Chapitre 2

Caractéristiques des vecteurs aléatoires

2.1 Espérance, variance et covariance d'un couple de variables aléatoires

2.1.1 Théorème de transfert, Espérance d'une somme

Théorème 2.1.1. Soient (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}^2$ continu de densité f et $\mathfrak{D}$ un ouvert de $\mathbb{R}^2$ contenant $X(\Omega) \times Y(\Omega)$ et $\varphi : \mathfrak{D} \to \mathbb{R}$. On suppose que $Z = \varphi(X, Y)$ est une variable aléatoire réelle à densité. Alors

$$Z \;\; admet \;\; une \;\; espérance \;\; \Longleftrightarrow \int_{\mathbb{R}^2} \mid \varphi(x, y) \mid f(x, y) dxdy \;\; existe.$$

De plus en cas d'existence :

$$\mathbb{E}(Z) = \int_{\mathbb{R}^2} \varphi(x, y) f(x, y) dxdy.$$

Théorème 2.1.2 (Cas des fonctions linéaires.)**.** Soit (X, Y) un couple de variables aléatoires continu de densité f de $(\mathbb{R}^2, \mathfrak{B}(R^2))$. On suppose que X et Y admettent une espérance. Alors $Z = aX + bY + c$ admet une espérance. De plus :

$$\mathbb{E}(Z) = \mathbb{E}(aX + bY + c] = a\mathbb{E}(X) + b\mathbb{E}(Y) + c.$$

Démonstration. En posant $\varphi(x, y) = ax + by + c$ on obtient

$$\begin{aligned}
\mathbb{E}\left(\varphi(X, Y)\right) &= a \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f(x, y) dy\right] dx + b \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f(x, y) dx\right] dy + c \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y) dxdy \\
&= a \int_{-\infty}^{\infty} x f_X(x) dx + b \int_{-\infty}^{\infty} y f_Y dy + c \\
&= a\mathbb{E}(X) + b\mathbb{E}(Y) + c
\end{aligned}$$

Théorème 2.1.3. Soit (X, Y) un couple de variables aléatoires discret et $\mathfrak{D}$ une partie contenant $X(\Omega) \times Y(\Omega)$ et $\phi : \mathfrak{D} \to \mathbb{R}$. On suppose que $Z = \phi(X, Y)$ est une variable aléatoire discrète. Lorsque $X(\Omega)$ et $Y(\Omega)$ sont finis, on a :

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

Théorème 2.1.4. Soit (X, Y) un couple de variables aléatoires (continu ou discret). On suppose que X et Y admettent une espérance. Alors

$$Si \;\; X \leq Y \;\; alors \;\; \mathbb{E}(X) \leq \mathbb{E}(Y).$$

Démonstration. Posons $Z = Y - X$: Z admet une espérance. De plus $Z \geq 0$ et donc $\mathbb{E}(Z) \geq 0$. Ainsi $\mathbb{E}(Y) - \mathbb{E}(X) \geq 0$.

Théorème 2.1.5. Soit (X, Y) un couple de variables aléatoires réelles indépendantes admettant une densité sur $(\Omega, \mathcal{A}, \mathbb{P})$ et admettant chacune une espérance. Alors $X.Y$ admet une espérance et :

$$\mathbb{E}(X.Y) = \mathbb{E}(X)\mathbb{E}(Y).$$

Démonstration. Pour tout $(x, y) \in \mathbb{R}^2$, on a

$$\mid xy \mid f_{X,Y}(x, y) = \mid x \mid f_X(x) \times \mid y \mid f_Y(y),$$

avec f_X et f_Y désignant les densités de X et Y, on sait que (X, Y) admet pour densité la fonction $f_{X,Y}$ définie par : $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. X et Y admettent une espérance, les intégrales : $\int_{-\infty}^{\infty} \mid x \mid f_X(x)$ et $\int_{-\infty}^{\infty} \mid y \mid f_Y(y)$ sont convergentes. La fonction définie par : $\phi_{X,Y}(x, y) = xyf_{X,Y}(x, y)$ est donc intégrable. De plus :

$$\mathbb{E}(X.Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf_{X,Y}(x, y)dxdy = \left(\int_{-\infty}^{\infty} xf_X(x)dx \right) \left(\int_{-\infty}^{\infty} yf_Y(y)dy \right) = \mathbb{E}\left(X\right) \mathbb{E}\left(Y\right).$$

2.1.2 Covariance et coefficient de corrélation

Proposition 2.1.6. Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}^2$, admettant chacune un moment d'ordre 2. Alors $X.Y$ admet une espérance.

- Le nombre $Cov(X, Y) = \mathbb{E}\left(X.Y\right) - \mathbb{E}\left(X\right) \mathbb{E}\left(Y\right)$ est appelé covariance de X et Y.
- On dit que X et Y sont corrélées lorsque $Cov(X, Y) \neq 0$.
- $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ est appelé le coefficient de corrélation de X et de Y.

Théorème 2.1.7. . Soient X , Y X' et Y' des v.a.r définies sur un même espace probabilisé, admettant chacune un moment d'ordre 2 et soit λ un réel. Alors :

- $Cov(X, Y) = Cov(Y, X)$.
- $Cov(X + X', Y) = Cov(X, Y) + Cov(X', Y)$.

- $Cov(X, Y + Y') = Cov(X, Y) + Cov(X, Y')$, $Cov(\lambda X, Y) = Cov(X, \lambda Y) = \lambda Cov(X, Y)$.
- $Cov(X, X) = Var(X)$.
- X et Y sont indépendantes $\Rightarrow Cov(X, Y) = 0$
- $Cov(X, Y) = \mathbb{E}\left(X - \mathbb{E}\left(X\right)\right)\left(Y - \mathbb{E}\left(Y\right)\right)$.
- (inégalité de Schartz) $\mid Cov(X, Y) \mid \leq \sigma_X \sigma_Y$.
- Si X et Y sont indépendantes $\Rightarrow \rho(X, Y) = 0$.

- $-1 \leq \rho(X, Y) \leq 1$.

Proposition 2.1.8. Soit $(X_i)_{i \in \mathbb{N}}$ une famille de variables aléatoires quelconques dans L^2. On a

(i) $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$

(ii) $X_1 \perp X_2 \Rightarrow Cov(X_1, X_2) = 0$ mais $Cov(X_1, X_2) = 0 \nRightarrow X_1 \perp X_2$

(iii)

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var\left(X_i\right) + 2\sum_{i \neq j} Cov(X_i, X_j).$$

Démonstration.

(i)

$$\begin{aligned} Var(X_1 + X_2) &= \mathbb{E}\left[X_1 + X_2 - \left[\left(\mathbb{E}(X_1) + \mathbb{E}(X_2)\right)\right]^2\right. \\ &= \mathbb{E}\left[(X_1 - \mathbb{E}X_1) + (X_2 - \mathbb{E}X_2)\right]^2 \\ &= \mathbb{E}\left[X_1 - \mathbb{E}X_1\right]^2 + \mathbb{E}\left[X_1 - \mathbb{E}X_1\right]^2 + 2\mathbb{E}\left[X_1 - \mathbb{E}X_1\right]\left[X_2 - \mathbb{E}X_2\right] \\ &= Var\left(X_1\right) + Var\left(X_2\right) + 2Cov\left(X_1, X_2\right). \end{aligned}$$

(ii) Si $X_1 \perp X_2$, on a dans le cas continu $f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$ et l'on a donc

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_X(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 \left[\int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2\right] = \mathbb{E}(X_1)\mathbb{E}(X_2),$$

ce qui donne

$$Cov(X_1, X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = 0.$$

Exemple 2.1.1. Soit X une v.a.d dont la loi est donnée par :

k	-2	-1	0	1	2
$\mathbb{P}(X_1 = k)$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$

1. On note $X_2 = X_1^2$. Déterminer les lois de X_1 et X_2 ainsi que celle du couple (X_1, X_2).
2. X_1 et X_2 sont-elles indépendantes ?
3. Calculer $Cov(X_1, X_2)$ et faire une remarque sur ce résultat.

Solution.

TABLE 2.1 – loi conjointe du couple aléatoire (X_1, X_2) .

X_1/X_2	0	1	4
-2	0	$\frac{1}{6}$	$\frac{1}{6}$
-1	0	$\frac{1}{6}$	0
0	$\frac{1}{6}$	0	0
1	0	$\frac{1}{4}$	0
2	0	0	$\frac{1}{6}$

TABLE 2.2 – loi de X_2

k	0	1	4
$\mathbb{P}(X_2 = k)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

TABLE 2.3 – loi de $Z = X_1X_2$

k	-8	-1	0	1	8
$\mathbb{P}(Z = k)$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$

$\mathbb{E}[XY] = 0$, *on en déduit que :*

$$Cov(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = 0.$$

Ainsi les variables X_1 et X_2 ne sont pas indépendantes mais pourtant leur covariance est nulle. Elles sont seulement non-corrélées.

La loi conjointe du couple aléatoire (X_1, X_2) est donnée par le table 2.4 :

On a $\mathbb{P}([X_1 = 1] \cap [X_2 = 0]) = 0$ et $\mathbb{P}(X_1 = 1)\,\mathbb{P}(X_2 = 0) \neq 0$, donc les variables X_1 et X_2 ne sont pas indépendantes.

$\mathbb{E}(X_1) = 0$ et $\mathbb{E}(X_2) = \frac{11}{6}$. Pour le calcul de la covariance, on a besoin de la loi de $Z = X_1X_2$, voir tableau 2.3.

$\mathbb{E}[XY] = 0$, *on en déduit que :*

$$Cov(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = 0.$$

Ainsi les variables X_1 et X_2 ne sont pas indépendantes mais pourtant leur covariance est nulle. Elles sont seulement non-corrélées.

TABLE 2.4 – loi conjointe du couple aléatoire (X_1, X_2) .

$X_1 \backslash X_2$	0	1	4
-2	0	$\frac{1}{6}$	$\frac{1}{6}$
-1	0	$\frac{1}{6}$	0
0	$\frac{1}{6}$	0	0
1	0	$\frac{1}{4}$	0
2	0	0	$\frac{1}{6}$

TABLE 2.5 – loi de X_2

k	0	1	4
$\mathbb{P}(X_2 = k)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

2.2 Espérance, variance et covariance d'un vecteur aléatoire

Définition 2.1. Soit X un vecteur aléatoire de composantes $X_1, X_2, ..., X_n$. Le vecteur X est dit intégrable si chaque composante X_i est intégrable c'est-à-dire si $\mathbb{E}(\| X \|) < 1$ avec $\| \,.\, \|$ la norme euclidienne définie sur $\mathbb{R}^n$. On appelle alors espérance mathématique de X le vecteur (colonne) de $\mathbb{R}^n$ de composantes

$$\mathbb{E}(X) = \mu_X = \begin{pmatrix} \mathbb{E}(X_1) \\ . \\ . \\ . \\ \mathbb{E}(X_n) \end{pmatrix} = \begin{pmatrix} \mu_{X_1} \\ . \\ . \\ . \\ \mu_{X_n} \end{pmatrix}.$$

Le vecteur X est de carré intégrable (admet un moment d'ordre deux fini) si chaque composante X_i est de carré intégrable ou encore si $\mathbb{E}(\| X \|^2) < \infty$.

Définition 2.2. .Soit X un vecteur aléatoire quelconque de carré intégrable. On appelle matrice de covariance de X la matrice carrée d'ordre n :

$$\Sigma_X = \begin{pmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_n} \\ \sigma_{X_2X_1} & \sigma^2_{X_2} & \cdots & \sigma_{X_2X_1} \\ . & . & \cdots & . \\ . & . & . & . \\ . & . & . & . \\ \sigma_{X_1X_n} & \sigma_{X_nX_2} & \cdots & \sigma^2_{X_n} \end{pmatrix}$$

avec

$$\sigma_{X_iX_j} = Cov[X_i, X_j] = \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] = \mathbb{E}[X_iX_j] - \mathbb{E}(X_{_i})\mathbb{E}(X_j)$$

et

$$\sigma^2_{X_i} = Var(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}^2(X_i), i, j = 1, ..., n.$$

La matrice Σ_X reprend ainsi les variances des différentes variables sur sa diagonale et les covariances entre paires de variables hors de sa diagonale.

Définition 2.3. De manière générale on peut écrire que :

$$\Sigma_X = Cov[X] = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^t] = \mathbb{E}[XX^t] - \mathbb{E}(X)\mathbb{E}(X^t)$$

Théorème 2.2.1. La matrice de covariance est une matrice symétrique semi-définie positive.

Démonstration. Par construction la matrice de covariance est symétrique. Elle s'écrit comme le produit d'un vecteur et de sa transposée (l'espérance s'applique ensuite à chaque composante et ne change donc pas la symétrie). Pour le deuxième point il suffit de remarquer que si Σ_X est la matrice de covariance du vecteur aléatoire X et si V est un vecteur constant de $\mathbb{R}^n$, alors

$$V^t \sum_X V = Var(\upsilon_1 X_1 + ... + \upsilon_d X_n) \geq 0.$$

Théorème 2.2.2. . Si X est un vecteur (colonne) aléatoire de $\mathbb{R}^n$ de vecteur moyenne μ_X et de matrice de covariance Σ_X. Alors si A est une matrice réelle $q \times p$, le vecteur aléatoire AX de $\mathbb{R}^q$ a pour vecteur moyenne $A\mu_X$ et pour matrice de covariance $A\Sigma_X A^t$.

Théorème 2.2.3. . *Soit X un vecteur aléatoire quelconque de carré intégrable, alors*

$$X_1 \perp X_2 \perp ... \perp X_n \Longrightarrow \Sigma_X = \begin{pmatrix} \sigma^2_{X_1} & 0 & ... & 0 \\ 0 & \sigma^2_{X_2} & ... & 0 \\ . & . & ... & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & ... & \sigma^2_{X_n} \end{pmatrix}$$

Définition 2.4. Soit X un vecteur aléatoire de carré intégrable. On appelle matrice de corrélation de X la matrice carrée d'ordre n donnée par :

$$R = Corr[X] = \begin{pmatrix} 1 & \rho_{X_1X_2} & ... & \rho_{X_1X_n} \\ \rho_{X_2X_1} & 1 & ... & \rho_{X_2X_n} \\ . & . & ... & . \\ . & . & . & . \\ . & . & . & . \\ \rho_{X_1X_n} & \rho_{X_nX_2} & ... & 1 \end{pmatrix}$$

avec

$$\rho_{X_iX_j} = \frac{Cov[X_i, X_j]}{\sigma_{X_1}\sigma_{X_2}}, i \neq j.$$

Propriété 2.2.4.

- La matrice R est semi-définie positive.
- R prenant toujours des valeurs égales à 1 sur sa diagonale, les valeurs hors de sa diagonales étant égales aux coefficients de corrélation entre les variables prises deux à deux.
- Lorsque les variables sont mutuellement indépendantes cette matrice est égale à la matrice identité puisque toutes les corrélations sont nulles.
- Si l'on fait appel au calcul matriciel, on vérifie aisément que les matrices R et Σ_X sont liées par la relation :

$$\Sigma_X = SRS \ \ avec \ \ S = \begin{pmatrix} \sigma_{X_1} & 0 & ... & 0 \\ 0 & \sigma_{X_2} & ... & \\ . & . & ... & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & ... & \sigma_{X_n} \end{pmatrix}$$

Proposition 2.2.5. Si $X \sim \mathcal{M}(n, p)$, alors

1. $\mathbb{E}(X) = (np_1, ..., np_k)$, la matrice de variance-covariance de X est égale à

$$\sum_X = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & ... & -np_1p_k \\ -np_1p_2 & np_2(1-p_2) & ... & -np_2p_k \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ -np_1p_k & -np_2p_k & ... & np_k(1-p_k) \end{pmatrix}.$$

2. La matrice de corrélation R est donnée par :

$$R = \begin{pmatrix} 1 & -\sqrt{\frac{p_1p_2}{(1-p_1)(1-p_2)}} & ... & -\sqrt{\frac{p_1p_k}{(1-p_1)(1-p_k)}} \\ -\sqrt{\frac{p_1p_2}{(1-p_1)(1-p_2)}} & 1 & ... & -\sqrt{\frac{p_2p_k}{(1-p_2)(1-p_k)}} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ -\sqrt{\frac{p_1p_k}{(1-p_1)(1-p_k)}} & -\sqrt{\frac{p_2p_k}{(1-p_2)(1-p_k)}} & ... & 1 \end{pmatrix}.$$

Espérance et (co)variance conditionnelle

Espérance conditionnelle d'un couple aléatoire-cas discret

Définition 2.5. Soit (X, Y) un couple de variables aléatoires. Supposons que Y est intégrable. La variable aléatoire qui prend les valeurs $\mathbb{E}[Y/X = x_i]$ avec les probabilités p_i est appelée espérance conditionnelle de Y sachant X et notée $\mathbb{E}[Y/X]$.

Remarque 2.2.1. Il est important de noter que l'espérance conditionnelle $\mathbb{E}[Y \diagup X]$ est en général une variable aléatoire et non pas une quantité déterministe. On peut l'interpréter comme la valeur moyenne prise par Y lorsque l'on connaît X.

Exemple 2.2.1. Soit $Y \sim \mathcal{P}(\alpha)$ et $Z \sim \mathcal{P}(\beta)$ deux variables aléatoires de loi de Poisson indépendantes. La loi de $X = Y + Z$ suit également une loi de Poisson de paramètre $\alpha + \beta$. On s'intéresse ici à la loi de Y sachant X. Soit $n \in N$, déterminons la loi de Y sachant $X = n$. Puisque $X = Y + Z$, il est clair que, sachant que $X = n$, Y est à valeurs dans $\{0, 1, ..., n\}$. Soit donc $k \in \{0, 1, ..., n\}$, on a

$$\mathbb{P}\left(Y = k/X = n\right) = \frac{\mathbb{P}\left(Y = k, X = n\right)}{\mathbb{P}\left(X = n\right)} = \frac{\mathbb{P}\left(Y = k, Z = n - k\right)}{\mathbb{P}\left(X = n\right)} = \frac{\mathbb{P}\left(Y = k\right)\mathbb{P}\left(Z = n - k\right)}{\mathbb{P}\left(X = n\right)}.$$

On obtient alors grâce aux fonctions de masse des lois de Poisson

$$\mathbb{P}\left(Y = k/X = n\right) = \begin{pmatrix} \text{k} \\ \text{n} \end{pmatrix} \left(\frac{\alpha}{\alpha + \beta}\right)^k \left(\frac{\beta}{\alpha + \beta}\right)^{n-k}.$$

Ainsi, sachant $X = n$, Y suit une loi binomiale $\mathcal{B}(n, \frac{\alpha}{\alpha+\beta})$.

L'espérance de Y sachant $X = n$ est l'espérance d'une loi binomiale $\mathcal{B}(n, \frac{\alpha}{\alpha+\beta})$. On a donc pour tout $n \geq 0$:

$$\mathbb{E}[Y \diagup X = n] = \frac{\alpha n}{\alpha + \beta}.$$

Puisque ceci est vrai pour tout n, l'espérance conditionnelle de Y sachant X est :

$$\mathbb{E}[Y \diagup X] = \frac{\alpha X}{\alpha + \beta},$$

qui est bien une fonction de X et donc une variable aléatoire. On peut donc calculer l'espérance de l'espérance conditionnelle.

Théorème 2.2.6. ***E'espérance totale.*** Si Y est intégrable, alors la variable $\mathbb{E}[Y/X]$ est également intégrable et on a :

$$\mathbb{E}[\mathbb{E}[Y/X]] = \mathbb{E}[Y].$$

Exemple 2.2.2. Toujours sur l'exemple précédent, on a

$$\mathbb{E}[\mathbb{E}[Y \diagup X]] = \frac{\alpha}{\alpha + \beta}\mathbb{E}[X],$$

L'espérance d'une loi de Poisson de paramètre $\alpha+\beta$ est $\alpha+\beta$, on retrouve donc bien

$$\mathbb{E}[\mathbb{E}[Y/X]] = \frac{\alpha}{\alpha+\beta}\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}(\alpha+\beta) = \alpha = \mathbb{E}[Y].$$

On vient de voir que dans le cas général, l'espérance conditionnelle $\mathbb{E}[Y/X]$ est une variable aléatoire. Il existe cependant un cas particulier où ce n'est pas le cas : lorsque X et Y sont indépendantes.

Espérance conditionnelle d'un couple aléatoire-cas absolument continu

Définition 2.6. La variable aléatoire qui prend les valeurs $\mathbb{E}[Y/X=x]$ avec la densité f_X est appelée espérance conditionnelle de Y sachant X. On la note $\mathbb{E}[Y/X]$. Pour x fixé, l'espérance conditionnelle de Y sachant $X=x$ est

$$\mathbb{E}[Y/X=x] = \int y f_{Y/X=x}(y)dy.$$

La fonction $\varphi : x \to \mathbb{E}[Y/X=x]$ est une fonction réelle d'une variable réelle. $\varphi(X)$ est donc une variable aléatoire.

Théorème 2.2.7. ***d'espérance totale.*** *Si Y est intégrable, alors la variable aléatoire $\mathbb{E}[Y/X]$ l'est aussi et on a :*

$$\mathbb{E}[\mathbb{E}[Y/X]] = \mathbb{E}[Y].$$

Exemple 2.2.3. On considère $(X,Y)^t$ un couple de variables aléatoires de loi uniforme sur le triangle
$\mathcal{T} = \{(x,y) : 0 < y < x < 1\}$. Le vecteur $(X,Y)^t$ admet donc comme densité

$$f_{X,Y}(x,y) = 21_{\mathcal{T}}(x,y).$$

Les densités marginales sont données par :

$$f_X(x) = 2x1_{[0,1]}(x), f_Y(y) = 2(1-y)1_{[0,1]}(y).$$

sachant que $X=x$. Nous voyons sur la Figure 2.1 que Y prend ses valeurs sur le segment $]0,x[$, de plus la loi du couple (X,Y) étant uniforme, tous les segments de même longueur incluent dans $]0,x[$ ont la même probabilité.

$$f_{Y/X=x}(y) = \frac{1}{x}1_{]0,x[}(y) = \frac{21_{0<y<x}}{2x} = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

$$\mathbb{E}[Y/X=x] = \int_{\mathbb{R}} y f_{Y/X=x}(y)dy = \frac{x}{2},$$

d'où, $\mathbb{E}[Y/X] = \frac{X}{2}$. On retrouve bien l'espérance de Y par le théorème précédent

$$\mathbb{E}[\mathbb{E}[Y/X]] = \frac{1}{2}\mathbb{E}[X] = \frac{1}{3} = \mathbb{E}[Y].$$

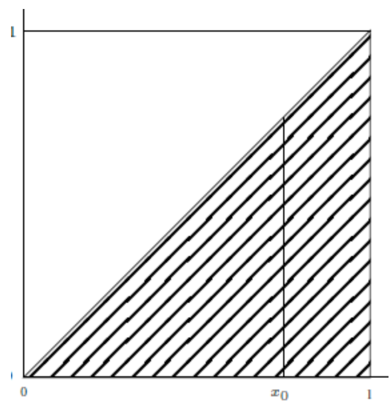

FIGURE 2.1 – Densité conditionnelle de $Y/X = x$.

La notion d'espérance et variance conitionnelle se généralise aisément au cas du vecteur aléatoire. On peut y rajouter la notion de covariance et de corrélation conditionnelle. Si X_a et X_b forment une partition d'un vecteur aléatoire X, l'espérance des espérances des variables $X_i/X_b = x_b$ s'obtient facilement par

Définition 2.7. Pour un vecteur aléatoire $X = (X_a, X_b)^t$ avec $X_a = (X_1, ..., X_{n_a})^t$, les espérances conditionnelles des variables $X_i/X_b = x_b$ et $X_i/X_a = x_a$ sont données par :

$$\mathbb{E}(X_i/x_b) = \mu_{X_i/x_b} = \begin{cases} \sum_i x_i \mathbb{P}(x_i/X_b = x_b) & \forall i, ..., n_a \quad \textit{cas discret} \\ \int_{-\infty}^{+\infty} x_i f(x_i/x_b) dx_i & \forall i, ..., n_a \quad \textit{cas continu.} \end{cases}$$

$$\mathbb{E}(X_i/x_a) = \mu_{X_i/x_a} = \begin{cases} \sum_i x_i \mathbb{P}(x_i/x_b) & \forall i, ..., n_b \quad \textit{cas discret} \\ \int_{-\infty}^{+\infty} x_i f(x_i/X_b = x_b) dx_i & \forall i, ..., n_b \quad \textit{cas continu.} \end{cases}$$

Définition 2.8. Les variances et covariances conditionnelles des variables $(X_i/X_b = x_b)_{1\leq i\leq n_a}$ et $(X_i/X_a = x_a)_{1\leq i\leq n_b}$ sont données par :

$$\sigma^2_{X_i/x_b} = \begin{cases} \left(\sum_i x_i^2 \mathbb{P}(x_i/x_b)\right) - \mu^2_{X_i/x_b} & \forall i, ..., n_a \quad \textit{cas discret} \\ \int_{-\infty}^{+\infty} x_i^2 f(x_i/x_b) dx_i - \mu^2_{X_i/x_b} & \forall i, ..., n_a \quad \textit{cas continu} \end{cases}$$

$$\sigma_{X_iX_j/x_b} = \begin{cases} \left(\sum_i \sum_j x_i x_j \mathbb{P}(x_i, x_j/x_b)\right) - \mu_{X_i/x_b}\mu_{X_j/x_b} & \forall i, ..., n_a \quad \textit{cas discret} \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_i x_j f(x_i, x_j/x_b) dx_i dx_j - \mu_{X_i/x_b}\mu_{X_j/x_b} & \forall i, ..., n_a \quad \textit{cas continu} \end{cases}$$

Définition 2.9. La matrice de variance covariances conditionnelle et le vecteur espérance conditionnel sont donnés par :

$$\mathbb{E}(X_a/X_b = x_b) = \mu_{X_a/x_b} = \begin{pmatrix} \mu_{X_1/x_b} \\ . \\ . \\ . \\ \mu_{X_{n_a}/x_b} \end{pmatrix}, \Sigma_{X_a/x_b} = \begin{pmatrix} \sigma^2_{X_1/x_b} & \sigma_{X_1X_2/x_b} & \cdots & \sigma_{X_1X_{n_a}/x_b} \\ \sigma_{X_2X_1/x_b} & \sigma^2_{X_2/x_b} & \cdots & \sigma_{X_2X_{n_a}/x_b} \\ . & . & \cdots & . \\ . & . & . & . \\ . & . & . & . \\ \sigma_{X_{n_a}X_1/x_b} & \sigma_{X_{n_a}X_2/x_b} & \cdots & \sigma^2_{X_{n_a}/x_b} \end{pmatrix}.$$

On peut également définir la matrice de corrélation conditionnelle :

$$R_{X_a/x_b} = \begin{pmatrix} 1 & \rho_{X_1X_2/x_b} & \cdots & \rho_{X_1X_{n_a}/x_b} \\ \rho_{X_2X_1/x_b} & 1 & \cdots & \rho_{X_2X_{n_a}/x_b} \\ . & . & \cdots & . \\ . & . & . & . \\ . & . & . & . \\ \rho_{X_{n_a}X_1/x_b} & \rho_{X_{n_a}X_2/x_b} & \cdots & 1 \end{pmatrix}.$$

Les coefficients de corrélation conditionnels sont :

$$\rho_{X_iX_j/x_b} = \frac{\sigma_{X_iX_j/x_b}}{\sigma_{X_i/x_b}\sigma_{X_j/x_b}}.$$

2.3 Recherche de densité , changement de vecteur aléatoire-fonction d'un vecteur aléatoire

Position du problème

Un problème important est le suivant : Soit X une variable aléatoire réelle, admettant la densité f_X. Soit h une fonction mesurable, de sorte que $Y = h(X)$ soit aussi une variable aléatoire. Est-ce que Y admet une densité, et si oui, comment la calculer ?
Il convient d'abord de remarquer que cette densité n'existe pas toujours. Si par exemple $h(x) = a$ pour tout x, la loi de Y est la masse de Dirac en a, qui n'a pas de densité.
Pour résoudre ce problème, l'idée consiste à essayer de mettre $\mathbb{E}(g(Y)) = \mathbb{E}(g \circ h(X))$ sous la forme $\int g(y)f_Y(y)dy$ pour une fonction convenable f_Y , et une classe de fonctions g suffisamment grande. La fonction f_Y sera alors la densité recherchée.

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} (g \circ h(x))f_X(x)dx. \tag{2.1}$$

Nous souhaitons faire le changement de variable $y = h(x)$ dans cette intégrale. Cela nécessite que h soit dérivable et bijective "par morceaux", et il faut faire très attention aux domaines où h est croissante ou décroissante. Plutôt qu'exposer une théorie générale, donnons des exemples.

Exemple 2.3.1.

1. Soit $Y = aX + b$, où a, b sont deux constantes. Si $a = 0$, nous avons alors $Y = b$ et la loi de Y est la masse de Dirac en b.
   Si au contraire $a \neq 0$, nous faisons le changement de variable $y = ax + b$ dans (2.1), ce qui donne :
   $$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(ax+b) f_X(x) dx = \int_{-\infty}^{\infty} g(y) f_X\left(\frac{y-b}{a}\right) \frac{1}{\mid a \mid} dy.$$
   Par exemple :
   - Si X suit la loi $\mathcal{N}(\mu, \sigma^2)$, alors $\frac{X-\mu}{\sigma}$ suit la loi normale $\mathcal{N}(0,1)$.
   - Si X suit la loi $\mathcal{N}(0,1)$, alors $aX + b$ suit la loi normale $\mathcal{N}(b, a^2)$.
   - Si X suit la loi uniforme sur $[\alpha, \beta]$, alors $aX + b$ suit la loi uniforme sur $[a\alpha + b, a\beta + b]$.
2. Soit $Y = X^2$. La fonction $x \to x^2$ est décroissante sur $\mathbb{R}_-$ et croissante sur $\mathbb{R}^+$. Le changement de variable $y = x^2$ donne alors
   $$\begin{aligned} \mathbb{E}(g(Y)) &= \int_{-\infty}^{0} g(x^2) f_X(x) dx + \int_{0}^{+\infty} g(x^2) f_X(x) dx \\ &= \int_{-\infty}^{0} g(y) f_X(-\sqrt{y}) \frac{dy}{2\sqrt{y}} + \int_{0}^{+\infty} g(y) f_X(-\sqrt{y}) \frac{dy}{2\sqrt{y}}. \end{aligned}$$
   Nous avons donc :
   $$f_Y(y) = (f_X(-\sqrt{y}) + f_X(\sqrt{y})) \frac{1}{2\sqrt{y}}. \tag{2.2}$$

Cas général

Dans le cas des vecteurs aléatoires, l'idée est la même. Considérons une fonction h définie en tout point d'un ouvert $\mathfrak{D} \subset \mathbb{R}^n$ et à valeurs dans $\mathbb{R}^m$. Pour $x = (x_1, ..., x_n)^t \in \mathbb{R}^n$, $h(x)$ est un élément de $\mathbb{R}^m$. Cet élément peut s'écrire sous la forme $(h_1(x), h_2(x), ...h_m(x))^t$. On définit ainsi n fonctions $h_1, ..., h_m$, chaque fonction h_i étant définie sur $\mathfrak{D}$ et à valeurs dans $\mathbb{R}$ pour $i = 1, ..., n$. Nous poserons alors $h = (h_1, h_2, ...h_m)$. Soit le vecteur aléatoire Y, dont chacune des variables peut être exprimée comme une fonction des variables contenues dans un autre vecteur aléatoire X avec

$$\begin{cases} Y_1 = h_1(X_1, ..., X_n) = h_1(X) \\ . \\ . \\ . \\ Y_m = h_m(X_1, ..., X_n) = h_m(X). \end{cases}$$

Plusieurs cas sont à considérer :

a. $m > n$: Le vecteur Y n'admet pas de densité.

b. Cas des fonctions inversibles $m = n$: Le système d'équation $Y = h(X)$ résolu par rapport aux variables X admet une solution unique. La transformation inverse est dans ce cas univoque, avec :

$$\left\{\begin{array}{l} Y_1 = h_1(X_1, ..., X_n) = h_1(X) \\ . \\ . \\ . \\ Y_n = h_m(X_1, ..., X_n) = h_n(X) \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} X_1 = g_1(Y_1, ..., Y_n) = g_1(Y) \\ . \\ . \\ . \\ X_n = g_n(Y_1, ..., Y_n) = g_n(Y) \end{array}\right.$$

Définition 2.10. Soit une application $h = (h_1, h_2, ...h_n)^t$ définie en tout point d'un ouvert $\mathfrak{D} \in \mathbb{R}^n$ et à valeurs dans $\mathbb{R}^n$. Si h est différentiable sur $\mathfrak{D}$, les matrices jacobiennes de h en $x = (x_1, ..., x_n)^t$ sont données par :

- $J_{x \longrightarrow y}$: est la matrice de la transformation permettant de passer des variables X aux variables Y.
- $J_{y \longrightarrow x}$ est la matrice jacobienne de la transformation inverse permettant de passer des variables Y aux variables X, avec

$$J_{x \longrightarrow y} = \begin{pmatrix} \frac{\partial g_1(y)}{\partial y_1} & .... & \frac{\partial g_1(y)}{\partial y_n} \\ . & ... & . \\ . & ... & . \\ . & ... & . \\ \frac{\partial g_n(y)}{\partial y_1} & ... & \frac{\partial g_n(y)}{\partial y_n} \end{pmatrix}; J_{y \longrightarrow x} = \begin{pmatrix} \frac{\partial h_1(x)}{\partial x_1} & .... & \frac{\partial h_1(x)}{\partial x_n} \\ . & ... & . \\ . & ... & . \\ . & ... & . \\ \frac{\partial h_n(x)}{\partial x_1} & ... & \frac{\partial h_n(x)}{\partial x_n} \end{pmatrix}$$

Théorème 2.3.1. Soient un vecteur aléatoire X de dimension n, absolument continu de densité f_X et $Y = h(X)$ où h est une application bijective continûment différentiables d'un ouvert $\mathfrak{D} \in \mathbb{R}^n$ dans $\mathbb{R}^m$ avec $\mathbb{P}[X \in \mathfrak{D}]$. Soit l'application g définie sur $h(\mathfrak{D})$, à valeurs dans $\mathfrak{D}$ et réciproque de h. Si le déterminant des matrices jacobiennes $det(J_{y \to x})$ et $det(J_{y \to x})$ ne s'annullent en aucun point $x \in \mathfrak{D}$, alors le vecteur aléatoire Y est absolument continu et admet pour densité :

$$f_Y(y) = \left\{\begin{array}{l} \mid det(J_{x \to y}) \mid f_X \circ (h^{-1}(y)) = \mid det(J_{x \to y}) \mid f_X(g(y)) \\ \frac{1}{|det(J_{y \to x})|} f_X \circ (h^{-1}(y)) = \frac{1}{|det(J_{y \to x})|} f_X(g(y)). \end{array}\right. \tag{2.3}$$

Exemple 2.3.2. *[3].* Une machine est destinée à fabriquer des briques dont la longueur L, la largeur l et la hauteur h sont fixées par un réglage, de manière à obtenir $L = 190$, $l = 90$ et $h = 50$ (en millimètres). On admet que l'erreur de réglage sur chacune des ces dimesions est distribuée uniformément entre -0.5 et $+0.5$ millimètre, ces erreurs étant indépendantes entre elles. Quelle sera la distribution conjointe des surfaces pour les trois faces de la brique si l'on calcule ces surfaces en faisant le produit des dimensions correspondantes ?
Si (X_1, X_2, X_3) est le vecteur des longueur, largeur et hauteur, on peut supposer que

$X_1 \perp X_2 \perp X_3$. L'erreur de réglage étant de ± 0.5 millimètre, on a $X_1 \sim \mathcal{U}_{[189.5,190.5]}$, $X_2 \sim \mathcal{U}_{[89.5,90.5]}$ et $X_3 \sim \mathcal{U}_{[49.5,50.5]}$. Ce qui montre que

$$f_{X_1}(x_1) = \begin{cases} 1 \ \ si \ \ x_1 \in [189.5, 190.5] \\ \quad 0 \quad sinon. \end{cases}, f_{X_2}(x_2) = \begin{cases} 1 \ \ si \ \ x_1 \in [89.5, 90.5] \\ \quad 0 \quad sinon. \end{cases},$$

$$f_{X_3}(x_3) = \begin{cases} 1 \ \ si \ \ x_1 \in [49.5, 50.5] \\ \quad 0 \quad sinon. \end{cases},$$

Ce que l'on demande est la distribution conjointe de $Y = (Y_1, Y_2, Y_3)$, avec

$$Y_1 = X_1X_2, Y_2 = X_1X_3 \ \ et \ \ \ Y_3 = X_2X_3.$$

On a donc :

$$\begin{cases} Y_1 = h_1(X) = X_1X_2 \\ Y_2 = h_2(X) = X_1X_3 \\ Y_3 = h_3(X) = X_2X_3 \end{cases} \Leftrightarrow \begin{cases} X_1 = g_1(Y) = \sqrt{\frac{Y_1Y_2}{Y_3}} \\ X_2 = g_2(Y) = \sqrt{\frac{Y_1Y_3}{Y_2}} \\ X_3 = g_3(Y) = \sqrt{\frac{Y_2Y_3}{Y_1}} \end{cases}$$

La matrice jacobienne $J_{y \to x}$ est donné par

$$J_{y \to x} = \begin{pmatrix} x_2 & x_1 & 0 \\ x_3 & 0 & x_1 \\ 0 & x_3 & x_2 \end{pmatrix}.$$

Ce qui donne $\mid det(J_{y \to x}) \mid = \mid -2x_1x_2x_3 \mid$. Puisque l'on a $x_1 = \sqrt{\frac{y_1y_2}{y_3}}$, $x_2 = \sqrt{\frac{y_1y_3}{y_2}}$ *et* $x_3 = \sqrt{\frac{y_2y_3}{y_1}}$. *On a* $\mid det(J_{x \to y}) \mid = 2\sqrt{y_1y_2y_3}$. *Puisque* $X_1 \perp X_2 \perp X_3$, *on a* $f_X(x) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3) = 1$, *soit* $f_X(g(y)) = 1$. En combinant ces résultats, on obtient :

$$f_Y(y) = \frac{1}{2\sqrt{y_1y_2y_3}}$$

sur le domaine des valeurs possibles pour y et vaut 0 ailleurs.

c. On remarquera que pour que cette méthode puisse fonctionner, il faut que le nombre de variables pour X et Y soit identique, le déterminant n'étant défini que pour des matrices carrées. Ceci implique que la méthode ne permet pas telle quelle d'obtenir la distribution conjointe de fonctions d'un vecteur aléatoire si le nombre de fonctions est inférieur au nombre de variables pour X, ce qui est un cas que l'on rencontrera fréquemment.

Solution :

On commence par compléter Y, en essayant de construire une application h' de $\mathbb{R}^n$ dans $\mathbb{R}^n$ dont les m premières composantes coïncident avec les composantes de h, et pour laquelle (2.3) s'applique. Nous obtenons ainsi la densité $f_{Y'}$ de $Y' = h'(X)$. Puis nous appliquons l'extension évidente de (2.3) :

$$f_Y(y_1, ..., y_m) = \int_{-\infty}^{\infty} ... \int_{\mathbb{R}^{n-m}} f_{Y'}(y_1, ..., y_m, y_{m+1}, ..., y_n) dy_{m+1}...dy_n. \tag{2.4}$$

Exemple 2.3.3. Soit $X = (U, V)$ un vecteur aléatoire dans $\mathbb{R}^2$, avec U et V indépendantes de loi $\Gamma(\alpha, \theta)$ et $\Gamma(\beta, \theta)$. Quelle est la densité de $Y = \frac{U}{U+V}$?.
Comme la dimension de Y est plus petite que celle de X, il faut d'abord completer Y. Nous prenons par exemple $Y' = (Y, Z)$, avec $Z = U + V$, ce qui correspond à $h(u, v) = \left(\frac{u+v}{v}, u + v\right)$. Cette application est bijective sur $A =]0, \infty[\times]0, \infty[$ dans $B =]0, 1[\times]0, \infty[$ et nous avons $h^{-1}(y, z) = g(y, z) = (yz, z(1 - y))$ qui a pour jacobien z. Comme

$$f_X(u, v) = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} v^{\beta-1} e^{-\theta(u+v)} 1_A(u, v).$$

$$f_{Y'}(y, z) = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha+\beta-1} y^{\alpha-1} (1 - y)^{\beta-1} e^{-\theta z} 1_B(y, z).$$

$$\begin{aligned} f_{Y(y)} &= \int_0^{\infty} f_{Y'}(y, z) dz \\ &= \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1 - y)^{\beta-1} \int_0^{\infty} z^{\alpha+\beta-1} e^{-\theta z} 1_B(y, z) dz \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1 - y)^{\beta-1} 1_{[0,1]}(y). \end{aligned}$$

La densité de Z est donnée par :

$$\frac{\theta^{\alpha+\beta}}{\Gamma(\alpha + \beta)} z^{\alpha+\beta-1} e^{-\theta z} 1_{]0,\infty[}(z).$$

Cas particulier

On peut résourdre ce problème en ajoutant des variables fictives , de manière à obtenir des dimensions identiques, Soit $Y_a = h(X)$ est de dimension m et X est de dimension n (avec $m < n$), le plus facile est de définir un vecteur $Y' = Y_b = X_b$ où

X_b sont les $n-m$ dernières variables de X, c'est à dire :

$$\left\{\begin{array}{l} Y_1 = h_1(X_1, ..., X_n) \\ . \\ . \\ . \\ Y_m = h_m(X_1, ..., X_n) \\ Y_{m+1} = X_{m+1} \\ . \\ . \\ . \\ Y_n = X_n. \end{array}\right. \Leftrightarrow Y = \begin{pmatrix} Y_a \\ Y_b \end{pmatrix} = \begin{pmatrix} h(X) \\ X_b \end{pmatrix}.$$

Si l'on partitionne la matrice jacobienne $J_{y\longrightarrow x}$ de façon correspondante , on a

$$J_{y\longrightarrow x} = \begin{pmatrix} \frac{\partial h_1(x)}{\partial x_1} & ... & \frac{\partial h_1(x)}{\partial x_m} & \frac{\partial h_1(x)}{\partial x_{m+1}} & ... & \frac{\partial h_1(x)}{\partial x_n} \\ . & ... & . & . & ... & . \\ . & ... & . & . & ... & . \\ . & ... & . & . & ... & . \\ \frac{\partial h_m(x)}{\partial x_1} & ... & \frac{\partial h_m(x)}{\partial x_m} & \frac{\partial h_m(x)}{\partial x_{m+1}} & ... & \frac{\partial h_m(x)}{\partial x_n} \\ \frac{\partial y_{m+1}}{\partial x_1} & ... & \frac{\partial y_{m+1}}{\partial x_m} & \frac{\partial y_{m+1}}{\partial x_{m+1}} & ... & \frac{\partial y_{m+1}}{\partial x_n} \\ . & ... & . & . & ... & . \\ . & ... & . & . & ... & . \\ . & ... & . & . & ... & . \\ \frac{\partial y_n(x)}{\partial x_1} & ... & \frac{\partial y_n(x)}{\partial x_m} & \frac{\partial y_n(x)}{\partial x_{m+1}} & ... & \frac{\partial y_n(x)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial h(x)}{\partial x_{ta}} & \frac{\partial h(x)}{\partial x_b^t} \\ 0 & I \end{pmatrix}$$

où 0 est une matrice nulle de dimension $(m-n)\times m$ et I est la matrice identité de dimension $(m-n)\times(m-1)$. On peut donc se contenter de calculer le déterminant de la matrice $\frac{\partial h(x)}{\partial x_a'}$ qui est de dimension $m\times m$ plutôt que de calculer celui de la matrice jacobienne $J_{y\longrightarrow x}$, de dimension $n\times n$. Sachant que $f_{Y_a}(y_a) = \int_{-\infty}^{\infty} f_Y(y_b)$, la fonction de densité de probabilité marginale est donc donnée par :

$$f_{Y_a}(y_a) = \frac{1}{\mid det(\frac{\partial h(x)}{\partial x_a^t}) \mid_{x=g(y)}} \int_{-\infty}^{\infty} f_X(g(y))dy_b.$$

Cas des fonctions linéaires

Si les fonctions $Y = h(X) = (h_1(X), ..., h_n(X))$ sont des fonctions linéaires des variables X, on peut les écrire sous la forme :

$$\left\{\begin{array}{l} Y_1 = h_1(X) = a_{11}X_1 + a_{12}X_2 + ... + a_{1n}X_n \\ . \\ . \\ . \\ Y_n = h_n(X) = a_{n1}X_1 + a_{n2}X_2 + ... + a_{nn}X_n \end{array}\right.$$

ce que l'on note plus facilement $Y = AX$, avec :

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

où les a_{ij} contenus dans la matrice A sont des valeurs réelles.

Propriété 2.3.2. *Il est alors facile de voir que pour la matrice jacobienne, on obtient $J_{y\longrightarrow x} = A$ et $J_{x\longrightarrow y} = A^{-1}$. La transformation inverse est univoque si l'inverse de A existe, et dans ce cas*

$$Y = h(X) = AX \Leftrightarrow X = g(Y) = A^{-1}Y$$

Exemple 2.3.4. *[3].* On s'interesse à un processus dont le nombre d'occurrence sur un intervalle de temps t suit une loi de Poisson de paramètre $\mu = \lambda t$ (par exemple le nombre de clients arrivant au guichet d'une banque, le nombre de particules émises par une substance radioactive,...).
Quelle sera la distribution pour le temps séparant la première occurrence de la troisième ?
Si le processus suit une loi de Poisson de paramètre $\mu = \lambda t$, le temps séparant deux occurrence successives suit une loi exponentielle de paramètre λ. Si X_1 et X_2 désignent respectivement les temps séparant la première occurrence de la deuxiéme et la deuxième occurrence de la troisième, on a de plus $X_1 \perp X_2$. Ce que l'on demande est la loi de probabilité pour $X_1 + X_2$.
Posons $X = (X_1, X_2)^t$, Y=$(Y_1, Y_2)^t$,$Y_1 = X_1 + X_2$ ainsi que la variable fictive $Y_2 = X_2$. On a donc

$$\left\{ \begin{array}{c} Y_1 = X_1 + X_2 \\ Y_2 = X_2 \end{array} \right. \Longleftrightarrow \left\{ \begin{array}{c} X_1 = Y_1 - Y_2 \\ X_2 = Y_2 \end{array} \right.$$

c'est-à- dire :

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}; A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

soit $\mid det(A) \mid= 1$. *Puisque $X_1 \perp X_2$, on a $f_X(x) = f_{X_1}(x_1)f_{X_1}(x_1)$ avec $f_{X_i}(x_i) = \lambda e^{\lambda x_i}(i = 1, 2)$ pour la loi exponentielle. On obtient donc :*

$$\begin{aligned} f_Y(y) &= \frac{1}{\mid det(A) \mid} f_X(A^{-1}Y) = f_{X_1}(y_1 - y_2)f_{X_2}(y_2) \\ &= \lambda e^{-\lambda(y_1 - y_2)}\lambda e^{-\lambda y_2} = \lambda^2 e^{-\lambda y_1}. \end{aligned}$$

Par conséquent,

$$f_Y(y) = \left\{ \begin{array}{c} \lambda^2 e^{-\lambda y_1} \quad pour \;\; y_1 \in [0, \infty[, \, y_2 \in [0, y_1[\\ 0 \quad ailleurs. \end{array} \right.$$

La densité marginale de Y_1 est donnée par :

$$\begin{aligned} f_{Y_1}(y_1) &= \int_0^\infty f_Y(y)dy_2 \int_0^\infty \lambda^2 e^{-\lambda y_1} dy_2 \\ &= \lambda^2 y_1 e^{-\lambda y_1}, y_1 \geq 0. \end{aligned}$$

Donc $Y \sim \gamma(2, \lambda)$.

Espérance et la matrice de variance covariance d'une fonction d'un vecteur aléatoire

Pour obtenir le vecteur espérance $\mathbb{E}(Y) = \mu_Y$ et la matrice de variance covariance Σ_Y d'un vecteur $Y = h(X)$, on peut calculer les espérances μ_{Y_i}, et les (co)variances $\sigma_{Y_iY_j}$ directement à partir de la loi de probabilité conjointe du vecteur X sans passer par les fonctions $\mathbb{P}_Y(y)$ ou $f_Y(y)$.

Proposition 2.3.3. Des formules équivalentes plus compactes pour la variance et la covariance sont obtenues en les exprimant en terme d'espérances :

$$Cov[h_i(X), h_j(X)] = \mathbb{E}[h_i(X)h_j(X)] - \mathbb{E}[h_i(X)]\mathbb{E}[h_j(X)] \tag{2.5}$$

$$Var[h_i(X)] = \mathbb{E}[h_i^2(X)] - \mathbb{E}^2[h_i(X)], \tag{2.6}$$

avec en particulier $Var[h_i(X)] = Cov[h_i(X), h_i(X)]$.

Cas des fonctions linéaires

Dans le cas où les fonctions $h(X)$ sont linéaires, on peut écrire $Y = AX$ et le calcul du vecteur espérance et la matrice de covariance peut se faire à l'aide du calcul matriciel. En se rappelons que $\mathbb{E}(a_{ij}X_j + b_j) = a_{ij}\mathbb{E}(X_j) + b_j$ et que $Var[a_{ij}X_j + b_j] = a_{ij}^2 Var(X_j)$ lorsque b_j est une constante, on obtient le résultat suivant :

$$Y = AX + b \Rightarrow \begin{cases} \mu_Y = A_{\mu_X} + b \\ \sum_Y = A \sum_X A^t \end{cases}$$