \boldsymbol{dvdu}$$

sont dites fonctions de répartition marginales des variables aléatoires X et Y.

Les fonctions

$$\boldsymbol{f_1}(\boldsymbol{x}) = \int_{-\infty}^{+\infty} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{v}) \boldsymbol{dv}$$

$$\boldsymbol{f_2}(\boldsymbol{y}) = \int_{-\infty}^{+\infty} \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{y}) \boldsymbol{du}$$

sont dites fonctions **densités de probabilité marginales des variables aléatoires X et Y.**

La densité conditionnelle de $\boldsymbol{x}$ **si** $\boldsymbol{Y} = \boldsymbol{y_0}$ est définie par

$$f(x/y_0) = \frac{f(x,y)}{f_2(y_0)}$$

La densité conditionnelle de $\boldsymbol{y}$ **si** $\boldsymbol{X = x_0}$ est définie par

$$f(y/x_0) = \frac{f(x,y)}{f_1(x_0)}$$

Les deux variables aléatoires X et Y sont dites indépendantes si et seulement si :

$$P[X \leq x, Y \leq y] = P(X \leq x), P(Y \leq y)$$

Ou

$$F(x,y) = F_1(x)F_2(x)$$

ou

$$f(x,y) = f_1(x)f_2(x)$$

L'espérance mathématique

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} uvf(v,v)dudv$$

En cas d'indépendance des deux variables aléatoires X et Y, on obtient

$$E(XY) = E(X)E(Y)$$

4- Covariance, coefficient de corrélation de X et Y .

$E(X)$ et $E(Y)$ existe on a:

$\boldsymbol{E(aX + bY) = aE(X) + bE(Y)}$ ou $a\ et\ b\ \in \mathbb{R}^2$

covariance des variables aléatoires X et Y :

$$cov(X,Y) = E\left[\left(X - E(X)\left(Y - E(Y)\right)\right)\right] = E(XY) - E(X)E(Y)$$

Les variables aléatoires X et Y sont indépendantes

$$E(XY) = E(X)E(Y)$$

$$\text{et } cov(X,Y) = 0$$

Le nombre réel r est défini par $\boldsymbol{r} = \frac{\boldsymbol{cov(X,Y)}}{\boldsymbol{\sigma_x \sigma_y}}$

ou $\boldsymbol{\sigma_x^2 = V(X)}$ et $\boldsymbol{\sigma_y^2 = V(Y)}$ est dit coefficient de corrélation de X et Y.

sI X et Y sont indépendantes on a $r = 0$. En général $-1 \leq r \leq +1$.

$\boldsymbol{V(aX + bY) = a^2V(X) + 2abcov(X,Y) + \ b^2V(Y)}$ ou $a \ et \ b \ \in \mathbb{R}^2$

Si X et Y sont indépendantes on a : $V(X + Y) = V(X) + V(Y)$

5- Changement de variables

si $f(x, y)$ est la densité de probabilité du couple de variables aléatoires (X, Y) et $U = l_1(X, Y)$, $V = l_2(X, Y)$ sont telles que chaque couple de valeurs (X, Y) correspond un et un seul couple (U, V) et inversement: soient $X = \psi_1(U, V)$ et $Y = \psi_2(U, V)$, la fonction densité de probabilité $h(u, v)$ des variables aléatoires U et V est définie par :

$$h(u, v) = f(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| = f[\psi_1(u, v), \psi_2(u, v)]|J|$$

Ou

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

est ***la jacobien*** ou déterminant de **jacobi**.

<u>Exercice</u>

L'un des résultats d'une enquête sociologique sur la violence familiale peut s'exprimer en considérant un couple (X, Y) de variable aléatoires discrètes, ou X représente la violence subie par un individu durant son enfance, mesurée par la même échelle 0-1-2, et Y la violence qu'il exerce dans sa famille, mesurée par la même échelle 0-1-2. Les probabilités « conjointes » estimées sont données par le tableau ci-dessous.

	Y=0	Y=1	Y=2
X=0	0.86	0.04	0.00
X=1	0.03	0.02	0.02
X=2	0.01	0.01	0.01

- Calculer le coefficient de corrélation de X et Y.

<u>Solution A</u>

Le tableau suivant donne l'ensemble des probabilités, incluant celles des lois marginales :

	Y=0	Y=1	Y=2	Total
X=0	0.86	0.04	0.00	0.90
X=1	0.03	0.02	0.02	0.07
X=2	0.01	0.01	0.01	0.03
Total	0.90	0.07	0.03	1

On calcule alors, avec les formules habituelles

Espérance	Variance	L'écart-type
$E(X) = \sum_{i=1}^{m} P_i x_i$	$V(X) = \sum_{i=1}^{m} P_i \ x_i^2$	$\sigma(X) = \sqrt{V(X)}$
$E(Y) = \sum_{i=1}^{n} P_j y_j$	$V(Y) = \sum_{i=1}^{n} P_j y_j^2$	$\sigma(Y) = \sqrt{V(Y)}$

$E(X) = 0.13,\ V(X) = 0.1731\ ;\ \delta(X) = 0.416$

Puis

$E(Y) = 0.13,\ V(Y) = 0.1731\ ;\ \delta(Y) = 0.416$

La covariance se calcule par :

$$E(XY) = \sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} x_i \ \ y_j$$

$$\boldsymbol{cov(X,Y) = E\left[\left(X - E(X)(Y - E(Y))\right)\right] = E(XY) - E(X)E(Y) = 0.1031}$$

Et on obtient enfin le coefficient de corrélation

$$\rho = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} = 0.596$$

<u>Exercice B</u>

Un couple de variables aléatoires (X, Y) suit une loi de répartition de densité :

$$f(x) = \begin{cases} kxy e^{-\left(x^2+y^2\right)} & Si\ x > 0 \quad et \quad y > 0 \\ 0 & ailleurs \end{cases}$$

1° Calculer le coefficient k.
2° Trouver les lois marginales de X et Y.

3° Calculer les densités conditionnelles de X si Y=y et de Y si X=x.
4° Calculer E(X), E(Y), V(X) et V(Y).
5° Trouver la matrice de covariance de X et Y.

solution

La fonction $f(x,y)$ est une densité de probabilité si $f(x,y) \geq 0$ et $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$

D'une part $k \geq 0$ et d'autre part $k\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy e^{-x^2} e^{-y^2} dxdy = 1$

On a : $k\int_{0}^{+\infty} x e^{-x^2} dx \int_{-\infty}^{+\infty} y e^{-y^2} dy = 1$ ce qui donne $k = 4$

2) Les fonctions densités de probabilité marginales sont :

$$f(x)_1 = \int_{0}^{+\infty} f(x,y)dy = \int_{-\infty}^{+\infty} 4xye^{-x^2-y^2} dy = 2xe^{-x^2} \text{ si } x > 0$$

$$f(x)_1 = 0 \text{ si } x < 0$$

Par symétrie an a $f_2(y) = 2ye^{-y^2}$ si $y > 0$

$$f_2(x) = 0 \text{ si } y \leq 0$$

1) La densité conditionnelle de X si Y=y est

$$f_1(x/y) = \frac{f(x,y)}{f_2(y)} = \begin{cases} 2xe^{-x^2} & Si\, x > 0 \\ 0 & Si\, x \leq 0 \end{cases}$$

De même

$$f_2(y/x) = \frac{f(x,y)}{f_1(x)} = \begin{cases} 2ye^{-y^2} & Si\, y > 0 \\ 0 & Si\, y \leq 0 \end{cases}$$

On remarque que les variables aléatoires X et Y sont indépendantes puisque

$f_1(x/y) = f_1(x)$ et $f_2(y/x) = f_2(x)$

Une autre condition est vérifiée aussi :

$$f(x,y) = f_1(x) f_2(y)$$

4) $E(X) = \int_{0}^{+\infty} 2x^2 e^{-x^2} dx = \int_{-\infty}^{+\infty} x e^{-x^2} dx = \frac{\sqrt{\Pi}}{2}$

De même $E(Y) = \frac{\sqrt{\Pi}}{2}$

Pour le calcule de la variance, on cherche d'abord :

$$E(X^2)=\int_0^{+\infty} 2x^3 e^{-x^2}\,dx = 1 \quad \text{donc} \quad V(X)=1-\frac{\Pi}{4}$$

De même $V(X)=1-\frac{\Pi}{4}$

2) Puisque X et Y sont des variables aléatoires indépendantes on a $Cov(X,Y)=0$
Et la matrice des covariances s'écrit :

$$\begin{Vmatrix} 1-\frac{\Pi}{4} & 0 \\ 0 & 1-\frac{\Pi}{4} \end{Vmatrix}$$