MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE A.MIRA DE BEJAIA

FACULTE DES SCIENCES EXACTES

DEPARTEMENT DE MATHEMATIQUES

Polycopié de cours

Variables aléatoires à plusieurs dimensions

N.SAADI

Année 2018-2019

Avant propos

Cet ouvrage s'adresse aussi bien aux chercheurs et professeurs qu'aux étudiants de premier cycle des universités. On y aborde les notions essentielles de la théorie des probabilités dans le cas multidimensionnel, pour son application dans différents domaines des sciences. Le lecteur est supposé connaître les notions fondamentales de la théorie des probabilités dans le cas unidimensionnel et le calcul intégral, y compris les notions de bases sur l'algèbre linéaire.

Les bases mathématiques sont développées de la manière la plus naturelle possible, de sorte à pouvoir disposer rapidement des résultats essentiels. Les notions plus avancées et plus difficiles, comme celles résultant des différentes définitions de convergence dans le cas multidimensionnel sont également présentés.

Ce polycopié est issu d'un cours donné aux étudiants de deuxième année Licence Mathématiques de l'université A.MIRA de Béjaia. Il est composé de 5 chapitres enrichis d'une liste d'éxercices assez variés couvrant différents aspects des concepts étudiés.

Table des matières

Chapitre 1

Généralités sur les vecteurs aléatoires

1.1 Introduction

La notion de vecteurs aléatoires intervient lorsqu'on s'intéresse à mesurer différentes grandeurs numériques dépendants de l'issue d'une même expérience aléatoire. Par exemple, dans l'observation d'une population quelconque on peut s'intéresser à mesurer l'âge, la taille, le poids... de chaque individu. On forme alors un vecteur d'observations aléatoires. Si on surveille une particule M évoluant dans le plan ou dans l'espace, on peut s'intéresser à sa position en fonction du temps déterminée par le vecteur de ses coordonnées aléatoires.

1.2 Rappels sur les variables aléatoires unidimensionnelles

L'espace $(\Omega, \mathcal{A}, \mathbb{P})$

On désigne par Ω l'ensemble des épreuves ou événements élémentaires w et par $\mathcal{A}$ une tribu sur Ω, c'est-à-dire une classe de parties de Ω qui vérifie les axiomes suivantes :

- (A_1) : $\varnothing \in \mathcal{A}$ et $\Omega \in \mathcal{A}$;
- (A_2) : $\mathcal{A}$ est stable par passage au complémentaire : $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$;
- (A_3) : $\mathcal{A}$ est stable par réunion et intersection dénombrables i.e. Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{A}$, alors $\bigcup_{n\in\mathbb{N}} A_n$ et $\bigcap_{n\in\mathbb{N}} A_n$ sont dans $\mathcal{A}$.

Définition 1.1. *L'ensemble des parties de Ω, $\mathcal{P}(\Omega)$ est un exemple de tribu sur Ω. Un élément A d'une tribu $\mathcal{A}$ sera appelé événement aléatoire .*

Définition 1.2. *On appelle probabilité définie sur l'espace $(\Omega, \mathcal{A})$ toute application $\mathbb{P} : \mathcal{A} \to [0, 1]$ satisfaisant les 3 axiomes suivantes :*

(i) $0 \leq \mathbb{P}(A) \leq 1, \forall A \in \mathcal{A}$.

(ii) $\mathbb{P}(\Omega) = 1$.

(iii) *Pour toute suite $(A_i)_{i\in\mathbb{N}^*}$ d'événements de $\mathcal{A}$ deux à deux disjoints (incompatibles) :*

$$\mathbb{P}\left(\bigcup_{i\in\mathbb{N}^*} A_i\right) = \sum_{i\in\mathbb{N}^*} \mathbb{P}(A_i)$$

Proposition 1.2.1. ***(Propriétés générales d'une probabilité)****. Toute probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{A})$ vérifie les proprirétés suivantes :*

1. $\mathbb{P}(\varnothing) = 0$.
2. $\forall A \in \mathcal{A}, \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
3. $\forall A \in \mathcal{A}, \forall B \in \mathcal{A}, A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.
4. $\forall A \in \mathcal{A}, \forall B \in \mathcal{A}, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Définition 1.3. *Le couple $(\Omega, \mathcal{A})$ est appelé espace probabilisable.*

Définition 1.4. *Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé espace probabilisé.*

Définition 1.5. *On appelle espace mesurable tout couple (E, ε) formé par un ensemble E et une tribu ε sur E.*

Définition 1.6. *Soit $\mathcal{C}$ un sous ensemble de $\mathbb{P}(E)$. La tribu engendrée par $\mathcal{C}$ est la plus petite tribu contenant $\mathcal{C}$. Cette tribu peut également être définie comme l'intersection de toutes les tribus contenant $\mathcal{C}$. On la note $\sigma(\mathcal{C})$.*

Définition 1.7. *Si $E = \mathbb{R}$, on appelle tribu borélienne $\mathcal{B}(\mathbb{R})$ la tribu ε engendrée par la classe des intervalles de $\mathbb{R}$ de la forme $]-\infty, a[$ pour $a \in \mathbb{R}$.*

Définition 1.8. *Soient $(\Omega, \mathcal{A})$ et (E, ε) deux espaces masurables. On dit que f est une application mesurable de $(\Omega, \mathcal{A})$ dans (E, ε) si l'image réciproque par f de tout élément de ε est un élément de $\mathcal{A}$. Autrement dit, f est mesurable si $f^{-1}(\varepsilon) \subset \mathcal{A}$.*

Définition 1.9. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et (E, ε) un espace mesurable. Une variable aléatoire X est une application mesurable de $(\Omega, \mathcal{A})$ dans (E, ε) c'est -à- dire une application $X : \Omega \to E$ telle que pour tout borélien B de ε l'ensemble*

$$X^{-1}(B) = \{w \in \Omega : X(w) \in B\} \subset \mathcal{A}.$$

On distingue les cas suivants :

a. Si E est dénombrable, la variable aléatoire X est dite discrète. Dans ce cas, E est muni de la tribu $\mathcal{P}(E)$ et X vérifie

$$\forall x \in E : \{\omega \in \Omega : X = x\} \in \mathcal{A}.$$

b. Si $E = \mathbb{R}$, la variable aléatoire X est dite variable aléatoire réelle continue. En général, $\mathbb{R}$ est muni de sa tribu de Borel $\varepsilon = \mathfrak{B}(\mathbb{R})$, dans ce cas X est une variable aléatoire réelle continue de $(\Omega, \mathcal{A}, \mathbb{P})$ dans $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$.

c. Si $E = \mathbb{R}^n$, on parle de variables aléatoires réelles multidimensionnelles ou vectorielles ou encore de vecteurs aléatoires réels. En général, $\mathbb{R}^n$ est muni de sa tribu de Borel $\varepsilon = \mathfrak{B}(\mathbb{R}^n)$, dans ce cas X est un vecteur aléatoire de $(\Omega, \mathcal{A}, \mathbb{P})$ dans $(\mathbb{R}, \mathfrak{B}(\mathbb{R}^n))$. On utilisera les abréviations (v.a), (v.a.r), (v.a.d) pour désigner une variable aléatoire, une variable aléatoire réelle et une variable aléatoire discrète respectivement.

Définition 1.10. *Si X est une v.a.r et $\mathbb{P}_X : \mathfrak{B}(\mathbb{R}) \to [0,1]$ définie par $\mathbb{P}_X(B) = \mathbb{P}\left(X^{-1}(B)\right)$ une mesure de probabilité sur $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. Cette mesure de probabilité est appelée loi de probabilité de X. On dit également que X suit la loi de probabilité $\mathbb{P}_X$ et on note*

$$X \sim \mathbb{P}_X.$$

Définition 1.11. *la tribu borélienne $\mathfrak{B}(\mathbb{R}^n)$ sur $\mathbb{R}^n$ est la tribu engendrée par les ensembles de la forme*

$$\prod_{i=1}^{n} (]-\infty, x_i[) \; o\grave{u} x_i \in \mathbb{R}, \forall i = 1, ..., n.$$

1.3 Couples de variables aléatoires réelles

Définition 1.12. *Un couple $Z = (X, Y)$ de variables aléatoires réelles est une application mesurable Z définie par :*

$$\begin{aligned} Z : \Omega &\to \mathbb{R}^2 \\ w &\longmapsto Z(w) = (X(w), Y(w)). \end{aligned}$$

Proposition 1.3.1. **(i)** *Etant donné un couple (X, Y) de variables aléatoires de $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$ et $B \in \mathfrak{B}(\mathbb{R}^2)$, l'ensemble $\{w \in \Omega \diagup (X(w), Y(w)) \in B\}$ est un événement.*

(ii) *Inversement, soit une application $(X, Y) : \Omega \to \mathbb{R}^2$ telle que :*

$$\forall B \in \mathfrak{B}(\mathbb{R}^2), \{w \in \Omega \diagup (X(w), Y(w)) \in B\} \in \mathcal{A}.$$

Alors (X, Y) est un couples de v.a de $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$ X et Y sont des variables aléatoires réelles.

Notation

Etant donné un couple (X, Y) de variables aléatoires de $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$, on note pour tout

$$B \in \mathfrak{B}(\mathbb{R}^2), ((X, Y) \in B) = \{w \in \Omega \diagup (X(w), Y(w)) \in B\}.$$

1.3.1 Probabilité image

Théorème 1.3.2. *Soient* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et* (X, Y) *un couple de variables aléatoires réelles* Ω *. Alors l'application* $\mathbb{P}_{XY} : \mathfrak{B}(\mathbb{R}^2) \to [0,1]$ *définie par :*

$$\forall B \in \mathfrak{B}(\mathbb{R}^2), \mathbb{P}_{XY}(B) = \mathbb{P}\left((X, Y) \in B\right)$$

est une probabilité sur $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2)$ *appelée probabilité image de* $\mathbb{P}$ *par* (X, Y)*.*

Définition 1.13. *Soit un espace probabilisé* $(\Omega, \mathcal{A}, \mathbb{P})$*. On dit que deux applications de* Ω *dans* $\mathbb{R}$ *forment un couple de variables aléatoires réelles de* $\mathfrak{B}(\mathbb{R}^2)$ *si, pour tout couple* (x, y) *de réels ,* $[X \leq x]$ *et* $[Y \leq y]$ *sont des éléments de* $\mathcal{A}$*.*

Définition 1.14. *Soient* (X_1, X_2) *et* (Y_1, Y_2) *deux couples de variables aléatores réelles de* $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$*. On dit que* (X_1, X_2) *et* (Y_1, Y_2) *sont équidistribués lorsque*

$$\mathbb{P}_{X_1 X_2}(B) = \mathbb{P}_{Y_1 Y_2}(B), \forall B \in B(\mathbb{R}^2).$$

1.3.2 Fonction de répartition d'un couple de variables aléatoires

Notations

Soit (X, Y) un couple de variables aléatoires réelles de $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$. Etant donnés B et C dans $\mathfrak{B}(\mathbb{R})$, $B \times C$ est un borélien de $\mathbb{R}^2$. On note alors :

$$\mathbb{P}\left((X, Y) \in B \times C)\right) = \mathbb{P}\left(X \in B, Y \in C\right) = \mathbb{P}\left((X \in B) \cap (Y \in C)\right).$$

$$\mathbb{P}\left(X \leq x, Y \leq y\right) = \mathbb{P}\left(X \in]-\infty, x], Y \in]-\infty, y]\right).$$

$$\mathbb{P}\left(X < x, Y < y\right) = \mathbb{P}\left(X \in]-\infty, x[, Y \in]-\infty, y[\right).$$

Définition 1.15. *Soit* (X, Y) *un couple de variables aléatoires de* $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$*. On appelle fonction de répartition du couple* (X, Y)*, l'application*

$$F_{X,Y} : \mathbb{R}^2 \to [0,1]$$

définie par :

$$\forall (x, y) \in \mathbb{R}^2, F_{X,Y}(x, y) = P\left([X \leq x] \cap [Y \leq y]\right) = \mathbb{P}\left(X \leq x, Y \leq y\right).$$

On dit aussi que $F_{X,Y}$ est la loi du couple (X, Y).

Théorème 1.3.3. Soit (X, Y) un couple de variables aléatoires réelles sur l'espace $(\Omega, \mathcal{A}, \mathbb{P})$. Alors :

(i) Pour tout $(a, b) \in \mathbb{R}^2$ et tout $(h, k) \in \mathbb{R}_+^2$, on a

$$\mathbb{P}(a < X \leq a+h, b < Y \leq b+k) = F_{XY}(a+h, b+k) - F_{XY}(a, b+k) - F_{XY}(a+h, b) + F_{XY}(a, b).$$

(ii) $F_{X,Y}$ *est croissante sur* $\mathbb{R}$, et continue à droite par rapport à chacune de ses variables.

(iii) *Pour tout* y *réel fixé,*

$$\lim_{x\to-\infty} F_{X,Y}(x,y) = 0, \lim_{x\to+\infty} F_{X,Y}(x,y) = F_Y(y).$$

Pour tout x réel fixé,

$$\lim_{x\to-\infty} F_{X,Y}(x,y) = 0, \lim_{y\to+\infty} F_{X,Y}(x,y) = F_X(x).$$

(iv)

$$\lim_{x\to+\infty}(\lim_{y\to+\infty} F_{X,Y}(x,y)) = \lim_{y\to+\infty}(\lim_{x\to+\infty} F_{X,Y}(x,y)) = 1.$$

Démonstration. (i) Pour tout $(a,b)\in\mathbb{R}^2$ et tout $(h,k)\in\mathbb{R}_+^2$, on a

$$\mathbb{P}(a<X\leq a+h, b<Y\leq b+k) = \mathbb{P}\left((X,Y)\in]a,a+h]\times]b,b+k]\right) = \mathbb{P}_{X,Y}\left(]a,a+h]\times]b,b+k]\right).$$

Par suite :

$$\begin{aligned}\mathbb{P}_{X,Y}\left(]a,a+h]\times]b,b+k]\right) &= \mathbb{P}_{X,Y}\left(]-\infty,a+h]\times]-\infty,b+k]\right) - \mathbb{P}_{X,Y}\left(]-\infty,a+h]\times]-\infty,b]\right)\\ &\quad -\mathbb{P}_{X,Y}\left(]-\infty,a]\times]-\infty,b+k]\right) + \mathbb{P}_{X,Y}\left(]-\infty,a]\times]-\infty,b]\right)\\ &= F_{XY}(a+h,b+k) - F_{XY}(a,b+k) - F_{XY}(a+h,b) + F_{XY}(a,b).\end{aligned}$$

(ii) Fixons par exemple y et soit a et x deux réels tels que $a\leq x$. $[X\leq a]\subset[X\leq x]$ *d'où,*

$$[X\leq a]\cap[Y\leq y]\subset[X\leq x]\cap[Y\leq y].$$

On en déduit que

$$\mathbb{P}\left[(X\leq a)\cap(Y\leq y)\right]\leq\mathbb{P}\left[(X\leq x)\cap(Y\leq y)\right].$$

C'est-à-dire :

$$F_{X,Y}(a,y)\leq F_{X,Y}(x,y).$$

D'autre part,

$$\begin{aligned}F_{X,Y}(x,y)-F_{X,Y}(a,y) &= \mathbb{P}\left[(X\leq x))\cap(Y\leq y)\right]-\mathbb{P}\left[(X\leq a))\cap(Y\leq y)\right]\\ &= \mathbb{P}\left[((X\leq x)-(X\leq a))\cap(Y\leq y)\right]\leq\mathbb{P}\left[(X\leq x)-(X\leq a)\right].\end{aligned}$$

Or,

$$\mathbb{P}\left[(X\leq x)-(X\leq a)\right] = \mathbb{P}\left(a<X\leq x\right) = F_X(x)-F_X(a).$$

Ainsi, $0\leq F_{X,Y}(x,y)-F_{X,Y}(a,y)\leq F_X(x)-F_X(a)$.
Par conséquent,

$$\lim_{x\to a^+}\left(F_{X,Y}(x,y)-F_{X,Y}(a,y)\right)=0.$$

$F_{X,Y}$ *est continue à droite par rapport à* x, pour y fixé.
On raisonne de même pour y.

(iii) Soit y réel fixé,

$$\mathbb{P}\left[(X \leq x) \cap (Y \leq y)\right] \leq \mathbb{P}(X \leq x)$$

d'où : $F_{X,Y}(x,y)) \leq F_X(x)$ *et par suite,*

$$\lim_{x \to -\infty} F_{X,Y}(x,y) = 0.$$

$$\begin{aligned} F_Y(y) - F_{X,Y}(x,y) &= \mathbb{P}(Y \leq y) - \mathbb{P}[(X \leq x) \cap (Y \leq y)] \\ &= \mathbb{P}[(Y \leq y) - (X \leq x)] \\ &= \mathbb{P}[(Y \leq y) \cap (X > x)] \leq \mathbb{P}(X > x). \end{aligned}$$

Autrement dit, $0 \leq F_Y(y) - F_{X,Y}(x,y) \leq 1 - F_X(x)$.
Par suite,

$$\lim_{x \to +\infty} [F_Y(y) - F_{X,Y}(x,y)] = 0.$$

(iv) Résulte immédiatement de (iii) par passage à la limite.

1.3.3 Indépendance de variables aléatoires

Soit (X,Y) un couple de variables aléatoires. On dit que X et Y sont indépendantes lorsque pour tout couple (A,B) de Boréliens de $\mathfrak{B}(\mathbb{R}^2)$, les événements $[X \in A]$ et $[Y \in B]$ sont indépendants. Autrement dit, les trois propriètès suivantes sont équivalentes.

Propriété 1.3.4. *1. X et Y indépendantes.*

2. $\forall (A,B) \in \mathfrak{B}(\mathbb{R}^2), \mathbb{P}\left(X \in A, Y \in B\right) = \mathbb{P}\left(X \in A\right)\mathbb{P}\left(Y \in B\right)$.
3. $\forall (A,B) \in \mathfrak{B}(\mathbb{R}^2), \mathbb{P}_{XY}\left(A \times B\right) = \mathbb{P}_X\left(A\right)\mathbb{P}_Y\left(B\right)$.

Proposition 1.3.5. *Soient X et Y deux variables aléatoires réelles. Alors X et Y sont indépendantes si et seulement si :*

$$F_{X,Y}(x,y) = F_X(x)F_Y(y).$$

1.3.4 Cas de variables aléatoires discrètes

Considérons un espace probabilisé modélisant une expérience aléatoire de deux variables aléatoires discrètes définies sur l'espace probabilisable $(\Omega, \mathcal{A})$. Chacune de ces variables prend donc ses valeurs sur un sous ensemble au plus dénombrable de $\mathbb{R}$. Nous posons donc
$X(\Omega) = \{x_i, i \in I\}$ et $Y(\Omega) = \{y_j, j \in J\}$, où I et J sont deux ensembles de $\mathbb{N}$. Les valeurs x_i sont supposées distinctes lorsque i parcourt I. De même, nous supposons que les valeurs y_j sont distinctes lorsque j parcourt J. L'ensemble des valeurs prises par le couple (X,Y) ou domaine de variation de (X,Y) est alors

$$X(\Omega) \times Y(\Omega) = \{(x_i, y_j) : i \in I, j \in J\}.$$

1.3.5 Lois conjointes d'un couple de variables aléatoires discrètes

Proposition 1.3.6. *Soit (X, Y) un couple à valeurs discrètes (v.a.d).*

(i) *Si I et J sont finis, on a*

$$\sum_{i\in I}\sum_{j\in J} P[X = x_i, Y = y_j] = 1.$$

(ii) *Pour tout événement B : $\mathbb{P}((X,Y) \in B) = \sum_{(x_i,y_j)\in B\cap X(\Omega)\times Y(\Omega)} P[X = x_i, Y = y_j]$.*

Définition 1.16. *Etant donné un couple (X, Y) de variables aléatoires discrètes (v.a.d) , on appelle loi conjoite du couple (X, Y), l'application qui à tout $(x_i, y_j) \in X(\Omega) \times Y(\Omega)$ associe*

$$\mathbb{P}_{ij} = \mathbb{P}_{_{XY}}(x_i, y_j) = \mathbb{P}\left(X = x_i, Y = y_j\right). \tag{1.1}$$

Lois marginales

Théorème 1.3.7. *Soit (X, Y) un couple de variables aléatoires discrètes. Alors les lois marginales de X et Y sont :*

$$\forall x_i \in X(\Omega), \mathbb{P}_{i\bullet} = \mathbb{P}(X = x_i) = \sum_{j\in J}\mathbb{P}\left([X = x_i, Y = y_j]\right) = \sum_{j\in J}\mathbb{P}_{ij}, \tag{1.2}$$

$$\forall y_j \in Y(\Omega), \mathbb{P}_{\bullet j} = \mathbb{P}(Y = y_{_j}) = \sum_{i\in I}\mathbb{P}\left([X = x_i, Y = y_j]\right) = \sum_{i\in I}\mathbb{P}_{ij}. \tag{1.3}$$

Nous pouvons aussi calculer la fonction de répartition du couple (X, Y) et nous obtenons :

$$\begin{aligned} F_{_{XY}}(x,y) &= \mathbb{P}[X \leq x, Y \leq y] \\ &= \sum_{\substack{(i,j)\in I\times J \\ x_i\leq x, y_j\leq y}} \mathbb{P}_{ij} = \sum_{\substack{i\in I \\ x_i\leq x}}\left(\sum_{\substack{j\in J \\ y_j\leq y}}\mathbb{P}_{ij}\right) = \sum_{\substack{j\in J \\ y_j\leq y}}\left(\sum_{\substack{i\in I \\ x_i\leq x}}\mathbb{P}_{ij}\right). \end{aligned}$$

Remarque 1.3.1. *Les lois marginales ne permettent pas de déterminer la loi : la connaissance de la loi conjointe permet de calculer les lois marginales. Mais la réciproque est généralement fausse.*

Exemple 1.3.1. *Considérons un couple (X, Y) de variables aléatoires discrètes, chacune à valeurs dans $\{0, 1\}$. Supposons que la loi de ce couple soit :*

$$\mathbb{P}[X = 0, Y = 0] = \beta, \;\; \mathbb{P}[X = 0, Y = 1] = \frac{1}{2} - \beta$$

$$\mathbb{P}[X=1,Y=0]=\frac{1}{2}-\beta, \mathbb{P}[X=1,Y=1]=\beta.$$

où $0 \leq \beta \leq \frac{1}{2}$*. Nous déterminons les marginales de* X *et de* Y *en appliquant (1.2) et (1.3) on obtient :*

$$\mathbb{P}[X=0]=\mathbb{P}[X=0,Y=0]+\mathbb{P}[X=0,Y=1]=\frac{1}{2}.$$

$$\mathbb{P}[X=1]=\mathbb{P}[X=1,Y=0]+\mathbb{P}[X=1,Y=1]=\frac{1}{2}.$$

La loi de X *est donc de Bernoulli de paramètre* $\frac{1}{2}$*. De même, on a :*

$$\mathbb{P}[Y=0]=\mathbb{P}[X=0,Y=0]+\mathbb{P}[X=0,Y=1]=\frac{1}{2}.$$

$$\mathbb{P}[Y=1]=\mathbb{P}[X=0,Y=0]+P[X=1,Y=1]=\frac{1}{2}.$$

Les deux variables X *et* Y *suivent donc des lois de Bernoulli qui ne dépendent pas de* β*, alors que la loi conjointe, elle, en dépend. Il est donc impossible de remonter à la loi conjointe du couple* (X,Y) *à partir de la seule donnée des lois marginales de* X *et* Y*.*

Théorème 1.3.8. *Soit* (X,Y) *un couple de v.a.d.* X *et* Y *sont indépendantes si et seulement si pour tout* $(x_i,y_i)\in X(\Omega)\times Y(\Omega)$*, on a :*

$$\mathbb{P}[X=x_i,Y=y_j]=\mathbb{P}(X=x_i)\,\mathbb{P}(Y=y_j).$$

Démonstration. Supposons que les variables X et Y sont indépendantes. Classons les x_i par ordre croissant ainsi que les y_i. L'événement $[X_i=x_i]$ s'écrit encore $[X\leq x_i]\backslash[X\leq x_{i-1}]$ parceque $x_{i-1}<x_i$. De même, nous avons $[Y=y_j]=[Y\leq y_j]\backslash[Y\leq y_{j-1}]$. *Ainsi*

$$\mathbb{P}[X=x_i]=\mathbb{P}[X\leq x_i]-\mathbb{P}[X\leq x_{i-1}]=F_X(x_i)-F_X(x_{i-1}) \tag{1.4}$$

et, de la même manière,

$$\mathbb{P}[Y=y_j]=\mathbb{P}[Y\leq y_j]-\mathbb{P}[Y\leq y_{j-1}]=F_Y(y_j)-F_Y(y_{j-1}). \tag{1.5}$$

Si nous faisons le produit $\mathbb{P}[X=x_i]\times\mathbb{P}[Y=y_j]$, nous obtenons donc :

$$\begin{aligned}\mathbb{P}(X=x_i)\,\mathbb{P}(Y=y_j) &= F_X(x_i)F_Y(y_j)-F_X(x_i)F_Y(y_{j-1})-F_X(x_{i-1})F_Y(y_j)+F_X(x_{i-1})F_Y(y_{j-1}) && (1.6)\\ &= F_{XY}(x_i,y_j)-F_{XY}(x_i,y_{j-1})-F_{XY}(x_{i-1},y_j)+F_{XY}(x_{i-1},y_{j-1}) && (1.7)\end{aligned}$$

où la dernière égalité provient de l'indépendance de X et de Y. Par définition de la fonction de répartition et en prenant en compte que :

$$\begin{aligned}([X\leq x_i]\cap[Y\leq y_j])\setminus([X\leq x_i]\cap[Y\leq y_{j-1}]) &= [X=x_i]\cap([Y\leq y_j]\backslash[Y\leq y_{j-1}]) && (1.8)\\ &= [X\leq x_i]\cap[Y=y_j]. && (1.9)\end{aligned}$$

Nous calculons :

$$\begin{aligned}F_{XY}(x_i,y_j)-F_{XY}(x_i,y_{j-1}) &= \mathbb{P}\left([X\le x_i]\cap[Y\le y_j]\right)-\mathbb{P}\left([X\le x_i]\cap[Y\le y_{j-1}]\right)\\ &= \mathbb{P}\left([X\le x_i]\cap[Y\le y_j]\right)\diagup\left([X\le x_i]\cap[Y\le y_{j-1}]\right)\\ &= \mathbb{P}\left([X\le x_i]\cap[Y=y_j]\right).\end{aligned}$$

De la même manière, on établit l'égalité :

$$F_{XY}(x_{i-1},y_j)-F_{XY}(x_{i-1},y_{j-1})=\mathbb{P}\left([X\le x_{i-1}]\cap[Y=y_j]\right). \tag{1.10}$$

En reportant (1.10) et (1.9) dans (1.6), il vient :

$$\begin{aligned}\mathbb{P}[X=x_i]\mathbb{P}[Y=y_j] &= \mathbb{P}\left([X\le x_i]\cap[Y=y_j]\right)-\mathbb{P}\left([X\le x_{i-1}]\cap[Y=y_j]\right)\\ &= \mathbb{P}\left([X\le x_i]\cap[Y=y_j]\right)/\left([X\le x_{i-1}]\cap[Y=y_j]\right).\end{aligned}$$

Puisque

$$\left([X\le x_i]\cap[Y=y_j]\right)\setminus\left([X\le x_{i-1}]\cap[Y=y_j]\right)=[X\le x_i]\cap[Y=y_j],$$

nous obtenons finalement *:*

$$\mathbb{P}\left([X=x_i]\cap[Y=y_j]\right)=\mathbb{P}\left([X=x_i]\right)\mathbb{P}\left([Y=y_j]\right).$$

Si nous calculons maintenant la fonction de répartition du couple (X,Y), nous trouvons que :

$$\begin{aligned}F_{XY}(x,y) &= \sum_{\substack{(i,j)\in I\times J\\ x_i\le x, y_j\le y}}\mathbb{P}[X=x_i,Y=y_j]\\ &= \sum_{\substack{i\in I\\ x_i\le x}}\sum_{\substack{j\in J\\ y_j\le y}}\mathbb{P}[X=x_i]\mathbb{P}[Y=y_j]\\ &= \sum_{\substack{i\in I\\ x_i\le x}}\mathbb{P}[X=x_i]\sum_{\substack{j\in J\\ y_j\le y}}\mathbb{P}[Y=y_j]\\ &= F_X(x)\times F_Y(y).\end{aligned}$$

La fonction de répartition du couple (X,Y) est le produit des fonctions de répartition de X et de Y. Nous dirons donc que les variables aléaatoires X et Y sont indépendantes car les lois marginales de X et de Y ne dépendent aucunement l'une de l'autre.

1.3.6 Lois conditionnelles

Définition 1.17. *Etant donné un couple (X,Y) de v.a.d et $y_j\in Y(\Omega)$ tel que $\mathbb{P}(Y=y_j)\neq 0$. On appelle loi conditionnelle de X sachant $[Y=y_j]$, l'application définie sur $X(\Omega)$ par :*

$$\mathbb{P}\left(X=x_i/Y=y_j\right)=\frac{\mathbb{P}(X=x_i,Y=y_j)}{\mathbb{P}(Y=y_j)}. \tag{1.11}$$

De même, si $x_i \in X(\Omega)$ est tel que $\mathbb{P}(X = x_i) \neq 0$, on appelle loi conditionnelle de Y sachant $[X = x_i]$, l'application définie sur $Y(\Omega)$ par :

$$\mathbb{P}\left(Y = y_j/X = x_i\right) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(X = x_i)}. \tag{1.12}$$

Théorème 1.3.9. ***Théorème des probabilités composées [?].*** *Soit (X, Y) un couple de v.a.d . Pour tout $(x_i, y_j) \in X(\Omega) \times Y(\Omega)$, on a*

$$\mathbb{P}(X = x_i, Y = y_j) = \begin{cases} P(X{=}x_i/Y = y_j)\mathbb{P}(Y = y_j) & si\ \mathbb{P}(Y = y_j) \neq 0, \\ 0 & sinon. \end{cases} \tag{1.13}$$

Théorème 1.3.10. ***Théorème des probabilités totales [?].*** *Soit (X, Y) un couple de v.a.d . Pour tout $(x_i, y_j) \in X(\Omega) \times Y(\Omega)$, on a*

$$\mathbb{P}(X = x_i) = \sum_{j \in J} \mathbb{P}\left(X = x_i/Y = y_j\right)\mathbb{P}(Y = y_j); \mathbb{P}(Y = y_j) = \sum_{i \in I} \mathbb{P}\left(Y/X = x_i\right)\mathbb{P}(X = x_i).$$

Exercice. Un lot de n articles présente un mélange des poduits de trois usines : n_1 articles de l'usine U_1, n_2 articles de l'usine U_2 et n_3 articles de l'usine U_3. Pour les articles de l'usine U_1, la probabilité de fonctionner sans défaillance pendant un temps τ est p_1, de l'usine U_2 est p_2 et de l'usine U_3 est p_3. On tire au hasard un article, calculer la probabilité que l'article fonctionnera sans défaillance pendant un temps τ.

Solution : Soit $A = \{$ article tiré fonctionnera sans défaillance pendant un temps $\tau\}$ et pour $i = 1, 2, 3$, $B_i = \{$article tiré appartient à l'usine $U_i\}$. On cherche à calculer $\mathbb{P}(A)$. Par la propriété des probabilités totale , on a

$$\mathbb{P}(A) = \sum_{i=1}^{3} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{3} \mathbb{P}(B_i)\mathbb{P}(A/B_i), B_i \neq \varnothing, \bigcup_i B_i = \Omega, B_i \cap B_j = \varnothing, i \neq j, i = 1, 2, 3.$$

Or, pour $i = 1, 2, 3$, $\mathbb{P}(B_i) = \frac{n_i}{n}$ et $\mathbb{P}(A/B_i) = p_i$, d'où :

$$\mathbb{P}(A) = \sum_{i=1}^{3} \frac{n_i}{n} p_i = \frac{n_1 p_1 + n_2 p_2 + n_3 p_3}{n}.$$

Théorème 1.3.11. ***Théorème de Bayes [?].*** *Soit (X, Y) un couple de variables aléatoires discrètes. Alors pour tout $(x_i, y_j) \in X(\Omega) \times Y(\Omega)$, on a*

$$\mathbb{P}(X = x_i/Y = y_j) = \frac{\mathbb{P}(Y = y_j/X = x_i)\mathbb{P}(X = x_i)}{\sum_i \mathbb{P}(Y/X = x_i)\mathbb{P}(X = x_i)};\ \mathbb{P}(Y = y_j/X = x_j) = \frac{\mathbb{P}(X = x_i/Y = y_j)\mathbb{P}(Y = y_j)}{\sum_j \mathbb{P}(X = x_i/Y = y_j)\mathbb{P}(Y = y_j)} \tag{1.14}$$

Conditionnement d'une variable aléatoire discrète par un événement de probabilité non nulle

Soient $A \in \mathcal{A}$ tel que $\mathbb{P}(A) > 0$ et une variable aléatoire X et à valeurs discrètes dans $\{x_i : i \in I\}$ où I est au plus dénombrable et les valeurs x_i sont distinctes deux à deux. On a :

$$\sum_{i\in I} \mathbb{P}[X = x_i/A] = \sum_{i\in I} \mathbb{P}\left([X = x_i] \cap A\right)/\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i\in I}(X = x_i) \cap A\right)/\mathbb{P}(A). \qquad (1.15)$$

Exemple 1.3.2. *On considère ici deux variables aléatoires discrètes X et Y dont la loi jointe est donnée dans le tableau suivant :*

$X \setminus Y$	0	1
0	$\frac{2}{8}$	$\frac{3}{8}$
1	$\frac{1}{8}$	$\frac{2}{8}$

Supposons que l'événement $[X = 0]$ est réalisé et déterminons la loi de Y . On a :

$$\mathbb{P}\left(Y = 0/X = 0\right) = \frac{\mathbb{P}\left(X = 0, Y = 0\right)}{\mathbb{P}(X = 0)} = \frac{2}{5}$$

et

$$\mathbb{P}\left(Y = 1/X = 0\right) = \frac{3}{5}.$$

On vient ainsi de définir une nouvelle variable aléatoire notée $[Y/X = 0]$ qui suit une loi de Bernoulli de paramètre $\frac{3}{5}$. De même il est facile de voir que la variable aléatoire $[Y/X = 1]$ suit une loi de Bernoulli de paramètre $\frac{2}{3}$. Ces lois sont appelées lois conditionnelles de Y sachant X.

1.3.7 Cas des V.a.r absolument continues

Définition 1.18. *Une application de $\mathbb{R}^2$ dans $\mathbb{R}$ est une densité de probabilité dans $\mathbb{R}^2$ si elle est mesurable, positive ou nulle et vérifie la condition de normalisation :*

$$\int_{\mathbb{R}^2} f(x, y)dxdy = 1. \qquad (1.16)$$

Définition 1.19. *Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}^2$. On dit que le couple (X, Y) est absolument continu s'il existe une densité de probabilité $f_{X,Y}$ sur $\mathbb{R}^2$ telle que, pour tout couple (x, y) de réels :*

$$F_{X,Y}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(\mu, \nu)d\mu d\nu.$$

Si l'on connait la densité de probabilité $f_{X,Y}$ d'un couple (X, Y) absolument continu de variables aléatoires, on accéde donc à la fonction de répartition $F_{X,Y}$ de ce couple en intégrant la densité, variable après variable, en choisissant l'ordre d'intégration à sa guise puisque, pour tout $x, y \in \mathbb{R}$:

$$F_{X,Y}(x, y) = \int_{-\infty}^{y}\left(\int_{-\infty}^{x} f_{X,Y}(\mu, \nu)d\mu\right)d\nu = \int_{-\infty}^{x}\left(\int_{-\infty}^{y} f_{X,Y}(\mu, \nu)d\nu\right)d\mu.$$

Exemple 1.3.3. *Soit (X, Y) un couple de variables aléatoires dont la loi est donnée par la densité :*

$$f_{X,Y}(x, y) = \frac{xy}{2} 1_{0 \leq y \leq x \leq 2}.$$

Déterminons la fonction de répartition au point (x_0, y_0). Il faut distinguer plusieurs cas. Par exemple, si $0 \leq y_0 \leq x_0 \leq 2$.

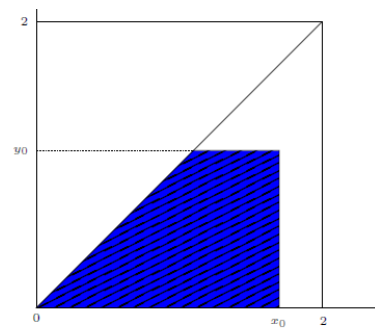

FIGURE 1.1 – Domaine d'intégration

La fonction de répartition s'obtienne en intégrant la densité sur le domaine hachuré 1.1. On peut calculer cette intégrale de deux manières différentes.

$$F_{(X,Y)}(x_0, y_0) = \int_0^{y_0} \int_y^{x_0} \frac{xy}{2} dxdy = \frac{y_0^2}{16}(2x_0^2 - y_0^2), 0 \leq y_0 \leq x_0 \leq 2$$

où

$$F_{(X,Y)}(x_0, y_0) = \int_0^{y_0} \int_0^{x_0} \frac{xy}{2} dxdy + \int_{y_0}^{x_0} \int_0^{y_0} \frac{xy}{2} dydx = \frac{y_0}{16}(2x_0^2 - y_0^2)$$

En répétant ces calculs sur les différents ensembles, on obtient

$$F_{X,Y}(x, y) = \begin{cases} 0 & si\ x < 0\ ou\ y < 0 \\ \frac{y^2}{16}(2x^2 - y^2) & si 0 \leq y \leq x \leq 2 \\ \frac{y^2}{16}(8 - y^2) & si\ 0 < y \leq 2 \leq x \\ \frac{x^4}{16} & si\ 0 \leq x \leq y\ et\ x < 2 \\ 1 & si\ x \leq 2\ et\ y \leq 2 \end{cases} \tag{1.17}$$

Distributions marginales

Définition 1.20. *Les fonctions de répartition marginales $F_X(x)$ et $F_Y(y)$ sont les fonctions de répartition de chacune des deux variables considérées séparément, avec :*

$$F_X(x) = \mathbb{P}(X \leq x); F_Y(y) = \mathbb{P}(Y \leq y).$$

Elles peuvent s'obtenir directement à partir de la fonction de répartition conjointe, avec

$$F_X(x) = F_{X,Y}(x, +\infty); F_Y(y) = F_{X,Y}(+\infty, y).$$

Démonstration. *Posons les événements* $A \equiv (X \leq x)$ *et* $B \equiv (Y \leq y)$, *de telle sorte que* $F_{X,Y}(x,y) = \mathbb{P}(A \cap B)$. *L'événement* $\Omega \equiv (Y \leq \infty)$ *est l'événement certain, et donc* $F_{X,Y}(x,\infty) = \mathbb{P}(A \cap \Omega) = \mathbb{P}(A) = F_X(x)$. *L'autre résultat se démontre de manière similaire.*

Proposition 1.3.12. *Soit* (X, Y) *un couple de variables aléatoires à valeurs dans* $\mathbb{R}^2$ *absolument continu. Soient* $F_{X,Y}$ *la fonction de répartition du couple* (X, Y) *et* $f_{X,Y}$ *la densité de probabilité de ce couple.*

i) $\mathbb{P}[(X,Y) \in B] = \int_B f_{X,Y}(x,y)dxdy$, *pour tout* $B \in \mathfrak{B}(\mathbb{R}^2)$.

ii) *Les variables aléatoires sont absolument continues de densité de probabilité respectives :*

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy \;\; et \;\; f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)dx.$$

Ces densités de probabilité sont appelées densités marginales de X *et de* Y *respectivement.*

iii) $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial y \partial x}(x,y)$ *pour presque tout* $(x,y) \in \mathbb{R}^2$.

Exemple 1.3.4. *Soit la fonction* $f(x,y)$ *suivante :*

$$f_{X,Y}(x,y) = \begin{cases} 2(x+y-2xy) & (x,y) \in [0,1] \times [0,1]; \\ 0 & \textit{ailleurs.} \end{cases} \tag{1.18}$$

1. *On demande de vérifier qu'il s'agit d'une fonction de densité de probabilié.*
2. *Calculer la probabilité de l'événement* $B \equiv \left(\mid X - Y \mid \leq \frac{1}{2}\right)$.
3. *Calculer les lois marginales de* X *et* Y.

1. *La fonction* $f_{X,Y}(x,y)$ *est une densité de proailité à condition que* $f_{X,Y}(x,y) \geq 0, \forall (x,y) \in \mathbb{R}^2$ *et que* $\int_{\mathbb{R}^2} f(x,y)dxdy = 1$. *La première condition est vérifiée, puisque* $2(x+y-2xy) \geq 0$ *pour tout* $(x,y) \in [0,1] \times [0,1]$. *Pour la seconde condition, on obtient :*

$$\int_0^1 \left(\int_0^1 2(x+y-2xy)dy \right) dx = \int_0^1 \left[2xy + y^2 - 2xy^2\right]_0^1 dx = \int_0^1 dx = 1.$$

Ce qui montre que $f_{X,Y}(x,y)$ *est bien une fonction densité de probabilité.*

2. *On peut calculer :*

$$\mathbb{P}(B) = \int_0^{\frac{1}{2}} \int_0^{x+\frac{1}{2}} f_{X,Y}(x,y)dydx + \int_{\frac{1}{2}}^1 \int_{x-\frac{1}{2}}^1 f_{X,Y}(x,y)dydx$$

Sur base de la symétrie de la fonction $f_{X,Y}(x,y)$, *on voit cependant que* $\mathbb{P}(\Omega) = \mathbb{P}(B) + 2\mathbb{P}(A) = 1$, *avec :*

$$\begin{aligned} \mathbb{P}(A) &= \int_{\frac{1}{2}}^1 \int_0^{x-\frac{1}{2}} f_{X,Y}(x,y)dydx = \int_{\frac{1}{2}}^1 \left[2xy + y^2 - 2xy^2\right]_0^{x-\frac{1}{2}} dx \\ &= \int_{\frac{1}{2}}^1 \left[-2x^3 + 5x^2 - \frac{5x}{2} + \frac{1}{4}\right] dx = \left[-\frac{x^4}{2} + \frac{5x^3}{3} - \frac{5x^2}{4} + \frac{x}{4}\right]_{\frac{1}{2}}^1 = \frac{17}{96} \end{aligned}$$

et l'on obtient $\mathbb{P}(B) = 1 - 2\frac{17}{96} = \frac{31}{48}$.

3. *Sur la base de la formule $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$, on obtient :*

$$f_X(x) = \int_0^1 2\left[x + y - 2xy\right] dy = \left[2xy + y^2 - 2xy\right]_0^1 = 1, \forall x \in [0,1],$$

tandis qu'elle vaut 0 ailleurs. La fonction $f(x,y)$ étant symétrique par rapport aux variables X et Y, on obtient directement $f_Y(y) = 1, \forall y \in [0,1]$ et vaut 0 ailleurs. On a donc $X \sim U_{[0,1]}$ et $Y \sim U_{[0,1]}$.

Lois conditionnelles

Définition 1.21. *Soit un couple (X,Y) absolument continu de variable aléatoires réelles, $f_{X,Y}$ la densité de probabilité de ce couple, f_X et f_Y les densités de probabilié marginales respectives de X et de Y. Pour tout $(x,y) \in \mathbb{R}^2$, on définit les densités de probabilité conditionnelles de X sachant $[Y = y]$ et de Y sachant $[X = x]$ par :*

$$f_{X/Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, f_Y(y) > 0; f_{Y/X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}, f_X(x) > 0. \tag{1.19}$$

et les fonctions de répartition conditionnelles de X sachant $[Y = y]$ et de Y sachant $[X = x]$ par :

$$F_{X/Y=y}(x) = \int_{-\infty}^{x} f_{X/Y=y}(t)dt, F_{Y/X=x}(y) = \int_{-\infty}^{y} f_{Y/X=x}(t)dt. \tag{1.20}$$

Conditionnement d'une variable aléatoire absolument continue par un événement de probabilité non nulle

Proposition 1.3.13. *Soient X une variable aléatoire réelle et un événement $A \in \mathfrak{B}(\mathbb{R})$ tel que $\mathbb{P}(A) \neq 0$. L'application $F_{X/A}$ de $\mathbb{R}$ dans $[0,1]$ qui associé à tout $x \in \mathbb{R}$ la probabilité conditionnelle $F_{X/A}(x) = \mathbb{P}[X \leq x/A]$ est une fonction de répartition appelée fonction de répartition conditionnelle de X sachant A.*

Lemme 1.3.14. *Soient un espace probabilisé $(\Omega, \mathfrak{B}(\mathbb{R}^2), \mathbb{P})$ et $A \in \mathfrak{B}(\mathbb{R}^2)$ tel que $0 < \mathbb{P}(A) < 1$ et une variable aléatoire X définie sur l'espace probabilisable $(\Omega, \mathfrak{B}(\mathbb{R}^2)$ admettant des fonctions de répartition conditionnelle $F_{X\diagup A}$ et $F_{X\diagup A^c}$ absolument continues de densité respectives $f_{X/A}$ et f_{X/A^c}. La variable aléatoire X est absolument continue de densité*

$$f_X(x) = \mathbb{P}(A)f_{X/A}(x) + \mathbb{P}(A^c)f_{X/A^c}(x). \tag{1.21}$$

Démonstration. *Par définition de la fonction de répartition conditionnelle $F_{X/A}$ de X sachant A, nous avons :*

$$F_{X/A}(x) = \frac{\mathbb{P}\left([X \leq x] \cap A\right)}{\mathbb{P}(A)}.$$

Si $F_{X/A}$ admet une densité $f_{X/A}$, alors nous avons aussi :

$$F_{X/A}(x) = \int_{-\infty}^{x} f_{X/A}(t)dt.$$

Nous déduisons de ces égalités que :

$$\mathbb{P}\left([X \leq x] \cap A\right) = \int_{-\infty}^{x} \mathbb{P}(A) f_{X/A}(t)dt.$$

Puisque $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, *un raisonnement analogue conduit à l'égalité :*

$$\mathbb{P}\left([X \leq x] \cap A^c\right) = \int_{-\infty}^{x} \mathbb{P}(A^c) f_{X/A^c}(t)dt.$$

Puisque les ensembles $[X \leq x] \cap A$ *et* $[X \leq x] \cap A^c$ *sont disjoints et d'union égale à* $[X \leq x]$, *la fonction de répartition* F_X *de* X *s'écrit :*

$$F_X(x) = \mathbb{P}[X \leq x] = \mathbb{P}\left([X \leq x] \cap A\right) + \mathbb{P}\left([X \leq x] \cap A^c\right).$$

Il vient donc :

$$F_X(x) = \int_{-\infty}^{x} \left(\mathbb{P}(A) f_{X/A}(t) + \mathbb{P}(A^c) f_{X/A^c}(t)\right) dt = \int_{-\infty}^{x} f_X(t)dt.$$

D'où le résultat.

Exemple 1.3.5. *Soit la fonction densité* $f(x, y)$ *donnée par :*

$$f_{X,Y}(x, y) = \begin{cases} 6x & \text{pour } 0 \leq x \leq y \leq 1; \\ 0 & \text{ailleurs.} \end{cases}$$

Les distributions marginales sont :

$$f_X(x) = \begin{cases} \int_x^1 6xdy = 6x(1-x) & \text{pour } x \in [0, 1]; \\ 0 & \text{ailleurs.} \end{cases}$$

$$f_Y(y) = \begin{cases} \int_0^y 6xdx = 3y^2 & \text{pour } y \in [0, 1]; \\ 0 & \text{ailleurs.} \end{cases}$$

Les fonctions de répartition marginales :

$$F_X(x) = \begin{cases} 0 & x < 0; \\ \int_0^x 6u(1-u)du = 3x^2 - 2x^3 & \text{pour } x \in [0, 1]; \\ 1 & x > 1. \end{cases}$$

$$F_Y(y) = \begin{cases} 0 & y < 0; \\ \int_0^y 3v^2dv = y^3 & \text{pour } y \in [0, 1]; \\ 1 & y > 1. \end{cases}$$

Les densités conditionnelles sont alors :

$$f_{X/Y=y}(x) = \begin{cases} \frac{2x}{y^2} & \text{pour } 0 \leq x \leq y \leq 1; \\ 0 & \text{ailleurs.} \end{cases}$$

$$f_{Y/X=x}(y) = \begin{cases} \frac{1}{1-x} & \textit{pour } 0 \leq x \leq y \leq 1; \\ 0 & \textit{ailleurs} \end{cases}$$

et les distributions conditionnelles :

$$F_{X/Y=y}(x) = \begin{cases} 0 & x < 0; \\ \int_0^x \frac{2u}{y^2} du = \frac{x^2}{y^2} & 0 \leq x \leq y \leq 1; \\ 1 & x > y. \end{cases}$$

$$F_{Y/X=x}(y) = \begin{cases} 0 & y < x; \\ \int_0^y \frac{1}{1-x} dv = \frac{y}{1-x} & 0 \leq x \leq y \leq 1; \\ 1 & y > 1. \end{cases}$$

Proposition 1.3.15. *Soient deux variables aléatoires réeeles X et Y absolument continues de densité respective f_X et f_Y. Les variables aléatoires X et Y sont indépendantes si et seulement si le couple (X, Y) est absolument continu de densité :*

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

Démonstration. *Tout d'abord, remarquons que le produit $f_{XY} = f_X \times f_Y$ est une densité de probabilité puisque :*

$$\int f(x, y)dxdy = \left(\int_{\mathbb{R}} f_X(x)dx \right) \left(\int_{\mathbb{R}} f_Y(y)dy \right) = 1.$$

Puisuqe $f_X \times f_Y$ est une densité deprobabilité du couple (X, Y), la fonction de répartition associée s'écrit :

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) du dv \\ &= \left(\int_{-\infty}^x f_X(u)du \right) \left(\int_{-\infty}^y f_Y(v)dv \right) \\ &= F_X(x) \times F_Y(y). \end{aligned}$$

Réciproquement, supposons que X et Y soient indépendantes. Considérons alors la fonction $f = f_X \times f_Y$. Soit $(x, y) \in \mathbb{R}^2$. Nous avons :

$$\begin{aligned} F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv &= \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) du dv \\ &= \left(\int_{-\infty}^x f_X(u)du \right) \left(\int_{-\infty}^y f_Y(v)dv \right) \\ &= F_X(x) F_Y(y). \end{aligned}$$

Puisque les variables aléatoires X et Y sont supposées indépendantes,

$$F_X(x) \times F_Y(y) = F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(y, v) du dv.$$

Exemple 1.3.6. *Soit $Z = (X, Y)^t$ un couple aléatoire de loi uniforme sur $\mathcal{D} = [-1, 1] \times [-1, 1]$. On a alors :*

$$f_{X,Y}(x, y) = \frac{1}{4} 1_{\mathcal{D}}(x, y) = \frac{1}{4} 1_{[-1,1]}(x) 1_{[-1,1]}(y) = f_X(x) f_Y(y)$$

avec

$$f_X(x) = \frac{1}{2} 1_{[-1,1]}(x);\, f_Y(y) = \frac{1}{2} 1_{[-1,1]}(y).$$

Théorème 1.3.16. ***Théorème des probabilité composées.*** *Soit (X, Y) un couple de variables aléatoires réelles (v.a.r) continu défini sur $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$. Alors pour tout $(x, y) \in \mathbb{R}^2$, on a*

$$f_{X,Y}(x, y) = f_{Y/X=x}(y) f_X(x) = f_{X/Y=y}(x) f_Y(y). \tag{1.22}$$

Théorème 1.3.17. ***Théorème des probabiités totales .*** *Soit (X, Y) un couple de v.a.r continu . Pour tout $(x, y) \in \mathbb{R}^2$, on a*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X/Y}(x) f_Y(y) dy; \quad f_Y(y) = \int_{-\infty}^{\infty} f_{Y/X}(y) f_X(x) dx.$$

Théorème 1.3.18. ***Théorème de Bayes .*** *Soit (X, Y) un couple de variables aléatoires réelles. Pour tout $(x, y) \in \mathbb{R}^2$, on a*

$$f_{(X/Y=y)}(x) = \frac{f_{(Y/X=x)}(y) f_X(x)}{\int_{-\infty}^{\infty} f_{(Y/X=x)}(y) f_X(x) dx};\, f_{(Y/X=x)}(y) = \frac{f_{(X/Y=y)}(x) f_Y(y)}{\int_{-\infty}^{\infty} f_{(X/Y=y)}(x) f_Y(y) dy}.$$

1.3.8 Transformation d'un couple aléatoire

Théorème 1.3.19. Soit (X, Y) un couple de variables aléatoires discrètes 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 définie sur $B \subseteq X(\Omega) \times Y(\Omega)$. Lorsque $X(\Omega)$ et $Y(\Omega)$ sont finis, on a

$$\mathbb{P}_Z(z) = \sum_{(x,y) \in B} \mathbb{P}(X = x, Y = y). \tag{1.23}$$

Théorème 1.3.20. *Soit (X, Y) un couple de variables aléatoires continues 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 continue définie sur $B \subseteq \mathfrak{B}(\mathbb{R}^2)$, alors*

$$F_Z(z) = \int_{(x,y) \in B} f(x, y) dy dx. \tag{1.24}$$

Quelques cas particuliers

Quelle que soit la fonction $Z = \phi(X, Y)$, il est toujours possible de déterminer les fonctions F_Z et f_Z en utilisant le procédé général exposé ci-dessus. On remarquera cependant que, dans certaines situations, la fonction f_Z peut s'obtenir directement sans passer par le calcul de F_Z.

Cas $n°1$: $Z = X + Y$ avec X et Y sont indépendantes

Théorème 1.3.21. *Pour une somme de deux variables aléatoires indépendantes, la fonction de masse de probabilité est la convolution des fonctions de masses de probabilité marginales pour X et Y. C'est-à-dire,*

$$\mathbb{P}_Z(z) = \begin{cases} \sum\limits_i \mathbb{P}_X(x_i)\mathbb{P}_Y(z-x_i) \\ \sum\limits_j \mathbb{P}_X(z-y_j)\mathbb{P}_Y(y_j) \end{cases} ; f_Z(z) = \begin{cases} \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx \\ \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy. \end{cases} ;$$

Démonstration. *L'événement $Z \leq z$ est équivalent à $X + Y \leq z$, et l'on obtient dans le cas continu*

$$F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)dydx = \int_{-\infty}^{\infty} f_X(x)\left[\int_{-\infty}^{z-x} f_Y(y)dy\right]dx = \int_{-\infty}^{\infty} f_X(x)F_Y(z-x)dx.$$

Dans le cas continu. En dérivant , on a

$$\begin{aligned} f_Z(z) &= \frac{dF_Z(z)}{dz} = \int_{-\infty}^{\infty} f_X(x)\frac{\partial F_Y(z-x)}{\partial z}dx \\ &= \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx. \end{aligned}$$

La démonstration de l'autre résultat est similaire avec y au lieu de x.

Cas $n°2$: $Z = \frac{X}{Y}$ avec X et Y quelconques

Théorème 1.3.22. *Pour un rapport de deux variables quelconques, la loi de probabilité peut s'obtenir directement à partir de la fonction de probabilité conjointe grâce aux formules :*

$$\mathbb{P}_Z(z) = \sum_j \mid y_j \mid \mathbb{P}(y_j z, y_j); f_Z(z) = \int_{-\infty}^{\infty} \mid y \mid f(yz, y)dy.$$

Cas $n°3$: $Z = \max(X, Y)$ ou $Z = \min(X, Y)$ avec X et Y sont indépendantes

Théorème 1.3.23. • *Si $Z = \max(X, Y)$ avec X et Y sont indépendantes, alors :*

$$\begin{aligned} F_Z(z) &= F_X(z)F_Y(z). \\ f_Z(z) &= f_X(z)F_Y(z) + F_X(z)f_Y(z). \end{aligned}$$

• *Si $Z = \min(X, Z)$ avec X et Y sont deux variables aléatoires indépendantes, alors :*

$$1 - F_Z(z) = (1 - F_X(z))(1 - F_Y(z)). \tag{1.25}$$

$$f_Z(z) = f_X(z)(1 - F_Y(z)) + (1 - F_X(z))f_Y(z). \tag{1.26}$$

Démonstration. • *L'événement $[Z \leq z]$ est équivalent à $[(X \leq z) \cap (Y \leq z)]$, puisque tous les couples (x, y) pour lesquels $x \leq z$ et $y \leq z$ satisfont la condition $\max(x, y) \leq z$. On a donc $F_Z(z) = F_{X,Y}(z, z)$. Si X et Y sont indépendantes, on a $F_{X,Y}(z, z) = F_X(z)F_Y(z)$. Dans le continu, on obtient*

$$f_Z(z) = \frac{dF_{X,Y}(z,z)}{dz} = \frac{d[F_X(z)F_Y(z)]}{dz} = F_Y(z)\frac{dF_X(z)}{dz} + F_X(z)\frac{dF_Y(z)}{dx} = F_Y(z)f_X(z) + F_X(z)f_Y(z).$$

• *L'événement $[Z \leq z]$ est équivalent à $[(X \leq z) \cup (Y \leq z)]$, puisque tous les couples (x, y) pour lesquels $x \leq z$ ou $y \leq z$ satisfont la condition $\min(x, y) \leq z$.
On peut donc également écrire que $(Z \leq z) = (X \leq z \cup Y \leq z)$,
c'est-à-dire $(Z > z) = (X > z \cap Y > z)$. Puisque X et Y sont indépendantes, cela revient à écrire $1 - F_Z(z) = (1 - F_X(z))(1 - F_Y(z))$. En dérivant cette expression par rapport à z, on trouve la formule proposée pour $f_Z(z)$.*

1.4 Vecteur aléatoire à plusieurs dimensions

Définition 1.22. *On considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et n variables aléatoires $X_1, X_2, ..., X_n$ définies sur Ω à valeurs dans $\mathbb{R}$. On considère l'application X de $(\Omega, \mathcal{A}, \mathbb{P})$ dans l'espace mesurable $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ définie pour tout $w \in \Omega$ par :*

$$\begin{aligned} X : \Omega &\to \mathbb{R}^n \\ w &\to X(w) = (X_1(w), X_2(w), ..., X_n(w))^t. \end{aligned}$$

On dit que X est un vecteur aléatoire défini sur $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$

Lemme 1.4.1. *Avec les notations de la définition précédente, $X = (X_1, X_2, ..., X_n)^t$ est un vecteur aléatoire si est seulement si*

$$\{w \in \Omega, X_1(w) \leq x_1, X_2(w) \leq x_2, ..., X_n(w) \leq x_n\} \in \mathfrak{B}(\mathbb{R}^n) \text{ pour tout } n\text{−uplet } (x_1, x_2, ..., x_n)$$

de réels.

On posera :

$$[X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n] = \{w \in \Omega : X_1(w) \leq x_1, X_2(w) \leq x_2, ..., X_n(w) \leq x_n\}.$$

Définition 1.23. ***Loi de probabilité d'un vecteur aléatoire*** *La loi de probabilité d'un vecteur aléatoire X est la probabilité image de $\mathbb{P}$ par X notée $\mathbb{P}_X$ et définie par*

$$\forall B \in \mathfrak{B}(\mathbb{R}^n), \mathbb{P}_X(B) = \mathbb{P}\left(X^{-1}(B)\right).$$

La probabilité $\mathbb{P}_X$ ainsi définie est une mesure de probabilité sur l'espace probabilisable $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ appelée aussi loi conjointe de $(X_1, X_2, ..., X_n)^t$: On appelle ième loi marginale de X et on note $\mathbb{P}_{X_i}$ la loi de la ième composante $X_i (i = 1, 2, ..., n)$.

Définition 1.24. ***Fonction de répartition conjointe.*** *Soit $X = (X_1, X_2, ..., X_n)^t$ un vecteur aléatoire à valeurs dans $\mathbb{R}^n$. Pour tout $x = (x_1, ..., x_n)^t \in \mathbb{R}^n$, on définit la fonction de répartition $F_X(x)$ de X en x en posant :*

$$\begin{aligned} F_X(x) &= F_X(x_1, x_2, ..., x_n) = \mathbb{P}[X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n] \\ &= \mathbb{P}\left([X_1 \leq x_1] \cap [X_2 \leq x_2] \cap ... \cap [X_n \leq x_n]\right). \end{aligned}$$

Propriété 1.4.2. • *Si $\forall\, i = 1, ..., n$ on a $x_i = \infty$, alors $F_X(+\infty) = \mathbb{P}(\Omega) = 1$.*

- *Si $\exists i = 1, ..., n$ tel que $x_i = -\infty$, alors $F_X(x) = \mathbb{P}(\varnothing) = 0$.*

Démonstration. *L'événement $[X_i \leq +\infty]$ est certain pour la variable X_i, tandis que $[X_i \leq -\infty]$ est l'événement impossible. On a donc $F_X(+\infty, ..., +\infty) = \mathbb{P}(\Omega) = 1$, tandis que $F_X(x_1, ..., x_n) = \mathbb{P}(\varnothing) = 0$ si au moins l'un des x_i est égale à $-\infty$.*

Définition 1.25. ***Vecteur aléatoire discret.*** *On appelle vecteur aléatoire discret à valeurs dans $\mathbb{R}^n$ toute application de la forme :*

$$\begin{aligned} X : \Omega &\to \mathbb{R}^n \\ w &\to X(w) = (X_1(w), X_2(w), ..., X_n(w))^t. \end{aligned}$$

où $X_1, ..., X_n$ sont des variables aléatoires discrètes.

Définition 1.26. ***Loi de probabilité conjointe.*** *On appelle loi conjointe du vecteur aléatoire $X = (X_1, ..., X_n)^t$ la fonction :*

$$\mathbb{P}_X : \begin{array}{c} X_1(\Omega) \times X_2(\Omega) \times ... \times X_n(\Omega) \to [0, 1] \\ (X_1, ..., X_n) \mapsto \mathbb{P}\left(X_1 = x_1, ..., X_n = x_n\right) \end{array}$$

où $[X_1 = x_1, ..., X_n = x_n] = [X_1 = x_1] \cap ... \cap [X_n = x_n]$.

Définition 1.27. *Soit X un vecteur aléatoire à valeurs discrètes. Toute fonction discrète est une loi de probabilité conjointe à condition que les deux propriétés suivantes soient vérifiées.*

- $\mathbb{P}_X(x) \geq 0, \forall x \in \mathbb{R}^n$.
- $\sum\limits_x \mathbb{P}_X(x) = \sum\limits_{x_1} ... \sum\limits_{x_n} \mathbb{P}_X(x_1, ..., x_n) = 1$.

Définition 1.28. ***Probabilité d'un événement (cas discret).*** *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace prbabilisable et $X = (X_1, X_2, ..., X_n)^t$ un vecteur aléatoire à valeurs discrètes de loi $\mathbb{P}_X$. La probabilité d'un événement $B \in \mathcal{A}$ est donnée par :*

$$\mathbb{P}(B) = \sum_{x \in B} \mathbb{P}_X(x).$$

Définition 1.29. ***Vecteur aléatoire absolument continu.*** Soit $X = (X_1, X_2, ..., X_n)^t$ un vecteur aléatoire à valeurs dans $\mathbb{R}^n$. On dit que X est absolument continu s'il existe une fonction mesurable positive ou nulle f_X, appelée densité , telle que :

$$\int_{\mathbb{R}^n} f_X(x)dx = \int_{\mathbb{R}} ... \int_{\mathbb{R}} f_X(x_1, ..., x_n)dx_1...dx_n = 1 \tag{1.27}$$

et

$$F_X(x) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_n} f_X(t_1, ..., t_n)dt_1....dt_n \tag{1.28}$$

Pour tout $(x_1, x_2, ..., x_n)^t \in \mathbb{R}^n$.

Définition 1.30. ***Probabilité d'un événement (cas continu).*** *Soit $X = (X_1, X_2, ..., X_n)^t$ un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ de densité f_X. La probabilité d'un événement $B \in \mathfrak{B}(\mathbb{R}^n)$ est donnée par :*

$$\mathbb{P}(B) = \int_B f_X(x)dx$$

Exemple 1.4.1. *Soit $X = (X_1, X_2)^t$ un vecteur aléatoire dont la densité est donnée par*

$$f_X(x_1, x_2) = \begin{cases} \frac{1}{2\sqrt{x_1x_2}} & si \;\; 0 < x_1 \leq x_2 < 1 \\ 0 & sinon. \end{cases}$$

On note $A =]-2, \frac{1}{4}[\times]-1, \frac{3}{2}[$. Déterminer $\mathbb{P}(X \in A)$.

$$\mathbb{P}(X \in A) = \int_0^{\frac{1}{4}} \int_{x_1}^1 \frac{1}{2\sqrt{x_1x_2}} dx_2dx_1 = \frac{3}{4}.$$

Proposition 1.4.3. Si le vecteur aléatoire X est absolument continu de fonction de répartition F_X, la densité de X est donnée par :

$$f_X(x_1, ..., x_n) = \frac{\partial^n F_X(x_1, ..., x_n)}{\partial x_1 \partial x_2 ... \partial x_n}.$$

1.5 Distributions marginales

Dans le cas d'un vecteur aléatoire comprenant n variables $(n > 1)$, on peut s'intéresser à la distribution conjointe d'un sous ensemble de m variables parmi les n variables de départ. Pour faciliter les écritures, définissons une partition du vecteur X, avec

$$X = (X_1, ...X_m, X_{m+1}, ..., X_n)^t = (X_a, X_b)^t,$$

où $X_a = (X_1, ...X_m)^t$ et $X_b = (X_{m+1}, ..., X_n)^t$ sont deux morceaux du vecteur aléatoire X, repectivement de dimension $n_a = m \; et \; n_b = n - m$.

Définition 1.31. ***Lois marginales d'un vecteur aléatoire discret.*** Soient deux vecteurs aléatoires $X_a = (X_1, ..., X_m)^t$ et $X_b = (X_{m+1}, ..., X_n)^t$ définis sur le même espace probabilisable $(\Omega, \mathcal{A})$. Si le vecteur $X = (X_1, X_2, ..., X_m, X_{m+1}, ...X_n)^t$ de dimension $m + n_b = n$ est un vecteur aléatoire discret de loi de probabilité $\mathbb{P}_X$, les vecteurs X_a et X_b sont des vecteurs aléatoires discrets et de loi de probabilité respectives :

$$\mathbb{P}_{X_a}(x_a) = \sum_{x_{m+1}} ... \sum_{x_n} \mathbb{P}_X(x) = \sum_{x_{m+1}} ... \sum_{x_n} \mathbb{P}_X(x_1, ..., x_n).$$

$$\mathbb{P}_{X_b}(x_b) = \sum_{x_1} ... \sum_{x_m} \mathbb{P}_X(x) = \sum_{x_1} ... \sum_{x_m} \mathbb{P}_X(x_1, ..., x_n).$$

avec $x_a = (x_1, ..., x_m)^t$ et $x_b = (x_{m+1}, ..., x_n)^t$.

Définition 1.32. ***Densités marginales d'un vecteur aléatoire absolument continu.***
Soient deux vecteurs aléatoires $X_a = (X_1, ..., X_m)^t$ et $X_b = (X_{m+1}, ..., X_n)^t$ définis sur le même espace probabilisable $(\Omega, \mathcal{A})$. Si le vecteur aléatoire $X = (X_1, X_2, ..., X_m, X_{m+1}, ...X_n)^t$ de dimension $m + n_b = n$ est absolument continu de densité de probabilité f_X, les vecteurs X_a et X_b sont absolument continus et de densité de probabilité respectives :

$$f_{X_a}(x_a) = \int_{\mathbb{R}^{n-m}} f_X(x)dx_b = \int_{\mathbb{R}^{n-m}} f_X(x_1, ..., x_n)dx_{m+1}...dx_n, \tag{1.29}$$

$$f_{X_b}(x_b) = \int_{\mathbb{R}^{m}} f_X(x)dx_a = \int_{\mathbb{R}^{m}} f_X(x_1, ..., x_n)dx_1...dx_m, \tag{1.30}$$

avec $x_a = (x_1, ..., x_m)^t$ *et* $x_b = (x_{m+1}, ..., x_n)^t$.

Distributions conditionnelles

Théorème 1.5.1. *Soit $X = (X_a, X_b)^t$ un vecteur aléatoire (discret ou continu), la fonction de (densité de) probabilité conditionnelle de X_a est donnée par :*

$$\mathbb{P}_{X_a/X_b=x_b}(x_a) = \frac{\mathbb{P}_X(x)}{\mathbb{P}_{X_b}(x_b)}; \quad f_{X_a/X_b=x_b}(x_a) = \frac{f_X(x)}{f_{X_b}(x_b)}, x = (x_1, ..., x_n).$$

Les fonctions $\mathbb{P}_{X_b/X_a}$ et f_{X_b/X_a} s'obtiennent de manière similaire.

Théorème 1.5.2. ***Théorème des probabilités totales.*** *Soit $X = (X_a, X_b)^t$ un vecteur aléatoire, alors*

$$\mathbb{P}_{X_a}(x_a) = \sum_{x_b} \mathbb{P}_{X_a/X_b=x_b}(x_a)\mathbb{P}(x_b); \quad f_{X_a}(x_a) = \int_{\mathbb{R}^{n-m}} f_{X_a/X_b=x_b}(x_a)f(x_b)dx_b.$$

Théorème 1.5.3. ***Théorème des probabilités composées.*** Soit $X = (X_a, X_b)^t$ un vecteur aléatoire (discret ou continu), alors

$$\mathbb{P}(x_1, ..., x_n) = \mathbb{P}(x_1)\mathbb{P}_{X_2/X_1=x_1}\mathbb{P}_{X_3/(x_1,x_2)}...\mathbb{P}_{X_n/(x_1,...,x_{n-1})}.$$

$$f_X(x_1, ..., x_n) = f(x_1)f_{X_2/X_1=x_1}f_{X_3/(x_1,x_2)}...f_{X_n/(x_1,...,x_{n-1})}.$$

Théorème 1.5.4. ***Théorème de Bayes .*** Soit $X = (X_a, X_b)^t$ un vecteur aléatoire de $(\mathbb{R}^n, \mathfrak{B}(R^n))$, on peut écrire que :

$$\mathbb{P}_{X_a}(x_a) = \frac{\mathbb{P}_{X_a/X_b=x_b}\mathbb{P}(x_b)}{\sum_{x_b}\mathbb{P}_{X_a/X_b=x_b}\mathbb{P}_{X_b}(x_b)}; f_{X_a}(x_a) = \frac{f_{X_a/X_b=x_b}(x_a)f_{X_b}(x_b)}{\int_{\mathbb{R}^{n-m}} f_{X_a/X_b=x_b}(x_a)f_{X_b}(x_b)dx_b}.$$

1.5.1 Indépendance

Définition 1.33. Soient $X = (X_1, ..., X_m)^t$ un vecteur aléatoire de dimension m et $Y = (Y_1, ..., Y_n)^t$, un vecteur aléatoire absolument continu. On dit que X et Y sont indépendants si pour tous les ensembles mesurables $A \in \mathfrak{B}_{\mathbb{R}^m}$ et $B \in \mathfrak{B}_{\mathbb{R}^n}$,

$$\mathbb{P}\left[X \in A, Y \in B\right] = \mathbb{P}\left[X \in A\right]\mathbb{P}\left[Y \in B\right].$$

Avec les notations de la définition précédente, soit le vecteur $X = (x_1, ..., x_m, x_{m+1}, ..., x_n)^t$. L'indépendance de X_a et X_b implique alors que :

$$F_X(x_a, x_b) = F_{X_a}(x_a)F_{X_b}(x_b). \tag{1.31}$$

Propriété 1.5.5. Pour tous les vecteurs $x \in \mathbb{R}^m$ et $y \in \mathbb{R}^n$. Nous retrouvons donc le même type de définition que celle donnée pour l'indépendance de deux variables aléatoires. Et comme pour les variables aléatoires, la réciproque est encore vraie ; si (1.32) est vraie, alors les vecteurs aléatoires X_a et X_b sont indépendants.

Définition 1.34. ***Indépendance mutuelle***. Soients $Y_1, ..., Y_l$ des vecteurs aléatoires de dimensions respectives $n_1, ..., n_l$ où $n_1, ..., n_l$ sont l enties naturels non nuls quelconques. On dit que ces vecteurs aléatoires sont mutuellement indépendants si, pour tous les ensembles mesurables $B_1 \in \mathfrak{B}(\mathbb{R}^{n_1}), ..., B_l \in \mathfrak{B}(\mathbb{R}^{n_l})$, on a :

$$\mathbb{P}\left[Y_1 \in B_1, ..., Y_l \in B_l\right] = \prod_{i}^{l}\mathbb{P}\left[Y_i \in B_i\right].$$

Définition 1.35. *Soient $X = (X_1, ..., X_m)^t$ un vecteur aléatoire discret de dimension m et $Y = (Y_1, ..., Y_n)^t$, un vecteur aléatoire discret de dimension n. On dit que X et Y sont indépendants si pour tous les ensembles mesurables A et B*

$$\mathbb{P}\left[X \in A, Y \in B\right] = \mathbb{P}\left[X \in A\right]\mathbb{P}\left[Y \in B\right].$$

Avec les notations de la définition précédente, soit le vecteur $X = (x_1, ..., x_m, x_{m+1}, ..., x_n)^t$. L'indépendance de X_a et X_b implique alors que :

$$\mathbb{P}_X\left(X_a = x_a, X_b = x_b\right) = \mathbb{P}\left(X_a = x_a\right)\mathbb{P}\left(X_b = x_b\right). \tag{1.32}$$

Théorème 1.5.6. ***Indépendance mutuelle de n variables aléatoires [3]***. *Soit* $X = (X_1, ..., X_n)^t$ un vecteur aléatoire (discret ou continu). Les variables aléatoires $X_1, X_2, ..., X_n$ seront dites mutuellement indépendantes (ce que l'on note aussi $X_1 \perp X_2 \perp ... \perp X_n$) si et seulement si

$$\mathrm{X}_1 \perp X_2 \perp ... \perp X_n \Leftrightarrow \begin{cases} \mathrm{P}_X(x_1, ..., x_n) = \mathbb{P}_{X_1}(x_1)\mathbb{P}_{X_2}(x_2)...\mathbb{P}_{X_n}(x_n) & \text{cas discret} \\ \mathrm{f}_X(x_1, ..., x_n) = f_{X_1}(x_1)f_{X_2}(x_2)...f_{X_n}(x_n) & \text{cas continu.} \end{cases}$$

1.5.2 Loi multinomiale

Exemple 1.5.1. On rappelle que la loi binomiale permet de modéliser le nombre de succès d'une expérience à 2 issues (succès ou échec) répétée n fois de façon indépendantes. La loi multinomiale généralise la loi binomiale. Plus précisément, suite à n répétitions indépendantes d'une expérience à k issues de probabilité respectives p_k, la loi multinomiale permet de modéliser le nombre de réalisations X_i de l'issue i. Supposons que l'on répète un nombre n fixé de fois une épreuve à l'issue de laquelle un et un seul des événements $A_1, ..., A_k$ est réalisé, ces répétitions étant indépendantes les unes des autres et les probabilités des A_i étant identiques d'une répétition à l'autre. Si l'on définit le vecteur $X = (X_1, ..., X_k)^t$ où X_i est le nombre de fois que l'événement A_i s'est réalisé, alors X suit une loi multinomiale de paramètre n et $p = (p_1, ..., p_k)^t$, et on note :

$$X \sim \mathcal{M}(n, p)$$

avec $p_i = \mathbb{P}(A_i) = \mathbb{P}[X_i \in A_i]$.

Définition 1.36. On dit que le vecteur $X = (X_1, ..., X_k)^t$ suit une loi multinomiale de paramètre $(n, p_1, ..., p_k)$ si pour $x = (x_1, ..., x_k)^t \in \mathbb{N}^k$ avec $\sum_{i=1}^{k} x_i = n$, on a

$$\mathbb{P}\left(X_1 = x_1, ..., X_k = x_k\right) = \frac{n!}{x_1!...x_k!} p_1^{x_1} ... p_k^{x_k}.$$

Exemple 1.5.2.

a. On effectue un sondage en choisissant au hasard n individus (avec remise) dans une population partagée en k catégories et on note X_i le nombre d'individus sélectionné dans la ième catégorie. Le vecteur $X = (X_1, ..., X_k)^t$ suit une loi multinomiale.

b. On retrouve la loi multinomiale dans diverses parties de la statistique telles que la régression logistique multiclasse ou le test du Chi-Deux.

c. On s'interesse à la couleur des yeux de n individus. Sur la base d'un très grand nombre d'individus, on a pu établir de façon très fiable la probabilité de chacune des couleurs possibles, à savoir $A_1 \equiv \ \ bruns$, $A_2 \equiv \ \ bleu$, $A_3 \equiv \ \ noisette$ et $A_4 \equiv \ \ vert$, avec $\mathbb{P}(A_1) = 0.37$, $\mathbb{P}(A_2) = 0.36$, $\mathbb{P}(A_1) = 0.16$ et $\mathbb{P}(A_1) = 0.11$, la variable couleur des yeux étant donc une variables catégorique nominale. Si l'on examine n individus pris au hasard dans cette population supposée suffisamment grande et que X_i désigne le nombre d'individus appartenant à la catégorie i, alors le vecteur $X = (X_1, X_2, X_3, X_4)^t$ suit une loi multinomiale.

Remarque 1.5.1. La loi multinomiale s'intéresse donc à une seule variable catégorique, mais il n' y aucune difficulté à prendre en compte plusieurs variables de manière simultanée ; il suffit de croiser entre elles les catégories de chacunes de ces variables. Le plus souvent, les probabilités de ces catégories croisées sont alors indicées en fonction de ces variables et présentées dans un tableau multidimensionnel.

Exemple 1.5.3. Supposons que l'on ait en plus déterminer la couleur des cheveux pour ces mêmes individus. On dispose donc des deux variables catégoriques nominales "couleur des yeux " et "couleur des cheveux", avec pour les cheveux $B_1 \equiv "noir"$, $B_2 \equiv "brun"$,$B_1 \equiv "roux"$ et $B_4 \equiv "blanc"$. Les probabilités $\mathbb{P}(A_i \cap B_j)$ des différentes combinaisons possibles sont données dans le tableau suivant :

i$/j$	1	2	3	4	P(A$_i$)
1	0.12	0.20	0.04	0.01	0.37
2	0.03	0.14	0.03	0.16	0.36
3	0.03	0.09	0.02	0.02	0.16
4	0.01	0.05	0.02	0.03	0.11
P(B$_j$)	0.19	0.48	0.11	0.22	1

On peut toujours utiliser le formalisme de la loi multinomiale en considérant que chacune des 16 catégories de cette loi est obtenue en croisant une des catégories des deux variables, avec $p_{ij} = \mathbb{P}(A_i \cap B_j)$ où $i, j = 1, ..., 4$.

1.6 Exercices du chapitre 1

Exercice 1.6.1. Soit le vecteur aléatoire $X = (X_1, X_2, X_3)^t$ de densité conjointe

$$f_X(x_1, x_2, x_3) = \begin{cases} 6x_1x_2^2x_3 & si \;\; 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_3 \leq 1. \\ 0 & ailleurs. \end{cases}$$

1. *Montrez que f_X est une densité de probabilité.*
2. *Trouvez les densités marginales de X_1, X_2 et X_3 de même que la densité conjointe de (X_1, X_2).*
3. *Les variables X_1, X_2 et X_3 sont- elles indépendantes ?*
4. *Calculer les densités conditionnelles*
   (i) $f_{(X_1,X_2)/X_3}$
   (ii) $f_{X_1/(X_2,X_3)}$.

Exercice 1.6.2. Soit $X = (X_1, X_2)$ un couple de variable aléatoire de densité de probabilite $f_{(X_1,X_2)}$ donnée par :

$$f_X(x_1, x_2) = k \exp\left(\frac{2x_1^2 + x_2^2 - 2x_1x_2}{2}\right), (x_1, x_2) \in \mathbb{R}^2.$$

1. Déterminer k pour que f soit une densité de probabilité.
2. Déterminer les lois marginales de X_1 et de X_2.
3. Déterminer la densité conditionnelle de X_2 sachant $[X_1 = x_1]$.
4. Les variables aléatoires X_1 et X_2 sont -elles indépendantes ?

Exercice 1.6.3. Soit $(X_1, X_2)^t$ un vecteur aléatoire suivant une loi multinomiale $\mathcal{M}(n, \frac{1}{3}, \frac{2}{3})$
1. Calculer $\mathbb{P}(X_1 = 1)$ et $\mathbb{P}(X_2 = 0)$
2. Les variables X_1 et X_2 sont-elles indépendantes ?

Exercice 1.6.4. Soient X_1 et X_2 des variables aléatoires discrètes de loi : $\forall (i, j) \in \mathbb{N}$

$$p_{ij} = \mathbb{P}((X_1 = i) \cap (X_2 = j)) = \frac{2}{k(k+1)}, k \geq 1, 1 \leq i \leq k, 1 \leq j \leq i.$$

1. Déterminer les lois marginales de X_1 et de X_2.
2. Déterminer la loi conditionnelle de X_1 sachant $\{X_2 = j\}$.
3. Déterminer la loi conditionnelle de X_2 sachant $\{X_1 = i\}$.
4. Les variables aléatoires X_1 et X_2 sont -elles indépendantes ?

Exercice 1.6.5. Soient X_1 et X_2 des variables aléatoires réelles à valeurs dans N telles que $\forall(i,j) \in \mathbb{N}$

$$p_{ij} = \mathbb{P}((X_1 = i) \cap (X_2 = j)) = \frac{\alpha}{2^{i+j}}.$$

1. Déterminer la valeur de α.
2. Montrer que X_1 et X_2 sont indépendantes.

Exercice 1.6.6. On considère deux variables aléatoires indépendantes X_1 et X_2 de loi exponentielle de paramètre $(\lambda_1 > 0, \lambda_2 > 0)$ respectivement.

– Quelle est la densité du vecteur aléatoire $(X_1, X_2)^t$?
– Soit $Y = \min(X_1, X_2)$ le minimum de ces deux variables aléatoires. Montrer que

$$\mathbb{P}(Y = X_1) = \mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

– Calculer la fonction de répartition de Y et en déduire que Y suit une loi exponentielle de paramètre $\lambda_1 + \lambda_2$.

Exercice 1.6.7. Soit $X = (X_1, X_2)^t$ un vecteur aléatoire de loi uniforme sur l'ensemble $\{]0,1[\times]0,1[\cup]1,2[\times]0,2[\}$.

– Représenter le support de X sur un graphique.
– Quelle est la densité de X ?
– Les variables X_1 et X_2 sont elles indépendantes ?

Exercice 1.6.8. Soit $(X_1, X_2)^t$ un vecteur aléatoire de loi uniforme sur le triangle

$$\mathcal{T} = \{(x_1, x_2) \in \mathbb{R}^2 \ \ tel \ \ que \ \ 0 < x_1 < x_2 < 1\}.$$

1. Représenter le support de X sur un graphique.
2. Calculer les densités marginales.
3. Les variables X_1 et X_2 sont-elles indépendantes ?
4. Calculer l'espérance X_1.

Exercice 1.6.9. *Soient X_1 et X_2 deux variables aléatoires indépendantes à valeurs dans $\{-1, 0, 1\}$. On suppose que la loi de X_1 est donnée par :*

$$\mathbb{P}(X_1 = -1) = p_1, \mathbb{P}(X_1 = 1) = p_2, p_i \in [0, \frac{1}{2}].$$

$$\mathbb{P}(X_1 = 0) = 1 - p_1 - p_2.$$

La variable aléatoire X_2 est uniforme sur $\{-1, 0, 1\}$ i.e.

$$\mathbb{P}(X_2 = -1) = \mathbb{P}(X_2 = 0) = \mathbb{P}(X_2 = 1) = \frac{1}{3}.$$

1. *Déterminer la loi de $S = X_1 + X_2$.*

2. *On définit la variable aléatoire Z par*

$$Z = \begin{cases} \theta_1 & \textit{si} \quad X_1 + X_2 = 0 \\ \theta_2 & \quad \textit{sinon.} \end{cases}$$

où θ_1 et θ_2 sont des réels distincts fixés.

a. *Montrer que la loi de Z ne dépend pas de p_1 et p_2.*

b. *Déterminer les valeurs de θ_1 et θ_1 de sorte que $E(Z) = 3$ et $V(Z) = 2$.*

Exercice 1.6.10. On pose

$$h(x) = \frac{1}{\Gamma(\alpha+1)} \exp(-x) x^{\alpha-1}, \alpha > 0$$

et

$$D = \{(x_1, x_2) \in \mathbb{R}^2, 0 < x_2 < x_1\}.$$

Soit $f(x_1, x_2) = h(x_1) 1_D(x_1, x_2)$.

1. Montrer que f est une densite de probabilité sur $\mathbb{R}^2$. On considere dans la suite un couple (X_1, X_2) de v.a.r. de densité f.
2. Les v.a. X_1 et X_2/X_1 sont-elles independantes ?
3. Quelle est la loi conditionnelle de X_2 sachant X_1 ?

Exercice 1.6.11. On note N la variable aléatoire comptant le nombre d'oeufs qu'un insecte donné pond. On suppose que N suit une loi de Poisson de paramètre $\lambda \in \mathbb{R}^+$. On suppose également que chaque oeuf donne naissance à un nouvel insecte avec probabilité p, indépendamment de l'éclosion des autres oeufs. On considère alors une famille $(X_i)_i$ de variables aléatoires de loi de Bernoulli de paramètre $p \in [0, 1]$. On suppose que les variables aléatoires $(N, X_1, X_2, ...)$ sont indépendantes et on note D le nombre de descendants de l'insecte.

1. Ecrire D en fonction des variables aléatoires N et X_i
2. Pour tout $(n, d) \in \mathbb{N}^2$, calculer $\mathbb{P}(D = d/N = n)$.
3. En déduire la loi de D et la loi de la variable aléatoire de $\mathbb{N}^2 Z$.

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

$$Z \text{ admet une espérance } \iff \int_{\mathbb{R}^2} \mid \varphi(x, y) \mid f(x, y) dx dy \text{ existe.}$$

De plus en cas d'existence :

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

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

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

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 $Cov(X_1, X_2) = 0 \not\Rightarrow X_1 \perp X_2$

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

(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 $Z = X_1 X_2$

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 :*

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

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 $Z = X_1 X_2$, voir tableau 2.3.

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

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

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 / 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 $\| \cdot \|$ 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

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

TABLE 2.6 – loi de $Z = X_1X_2$

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

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

$$V^t \sum_X V = Var(v_1X_1 + ... + v_dX_n) \geq 0.$$

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

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

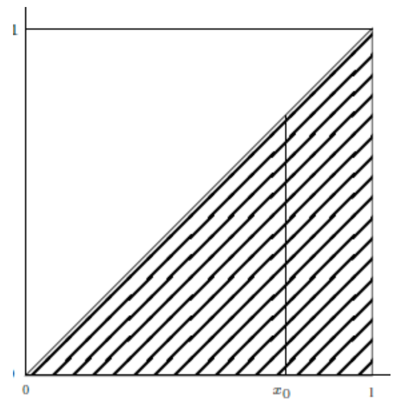

FIGURE 2.1 – Densité conditionnelle de $Y/X = x$.

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 :

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

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

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 :
   $$f_Y(y) = (f_X(-\sqrt{y}) + f_X(\sqrt{y})) \frac{1}{2\sqrt{y}}. \qquad (2.2)$$

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

$$Y_1 = X_1 X_2, Y_2 = X_1 X_3 \;\; et \;\;\; Y_3 = X_2 X_3.$$

On a donc :

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

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 $\mid det(J_{y \to x}) \mid = \mid -2x_1 x_2 x_3 \mid$. Puisque l'on a $x_1 = \sqrt{\frac{y_1 y_2}{y_3}}$, $x_2 = \sqrt{\frac{y_1 y_3}{y_2}}$ *et* $x_3 = \sqrt{\frac{y_2 y_3}{y_1}}$. *On a* $\mid det(J_{x \to y}) \mid = 2\sqrt{y_1 y_2 y_3}$. *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 :

$$f_Y(y) = \frac{1}{2\sqrt{y_1 y_2 y_3}}$$

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

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

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,

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

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 $\sigma_{Y_i Y_j}$ 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 :

$$Y = AX + b \Rightarrow \left\{ \begin{array}{l} \mu_Y = A_{\mu_X} + b \\ \sum_Y = A \sum_X A^t \end{array} \right.$$

2.4 Exercices du chapitre 2.

Exercice 2.4.1. Soit (X, Y) un couple aléatoire de densité

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \mid x \mid e^{\frac{-x^2(1+y^2)}{2}}, (x, y) \in \mathbb{R}^2.$$

1. Calculer la loi de X et la loi de Y. Les deux variables sont-elles indépendantes ?
2. Est-ce que X et XY sont indépendantes ? Calculer la loi de XY.
3. Quelle est la loi de $X(1 + Y)$?.

Exercice 2.4.2. Soit (X, Y) un couple aléatoire dont la loi adment la densité

$$f(x, y) = c(y^2 - x^2) \exp[-y] 1_{\{-y \leq x \leq y, y>0\}}(x, y)$$

où c est une constante.

1. Déterminer la valeur de la constante c.
2. X et Y sont- elles indépendantes ?
3. Pour $k \in \mathbb{N}$, calculer $\mathbb{E}(X^k Y^k]$. En déduire $Cov(X, Y)$. Que constate-t-on ?
4. On pose $S = X + Y$ et $U = Y - X$
   a. Déterminer la loi du couple (S, U)
   b. S, U sont -elles indépendantes ?

Exercice 2.4.3. Soient X et Y deux v.a.r. On suppose que la densité conditionnelle de X sachant $Y = y$ est la densité $1_{\mathbb{R}^+}(x) y^2 x e^{-xy}$ et que la loi de Y est de densité $\frac{1}{y^2} 1_{[1,+1[}(y)$. On pose $T = XY$.

1. Trouver la loi du couple (T, Y). Qu'en déduit-on ?
2. Trouver la loi conditionnelle de Y sachant $X = x$
3. Calculer $\mathbb{E}(Y \diagup X)$.

Exercice 2.4.4. Soit $X_1, ..., X_n$ des variables aléatoires indépendantes suivant la même loi de probabilité $f(x) = e^{-x} 1_{]0,\infty[}(x)$.
Pour $i = 1, 2, ..., n$, on pose : $S_i = \sum_{j=1}^{i} X_j$.

1. Déterminer la loi du couple (S_1, S_2).
2. S_1 et S_2 sont -elles indépendantes ?
3. Déterminer F_{S_1,S_2} la fonction de répartition du couple (S_1, S_2). En déduire F_{S_1} et F_{S_2}, les fonctions de répartition marginales de S_1 et S_2.
4. Déterminer la loi de $(S_1, ..., S_n)$. En déduire la densité de la loi marginale de S_n.
5. Pour $m \neq k$, $(1 \leq m \leq n, 1 \leq k \leq n)$, calculer $\rho(m, k)$ le coefficient de corrélation entre S_m et S_k.

Exercice 2.4.5. Soit X et Y deux variables aléatoires. On suppose que :

- La loi de Y admet la densité : $f_Y(y) = \frac{\alpha\beta^\alpha}{y^{\alpha+1}} 1_{]\beta,\infty[}(y)(\alpha > 0, \beta > 0)$.
- Conditionnellement à Y, X suit une loi uniforme sur $]-y, y[$.

1. Déterminer la densité de probabilité du couple (X, Y).
2. Déterminer la densité de la loi Y sachant $\{X = x\}$.
3. Calculer $\mathbb{E}\left[Y \diagup X = x\right]$.

Exercice 2.4.6. Soit X, Y et Z trois variables aléatoires réelles indépendantes suivant la même loi uniforme sur l'intervalle $]0, 1[$.
On pose :$U = XV$ et $V = Z^2 - XV$.

- Déterminer la loi de U.
- Déterminer la loi conditionnelle de V sachant $\{U = u\}$. En déduire la loi de V.

Chapitre 3

Fonctions caractéristiques

3.1 Transformé de Fourier d'une mesure bornée

Définition 3.1. *Soient $x = (x_1, ..., x_n)^t$ et $y = (y_1, ..., y_n)^t$ deux éléments de $\mathbb{R}^n$. Le produit scalaire usuel de x et y est donné par :*

$$< x, y >= x_1 y_1 + ... + x_n y_n.$$

Définition 3.2. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et les variables aléatoires considérées sont supposées définies sur cet espace. $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ est un espace mesurable où $\mathbb{R}^n$ est muni de la tribu de Borel $\mathfrak{B}(\mathbb{R}^n)$. Une variable aléatoire X à valeurs complexes est une application mesurable de la forme*

$$X = X_1 + iX_2$$

pour deux variables aléatoires réelles X_1 et X_2 appelées partie réelle et partie imaginaire de X respectivement.

Définition 3.3. *On dit qu'une variable aléatoire complexe X est intégrable si les v.a.r. X_1 et X_2 sont intégrables. Dans ce cas l'espérance de X est alors définie par :*

$$\mathbb{E}(X) = \mathbb{E}(X_1) + i\mathbb{E}(X_2).$$

Définition 3.4. ***Transformée de Fourier d'une mesure bornée.*** *Soit μ une mesure bornée sur $\mathbb{R}^n$ muni de sa mesure borélienne μ. On appelle transformée de Fourier de μ l'application $\hat{\mu}$ de $\mathbb{R}^n$ dans $\mathbb{C}$ définie par :*

$$\forall t \in \mathbb{R}^n : \hat{\mu}(t) = \int_{\mathbb{R}^n} e^{i<x,t>} \mu(dx).$$

Remarque 3.1.1. *La transformée de Fourier d'une mesure bornée est bien définie étant donné que*

$$\forall t \in \mathbb{R}^n : \int_{\mathbb{R}^n} | e^{i<x,t>} | \mu(dx) \leq \mu(\mathbb{R}^n) < \infty.$$

De plus, si la mesure bornée μ admet une densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$, alors

$$\forall t \in \mathbb{R}^n : \hat{\mu}(t) = \int_{\mathbb{R}^n} e^{i<x,t>} f(x)dx.$$

On définit alors la fonction caractéristique d'une variable aléatoire à partir de la transformée de Fourier de sa densité.

3.2 Fonctions caractéristiques d'un vecteur aléatoire

Définition 3.5. ***Fonction caractéristique.*** *Soit X un vecteur aléatoire de $\mathbb{R}^n$. On appelle fonction caractéristique de X et on note ϕ_X la transformée de Fourier de sa loi $\mathbb{P}_X$ c'est-à- dire la fonction vectorielle et à valeurs complexes définie par :*

$$\phi_X : \quad \begin{array}{c} \mathbb{R}^n \to \mathbb{C} \\ t \to \int_{\mathbb{R}^n} e^{i<x,t>} d\mathbb{P}_X. \end{array} .$$

La loi d'une variable aléatoire étant par construction une mesure de probabilité, ϕ_X existe pour toute variable aléatoire X. De plus, d'après le théorème de transfert $\phi_X(t) = \mathbb{E}\left[e^{i<X,t>}\right] = \mathbb{E}\left[e^{i\sum_{j=1}^n t_j x_j}\right] = \int_{-\infty}^{\infty} e^{it'x} f_X(x)dx$.

Définition 3.6. *La fonction caractéristique d'un vecteur aléatoire discret X de loi P_X est donnée par :*

$$\phi_X(t) = \mathbb{E}\left[e^{i<X,t>}\right] = \sum_j e^{it'x} \mathbb{P}_X(X_j = x_j).$$

Exemple 3.2.1. *Soit X une v.a. unidimensionnelle suivant une loi géométrique $G(p)$ avec $p \in]0,1[$. Calculons sa fonction caractéristique.*

$$\begin{aligned} \forall t \in \mathbb{R} : \phi_X(t) &= \sum_{k=1}^{\infty} p(1-p)^{k-1} e^{itk} \\ &= \frac{p}{1-p} \sum_{k=1}^{\infty} \left((1-p)e^{it}\right)^k \\ &= \frac{p}{1-p} . \frac{(1-p)e^{it}}{1-(1-p)e^{it}} = \frac{pe^{it}}{1-(1-p)e^{it}}. \end{aligned}$$

Exemple 3.2.2. *Soit X une variable aléatoire suivant une loi exponentielle $Exp(\lambda)$ de paramètre $\lambda \in \mathbb{R}_+^*$.*

$$\begin{aligned} \forall t \in \mathbb{R} : \phi_X(t) &= \int_0^{\infty} \lambda e^{-\lambda x} e^{itx} dx \\ &= \frac{\lambda}{\lambda - it}. \end{aligned}$$

Théorème 3.2.1. *Soit $X = (X_1, X_2, ..., X_n)^t$ un vecteur aléatoire à valeurs dans $\mathbb{R}^n$. La fonction caractéristique ϕ_X de X vérifie les propriétès suivantes :*

(i) $\forall u \in \mathbb{R}^n$: $\mid \phi_X(u) \mid \leq \phi_X(0) = 1$ *où* 0 *est le vecteur nul de* $\mathbb{R}^n$ *;*

(ii) $\forall j \in \{0, 1, ..., n\}$ *et* $\forall u_j \in \mathbb{R} : \phi_{X_j}(u_j) = \phi_X(0, ..., u_j, ..., 0)$

(iii) $\forall u \in \mathbb{R}^n$: $\overline{\phi_X(u)} = \phi_X(-u)$.

(iv) *Si on pose $Y = AX + b$ où A est une matrice $p \times n$ et b un vecteur de $\mathbb{R}^p$ on a, pour tout u dans $\mathbb{R}^p$:*

$$\phi_Y(u) = e^{i<u,b>}\phi_X(A^t u),$$

où A^t est la matrice transposée de A.

Démonstration. **(i)** *On a*

$$\begin{aligned} \mid \phi_X(u) \mid &= \mid \int e^{i<u,x>} d\mathbb{P}_X \mid \leq \int \mid e^{i<u,x>} \mid dP_X \\ &= \int dP_X(x) = 1 \end{aligned}$$

et

$$\phi_X(0) = 1.$$

(ii) *Par définition, on a :*

$$\phi_X(0) = (0, ..., u_j, ..., 0) = \mathbb{E}(e^{iu_j X_j}) = \phi_{X_j}(u_j).$$

(iii) *On peut écrire :*

$$\begin{aligned} \phi_X(-u) &= \mathbb{E}\left(e^{i<-u,X>}\right) = \mathbb{E}\left(e^{-i<u,X>}\right) \\ &= \mathbb{E}\left(\overline{e^{i<u,X>}}\right) = \overline{\mathbb{E}\left(e^{-i<u,X>}\right)} = \overline{\phi_X(u)}. \end{aligned}$$

(iv) *On a*

$$\begin{aligned} \phi_Y(u) &= \mathbb{E}\left(e^{i<u,Y>}\right) = \mathbb{E}\left(e^{i<u,AX>+i<u,b>}\right) \\ &= e^{i<u,b>}\mathbb{E}\left(e^{i<A'u,X>}\right) = e^{i<u,b>}\phi_X(A'u). \end{aligned}$$

Théorème 3.2.2. *Une famille $(X_j)_{j=1,...,n}$ constitue une famille de v.a.r. indépendantes si et seulement si, pour tout $u = (u_1, ..., u_n) \in \mathbb{R}^n$, on a :*

$$\phi_{(X_1,...,X_n)}(u) = \prod_{j=1}^{n} \phi_{X_j}(u_j).$$

Démonstration. *On a*

$$\phi(u) = \mathbb{E}\left[e^{i\sum_{j=1}^{n} u_j X_j}\right] = \mathbb{E}\left[e^{iu_1X_1} \times ... \times e^{iu_nX_n}\right].$$

D'autre parts, si $X_1 \perp X_2 \perp ... \perp X_n$, alors on a $h_1(X_1) \perp h_2(X_2) \perp ... \perp h_n(X_n)$avec $h_i(X_i) = e^{it_jX_j}$, ce qui implique le résultat

$$\mathbb{E}\left[h_1(X_1)...h_n(X_n)\right] = \mathbb{E}\left[h_1(X_1)\right]...\mathbb{E}\left[h_n(X_n)\right].$$

On obtient donc

$$\mathbb{E}\left[e^{iu_1X_1}...e^{iu_nX_n}\right] = \mathbb{E}\left[e^{iu_1X_1}\right]...\mathbb{E}\left[e^{iu_nX_n}\right],$$

c'est-à-dire

$$\phi_{(X_1,...,X_n)}(u) = \prod_{j=1}^{n} \phi_{X_j}(u_j).$$

Exercice 3.2.3. *Quelle est l'expression de la fonction caractéristique pour $Y = X_1 + X_2 + ... + X_n$ si les variables $X_1, X_2, ..., X_n$ sont indépendantes et identiquement distribuées (iid) ?*

Solution. *On a :*

$$\phi_Y(u) = \mathbb{E}(e^{iuY}) = \mathbb{E}\left[e^{iu(X_1+X_2+...+X_n)}\right].$$

Puisque $X_1 \perp X_2 \perp ... \perp X_n$, on a $\mathbb{E}\left[e^{iuX_1} \times ... \times e^{iuX_n}\right] = \mathbb{E}\left[e^{iuX_1}\right]...\mathbb{E}\left[e^{iuX_n}\right]$, on peut donc écrire $\phi_Y(u) = \phi_{X_1}(u) \times ... \times \phi_{X_n}(u)$. Puisque les variables sont identiquement distribuées, leur fonction caractéristique et identique et l'onbtient

$$\phi_Y(u) = [\phi_{X_1}(u)]^n.$$

3.3 Moments et fonction caractéristique

Moments de variables aléatoires réelles

Définition 3.7. *(**Espace** L^p). Soit p un réel tel que $1 \leq p < \infty$. On dit qu'une variable aléatoire X appartient à l'espace $L^p(\Omega, \mathcal{A}, \mathbb{P})$ si elle est de puissance p-intégrable, i.e. si*

$$\mathbb{E} \mid X \mid^p < \infty.$$

La norme sur sur $L^p(\Omega, \mathcal{A}, \mathbb{P})$ est :

$$\| X \|_p = \mathbb{E}(\mid X \mid^p)^{\frac{1}{p}} = \left(\int \mid X \mid^p d\mathbb{P}_X\right)^{\frac{1}{p}}.$$

Définition 3.8. *Pour une variable aléatoire X dans $L^p(\Omega, \mathcal{A}, \mathbb{P})$, on appelle*

- *Moment d'ordre p le réel :* $\mathbb{E}X^p$.
- *Moment absolu d'ordre p le réel :* $\mathbb{E} \mid X \mid^p$.
- *Moment centré d'ordre p le réel :* $\mathbb{E}((X - \mathbb{E}X)^p)$.

Théorème 3.3.1. ***(Inégalité de Hölder)**. Pour tous réels p et q tels que $p > 1$, $q > 1$ et$1/p + 1/q = 1$ et toutes $v.a.X$ et Y respectivement dans $L^p(\Omega, \mathcal{A}, \mathbb{P})$ et $L^q(\Omega, \mathcal{A}, \mathbb{P})$, on a*

$$\| XY \| \leq \| X \|_p \| Y \|_q .$$

Théorème 3.3.2. ***(Inégalité de Minkowski)**. Pour tous réels p et q tels que $p > 1$, $q > 1$ et $1/p + 1/q = 1$ et toutes $v.a.X$ et Y respectivement dans $L^p(\Omega, \mathcal{A}, \mathbb{P})$ et $L^q(\Omega, \mathcal{A}, \mathbb{P})$, on a*

$$\| X + Y \| \leq \| X \|_p + \| Y \|_q .$$

Définition 3.9. ***(Espace** L^2) L'inégalité de Hölder, vue précédemment, appliquée pour $p = q = 2$ est souvent appelée inégalité de Schwartz.*

Proposition 3.3.3. *Soit X et Y deux v.a.r. dans L^2. On a*

$$\| X + Y \| \leq \| X \|_2 \| Y \|_2 .$$

Définition 3.10. *(**N**ormes usuelles sur $\mathbb{R}^n$)*
- *Pour tout $x = (x_1, ..., x_n) \in \mathbb{R}^n$, $p \in]0, \infty[$: $x \mapsto \| x \|_p = (\sum_{i=1}^n \mid x_i \mid^p)^{\frac{1}{p}}$.*
- *Pour tout $x = (x_1, ..., x_n) \in \mathbb{R}^n$, $p \in]0, \infty[$: $x \mapsto \| x \|_\infty = \sup_i \mid x_i \mid$.*
- *Pour tout $x = (x_1, ..., x_n) \in \mathbb{R}^n$, on note $\| x \|_1 = \mid x_1 \mid + ... + \mid x_n \mid$. Norme L^1*

Théorème 3.3.4. *Soit $X = (X_1, X_2, ..., X_n)^t$ un vecteur aléatoire. Si $\| X \|_k$ admet une espérance finie pour $k \in \mathbb{N}^*$, alors ϕ_X est k fois continûment différentiable et on a :*

$$\frac{\partial^k \phi_X}{\partial u_{j_1} ....\partial u_{j_k}}(u) = i^k \mathbb{E}[X_{j_1} X_{j_2} ... X_{j_k} e^{i<u,X>}], \tag{3.1}$$

pour tout $u = (u_1, ..., u_n)^t \in \mathbb{R}^n$ et tout entiers $j_1, ..., j_k$ distincts choisis dans $\{1, 2, ..., k\}$. En particulier, on a

$$\frac{\partial^k \phi_X}{\partial u_{j_1} ....\partial u_{j_k}}(0) = i^k \mathbb{E}[X_{j_1} X_{j_2} ... X_{j_k}]. \tag{3.2}$$

Application

Calcul des moments $\mathbb{E}(X_1 ... X_n)$ en fonction des dérivées de ϕ_X en 0. Par exemple, si X est à valeurs réelles et de carré intégrable, on a

$$\mathbb{E}(X) = i\phi'_X(0); \mathbb{E}(X^2) = -\phi''_X(0).$$

3.4 Fonction génératrice

Définition 3.11. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On appelle fonction génératrice de X et on note G_X la fonction définie sur le disque unité du plan complexe par :*

$$\forall \mid u \mid \leq 1, G_X(u) = \mathbb{E}(u^X) = \sum_n u^n \mathbb{P}(X = n).$$

On note que

$$G_X(1) = 1 \ \ et \ \ \phi_X(u) = G_X\left(e^{iu}\right).$$

Proposition 3.4.1.

1. *La fonction génératrice G_X est continue sur son domaine de définition et est de classe C^1 sur le domaine $(|u| < 1)$.*
2. *La fonction génératrice G_X caractérise la loi $\mathbb{P}_X$. Soit pour tout $n \in \mathbb{N}$*

$$\mathbb{P}(X = n) = \frac{1}{n!} G_X^{(n)}(0).$$

3. *La fonction génératrice G_X admet une dérivée à gauche en 1 $G'_X(1)$ si et seulement si $\mathbb{E}(X)$ existe et est fini. Dans ce cas*

$$\mathbb{E}(X) = G'_X(1).$$

4. *Plus généralement, la fonction génératrice G_X admet une dérivée à gauche d'ordre $r(r \in \mathbb{N}^*)$ en 1 ssi le moment factoriel d'ordre r existe et est fini. Dans ce cas :*

$$\mathbb{E}\left(X(X-1)...(X-r+1)\right) = G_X^{(r)}.$$

En particulier

$$\mathbb{E}\left(X(X-1)\right) = G''_X(1), V(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2.$$

5. *Soit $X_1, X_2, ..., X_n$ des v.a entières indépendantes. On pose*

$$S_n = \sum_{k=1}^{n} X_k.$$

Alors

$$\forall \mid u \mid \leq 1, G_{S_n}(u) = \prod_{k=1}^{n} G_{X_k}(u).$$

3.4.1 Fonction génératrice des moments

Définition 3.12. *Soit X une variable aléatoire réelle. On considère la fonction de variable réelle u*

$$g_X(u) = \mathbb{E}(e^{uX}), u \in \mathbb{R}.$$

Si g_X est définie dans un voisinage ouvert de l'origine, elle est appelée fonction génératrice des moments de X.

Proposition 3.4.2. *Soit X une v.a de fonction génératrice des moments g_X.*

1. *La fonction g_X est convexe et caractérise la loi de X.*
2. *Pour tous réels a et b, on a*
$$g_{aX+b}(u) = e^{bu} g_X(au).$$
En particulier si la loi de X est symétrique alors g_X est paire.
3. *Soit (X, Y) un couple de variables aléatoires indépendantes de fonctions génératrices des moments g_X et g_Y respectivement. Alors la somme $X + Y$ admet une fonction génératrice des moments et*
$$g_{X+Y} = g_X g_Y.$$

Théorème 3.4.3. *[9] Soit X une v.a de fonction génératrice des moments g_X. On suppose que g_X est définie sur un intervalle ouvert I centré en 0. Alors*

1. *Pour tout $r \geq 1$*
$$\mathbb{E}(X^r) < \infty \;\; g_X^{(r)}(0) = \mathbb{E}(X^r).$$
2. *Pour tout $u \in I$*
$$g_X(u) = 1 + \sum_{r \geq 1} \frac{u^r}{r!} \mathbb{E}(X^r).$$

3.5 Exercice du chapitre 3

Exercice 3.5.1.

1. On s'intéresse à un processus dont le nombre d'occurrences sur un intervalle de temps t suit une loi de poisson de paramètre $\mu = \lambda t$.
   Quelle sera la loi de probabilité donnant le temps séparant n occurrences successives ?

Exercice 3.5.2. Soit λ un réel srictement positif. Pour tout (i, j) de $\mathbb{N}^2$. On pose

$$p_{ij} = \frac{1}{\lambda} \mathcal{C}_{i+j}^{i} \left(\frac{\lambda}{1 + 2\lambda} \right)^{i+j+1}.$$

1. Vérifier que la famille $(p_{ij}, (i, j) \in \mathbb{N}^2)$ définit une probabilité sur $\mathbb{N}^2$.
2. Soit (X, Y) un couple de variables aléatoires de loi conjointe :

$$\mathbb{P}(X = i, Y = j) = p_{ij}.$$

a. Vérifier que la fonction génératrice du couple (X, Y) est donnée par :

$$G_{X,Y}(s, t) = \frac{1}{1 + 2\lambda - \lambda(s + t)}, (s, t) \in [-1.1] \times [-1.1].$$

b. Déterminer les fonctions génératrices des lois marginales. En déduire que les variables X et Y suivent la même loi binomiale négative de paramètres à identifier.

c. On pose $Z = X + Y$. Calculer la fonction génératrice de la variable Z et identifier sa loi. Vérifier que la loi conditionnelle de X sachant $(Z = z)$ est binomiale dont on calculera les paramètres.

Exercice 3.5.3. Soient X_1 et X_2 deux variables aléatoires indépendantes de loi commune f et de fonction caractéristique ϕ.

1. Calculer la fonction caractéristique de $X_1 - X_2$.
2. Montrer que

$$\mathbb{P}(X_1 - X_2 = 0) = \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{T} | \phi(t) |^2 \, dt.$$

3. Montrer que si ϕ est de carré intégrable alors

$$\mathbb{P}(X_1 - X_2 = 0) = 0.$$

Exercice 3.5.4. Une urne contient des boules rouges, vertes, bleues, en proportion p_r, p_v, p_b avec $p_r + p_v + p_b = 1$: On tire n boules avec remise et on note X_n le vecteur aléatoire de composantes R_n ; V_n et B_n désignant respectivement les nombres de boules rouges, vertes et bleues.

1. Déterminer la loi du vecteur X_n.
2. Les variables aléatoires R_n, V_n et B_n sont-elles indépendantes ?
3. Calculer G_X, la fonction génératrice du vecteur X_n.
4. Soit N une variable aléatoire discrète indépendante de la suite $(X_n)n \geq 1$.
   On pose
$$R_N = R_n, V_N = V_n \;\; et \;\; B_N = B_n \;\; si \;\; N = n.$$
   a. Utiliser le résultat de la première question et la loi de la variable aléatoire N pour calculer la loi du vecteur $X_N = (R_N, V_N, B_N)$.

   b. Montrer que les composantes de X_N sont indépendantes dans le cas où N suit une loi de Poisson.

   c. On désigne par G_N la fonction génératrice de la variable N : Calculer la fonction génératrice G_{XN} du veteur X_N en fonction de G_N. Retrouver alors le résultat de la question précédente.

Chapitre 4

Vecteur aléatoire gaussien (normal)

Les vecteurs gaussiens sont associés aux lois gaussiennes multivariées, et de ce fait jouent un rôle important en probabilités et en statistique. Ils apparaissent naturellement comme des objets limites et serviront en particulier dans le prochain chapitre.

4.1 Vecteur aléatoire gaussien

Définition 4.1. Un vecteur aléatoire X à valeurs dans $\mathbb{R}^n$ est un vecteur gaussien si et seulement si toute combinaison linéaire de ses coordonnées est une variable aléatoire gaussienne i.e. $\forall \lambda \in \mathbb{R}^n, \sum_{j=1}^{n} \lambda_j X_j$ suit une loi gaussienne.

Puisque la loi du vecteur aléatoire gaussien X est totalement déterminée par la donnée du vecteur espérance μ et de la matrice de covariance Σ_X, on écrit $X \sim \mathcal{N}(\mu_X, \Sigma_X)$ et on dira que X suit la loi gaussienne $\mathcal{N}(\mu_X, \Sigma_X)$.

Remarque 4.1.1.

- Si les composantes X_i sont des variables aléatoires gaussienne indépendantes, le vecteur X est gaussien .
- Si les composantes X_i sont de loi gaussienne mais pas indépendantes, il se peut que X ne soit pas un vecteur gaussien. Prenons par exemple X_1 de loi $N(0,1)$, et

$$X_2 = \begin{cases} X_1 & si \quad |X_1| \leq 1 \\ -X_1 & sinon. \end{cases}$$

Alors $X_2 \sim \mathcal{N}(0,1)$, mais $X = (X_1, X_2)$ n'est pas un vecteur gaussien, puisque $0 < \mathbb{P}(X_1 + X_2 = 0) < 1$ (donc $X_1 + X_2$ ne suit pas une loi normale).

Proposition 4.1.1. *Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Tout vecteur aléatoire gaussien $X \sim \mathcal{N}(\mu, \Sigma)$ de dimension n à valeurs dans $\mathbb{R}^n$ vérifie les propriétés suivantes :*

1. *Chaque composante de X est une variable aléatoire gaussienne appartenant à $L^2(\Omega, \mathcal{A}, \mathbb{P})$;*
2. *Le vecteur aléatoire X est intégrable et $\mathbb{E}(X) = \mu_X = (\mu_1, ..., \mu_n)^t$;*
3. *Le vecteur aléatoire X admet une matrice de covariance $\Sigma_X = \Sigma$.*

Proposition 4.1.2. Soient $\mu \in \mathbb{R}^n$ et $\Sigma \in \mathcal{M}_{n\times n}$ une matrice symétrique positive. Il existe alors un vecteur gaussien X de vecteur moyenne μ_X et de matrice de covariance Σ_X.

Démonstration. Soit $Z = (Z_1, ..., Z_n)$ un vecteur de variables aléatoires réelles i.i.d. de loi $\mathcal{N}(0, 1)$.

- Z est un vecteur gaussien de vecteur moyenne $\mu_Z = 0_n$ et de matrice de covariance $\Sigma_Z = I_n$.
- Σ étant symétrique positive, il existe $B \in \mathcal{M}_{n\times n}$ symétrique telle que $\Sigma = B^2$.
  Posons $X = BZ + \mu$.
- X est un vecteur gaussien de moyenne et de matrice de covariance respectives

$$\mu_X = B\mu_Z + \mu = \mu; \Sigma_X = B\Sigma_Z B^t = B^2 = \Sigma.$$

Exemple 4.1.1. *Montrer qu'il existe un vecteur gaussien de matrice de covariance et de vecteur moyenne*

$$\mu_X = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}; \Sigma_X = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Solution. La matrice Σ_X est bien symétrique.
Pour tout $X = (x_1, x_2, x_3)^t \in \mathbb{R}^3$, nous avons

$$\begin{aligned} x^t\Sigma_X x &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= 2(x_1^2 + x_1x^2 + x_1x_3 + x_2^2 + x^2x_3 + x_3^2) \\ &= (x_1 + \frac{x_2}{2} + x_3^2)^2 + \frac{3}{4}(x_2 + \frac{x_3}{3})^2 + \frac{23}{36}x_3^2 \geq 0. \end{aligned}$$

En définitive, Σ_X est symétrique positive.
D'aprés la proposition précédente, nous en déduisons qu'il existe un vecteur gaussien de vecteur moyenne μ_X et de matrice de covariance Σ_X.

Exemple 4.1.2. ***Cas particulier***. *Soient $X_1, ..., X_n$ des variables aléatoires réelles indépendates de loi $\mathcal{N}(0, 1)$. Calculer le vecteur moyenne et la matrice de covariance de $X = (X_1, ..., X_n)$.*
On a pour tout $i \in \{1, ..., n\}$, $\mathbb{E}(X_i) = 0$, donc

$$\mu_X = (0, ..., 0)^t.$$

De plus, pour tout $i, j \in \{1, ..., n\}$,

$$Cov(X_i, X_j) = \begin{cases} 1 \ si \ \ i = j \\ 0 \ \ sinon. \end{cases}$$

Par conséquent, $\Sigma_X = I_n$ la matrice identité de $\mathcal{M}_{n\times n}(\mathbb{R})$.

Proposition 4.1.3. Si X est un vecteur aléatoire gaussien de vecteur espérances $\mu = (\mu_1, ..., \mu_n)^t$ et de matrice de covariance Σ_X , alors, pour tout λ dans $\mathbb{R}^n$, la v.a. $\lambda' X$ est de loi $\mathcal{N}(\lambda'\mu, \lambda'\Sigma_X\lambda)$.

Démonstration. On utilise d'abord le fait que, par définition d'un vecteur gaussien, la v.a. $\lambda' X$ est de loi normale. Il ne reste plus qu'à calculer son espérance et sa variance. On utilise alors les résultats vus au chapitre 2, pour obtenir :

$$\mathbb{E}(\lambda' X) = \lambda'\mathbb{E}(X) = \lambda'\mu_X$$

et

$$\Sigma_{\lambda' X} = \lambda'\Sigma_X\lambda.$$

On peut aussi caractériser un vecteur gaussien par sa fonction caractéristique, grâce à la proposition suivante.

Proposition 4.1.4. On dit que n variables aléatoires $X_1, ..., X_n$ sont conjointement gaussiennes, ou que le vecteur aléatoire $X = (X_1, ..., X_n)^t$ est gaussien de dimension n sur $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$, s'il existe un vecteur colonne $\mu \in \mathbb{R}^n$ et une matrice Σ_X de n lignes et n colonnes symétrique telle que la fonction caractéristiqe de X soit de la forme :

$$\Phi_X(u) = \exp[i < u, \mu >] \exp[-\frac{1}{2}u^t\Sigma u] \ \ pour \ \ tout \ \ u \in \mathbb{R}^n. \tag{4.1}$$

Démonstration.
avec

$$\mathbb{E}(Y) = \mathbb{E}\left[\sum_{j=1}^{n} u_j X_j\right] = \sum_{j=1}^{n} \mathbb{E}(X_j) = u'\mu$$

et

$$\begin{aligned} Var(Y) &= Var\left[\sum_{j=1}^{n} u_j X_j\right] = \sum_{j}^{n}\sum_{k}^{n} Cov(u_jX_j, u_kX_k) \\ &= \sum_{j}^{n}\sum_{k}^{n} u_ju_kCov(X_j, X_k) = u^t\Sigma u. \end{aligned}$$

($\Rightarrow$) Par définition, on a :

$$\phi_X(u) = \mathbb{E}[e^{iu^tX}] = \mathbb{E}[e^{iY}] = \phi_Y(1),$$

où $Y = \sum_{j=1}^{n} u_jX_j$. La variable aléatoire Y est une variable aléatoire gaussienne puisque c'est une combinaison linéaire des coordonnées du vecteur gaussien X.
On a

$$\phi_Y(1) = e^{i\mathbb{E}Y}e^{-\frac{Var(Y)}{2}},$$

avec

$$\mathbb{E}(Y) = \mathbb{E}\left[\sum_{j=1}^{n} u_j X_j\right] = \sum_{j=1}^{n} \mathbb{E}(X_j) = u'\mu$$

et

$$\begin{aligned} Var(Y) &= Var\left[\sum_{j=1}^{n} u_j X_j\right] = \sum_{j}^{n}\sum_{k}^{n} Cov(u_j X_j, u_k X_k) \\ &= \sum_{j}^{n}\sum_{k}^{n} u_j u_k Cov(X_j, X_k) = u^t \Sigma u. \end{aligned}$$

($\Leftarrow$) *Soit $\lambda = (\lambda_1, ..., \lambda_n)^t$. On pose $Y = \sum_{j=1}^{n} \lambda_j X_j = \lambda' X$. Montrons que Y est une variable aléatoire gaussienne. Pour $u \in \mathbb{R}$, la fonction caractéristique de Y est donnée par :*

$$\begin{aligned} \phi_Y(u) &= \mathbb{E}\left[e^{iu\lambda^t X}\right] = \phi_X(u\lambda) \\ &= e^{i(u\lambda)^t\mu - \frac{1}{2}(u\lambda)^t \Sigma_X u\lambda}. \\ &= e^{iu\lambda^t u - \frac{1}{2}u^2 \lambda^t \Sigma_X \lambda}. \end{aligned}$$

On reconnaît la fonction caractéristique de la loi gaussienne $\mathcal{N}(\lambda^t\mu, \lambda^t\Sigma\lambda)$. Donc Y est bien une variable aléatoire gaussienne.

Remarque 4.1.2. Un vecteur gaussien de matrice de covariance Σ_X telle que $det(\Sigma_X) = 0$ (i.e.Σ_X non inversible) est dit dégénéré et n'admet pas de densité par rapport à la mesure de Lebesgue dans $\mathbb{R}^n$.

Théorème 4.1.5. Un vecteur aléatoire $X \backsim \mathcal{N}(\mu_X, \Sigma_X)$ est absolument continu si et seulement si sa matrice de covariance Σ_X est inversible. Dans ce cas, la densité de probabilité de X est

$$f_X(x_1, ..., x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det(\Sigma_X)}} \exp\left[-\frac{1}{2}(x-\mu_X)^t \Sigma_X^{-1}(x-\mu_X)\right], avec$$

X=$(X_1, ..., X_n)^t$.

4.1.1 Quelques cas particuliers

- $n = 1$. *Posons $\mu_X = \mu$, $\Sigma_X = (\sigma^2)$. On obtient :*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(x-\mu)^2}{2\sigma^2}]$$

On retrouve la densite de la loi normale univariee de moyenne μ et variance σ^2.

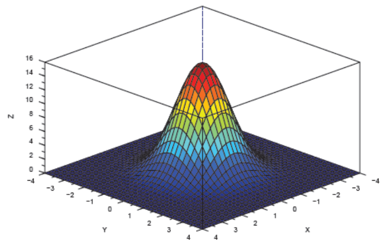

FIGURE 4.1 – Densité gaussienne en dimension deux

- $n = 2$. *Posons* $\mu_X = (\mu_1, \mu_2)^t$. *On peut écrire* Σ_X *sous la forme :*

$$\Sigma = \begin{pmatrix} \sigma_{X1}^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

avec $(\sigma_1 > 0)$ *et* $(\sigma_2 > 0)$. *On a alors* $(det\,(\sum_X)) = \sigma_1^2\sigma_2^2(1-\rho_{1,2}^2)$ *et* Σ_X *est une matrice définie positive puisque* $|\rho_{1,2}| < 1$. *De plus,*

$$\Sigma_X^{-1} = \frac{1}{det\,(\Sigma_X)} \begin{pmatrix} \sigma_2^2 & -\rho_{1,2}\sigma_1\sigma_2 \\ -\rho_{1,2}\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

On obtient donc

$$\begin{aligned} f(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho_{1,2}^2)}} \times \\ & \exp\left(\frac{1}{2(1-\rho_{1,2}^2)}\right)\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho_{1,2}\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]. \end{aligned}$$

4.2 Distributions marginales et conditionnelles

Supposons que l'on définisse la partition suivante :

$$X = \begin{pmatrix} X_a \\ X_b \end{pmatrix}; \mathbb{E}(X) = \mu_X = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}; \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix};$$

où

- μ_a et Σ_{aa} sont le vecteur espérance et la matrice de covariance du vecteur aléatoire X_a ;
- μ_b et Σ_{bb} sont le vecteur espérance et la matrice de covariance du vecteur aléatoire X_b ;
- $\Sigma_{ab} = \Sigma'_{ba}$ matrice des covariances entre les variables de X_a et les variables de X_b.

Théorème 4.2.1. *[3].* Soit X un vecteur aléatoire absolument continu. Si $X = (X_a, X_b)^t \sim \mathcal{N}(\mu, \Sigma_X)$, alors

$$X_a \sim \mathcal{N}(\mu_a, \Sigma_{aa}) \ ; \ X_b \sim \mathcal{N}(\mu_a, \Sigma_{bb}).$$

Démonstration. Soit $u \in \mathbb{R}^n$, on sait que $\phi(u) = e^{iu^t\mu - \frac{1}{2}u^t\Sigma_X u}$. D'autre part le partitionnemment des vecteurs u, μ_X et Σ_X. On peut écrire $\phi(u) = \phi(u_a, u_b)$ avec

$$u^t\mu = (u_a, u_b)\begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} = u_a^t\mu_a + u_b^t\mu_b, u^t \sum u = \begin{pmatrix} u_a \\ u_b \end{pmatrix}^t \begin{pmatrix} \sum_{aa} & \sum_{ab} \\ \sum_{ba} & \sum_{bb} \end{pmatrix} \begin{pmatrix} u_a \\ u_b \end{pmatrix}.$$

La fonction caractéristique marginale de X_a est donnée par

$$\phi_{X_a}(u_a) = \phi_X(u_a, 0),$$

c'est-à dire en posant $u_b = 0$ dans les expressions précédentes. En effectuant les produits matriciels. On obtient ainsi

$$\phi_{X_a}(u_a) = e^{iu_a^t\mu_a - \frac{1}{2}u_a^t \sum_{aa} u_a},$$

qui est bien la fonction caractéristique d'un vecteur normal dont le vecteur espérance est μ_a et la matrice de covariance est $\sum_{aa}$. La démonstration pour X_b est similaire en posant $u_a = 0$.

Définition 4.2. *Les fonctions de densité de probabilité marginales des vecteurs X_a et X_b sont données par :*

$$f_{X_a}(x_a) = \frac{1}{\sqrt{(2\pi)^{n_a} det(\sum_{aa})}} \exp\left(-\frac{1}{2}(x_a - \mu_a)^t \sum_{aa}^{-1}(x_a - \mu_a)\right);$$

$$f_{X_b}(x_b) = \frac{1}{\sqrt{(2\pi)^{n_b} det(\sum_{bb})}} \exp\left(-\frac{1}{2}(x_b - \mu_b)^t \sum_{bb}^{-1}(x_b - \mu_b)\right).$$

Théorème 4.2.2. *[3].* Soit X un vecteur aléatoire absolument continu. Si $X = (X_a, X_b)^t \sim \mathcal{N}(\mu, \Sigma)$, alors :

$$[X_a/X_b = x_b] \sim \mathcal{N}(\mu_{a/b}, \Sigma_{a/b}) \ ; \ X_b/X_a = x_a \sim \mathcal{N}(\mu_{b/a}, \Sigma_{b/a}),$$

avec :

$$\mu_{a/b} = \mu_a + \Sigma_{ab}^{-1}\Sigma_{bb}^{-1}(x_b - \mu_b) \ ; \ \Sigma_{a/b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba};$$

$$\mu_{b/a} = \mu_b + \Sigma_{ba}^{-1}\Sigma_{aa}^{-1}(x_a - \mu_a) \ ; \ \Sigma_{b/a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}.$$

Démonstration. *Par définition, la loi conditionnelle* $f_{X_b/X_a=x_a}(x_b) = \frac{f_X(x)}{f(x_a)}$. *D'autre part, en partitionnant les vecteurs* x, μ *et la matrice* Σ, *on peut écrire :*

$$det\left(\Sigma\right) = det\begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} = det\left(\Sigma_{aa}\right) det\left(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)$$

$$(x-\mu)^t\Sigma^{-1}(x-\mu) = det\left(\Sigma^{-1}\right)\begin{pmatrix} x_a-\mu_a \\ x_b-\mu_b \end{pmatrix}^t \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} \begin{pmatrix} x_a-\mu_a \\ x_b-\mu_b \end{pmatrix}$$

avec

$$det\left(\Sigma_X^{-1}\right) = \frac{1}{det\left(\Sigma_X\right)}$$

et

$$\begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -\Sigma_{aa}^{-1}\Sigma_{ab} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \Sigma_{aa}^{-1} & 0 \\ -S^{-1}\Sigma_{aa}\Sigma_{aa}^{-1} & S^{-1} \end{pmatrix}$$

avec

$$S = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}.$$

On obtient :

$$f(x_b/x_a) = \frac{1}{(2\pi)^{n_b} det\left(\Sigma_{a/b}\right)} e^{-\frac{1}{2}(x_b-\mu_{b/a})^t\Sigma_{b/a}(x_b-\mu_{a/b})}.$$

Si l'on pose

$$\mu_{a/b} = \mu_a + \Sigma_{ab}^{-1}\Sigma_{bb}^{-1}(x_b-\mu_b) \;\; ; \;\; \Sigma_{a/b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba},$$

l'expression $f_{X_b/X_a=x_a(x_b)}$ *est bien celle de la fonction de densité de probabilité d'un vecteur normal dont le vecteur espérance est* $\mu_{a/b}$ *est la matrice de covariance est* $\Sigma_{a/b}$.

Exercice 4.2.3. *[3].* On a mesuré simultanément la hauteur X_1 (en centimètre), le logarithme népérien X_2 du poids (en kilogrammes) et le pourcentage X_3 de graisse (en pourcentage % du poids total) pour un grand nombre d'hommes de type européen agés de 50 à 80 ans . Ceci a permis de déterminer les espérances et les écarts -type données dans le tableau 4.1 : Les corrélations entre ces variables étant $\rho_{X_1X_2} = 0.49$, $\rho_{X_1X_3} = 0.07$

TABLE 4.1 –

i	1	2	3
μ_{X_i}	174	4.41	21.1
σ_{X_i}	6.5	0.15	5

et $\rho_{X_2X_3} = 0.49$. On suppose que $X = (X_1, X_2, X_3)^t$ forme un vecteur aléatoire normal.

1. Quelle est la corrélation entre la taille et le poids lorsque le pourcentage de graisse est fixé ?

2. Quelle est la corrélation entre la taille et le pourcentage de graisse lorsque le poids est fixé ?
3. La variable X_3 est laborieuse à mesurer, en envisage de la prédire sur base de x_1 et x_2 faciles à obtenir. Quelle sera la valeur attendue pour X_3 si l'on sait que $X_1 = x_1$ et $X_2 = x_2$?
4. Quel est la distribution du pourcentage de graisse chez un individu qui mesure 180 centimètre et pèse 68 Kilogrammes ?

Solution.

$$\Sigma_{a/b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} = \begin{pmatrix} 42 & 0.436 \\ 0.436 & 0.0141 \end{pmatrix}.$$

Soit :

$$\rho_{X_1,X_2/X_3} = \frac{0.436}{\sqrt{(42)(0.0141)}} = 0.57$$

c'est le coefficient de corrélation entre la taille et le poids conditionnellement au pourcentage de graisse.

En posant $X_a = (X_1, X_3)^t$ *et* $X_b = X_2$,

$$\Sigma_{aa} = \begin{pmatrix} 42.3 & 2.28 \\ 2.28 & 25 \end{pmatrix}; \Sigma_{ab} = 0.0225; \Sigma_{ab} = \Sigma_{ba}^t = \begin{pmatrix} 0.478 \\ 0.458 \end{pmatrix}.$$

et l'on obtient :

$$\Sigma_{a/b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} = \begin{pmatrix} 32.1 & -7.44 \\ -7.44 & 46.7 \end{pmatrix}$$

soit :

$$\rho_{X_1,X_3/x_2} = \frac{-7.44}{\sqrt{(32.1)(46.7)}} = -0.53$$

c'est le coefficient de corrélation entre la taille et le pourcentage de graisse conditionnellement au poids

En posant $X_a = (X_1, X_2)^t$ *et* $X_b = X_3$*, on peut écrire*

$$\begin{aligned} \mu_{b/a} &= \;= 21.1 + (2.28, 0.458)\begin{pmatrix} 42.3 & 0.478 \\ 0.478 & 0.0225 \end{pmatrix}^{-1}\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 4.41 \\ 174 \end{pmatrix}\right) \\ &= -0.232x_1 + 25.3x_2 - 49.9. \end{aligned}$$

En posant $x_1 = 180$ *et* $x_2 = ln(68)$ *dans la formule précédente, on obtient :*

$$\mu_{b/a} = (-0.232)(180) + (25.3)ln(68) - 49.9 = 150.9$$

Soit un pourcentage moyen $\mu_{X_3/(x_1,x_2)}$ *en graisse inférieur à la moyenne* μ_{X_3}*. La variance quant à elle ne dépend ni de* x_1 *ni de* x_2 *et est donnée par :*

$$\Sigma_{b/a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} = 14$$

soit une valeur pour $\sigma^2_{X_3/(x_1,x_2)}$ inférieure à la valeur $\sigma^2_{X_3}$, ce qui traduit la réduction de l'incertitude sur X_3 obtenue en utilisant l'information donnée par x_1 et x_2. On a donc $X_3\diagup(180, ln(68)) \sim \mathcal{N}(15.09, 14)$ sous l'hpothèse du vecteur gaussien.

1. $X = (X_1, X_2, X_3)^t \sim \mathcal{N}(\mu_X, \Sigma_X)$, avec :

$$\mu = \begin{pmatrix} 174 \\ 4.41 \\ 21.1 \end{pmatrix}; S = \begin{pmatrix} 6.5 & 0 & 0 \\ 0 & 0.15 & 0 \\ 0 & 0 & 5 \end{pmatrix}; R = \begin{pmatrix} 1 & 0.49 & 0.07 \\ 0.49 & 1 & 0.61 \\ 0.07 & 0.61 & 1 \end{pmatrix}$$

   ce qui donne :

$$\Sigma_X = SRS = \begin{pmatrix} 42.3 & 0.478 & 2.28 \\ 0.478 & 0.0225 & 0.0458 \\ 2.28 & 0.0458 & 25 \end{pmatrix}.$$

   En posant $X_a = (X_1, X_2)$ et $X_b = X_3$, on peut écrire

$$\Sigma_{aa} = \begin{pmatrix} 42.3 & 0.478 \\ 0.478 & 0.0225 \end{pmatrix}; \Sigma_{bb} = 25; \Sigma_{ba} = \Sigma_{ab} = \begin{pmatrix} 2.28 \\ 0.458 \end{pmatrix}$$

   et l'on obtient :

$$\Sigma_{a/b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} = \begin{pmatrix} 42 & 0.436 \\ 0.436 & 0.0141 \end{pmatrix}.$$

   Soit :

$$\rho_{X_1,X_2/X_3} = \frac{0.436}{\sqrt{(42)(0.0141)}} = 0.57$$

   c'est le coefficient de corrélation entre la taille et le poids conditionnellement au pourcentage de graisse.

2. En posant $X_a = (X_1, X_3)^t$ et $X_b = X_2$,

$$\Sigma_{aa} = \begin{pmatrix} 42.3 & 2.28 \\ 2.28 & 25 \end{pmatrix}; \Sigma_{ab} = 0.0225; \Sigma_{ab} = \Sigma_{ba}^t = \begin{pmatrix} 0.478 \\ 0.458 \end{pmatrix}.$$

   et l'on obtient :

$$\Sigma_{a/b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} = \begin{pmatrix} 32.1 & -7.44 \\ -7.44 & 46.7 \end{pmatrix}$$

   soit :

$$\rho_{X_1,X_3/x_2} = \frac{-7.44}{\sqrt{(32.1)(46.7)}} = -0.53$$

   c'est le coefficient de corrélation entre la taille et le pourcentage de graisse conditionnellement au poids

3. En posant $X_a = (X_1, X_2)^t$ et $X_b = X_3$, on peut écrire

$$\begin{aligned}\mu_{b/a} &= = 21.1 + (2.28, 0.458)\begin{pmatrix} 42.3 & 0.478 \\ 0.478 & 0.0225 \end{pmatrix}^{-1}\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 4.41 \\ 174 \end{pmatrix}\right) \\ &= -0.232x_1 + 25.3x_2 - 49.9.\end{aligned}$$

4. En posant $x_1 = 180$ et $x_2 = ln(68)$ dans la formule précédente, on obtient :

$$\mu_{b/a} = (-0.232)(180) + (25.3)ln(68) - 49.9 = 150.9$$

Soit un pourcentage moyen $\mu_{X_3/(x_1,x_2)}$ en graisse inférieur à la moyenne μ_{X_3}. La variance quant à elle ne dépend ni de x_1 ni de x_2 et est donnée par :

$$\Sigma_{b\diagup a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} = 14$$

soit une valeur pour $\sigma^2_{X_3/(x_1,x_2)}$ inférieure à la valeur $\sigma^2_{X_3}$, ce qui traduit la réduction de l'incertitude sur X_3 obtenue en utilisant l'information donnée par x_1 et x_2. On a donc $X_3\diagup(180, ln(68)) \sim \mathcal{N}(15.09, 14)$ sous l'hpothèse du vecteur gaussien.

4.3 Indépendance

Pour un vecteur normal, l'indépendance entre variables est équivalente à l'absence de corrélation entre ces variables.

Théorème 4.3.1. Soit X un vecteur gaussien dans $\mathbb{R}^n$: Pour que ses composantes $X_1, ..., X_n$ soient indépendantes, il faut et il suffit que la matrice de covariance soit diagonale.

Démonstration. Il suffit, bien sûr, de montrer la réciproque. Supposons donc que Σ_X soit diagonale, i.e.

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

Comme X est un vecteur gaussien de loi $N(\mu, \Sigma)$, chacune de ses composantes X_j , pour $j = 1, ..., n$ est de loi normale $N\left(\mu_j, \sigma_j^2\right)$ et de fonction caractéristique :

$$\phi_{X_j}(\lambda_j) = \exp\left[i\lambda_j\mu_j - \frac{1}{2}\sigma_j^2\lambda_j^2\right], \lambda_j \in \mathbb{R}. \tag{4.2}$$

Par ailleurs, la fonction caractéristique du vecteur X est, pour tout λ dans $\mathbb{R}^n$:

$$\begin{aligned}
\phi_X(\lambda) &= \exp\left[i\lambda'\mu - \frac{1}{2}\lambda'\Sigma_X\lambda\right] \\
&= \exp\left[i\sum_{j=1}^{n}\lambda_j\mu_j - \frac{1}{2}\sum_{j=1}^{n}\lambda_j^2\sigma_j^2\right] \\
&= \exp\left[\sum_{j=1}^{n}\left(i\lambda_j\mu_j - \frac{1}{2}\lambda_j^2\sigma_j^2\right)\right] \\
&= \prod_{j=1}^{n}\exp\left[i\lambda_j\mu_j - \frac{1}{2}\lambda_j^2\sigma_j^2\right] = \prod_{j=1}^{n}\phi_{X_j}(\lambda_j).
\end{aligned}$$

Proposition 4.3.2. *[3].*

- $X_1 \perp ... \perp X_{n_a} \diagup x_b \Leftrightarrow \Sigma_{a\diagup b} = \begin{pmatrix} \sigma^2_{X_1\diagup x_b} & 0 & ... & 0 \\ 0 & \sigma^2_{X_2\diagup x_b} & ... & 0 \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ 0 & 0 & ... & \sigma^2_{X_{n_a}\diagup x_b} \end{pmatrix}$
- $X_a \perp X_b \Leftrightarrow \Sigma_{ab} = \Sigma_{ba}^t$.

  Si $X = (X_a, X_b)^t \sim N(\mu_X, \Sigma_X)$ avec $\Sigma_X = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$, alors

- $X_1 \perp ... \perp X_{n_a} \Leftrightarrow \Sigma_{aa} = \begin{pmatrix} \sigma^2_{X_1} & 0 & ... & 0 \\ 0 & \sigma^2_{X_2} & ... & 0 \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ 0 & 0 & ... & \sigma^2_{X_{n_a}} \end{pmatrix}$
- $X_1 \perp ... \perp X_{n_a} \diagup x_b \Leftrightarrow \Sigma_{a\diagup b} = \begin{pmatrix} \sigma^2_{X_1\diagup x_b} & 0 & ... & 0 \\ 0 & \sigma^2_{X_2\diagup x_b} & ... & 0 \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ 0 & 0 & ... & \sigma^2_{X_{n_a}\diagup x_b} \end{pmatrix}$
- $X_a \perp X_b \Leftrightarrow \Sigma_{ab} = \Sigma_{ba}^t$.

Exemple 4.3.1. *Soit $X = (X_1, X_2, X_3)^t$ un vecteur gaussien de matrice de covariance*

$$\Sigma_X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 4 \end{pmatrix}.$$

Montrer que X_1 et $X_2 - X_3$ sont indépendants.
Solution. *Le vecteur $Y = (X_1, X_2 - X_3)^t$ est un vecteur gaussien, car*

$$Y = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} X.$$

De plus

$$\begin{aligned} Cov(Y_1, Y_2) &= Cov(X_1, X_2 - X_3) = Cov(X_1 X_2) - Cov(X_1, X_3) \\ &= 1 - 1 = 0. \end{aligned}$$

Par conséquent, $Y_1 = X_1$ et $Y_2 = X_2 - X_3$ sont indépendants.

4.3.1 Transformation linéaire d'un vecteur gaussien

Proposition 4.3.3. *La transformée d'un vecteur gaussien de $\mathbb{R}^n$ par une application linéaire de $\mathbb{R}^n$ vers $\mathbb{R}^p$ est encore un vecteur gaussien.*

Démonstration.

$$\begin{aligned} \phi_Y(\lambda) &= \phi_{AX}(\lambda) = \mathbb{E}\left(e^{i<\lambda, AX>}\right) = \mathbb{E}\left(e^{i<A'\lambda, X>}\right) \\ &= \phi_X(A^t\lambda) = \exp\left[i\lambda^t A\lambda - \frac{1}{2}\lambda^t A \sum_X A^t \lambda\right]. \end{aligned}$$

Par caractérisation, le vecteur Y est donc un vecteur gaussien dans $\mathbb{R}^p$ de vecteur des espérances $A\mu$ et de matrice de covariance $A\Sigma_X A^t$, i.e.

$$Y \sim \mathcal{N}\left(A\mu_X, A\Sigma_X A^t\right).$$

Soit X un vecteur gaussien de $\mathbb{R}^n$, de vecteur des espérances μ et de matrice de covariance Σ_X. Soit A la matrice associée à une transformation linéaire quelconque de $\mathbb{R}^n$ vers $\mathbb{R}^p$. La matrice A est donc de dimension $p \times n$. Calculons la fonction caractéristique du vecteur aléatoire $Y = AX$. D'après ce que l'on a vu au chapitre précédent, pour tout λ de $\mathbb{R}^p$, on a :

$$\begin{aligned} \phi_Y(\lambda) &= \phi_{AX}(\lambda) = \mathbb{E}\left(e^{i<\lambda, AX>}\right) = \mathbb{E}\left(e^{i<A'\lambda, X>}\right) \\ &= \phi_X(A^t\lambda) = \exp\left[i\lambda^t A\lambda - \frac{1}{2}\lambda^t A \sum_X A^t \lambda\right]. \end{aligned}$$

Par caractérisation, le vecteur Y est donc un vecteur gaussien dans $\mathbb{R}^p$ de vecteur des espérances $A\mu$ et de matrice de covariance $A\Sigma_X A^t$, i.e.

$$Y \sim \mathcal{N}\left(A\mu_X, A\Sigma_X A^t\right).$$

Exercice 4.3.4. *Soit* $X = (X_1, ..., X_4)^t$ *un vecteur gaussien tel que*

$$\mu_X = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 2 \end{pmatrix}; \Sigma_X = \begin{pmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Montrer que $Y = (X_1 + 2X_2 + 1, X_2 - X_3)^t$ *est un vecteur gaussien et donner sa loi en fonction de* μ_X *et* Σ_X.

Solution. Nous avons

$$Y = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix} X + \begin{pmatrix} 1 \\ 0 \end{pmatrix} = AX + b.$$

Donc Y est un vecteur gaussien de loi $\mathcal{N}(\mu_X, \Sigma_Y)$, où

$$\mu_Y = A\mu_X + b = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad et \quad \Sigma_Y = A\Sigma_X A^t = \begin{pmatrix} 14 & 3 \\ 3 & 2 \end{pmatrix}.$$

4.4 Exercices du chapitre 4.

Exercice 4.4.1. Considérons la matrice :

$$\Sigma = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}.$$

1. Montrer q'il existe un vecteur $(X_1, X_2, X_3)^t$ de moyenne $\mu = (0, 1, 1)^t$ et de matrice de covariance Σ.
2. Préciser les relations d'indépendance entre les coordonnées de X.
3. Posons $Y_1 = X_1 + X_2$, $Y_2 = \alpha X_1 + 2X_2 + \beta X_3$ et $Y_3 = -X_1 - X_2 + X_3$.
   a. Montrer que $Y = (Y_1, Y_2, Y_3)^t$ est un vecteur gaussien dont on donnera les paramètres.
   b. Les variables aléatoires Y_1, Y_2 et Y_3 sont-elles mutuellement indépendantes ?
   c. Pour quelles valeurs de (α, β), les variables Y_2 et Y_3 sont-elles indépendantes ?

Exercice 4.4.2. Soit $X = (X_1, X_2, X_3) \sim N(0, \Sigma)$, où

$$\sum = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 5 & 1 \\ -2 & 1 & 5 \end{pmatrix}.$$

1. Montrer que $det(\Sigma) = 0$. Le vecteur X possède -il une densité dans $\mathbb{R}^3$?
2. Trouver $a \in \mathbb{R}$ pour que X_1 et $Y = X_2 - aX_1$ soient independantes. Calculer $Var(Y)$ et en deduire la loi de (X_1, Y).
3. Trouver la loi conditionnelle de X_2 sachant X_1.

Exercice 4.4.3. Soit $X = (X_1, X_2, X_3)^t \sim N(0, \Sigma_X)$, où

$$\sum = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

1. Quelle est la loi du couple (X_1, X_2)?
2. Déterminer a un reel tel que $Y = aX_1 + X_2$ est indépendante de X_1. Que vaut $\mathbb{E}[Y]$? $Var(Y)$?.
3. En déduire $\mathbb{E}[X_2 \diagup X_1]$. Quelle est la loi conditionnelle de X_2 sachant X_1?
4. Déterminer un réel β tels que $Z = \beta X_1 + X_3$ est indépendante de X_1. En deduire

$$\mathbb{E}[X_3 \diagup X_1], \mathbb{E}[X_3^2 \diagup X_1].$$

5. Calculer E[$\mathrm{X}_1^2 X_2 + X_3^2 X_1 \diagup X_1$].

Exercice 4.4.4. Posons $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, *où* $\mid \rho \mid < 1$.

1. Montrer qu'il existe un vecteur gaussien $(X_1, X_2)^t$ de moyenne nulle et de matrice de covariance Σ_X.
2. On pose :

   a. $Y_1 = \frac{1}{\sqrt{2}}(X_1 + X_2)$, $Y_2 = \frac{1}{\sqrt{2}}(X_1 - X_2)$.

   b. Montrer que $Y = (Y_1, Y_2)^t$ est un vecteur gaussien et calculer la matrice de covariance du vecteur Y.

   c. Les variables aléatoires Y_1 et Y_2 sont-elles indépendantes ?

   d. Montrer que la loi de Y admet une densité f_Y.

Exercice 4.4.5. Soit X un vecteur gaussien de matrice de covariance Σ et d'espérance μ. Nous supposons que $\Sigma = ADA^t$ où D est diagonale et A orthogonale. Nous considérons le vecteur aléatoire $Y = A^t(X - \mu)$.

1. Montrer que Y est un vecteur gaussien
2. Déterminer l'espérance μ_Y de Y.
3. Déterminer la matrice de covariance Σ_Y de Y . En déduire que les variables Y_k sont indépendantes.

Exercice 4.4.6. Soit $(X_1, X_2)^t$ un vecteur gaussien centré, avec $\mathbb{E}(X_1^2) = 4$ et $\mathbb{E}(X^2) = 1$ et tel que les variables $2X_1 + X_2$ et $X_1 - 3X_2$ sont indépendantes.

1. Déterminer la matrice de covariance de (X_1, X_2).
2. Montrer que le vecteur $(X_1 + X_2, 2X_1 - X_2)$ est également gaussien, puis déterminer sa matrice de covariance.

Exercice 4.4.7. Soit $(X_1, X_2, X_3)^t \in \mathbb{R}^3$ un vecteur gaussien. On pose $U = X_1 + X_2 + X_3$ et $V = X_1 - X_2$.

1. Montrer que $(U, V) \in \mathbb{R}^2$ est gaussien
2. A quelle condition sur la matrice de covariance de $(X_1, X_2, X_3)^t$ les variables U et V sont-elles indépendantes ?

Exercice 4.4.8. Existe t-il un vecteur gaussien de $\mathbb{R}^3$ dont la matrice de covariance est

$$\Sigma_X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 3 & 4 \\ 0 & 4 & 3 \end{pmatrix}.$$

Exercice 4.4.9. Soit $X = (X_1, ..., X_n)^t$ un vecteur aléatoire gaussien, centré, réduit, $\mathcal{N}(0, I_n)$. Pour tout $i \in 1, ..., n$, on pose $Y_i = X_1 + ... + X_i - X_{i+1}$ (avec la convention $X_{n+1} = 0$). Les v.a. $Y_1, ..., Y_n$ sont-elles indépendantes ?

Exercice 4.4.10. Soient n variables aléatoires réelles $X_1, X_2, ..., X_n$ indépendantes et identiquement distribuées selon une loi normale centrée de variance unité.On définit deux nouvelles variables aléatoires :

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i \;\; et \;\; V_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2.$$

Soit K une matrice centrée d'ordre n orthogonale et dont les termes de la première ligne sont tous égaux à $\frac{1}{n}$. Soit le vecteur aléatoire $Y = AX$ où $X = (X_1, ..., X_n)^t$.

1. Déterminer la loi de Y.
2. Exprimer M_n et V_n en fonction des composantes de Y.
3. Les variables V_n et M_n sont elles indépendantes ?
4. Donner la loi de M_n.
5. Montrer que si X est une variable aléatoire qui suit une loi normale centrée de variance unité, alors X^2 suit une loi gamma $\gamma(\frac{1}{2}, \frac{1}{2})$. On rappelle que $Y \sim \gamma(\theta, \lambda)$ si Y est absolument continue et admet pour densité :
$$f_Y(y) = \frac{\lambda^\theta}{\Gamma(\theta)} y^{\theta-1}\exp[-\lambda y], y \in]0, \infty[,$$
avec $\Gamma(\theta) = \int_0^\infty x^{\theta-1}e^{-x}dx$.
6. Soient deux variables aléatoires Z_1 et Z_2 et de loi respectivement $\gamma(\theta_1, \lambda)$ et $\gamma(\theta_2, \lambda)$. Déterminer la loi de $Z = Z_1 + Z_2$
7. Déduire des questions précédentes la loi de V_n.

Exercice 4.4.11. Soient X_1 et X_2 deux variables aléatoires indépendantes de loi commune f et de fonction caractéristique ϕ.

1. On suppose que X_1 est de loi gaussienne standard. On pose $X_3 = ZX_1$ pour une variable aléatoire Z de loi uniforme sur $\{-1, 1\}$ et indépendante de X_1 sur $[-1, 1]$:
2. Vérifirer que X_3 est de loi gaussienne standard.
3. Montrer que
$$\mathbb{P}(X_1 - X_2 = 0) = \mathbb{P}(X_1 - X_3 = 0) = \frac{1}{2}.$$
4. Que peut on dire de l'indépendance des variables aléatoires X_1 et X_3 ? Le vecteur (X_1, X_3) est-il gaussien ?

Exercice 4.4.12. Soit (X, Y) un vecteur gaussien de matrice de covariance

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}.$$

1. Verifier que la matrice est bien une matrice de covariance.
2. Déterminer $\mathbb{E}[X \diagup Y]$.
3. Déterminer la loi de X conditionnellement à Y.

Chapitre 5

Convergence des suites de vecteurs aléatoires

5.1 Rappel : Modes de convergence des suites de variables aléatoires

5.1.1 Notations

On considère $(X_n)_{n\in\mathbb{N}}$ (éventuellement $(Y_n)_{n\in\mathbb{N}}$) une suite de variables aléatoires définies sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et X (éventuellement Y) une variable aléatoire définie sur le même espace.
On note $(F_n)_{n\in\mathbb{N}}$ la suite des fonctions de répartitions de $(X_n)_{n\in\mathbb{N}}$ et F celle de X.
On note $(\phi_n)_{n\in\mathbb{N}}$ la suite des fonctions caractéristiques de $(X_n)_{n\in\mathbb{N}}$ et ϕ celle de X.

5.1.2 Convergence en loi

Définition 5.1. $X_n \overset{\mathcal{L}}{\to} X$: *X_n converge en loi vers X si et seulement si*

$$\lim_{n\to\infty} F_n(x) = F(x),$$

en tout point de continuité de F.

Exemple 5.1.1. *Soit X_n une suite de v.a. de loi $\mathcal{B}(n, \frac{\lambda}{n})$. Alors, lorsque n tend vers l'infini,*

$X_n \xrightarrow{\mathcal{L}} X$, *où* X *suit une loi* $\mathcal{P}(\lambda)$. *Soit* $x \in \{0, ..., n\}$.

$$\begin{aligned}
\mathbb{P}(X_n = x) &= C_n^x \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n(n-1)...(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^x} \\
&= \frac{\lambda^x}{x!} \left(1 - \frac{1}{n}\right) ... \left(1 - \frac{x-1}{n}\right) \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^x}.
\end{aligned}$$

$$\lim_{n \longrightarrow \infty} \mathbb{P}(X_n = x) = \frac{\lambda^x}{x!} \lim_{n \longrightarrow \infty} \left(1 - \frac{1}{n}\right)^n = \frac{\lambda^x}{x!} e^{-\lambda}.$$

Donc

$$\lim_{n \longrightarrow \infty} F_n(x) = \lim_{n \longrightarrow \infty} \sum_{k=0}^{x} \mathbb{P}(X_n = k) = \mathbb{P}(X \leq x),$$

où X *suit une loi de Poisson de paramètre* λ.

Théorème 5.1.1. ***Théorème de Paul Levy.***

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \lim_{n \to \infty} \phi_n(x) = \phi(x).$$

Lemme 5.1.2. ***Lemme de Portmanteau.***
$X_n \xrightarrow{\mathcal{L}} X$ *:* $(X_n)_{n \in \mathbb{N}}$ *converge en loi vers* X *si pour toute fonction* f *réelle, continue et bornée sur* $\mathbb{R}$*, on a*

$$\lim_{n \to \infty} \mathbb{E}\left[f(X_n)\right] = \mathbb{E}\left[f(X)\right].$$

Théorème 5.1.3. *[14]*

$$\forall c \in \mathbb{R}, X_n \xrightarrow{\mathcal{L}} X, Y_n \xrightarrow{\mathcal{L}} c \Rightarrow (X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c).$$

Théorème 5.1.4. ***Téorème Slutsky.***
$\forall c \in \mathbb{R}, X_n \xrightarrow{\mathcal{L}} X, Y_n \xrightarrow{\mathcal{L}} c$, *alors*

1. $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$.
2. $X_n Y_n \xrightarrow{\mathcal{L}} Xc$.
3. $\frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X}{c}, c \neq 0$.

Convergence en probabilité

Définition 5.2. $X_n \xrightarrow{\mathbb{P}} X$ *:* $(X_n)_{n \in \mathbb{N}}$ *converge en probabilité vers* X *si et seulement si*

$$\forall \epsilon > 0 \;\; \mathbb{P}(\mid X_n - X \mid > \epsilon) \overset{n \longrightarrow \infty}{\longrightarrow} 0.$$

Théorème 5.1.5. ***Condition suffisante de convergence en probabilité.***

$$\lim_{n\to\infty}\mathbb{E}(X_n)=c,\ \lim_{n\to\infty}Var(X_n)=0\Rightarrow X_n\xrightarrow{\mathbb{P}}c, c<\infty.$$

Propriété 5.1.6.

$$X_n\overset{\mathbb{P}}{\to}X, Y_n\overset{\mathbb{P}}{\to}Y\Rightarrow X_n+Y_n\overset{\mathbb{P}}{\to}X+Y.$$

5.1.3 Convergence presque- sûre

Théorème 5.1.7. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles.*

1. *Si X_n converge en loi vers une variable aléatoire réelle X, alors ϕ_n converge vers ϕ, la fonction caractéristique de X.*
2. *Si ϕ_n converge simplement vers une fonction ϕ et si ϕ est continue en 0, alors X_n converge en loi vers une variable aléatoire réelle X dont la fonction caractéristique est ϕ.*

Propriété 5.1.8. $X_n\overset{p.s.}{\to}X$: *$(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers X si et seulement si*

$$\mathbb{P}\left(\{w\in\Omega : X_n(w)\overset{n\longrightarrow\infty}{\longrightarrow}X(w)\}\right)=1.$$

2.

$$X_n\overset{p.s.}{\to}X\Leftrightarrow\mathbb{P}\left(\lim_{n\longrightarrow\infty}X_n=X\right)=1.$$

3.

$$X_n\overset{p.s.}{\to}X\Leftrightarrow\{w\in\Omega : \lim_{n\to\infty}X_n(w)\neq X(w)=\varnothing\}.$$

Proposition 5.1.9. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires et soit g une fonction continue.*

1. *Si $X_n\overset{p.s}{\to}X$ alors $g(X_n)\overset{p.s}{\to}g(X)$.*
2. *Si $X_n\overset{\mathbb{P}}{\to}X$ alors $g(X_n)\overset{\mathbb{P}}{\to}g(X)$.*

5.2 Convergence en loi des suites de vecteurs aléatoires

Pour tout vecteur aléatoire $X=(X_1,...,X_d)^t$ à valeurs dans $\mathbb{R}^d$, on note F_X sa fonction de répartition et ϕ_X sa fonction caractéristique.
Pour toute fonction g, on pose

$$C_g=\{x=(x_1,...,x_d)\in\mathbb{R}^d : g\ \ continue\ \ en\ \ x\}.$$

Définition 5.3. *Soient $(X_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$ et X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. La suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si, pour toute fonction g de $\mathbb{R}^d$ vers $\mathbb{R}$, continue et bornée*

$$\lim_{n\to\infty}\mathbb{E}\left(g(X_n)\right)=\mathbb{E}\left(g(X)\right).$$

On note $X_n\overset{\mathcal{L}}{\to}X$ et on dit aussi parfois que la loi de X_n converge vers celle de X.

Proposition 5.2.1. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. La suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si et seulement si*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

en tout point de continuité de F i.e. pour tout $x \in C_F$.

Proposition 5.2.2. *Soit (X_n) une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. Les affirmations suivantes sont équivalentes.*

1. $F_{X_n}(x) \overset{n\to\infty}{\longrightarrow} F_X(x)$ *pour tout* $x \in C_F$.
2. $\mathbb{E}\left(g(X_n)\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left(g(X)\right)$ *pour toute fonction mesurable $g : \mathbb{R}^d \to \mathbb{R}$ continue et bornée.*
3. $\mathbb{E}\left(g(X_n)\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left(g(X)\right)$ *pour toute fonction mesurable $g : \mathbb{R}^d \to \mathbb{R}$ uniformément continue et bornée.*
4. $\mathbb{E}\left(g(X_n)\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left(g(X)\right)$ *pour toute fonction mesurable $g : \mathbb{R}^d \to \mathbb{R}$ continue à support compact.*
5. $\mathbb{E}\left(g(X_n)\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left(g(X)\right)$ *pour toute fonction mesurable $g : \mathbb{R}^d \to \mathbb{R}$ mesurable telle que $\mathbb{P}\left(X \in C_g\right) = 1$.*

5.3 Convergence en probabilité

Définition 5.4. *Soit (X_n) et X des vecteurs aléatoires à valeurs dans $\mathbb{R}^p$. On dit que (X_n) converge en probabilité vers le vecteur aléatoire X si ses p composantes $X_{1,1}, ..., X_{n,p}$ convergent en probabilité vers les composantes $X_1, , ..., X_p$ de X.*

Définition 5.5. *Soit $\parallel . \parallel$ une norme quelconque dans $\mathbb{R}^p$. Soient (X_n) et X deux vecteurs aléatoires à valeurs dans $\mathbb{R}^p$. La suite (X_n) converge en probabilité vers X si, et seulement si,*

$$\parallel X_n - X \parallel \overset{\mathbb{P}}{\to} 0, quand\ \ n \to \infty.$$

Théorème 5.3.1. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$.*

$$Si\ \ X_n \overset{\mathbb{P}}{\to} X\ \ alors\ \ X_n \overset{\mathcal{L}}{\to} X.$$

Proposition 5.3.2. *Considérons des suites (X_n) et (Y_n) de v.a.r. Si on a les convergences :*

$$X_n \overset{\mathbb{P}}{\to} X$$

et

$$Y_n \overset{\mathbb{P}}{\to} Y$$

et si g est une fonction continue de $\mathbb{R}^2$ dans $\mathbb{R}$, alors

$$g(X_n, Y_n) \overset{\mathbb{P}}{\to} g(X, Y).$$

Proposition 5.3.3. *[14]. Soient (X_n) et (Y_n) deux v.a.r. Si on a les convergences :*

$$X_n \xrightarrow{\mathbb{P}} X$$

$$Y_n \xrightarrow{\mathbb{P}} Y,$$

alors

$$X_n + \lambda Y_n \xrightarrow{\mathbb{P}} X + \lambda Y \ \ pourtout \ \ \lambda \ \ dans \ \ \mathbb{R},$$

et

$$X_n + Y_n \xrightarrow{\mathbb{P}} X + \lambda Y.$$

Corollaire 5.3.4. *[14]. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires.*

$$Si \ \ X_n \xrightarrow{\mathbb{P}} c \ \ alors \ \ X_n \xrightarrow{\mathcal{L}} c.$$

5.3.1 Propriétés de la convergence en loi

Proposition 5.3.5. *Soit $f : \mathbb{R}^n \to \mathbb{R}^k$ une fonction mesurable. Si $X_n \xrightarrow{\mathcal{L}} X$ et si $\mathbb{P}(X \in C_g) = 1$, alors $f(X_n) \xrightarrow{\mathcal{L}} f(X)$.*

Démonstration. *On utilise le critère 2 de la proposition 5.2.2. Soit $g : \mathbb{R}^k \to \mathbb{R}$ continue bornée. La fonction gof est bornée et elle est continue sur C_f Par conséquent, on a*

$$\mathbb{P}(X \in C_{gof}) = 1 \ \ et \ \ gof : \mathbb{R}^n \to \mathbb{R}$$

est mesurable bornée donc d'après le critère 3 de la proposition 5.2.2,

$$\mathbb{E}_g(f(X_n)) \to \mathbb{E}_g(f(X)).$$

Exemple 5.3.1. • *Si $X_n \xrightarrow{\mathcal{L}} X \sim N(0,1)$ alors $\frac{1}{X_n} \xrightarrow{\mathcal{L}} \frac{1}{X}$ car*

$$\mathbb{P}\left(X \in \{x : x \to x^{-1} \ \ est \ \ continue \ \ en \ \ x \ \ \}\right).$$

• *Si $X_n \xrightarrow{\mathcal{L}} X \sim \mathcal{N}(0,1)$, alors*

$$X_n^2 \xrightarrow{\mathcal{L}} \chi_1^2.$$

Théorème 5.3.6. *Soit (X_n) et Xdes vecteurs aléatoires de $\mathbb{R}^n$, absolument continus de densité f_n et f par rapport à la mesure de Lebesgue dans $\mathbb{R}^n$. Si on a,*

$$\lim_n f_n = f,$$

alors

$$X_n \xrightarrow{\mathcal{L}} X.$$

5.4 Convergence en loi et fonction caractéristique

Soit (X_n) une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^n$. Pour tout $n \in \mathbb{N}$, on note ϕ_{X_n} la fonction caractéristique de X_n.

Théorème 5.4.1. *La suite de vecteurs aléatoires (X_n) converge en loi vers X si et sulement si pour tout $t \in \mathbb{R}^d$,*

$$\phi_{X_n}(t) \to \phi_X(t).$$

Théorème 5.4.2. ***(Théorème de Cramer-Wold)**. Soit (X_n) une suite de vecteurs alétoires à valeurs dans $\mathbb{R}^d$ avec $d > 1$. (X_n) converge en loi vers X si et sulement si pour tout $\lambda \in \mathbb{R}^d$,*

$$\lambda' X_n = \sum_{i=1}^{d} \lambda_i X_n^{(i)} \xrightarrow{\mathcal{L}} \sum_{i=1}^{d} \lambda_i X^{(i)} = \lambda' X.$$

5.5 Théorème central limite vectoriel

Le théorème suivant est fondamental et très souvent utilisé en Statistique. Notons $X_{n,1}, ..., X_{n,d}$ les d coordonnées d'un vecteurs X_n dans $\mathbb{R}^d$.
Notons $\overline{X_n}$ le vecteur des moyennes des composantes des n premiers vecteurs de la suite X_n, i.e

$$\overline{X_n} = \begin{pmatrix} \frac{1}{n}\sum_{j=1}^{n} X_{j,1} \\ . \\ . \\ . \\ \frac{1}{n}\sum_{j=1}^{n} X_{j,d} \end{pmatrix}.$$

Théorème 5.5.1. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires de $\mathbb{R}^d$ i.i.d. admettant un moment d'ordre 2. On note μ leur espérance et Σ leur matrice de variance-covariance. Alors,*

$$\sqrt{n}\left(\overline{X_n} - \mu\right) \xrightarrow{\mathcal{L}} \mathcal{N}_d\left(0, \Sigma_X\right).$$

Pour démontrer ce théorème nous utiliserons son cas particulier, correspondant au cas unidimensionnel.

Théorème 5.5.2. *Soit $(Z_n)_{n\in\mathbb{N}}$ une suite de v.a.r. indépendantes, dans L_2 et de même loi de moyenne μ et de variance σ^2. On a alors*

$$\sqrt{n}\left(\frac{\frac{1}{n}\sum_{j=1}^{n} Z_j - \mu}{\sigma}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Démonstration. *Notons*

$$\begin{aligned} Y_n &= \sqrt{n}\left(\frac{\frac{1}{n}\sum_{j=1}^n Z_j - \mu}{\sigma}\right) = \frac{\sqrt{n}}{n}\left(\sum_{j=1}^n \left(\frac{(Z_n-\mu)}{\sigma}\right)\right) \\ &= \sum_{j=1}^n U_j \end{aligned}$$

où les v.a.r U_j sont définies par :

$$U_j = \frac{Z_j - \mu}{\sigma},$$

pour $j = 1, ..., n$. Ces dernières sont, par hypothèse, centrées réduites, de même loi et indépendantes. On peut alors écrire :

$$\begin{aligned} \phi_{Y_n}(t) &= \mathbb{E}(e^{itY_n}) = \mathbb{E}\left(e^{it\frac{1}{\sqrt{n}}\sum_{j=1}^n U_j}\right) \\ &= \prod_{j=1}^n \phi_{U_j}\left(\frac{t}{\sqrt{n}})\right) = \left(\phi\left(\frac{t}{n}\right)\right)^n \end{aligned}$$

(car les v.a sont indépendantes et identiquement distribuées), où ϕ est la fonction caractéristique des v.a.r. U_j . Or, en utilisant les propriètès de la fonction caractéristique, on a :

$$\varphi'(0) = i\mathbb{E}(U_j) = 0$$

et

$$\phi''(0) = i^2\mathbb{E}\left(U_j^2\right) = -Var\left(U_j\right) = -1.$$

Le développement de Taylor de ϕ à l'ordre 2 et en 0, est alors :

$$\begin{aligned} \phi(u) &= \phi(0) + \phi'(0)u + \frac{1}{2}\phi''(0)u^2 + u^2\varepsilon(u) \\ &= 1 - \frac{u^2}{2} + u^2\varepsilon(u), \end{aligned}$$

avec ε une fonction telle que $\lim_{u\to 0}\varepsilon(u) = 0$. Ainsi, on peut écrire :

$$\phi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)$$

et

$$\phi_{Y_n}(t) = \left(1 - \frac{\frac{t^2}{2} + t^2\varepsilon\left(\frac{t}{\sqrt{n}}\right)}{n}\right)^n.$$

dont on tire aisément la convergence :

$$\phi_{Y_n}(t) \to e^{-\frac{t^2}{2}},$$

$$Y_n \xrightarrow{\mathcal{L}} Z,$$

où X est une variable aléatoire de loi $\mathcal{N}(0,1)$.

Démonstration. *du théorème 5.5.1. Pour tout u dans $\mathbb{R}^d$, notons $Z_n = u'X_n$. Par hypothèse, les v.a.r. constituant la suite (Z_n) sont donc indépendantes, de même loi, d'espérance $\mathbb{E}Z_n = u'u$ et de variance*

$$Var(Z_n) = u'\Sigma u.$$

En utilisant le thèorème unidimensionnel, il vient alors :

$$\sqrt{n}\left(\frac{\frac{1}{n}\sum_{j=1}^{n} Z_j - uu'}{\sqrt{u'\sum u}}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Cette convergence peut être réécrite sous la forme :

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} Z_j - uu'\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, u'\Sigma u\right) \Leftrightarrow u'.\sqrt{n}\left(\overline{X_n} - u\right) \to^{\mathcal{I}} u'X.$$

où X est un vecteur aléatoire gaussien $N_d(0,\Sigma)$.

Ce résultat étant vrai pour tout u dans $\mathbb{R}^d$, le théorème de Cramer-Wold nous permet de conclure que

$$\sqrt{n}\left(\overline{X_n} - \mu\right) \xrightarrow{\mathcal{L}} X,$$

quand n tend vers $+\infty$.

5.6 Théorème de Cramer

Théorème 5.6.1. *Soit g de $\mathbb{R}^d$ dans $\mathbb{R}^k$ une fonction de classe C^1 au voisinage de μ. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de vecteurs aléatoires dans $\mathbb{R}^d$ telle que*

$$\sqrt{n}(U_n - \mu) \xrightarrow{\mathcal{L}} U \tag{5.1}$$

alors

$$\sqrt{n}(g(U_n) - g(\mu)) \xrightarrow{\mathcal{L}} (J(g(\mu))^T U. \tag{5.2}$$

le jacobien est :

$$\nabla(g(\mu)) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & . \quad . \quad . & \frac{\partial g_k}{\partial x_1} \\ . & & . \\ . & & . \\ . & & . \\ \frac{\partial g_1}{\partial x_d} & . \quad . & \frac{\partial g_k}{\partial x_d} \end{pmatrix}.$$

Corollaire 5.6.2. *Soit g de $\mathbb{R}^d$ dans $\mathbb{R}^k$ une fonction de classe C^1 au voisinage de μ. Si*

$$\sqrt{n}(X_n - X) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \Sigma\right) \tag{5.3}$$

alors

$$\sqrt{n}(g(X_n) - g(X)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \nabla(g(\mu))^t \sigma \nabla(g(\mu))\right). \tag{5.4}$$

Exemple 5.6.1. *Si $(X_n)_{n\in N}$ est une suite iid de L^2. Posons $\overline{X_n} = \frac{1}{n}\sum_{j=1}^n X_j$. D'après le théorème central limite,*

$$\sqrt{n}\left(\overline{X_n} - \mu\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2\right)$$

d'où, pour toute fonction g de classe C^1 au voisinage de μ , on a

$$\sqrt{n}(g(\overline{X_n}) - g(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, (\nabla(g(\mu))^2\sigma^2\right)$$

En prenant $g(x) = x^2$, on obtient

$$\sqrt{n}(X_n^2 - \mu^2) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\mu^2\sigma^2\right).$$

La limite est une loi gaussienne dégénérée si $\mu = 0$.

5.7 Théorème de Cochran, lois du χ^2.

Dans tout ce paragraphe, nous nous placerons dans $\mathbb{R}^n$ muni du produit scalaire euclidien et on notera $\|\ .\ \|$ la norme euclidienne dans $\mathbb{R}^n$.

Théorème 5.7.1. *Soit X un vecteur gaussien de $\mathbb{R}^n$ tel que $\mu = \mathbb{E}(X)$ et $Var(X) = I_d$. La loi de $\|\ X\ \|^2$ ne dépend que de n et $\|\ .\ \|$. On note*

$$\|\ X\ \|^2 \sim \chi^2\left(n, \|\ \mu\ \|^2\right).$$

et on dit que $\|\ X\ \|^2$ suit une loi du χ^2 (qui est décentrée si $\|\ \mu\ \| \neq 0$). L'entier d est le nombre de degrés de liberté. Lorsque $\|\ \mu\ \| = 0$, on note plus simplement $\|\ \mu\ \|^2 \sim (d)$.

Démonstration. *Soit $Y \in \mathbb{R}^d$ tel que $Y \rightsquigarrow \mathcal{N}(\mu', I_d)$ avec $\|\ \mu\ \| = \|\ \mu'\ \|$*

Il existe U matrice orthogonale telle que $\mu = U\mu'$. Donc $UY \sim N(\mu; I_d) \backsim X$ et

$$\|\ Y\ \|^2 = \|\ UY\ \|^2 \rightsquigarrow \|\ X\ \|^2$$

Théorème 5.7.2. ***(de Cochran)****. Soit $E_1 \oplus ... \oplus E_r$ une décomposition de $\mathbb{R}^d$ en sous-espaces deux à deux orthogonaux de dimensions respectives $d_1, d_2, ..., d_r$. Si $X \sim \mathcal{N}(\mu, I_n)$, Les vecteurs aléatoires $X_{E_1}, ..., X_{E_r}$, projections orthogonales de X sur $E_1 \oplus ... \oplus E_r$ sont indépendantes, les variables aléatoires $\|\ X_{E_1}\ \|^2, ..., \|\ X_{E_r}\ \|^2$ sont indépendantes et*

$$\left(\|\ X_{E_1}\ \|^2, ..., \|\ X_{E_r}\ \|^2\right) \longrightarrow \left(\chi^2(d_1, \mu_{\|E_1\|}), ..., \chi^2(d_r, \mu_{\|E_r\|})\right),$$

où $\mu_{E_1}, ..., \mu_{E_r}$ sont les projections de μ sur $E_1, ..., E_r$.

5.8 Exercices du chapitre 5.

Exercice 5.8.1. ***Théorème central limite multidimensionnel.***

1. *Soit* $X = (X_1, ..., X_n)$ *un vecteur aléatoire à valeur dans* $\mathbb{R}^n$*. Montrer que sa fonction caractéristique* ϕ_X *admet le développement limité suivant :*

$$\phi_X(t) = 1 + i\sum_{j}^{n} t_j \mathbb{E}(X_j) - \sum_{k,j}^{n} \mathbb{E}(X_k X_j) t_j t_k + o(\| t \|^2).$$

2. *Soit* $(X_n)_{n\geq 1}$ *une suite de vecteurs aléatoires à valeurs dans* $\mathbb{R}^n$*, indépendants, centrés de même loi. On note* $\sum$ *leur matrice de covariance commune et* $S_n = X_1 + ... + X_n$*.*
   – *Calculer la fonction caractéristique* $\frac{S_n}{\sqrt{n}}$ *en fonction de celle de* X_1*.*
   – *Montrer que* $\frac{S_n}{\sqrt{n}}$ *converge en loi et déterminer la loi limite.*

Exercice 5.8.2. ***Application du Théorème de Cochran.***
Dans cet exercice $X = (X_1, X_2, X_3, X_4)^t$ *désigne un vecteur gaussien de moyenne* $(0, 0, 0, 0)^t$ *et de matrice de variance-covariance identité.*

1. *Quelle est la loi de* $\frac{(X_2-X_1)^2}{2} + \frac{(X_2+X_1)^2}{2}$ *?*
2. *Quelle est la loi de* $\frac{(X_2-X_1)^2}{(X_2+X_1)^2}$
3. *Quel est le projeté orthogonal* $P_E(X)$ *de* X *sur l'espace vectoriel engendré par les vecteurs* $(1, 1, 0, 0)^t$ *et* $(0, 0, 1, 1)^t$ *?*
4. *Quelle est la loi de* $\| P_E(X) \|^2$*?*
5. *Quelle est la loi de* $\frac{\|P_E(X)\|^2}{\|X-P_E(X)\|^2}$*?*

Exercice 5.8.3. ***Convergence en probabiité et convergence en loi.***

1. *Soient* $(X_n)_{n\in\mathbb{N}}$ *et* $(Y_n)_{n\in\mathbb{N}}$ *deux suites de variables aléatoires. On suppose que* $(X_n)_{n\in\mathbb{N}}$ *converge en loi vers une variable* X *et* $(Y_n)_{n\in\mathbb{N}}$ *converge en loi vers une constante déterministe* c*.*
   Montrer que la suite de vecteurs aléatoires $(X_n, Y_n)_{n\in\mathbb{N}}$ *converge en loi vers* (X, c)*.*
2. *Application : Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires indépendantes identiquement distribuées, de moyenne* μ *et de variance finie. On note* $\overline{X_n} = \frac{X_1+...+X_n}{n}$ *leur moyenne empirique et* $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ *leur variance empirique. On pose* $T_n = \sqrt{n}\frac{\overline{X_n}-\mu}{S_n}$*. Montrer que la suite*
   (T_n) *converge en loi vers une variable aléatoire dont on déterminera la loi.*
3. *Trouver un exemple de deux suites de variables aléatoires* $(X_n)_{n\in\mathbb{N}}$ *et* $(Y_n)_{n\in\mathbb{N}}$ *qui convergent chacune en loi mais telles que la suite de couples de variables aléatoires* $(X_n, Y_n)_{n\in\mathbb{N}}$ *ne converge pas en loi.*

Bibliographie

[1] *Anne, P., Claude, M.* Cours de Probabilités : Modèles et Applications, *Université de Nantes, 2010.*

[2] *Barbe, P., Ledoux, M.* Probabilités, *EDP Sciences, 2007.*

[3] *Bogaert, P.* Introduction au calcul des probabilités, cours et exercices corrigés, *Boeck et Lacier s.a , 2006.*

[4] *Billingsley, P.* Probabilities. 2. Measure theory. I. Title. II. Series. QA273.B575 1995.

[5] *Delmas, J,F., Jourdain, B.* Modèles aléatoires : applications aux sciences de l'ingénieur et du vivant, *Springer, 2006.*

[6] *Feller, W.* An introduction to Probability Theory and its Applications, Vol. 1. 3rd Edition, Wiley series in probabilities and mathematical statistics, 1968.

[7] *Foata, D., Fuchs, A.* Calcul des probabilités : 2 ème édition, *Dunod,2003.*

[8] *Ghorbanzandh, D.* Probabilités, Exercices corrigés, *Edition Techninip, Paris.*

[9] *Jacod, J., Protter, P.* L'essentiel en théorie des probabilités, *Cassini,2003.*

[10] *Jacques, J.* Vecteurs Gaussiens, *Bordeaux 1, 2014.*

[11] *Jean, Y.* Probabilités 2,ISBN 978-2-84225-144-4 *Cassini.*

[12] *Laurent, R.* Cours de Probabilités générales, *Université Rennes 2.*

[13] *Méléard, S.* Introduction à la théorie et au calcul des probabilités, *École Polytechnique, Paris.*

[14] *Monfort, A.* Cours de Probabilités, *Economica, 1996.*

[15] *Pastor, D., Sinte, C.* Probabilités pour l'ingénieur, des fondaments aux calculs, *Lavoisier, 1998.*