

Université Rennes 2
Licence MASS 3

Espérance conditionnelle & Chaînes de Markov

Arnaud GUYADER

Table des matières

1	Espérance conditionnelle	1
1.1	Cas discret	1
1.2	Cas absolument continu	7
1.3	Applications	14
1.3.1	Probabilités conditionnelles	14
1.3.2	La régression	15
1.4	Interprétation géométrique de l'espérance conditionnelle	20
1.5	Espérance conditionnelle : le cas général	27
1.6	Exercices	28
2	Vecteurs gaussiens et conditionnement	63
2.1	Rappels sur les vecteurs gaussiens	63
2.1.1	Variables gaussiennes	63
2.1.2	Vecteurs gaussiens	67
2.2	Conditionnement des vecteurs gaussiens	78
2.2.1	Conditionnement pour un couple gaussien	78
2.2.2	Hyperplan de régression	80
2.2.3	Espérance conditionnelle gaussienne	83
2.3	Exercices	86
3	Chaînes de Markov	113
3.1	Définition d'une chaîne de Markov	113
3.2	Equations de Chapman-Kolmogorov	115
3.3	Classification des états	119
3.4	Comportement asymptotique	123
3.5	Exercices	133
A	Annales	147

Chapitre 1

Espérance conditionnelle

Introduction

L'espérance conditionnelle est un outil d'usage constant en probabilités et statistiques. Néanmoins, sa définition dans le cas général n'est pas simple. C'est pourquoi ce chapitre présente l'idée par étapes et de façon intuitive : cas discret, cas absolument continu, interprétation géométrique dans L^2 et enfin extension à L^1 .

1.1 Cas discret

On considère un couple aléatoire discret (X, Y) , c'est-à-dire une application mesurable

$$(X, Y) : \begin{cases} (\Omega, \mathcal{F}, \mathbb{P}) & \rightarrow \mathcal{X} \times \mathcal{Y} \\ \omega & \mapsto (X(\omega), Y(\omega)) \end{cases}$$

avec les ensembles $\mathcal{X} = (x_i)_{i \in I}$ et $\mathcal{Y} = (y_j)_{j \in J}$ au plus dénombrables (i.e. finis ou dénombrables). Autrement dit, les ensembles d'indices I et J sont au plus dénombrables : penser à des ensembles finis, à \mathbb{N} , à \mathbb{Z} . Pour calculer des quantités liées à ce couple aléatoire, il faut bien sûr connaître la probabilité de tomber sur un couple (x_i, y_j) . Nous adoptons la notation :

$$p_{ij} = \mathbb{P}(X = x_i, Y = y_j).$$

La suite double $(p_{ij})_{i \in I, j \in J}$ est appelée loi jointe du couple (X, Y) . Il est clair que :

$$\begin{cases} 0 \leq p_{ij} \leq 1 \\ \sum_{i \in I, j \in J} p_{ij} = 1 \end{cases}$$

Exemple. On tire deux chiffres au hasard, indépendamment et de façon équiprobable entre 1 et 3. Soit X le maximum des chiffres obtenus et Y la somme des chiffres obtenus. La loi jointe du couple (X, Y) se représente sous forme d'un tableau (voir figure 1.1).

Définition 1 (Lois marginales)

Soit (X, Y) un couple aléatoire. Les variables aléatoires X et Y sont dites marginales. La loi de X , dite loi marginale, est entièrement déterminée par les probabilités $p_{i.}$ de tomber sur les points x_i :

$$p_{i.} = \mathbb{P}(X = x_i) = \sum_{j \in J} \mathbb{P}(X = x_i, Y = y_j) = \sum_{j \in J} p_{ij}$$

De même pour la loi marginale de Y et les probabilités $p_{.j}$ de tomber sur les points y_j :

$$p_{.j} = \mathbb{P}(Y = y_j) = \sum_{i \in I} \mathbb{P}(X = x_i, Y = y_j) = \sum_{i \in I} p_{ij}$$

$X \backslash Y$	2	3	4	5	6
1	$\frac{1}{9}$	0	0	0	0
2	0	$\frac{2}{9}$	$\frac{1}{9}$	0	0
3	0	0	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

FIGURE 1.1 – Loi jointe pour le max et la somme.

Exemple. Pour l'exemple précédent, on calcule aisément les lois marginales de X et Y : il suffit de sommer sur chaque ligne pour la loi de X et sur chaque colonne pour la loi de Y (voir figure 1.1).

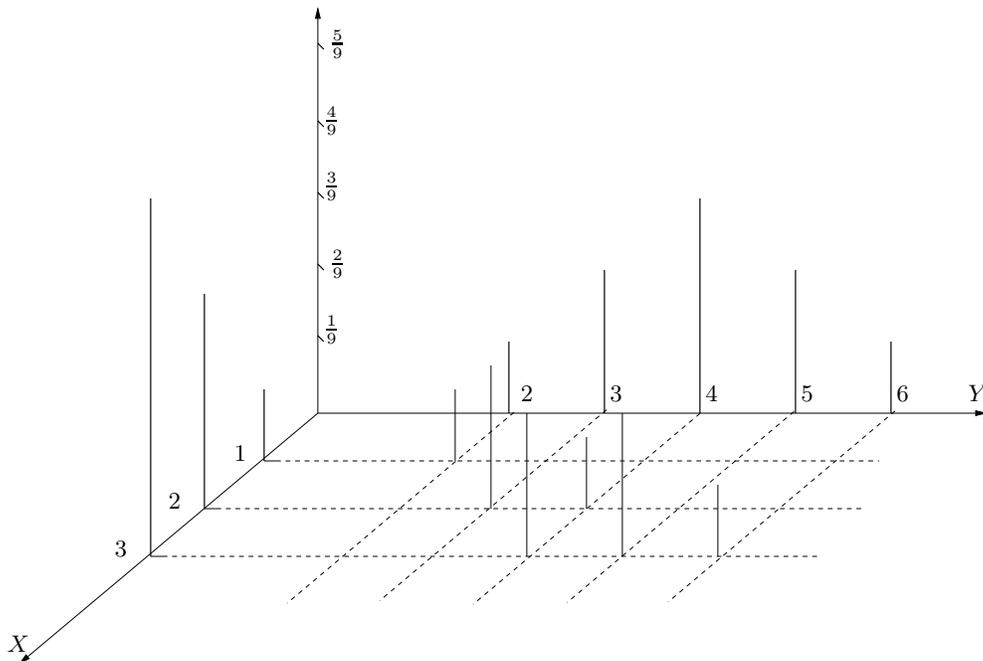


FIGURE 1.2 – Loi jointe et lois marginales pour le max et la somme.

Achtung! La connaissance des lois marginales ne suffit pas à déterminer la loi du couple (X, Y) . Autrement dit, on peut trouver deux couples (X_1, Y_1) et (X_2, Y_2) n'ayant pas même loi jointe, mais tels que les lois de X_1 et X_2 soient égales, ainsi que les lois de Y_1 et Y_2 (cf. figure 1.3).

La situation agréable est celle où les variables marginales X et Y sont indépendantes. Celle-ci se vérifie facilement une fois connues la loi jointe et les lois marginales.

Proposition 1 (Lois marginales et indépendance)

Les variables aléatoires marginales X et Y sont indépendantes si et seulement si :

$$\forall (i, j) \in I \times J \quad \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j),$$

Y ₁ X ₁ \	1	2	3
1	0	$\frac{1}{4}$	0
2	$\frac{1}{4}$	0	$\frac{1}{4}$
3	0	$\frac{1}{4}$	0

Y ₂ X ₂ \	1	2	3
1	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
2	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
3	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$

FIGURE 1.3 – Mêmes lois marginales mais loi jointe différente.

c'est-à-dire avec nos notations : $p_{ij} = p_i p_j$.

Exemples :

1. Sur l'exemple précédent du max et de la somme, il est clair que X et Y ne sont pas indépendantes puisque par exemple :

$$p_{12} = \frac{1}{9} \neq p_{1.} \times p_{.2} = \frac{1}{9} \times \frac{1}{9} = \frac{1}{81}$$

2. Jeu de cartes : on tire une carte au hasard dans un jeu de 32 cartes. Le résultat de ce tirage est représenté par le couple aléatoire (X, Y) , où X est la couleur et Y la valeur. Autrement dit, X appartient à l'ensemble {Pique, Cœur, Carreau, Trèfle} et Y à l'ensemble {7, 8, 9, 10, Valet, Dame, Roi, As}. Il est clair que :

$$\forall (i, j) \in I \times J \quad \mathbb{P}(X = x_i, Y = y_j) = \frac{1}{32} = \frac{1}{4} \times \frac{1}{8} = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j),$$

donc X et Y sont indépendantes.

Remarque. Soit $i \in I$ fixé. Notons qu'on peut avoir $p_{ij} = 0$, c'est-à-dire que l'événement $\{X = x_i, Y = y_j\}$ ne se réalise jamais. Par contre, on exclut le cas où $p_{i.} = 0$: ceci signifierait que X ne prend jamais la valeur x_i , auquel cas cette valeur n'aurait rien à faire dans \mathcal{X} .

Puisque chacune des probabilités $p_{i.}$ est non nulle, on peut définir la probabilité conditionnelle de $Y = y_j$ sachant $X = x_i$ par la formule :

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

Définition 2 (Probabilités conditionnelles)

Soit $x_i \in \mathcal{X}$. La loi conditionnelle de Y sachant $X = x_i$ est la loi discrète prenant les valeurs y_j avec les probabilités $p_{j|i} = \mathbb{P}(Y = y_j | X = x_i)$.

Exemple : Lois de Poisson.

Soit $Y \sim \mathcal{P}(\alpha)$ et $Z \sim \mathcal{P}(\beta)$ deux variables aléatoires de Poisson indépendantes. On s'intéresse à leur somme $X = Y + Z$. X est bien sûr une variable aléatoire. On rappelle que Y suit une loi de Poisson de paramètre α si Y est à valeurs dans \mathbb{N} , avec (voir aussi figure 1.4) :

$$\forall n \in \mathbb{N} \quad \mathbb{P}(Y = n) = e^{-\alpha} \frac{\alpha^n}{n!}.$$

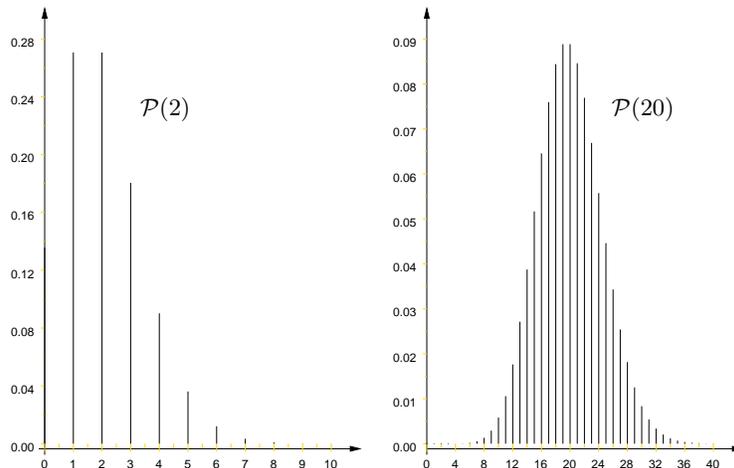


FIGURE 1.4 – Lois de Poisson de paramètres 2 et 20.

1. Loi de X ?

La variable aléatoire X est à valeurs dans \mathbb{N} en tant que somme de variables aléatoires à valeurs dans \mathbb{N} . On commence par déterminer sa loi. Soit donc $n \in \mathbb{N}$ fixé, alors :

$$\mathbb{P}(X = n) = \mathbb{P}(Y + Z = n) = \mathbb{P}\left(\bigcup_{k=0}^n \{Y = k, Z = n - k\}\right) = \sum_{k=0}^n \mathbb{P}(Y = k, Z = n - k).$$

Or Y et Z sont indépendantes, donc :

$$\mathbb{P}(X = n) = \sum_{k=0}^n \mathbb{P}(Y = k)\mathbb{P}(Z = n - k) = \sum_{k=0}^n e^{-\alpha} \frac{\alpha^k}{k!} e^{-\beta} \frac{\beta^{n-k}}{(n-k)!} = \frac{e^{-(\alpha+\beta)}}{n!} \sum_{k=0}^n \binom{n}{k} \alpha^k \beta^{n-k},$$

et on reconnaît la formule du binôme :

$$\mathbb{P}(X = n) = e^{-(\alpha+\beta)} \frac{(\alpha + \beta)^n}{n!}.$$

C'est-à-dire que X suit une loi de Poisson de paramètre $(\alpha + \beta)$. Ce résultat se généralise d'ailleurs sans problème : si les $X_i \sim \mathcal{P}(\lambda_i)$ sont globalement indépendantes, alors leur somme S suit encore une loi de Poisson :

$$S \sim \mathcal{P}(\lambda_1 + \dots + \lambda_n).$$

D'un point de vue théorie de la mesure, on vient de montrer que le produit de convolution de deux lois de Poisson est une loi de Poisson.

2. Loi de Y sachant X ?

Soit $n \in \mathbb{N}$, déterminons la loi de Y sachant $X = n$. Puisque $X = Y + Z$, il est clair que, sachant $X = n$, Y est à valeurs dans $\{0, 1, \dots, n\}$. Soit donc $k \in \{0, 1, \dots, n\}$:

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

Et il suffit alors d'exprimer tout ceci grâce aux lois de Poisson, ce qui donne après simplifications :

$$\mathbb{P}(Y = k | X = n) = \binom{n}{k} \left(\frac{\alpha}{\alpha + \beta}\right)^k \left(\frac{\beta}{\alpha + \beta}\right)^{n-k}.$$

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

Revenons au cas général et supposons que Y soit intégrable. Si X est figée à x_i , il est naturel de considérer la valeur moyenne de la variable aléatoire Y lorsque $X = x_i$: c'est ce qu'on appelle l'espérance conditionnelle de Y sachant $X = x_i$. Elle s'écrit :

$$\mathbb{E}[Y|X = x_i] = \sum_{j \in J} p_{j|i} y_j$$

Or on sait que X prend la valeur $X = x_i$ avec la probabilité p_i .

Définition 3 (Espérance conditionnelle)

Supposons Y 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 $\mathbb{E}[Y|X]$.

Nota Bene. Il faut noter qu'en général l'espérance conditionnelle $\mathbb{E}[Y|X]$ est une **variable aléatoire** et non un nombre. On peut l'interpréter comme la valeur moyenne prise par Y lorsque l'on connaît X . Elle pourra donc s'écrire comme une fonction de X .

Exemple. On reprend l'exemple précédent. L'espérance de Y sachant $X = n$ est l'espérance d'une loi binômiale $\mathcal{B}(n, \frac{\alpha}{\alpha+\beta})$. Donc pour tout $n \geq 0$:

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

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

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

qui est bien une fonction de X , donc une variable aléatoire, et non un nombre.

Théorème 1 (Calcul d'espérance par conditionnement)

Si Y est intégrable, alors la variable aléatoire $\mathbb{E}[Y|X]$ aussi et on a :

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

Exemple. Sur l'exemple précédent, les paramètres α et β étant des constantes, on peut écrire :

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

or l'espérance d'une loi de Poisson de paramètre $(\alpha + \beta)$ est tout simplement $(\alpha + \beta)$, donc :

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

Preuve. C'est l'âne qui trotte :

$$\mathbb{E}[\mathbb{E}[Y|X]] = \sum_{i \in I} p_i \mathbb{E}[Y|X = x_i] = \sum_{i \in I} p_i \left(\sum_{j \in J} p_{j|i} y_j \right),$$

or $p_{j|i} = \frac{p_{ij}}{p_i}$, donc :

$$\mathbb{E}[\mathbb{E}[Y|X]] = \sum_{i \in I} \left(\sum_{j \in J} p_{ij} y_j \right) = \sum_{j \in J} \left(\sum_{i \in I} p_{ij} \right) y_j,$$

or, par définition, $p_{.j} = \sum_{i \in I} p_{ij}$, donc :

$$\mathbb{E}[\mathbb{E}[Y|X]] = \sum_{j \in J} p_{.j} y_j = \mathbb{E}[Y].$$

■

Remarque. Ce résultat permet souvent de calculer l'espérance de Y en deux étapes : on exprime d'abord $\mathbb{E}[Y|X]$ comme une fonction $\varphi(X)$ de la variable aléatoire X . Puis, si cette fonction φ et la loi de X sont "assez simples", on calcule $\mathbb{E}[\varphi(X)]$. Voir par exemple les exercices "Un dé et une pièce" et "Somme aléatoire de variables aléatoires" en fin de chapitre.

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

Propriétés 1 (Espérance conditionnelle et indépendance)

Si Y est intégrable, si X et Y sont indépendantes, alors la variable aléatoire $\mathbb{E}[Y|X]$ est constante, égale à $\mathbb{E}[Y]$.

Preuve. Si X et Y sont indépendantes, alors pour tout couple $(i, j) \in I \times J$:

$$p_{ij} = p_i \cdot p_{.j}.$$

On en déduit que :

$$\forall (i, j) \in I \times J \quad p_{j|i} = p_{.j},$$

donc pour tout $x_i \in \mathcal{X}$:

$$\mathbb{E}[Y|X = x_i] = \sum_{j \in J} p_{j|i} y_j = \sum_{j \in J} p_{.j} y_j = \mathbb{E}[Y],$$

or par définition $\mathbb{E}[Y|X]$ est la variable aléatoire qui prend les valeurs $\mathbb{E}[Y|X = x_i]$ avec les probabilités p_i . On en déduit que $\mathbb{E}[Y|X]$ est la variable aléatoire constante égale à $\mathbb{E}[Y]$.

■

Dans de nombreuses situations, on désire calculer la valeur moyenne prise par une fonction du couple (X, Y) , c'est-à-dire :

$$\mathbb{E}[h(X, Y)] = \int_{\Omega} h(X, Y) d\mathbb{P},$$

où h est une fonction de \mathbb{R}^2 dans \mathbb{R} . Par exemple si on veut calculer la moyenne de la somme de deux variables, ou la moyenne de leur produit.

Rappel : Théorème de transfert

Sous réserve d'intégrabilité, le théorème de transfert assure que l'espérance précédente s'écrit comme une somme double :

$$\mathbb{E}[h(X, Y)] = \sum_{i \in I, j \in J} h(x_i, y_j) p_{ij}$$

Le cas simple est celui où, d'une, h se décompose en produit :

$$h(x, y) = f(x)g(y),$$

et, de deux, X et Y sont indépendantes. Dans ce cas, on a immédiatement :

$$\mathbb{E}[h(X, Y)] = \left(\sum_{i \in I} f(x_i) p_i \right) \left(\sum_{j \in J} g(y_j) p_{.j} \right) = \mathbb{E}[f(X)] \mathbb{E}[g(Y)],$$

c'est-à-dire qu'il suffit de calculer deux espérances discrètes classiques.

Dans le cas général, h ne se décompose pas aussi simplement et les variables X et Y ne sont pas indépendantes. Néanmoins, sous les hypothèses usuelles d'intégrabilité, on peut toujours écrire :

$$\mathbb{E}[h(X, Y)] = \sum_{i \in I} \left(\sum_{j \in J} h(x_i, y_j) p_{j|i} \right) p_i = \sum_{i \in I} \mathbb{E}[h(x_i, Y) | X = x_i] \mathbb{P}(X = x_i)$$

Ceci est une autre façon de dire que :

$$\mathbb{E}[h(X, Y)] = \mathbb{E}[\mathbb{E}[h(X, Y) | X]]$$

où $\mathbb{E}[h(X, Y) | X]$ est la variable aléatoire qui prend les valeurs $\mathbb{E}[h(x_i, Y) | X = x_i]$ avec les probabilités p_i . On a ainsi ramené le calcul d'une somme double à deux calculs de sommes simples.

1.2 Cas absolument continu

Pour une variable aléatoire réelle X , les deux situations classiques sont les suivantes : X est discrète ou X est absolument continue, c'est-à-dire qu'elle admet une densité. Dans le paragraphe précédent, on a vu le pendant d'une loi discrète pour un couple aléatoire. Etudions maintenant l'analogie d'une loi absolument continue pour un couple aléatoire (X, Y) à valeurs dans \mathbb{R}^2 (ou un sous-ensemble de \mathbb{R}^2).

Par définition, la loi jointe $P_{X,Y}$ du couple est la mesure de probabilité sur $(\mathbb{R}^2, \mathcal{B}_2)$ définie par :

$$\forall B \in \mathcal{B}_2, \quad P_{X,Y}(B) = \mathbb{P}((X, Y) \in B),$$

que l'on peut voir comme la probabilité que le point aléatoire M de coordonnées (X, Y) tombe dans l'ensemble borélien B .

Définition 4 (Loi jointe absolument continue)

On dit que la loi $P_{X,Y}$ est absolument continue¹ s'il existe une fonction mesurable $f : (\mathbb{R}^2, \mathcal{B}_2) \rightarrow (\mathbb{R}, \mathcal{B})$ telle que :

$$\forall B \in \mathcal{B}_2, \quad P_{X,Y}(B) = \iint_B f(x, y) dx dy.$$

La fonction f est appelée densité de probabilité du couple (X, Y) . On la note parfois $f_{X,Y}$.

Pour qu'une fonction f soit une densité de probabilité, il faut et il suffit qu'elle soit positive et intégrée à 1 :

$$\begin{cases} f(x, y) \geq 0 \\ \iint_{\mathbb{R}^2} f(x, y) dx dy = 1 \end{cases}$$

Remarque. En pratique, dans tout ce paragraphe, on peut faire le parallèle avec ce qui a été vu dans le cas discret : il suffit de remplacer x_i par x , y_j par y , p_{ij} par $f(x, y)$ et les sommes par des intégrales.

Exemple. On considère un couple (X, Y) de densité :

$$f(x, y) = 2e^{-(x+y)} \mathbb{1}_{\{0 \leq x \leq y\}}$$

1. sous-entendu : par rapport à la mesure de Lebesgue sur \mathbb{R}^2 .

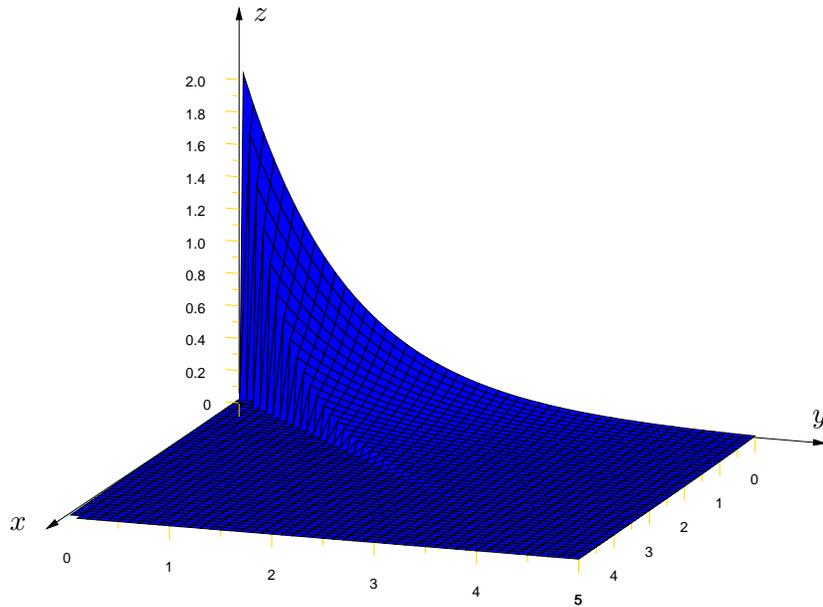


FIGURE 1.5 – Représentation de la densité jointe $f(x, y) = 2e^{-(x+y)} \mathbb{1}_{\{0 \leq x \leq y\}}$.

On vérifie que ceci définit bien une densité de probabilité sur \mathbb{R}^2 . En effet, f est positive et par le théorème de Fubini-Tonelli, on a pour le calcul de l'intégrale double :

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_0^{+\infty} \left(\int_0^y 2e^{-(x+y)} dx \right) dy = \int_0^{+\infty} \left(\int_x^{+\infty} 2e^{-(x+y)} dy \right) dx.$$

Prenons par exemple la première expression :

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_0^{+\infty} 2e^{-y} [-e^{-x}]_0^y dy = \int_0^{+\infty} (2e^{-y} - 2e^{-2y}) dy,$$

ce qui donne finalement :

$$\int_{\mathbb{R}^2} f(x, y) dx dy = [-2e^{-y} + e^{-2y}]_0^{+\infty} = 1.$$

La représentation de la densité f est donnée figure 1.5.

Comme dans le cas discret, on peut définir les lois des variables aléatoires marginales X et Y .

Proposition 2 (Lois marginales)

Si le couple (X, Y) est absolument continu, les variables marginales X et Y sont absolument continues et la densité jointe $f(x, y)$ détermine les densités marginales $f(x)$ et $f(y)$:

$$f(x) = f_X(x) = \int_{\mathbb{R}} f(x, y) dy \quad \& \quad f(y) = f_Y(y) = \int_{\mathbb{R}} f(x, y) dx$$

Convention. Suivant le contexte, la densité marginale de X sera notée $f(x)$ ou f_X , mais rarement $f_X(x)$, qui est lourdingue. Idem pour Y .

Exemple. Pour l'exemple précédent, on obtient (voir figure 1.6) :

$$\begin{cases} f(x) = 2e^{-2x} \mathbb{1}_{[0,+\infty[}(x) \\ f(y) = 2e^{-y}(1 - e^{-y}) \mathbb{1}_{[0,+\infty[}(y) \end{cases}$$

Une fois connues les lois marginales, on peut effectuer les calculs usuels sur les variables aléatoires absolument continues. Par exemple, sous réserve d'intégrabilité, l'espérance de X est alors simplement :

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx.$$

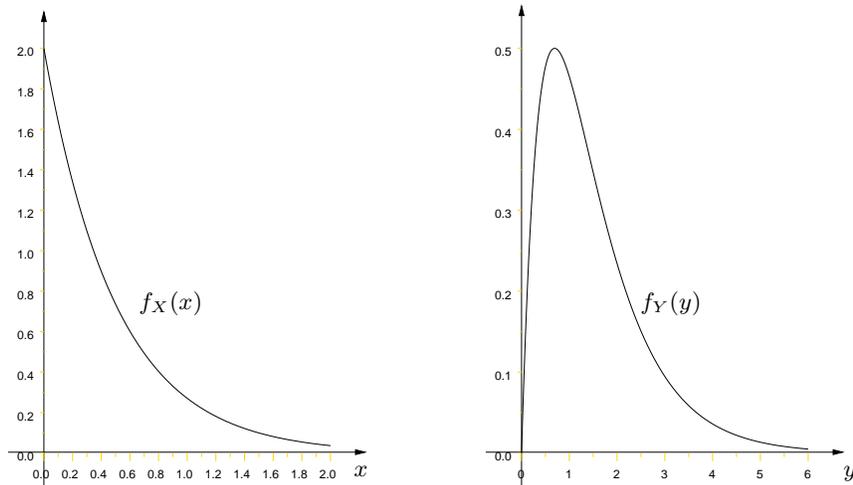


FIGURE 1.6 – Représentation des densités marginales $f(x)$ et $f(y)$.

Chausse-trappe². Pour l'exemple précédent, puisque X suit une loi exponentielle $\mathcal{E}(2)$, on a $\mathbb{E}[X] = \frac{1}{2}$. On rappelle au passage que si $X \sim \mathcal{E}(\lambda)$, c'est-à-dire si X a pour densité :

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,+\infty[}(x),$$

alors $\mathbb{E}[X] = \frac{1}{\lambda}$. Les Anglo-Saxons adoptent la convention inverse : pour eux, la variable aléatoire T suit une loi exponentielle de paramètre θ si T a pour densité :

$$f(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}} \mathbb{1}_{[0,+\infty[}(t),$$

auquel cas on a bien sûr tout simplement $\mathbb{E}[T] = \theta$. Lorsqu'on veut simuler des lois exponentielles à l'aide d'un logiciel, il faut donc faire attention à la convention utilisée par celui-ci.

Dans le cas général, par définition, les variables aléatoires X et Y sont indépendantes si pour tout couple de boréliens B et B' de \mathbb{R} , on a :

$$\mathbb{P}(X \in B, Y \in B') = \mathbb{P}(X \in B)\mathbb{P}(Y \in B'),$$

ou encore si pour toutes fonctions bornées (ou positives) g et h de \mathbb{R} dans \mathbb{R} :

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)].$$

Si la loi jointe est absolument continue, l'indépendance se vérifie de façon simple.

2. Les cuistres écrivent plutôt "chasse-trape", les deux orthographe étant acceptées.

Proposition 3 (Indépendance)

Avec les notations précédentes, les variables aléatoires X et Y sont indépendantes si et seulement si pour tout couple $(x, y) \in \mathbb{R}^2$:

$$f(x, y) = f(x)f(y).$$

Exemple. Pour l'exemple précédent, X et Y ne sont pas indépendantes puisque :

$$f(0, 0) = 2 \neq f_X(0)f_Y(0) = 0.$$

Remarque. Le raisonnement sur les supports permet parfois de conclure rapidement à la non-indépendance. Le support de la loi de X est l'adhérence de l'endroit où X a des chances de tomber :

$$\text{Supp}(X) = \text{Adh}\{x \in \mathbb{R} : f(x) \neq 0\}.$$

C'est généralement un intervalle fermé I_X . On définit de même le support I_Y de la loi de Y . Mais alors, si X et Y sont indépendantes, le support du couple (X, Y) est $I_X \times I_Y$, produit cartésien de I_X par I_Y . C'est-à-dire, en général, un pavé (fermé) de \mathbb{R}^2 . Donc si le support du couple (X, Y) n'est pas un pavé, X et Y ne sont pas indépendantes.

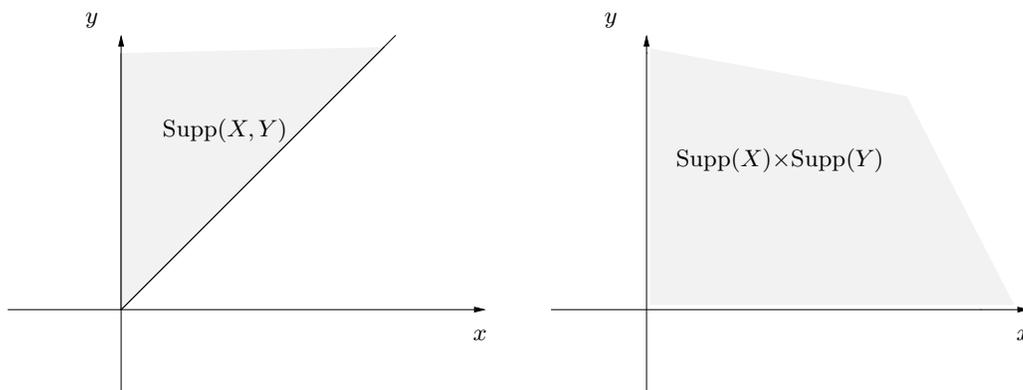


FIGURE 1.7 – Support du couple (X, Y) (à gauche) et produit cartésien des supports de X et de Y (à droite).

Exemple. Pour l'exemple précédent, le support de X est égal au support de Y , à savoir $\mathbb{R}^+ = [0, +\infty[$. Donc si X et Y étaient indépendantes, le support du couple (X, Y) serait le pavé $\mathbb{R}^+ \times \mathbb{R}^+$. Or le support de (X, Y) est :

$$\text{Supp}(X, Y) = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq y\} \neq \mathbb{R}^+ \times \mathbb{R}^+,$$

donc X et Y ne sont pas indépendantes (voir figure 1.7).

On veut maintenant définir l'analogue des probabilités conditionnelles vues dans le cas discret.

Définition 5 (Lois conditionnelles)

La densité conditionnelle de Y sachant $X = x$ est :

$$f(y|x) = f_{Y|X}(y|x) = \begin{cases} \frac{f(x,y)}{f(x)} & \text{si } f(x) > 0 \\ 0 & \text{si } f(x) = 0 \end{cases}$$

Interprétation graphique. Pour la représentation de la densité conditionnelle $f(y|x_0)$, il suffit de faire une coupe de la surface définie par la densité jointe $f(x, y)$ par le plan d'équation $x = x_0$. On obtient ainsi la fonction $y \mapsto f(x_0, y)$: au facteur de normalisation $f(x_0)$ près, ceci donne une idée de la densité conditionnelle $f(y|x_0)$.

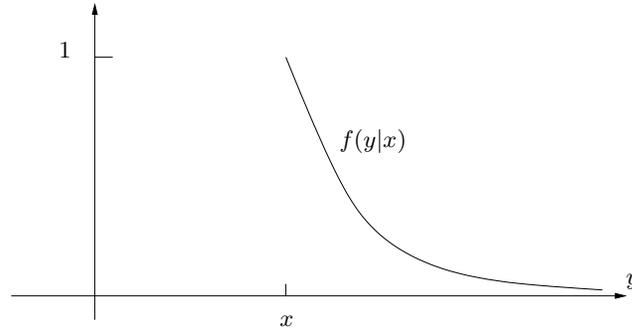


FIGURE 1.8 – Densité conditionnelle $f(y|x) = e^{-(y-x)} \mathbb{1}_{\{y \geq x\}}$.

Exemple. Pour l'exemple précédent, pour tout $x \geq 0$, on a :

$$f(y|x) = e^{-(y-x)} \mathbb{1}_{\{y \geq x\}},$$

c'est-à-dire que, conditionnellement à $X = x$, Y suit une loi exponentielle de paramètre 1 traduite sur l'intervalle $[x, +\infty[$. Ceci est illustré figure 1.8.

Ainsi définie, pour tout $x \geq 0$, la fonction $f(\cdot|x)$ est une densité de probabilité, c'est-à-dire qu'elle est positive et somme à 1. Les relations déjà vues dans le cas discret entre marginales et conditionnelles sont encore valables : il suffit de remplacer les sommes discrètes par des intégrales. Ainsi on a par exemple :

$$f(y) = \int_{\mathbb{R}} f(y|x) f(x) dx.$$

De plus, si les variables aléatoires X et Y sont indépendantes, on a bien sûr $f_{X|Y} = f_X$ et $f_{Y|X} = f_Y$.

On veut maintenant définir l'espérance conditionnelle. Pour x fixé, l'espérance conditionnelle de Y sachant $X = x$ est :

$$\mathbb{E}[Y|X = x] = \int_{\mathbb{R}} y f(y|x) dy.$$

La fonction

$$\varphi : x \mapsto \varphi(x) = \mathbb{E}[Y|X = x]$$

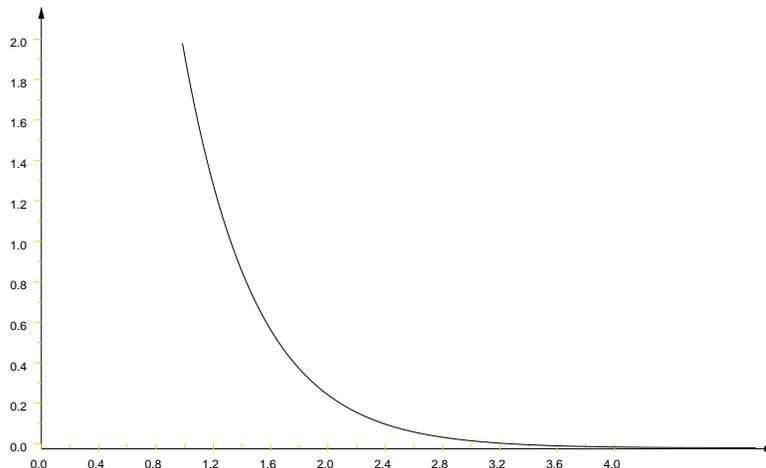
est une fonction réelle de la variable réelle. $\varphi(X)$ est donc une variable aléatoire : c'est l'espérance conditionnelle de Y sachant X .

Définition 6 (Espérance conditionnelle)

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 et on la note $\mathbb{E}[Y|X]$.

Exemple. Pour l'exemple précédent, on obtient pour tout $x \geq 0$:

$$\mathbb{E}[Y|X = x] = (x + 1) \mathbb{1}_{\{x \geq 0\}},$$

FIGURE 1.9 – Densité de la variable aléatoire $\mathbb{E}[Y|X]$.

donc $\mathbb{E}[Y|X] = X + 1$. Or on a vu que $X \sim \mathcal{E}(2)$, donc la variable aléatoire $\mathbb{E}[Y|X]$ suit une loi exponentielle de paramètre 2 translatée sur l'intervalle $[1, +\infty[$ (voir figure 1.9).

Propriétés 2 (Calcul d'espérance par conditionnement)

Si Y est intégrable, alors la variable aléatoire $\mathbb{E}[Y|X]$ aussi et on a :

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

Preuve. La preuve est la même que dans le cas discret. La variable aléatoire $\mathbb{E}[Y|X]$ prend les valeurs $\mathbb{E}[Y|X = x]$ avec densité $f(x)$, donc son espérance vaut :

$$\mathbb{E}[\mathbb{E}[Y|X]] = \int_{\mathbb{R}} \mathbb{E}[Y|X = x]f(x) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} yf(y|x) dy \right) f(x) dx,$$

donc d'après le théorème de Fubini :

$$\mathbb{E}[\mathbb{E}[Y|X]] = \int_{\mathbb{R}} y \left(\int_{\mathbb{R}} f(y|x)f(x) dx \right) dy,$$

et puisque $f(y) = \int_{\mathbb{R}} f(y|x)f(x) dx$, on retrouve bien :

$$\mathbb{E}[\mathbb{E}[Y|X]] = \int_{\mathbb{R}} yf(y) dy = \mathbb{E}[Y].$$

■

Exemple. Pour l'exemple précédent, on a obtenu $\mathbb{E}[Y|X] = X + 1$, avec $X \sim \mathcal{E}(2)$, d'où :

$$\mathbb{E}[Y] = \mathbb{E}[X + 1] = \mathbb{E}[X] + 1 = \frac{3}{2},$$

résultat que l'on retrouve bien en considérant la loi marginale de Y :

$$\mathbb{E}[Y] = \int_{\mathbb{R}} yf(y) dy = \int_{\mathbb{R}^+} y(2e^{-y} - 2e^{-2y}) dy = 2 - \frac{1}{2} = \frac{3}{2}.$$

Dans le cas général, on retrouve alors pour les couples absolument continus les propriétés vues pour les couples discrets. On commence par définir l'espérance conditionnelle d'un couple sachant l'une des variables. Soit $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction, l'espérance mathématique de la variable aléatoire $h(X, Y)$ est définie si :

$$\int_{\mathbb{R}^2} |h(x, y)| f(x, y) dx dy < +\infty,$$

auquel cas elle vaut :

$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}^2} h(x, y) f(x, y) dx dy,$$

que l'on peut encore écrire :

$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x, y) f(y|x) dy \right) f(x) dx.$$

La définition suivante est alors naturelle.

Définition 7 (Espérance conditionnelle d'un couple)

L'espérance conditionnelle de $h(X, Y)$ sachant $X = x$ est :

$$\mathbb{E}[h(X, Y)|X = x] = \int_{\mathbb{R}} h(x, y) f(y|x) dy = \mathbb{E}[h(x, Y)|X = x].$$

L'espérance conditionnelle de $h(X, Y)$ sachant X , notée $\mathbb{E}[h(X, Y)|X]$, est la variable aléatoire qui prend les valeurs $\mathbb{E}[h(x, Y)|X = x]$ avec la densité de probabilité $f(x)$.

On peut alors énumérer différentes propriétés de l'espérance conditionnelle. Dans ce qui suit, on ne considère que des "bonnes fonctions", c'est-à-dire telles qu'on n'ait pas de problème d'intégrabilité.

Propriétés 3 (Propriétés de l'espérance conditionnelle)

Sous réserve d'intégrabilité des variables aléatoires, on a les propriétés suivantes :

– Calcul d'espérance par conditionnement :

$$\mathbb{E}[\mathbb{E}[h(X, Y)|X]] = \int_{\mathbb{R}} \mathbb{E}[h(x, Y)|X = x] f(x) dx = \mathbb{E}[h(X, Y)].$$

– Indépendance : si X et Y sont indépendantes, alors $\mathbb{E}[g(Y)|X] = \mathbb{E}[g(Y)]$. En particulier, $\mathbb{E}[Y|X] = E[Y]$.

– On a $\mathbb{E}[g(X)|X] = g(X)$. En particulier $\mathbb{E}[X|X] = X$.

– Linéarité :

$$\mathbb{E}[\alpha g(X) + \beta h(Y)|X] = \alpha \mathbb{E}[g(X)|X] + \beta \mathbb{E}[h(Y)|X] = \alpha g(X) + \beta \mathbb{E}[h(Y)|X].$$

– Linéarité(bis) :

$$\mathbb{E}[g(X)h(Y)|X] = g(X)\mathbb{E}[h(Y)|X].$$

Preuve. Toutes les démonstrations se font sans difficulté en revenant à la définition de l'espérance conditionnelle. Pour la première relation, il suffit par exemple de dire que :

- La variable aléatoire $\mathbb{E}[h(X, Y)|X]$ prend les valeurs $\mathbb{E}[h(x, Y)|X = x]$ avec densité de probabilité $f(x)$. Donc son espérance vaut :

$$\mathbb{E}[\mathbb{E}[h(X, Y)|X]] = \int_{\mathbb{R}} \mathbb{E}[h(x, Y)|X = x] f(x) dx.$$

Par ailleurs, on a pour tout réel x :

$$\mathbb{E}[h(x, Y)|X = x] = \int_{\mathbb{R}} h(x, y) f(y|x) dy,$$

d'où il vient, puisque $f(x, y) = f(y|x)f(x)$:

$$\mathbb{E}[\mathbb{E}[h(X, Y)|X]] = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x, y) f(y|x) dy \right) f(x) dx = \int_{\mathbb{R}^2} h(x, y) f(x, y) dx dy,$$

et on reconnaît $\mathbb{E}[h(X, Y)]$.

- Und so weiter...

■

Remarque. Tout comme l'espérance classique, l'espérance conditionnelle est linéaire. La dernière propriété est assez spectaculaire : du point de vue de l'espérance conditionnelle, toute fonction de la variable aléatoire X se comporte comme une constante, on peut donc la sortir du crochet.

1.3 Applications

Toute cette section est valable aussi bien dans le cas discret que dans le cas absolument continu. Ce n'est que par souci de simplification qu'on se place parfois dans l'une des deux situations.

1.3.1 Probabilités conditionnelles

Soit A un événement qui s'exprime en fonction de X et Y , par exemple :

$$A = \{X < Y\} = \{\omega \in \Omega : X(\omega) < Y(\omega)\}.$$

On peut écrire sa probabilité comme l'espérance d'une indicatrice :

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A] = \mathbb{E}[\mathbb{1}_{\{X < Y\}}] = \int_{\mathbb{R}^2} \mathbb{1}_{\{x < y\}}(x, y) f(x, y) dx dy,$$

or il est souvent plus facile de calculer cette quantité en commençant par geler l'une des variables et en intégrant par rapport à l'autre. C'est le principe du conditionnement.

Définition 8 (Probabilités conditionnelles)

La probabilité conditionnelle de l'événement A sachant $X = x$ est la quantité :

$$\mathbb{P}(A|X = x) = \mathbb{E}[\mathbb{1}_A|X = x] = \int_{\mathbb{R}} \mathbb{1}_A(x, y) f(y|x) dy.$$

La probabilité conditionnelle de A sachant X , notée $\mathbb{P}(A|X)$, est la variable aléatoire prenant les valeurs $\mathbb{P}(A|X = x)$ avec densité $f(x)$.

Remarque. Il faut noter que, tout comme l'espérance conditionnelle $\mathbb{E}[Y|X]$, la probabilité conditionnelle $\mathbb{P}(A|X)$ est une **variable aléatoire**.

D'après les résultats de la section précédente, on peut alors appliquer la technique de calcul d'espérance par conditionnement.

Proposition 4 (Calcul de probabilité par conditionnement)

$$\mathbb{P}(A) = \int_{\mathbb{R}} \mathbb{P}(A|X = x)f(x) dx.$$

Preuve. Il suffit d'appliquer le calcul d'espérance par conditionnement (propriétés 3) à la fonction $h(x, y) = \mathbb{1}_{\{(x,y) \in A\}}$. ■

Pour revenir à l'exemple initial, on a alors :

$$\mathbb{P}(A|X = x) = \mathbb{P}(\mathbb{1}_{\{X < Y\}}|X = x) = \mathbb{E}[\mathbb{1}_{\{X < Y\}}|X = x] = \mathbb{E}[\mathbb{1}_{\{x < Y\}}|X = x] = \mathbb{P}(Y > x|X = x).$$

En particulier, si X et Y sont indépendantes : $\mathbb{P}(x < Y|X = x) = \mathbb{P}(x < Y)$.

Exemple. Soit X et Y deux variables aléatoires indépendantes, avec $X \sim \mathcal{E}(\lambda)$ et $Y \sim \mathcal{E}(\mu)$. On veut calculer la probabilité $\mathbb{P}(X < Y)$. On écrit donc :

$$\mathbb{P}(X < Y) = \int_{\mathbb{R}} \mathbb{P}(x < Y|X = x)f(x) dx = \int_{\mathbb{R}^+} \mathbb{P}(x < Y|X = x)\lambda e^{-\lambda x} dx.$$

Or on a par hypothèse d'indépendance :

$$\mathbb{P}(x < Y|X = x) = \mathbb{P}(x < Y) = 1 - F_Y(x) = e^{-\mu x}.$$

D'où il vient :

$$\mathbb{P}(X < Y) = \int_{\mathbb{R}^+} \lambda e^{-(\lambda+\mu)x} dx = \frac{\lambda}{\lambda + \mu}.$$

Ce type de résultat intervient constamment dans l'étude des processus markoviens de sauts (par exemple les files d'attente).

1.3.2 La régression

Approximation d'une variable aléatoire par une constante

Soit Y une variable aléatoire de carré intégrable. On veut approcher Y par une constante. Si on s'intéresse à l'erreur quadratique, la solution est donnée par l'espérance.

Proposition 5 (Approximation par une constante)

Supposons Y telle que $\mathbb{E}[Y^2] < +\infty$. Parmi tous les réels a , la quantité $\mathbb{E}[(Y - a)^2]$ est minimale lorsque $a = \mathbb{E}[Y]$. C'est-à-dire :

$$\min_{a \in \mathbb{R}} \mathbb{E}[(Y - a)^2] = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \text{Var } Y.$$

Preuve. Notons $m = \mathbb{E}[Y]$, alors pour tout réel a :

$$\mathbb{E}[(Y - a)^2] = \mathbb{E}[(Y - m) + (m - a)]^2.$$

On utilise la linéarité de l'espérance :

$$\mathbb{E}[(Y - a)^2] = \mathbb{E}[(Y - m)^2] + 2(m - a)\mathbb{E}[(Y - m)] + (m - a)^2$$

Or $\mathbb{E}[(Y - m)] = 0$ et par suite :

$$\mathbb{E}[(Y - a)^2] = \mathbb{E}[(Y - m)^2] + (m - a)^2.$$

Cette quantité est minimale lorsque $a = m = \mathbb{E}[Y]$. ■

Remarque. Si on considère l'erreur en norme L^1 , c'est-à-dire si on cherche le réel a tel que $\mathbb{E}|Y - a|$ soit minimale, on obtient non pas la moyenne de Y , mais sa médiane.

Rappel. Si Y admet un moment d'ordre 2, i.e. si $\mathbb{E}[Y^2] < +\infty$, alors Y admet un moment d'ordre 1, i.e. $\mathbb{E}|Y| < +\infty$. De manière générale, si Y admet un moment d'ordre $p \in \mathbb{N}^*$, i.e. si $\mathbb{E}[|Y|^p] < +\infty$, alors Y admet un moment d'ordre k pour tout $k \in \{1, \dots, p\}$. Plus précisément, on a l'inégalité suivante entre moments :

$$1 \leq p \leq q \Rightarrow \mathbb{E}[|Y^p|]^{\frac{1}{p}} \leq \mathbb{E}[|Y^q|]^{\frac{1}{q}}.$$

Approximation d'une variable aléatoire par une droite

On considère un couple aléatoire (X, Y) , dont on connaît la loi jointe. On suppose que les variables marginales X et Y sont toutes deux de carré intégrable. Parler de la covariance du couple (X, Y) a alors un sens puisque :

$$\forall \omega \in \Omega \quad |X(\omega)Y(\omega)| \leq \frac{1}{2}(X^2(\omega) + Y^2(\omega)),$$

ce qui implique en passant aux espérances :

$$\mathbb{E}|XY| \leq \frac{1}{2}\mathbb{E}[X^2 + Y^2] = \frac{1}{2}(\mathbb{E}[X^2] + \mathbb{E}[Y^2]) < +\infty.$$

Supposons qu'on observe des réalisations de la variable aléatoire X , mais pas celles de Y : pour chaque valeur x de X , on voudrait deviner la valeur y de Y . Ceci est impossible, puisque même sachant $X = x$, Y est aléatoire. On peut néanmoins chercher à faire l'erreur la plus petite possible en moyenne.

Une idée simple est d'approcher Y par une fonction affine de X , c'est-à-dire chercher la variable aléatoire $Y = aX + b$ la plus proche possible, en moyenne, de Y . Pour l'erreur quadratique, on cherche donc à minimiser :

$$\mathbb{E}[(Y - (aX + b))^2] = \mathbb{E}[X^2]a^2 + 2\mathbb{E}[X]ab + b^2 - 2\mathbb{E}[XY]a - 2\mathbb{E}[Y]b + \mathbb{E}[Y^2] = \phi(a, b).$$

Cette fonction ϕ des deux variables a et b est en fait issue d'une forme quadratique définie positive et atteint son minimum au point :

$$\begin{cases} a &= \frac{\text{Cov}(X, Y)}{\sigma^2(X)} \\ b &= \mathbb{E}[Y] - \frac{\text{Cov}(X, Y)}{\sigma^2(X)}\mathbb{E}[X] \end{cases}$$

comme le montre la factorisation de Gauss :

$$\phi(a, b) = (b - (\mathbb{E}[Y] - \mathbb{E}[X]a))^2 + \left(\sigma(X)a - \frac{\text{Cov}(X, Y)}{\sigma(X)} \right)^2 + \sigma^2(Y) - \frac{\text{Cov}^2(X, Y)}{\sigma^2(X)}.$$

En notant ρ le coefficient de corrélation linéaire :

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

l'erreur minimale moyenne, appelée erreur quadratique, est donc :

$$\min_{a, b} \mathbb{E}[(Y - (aX + b))^2] = \sigma^2(Y)(1 - \rho^2).$$

Noter que ceci suppose $\sigma(X) \neq 0$: si $\sigma(X) = 0$, la variable aléatoire X est presque sûrement constante, donc chercher à approcher Y par une fonction affine de X revient à approcher Y par une constante c . On a vu précédemment que le mieux est de prendre $c = \mathbb{E}[Y]$.

On voit que l'erreur quadratique est d'autant plus faible que :

1. La valeur absolue $|\rho|$ du coefficient de corrélation linéaire est proche de 1. En général, on sait simplement que :

$$-1 \leq \rho \leq 1.$$

Lorsque $\rho = 0$, X et Y sont dits décorrélés et le mieux qu'on puisse faire est à nouveau d'approcher Y par sa moyenne $\mathbb{E}[Y]$. Lorsque $\rho = \pm 1$, X et Y sont liés par une relation affine et l'erreur est nulle.

2. La variable Y est peu dispersée autour de sa moyenne.

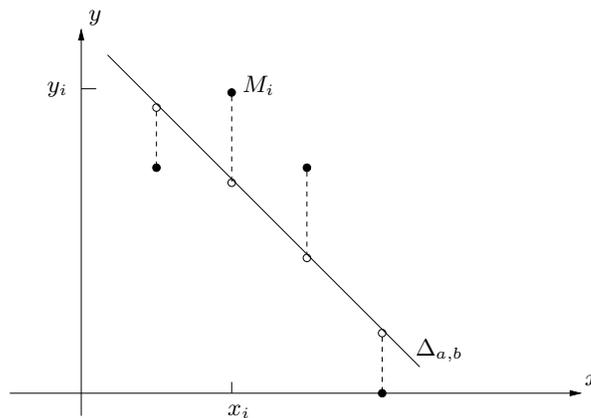


FIGURE 1.10 – Droite de régression.

Remarque. En statistiques, c'est-à-dire dans la vraie vie, on ne connaît pas la loi du couple (X, Y) . On dispose simplement d'un nuage de points $(x_i, y_i)_{1 \leq i \leq n}$. Or, si on cherche la droite d'équation $y = ax + b$ qui minimise la somme des carrés :

$$\sum_{i=1}^n (y_i - (ax_i + b))^2,$$

on retrouve la droite de régression ci-dessus en considérant les espérances, variances et covariances empiriques (voir figure 1.10).

Approximation d'une variable aléatoire par une fonction

On reprend le problème précédent. On considère donc un couple aléatoire (X, Y) dont on connaît la loi jointe, avec Y de carré intégrable. On suppose toujours qu'on n'observe que la variable aléatoire X et qu'on veut approcher au mieux la variable aléatoire Y . Par rapport à ce qui précède, on ne se restreint pas à une approximation de Y par une fonction affine de X : parmi toutes les fonctions possibles, on cherche la fonction de X qui approche le mieux Y . On retrouve alors l'espérance conditionnelle.

Définition 9 (Courbe de régression)

La courbe $x \mapsto y = \mathbb{E}[Y|X = x]$ est appelée courbe de régression de Y en X .

Par exemple, si X et Y sont indépendantes, la courbe de régression de Y en X est tout simplement la droite horizontale $y = \mathbb{E}[Y]$.

Exemples.

1. On considère à nouveau le couple (X, Y) de densité :

$$f(x, y) = 2e^{-(x+y)} \mathbb{1}_{\{0 \leq x \leq y\}}$$

La fonction de régression de Y en X est la fonction :

$$u : \begin{cases} [0, +\infty[& \rightarrow \mathbb{R} \\ x & \mapsto (x + 1) \end{cases}$$

Sur cet exemple, on voit que la courbe de régression coïncide avec la droite de régression : ceci n'est pas vrai en général, comme le montre l'exemple suivant.

2. Soit (X, Y) un couple aléatoire de densité jointe :

$$f(x, y) = \frac{12}{5}y(2 - y - x)\mathbb{1}_{]0,1]^2}(x, y).$$

Quelques calculs permettent de montrer que pour tout $x \in]0, 1[$:

$$\mathbb{E}[Y|X = x] = \frac{5 - 4x}{8 - 6x}.$$

La courbe de régression est donc un morceau d'hyperbole (voir figure 1.11).

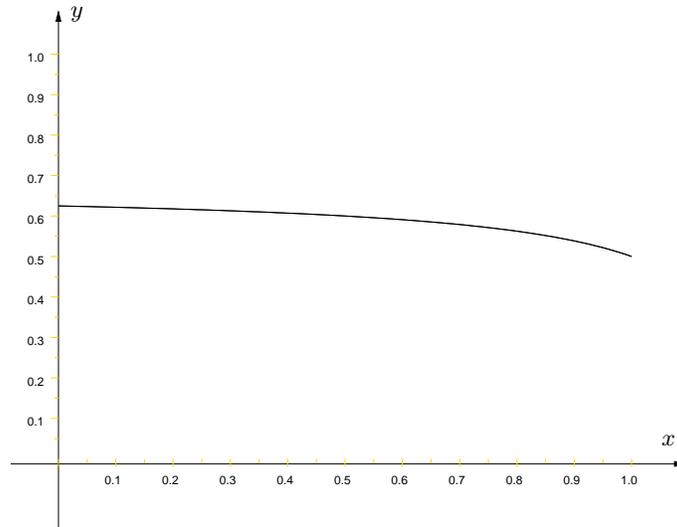
La fonction de régression vérifie une propriété de minimalité souvent utilisée en statistiques. On considère une fonction $u : \mathbb{R} \rightarrow \mathbb{R}$ et la quantité :

$$\mathbb{E}[(Y - u(X))^2].$$

Cette quantité varie lorsque la fonction u varie. Quand est-elle minimale ? Ce genre de problème, dit de calcul des variations, est en général difficile : on cherche une fonction minimisant un certain critère (et non un point de l'espace de dimension finie \mathbb{R}^n comme en optimisation classique). Mais ici tout est simple, au moins d'un point de vue théorique...

Théorème 2 (Espérance conditionnelle et régression)

Supposons Y telle que $\mathbb{E}[Y^2] < +\infty$. Parmi toutes les fonctions $u : \mathbb{R} \rightarrow \mathbb{R}$, l'erreur d'approximation $\mathbb{E}[(Y - u(X))^2]$ est minimale lorsque u est la fonction de régression $x \mapsto \mathbb{E}[Y|X = x]$, i.e. lorsque $u(X) = \mathbb{E}[Y|X]$.

FIGURE 1.11 – Courbe de régression $y = \frac{5-4x}{8-6x}$.

La preuve est calquée sur celle déjà vue pour l'approximation de Y par une constante.

Preuve. Notons $m(X) = \mathbb{E}[Y|X]$, alors pour toute fonction $u : \mathbb{R} \rightarrow \mathbb{R}$, on peut écrire :

$$\mathbb{E}[(Y - u(X))^2] = \mathbb{E}[(Y - m(X)) + (m(X) - u(X))]^2$$

On utilise la linéarité de l'espérance :

$$\mathbb{E}[(Y - u(X))^2] = \mathbb{E}[(Y - m(X))^2] + 2\mathbb{E}[(Y - m(X))(m(X) - u(X))] + \mathbb{E}[(u(X) - m(X))^2].$$

Or le calcul d'espérance par conditionnement assure que :

$$\mathbb{E}[(Y - m(X))(m(X) - u(X))] = \mathbb{E}[\mathbb{E}[(Y - m(X))(m(X) - u(X))|X]],$$

et puisque $m(X) - u(X)$ est une fonction de X , on sait que :

$$\mathbb{E}[(Y - m(X))(m(X) - u(X))] = \mathbb{E}[\mathbb{E}[(Y - m(X))|X](m(X) - u(X))],$$

or par linéarité de l'espérance conditionnelle et puisque $\mathbb{E}[m(X)|X] = m(X) = \mathbb{E}[Y|X]$, on en déduit que :

$$\mathbb{E}[(Y - m(X))|X] = \mathbb{E}[Y|X] - \mathbb{E}[m(X)|X] = \mathbb{E}[Y|X] - m(X) = 0.$$

On a donc obtenu :

$$\mathbb{E}[(Y - u(X))^2] = \mathbb{E}[(Y - m(X))^2] + \mathbb{E}[(u(X) - m(X))^2].$$

Cette quantité est minimale lorsque $u(X) = \mathbb{E}[Y|X]$. ■

On donne un nom au minimum obtenu.

Définition 10 (Erreur quadratique moyenne)

La quantité

$$\sigma^2 = \min_u \mathbb{E}[(Y - u(X))^2] = \mathbb{E}[(Y - \mathbb{E}[Y|X])^2]$$

est appelée l'erreur quadratique moyenne, ou la variance résiduelle, ou encore le résidu.

Nous allons voir que cette dernière approche peut s'interpréter géométriquement et fournir une représentation très commode de l'espérance conditionnelle.

1.4 Interprétation géométrique de l'espérance conditionnelle

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On note $L^2(\Omega, \mathcal{F}, \mathbb{P})$ ou plus simplement $L^2(\Omega)$ l'ensemble des variables aléatoires $X : \Omega \rightarrow \mathbb{R}$ de carré intégrable, i.e. telles que :

$$\mathbb{E}[X^2] < +\infty.$$

On convient de plus d'identifier deux variables aléatoires presque sûrement égales.

Propriétés 4 (Propriétés de $L^2(\Omega)$)

- Si X et Y appartiennent à $L^2(\Omega)$, alors XY est intégrable.
- $L^2(\Omega)$ est un espace vectoriel.

Preuve. On a déjà vu dans le paragraphe sur la régression que si X et Y sont de carré intégrable, la variable aléatoire XY est elle aussi intégrable. Alors si α et β sont deux scalaires, la variable aléatoire $(\alpha X + \beta Y)$ est dans $L^2(\Omega)$ puisque :

$$(\alpha X + \beta Y)^2 = \alpha^2 X^2 + 2\alpha\beta XY + \beta^2 Y^2,$$

qui est une somme de variables aléatoires intégrables. Ceci fait de $L^2(\Omega)$ un sous-espace vectoriel de l'espace des variables aléatoires sur $(\Omega, \mathcal{F}, \mathbb{P})$. ■

A part dans le cas où l'espace Ω est fini, l'espace $L^2(\Omega)$ est de dimension infinie. Nous allons maintenant voir en quoi $L^2(\Omega)$ "ressemble" à l'espace usuel \mathbb{R}^n muni de la norme euclidienne. Rappelons qu'un produit scalaire est une forme bilinéaire symétrique définie positive, dont on peut déduire une norme.

Proposition 6 (Le produit scalaire dans $L^2(\Omega)$)

L'application

$$\langle \cdot, \cdot \rangle : \begin{cases} L^2(\Omega) \times L^2(\Omega) & \rightarrow \mathbb{R} \\ (X, Y) & \mapsto \langle X, Y \rangle = \mathbb{E}[XY] \end{cases}$$

est un produit scalaire sur $L^2(\Omega)$. La norme associée est :

$$\|X\| = \sqrt{\mathbb{E}[X^2]}.$$

Remarque. On écrit la norme $\|X\|$ plutôt que $\|X\|_2$ afin de ne pas alourdir les notations.

Preuve. L'intégrabilité de XY a été vue ci-dessus donc l'application est bien définie sur l'espace produit $L^2(\Omega) \times L^2(\Omega)$. La bilinéarité et la symétrie sont évidentes. Soit maintenant $X \in L^2(\Omega)$, il est clair que $\langle X, X \rangle = \mathbb{E}[X^2] \geq 0$. Supposons que $\mathbb{E}[X^2] = 0$, alors par l'inégalité de Tchebychev, pour tout $n \in \mathbb{N}^*$:

$$\mathbb{P}\left(X \geq \frac{1}{n}\right) \leq n^2 \mathbb{E}[X^2] = 0,$$

d'où l'on déduit par sous-sigma-additivité d'une mesure de probabilité :

$$\mathbb{P}(X > 0) = \mathbb{P}\left(\bigcup_{n=1}^{+\infty} \left\{X \geq \frac{1}{n}\right\}\right) \leq \sum_{n=1}^{+\infty} \mathbb{P}\left(X \geq \frac{1}{n}\right) = 0,$$

c'est-à-dire que X est presque sûrement égale à 0. Puisqu'on a convenu au début de ce paragraphe d'identifier deux variables presque sûrement égales, on a donc $X = 0$. Ainsi $\langle \cdot, \cdot \rangle$ est bien une forme bilinéaire symétrique définie positive, c'est-à-dire un produit scalaire.

■

Dans ce cadre, dire que les variables aléatoires X et Y sont orthogonales pour le produit scalaire $\langle \cdot, \cdot \rangle$ signifie que $\mathbb{E}[XY] = 0$. Dans le cas de variables centrées, l'orthogonalité correspond donc à la non-corrélation. On récupère automatiquement les propriétés d'une norme issue d'un produit scalaire, vues en cours d'algèbre linéaire.

Propriétés 5 (Propriétés de la norme dans $L^2(\Omega)$)

- *Positivité* : $\|X\| \geq 0$ et $\|X\| = 0$ seulement pour $X = 0$.
- *Homogénéité* : pour tout réel a , $\|aX\| = |a|\|X\|$.
- *Inégalité triangulaire* : $\|X + Y\| \leq \|X\| + \|Y\|$.
- *Inégalité de Cauchy-Schwarz* : $\langle X, Y \rangle \leq \|X\|\|Y\|$.

La norme $\|\cdot\|$ permet de mesurer la distance entre deux variables aléatoires :

$$d(X, Y) = \|Y - X\|$$

On l'appelle la distance en moyenne quadratique entre X et Y . On peut alors associer à cette norme la notion de convergence en moyenne quadratique : la suite de variables aléatoires $(X_n)_{n \geq 0}$ converge en moyenne quadratique vers la variable aléatoire X et on note :

$$X_n \xrightarrow{L^2} X$$

si $\lim_{n \rightarrow \infty} \|X - X_n\| = 0$, c'est-à-dire si

$$\lim_{n \rightarrow \infty} \mathbb{E}[(X - X_n)^2] = 0.$$

Nous donnons maintenant un résultat dû à F. Riesz et d'une grande importance théorique.

Théorème 3 (Espace de Hilbert)

L'espace $L^2(\Omega)$ est complet, c'est-à-dire que toute suite $(X_n)_{n \geq 0}$ de $L^2(\Omega)$ vérifiant le critère de Cauchy converge en moyenne quadratique vers une variable aléatoire X de $L^2(\Omega)$.

Comme en analyse, l'intérêt du critère de Cauchy dans un espace complet est de permettre de montrer la convergence d'une suite sans connaître sa limite.

Preuve. Soit $(X_n)_{n \geq 0}$ une suite de Cauchy dans $L^2(\Omega)$. Il existe donc une suite d'indices $(n_k)_{k \geq 0}$ telle que :

$$\forall n \geq n_k, \forall p \geq 0 \quad \|X_{n+p} - X_n\| \leq \frac{1}{2^k}.$$

En particulier, la suite de variables aléatoires $(Y_k)_{k \geq 0}$ définie par $Y_0 = X_{n_0}$ et :

$$\forall k \geq 1 \quad Y_k = X_{n_k} - X_{n_{k-1}}$$

est telle que :

$$\sum_{k=0}^{+\infty} \|Y_k\| \leq \|X_{n_0}\| + \sum_{k=0}^{+\infty} \frac{1}{2^k} = \|X_{n_0}\| + 2 < +\infty.$$

On va montrer que $(Y_k)_{k \geq 0}$ converge presque sûrement vers une variable aléatoire X . Considérons la variable aléatoire Z définie pour tout $\omega \in \Omega$ par :

$$Z(\omega) = \lim_{K \rightarrow +\infty} Z_K(\omega) = \lim_{K \rightarrow +\infty} \sum_{k=0}^K |Y_k(\omega)| = \sum_{k=0}^{+\infty} |Y_k(\omega)| \leq +\infty,$$

avec les variables aléatoires Z_K définie pour tout $\omega \in \Omega$ par :

$$Z_K(\omega) = \sum_{k=0}^K |Y_k(\omega)| < +\infty.$$

Autrement dit, la suite croissante de variables aléatoires positives $(Z_K)_{K \geq 0}$ converge presque sûrement vers Z . Il en va de même de la suite $(Z_K^2)_{K \geq 0}$ vers Z^2 et on peut donc appliquer le théorème de Beppo Lévi :

$$\mathbb{E}[Z^2] = \lim_{K \rightarrow +\infty} \mathbb{E}[Z_K^2].$$

Mais par l'inégalité triangulaire on a aussi pour tout $K \geq 0$:

$$\mathbb{E}[Z_K^2] = \|Z_K\|^2 \leq \left(\sum_{k=0}^K \|Y_k\| \right)^2 \leq \left(\sum_{k=0}^{+\infty} \|Y_k\| \right)^2 < +\infty.$$

On en déduit que $\mathbb{E}[Z^2] < +\infty$. Mais alors par Tchebychev, pour tout $n \in \mathbb{N}^*$:

$$\mathbb{P}(Z \geq n) \leq \frac{\mathbb{E}[Z^2]}{n^2},$$

d'où l'on déduit par continuité monotone décroissante d'une mesure de probabilité :

$$\mathbb{P}(Z = +\infty) = \mathbb{P}\left(\bigcap_{n=1}^{+\infty} \{Z \geq n\}\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(Z \geq n),$$

ce qui donne :

$$\mathbb{P}(Z = +\infty) \leq \lim_{n \rightarrow +\infty} \frac{\mathbb{E}[Z^2]}{n^2} = 0,$$

c'est-à-dire que Z est presque sûrement finie. Puisqu'on convient de confondre deux variables aléatoires presque sûrement égales, quitte à remplacer $Z(\omega)$ par 0 aux éventuels points ω où on aurait $Z(\omega) = +\infty$, on peut donc considérer que pour tout $\omega \in \Omega$, on a :

$$\sum_{k=0}^{+\infty} |Y_k(\omega)| < +\infty.$$

En particulier, la série de variables aléatoires $\sum_{k \geq 0} Y_k$ est absolument convergente sur Ω , donc simplement convergente, et il existe une variable aléatoire X finie sur Ω telle que pour tout $\omega \in \Omega$:

$$X(\omega) = \sum_{k=0}^{+\infty} Y_k(\omega).$$

La variable aléatoire X est de carré intégrable puisque par définition, pour tout $\omega \in \Omega$:

$$|X(\omega)| \leq Z(\omega),$$

et on a prouvé ci-dessus que $Z \in L^2(\Omega)$. Il reste à voir que :

$$\left\| X - \sum_{k=0}^K Y_k \right\| = \left\| \sum_{k=K+1}^{+\infty} Y_k \right\| \leq \sum_{k=K+1}^{+\infty} \|Y_k\| \leq \frac{1}{2^K} \xrightarrow{K \rightarrow +\infty} 0.$$

On revient enfin à la suite $(X_n)_{n \geq 0}$. Pour tout $n \geq n_K$, on a par inégalité triangulaire :

$$\|X - X_n\| \leq \|X - X_{n_K}\| + \|X_{n_K} - X_n\|,$$

or il est clair que :

$$X_{n_K} = \sum_{k=0}^K Y_k,$$

d'où il vient :

$$\|X - X_n\| \leq \frac{1}{2^K} + \frac{1}{2^K} = \frac{1}{2^{K-1}} \xrightarrow{K \rightarrow +\infty} 0,$$

et la suite $(X_n)_{n \geq 0}$ converge en moyenne quadratique vers la variable aléatoire X . ■

Remarque. Qu'a-t-on utilisé comme outils ? Beppo Lévi et la continuité monotone décroissante, or ceux-ci sont valables dans tout espace mesuré (Ω, \mathcal{F}, m) . Ainsi, le résultat qu'on vient d'établir est en fait très général : on le retrouve par exemple dans l'espace classique d'intégration $(\mathbb{R}, \mathcal{B}, \lambda)$ lorsqu'on considère les fonctions f de carré intégrable sur \mathbb{R} . On le retrouve aussi dans l'espace des suites $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ muni de la mesure de comptage, autrement dit dans l'étude des séries numériques $\sum_{n \geq 0} u_n$ telles que $\sum_{n=0}^{+\infty} u_n^2 < +\infty$. Dans ce dernier espace, le produit scalaire est :

$$\langle u, v \rangle = \sum_{n=0}^{+\infty} u_n v_n.$$

Un espace vectoriel ayant un produit scalaire et complet pour la norme induite par celui-ci est appelé espace de Hilbert. Ces espaces jouissent de nombreuses propriétés. Celle qui suit est cruciale.

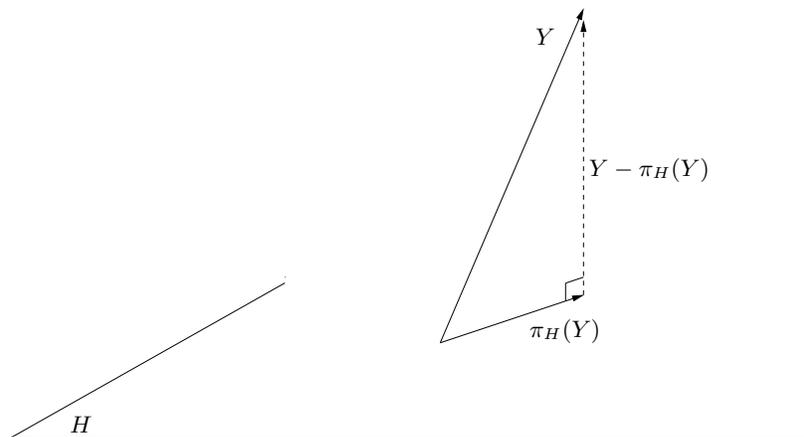


FIGURE 1.12 – Projeté orthogonal $\pi_H(Y)$ de Y sur le sous-espace H .

Théorème 4 (Théorème de la projection orthogonale)

Soit H un sous-espace fermé de $L^2(\Omega)$. Pour tout Y de $L^2(\Omega)$, il existe une unique variable aléatoire de H , notée $\pi_H(Y)$, qui soit à plus courte distance de Y . On l'appelle le projeté orthogonal de Y sur H et elle est entièrement caractérisée par la double propriété :

$$\begin{cases} \pi_H(Y) \in H \\ Y - \pi_H(Y) \perp H \end{cases}$$

L'application $Y \mapsto \pi_H(Y)$ est linéaire sur $L^2(\Omega)$. De plus, $\pi_H(Y) = Y$ ssi $Y \in H$, et $\|\pi_H(Y)\|_2 < \|Y\|$ si $Y \notin H$ (voir figure 1.12).

Preuve. Notons :

$$d = \inf_{X \in H} d(X, Y).$$

Rappelons l'identité du parallélogramme, valable dès qu'on considère un produit scalaire (cf. figure 1.13) :

$$\|U + V\|^2 + \|U - V\|^2 = 2(\|U\|^2 + \|V\|^2).$$

En considérant deux variables aléatoires X_1 et X_2 de H , on l'applique à $U = (X_1 - Y)$ et $V = (X_2 - Y)$:

$$4 \left\| \frac{X_1 + X_2}{2} - Y \right\|^2 + \|X_1 - X_2\|^2 = 2(\|X_1 - Y\|^2 + \|X_2 - Y\|^2).$$

Mais, puisque H est un sous-espace de $L^2(\Omega)$, la variable aléatoire milieu $(X_1 + X_2)/2$ appartient à H et vérifie :

$$\left\| \frac{X_1 + X_2}{2} - Y \right\| \geq d,$$

d'où l'on déduit que pour tout couple (X_1, X_2) de H :

$$\|X_1 - X_2\|^2 \leq 2(\|X_1 - Y\|^2 + \|X_2 - Y\|^2) - 4d^2. \quad (1.1)$$

Cette inégalité prouve que l'inf ne peut être atteint par deux éléments distincts X_1 et X_2 de H , sinon on aurait $\|X_1 - X_2\|^2 < 0$. Par définition d'une borne inf, il existe une suite (X_n) de H telle que :

$$\lim_{n \rightarrow +\infty} d(X_n, Y) = d.$$

Si on applique l'inégalité 1.1, on obtient pour tout couple d'entiers naturels (n, p) :

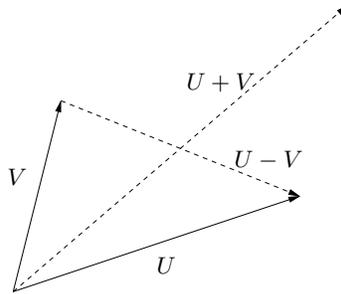


FIGURE 1.13 – Identité du parallélogramme : $\|U + V\|^2 + \|U - V\|^2 = 2(\|U\|^2 + \|V\|^2)$.

$$\|X_{n+p} - X_n\|^2 \leq 2(\|X_{n+p} - Y\|^2 + \|X_n - Y\|^2) - 4d^2.$$

Pour n assez grand, cette quantité peut être rendue arbitrairement petite, indépendamment de p , ce qui prouve que (X_n) est une suite de Cauchy de H , sous-espace de $L^2(\Omega)$. Par le théorème de Riesz, elle est donc convergente vers une variable aléatoire de $L^2(\Omega)$. Mais puisque H est fermé par hypothèse, celle-ci appartient nécessairement à H : on la note $\pi_H(Y)$ et on l'appelle le projeté orthogonal de Y sur H .

Montrons que $(Y - \pi_H(Y))$ est orthogonale à toute variable aléatoire X de H . Soit donc $X \in H$, alors pour tout réel α , on a aussi $\alpha X \in H$ et par définition de la borne inf, on a donc :

$$\|\pi_H(Y) + \alpha X - Y\|^2 \geq \|\pi_H(Y) - Y\|^2,$$

ce qui s'écrit en développant avec le produit scalaire :

$$\forall \alpha \in \mathbb{R} \quad \|X\|^2 \alpha^2 + 2\langle \pi_H(Y) - Y, X \rangle \alpha \geq 0.$$

Ce trinôme en α ne peut être de signe constant que si $\langle \pi_H(Y) - Y, V \rangle = 0$.
Les deux propriétés

$$\begin{cases} \pi_H(Y) \in H \\ Y - \pi_H(Y) \perp H \end{cases}$$

caractérisent bien $\pi_H(Y)$, car si X est un autre élément de H ayant ces deux propriétés, alors $(\pi_H(Y) - X) = (Y - \pi_H(Y)) + (\pi_H(Y) - X)$ est dans H et orthogonal à tout élément de H donc à lui-même, donc c'est le vecteur nul. Cette caractérisation permet de voir facilement que la projection π_H est un opérateur linéaire. Par ailleurs, il est clair que si $Y \in H$, alors $\pi_H(Y) = Y$. Enfin, puisque $(Y - \pi_H(Y))$ et $\pi_H(Y)$ sont orthogonaux, la relation de Pythagore dans le triangle rectangle d'hypoténuse Y implique que $\|\pi_H(Y)\| < \|Y\|$, sauf si $Y = \pi_H(Y) \in H$. ■

Remarque. Le théorème de projection est encore valable si on considère un ensemble convexe fermé non vide \mathcal{C} à la place du sous-espace vectoriel fermé H .

Considérons maintenant une variable aléatoire X . On lui associe :

$$L^2(X) \triangleq \{u(X) \text{ avec } u : \mathbb{R} \rightarrow \mathbb{R} \text{ borélienne telle que } \mathbb{E}[u^2(X)] < +\infty\},$$

ensemble des variables aléatoires qui s'expriment comme fonctions de X et de carré intégrable. Il est clair que $L^2(X)$ est un sous-espace de $L^2(\Omega)$. On peut de plus montrer que c'est un sous-espace fermé de $L^2(\Omega)$.

Lemme 1 ($L^2(X)$ sous-espace fermé de $L^2(\Omega)$)

Soit X une variable aléatoire, alors l'espace $L^2(X)$ est un sous-espace fermé de l'espace $L^2(\Omega)$ des variables aléatoires de carré intégrable.

Preuve. Notons \mathbb{P}_X la probabilité image de la variable aléatoire X , c'est-à-dire la loi de probabilité définie sur \mathbb{R} muni de la tribu borélienne \mathcal{B} par :

$$\forall B \in \mathcal{B} \quad \mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

Ceci fait de $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ un espace probabilisé. Soit $E = L^2(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ l'espace des variables aléatoires de carré intégrable sur cet espace. Ce sont les fonctions $f : \mathbb{R} \rightarrow \mathbb{R}$ boréliennes telles que :

$$\int_{\mathbb{R}} f^2(x) \mathbb{P}_X(dx) < +\infty.$$

Mais par le théorème de transfert, ceci est équivalent à dire que :

$$\int_{\Omega} f^2(X) d\mathbb{P} = \mathbb{E}[f^2(X)] < +\infty.$$

Autrement dit, l'application linéaire

$$\Psi : \begin{cases} E & \rightarrow L^2(X) \\ f & \mapsto f(X) \end{cases}$$

est un isomorphisme de E vers $L^2(X)$. De plus, toujours par le théorème de transfert, on a :

$$\langle f, g \rangle_E = \int_{\mathbb{R}} f(x)g(x) \mathbb{P}_X(dx) = \mathbb{E}[f(X)g(X)] = \langle f(X), g(X) \rangle_{L^2(\Omega)},$$

donc Ψ conserve le produit scalaire. Finalement, Ψ est une isométrie. Mais par le théorème 3 on sait que $E = L^2(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$ est un espace de Hilbert. En particulier E est fermé. Par l'isométrie Ψ , cette propriété passe à $L^2(X)$, qui est donc bien un sous-espace fermé de $L^2(\Omega)$.



Par suite on peut appliquer le résultat de projection orthogonale ci-dessus. Soit donc Y une autre variable aléatoire de carré intégrable : il existe une unique variable aléatoire $\pi_{L^2(X)}(Y)$ dans le sous-espace $L^2(X)$ qui soit à plus courte distance de Y . Mais, dans les cas discret et continu, on a déjà traité ce problème dans le paragraphe sur la régression : c'est exactement l'espérance conditionnelle de Y sachant X . C'est pourquoi il est naturel d'en partir comme définition.

Définition 11 (Espérance conditionnelle)

Soit (X, Y) un couple aléatoire, avec $Y \in L^2(\Omega)$. L'espérance conditionnelle de Y sachant X , notée $\mathbb{E}[Y|X]$, est la projection orthogonale de Y sur le sous-espace $L^2(X)$ des variables aléatoires fonctions de X et de carré intégrable.

Ainsi l'espérance conditionnelle de Y sachant X admet une interprétation géométrique très simple (cf. figure 1.14). Cette interprétation est fructueuse, car elle permet de retrouver sans effort certaines propriétés usuelles de l'espérance conditionnelle (un bête dessin et l'affaire est entendue).

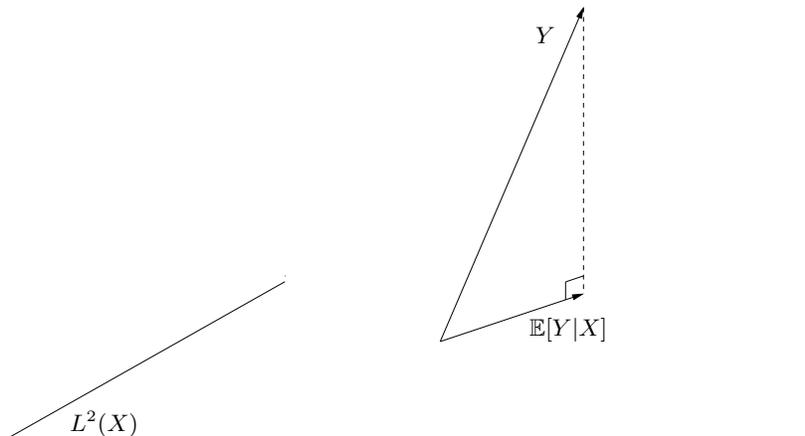


FIGURE 1.14 – L'espérance conditionnelle comme projection orthogonale.

Propriétés 6

Soit (X, Y) un couple aléatoire, avec $Y \in L^2(\Omega)$.

- Distance minimale : $\forall Z \in L^2(X), \|Y - \mathbb{E}[Y|X]\| \leq \|Y - Z\|$.
- Orthogonalité : $\forall Z \in L^2(X), \langle Y - \mathbb{E}[Y|X], Z \rangle = 0$.
- Orthogonalité(bis) : $\forall Z \in L^2(X), \langle Y, Z \rangle = \langle \mathbb{E}[Y|X], Z \rangle$.
- Pythagore : $\|Y\|^2 = \|\mathbb{E}[Y|X]\|^2 + \|Y - \mathbb{E}[Y|X]\|^2$.
- Pythagore(bis) : $\|\mathbb{E}[Y|X]\| \leq \|Y\|$, avec égalité si et seulement si Y est une fonction de X .
- Linéarité : Soit Y_1 et Y_2 de carrés intégrables, α et β deux réels, alors :

$$\mathbb{E}[\alpha Y_1 + \beta Y_2 | X] = \alpha \mathbb{E}[Y_1 | X] + \beta \mathbb{E}[Y_2 | X].$$

- Linéarité(bis) : si $u : \mathbb{R} \rightarrow \mathbb{R}$ est bornée, alors $\mathbb{E}[u(X)Y | X] = u(X)\mathbb{E}[Y | X]$.
- Positivité : Si $Y \geq 0$, alors $\mathbb{E}[Y | X] \geq 0$.
- Positivité(bis) : si Y_1 et Y_2 sont de carrés intégrables, avec $Y_1 \leq Y_2$, alors $\mathbb{E}[Y_1 | X] \leq \mathbb{E}[Y_2 | X]$.
- Calcul d'espérance par conditionnement : $\mathbb{E}[\mathbb{E}[Y | X]] = \mathbb{E}[Y]$.
- Espérance conditionnelle et indépendance : si X et Y sont indépendantes, alors $\mathbb{E}[Y | X] = \mathbb{E}[Y]$.

Remarques.

1. Par commodité des notations, toutes les propriétés ont été énoncées en termes de produits scalaires et de normes. Cependant, il faut savoir les lire aussi bien en termes d'espérances et d'espérances conditionnelles. De même, chaque fois qu'on écrit $Z \in L^2(X)$, il faut lire $Z = u(X)$, avec $u(X) \in L^2(\Omega)$. Par exemple, la propriété d'orthogonalité(bis) s'écrit encore : pour toute fonction u telle que la variable aléatoire $u(X)$ soit de carré intégrable, on a :

$$\mathbb{E}[u(X)Y] = \mathbb{E}[u(X)\mathbb{E}[Y|X]].$$

2. Le théorème de projection orthogonale permet de comprendre toutes les méthodes dites de moindres carrés en statistiques. Dans ces applications, l'espace H est engendré par les variables observables (X_1, \dots, X_n) , ou variables explicatives. Partant de celles-ci, on cherche à estimer (ou à expliquer) une autre variable Y , non observée. L'idée est de chercher une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que la variable aléatoire $f(X_1, \dots, X_n)$ approche le mieux possible Y . Les méthodes de moindres carrés sont basées sur la distance L^2 : dans ce cas, ce qui a été vu ci-dessus se généralise sans problème et l'unique solution au problème est l'espérance conditionnelle de Y sachant le n -uplet (X_1, \dots, X_n) .

1.5 Espérance conditionnelle : le cas général

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On note $L^1(\Omega, \mathcal{F}, \mathbb{P})$, ou plus simplement $L^1(\Omega)$, l'ensemble des variables aléatoires $X : \Omega \rightarrow \mathbb{R}$ intégrables, i.e. telles que :

$$\mathbb{E}|X| < +\infty,$$

en convenant toujours d'identifier deux variables aléatoires presque sûrement égales.

Dans l'étude des cas discret et continu, pour définir l'espérance conditionnelle de Y sachant X , on a vu qu'il suffisait de supposer Y intégrable. C'est pourquoi si on veut donner une définition générale, l'interprétation géométrique de $L^2(\Omega)$ n'est pas complètement satisfaisante. Néanmoins, c'est celle qu'il faudra garder en tête pour se souvenir de toutes les propriétés usuelles. Du reste, pour la définition qui suit, on part de la propriété de projection déjà vue.

Définition 12 (Espérance conditionnelle)

Soit (X, Y) un couple aléatoire, avec Y intégrable. L'espérance conditionnelle de Y sachant X est l'unique variable aléatoire fonction de X , notée $\mathbb{E}[Y|X]$, telle que pour toute fonction bornée $u : \mathbb{R} \rightarrow \mathbb{R}$, on ait :

$$\mathbb{E}[u(X)Y] = \mathbb{E}[u(X)\mathbb{E}[Y|X]].$$

Ainsi il existe une fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mesurable telle que $\mathbb{E}[Y|X] = \varphi(X)$. On retrouve alors toutes les propriétés vues ci-dessus non propres à l'espace $L^2(\Omega)$.

Propriétés 7

Soit (X, Y) un couple aléatoire, avec $Y \in L^1(\Omega)$.

- Cas d'égalité : si $Y = g(X)$ est fonction de X , alors $\mathbb{E}[Y|X] = Y$. En particulier $\mathbb{E}[X|X] = X$.
- Linéarité : Soit Y_1 et Y_2 intégrables, α et β deux réels, alors :

$$\mathbb{E}[\alpha Y_1 + \beta Y_2|X] = \alpha \mathbb{E}[Y_1|X] + \beta \mathbb{E}[Y_2|X]$$

- Linéarité(bis) : si $u : \mathbb{R} \rightarrow \mathbb{R}$ est bornée, alors $\mathbb{E}[u(X)Y|X] = u(X)\mathbb{E}[Y|X]$.
- Positivité : Si $Y \geq 0$, alors $\mathbb{E}[Y|X] \geq 0$.
- Positivité(bis) : si Y_1 et Y_2 sont intégrables, avec $Y_1 \leq Y_2$, alors $\mathbb{E}[Y_1|X] \leq \mathbb{E}[Y_2|X]$.
- Calcul d'espérance par conditionnement : $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$.
- Espérance conditionnelle et indépendance : si X et Y sont indépendantes, alors $\mathbb{E}[Y|X] = \mathbb{E}[Y]$.

1.6 Exercices

“Là où il y a une volonté, il y a un chemin.” Lénine.

Exercice 1.1 (Let's make a deal)

Vous participez à un jeu où l'on vous propose trois portes au choix. L'une des portes cache une voiture à gagner, et chacune des deux autres une chèvre. Vous choisissez une porte, mais sans l'ouvrir! L'animateur, qui sait où est la voiture, ouvre une autre porte, derrière laquelle se trouve une chèvre. Il vous donne maintenant la possibilité de vous en tenir à votre choix initial, ou de changer de porte. Qu'avez-vous intérêt à faire?

Remarque : C'est un problème auquel étaient confrontés les invités du jeu télévisé "Let's make a deal" de Monty Hall (animateur et producteur américain), sauf que les lots de consolation n'étaient pas des chèvres.

Corrigé

Supposons, sans perte de généralité, la configuration suivante : (V, C, C) , c'est-à-dire que la voiture est derrière la porte 1, les chèvres derrière les portes 2 et 3. Le jeu se déroule alors comme suit :

1. Sans changement de porte :
 - (a) le spectateur choisit la porte 1, donc l'animateur ouvre indifféremment l'une des deux autres portes, et le spectateur gagne.
 - (b) le spectateur choisit la porte 2, donc l'animateur ouvre la porte 3, et le spectateur perd.
 - (c) le spectateur choisit la porte 3, donc l'animateur ouvre la porte 2, et le spectateur perd.
2. Avec changement de porte :
 - (a) le spectateur choisit la porte 1, l'animateur ouvre indifféremment l'une des deux autres portes, le spectateur ouvre l'autre et perd.
 - (b) le spectateur choisit la porte 2, donc l'animateur ouvre la porte 3, le spectateur ouvre la porte 1 et gagne.
 - (c) le spectateur choisit la porte 3, donc l'animateur ouvre la porte 2, le spectateur ouvre la porte 1 et gagne.

Bilan des courses : s'il change de porte, il gagne 2 fois sur 3, sinon seulement 1 fois sur 3. Il vaut donc mieux changer de porte!

Exercice 1.2 (Le problème du dépistage)

1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Soit (H_1, \dots, H_n) une partition de Ω en n événements de probabilités non nulles. Soit $A \in \mathcal{F}$ tel que $\mathbb{P}(A) > 0$. Rappeler la formule de Bayes (encore appelée formule de probabilité des causes, les H_i étant les causes possibles et A la conséquence).
2. Application : Test de dépistage
 Une maladie est présente dans la population, dans la proportion d'une personne malade sur 1000. Un responsable d'un grand laboratoire pharmaceutique vient vous vanter son nouveau test de dépistage : si une personne est malade, le test est positif à 99%. Néanmoins, sur une personne non malade, le test est positif à 0.2%. Ces chiffres ont l'air excellent, vous ne pouvez qu'en convenir. Toutefois, ce qui vous intéresse, plus que les résultats présentés par le laboratoire, c'est la probabilité qu'une personne soit réellement malade lorsque son test est positif. Calculer cette probabilité.

Corrigé

1. Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé, (H_1, \dots, H_n) une partition de Ω en n événements de probabilités non nulles et $A \in \mathcal{F}$ tel que $\mathbb{P}(A) > 0$, la formule de Bayes (dite de probabilité des causes) dit que pour tout j entre 1 et n :

$$\mathbb{P}(H_j|A) = \frac{\mathbb{P}(A|H_j)\mathbb{P}(H_j)}{\sum_{i=1}^n \mathbb{P}(A|H_i)\mathbb{P}(H_i)}.$$

2. Application : Test de dépistage

Si on note A l'événement : "Le test est positif", et H l'événement : "La personne est malade", on cherche donc la probabilité $\mathbb{P}(H|A)$ et la formule de Bayes donne :

$$\mathbb{P}(H|A) = \frac{\mathbb{P}(A|H)\mathbb{P}(H)}{\mathbb{P}(A|H)\mathbb{P}(H) + \mathbb{P}(A|\bar{H})\mathbb{P}(\bar{H})}.$$

D'après l'énoncé, on a $\mathbb{P}(H) = 1/1000$, $\mathbb{P}(A|H) = 0.99$, $\mathbb{P}(A|\bar{H}) = 0.002$, les autres probabilités intervenant dans la formule de Bayes s'en déduisant facilement. Ceci donne $\mathbb{P}(H|A) \approx 1/3$. Le test n'est donc pas si fiable que ça ! Il n'empêche qu'il peut servir, en pratique, à faire une première sélection avant d'effectuer un second test plus fiable (mais plus coûteux) sur les patients pour lesquels ce premier test est positif.

Exercice 1.3 (Composition de familles)

Une population est composée de familles de 0, 1, 2 ou 3 enfants. Il y a une famille sans enfant pour 3 de 1 enfant, 4 de 2 enfants et 2 de 3 enfants. On suppose que les deux sexes sont équiprobables et qu'ils sont indépendants pour deux enfants différents.

1. Donner les probabilités de nombres d'enfants par famille p_0, p_1, p_2, p_3 .
2. On choisit une famille au hasard : quelle est la probabilité qu'il n'y ait aucun garçon ?
3. Toujours pour une famille choisie au hasard, quelle est la probabilité qu'elle ait 2 enfants sachant qu'elle n'a aucun garçon ?

Corrigé

1. On a d'après le texte : $p_1 = 3p_0$, $p_2 = 4p_0$ et $p_3 = 2p_0$. Puisque la somme des p_i fait 1, on en déduit que :

$$p = [p_0, p_1, p_2, p_3] = \left[\frac{1}{10}, \frac{3}{10}, \frac{4}{10}, \frac{2}{10} \right].$$

2. Notons G l'événement : "Il y a au moins un garçon dans la famille." On cherche donc $\mathbb{P}(\bar{G})$. Nous allons utiliser la formule des probabilités totales via la partition $\Omega = \{E_0, E_1, E_2, E_3\}$ suivant le nombre d'enfants par famille :

$$\mathbb{P}(\bar{G}) = \sum_{i=0}^3 \mathbb{P}(\bar{G}|E_i)\mathbb{P}(E_i) = \sum_{i=0}^3 \mathbb{P}(\bar{G}|E_i)p_i,$$

où il reste à voir que pour tout i on a $\mathbb{P}(\bar{G}|E_i) = (1/2)^i$. Finalement on obtient $\mathbb{P}(\bar{G}) = 3/8$.

3. On cherche cette fois la probabilité $\mathbb{P}(E_2|\bar{G})$, il suffit d'inverser le conditionnement :

$$\mathbb{P}(E_2|\bar{G}) = \frac{\mathbb{P}(\bar{G}|E_2)\mathbb{P}(E_2)}{\mathbb{P}(\bar{G})}.$$

D'après la question précédente, on sait que $\mathbb{P}(\bar{G}) = 3/8$, et d'après la première question $\mathbb{P}(E_2) = p_2 = 4/10$. On arrive donc à $\mathbb{P}(E_2|\bar{G}) = 4/15$.

Exercice 1.4 (L'ivresse du gardien de nuit)

Un gardien de nuit a 10 clés, dont une seule marche, pour ouvrir une porte. Il emploie deux méthodes. Méthode A : à jeun, il retire du trousseau les clés déjà essayées ; méthode B : ivre, il remet la clé dans le trousseau après chaque essai.

1. Méthode A : on appelle p_n la probabilité qu'il faille n essais pour ouvrir la porte. Déterminer p_n .
2. Méthode B : on appelle q_n la probabilité qu'il faille n essais pour ouvrir la porte. Déterminer q_n .
3. Le gardien est ivre un jour sur trois. Un jour, après avoir essayé 8 clés, le gardien n'a toujours pas ouvert la porte. Quelle est la probabilité qu'il soit ivre ?

Corrigé

1. Méthode A : on appelle p_n la probabilité qu'il faille n essais pour ouvrir la porte. Puisqu'il retire chaque clé après un essai infructueux, il est clair que n peut prendre les valeurs de 1 à 10. On peut calculer les probabilités de proche en proche : la probabilité p_1 est clairement $p_1 = 1/10$. Pour qu'il ouvre la porte au deuxième essai, il faut qu'il se soit trompé au premier, ce qui arrive avec probabilité $9/10$ et qu'il ait réussi au second, ce qui arrive avec probabilité $1/9$, donc à nouveau $p_2 = 1/10$. En itérant ce raisonnement, on voit sans peine que pour tout n entre 1 et 10, $p_n = 1/10$. Nous parlerons dans ce cas de loi uniforme sur l'ensemble $\{1, \dots, 10\}$.

Remarque : on pouvait obtenir ce résultat par un autre raisonnement : les 10 clés du trousseau arrivent dans un certain ordre et il n'y a aucune raison que la clé qui ouvre la porte soit à une position plutôt qu'à une autre, donc le nombre d'essais nécessaires pour ouvrir la porte est équiréparti entre 1 et 10.

2. Méthode B : cette fois, le nombre n d'essais nécessaire peut prendre toute valeur de \mathbb{N}^* . La probabilité q_1 est à nouveau $q_1 = 1/10$. Pour qu'il ouvre la porte au deuxième essai, il faut qu'il se soit trompé au premier, ce qui arrive avec probabilité $9/10$, et qu'il ait réussi au second, ce qui arrive avec probabilité $1/10$, donc $q_2 = 1/10 \times 9/10$. En itérant ce raisonnement, on voit que :

$$\forall n \in \mathbb{N}^* \quad q_n = \frac{1}{10} \left(\frac{9}{10} \right)^{n-1}.$$

On dit dans ce cas que le nombre d'essais suit une loi géométrique de paramètre $1/10$.

3. Notons $\{N > 8\}$ l'événement : "Après 8 essais, la porte n'est toujours pas ouverte" et, conformément à ce qui précède, A (resp. B) l'événement : "Le gardien est à jeun (resp. ivre)." Notons au passage que $A = \overline{B}$. On cherche donc $\mathbb{P}(B|\{N > 8\})$. On utilise la formule de Bayes :

$$\mathbb{P}(B|\{N > 8\}) = \frac{\mathbb{P}(\{N > 8\}|B)\mathbb{P}(B)}{\mathbb{P}(\{N > 8\}|A)\mathbb{P}(A) + \mathbb{P}(\{N > 8\}|B)\mathbb{P}(B)}.$$

Le texte nous apprend que $\mathbb{P}(B) = 1/3$, donc $\mathbb{P}(A) = 2/3$. Avec des notations naturelles, on obtient d'une part :

$$\mathbb{P}(\{N > 8\}|B) = \sum_{n=9}^{+\infty} q_n = \frac{1}{10} \sum_{n=9}^{+\infty} \left(\frac{9}{10} \right)^{n-1} = \left(\frac{9}{10} \right)^8,$$

puisqu'on a reconnu une série géométrique de raison $9/10$. Plus simple encore :

$$\mathbb{P}(\{N > 8\}|A) = p_9 + p_{10} = \frac{2}{10}.$$

Il vient donc $\mathbb{P}(B|E_8) \approx 0,518$.

Exercice 1.5 (Memento)

1. On dit que la variable aléatoire discrète X suit une loi géométrique de paramètre $p \in]0, 1[$ si X est à valeurs dans \mathbb{N}^* , avec $\mathbb{P}(X = k) = p(1 - p)^{k-1}$. Soit $n \in \mathbb{N}$, déterminer $\mathbb{P}(X > n)$. Montrer que X vérifie la propriété suivante, dite d'absence de mémoire :

$$\forall (m, n) \in \mathbb{N}^2 \quad \mathbb{P}(X > n + m \mid X > m) = \mathbb{P}(X > n).$$

2. Rappeler la densité d'une loi exponentielle de paramètre $\lambda > 0$, ainsi que sa fonction de répartition. Montrer que X vérifie :

$$\forall t \geq 0, \forall s \geq 0 \quad \mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s),$$

c'est-à-dire la propriété d'absence de mémoire.

3. Application : la durée de vie T en années d'une télévision suit une loi exponentielle de moyenne 8 ans. Vous possédez une telle télévision depuis 2 ans, quelle est la probabilité que sa durée de vie soit encore d'au moins 8 ans à partir de maintenant ?

Corrigé

1. Soit $X \sim \mathcal{G}(p)$ loi géométrique de paramètre $p \in]0, 1[$. On a :

$$\mathbb{P}(X > n) = \sum_{k=n+1}^{+\infty} \mathbb{P}(X = k) = \sum_{k=n+1}^{+\infty} p(1 - p)^{k-1} = p \sum_{k=n+1}^{+\infty} (1 - p)^{k-1},$$

où l'on reconnaît une somme géométrique, donc :

$$\mathbb{P}(X > n) = p \frac{(1 - p)^n}{1 - (1 - p)} = (1 - p)^n.$$

Par définition de la probabilité conditionnelle, on a alors $\forall (m, n) \in \mathbb{N} \times \mathbb{N}$:

$$\mathbb{P}(X > n + m \mid X > m) = \frac{\mathbb{P}(\{X > n + m\} \cap \{X > m\})}{\mathbb{P}(X > m)} = \frac{\mathbb{P}(X > n + m)}{\mathbb{P}(X > m)},$$

puisque l'événement $\{X > n + m\}$ implique l'événement $\{X > m\}$. Grâce au calcul précédent, on a donc :

$$\mathbb{P}(X > n + m \mid X > m) = \frac{(1 - p)^{n+m}}{(1 - p)^m} = (1 - p)^n = \mathbb{P}(X > n).$$

2. Si la variable aléatoire X suit une loi exponentielle de paramètre λ , sa densité est :

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x \geq 0\}}.$$

Sa fonction de répartition F vaut :

$$F(x) = (1 - e^{-\lambda x}) \mathbb{1}_{\{x \geq 0\}}.$$

Pour tout $s \geq 0$, on a donc :

$$\mathbb{P}(X > s) = 1 - \mathbb{P}(X \leq s) = 1 - F(s) = e^{-\lambda s}.$$

D'autre part, par le même raisonnement qu'en question précédente, il vient pour tout couple (s, t) de réels positifs :

$$\mathbb{P}(X > t + s \mid X > t) = \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(X > s),$$

donc la loi exponentielle n'a pas de mémoire.

3. Application : la probabilité cherchée s'écrit

$$\mathbb{P}(X > 2 + 8 | X > 2) = \mathbb{P}(X > 8) = e^{-1} \approx 0.37.$$

Exercice 1.6 (Loi de succession de Laplace)

On dispose de $(N + 1)$ urnes, numérotées de 0 à N . L'urne k contient k boules rouges et $(N - k)$ boules blanches. On choisit une urne au hasard. Sans connaître son numéro, on en tire n fois de suite une boule, avec remise après chaque tirage.

1. Quelle est la probabilité que le tirage suivant donne encore une boule rouge sachant que, au cours des n premiers tirages, seules des boules rouges ont été tirées? Indication : on pourra noter E_n (respectivement E_{n+1}) le fait de tirer n (respectivement $(n + 1)$) boules rouges à la suite et décomposer ces deux événements sur la partition (U_0, \dots, U_N) formée par les urnes.
2. Calculer la limite de cette probabilité lorsque N tend vers l'infini. (Rappel sur les sommes de Riemann : si f est continue sur $[0, 1]$, alors $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(k/n) = \int_0^1 f(x) dx$.)

Corrigé

1. La probabilité cherchée s'écrit, en suivant l'indication de l'énoncé :

$$p_N = \mathbb{P}(E_{n+1} | E_n) = \frac{\mathbb{P}(E_{n+1} \cap E_n)}{\mathbb{P}(E_n)} = \frac{\mathbb{P}(E_{n+1})}{\mathbb{P}(E_n)},$$

la dernière égalité venant de ce que $E_{n+1} \subseteq E_n$. Les deux termes se traitent alors de la même façon, en décomposant sur la partition $\{U_0, \dots, U_N\}$:

$$\mathbb{P}(E_n) = \sum_{k=0}^N \mathbb{P}(E_n | U_k) \mathbb{P}(U_k) = \frac{1}{N + 1} \sum_{k=0}^N \mathbb{P}(E_n | U_k),$$

le terme $\frac{1}{N+1}$ venant de l'équiprobabilité pour le choix de l'urne dans laquelle on pioche. Il reste à voir que si on pioche dans l'urne U_k , la probabilité de tirer 1 boule rouge est k/N donc la probabilité de tirer n boules rouges à la suite est $(k/N)^n$. On a donc :

$$p_N = \frac{\frac{1}{N+1} \sum_{k=0}^N (k/N)^{n+1}}{\frac{1}{N+1} \sum_{k=0}^N (k/N)^n}.$$

2. Pour trouver la limite de (p_N) lorsque le nombre N d'urnes tend vers l'infini, il suffit d'appliquer le résultat sur les sommes de Riemann :

$$\frac{1}{N + 1} \sum_{k=0}^N (k/N)^n = \frac{N}{N + 1} \left(\frac{1}{N} \sum_{k=1}^N (k/N)^n \right) \xrightarrow{N \rightarrow \infty} \int_0^1 x^n dx = \frac{1}{n + 1}.$$

On en déduit :

$$\lim_{N \rightarrow \infty} p_N = \frac{n + 1}{n + 2}.$$

Exercice 1.7 (Transmission bruitée)

Un message doit être transmis d'un point à un autre à travers N canaux successifs. Ce message peut prendre deux valeurs, 0 ou 1. Durant le passage par un canal, le message a la probabilité $p \in]0, 1[$ d'être bruité, c'est-à-dire d'être transformé en son contraire, et $(1 - p)$ d'être transmis fidèlement. Les canaux se comportent indépendamment les uns des autres.

1. Notons I_n l'événement : "en sortie de n -ème canal, le message est le même que celui transmis initialement." Exprimer $\mathbb{P}(I_{n+1})$ en fonction de $\mathbb{P}(I_n)$ et de p .

2. En notant $p_n = \mathbb{P}(I_n)$, donner une relation de récurrence entre p_{n+1} et p_n . Que vaut p_1 ?
3. On considère une suite $(u_n)_{n \geq 1}$ vérifiant la relation de récurrence :

$$u_{n+1} = (1 - 2p)u_n + p.$$

Une telle suite est dite arithmético-géométrique. Vérifier que la suite $(v_n)_{n \geq 1}$, définie par $v_n = u_n - \frac{1}{2}$, est géométrique. En déduire v_n en fonction de p et v_1 .

4. En déduire p_n en fonction de p pour tout $n \in \{1, \dots, N\}$.
5. Que vaut $\lim_{N \rightarrow +\infty} p_N$? Qu'est-ce que ce résultat a d'étonnant à première vue ?

Corrigé

1. Pour que l'événement I_{n+1} ait lieu, de deux choses l'une : ou bien I_n était réalisé et le message a été bien transmis dans le $(n+1)$ -ème canal, ou bien $\overline{I_n}$ était réalisé et le message a été mal transmis dans le $(n+1)$ -ème canal. C'est en fait la formule des probabilités totales qui s'applique ici :

$$\mathbb{P}(I_{n+1}) = \mathbb{P}(I_{n+1}|I_n)\mathbb{P}(I_n) + \mathbb{P}(I_{n+1}|\overline{I_n})\mathbb{P}(\overline{I_n}),$$

c'est-à-dire :

$$\mathbb{P}(I_{n+1}) = (1 - p)\mathbb{P}(I_n) + p(1 - \mathbb{P}(I_n)).$$

2. On a donc la relation de récurrence :

$$p_{n+1} = (1 - p)p_n + p(1 - p_n) = (1 - 2p)p_n + p.$$

La condition initiale est $p_1 = 1 - p$, probabilité que le message n'ait pas été bruité dans le premier canal.

3. On écrit :

$$v_{n+1} = u_{n+1} - \frac{1}{2} = (1 - 2p)u_n + p - \frac{1}{2},$$

et en remplaçant u_n par $v_n + \frac{1}{2}$, il vient $v_{n+1} = (1 - 2p)v_n$, donc la suite $(v_n)_{n \geq 1}$ est géométrique de raison $(1 - 2p)$. On en déduit :

$$\forall n \in \{1, \dots, N\} \quad v_n = (1 - 2p)^{n-1}v_1.$$

4. On a la même relation pour p_n que pour $u_n = v_n + \frac{1}{2}$ et puisque $p_1 = (1 - p)$, on en déduit que :

$$\forall n \in \{1, \dots, N\} \quad p_n = \frac{1}{2} + \left(\frac{1}{2} - p\right)(1 - 2p)^{n-1}.$$

5. Pour déterminer $\lim_{N \rightarrow +\infty} p_N$, on peut distinguer 3 cas :

- (a) $p = 0$: la transmission est fiable et on retrouve bien sûr $p_N = 1$ pour tout N .
- (b) $p = 1$: chaque passage dans un canal change de façon certaine le message, donc p_N dépend de la parité du nombre de canaux : $p_{2N} = 1$ et $p_{2N+1} = 0$.
- (c) $0 < p < 1$: contrairement aux deux situations précédentes, on est dans le cas d'un bruitage aléatoire. On remarque que $\lim_{N \rightarrow +\infty} (1 - 2p)^{N-1} = 0$ et $\lim_{N \rightarrow +\infty} p_N = \frac{1}{2}$. Ceci signifie que dès que le nombre de canaux devient grand, on est incapable de retrouver le message initial de façon fiable : autant tirer à pile ou face ! C'est le fameux principe du téléphone arabe.

Exercice 1.8 (La roulette de la lose)

Deux joueurs A et B jouent une succession de parties de pile ou face. A chaque coup, A a la probabilité $p \in]0, 1[$ de gagner, auquel cas B lui donne 1€, sinon le contraire. Les joueurs A et B disposent en début de partie de 50€ chacun. La partie s'arrête lorsque l'un des deux est ruiné. On cherche la probabilité que A finisse ruiné. Pour tout $n \in \{0, \dots, 100\}$, on note p_n la probabilité que A finisse ruiné s'il commence avec n € et B avec $(100 - n)$ €.

1. Que valent p_0 et p_{100} ?
2. Notons R_n l'événement : “ A finit ruiné en commençant avec n €”, c'est-à-dire que $p_n = \mathbb{P}(R_n)$. Décomposer $\mathbb{P}(R_n)$ en conditionnant par le résultat de la première partie, de façon à obtenir une relation de récurrence entre p_{n+1} , p_n et p_{n-1} .
3. On admet que la solution de cette équation est de la forme :

$$p_n = \alpha + \beta \left(\frac{1-p}{p} \right)^n.$$

Déterminer α et β .

4. En déduire la probabilité que A finisse ruiné.
5. De passage à Dinard, vous rentrez au casino et jouez à la roulette : il y a 18 numéros rouges, 18 numéros noirs et 1 numéro vert, le zéro. Vous jouez rouge pour 1€ à chaque fois. Vous commencez avec 50€ et vous arrêtez si vous avez 100€ ou si vous êtes ruiné. Pourquoi valait-il mieux aller baguenauder sur les sentiers côtiers ce jour-là ?
6. Sachant que vous commencez avec 50€ et que vous ne partirez que ruiné ou avec 100€ en poche, quelle tactique vaut-il mieux adapter pour maximiser vos chances de succès ?

Corrigé

Pour tout $n \in \{0, \dots, 100\}$, on note p_n la probabilité que A finisse ruiné s'il commence avec n € et B avec $(100 - n)$ €.

1. On a bien sûr $p_0 = 1$ et $p_{100} = 0$.
2. Supposons que A commence avec n € avec $0 < n < 100$: à la première partie, ou bien il gagne (ce qui arrive avec probabilité p) et la probabilité qu'il se ruine ensuite devient p_{n+1} , ou bien il perd (ce qui arrive avec probabilité $(1-p)$) et la probabilité qu'il se ruine ensuite devient p_{n-1} . La formule des probabilités totales s'écrit donc :

$$p_n = p \times p_{n+1} + (1-p) \times p_{n-1}.$$

3. Si pour tout $n \in \{0, \dots, 100\}$, on admet que :

$$p_n = \alpha + \beta \left(\frac{1-p}{p} \right)^n,$$

il nous reste simplement à déterminer α et β grâce aux conditions aux bords $p_0 = 1$ et $p_{100} = 0$. Notons $\theta = \frac{1-p}{p}$ afin d'alléger les notations. On a donc à résoudre le système linéaire de deux équations à deux inconnues :

$$\begin{cases} \alpha + \beta = 1 \\ \alpha + \beta\theta^{100} = 0 \end{cases} \iff \begin{cases} \alpha = \frac{\theta^{100}}{\theta^{100}-1} \\ \beta = \frac{-1}{\theta^{100}-1} \end{cases}$$

Ceci donne finalement :

$$\forall n \in \{0, \dots, 100\} \quad p_n = \frac{\theta^{100} - \theta^n}{\theta^{100} - 1}.$$

4. La probabilité que A finisse ruiné en commençant avec 50€ est donc $p_{50} = \frac{\theta^{100} - \theta^{50}}{\theta^{100} - 1}$.
5. A la roulette, la probabilité de gain à chaque partie est $p = 18/37$, donc $\theta = 19/18$, et la probabilité de finir ruiné est : $p_{50} \approx 94\%$. Il valait mieux en effet aller se promener ce jour-là...
6. Tant qu'à être prêt à perdre 50€, le mieux (ou plutôt : le moins pire) est de les miser en une seule fois. La probabilité de finir ruiné est alors simplement $p = 18/37$.

Exercice 1.9 (Un dé et une pièce)

On lance un dé équilibré, puis une pièce de monnaie non biaisée un nombre de fois égal au résultat du dé. Soit X le résultat du dé et Y le nombre de "Pile" amenés par la pièce de monnaie.

1. Déterminer la loi jointe du couple (X, Y) .
2. Soit $n \in \{1, \dots, 6\}$. Quelle est la loi de Y sachant $X = n$?
3. En déduire $\mathbb{E}[Y|X = n]$, puis $\mathbb{E}[Y|X]$.
4. Calculer $\mathbb{E}[Y]$.

Corrigé

1. Le couple (X, Y) est à valeurs dans l'ensemble E défini comme suit :

$$E = \{(n, k) : 1 \leq n \leq 6, 0 \leq k \leq n\}.$$

Pour tout couple (n, k) de E , on a alors :

$$\mathbb{P}((X, Y) = (n, k)) = \mathbb{P}(X = n)\mathbb{P}(Y = k|X = n) = \frac{1}{6} \times \frac{\binom{n}{k}}{2^n}.$$

2. Soit $n \in \{1, \dots, 6\}$, alors la loi de Y sachant $X = n$ est la loi binomiale $\mathcal{B}(n, 1/2)$.
3. On en déduit que $\mathbb{E}[Y|X = n] = n/2$, et plus généralement $\mathbb{E}[Y|X] = X/2$.
4. Il vient donc $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[X/2] = \mathbb{E}[X]/2$. Or X suit une loi uniforme sur $\{1, \dots, 6\}$ donc $\mathbb{E}[X] = (1 + 6)/2 = 7/2$. Finalement $\mathbb{E}[Y] = 7/4$.

Exercice 1.10 (Minimum et maximum)

On tire deux variables U et V de façon indépendante et uniformément dans l'ensemble $\{1, 2, 3, 4, 5\}$. On en déduit les variables aléatoires $X = \min(U, V)$ et $Y = \max(U, V)$.

1. Déterminer la loi jointe du couple (U, Y) .
2. Déterminer $\mathbb{E}[U|Y = n]$, pour $n \in \{1, 2, 3, 4, 5\}$.
3. En déduire $\mathbb{E}[U|Y]$.
4. Déterminer $\mathbb{E}[Y|U]$.
5. Déterminer de même $\mathbb{E}[U|X]$ et $\mathbb{E}[X|U]$.

Corrigé

1. La loi jointe du couple (U, Y) est donnée figure 1.15.
2. Soit $n \in \{1, 2, 3, 4, 5\}$ fixé. Alors si $Y = n$, puisque Y est le maximum de U et V , il est clair que U peut prendre les valeurs de 1 à n . On a donc :

$$\mathbb{E}[U|Y = n] = \mathbb{P}(U = 1|Y = n) + 2\mathbb{P}(U = 2|Y = n) + \dots + n\mathbb{P}(U = n|Y = n).$$

Il reste à préciser les probabilités :

$$\mathbb{P}(U = k|Y = n) = \frac{\mathbb{P}(U = k, Y = n)}{\mathbb{P}(Y = n)}.$$

$U \backslash Y$	1	2	3	4	5
1	1/25	1/25	1/25	1/25	1/25
2	0	2/25	1/25	1/25	1/25
3	0	0	3/25	1/25	1/25
4	0	0	0	4/25	1/25
5	0	0	0	0	5/25

FIGURE 1.15 – Loi jointe pour le couple (U, Y) .

Puisqu'on connaît la loi jointe, il reste à préciser la loi marginale de Y , c'est-à-dire sommer sur les colonnes dans le tableau de la question précédente. Ce qui donne :

$$\mathbb{P}(Y = n) = \mathbb{P}(U = 1, Y = n) + \dots + \mathbb{P}(U = n, Y = n) = \frac{1}{25}(n-1) + \frac{1}{25}n = \frac{2n-1}{25}.$$

Ainsi, on obtient pour la loi conditionnelle de U sachant Y :

$$\mathbb{P}(U = k|Y = n) = \begin{cases} 1/(2n-1) & \text{si } 1 \leq k \leq (n-1) \\ n/(2n-1) & \text{si } k = n \end{cases}$$

Au total, on obtient :

$$\mathbb{E}[U|Y = n] = (1 + \dots + (n-1))\frac{1}{2n-1} + n\frac{n}{2n-1}.$$

La première somme, entre parenthèses, est arithmétique de raison 1, donc :

$$1 + \dots + (n-1) = \frac{n(n-1)}{2},$$

d'où finalement :

$$\mathbb{E}[U|Y = n] = \frac{n(n-1)}{2(2n-1)} + \frac{n^2}{2n-1} = \frac{n(3n-1)}{2(2n-1)}.$$

Remarque : quand vous arrivez ici, après quelques calculs, pensez à vérifier que la formule fonctionne, par exemple pour $n = 1$ et $n = 2$.

3. On en déduit que :

$$\mathbb{E}[U|Y] = \frac{Y(3Y-1)}{2(2Y-1)}.$$

4. Pour déterminer $\mathbb{E}[Y|U]$, on commence par calculer $\mathbb{E}[Y|U = n]$ pour tout $n \in \{1, 2, 3, 4, 5\}$. Lorsque U vaut n , il est clair que Y peut prendre les valeurs $n, \dots, 5$. Comme ci-dessus, il faut donc commencer par préciser la loi marginale de U . Or U est obtenue en tirant un nombre au hasard entre 1 et 5, donc U suit une loi uniforme sur l'ensemble $\{1, 2, 3, 4, 5\}$: $\mathbb{P}(U = n) = \frac{1}{5}$. On en déduit que :

$$\mathbb{P}(Y = k|U = n) = \begin{cases} 1/5 & \text{si } (n+1) \leq k \leq 5 \\ n/5 & \text{si } k = n \end{cases}$$

On en déduit l'espérance conditionnelle de Y sachant $U = n$:

$$\mathbb{E}[Y|U = n] = n\frac{n}{5} + ((n+1) + \dots + 5)\frac{1}{5}.$$

On reconnaît à nouveau une somme arithmétique dans la parenthèse :

$$(n+1) + \dots + 5 = \frac{(n+6)(5 - (n+1) + 1)}{2} = \frac{(n+6)(5-n)}{2},$$

et finalement on obtient :

$$\mathbb{E}[Y|U = n] = \frac{n^2 - n + 30}{10}.$$

Et l'espérance conditionnelle de Y sachant U est donc :

$$\mathbb{E}[Y|U] = \frac{U^2 - U + 30}{10}.$$

$X \backslash U$	1	2	3	4	5
1	5/25	0	0	0	0
2	1/25	4/25	0	0	0
3	1/25	1/25	3/25	0	0
4	1/25	1/25	1/25	2/25	0
5	1/25	1/25	1/25	1/25	1/25

FIGURE 1.16 – Loi jointe pour le couple (U, X) .

5. Pour déterminer $\mathbb{E}[U|X]$, on reprend pas à pas le raisonnement vu ci-dessus. La loi jointe du couple aléatoire (U, X) est représentée figure 1.16. Pour tout n entre 1 et 5, on a cette fois :

$$\mathbb{E}[U|X = n] = n\mathbb{P}(U = n|X = n) + \dots + 5\mathbb{P}(U = 5|X = n).$$

Pour la loi marginale de X , on a : $\mathbb{P}(X = n) = \frac{11-2n}{25}$. Ce qui donne pour la loi conditionnelle de U sachant $X = n$:

$$\mathbb{P}(U = k|X = n) = \begin{cases} 1/(11-2n) & \text{si } (n+1) \leq k \leq 5 \\ (6-n)/(11-2n) & \text{si } k = n \end{cases}$$

On a donc :

$$\mathbb{E}[U|X = n] = n\frac{6-n}{11-2n} + ((n+1) + \dots + 5)\frac{1}{11-2n} = \frac{30 + 11n - 3n^2}{22 - 4n}.$$

Donc finalement :

$$\mathbb{E}[U|X] = \frac{30 + 11X - 3X^2}{22 - 4X}.$$

Pour calculer l'espérance conditionnelle de X sachant U , on a deux possibilités : ou bien on reprend la méthode plan-plan ci-dessus, ou bien on pense à une ruse de sioux. Il suffit en effet de remarquer, puisque l'espérance conditionnelle est linéaire, que :

$$\mathbb{E}[X + Y|U] = \mathbb{E}[X|U] + \mathbb{E}[Y|U],$$

or $X + Y = U + V$, puisque si X est égal à U , Y est égal à V et vice-versa. Donc :

$$\mathbb{E}[X + Y|U] = \mathbb{E}[U + V|U] = \mathbb{E}[U|U] + \mathbb{E}[V|U],$$

et on utilise les propriétés classiques de l'espérance conditionnelle : $\mathbb{E}[U|U] = U$ d'une part, et $\mathbb{E}[V|U] = \mathbb{E}[V]$ d'autre part, puisque U et V sont indépendantes. Si on fait les comptes, on a donc obtenu :

$$\mathbb{E}[X|U] = U + \mathbb{E}[V] - \mathbb{E}[Y|U].$$

Or $\mathbb{E}[V] = 3$ puisque V suit une loi uniforme et on a calculé $\mathbb{E}[Y|U]$ ci-dessus. Finalement :

$$\mathbb{E}[X|U] = \frac{11U - U^2}{10},$$

et tout est dit.

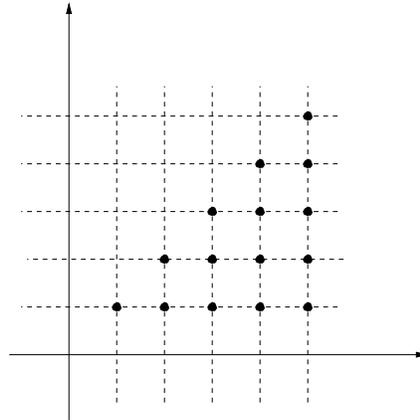


FIGURE 1.17 – Tirage uniforme dans un triangle.

Exercice 1.11 (Des points dans \mathbb{N}^2)

On tire un point de façon uniforme parmi ceux de la figure 1.17. Ceci donne un couple aléatoire (X, Y) dont la loi jointe est : $p_{ij} = \mathbb{P}(X = i, Y = j) = \frac{1}{15}$, $1 \leq j \leq i \leq 5$.

1. Donner les lois marginales de X et de Y .
2. Soit $j \in \{1, \dots, 5\}$ fixé : donner la loi conditionnelle de X sachant $Y = j$, c'est-à-dire $\mathbb{P}(X = i|Y = j)$.
3. Calculer $\mathbb{E}[X|Y = j]$, en déduire $\mathbb{E}[X|Y]$, puis $\mathbb{E}[X]$ en fonction de $\mathbb{E}[Y]$.
4. Déterminer de même la loi conditionnelle de Y sachant $X = i$.
5. Calculer $\mathbb{E}[Y|X = i]$, en déduire $\mathbb{E}[Y|X]$, puis $\mathbb{E}[Y]$ en fonction de $\mathbb{E}[X]$.

6. Dédurre des questions précédentes $\mathbb{E}[X]$ et $\mathbb{E}[Y]$.
7. Généralisation : soit $N \in \mathbb{N}^*$ fixé, reprendre les questions précédentes en remplaçant 5 par N (on ne demande pas le détail des calculs, uniquement les résultats).

Corrigé

Le corrigé est donné en annexe (sujet de juin 2006).

Exercice 1.12 (Somme aléatoire de variables aléatoires)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires admettant la même espérance $m = \mathbb{E}[X_1]$. Soit N une variable aléatoire à valeurs dans \mathbb{N}^* indépendante de la suite $(X_n)_{n \geq 1}$. On pose $S_n = \sum_{k=1}^n X_k$. On s'intéresse dans cet exercice à la variable aléatoire S_N .

1. Si $N \sim \mathcal{G}(1/2)$ et les variables X_n sont équiprobables sur $\{1, \dots, 6\}$, donner une façon de simuler S_N à l'aide d'un dé et d'une pièce.
2. Déterminer $\mathbb{E}[S_N | N = n]$. En déduire $\mathbb{E}[S_N | N]$.
3. Que vaut $\mathbb{E}[S_N]$?
4. Application : la fièvre acheteuse. Le nombre de clients se rendant dans un magasin donné dans l'espace d'une journée est une variable aléatoire de moyenne 50. La somme dépensée par chacun des clients est aussi une variable aléatoire de moyenne 20€. Avec des hypothèses raisonnables, quel est le chiffre d'affaires quotidien moyen du magasin ?

Corrigé

1. Tout d'abord, il suffit de lancer une pièce équilibrée et de compter le nombre de lancers nécessaires pour voir apparaître Pile. On appelle N ce nombre, on sait qu'il suit une loi géométrique de paramètre $1/2$. Il suffit alors de lancer N fois un dé équilibré à 6 faces et de faire la somme des résultats obtenus pour obtenir S_N .
2. Si $N = n$, alors :

$$S_N = S_n = X_1 + \dots + X_n$$

est la somme de n variables de même moyenne m , donc :

$$\mathbb{E}[S_N | N = n] = nm.$$

On en déduit que :

$$\mathbb{E}[S_N | N] = mN.$$

3. La technique de calcul d'espérance par conditionnement permet alors d'écrire que :

$$\mathbb{E}[S_N] = \mathbb{E}[\mathbb{E}[S_N | N]] = m\mathbb{E}[N].$$

4. Application : la fièvre acheteuse. Le nombre de clients se rendant dans un magasin donné dans l'espace d'une journée est une variable aléatoire de moyenne 50. La somme dépensée par chacun des clients est aussi une variable aléatoire de moyenne 20€. En supposant que ce que dépense chaque client (variable aléatoire X_i) est indépendant du nombre N de clients à passer dans le magasin dans la journée, on en déduit que le chiffre d'affaires quotidien moyen du magasin est :

$$\mathbb{E}[S_N] = 50 \times 20 = 1000\text{€}.$$

Exercice 1.13 (North by Northwest)

Soit X et Y deux variables aléatoires discrètes à valeurs dans \mathbb{N} . On suppose que $X \sim \mathcal{P}(\lambda)$, loi de Poisson de paramètre $\lambda > 0$. On suppose que, pour tout entier $n > 0$, la loi de Y sachant $X = n$ est la loi binômiale $\mathcal{B}(n, p)$; et que $Y = 0$ si $X = 0$.

1. Donner la loi jointe du couple aléatoire (X, Y) .
2. Montrer que Y suit une loi de Poisson de paramètre $p\lambda$.
3. Montrer que :

$$\forall n \geq k \geq 0 \quad \mathbb{P}(X = n | Y = k) = e^{-(1-p)\lambda} \frac{((1-p)\lambda)^{n-k}}{(n-k)!},$$

c'est-à-dire que, sachant $Y = k$, X suit une loi de Poisson translatée. En déduire $\mathbb{E}[X|Y = k]$ et de façon générale que :

$$\mathbb{E}[X|Y] = Y + \lambda(1-p).$$

4. Application : à un embranchement routier, le nombre X de véhicules arrivant en une heure suit une loi de Poisson $\mathcal{P}(100)$ (hypothèse courante dans ce genre de situation). Les véhicules ont alors le choix entre deux directions A ou B : ils choisissent A avec la même probabilité $1/3$, et ce de façon indépendante. Sachant qu'en une heure, on sait simplement que 100 voitures ont pris la direction A , quel est le nombre moyen de voitures qui sont passées par l'embranchement ?

Corrigé

1. Le couple (X, Y) est à valeurs dans l'ensemble E défini comme suit :

$$E = \{(n, k) : 0 \leq k \leq n\}.$$

Pour tout couple (n, k) de E , on a alors :

$$\mathbb{P}((X, Y) = (n, k)) = \mathbb{P}(X = n)\mathbb{P}(Y = k | X = n) = e^{-\lambda} \frac{\lambda^n}{n!} \times \binom{n}{k} p^k (1-p)^{n-k},$$

quantité qui s'écrit encore, puisque $n = k + (n - k)$ et $\lambda = p\lambda + (1-p)\lambda$:

$$\mathbb{P}((X, Y) = (n, k)) = e^{-p\lambda} \frac{(p\lambda)^k}{k!} \times e^{-(1-p)\lambda} \frac{((1-p)\lambda)^{n-k}}{(n-k)!}.$$

2. La variable Y est à valeurs dans \mathbb{N} et pour tout entier naturel k , on a :

$$\mathbb{P}(Y = k) = \sum_{n=k}^{+\infty} \mathbb{P}(X = n, Y = k) = e^{-p\lambda} \frac{(p\lambda)^k}{k!} \times e^{-(1-p)\lambda} \sum_{n=k}^{+\infty} \frac{((1-p)\lambda)^{n-k}}{(n-k)!},$$

et l'on réindexe la somme pour reconnaître la série de l'exponentielle :

$$\sum_{n=k}^{+\infty} \frac{((1-p)\lambda)^{n-k}}{(n-k)!} = \sum_{m=0}^{+\infty} \frac{((1-p)\lambda)^m}{m!} = e^{(1-p)\lambda}.$$

Tout compte fait, pour tout entier naturel k :

$$\mathbb{P}(Y = k) = e^{-p\lambda} \frac{(p\lambda)^k}{k!}.$$

En d'autres termes, Y suit une loi de Poisson de paramètre $p\lambda$.

3. Pour $n \geq k \geq 0$, on peut alors écrire :

$$\mathbb{P}(X = n|Y = k) = \frac{\mathbb{P}(X = n, Y = k)}{\mathbb{P}(Y = k)},$$

où numérateur et dénominateur ont été calculés dans les questions précédentes, ce qui permet d'aboutir à :

$$\forall n \geq k \geq 0 \quad \mathbb{P}(X = n|Y = k) = e^{-(1-p)\lambda} \frac{((1-p)\lambda)^{n-k}}{(n-k)!},$$

c'est-à-dire que, sachant $Y = k$, X suit en effet une loi de Poisson translatée. Pour en déduire $\mathbb{E}[X|Y = k]$ et pour reprendre la formule chère aux amateurs de westerns spaghettis, "Le monde se divise en deux catégories : ceux qui ont le revolver chargé, et ceux qui creusent", i.e. sans ou avec calculs. Optons pour la première option : lorsque $Y = k$, X suit une loi de Poisson de paramètre $(1-p)\lambda$ translatée de k , c'est-à-dire que $X - k$ suit exactement une loi de Poisson de paramètre $(1-p)\lambda$. On en déduit que $\mathbb{E}[X - k|Y = k] = (1-p)\lambda$, autrement dit $\mathbb{E}[X|Y = k] = k + (1-p)\lambda$ et plus généralement :

$$\mathbb{E}[X|Y] = Y + \lambda(1-p).$$

4. Nous sommes exactement dans le cadre d'application de ce qui précède, avec $\lambda = 100$, $p = 1/3$ et $k = 100$. Le nombre moyen de voitures qui sont passées par l'embranchement vaut donc :

$$\mathbb{E}[X|Y = 100] = 100 + (1 - 1/3) \times 100 \approx 167.$$

Un raisonnement moisi serait le suivant : pour une voiture qui prend la direction A , deux prennent la direction B , donc si on a vu passer 100 voitures en une heure partant vers A , 200 voitures ont dû prendre la direction B , donc au total 300 ont dû se présenter à l'embranchement. Où est la faute ? Le fait que les choix de A ou B sont indépendants implique que l'information sur l'un n'apporte aucune information sur l'autre. Donc quel que soit le nombre de voitures partant vers A en une heure, le nombre moyen de voitures vers B est inchangé et vaut environ 67. Formellement, ceci peut se voir en utilisant les propriétés classiques de l'espérance conditionnelle :

$$\mathbb{E}[X|Y] = \mathbb{E}[Y + Z|Y] = \mathbb{E}[Y|Y] + \mathbb{E}[Z|Y] = Y + \mathbb{E}[Z],$$

où la relation $\mathbb{E}[Z|Y] = \mathbb{E}[Z]$ vient de l'indépendance de Y et Z .

Exercice 1.14 (Espérance d'une variable géométrique)

Soit une urne contenant N boules noires et M boules blanches ($N, M \geq 1$). On pose $p = \frac{N}{N+M}$. On effectue une suite de tirages avec remise et on désigne par T le nombre de tirages nécessaires pour amener pour la première fois une boule noire.

1. Quelle est la loi de T ? Que vaut $\mathbb{E}[T]$?
2. On calcule ici l'espérance de T par une autre méthode. On introduit une variable X qui prend la valeur 0 ou 1 selon que la première boule tirée est blanche ou noire.
 - (a) Déterminer $\mathbb{E}[T|X = 1]$.
 - (b) Déterminer $\mathbb{E}[T|X = 0]$ en fonction de $\mathbb{E}[T]$.
 - (c) Via un calcul d'espérance par conditionnement, en déduire $\mathbb{E}[T]$.

Corrigé

1. La variable T est à valeurs dans \mathbb{N}^* et pour tout entier naturel non nul n , il faut $(n-1)$ échecs et finalement un succès pour que T vaille n . Puisque les tirages sont indépendants, ceci se traduit par

$$\mathbb{P}(T = n) = p(1-p)^{n-1}$$

On dit que T suit une loi géométrique de paramètre p et on note $T \sim \mathcal{G}(p)$. Le cadre général est le suivant : lors d'une expérience, un événement donné a la probabilité p de survenir ; si on répète cette expérience de façon indépendante jusqu'à ce que cet événement apparaisse, le nombre d'expériences nécessaires suit la loi géométrique de paramètre p . On prendra garde au fait que le paramètre p de cette loi ne correspond pas à la raison de la suite des probabilités $(\mathbb{P}(T = n))_{n \in \mathbb{N}^*}$, lequel vaut $(1-p)$.

Le calcul de l'espérance de T se fait alors comme suit :

$$\mathbb{E}[T] = \sum_{n=1}^{\infty} np(1-p)^{n-1} = p \sum_{n=1}^{\infty} n(1-p)^{n-1}.$$

Il suffit alors de se souvenir de la somme d'une série géométrique et de dériver terme à terme pour obtenir le résultat voulu :

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \Rightarrow \frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} nx^{n-1} \Rightarrow \frac{1}{p^2} = \sum_{n=1}^{\infty} n(1-p)^{n-1}$$

ce qui donne $\mathbb{E}[T] = 1/p$. Interprétation élémentaire : plus l'événement d'intérêt est rare, autrement dit plus p est faible, et en moyenne plus il faut attendre pour le voir apparaître.

2. (a) Il est clair que $\mathbb{E}[T|X = 1] = 1$.
 (b) Si $X = 0$, la première boule tirée est blanche et on revient au point de départ (hormis qu'on a déjà fait un tirage), ce qui se traduit par $\mathbb{E}[T|X = 0] = 1 + \mathbb{E}[T]$.
 (c) Tenant compte de ce que $\mathbb{P}(X = 0) = 1 - \mathbb{P}(X = 1) = 1 - p$, on peut alors écrire

$$\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T|X]] = \mathbb{E}[T|X = 0]\mathbb{P}(X = 0) + \mathbb{E}[T|X = 1]\mathbb{P}(X = 1) = (1 + \mathbb{E}[T])(1-p) + p$$

ce qui conduit bien à $\mathbb{E}[T] = 1/p$.

Exercice 1.15 (Germinal revival)

Un mineur est prisonnier dans un puits d'où partent trois tunnels. Le premier tunnel le mènerait à la sortie au bout de 3 heures de marche. Le second le ramènerait à son point de départ au bout de 5 heures de marche, de même que le troisième au bout de 7 heures. On suppose que les tunnels sont indiscernables et qu'à chaque fois qu'il est au point de départ, le mineur emprunte l'un des trois de façon équiprobable. On note T le nombre d'heures nécessaires pour sortir du puits. Soit $X \in \{1, 2, 3\}$ le numéro du tunnel que le prisonnier choisit à sa première tentative. Reprendre le raisonnement de l'exercice 1.14 (partie 2.) pour calculer le temps moyen qu'il faut au mineur pour sortir.

Corrigé

En s'inspirant de l'exercice 1.14, on obtient : $\mathbb{E}[T|X = 1] = 3$, $\mathbb{E}[T|X = 2] = \mathbb{E}[T] + 5$ et $\mathbb{E}[T|X = 3] = \mathbb{E}[T] + 7$. Par ailleurs, le calcul d'espérance par conditionnement donne :

$$\mathbb{E}[T] = \mathbb{E}[T|X = 1]\mathbb{P}(X = 1) + \mathbb{E}[T|X = 2]\mathbb{P}(X = 2) + \mathbb{E}[T|X = 3]\mathbb{P}(X = 3).$$

Mais puisque le prisonnier choisit au hasard parmi les trois tunnels, on a :

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \mathbb{P}(X = 3) = \frac{1}{3}.$$

Ainsi il vient :

$$\mathbb{E}[T] = \frac{1}{3}(3 + \mathbb{E}[T] + 5 + \mathbb{E}[T] + 7),$$

d'où l'on déduit :

$$\mathbb{E}[T] = 15.$$

Il faut en moyenne 15 heures au mineur pour sortir. Autant dire qu'il n'a pas le cul sorti des ronces...

Exercice 1.16 (Variable Y définie à partir de X)

On considère une variable aléatoire X à valeurs dans \mathbb{N}^* et telle que :

$$\forall i \in \mathbb{N}^* \quad \mathbb{P}(X = i) = \frac{2}{3^i}.$$

Soit Y une variable aléatoire telle que, sachant $X = i$, la loi de Y est l'équiprobabilité sur $\{i, i+1\}$.

1. Que vaut $\mathbb{E}[X]$?
2. Pour tout $i \in \mathbb{N}^*$, déterminer $\mathbb{E}[Y|X = i]$. En déduire $\mathbb{E}[Y|X]$, puis $\mathbb{E}[Y]$.
3. Calculer la loi jointe du couple (X, Y) .
4. Déterminer la loi de Y .
5. Pour tout $j \in \mathbb{N}^*$, déterminer $\mathbb{E}[X|Y = j]$. En déduire $\mathbb{E}[X|Y]$.
6. Calculer $\text{Cov}(X, Y)$.

Corrigé

1. On a vu que X suit une loi géométrique de paramètre $2/3$ donc $\mathbb{E}[X] = 3/2$.
2. Pour tout $i \in \mathbb{N}^*$, on a vu que :

$$\mathbb{E}[Y|X = i] = \frac{1}{2}(i + (i + 1)) = \frac{2i + 1}{2}.$$

On en déduit que

$$\mathbb{E}[Y|X] = \frac{2X + 1}{2},$$

et par suite

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \frac{1}{2}(2\mathbb{E}[X] + 1) = 2.$$

3. La loi jointe du couple (X, Y) est très simple puisqu'on connaît marginale et conditionnelle. Pour tout $i \in \mathbb{N}^*$, on a :

$$p_{i,j} = \begin{cases} 0 & \text{si } j \notin \{i, i+1\} \\ \frac{1}{3^i} & \text{si } j \in \{i, i+1\} \end{cases}$$

4. La variable aléatoire Y est à valeurs dans \mathbb{N}^* , avec

$$p_j = \mathbb{P}(Y = j) = \begin{cases} 1/3 & \text{si } j = 1 \\ \frac{4}{3^j} & \text{si } j \geq 2 \end{cases}$$

5. On commence par déterminer la loi conditionnelle de X sachant $Y = j$. Or si $Y = 1$, il est clair que X vaut 1, donc que $\mathbb{E}[X|Y = 1] = 1$. Si $Y = j > 1$, alors X ne peut valoir que j ou $(j - 1)$ et plus précisément :

$$\mathbb{P}(X = j - 1|Y = j) = \frac{\mathbb{P}(X = j - 1, Y = j)}{\mathbb{P}(Y = j)} = \frac{3}{4},$$

et de même :

$$\mathbb{P}(X = j|Y = j) = \frac{1}{4},$$

d'où l'on déduit :

$$\mathbb{E}[X|Y = j] = \frac{3(j-1)}{4} + \frac{j}{4} = \frac{4j-3}{4}$$

On est donc obligé de faire attention à la valeur 1 pour la variable aléatoire X :

$$\mathbb{E}[X|Y] = \mathbb{1}_{\{Y=1\}} + \frac{4Y-3}{4} \mathbb{1}_{\{Y>1\}}$$

6. On a enfin

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

or on a déjà vu que $\mathbb{E}[X] = 3/2$ et $\mathbb{E}[Y] = 2$ et

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY|X]] = \mathbb{E}[X\mathbb{E}[Y|X]] = \mathbb{E}\left[X \frac{2X+1}{2}\right] = \frac{1}{2}(2\mathbb{E}[X^2] + \mathbb{E}[X]).$$

Il reste à voir que $\mathbb{E}[X^2] = \text{Var}X + (\mathbb{E}[X])^2$, et à se souvenir (ou à recalculer) que la variance d'une loi géométrique de paramètre p est q/p^2 (donc ici $3/4$). On a donc :

$$\mathbb{E}[XY] = \frac{1}{2} \left(2 \left(\frac{3}{4} + \frac{9}{4} \right) + \frac{3}{2} \right) = \frac{15}{4}.$$

Finalement :

$$\text{Cov}(X, Y) = \frac{15}{4} - 3 = \frac{3}{4}.$$

Exercice 1.17 (Couple aléatoire)

Rappels sur les séries entières : Pour tout $x \in [-1, 1[$, on a :

$$\ln(1-x) = - \sum_{n=1}^{+\infty} \frac{x^n}{n},$$

Pour tout $x \in [-1, 1[$ et pour tout entier naturel k :

$$\frac{k!}{(1-x)^{k+1}} = \sum_{n=0}^{+\infty} \frac{(n+k)!}{n!} x^n.$$

On considère un couple aléatoire (X, Y) à valeurs dans $\mathbb{N}^2 \setminus \{(0, 0)\}$ dont la loi jointe est définie par :

$$\forall (i, j) \in \mathbb{N}^2 \setminus \{(0, 0)\} \quad \mathbb{P}(X = i, Y = j) = \frac{1}{\ln 2} \frac{(i+j-1)!}{i!j!3^i6^j}.$$

1. Calculer $\mathbb{P}(X = 0)$.
2. Pour tout $i \in \mathbb{N}^*$, calculer $\mathbb{P}(X = i)$.
3. Déterminer la loi de Y conditionnellement à $X = 0$. Calculer $\mathbb{E}[Y|X = 0]$.
4. Pour tout $i \in \mathbb{N}^*$, déterminer la loi de Y conditionnellement à $X = i$. Calculer $\mathbb{E}[Y|X = i]$.
5. En déduire $\mathbb{E}[Y|X]$.

6. En déduire l'égalité suivante :

$$\mathbb{E}[Y] - \frac{1}{5}\mathbb{E}[X] = \frac{1}{5 \ln 2}.$$

Corrigé

1. Lorsque $X = 0$, Y peut prendre les valeurs 1, 2, etc. On a donc :

$$\mathbb{P}(X = 0) = \sum_{j=1}^{+\infty} \mathbb{P}(X = 0, Y = j) = \sum_{j=1}^{+\infty} \frac{1}{\ln 2} \frac{(j-1)!}{j!6^j} = \frac{1}{\ln 2} \sum_{j=1}^{+\infty} \frac{1}{j6^j},$$

et il suffit alors d'appliquer la formule donnée en rappel pour obtenir :

$$\mathbb{P}(X = 0) = \frac{\ln 6 - \ln 5}{\ln 2}.$$

2. Lorsque $X = i > 0$, Y peut prendre les valeurs 0, 1, 2, etc. On a cette fois :

$$\mathbb{P}(X = i) = \sum_{j=0}^{+\infty} \frac{1}{\ln 2} \frac{(i+j-1)!}{i!j!3^i6^j} = \frac{1}{\ln 2} \frac{1}{i!3^i} \sum_{j=0}^{+\infty} \frac{(j+(i-1))!}{j!6^j},$$

et on applique la formule du rappel :

$$\mathbb{P}(X = i) = \frac{1}{\ln 2} \frac{1}{i!3^i} (i-1)! \frac{6^i}{5^i} = \frac{1}{\ln 2} \frac{2^i}{i5^i}.$$

3. Pour tout $j > 0$, on a donc :

$$\mathbb{P}(Y = j|X = 0) = \frac{\mathbb{P}(X = 0, Y = j)}{\mathbb{P}(X = 0)} = \frac{1}{(\ln 6 - \ln 5)j6^j}.$$

La valeur moyenne de Y sachant $X = 0$ est donc :

$$\mathbb{E}[Y|X = 0] = \sum_{j=1}^{+\infty} j\mathbb{P}(Y = j|X = 0) = \frac{1}{\ln 6 - \ln 5} \sum_{j=1}^{+\infty} \frac{1}{6^j},$$

et on reconnaît une série géométrique :

$$\mathbb{E}[Y|X = 0] = \frac{1}{5(\ln 6 - \ln 5)}.$$

4. Soit $i > 0$ fixé. Pour tout $j \geq 0$, on a :

$$\mathbb{P}(Y = j|X = i) = \frac{\mathbb{P}(X = i, Y = j)}{\mathbb{P}(X = i)} = \frac{(i+j-1)!5^i}{(i-1)!j!6^{i+j}}.$$

D'où l'on déduit l'espérance conditionnelle de Y sachant $X = i$:

$$\mathbb{E}[Y|X = i] = \sum_{j=0}^{+\infty} j\mathbb{P}(Y = j|X = i) = \frac{5^i}{(i-1)!6^i} \sum_{j=1}^{+\infty} \frac{((j-1)+i)!}{(j-1)!} \frac{1}{6^j},$$

ce qui s'écrit encore :

$$\mathbb{E}[Y|X = i] = \frac{5^i}{(i-1)!6^{i+1}} \sum_{n=0}^{+\infty} \frac{(n+i)!}{n!} \frac{1}{6^n},$$

c'est-à-dire, après simplifications :

$$\mathbb{E}[Y|X = i] = \frac{i}{5}.$$

5. Pour l'expression de l'espérance conditionnelle de Y sachant X , il faut donc faire attention à $X = 0$:

$$\mathbb{E}[Y|X] = \frac{1}{5(\ln 6 - \ln 5)} \mathbb{1}_{\{X=0\}} + \frac{1}{5}X \mathbb{1}_{\{X>0\}} = \frac{1}{5(\ln 6 - \ln 5)} \mathbb{1}_{\{X=0\}} + \frac{1}{5}X.$$

6. On en déduit l'espérance de Y en fonction de celle de X :

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \frac{1}{5(\ln 6 - \ln 5)} \mathbb{E}[\mathbb{1}_{\{X=0\}}] + \frac{1}{5} \mathbb{E}[X].$$

La variable aléatoire $\mathbb{1}_{\{X=0\}}$ est binaire, elle prend les valeurs 0 et 1 avec les probabilités respectives $\mathbb{P}(X > 0)$ et $\mathbb{P}(X = 0)$, donc son espérance est tout simplement :

$$\mathbb{E}[\mathbb{1}_{\{X=0\}}] = 0 \times \mathbb{P}(X > 0) + 1 \times \mathbb{P}(X = 0) = \mathbb{P}(X = 0) = \frac{\ln 6 - \ln 5}{\ln 2}.$$

On en déduit que :

$$\mathbb{E}[Y] = \frac{1}{5 \ln 2} + \frac{1}{5} \mathbb{E}[X],$$

ce qui est bien le résultat voulu.

Remarque. Une version plus générale de cet exercice se trouve dans l'ouvrage *Toutes les probabilités et les statistiques*, de Jacques Dauxois et Claudie Hassenforder, Ellipses, 2004.

Exercice 1.18 (Echauffement)

On considère la fonction f définie sur \mathbb{R}^2 par :

$$f(x, y) = e^{-(x+y)} \mathbb{1}_{\{x \geq 0, y \geq 0\}}$$

1. Vérifier que f est une densité sur \mathbb{R}^2 .
2. Soit (X, Y) un couple de densité f . Déterminer les marginales $f(x)$ et $f(y)$.
3. Calculer la covariance du couple (X, Y) .

Corrigé

1. Il est clair que f est une fonction positive. Par ailleurs, par le théorème de Fubini-Tonelli, le calcul de son intégrale double sur \mathbb{R}^2 se fait sans problème :

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = \int_0^{+\infty} \left(\int_0^{+\infty} e^{-(x+y)} dy \right) dx = \dots = 1,$$

et f est bien une densité sur \mathbb{R}^2 .

2. La densité $f(x)$ de la variable aléatoire X s'obtient en intégrant par rapport à y :

$$f(x) = \int_{\mathbb{R}} f(x, y) dy = \dots = e^{-x} \mathbb{1}_{[0, +\infty[}(x).$$

On voit donc que X suit une loi exponentielle de paramètre 1, ce que l'on note : $X \sim \mathcal{E}(1)$. Vu les rôles symétriques joués par X et Y , la variable aléatoire Y a la même loi : $Y \sim \mathcal{E}(1)$.

3. On remarque que :

$$\forall (x, y) \in \mathbb{R}^2 \quad f(x, y) = f(x)f(y),$$

donc X et Y sont indépendantes, donc leur covariance est nulle (rappelons que la réciproque est fautive en général, sauf dans le cas des vecteurs gaussiens).

Exercice 1.19 (Montée en puissance)

Soit (X, Y) un couple aléatoire de densité jointe :

$$f(x, y) = \frac{1}{y} e^{-\frac{x}{y} - y} \mathbb{1}_{]0, +\infty[^2}(x, y)$$

1. Déterminer la densité marginale $f(y)$ de Y .
2. En déduire la densité conditionnelle $f(x|y)$.
3. Que vaut $\mathbb{E}[X|Y = y]$. En déduire l'espérance conditionnelle de X sachant Y .
4. On considère cette fois : $f(x, y) = \frac{12}{5}x(2 - x - y) \mathbb{1}_{]0, 1[^2}(x, y)$. Montrer que

$$\mathbb{E}[X|Y] = \frac{5 - 4Y}{8 - 6Y}$$

Corrigé

1. La densité marginale de Y vaut :

$$f(y) = \int_0^{+\infty} \frac{1}{y} e^{-\frac{x}{y} - y} \mathbb{1}_{]0, +\infty[(y) dy,$$

ce qui donne après calculs :

$$f(y) = e^{-y} \mathbb{1}_{]0, +\infty[(y),$$

c'est-à-dire que $Y \sim \varepsilon(1)$.

2. On en déduit la densité conditionnelle $f(x|y)$. Pour tout $y > 0$:

$$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{1}{y} e^{-\frac{x}{y}} \mathbb{1}_{]0, +\infty[(x),$$

donc sachant $Y = y$, $X \sim \varepsilon(1/y)$.

3. On sait que si $X \sim \varepsilon(\lambda)$, alors $\mathbb{E}[X] = 1/\lambda$. Or sachant $Y = y$, $X \sim \varepsilon(1/y)$, donc :

$$\mathbb{E}[X|Y = y] = y,$$

et par suite : $\mathbb{E}[X|Y] = Y$.

4. On considère cette fois : $f(x, y) = \frac{12}{5}x(2 - x - y) \mathbb{1}_{]0, 1[^2}(x, y)$. Pour montrer que

$$\mathbb{E}[X|Y] = \frac{5 - 4Y}{8 - 6Y},$$

il suffit d'appliquer la méthode usuelle. Si vous n'arrivez pas à ce résultat, c'est que vous avez fait une faute de calcul quelque part !

Exercice 1.20 (Mai 2007)

Soit (X, Y) un couple aléatoire de densité jointe :

$$f(x, y) = cx(y - x)e^{-y} \mathbb{1}_{\{0 < x \leq y\}}.$$

1. Soit V une variable aléatoire qui suit une loi exponentielle de paramètre λ . Rappeler son moment d'ordre n , c'est-à-dire $\mathbb{E}[V^n]$.
2. Déterminer c pour que f soit effectivement une densité.
3. Calculer $f(x|y)$, densité conditionnelle de X sachant $Y = y$.
4. En déduire que $\mathbb{E}[X|Y] = Y/2$.

5. Calculer $f(y|x)$, densité conditionnelle de Y sachant $X = x$.
6. En déduire que $\mathbb{E}[Y|X] = X + 2$.
7. Déduire des questions 4 et 6 les quantités $\mathbb{E}[X]$ et $\mathbb{E}[Y]$.

Corrigé

Cf. annales en fin de polycopié.

Exercice 1.21 (Mai 2008)

On considère le couple aléatoire (X, Y) de densité la fonction f définie sur \mathbb{R}^2 par :

$$f(x, y) = \begin{cases} 2 & \text{si } 0 \leq x \leq y \leq 1 \\ 0 & \text{sinon} \end{cases}$$

1. Représenter f et vérifier qu'il s'agit bien d'une fonction de densité.
2. Les variables X et Y sont-elles indépendantes ?
3. Déterminer les lois marginales puis calculer $\mathbb{E}[X]$ et $\mathbb{E}[Y]$.
4. Calculer les lois conditionnelles $f(y|x)$ et $f(x|y)$.
5. Calculer $\mathbb{E}[X|Y]$.
6. En utilisant le résultat précédent et la valeur de $\mathbb{E}[Y]$ de la question 3, retrouver la valeur de $\mathbb{E}[X]$.

Corrigé

Cf. annales en fin de polycopié.

Exercice 1.22 (Changement de couple)

On considère la fonction suivante :

$$f(x, y) = e^{-y} \mathbb{1}_{\{0 < x < y\}}.$$

1. Vérifier que $f(x, y)$ définit une densité de probabilité sur \mathbb{R}^2 .
2. Calculer les densités marginales de X et Y . Les variables X et Y sont-elles indépendantes ?
3. Calculer $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\text{Cov}(X, Y)$.
4. Déterminer la densité conditionnelle $f_{Y|X}$ de Y sachant $X = x$. En déduire $\mathbb{E}[Y|X = x]$, puis l'espérance conditionnelle $\mathbb{E}[Y|X]$.
5. Déterminer de même $\mathbb{E}[X|Y]$.
6. Déterminer la loi jointe $f_{Z,T}$ du couple (Z, T) défini par :

$$\begin{cases} Z &= X + Y \\ T &= Y - X \end{cases}$$

7. En déduire les densités marginales de Z et T .

Corrigé

Cet exercice est corrigé en annexe, sujet de juin 2005.

Exercice 1.23 (Jeu de fléchettes)

Soit D le disque de centre $(0, 0)$ et de rayon 1, (X, Y) un point tiré uniformément dans D .

1. Donner la densité $f(x, y)$ du couple (X, Y) .
2. Déterminer les lois marginales de X et Y . En déduire $\mathbb{E}[X]$ et $\mathbb{E}[Y]$.
3. Les variables X et Y sont-elles indépendantes ?
4. Calculer la covariance du couple (X, Y) .
5. Déterminer la fonction de répartition $G(u)$, puis la densité $g(u)$, de la variable aléatoire $U = X^2 + Y^2$.
6. Calculer l'espérance de U . En déduire $\mathbb{E}[X^2]$, $\mathbb{E}[Y^2]$, la variance de X et celle de Y .
7. Déterminer la densité conditionnelle $f(\cdot|x)$ de Y sachant $X = x$. Calculer $\mathbb{E}[Y^2|X = x]$, puis $\mathbb{E}[X^2 + Y^2|X = x]$, puis $\mathbb{E}[X^2 + Y^2|X]$.
8. Un tireur tire sur la cible D : la loi du point d'impact (X, Y) sur la cible est uniforme. Au point d'impact est associée la distance au centre $L = \sqrt{X^2 + Y^2}$. Supposons qu'il tire n fois de façons indépendantes : ceci donne un n -uplet (L_1, \dots, L_n) de variables aléatoires. Soit $0 < a \leq 1$: calculer la probabilité que l'une au moins des fléchettes soit à distance inférieure à a du centre de la cible.

Corrigé

1. Puisque le point (X, Y) est tiré uniformément dans le disque D , la densité $f(x, y)$ du couple (X, Y) est tout simplement l'indicatrice du disque divisé par la surface de ce disque. C'est la généralisation d'une loi uniforme sur un segment de \mathbb{R} (indicatrice du segment divisé par sa longueur). Ainsi :

$$f(x, y) = \frac{1}{\pi} \mathbb{1}_D(x, y) = \frac{1}{\pi} \mathbb{1}_{\{x^2 + y^2 \leq 1\}}(x, y).$$

2. Si x est fixé entre -1 et 1 , y ne peut varier qu'entre $-\sqrt{1-x^2}$ et $+\sqrt{1-x^2}$ (faire un dessin !). On a alors :

$$f(x) = \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi} \sqrt{1-x^2} \mathbb{1}_{[-1,1]}(x).$$

Puisque l'abscisse X et l'ordonnée Y jouent des rôles symétriques, on a aussi :

$$f(y) = \frac{2}{\pi} \sqrt{1-y^2} \mathbb{1}_{[-1,1]}(y).$$

On en déduit :

$$\mathbb{E}[X] = \frac{2}{\pi} \int_{-1}^1 x \sqrt{1-x^2} dx,$$

et il y a deux façons de voir les choses : ou bien on y va brutalement et on reconnaît à peu de choses près la dérivée de $(1-x^2)^{3/2}$. Ou bien on ruse on voit que c'est l'intégrale d'une fonction impaire sur un domaine symétrique par rapport à 0, donc elle vaut 0 (faire un dessin). Ainsi $\mathbb{E}[X] = 0$, ce qui n'est pas étonnant : on lance les fléchettes aussi bien dans les abscisses négatives que positives. Puisque Y a même loi que X , on a aussi $\mathbb{E}[Y] = 0$.

3. Les variables X et Y ne sont pas indépendantes, puisque le support de la loi du couple (X, Y) n'est pas un pavé, mais un disque. On peut aussi le voir en vérifiant que la loi jointe $f(x, y)$ n'est pas égale au produit des marginales.
4. Par définition la covariance du couple (X, Y) est :

$$\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] = \iint_D xy f(x, y) dx dy,$$

et par le théorème de Fubini (on intègre une fonction continue sur un domaine borné donc no souci) :

$$\text{cov}(X, Y) = \frac{1}{\pi} \int_{-1}^{+1} x \left(\int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} y \, dy \right) dx = 0,$$

puisque pour tout x entre -1 et 1 , le segment $[-\sqrt{1-x^2}, +\sqrt{1-x^2}]$ est symétrique par rapport à 0 et la fonction $y \mapsto y$ est impaire. On en déduit que $\text{cov}(X, Y) = 0$ alors que X et Y ne sont pas indépendantes : Etonnov, niet ?

5. La variable aléatoire $(X^2 + Y^2)$ est à valeurs entre 0 et 1 et pour tout $u \in [0, 1]$, on a :

$$G(u) = \mathbb{P}(X^2 + Y^2 \leq u) = \mathbb{P}\left(\sqrt{X^2 + Y^2} \leq \sqrt{u}\right),$$

or $\sqrt{X^2 + Y^2}$ est la distance au centre d'un point M tiré au hasard dans le disque. La probabilité que celle-ci soit plus petite que \sqrt{u} correspond donc au rapport des surfaces entre le disque de centre O et de rayon \sqrt{u} et le disque D , c'est-à-dire que :

$$G(u) = \begin{cases} 0 & \text{si } u \leq 0 \\ u & \text{si } 0 \leq u \leq 1 \\ 1 & \text{si } u \geq 1 \end{cases}$$

Autrement dit U suit une loi uniforme sur $[0, 1]$, ce qu'on note $U \sim \mathcal{U}_{[0,1]}$. On en déduit que sa densité est l'indicatrice du segment $[0, 1]$: $g(u) = \mathbb{1}_{[0,1]}(u)$.

6. L'espérance de U vaut donc $1/2$. Puisque X et Y ont même loi, on a $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$, et puisque $U = (X^2 + Y^2)$, on a :

$$\mathbb{E}[U] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] = 2\mathbb{E}[X^2] \Rightarrow \mathbb{E}[X^2] = \frac{1}{2}\mathbb{E}[U] = \frac{1}{4}.$$

Les variances de X et Y sont identiques et :

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] = \frac{1}{4}.$$

7. La densité conditionnelle $f(\cdot|x)$ de Y sachant $X = x$ est :

$$f(y|x) = \frac{1}{2\sqrt{1-x^2}} \mathbb{1}_{[-\sqrt{1-x^2}, +\sqrt{1-x^2}]}(y).$$

On en déduit :

$$\mathbb{E}[Y^2|X = x] = \frac{1}{2\sqrt{1-x^2}} \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} y^2 \, dy = \frac{1-x^2}{3}.$$

Mais alors :

$$\mathbb{E}[X^2 + Y^2|X = x] = \mathbb{E}[X^2|X = x] + \mathbb{E}[Y^2|X = x] = x^2 + \frac{1-x^2}{3} = \frac{2x^2 + 1}{3},$$

et de façon générale :

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

8. La probabilité que l'une au moins des fléchettes soit à distance inférieure à a du centre de la cible est :

$$p_a = \mathbb{P}(\min(L_1, \dots, L_n) \leq a) = 1 - \mathbb{P}(\min(L_1, \dots, L_n) > a) = 1 - \mathbb{P}(L_1 > a, \dots, L_n > a),$$

or les variables aléatoires L_1, \dots, L_n sont indépendantes et identiquement distribuées, donc :

$$\mathbb{P}(L_1 > a, \dots, L_n > a) = \mathbb{P}(L_1 > a) \times \dots \times \mathbb{P}(L_n > a) = \mathbb{P}(L_1 > a)^n.$$

Mais on a alors :

$$\mathbb{P}(L_1 > a) = 1 - \mathbb{P}(L_1 \leq a) = 1 - \mathbb{P}(L_1^2 \leq a^2) = 1 - \mathbb{P}(X^2 + Y^2 \leq a^2) = 1 - G(a^2) = 1 - a^2.$$

Ainsi :

$$p_a = 1 - (1 - a^2)^n.$$

Exercice 1.24 (Lois exponentielles)

Soit X et Y deux variables aléatoires indépendantes suivant une loi exponentielle de même paramètre $\lambda > 0$.

1. Quelle est la loi jointe $f_{X,Y}$ du couple (X, Y) ?
2. Déterminer la loi jointe $f_{V,W}$ du couple (V, W) défini par :

$$\begin{cases} V &= X + Y \\ W &= X \end{cases}$$

3. En déduire la densité de V .
4. Calculer $f(w|v)$. Quelle loi reconnaît-on ?

Corrigé

Voir les annales, sujet de mai 2008.

Exercice 1.25 (Minimum de variables exponentielles)

1. On considère deux variables aléatoires indépendantes X_1 et X_2 exponentielles de paramètres respectifs λ_1 et λ_2 . Soit $Y = \min(X_1, X_2)$ le minimum de ces deux variables. Montrer que Y suit une loi exponentielle de paramètre $(\lambda_1 + \lambda_2)$ (on pourra utiliser les fonctions de répartition).
2. Montrer que :

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

Indication : on pourra calculer $\mathbb{P}(X_1 < X_2)$ en conditionnant par rapport à X_2 .

3. Deux guichets sont ouverts à une banque : le temps de service au premier (respectivement second) guichet suit une loi exponentielle de moyenne 20 (respectivement 30) minutes. Aude et Vincent sont convoqués à la banque pour s'expliquer sur leurs découverts respectifs : Aude choisit le guichet 1, Vincent le 2. Quelle est la probabilité que Aude sorte la première ?
4. En moyenne, combien de temps faut-il pour que les deux soient sortis ? Indication : le max de deux nombres, c'est la somme moins le min.

Corrigé

1. Notons F_Y la fonction de répartition de Y , alors :

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\min(X_1, X_2) \leq y) = 1 - \mathbb{P}(\min(X_1, X_2) > y),$$

ce qui s'écrit encore :

$$F_Y(y) = 1 - \mathbb{P}(\{X_1 > y\} \cap \{X_2 > y\}).$$

Or X_1 et X_2 sont indépendantes :

$$F_Y(y) = 1 - \mathbb{P}(X_1 > y)\mathbb{P}(X_2 > y) = 1 - e^{-\lambda_1 y} \mathbb{1}_{\mathbb{R}^+}(y) e^{-\lambda_2 y} \mathbb{1}_{\mathbb{R}^+}(y) = 1 - e^{-(\lambda_1 + \lambda_2)y} \mathbb{1}_{\mathbb{R}^+}(y),$$

c'est-à-dire que $Y \sim \mathcal{E}(\lambda_1 + \lambda_2)$.

2. On a :

$$\mathbb{P}(Y = X_1) = \mathbb{P}(X_1 < X_2) = \int_0^{+\infty} \mathbb{P}(X_1 < X_2 | X_2 = x) f_{X_2}(x) dx,$$

c'est-à-dire :

$$\mathbb{P}(Y = X_1) = \int_0^{+\infty} \mathbb{P}(X_1 < x | X_2 = x) \lambda_2 e^{-\lambda_2 x} dx.$$

Mais puisque X_1 et X_2 sont indépendantes, le premier terme dans l'intégrale est simplement :

$$\mathbb{P}(X_1 < x | X_2 = x) = \mathbb{P}(X_1 < x) = 1 - e^{-\lambda_1 x},$$

d'où l'on déduit :

$$\mathbb{P}(Y = X_1) = \int_0^{+\infty} (\lambda_2 e^{-\lambda_2 x} - \lambda_2 e^{-(\lambda_1 + \lambda_2)x}) dx = 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2}.$$

Finalement on a bien :

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

3. Rappelons qu'une exponentielle de moyenne 20 a pour paramètre $1/20$. La probabilité que Aude sorte la première est donc tout simplement :

$$p = \frac{1/20}{1/20 + 1/30} = \frac{3}{5}.$$

4. Soit X_a , respectivement X_v , le temps nécessaire pour que Aude, respectivement Vincent, sorte de la banque. On cherche donc à calculer $\mathbb{E}[\max(X_a, X_v)]$. Il suffit de remarquer que :

$$\max(X_a, X_v) = X_a + X_v - \min(X_a, X_v),$$

d'où par linéarité de l'espérance :

$$\mathbb{E}[\max(X_a, X_v)] = \mathbb{E}[X_a] + \mathbb{E}[X_v] - \mathbb{E}[\min(X_a, X_v)] = 20 + 30 - \frac{1}{1/20 + 1/30} = 38 \text{ min.}$$

Exercice 1.26 (Variable Y définie à partir de X)

Soit X une variable aléatoire de densité :

$$\frac{2}{(\ln 2)^2} \frac{\ln(1+x)}{1+x} \mathbb{1}_{[0,1]}(x)$$

Soit Y une variable aléatoire telle que la loi conditionnelle de Y sachant $X = x$ est :

$$\frac{1}{\ln(1+x)} \frac{1}{1+y} \mathbb{1}_{[0,x]}(y)$$

1. Donner la densité jointe du couple (X, Y) .
2. Les variables X et Y sont-elles indépendantes?
3. Quelle est la loi de Y ? Loi conditionnelle de X sachant Y ?
4. Déterminer l'espérance conditionnelle $\mathbb{E}[X|Y]$.

Corrigé

Cf. annales, sujet de juin 2006.

Exercice 1.27 (L'aiguille de Buffon)

On suppose qu'on lance une aiguille de longueur unité sur un parquet dont les lames sont elles-mêmes de largeur unité. On voudrait calculer la probabilité p que l'aiguille soit à cheval sur deux lames. On modélise le problème comme suit : la variable aléatoire X correspond à la distance du milieu de l'aiguille au bord de lame le plus proche, la variable aléatoire T correspond à l'angle entre l'aiguille et l'axe des abscisses (cf. figure 1.18, à gauche). On suppose que X est uniformément distribuée sur $[0, 1/2]$, T uniformément distribuée sur $[-\pi/2, \pi/2]$, et que ces deux variables sont indépendantes.

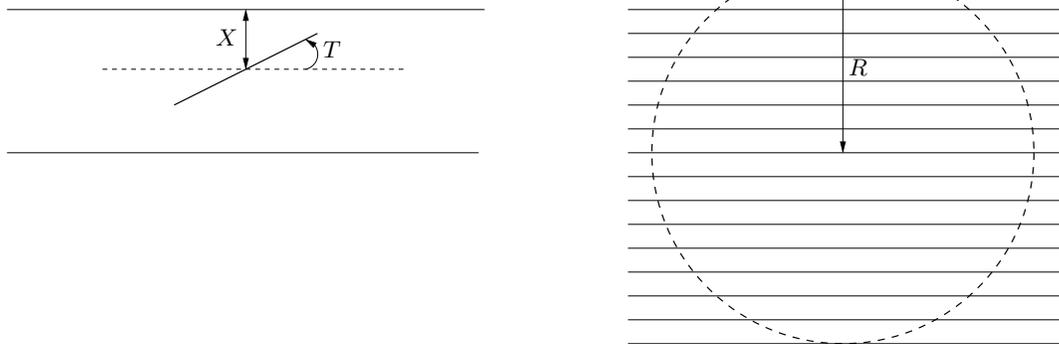


FIGURE 1.18 – Modélisation de l'expérience de Buffon (à gauche) et solution diabolique (à droite).

1. Expliquer pourquoi la probabilité cherchée peut s'écrire :

$$p = \mathbb{P} \left(X \leq \frac{1}{2} |\sin T| \right).$$

2. Soit $t \in [-\pi/2, \pi/2]$. Que vaut $\mathbb{P}(X \leq \frac{1}{2} |\sin t|)$?
3. Grâce à un calcul de probabilité par conditionnement, en déduire que $p = \frac{2}{\pi}$.
4. Méthode heuristique : on jette un très grand nombre d'aiguilles sur le parquet, de sorte qu'il y en ait dans toutes les directions. On peut donc les mettre bout à bout de façon à former un très grand cercle, de rayon R (cf. figure 1.18, à droite).
 - (a) Quel est approximativement le nombre N d'allumettes nécessaires pour former ce cercle?
 - (b) Quel est approximativement le nombre N_i de lames de parquet intersectées par ces allumettes?
 - (c) En faisant le rapport entre ces deux nombres, retrouver le résultat $p = \frac{2}{\pi}$.

1. Puisque l'aiguille est de longueur unité, la distance verticale entre ses deux extrémités est égale à $|\sin T|$. Dès lors, la distance verticale entre le milieu de l'aiguille et chacune de ses deux extrémités vaut $|\sin T|/2$. Pour que l'aiguille ne chevauche aucune lame, il faut et il suffit que ce nombre soit plus petit que la distance du milieu de l'aiguille au bord de lame le plus proche, notée X . Ainsi la probabilité cherchée s'écrit :

$$p = \mathbb{P}\left(X \leq \frac{1}{2}|\sin T|\right).$$

2. Soit $t \in [-\pi/2, \pi/2]$. Puisque X suit une loi uniforme sur $[0, 1/2]$, il vient

$$\mathbb{P}(X \leq \frac{1}{2}|\sin t|) = |\sin t|.$$

3. Grâce à un calcul de probabilité par conditionnement, en déduire que $p = \frac{2}{\pi}$.
4. Méthode heuristique : on jette un très grand nombre d'aiguilles sur le parquet, de sorte qu'il y en ait dans toutes les directions. On peut donc les mettre bout à bout de façon à former un très grand cercle, de rayon R (cf. figure 1.18, à droite).
 - (a) Quel est approximativement le nombre N d'allumettes nécessaires pour former ce cercle ?
 - (b) Quel est approximativement le nombre N_i de lames de parquet intersectées par ces allumettes ?
 - (c) En faisant le rapport entre ces deux nombres, retrouver le résultat $p = \frac{2}{\pi}$.

Exercice 1.28 (Casser un bâton en trois)

On casse un bâton en trois morceaux "au hasard" et on veut connaître la probabilité de pouvoir faire un triangle avec ces trois morceaux. On suppose pour simplifier les calculs que le bâton est de longueur unité.

1. Si on considère trois bâtons de longueurs a , b et c , à quelles conditions sur ces trois nombres peut-on construire un triangle avec ces trois bâtons ?
2. Première méthode : on tire uniformément entre 0 et 1 deux variables aléatoires indépendantes U et V . Représenter graphiquement, dans le carré $[0, 1] \times [0, 1]$ les couples admissibles. En déduire que la probabilité cherchée vaut $\frac{1}{4}$.
3. Seconde méthode : on casse d'abord le bâton en deux morceaux (tirage d'une variable aléatoire X uniforme sur $[0, 1]$), puis on choisit au hasard l'un des deux morceaux (pile ou face non biaisé), puis on recasse ce morceau en deux (tirage d'une variable uniforme Y). Déterminer la densité, notée $f(y|x)$, de Y sachant $X = x$. En déduire que la probabilité cherchée vaut $\ln 2 - \frac{1}{2} \approx 0.19$.
4. Pourquoi ne trouve-t-on pas le même résultat ?

Corrigé

1. La condition générale pour qu'on puisse construire un triangle est qu'aucune des deux longueurs ne soit supérieure à la somme des deux autres, ce qui donne un système de trois inéquations à satisfaire.
2. Les deux points U et V définissent les trois segments $[0, \min(U, V)]$, $[\min(U, V), \max(U, V)]$ et $[\max(U, V), 1]$, de longueurs respectives $\min(U, V)$, $\max(U, V) - \min(U, V)$ et $1 - \max(U, V)$. D'après la question précédente, le système suivant doit alors être satisfait :

$$\begin{cases} \min(U, V) \leq (\max(U, V) - \min(U, V)) + (1 - \max(U, V)) \\ \max(U, V) - \min(U, V) \leq \min(U, V) + (1 - \max(U, V)) \\ 1 - \max(U, V) \leq \min(U, V) + (\max(U, V) - \min(U, V)) \end{cases}$$

lequel est équivalent à :

$$\begin{cases} \min(U, V) \leq 1/2 \\ \max(U, V) - \min(U, V) \leq 1/2 \\ \max(U, V) \geq 1/2 \end{cases}$$

De par la symétrie des rôles joués par U et V , on peut supposer dans un premier temps $U \leq V$, c'est-à-dire se restreindre au triangle supérieur gauche du carré unité, étudier la région admissible dans ce triangle, et la seconde région s'en déduira par symétrie par rapport à la droite $u = v$. Or si $U \leq V$, le système ci-dessus devient

$$\begin{cases} U \leq 1/2 \\ V \leq U + 1/2 \\ V \geq 1/2 \end{cases}$$

La région admissible est représentée figure 1.19. En cassant un bâton en trois morceaux de

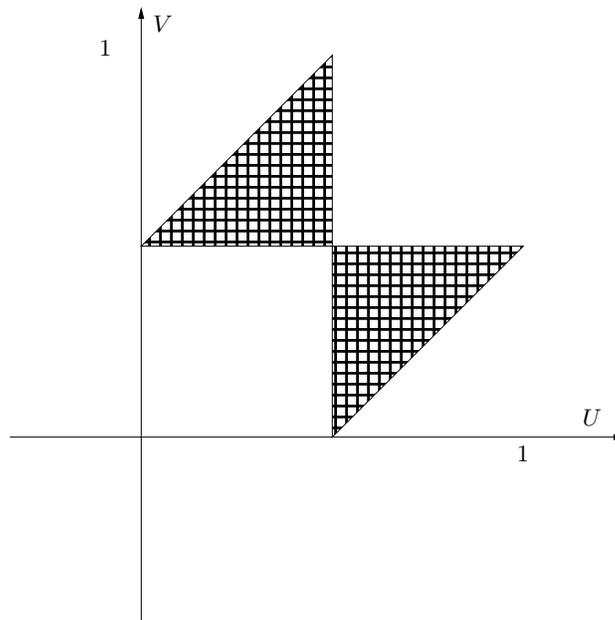


FIGURE 1.19 – Région admissible.

cette façon, la probabilité qu'on puisse faire un triangle correspond exactement à la surface de cette région, laquelle vaut $1/4$.

3. Seconde méthode : on casse d'abord le bâton en deux morceaux (tirage d'une variable aléatoire X uniforme sur $[0, 1]$), puis on choisit au hasard l'un des deux morceaux (pile ou face non biaisé), puis on recasse ce morceau en deux (tirage d'une variable uniforme Y). Déterminer la densité, notée $f(y|x)$, de Y sachant $X = x$. En déduire que la probabilité cherchée vaut $\ln 2 - \frac{1}{2} \approx 0.19$.
4. Pourquoi ne trouve-t-on pas le même résultat ?

Exercice 1.29 (Triplet aléatoire)

Soit (X, Y, Z) un triplet aléatoire. La loi marginale de X est donnée par :

$$f(x) = \frac{1}{6}x^3 e^{-x} \mathbb{1}_{]0, +\infty[}(x).$$

La loi conditionnelle de Y sachant $X = x$ est donnée par :

$$f(y|x) = 3\frac{y^2}{x^3} \mathbb{1}_{]0,x[}(y).$$

La loi conditionnelle de Z sachant $X = x$ et $Y = y$ est donnée par :

$$f(z|x, y) = 2\frac{y-z}{y^2} \mathbb{1}_{]0,y[}(z).$$

1. Soit $V \sim \mathcal{E}(\lambda)$. Donner pour tout $n \in \mathbb{N}$: $\mathbb{E}[V^n]$.
2. Représenter l'ensemble des valeurs prises par le triplet (X, Y, Z) .
3. Déterminer la densité jointe $f(x, y, z)$. En déduire la densité de Z .
4. Que vaut la densité jointe du couple (X, Y) conditionnellement à $Z = z$?
5. En déduire la densité de X sachant $Z = z$, notée $f(x|z)$, puis $\mathbb{E}[X|Z]$.
6. Soit $S = X + Y$ et $T = X - Y$. Déterminer la densité jointe $f_{S,T}$ du couple (S, T) . Quelle loi suit la variable aléatoire T ?

Exercice 1.30 (Couple mixte)

On rappelle que si $V \sim \mathcal{E}(\alpha)$, on a : $\mathbb{E}[V^n] = n!/\alpha^n$. Soit alors (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{N} \times [0, +\infty[$, tel que :

- la loi marginale de Y est exponentielle de paramètre 1 ;
- la loi conditionnelle de X sachant $Y = \lambda$ est une loi de Poisson de paramètre λ .

Puisque le couple n'est ni discret, ni absolument continu, on propose de noter $p(n, \lambda)$ sa loi jointe, $p(n) = \mathbb{P}(X = n)$ la loi marginale de X , $p(\lambda)$ la densité de Y , etc.

1. Déterminer la loi jointe du couple (X, Y) , c'est-à-dire $p(n, \lambda)$.
2. Déterminer la loi marginale de X , c'est-à-dire $p(n)$. Quel est le lien avec une loi géométrique classique ? Que vaut $\mathbb{E}[X]$?
3. Déterminer la densité conditionnelle de Y sachant $X = n$, c'est-à-dire $p(\lambda|n)$.
4. Déterminer l'espérance conditionnelle de Y sachant $X = n$, c'est-à-dire $\mathbb{E}[Y|X = n]$. En déduire $\mathbb{E}[Y|X]$.
5. Vérifier sur cet exemple la relation vue en cours dans les cas classiques :

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

Exercice 1.31 (Triplet exponentiel)

Soit X_1, X_2 et X_3 des variables i.i.d. de loi exponentielle de paramètre α . On pose :

$$\begin{cases} Y_1 &= X_2 - X_1 \\ Y_2 &= X_3 - X_1 \end{cases}$$

On s'intéresse à la densité jointe du couple (Y_1, Y_2) sachant $X_1 = x_1$. Montrer que pour tout triplet (x_1, y_1, y_2) de \mathbb{R}^3 , on a :

$$f(y_1, y_2|x_1) = \alpha^2 e^{-\alpha(2x_1+y_1+y_2)} \mathbb{1}_{\{y_1 \geq -x_1, y_2 \geq -x_1\}}.$$

Exercice 1.32 (Droite de régression en statistiques)

On considère les tailles et poids de dix enfants de six ans :

Taille	121	123	108	118	111	109	114	103	110	115
Poids	25	22	19	24	19	18	20	15	20	21

1. Calculer les espérances, variances et covariance empiriques pour cet échantillon.
2. Déterminer la droite de régression $y = \hat{a}x + \hat{b}$.
3. Quelle est l'erreur quadratique moyenne pour cet échantillon ?

Exercice 1.33 (Droite de régression en probabilités)

Soit (X, Y) un couple aléatoire de densité jointe :

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{2x^2 - 2xy + y^2}{2}}$$

1. Montrer que $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 2)$, lois normales centrées de variances respectives 1 et 2.
2. Montrer que la covariance du couple (X, Y) vaut 1.
3. En déduire l'équation de la droite de régression de Y en X : $y = ax + b$.
4. Montrer que, sachant $X = x$, Y suit une loi normale $\mathcal{N}(x, 1)$. En déduire la courbe de régression : $x \mapsto \mathbb{E}[Y|X = x]$.
5. Sachant $X = x$, on veut la probabilité que Y s'éloigne de $ax + b$ de plus de une unité, i.e. calculer :

$$\mathbb{P}(|Y - (aX + b)| > 1 | X = x).$$

Indication : si $V \sim \mathcal{N}(0, 1)$, alors $\mathbb{P}(|V| > 1) \approx 0.32$.

Exercice 1.34 (Droite de régression et courbe de régression)

Soit (X, Y) un couple aléatoire de densité jointe :

$$f(x, y) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y^2}{x^2} - 2y + x^2 + 2x\right)} \mathbb{1}_{\{x > 0\}}$$

1. Montrer que $X \sim \mathcal{E}(1)$, loi exponentielle de paramètre 1.
2. Calculer $f(y|x)$ pour montrer que, sachant $X = x$, Y suit une loi normale $\mathcal{N}(x^2, x^2)$.
3. En déduire la courbe de régression : $x \mapsto \mathbb{E}[Y|X = x]$.
4. Sachant $X = x$, donner une zone de confiance à 95% pour Y .
5. Déterminer l'équation de la droite de régression de Y en X .
6. Représenter graphiquement les résultats.

Exercice 1.35 (Droite de régression et points aberrants)

Douze personnes sont inscrites à une formation. Au début de la formation, ces stagiaires subissent une épreuve A notée sur 20. A la fin de la formation, elles subissent une épreuve B de niveau identique. Les résultats sont donnés dans le tableau suivant :

Epreuve A	3	4	6	7	9	10	9	11	12	13	15	4
Epreuve B	8	9	10	13	15	14	13	16	13	19	6	19

1. Représenter le nuage de points. Déterminer la droite de régression. Calculer le coefficient de corrélation. Commenter.
2. Deux stagiaires semblent se distinguer des autres. Les supprimer³ et déterminer la droite de régression sur les dix points restants. Calculer le coefficient de corrélation. Commenter.

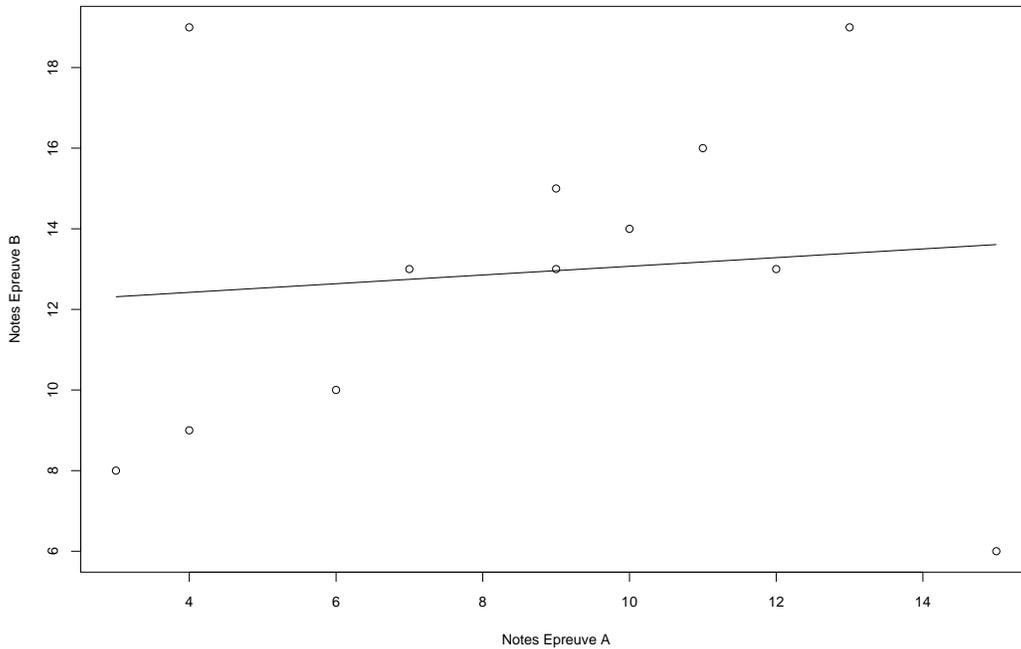


FIGURE 1.20 – Représentation des notes et droite de régression pour l'ensemble des 12 stagiaires.

Corrigé

1. Le nuage de points ainsi que la droite de régression sont représentés figure 1.20. On cherche à expliquer les notes à l'épreuve B, notées y_1, \dots, y_{12} à partir des notes à l'épreuve A, notées x_1, \dots, x_{12} . L'équation de la droite de régression est $y = \hat{a}x + \hat{b}$, avec :

$$\hat{a} = \frac{\sum_{i=1}^{12} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{12} (x_i - \bar{x})^2} \approx 0.11$$

rapport de la covariance empirique entre les notes à l'épreuve A et celles à l'épreuve B et de la variance empirique des notes à l'épreuve A. Pour l'ordonnée à l'origine, on a :

$$\hat{b} = \bar{y} - \hat{a}\bar{x} \approx 12.0$$

Le coefficient de corrélation linéaire vaut :

$$\hat{\rho} = \frac{\sum_{i=1}^{12} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{12} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{12} (y_i - \bar{y})^2}} \approx 0.10$$

Le coefficient proche de 0 pourrait laisser penser qu'il n'y a pas une forte corrélation linéaire entre les notes à l'épreuve A et les notes à l'épreuve B. De fait, sur la figure 1.20, la droite de régression ne semble pas représenter correctement le nuage de points.

2. On élimine les notes des deux derniers stagiaires, c'est-à-dire les deux dernières colonnes du tableau. Le nuage de points ainsi que la droite de régression sont représentés figure 1.21. L'équation de la droite de régression est encore $y = \hat{a}x + \hat{b}$, avec :

$$\hat{a} = \frac{\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{10} (x_i - \bar{x})^2} \approx 0.90$$

3. "Je ne crois aux statistiques que lorsque je les ai moi-même falsifiées." Winston Churchill.

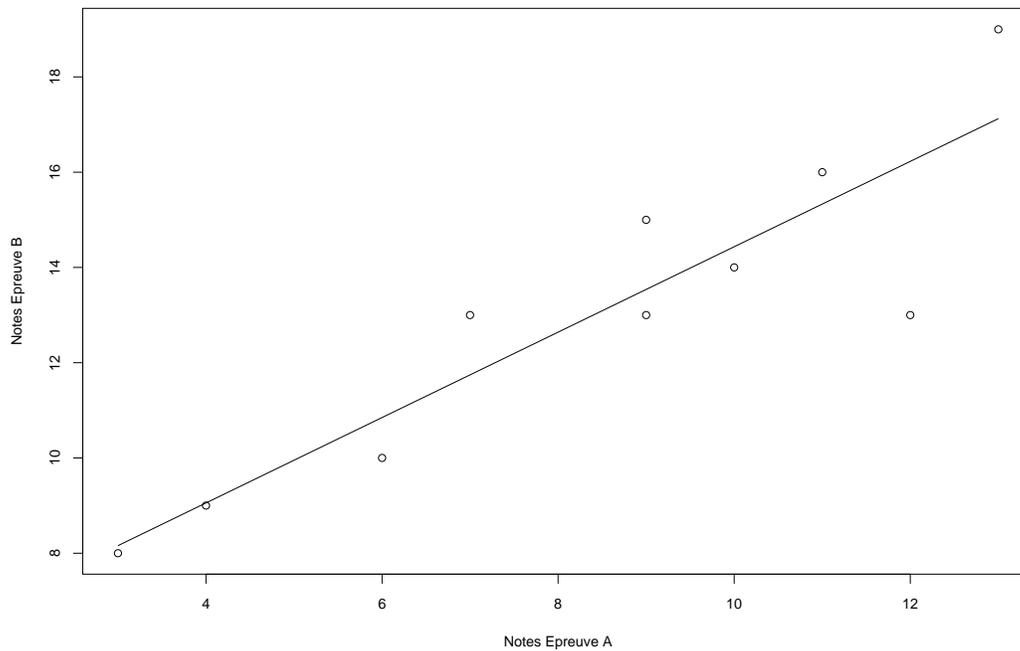


FIGURE 1.21 – Représentation des notes et droite de régression pour les 10 premiers stagiaires.

rapport de la covariance empirique entre les notes à l'épreuve A et celles à l'épreuve B et de la variance empirique des notes à l'épreuve A. Pour l'ordonnée à l'origine, on a :

$$\hat{b} = \bar{y} - \hat{a}\bar{x} \approx 5.5$$

Le coefficient de corrélation linéaire vaut :

$$\hat{\rho} = \frac{\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{10} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{10} (y_i - \bar{y})^2}} \approx 0.90$$

On obtient cette fois une forte corrélation linéaire puisque $\hat{\rho}$ est proche de 1. De même, sur la figure 1.21, la droite de régression est tout à fait représentative du nuage de points. Ainsi les notes des 2 derniers individus suffisaient à masquer la forte corrélation linéaire et à fausser complètement la régression linéaire pour expliquer la seconde note à partir de la première : ce sont ce qu'on appelle des individus aberrants. On trouvera la définition précise de cette notion dans le livre de Pierre-André Cornillon et Eric Matzner-Løber [8], paragraphe 4.1.2.

Exercice 1.36 (Un peu de prédiction)

Soit $\theta \in \mathbb{R}^*$ fixé. Soit $(Z_n)_{n \geq 0}$ une suite de variables aléatoires indépendantes centrées et de même variance σ^2 . On construit à partir de $(Z_n)_{n \geq 0}$ la suite de variables aléatoires $(X_n)_{n \geq 0}$ comme suit

$$\begin{cases} X_0 &= Z_0 \\ X_{n+1} &= Z_{n+1} - \theta Z_n \end{cases}$$

1. Pourquoi les X_n sont-elles de carrés intégrables ?
2. Montrer que la projection de X_{n+1} sur le sous-espace de L^2 engendré par les $(X_i)_{0 \leq i \leq n}$, noté dans le cours $\mathbb{E}[X_{n+1}|X_0, \dots, X_n]$, est :

$$\hat{X}_{n+1} = - \sum_{j=1}^{n+1} \theta^j X_{n+1-j}$$

On l'appelle encore le prédicteur des moindres carrés de X_{n+1} .

3. Calculer l'erreur quadratique moyenne, encore appelée erreur de prédiction, c'est-à-dire :

$$\mathbb{E}[(X_{n+1} - \hat{X}_{n+1})^2].$$

Exercice 1.37 (Un problème déterministe)

Le but de l'exercice est de déterminer deux nombres réels a et b qui minimisent l'intégrale :

$$\int_0^1 (e^x - ax - b)^2 dx.$$

On utilise pour ce faire une interprétation stochastique du problème. Considérons l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}_{[0,1]}, \lambda_{[0,1]})$. Dans ce contexte, une variable aléatoire est tout simplement une fonction borélienne $f : [0, 1] \rightarrow \mathbb{R}$. Si elle est intégrable sur $[0, 1]$, son espérance est :

$$\mathbb{E}[f] = \int_0^1 f(x) dx.$$

$H = L^2([0, 1], \mathcal{B}_{[0,1]}, \lambda_{[0,1]})$ est donc l'espace des fonctions boréliennes de carrés intégrables sur l'intervalle $[0, 1]$. Montrer que le problème de minimisation ci-dessus revient alors à déterminer une droite de régression. En déduire a et b .

Corrigé

Si on adopte les notations vues en régression dans le cours, la fonction identité $x \mapsto x$ correspond à la variable aléatoire X , tandis que la fonction $x \mapsto e^x$ correspond à la variable aléatoire Y .

De façon générale, faire une régression linéaire de la variable aléatoire Y sur la variable aléatoire X , c'est chercher les deux réels a et b tels que l'erreur quadratique moyenne faite en approchant Y par $aX + b$ soit minimale. On veut donc trouver :

$$\arg \min_{a,b} \mathbb{E} [(Y - (aX + b))^2],$$

ce qui, transposé dans notre contexte, s'écrit encore :

$$\arg \min_{a,b} \int_0^1 (e^x - ax - b)^2 dx.$$

Il nous suffit donc d'appliquer les formules habituelles donnant pente et ordonnée à l'origine en fonction des espérances, variances et covariance :

$$\begin{cases} a = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \\ b = \mathbb{E}[Y] - a\mathbb{E}[X] \end{cases}$$

Il reste à évaluer les quantités en jeu. L'espérance de X correspond à la valeur moyenne prise par la fonction identité sur $[0, 1]$:

$$\mathbb{E}[X] = \int_0^1 x dx = \frac{1}{2}.$$

De même pour l'espérance de Y :

$$\mathbb{E}[Y] = \int_0^1 e^x dx = e - 1.$$

La variance de X est comme d'habitude :

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

avec :

$$\mathbb{E}[X^2] = \int_0^1 x^2 dx = \frac{1}{3},$$

d'où finalement : $\text{Var}(X) = \frac{1}{12}$. De même, on a :

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Or les variables aléatoires X et Y sont liées par la relation $Y = e^X$, donc :

$$\mathbb{E}[XY] = \mathbb{E}[Xe^X] = \int_0^1 xe^x dx,$$

ce qui donne après une intégration par parties :

$$\mathbb{E}[XY] = 1 \Rightarrow \text{Cov}(X, Y) = \frac{1}{2}(3 - e).$$

On en déduit que les coefficients de la régression linéaire sont :

$$\begin{cases} a = 6(3 - e) \approx 1.69. \\ b = 2(2e - 5) \approx 0.87. \end{cases}$$

Autrement dit, sur l'intervalle $[0, 1]$, la meilleure approximation au sens de la norme L^2 de la fonction $x \mapsto e^x$ par une fonction affine est donnée par la droite (cf. figure 1.22) :

$$y = 6(3 - e)x + 2(2e - 5) \approx 1.69x + 0.87.$$

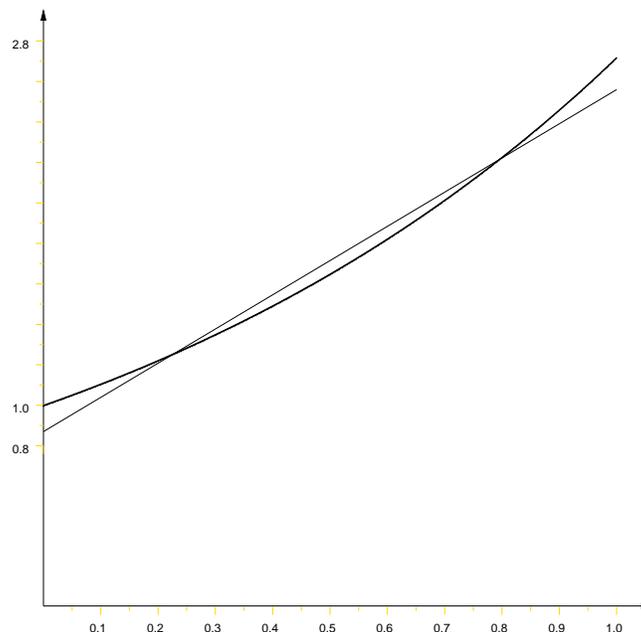


FIGURE 1.22 – Approximation en norme L^2 de $x \mapsto e^x$ par une fonction affine sur $[0, 1]$.

Remarque. L'approche brutale consisterait à voir la quantité à optimiser comme une fonction des deux variables a et b :

$$\Phi(a, b) = \int_0^1 (e^x - ax - b)^2 dx,$$

ce qui donne après développement et calculs :

$$\Phi(a, b) = \frac{a^2}{3} + b^2 + ab - 2a + 2(1 - e)b + \frac{1}{2}(e^2 - 1).$$

On effectue une factorisation “à la Gauss” :

$$\Phi(a, b) = \left(b - \left(e - 1 - \frac{a}{2} \right) \right)^2 + \frac{1}{12}(a - 6(3 - e))^2 - \left(\frac{7}{2}e^2 - 20e + \frac{57}{2} \right).$$

Cette quantité est minimale lorsqu'on annule les 2 carrés, c'est-à-dire lorsque :

$$\begin{cases} a = 6(3 - e) \\ b = e - 1 - \frac{a}{2} = 2(2e - 5) \end{cases}$$

On retrouve le résultat précédent. Happy end !

Chapitre 2

Vecteurs gaussiens et conditionnement

Introduction

Le calcul conditionnel s'exprime très simplement dans le cadre gaussien, puisque tout se ramène à du calcul matriciel. C'est pourquoi on dit qu'on est dans un cadre linéaire. C'est ce qui devrait ressortir de ce chapitre. Au préalable, il convient de faire quelques rappels sur les lois normales uni- et multi-dimensionnelles.

2.1 Rappels sur les vecteurs gaussiens

2.1.1 Variables gaussiennes

Dans toute la suite, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé.

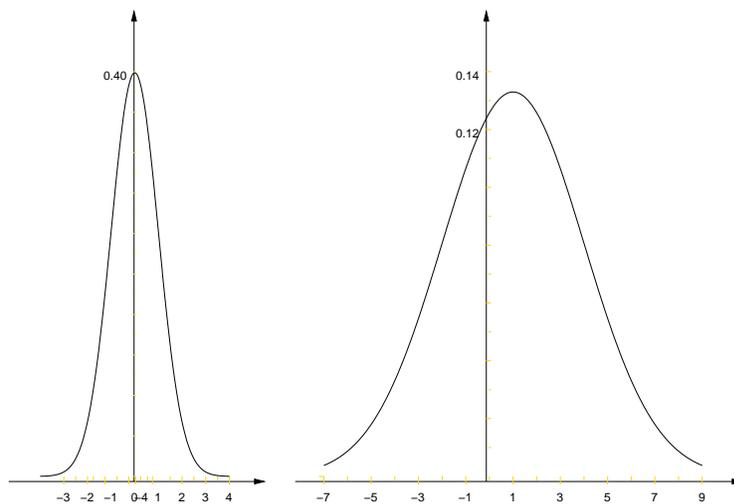


FIGURE 2.1 – Densités des lois normales $\mathcal{N}(0, 1)$ (à gauche) et $\mathcal{N}(2, 9)$ (à droite).

Définition 13 (Variable gaussienne)

On dit que la variable aléatoire $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ est gaussienne de moyenne m et de variance

$\sigma^2 > 0$ et on note $X \sim \mathcal{N}(m, \sigma^2)$, si X admet pour densité :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Des exemples de courbes en cloches sont donnés figure 2.1.

Remarques.

- Supposons qu'on tire des nombres selon une loi normale $\mathcal{N}(m, \sigma^2)$, par exemple avec un ordinateur. Alors plus l'écart-type σ est faible et plus on a des chances d'obtenir des résultats autour de la moyenne m : 68% de tomber à distance inférieure ou égale à σ , 95% de tomber à distance inférieure ou égale à 2σ , 99,7% de tomber à distance inférieure ou égale à 3σ . Ceci est illustré figure 2.2.
- La loi d'une variable gaussienne est complètement définie par la seule donnée de sa moyenne m et de sa variance σ^2 .
- Si la variance σ^2 est nulle, dire que $X \sim \mathcal{N}(m, 0)$ signifie que la variable aléatoire X est (quasi-) déterministe : elle ne prend presque sûrement que la valeur m .
- Il y a un lien très simple entre la loi normale centrée réduite et toute autre loi normale, puisque : si $X \sim \mathcal{N}(0, 1)$, alors $Y = \sigma X + m \sim \mathcal{N}(m, \sigma^2)$. On verra que cette propriété admet une généralisation vectorielle.

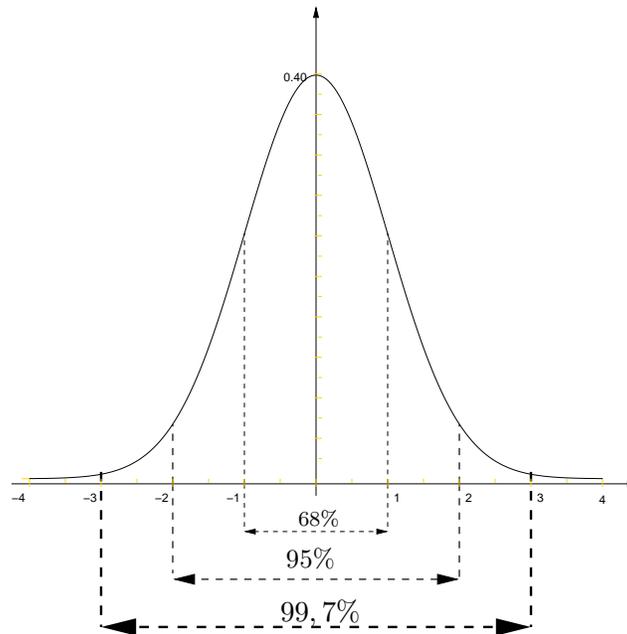


FIGURE 2.2 – Intervalles de confiance à 68%, 95% et 99,7% pour une $\mathcal{N}(0, 1)$.

On rappelle que la fonction caractéristique d'une variable aléatoire X est la fonction

$$\Phi_X : \begin{cases} \mathbb{R} & \rightarrow \mathbb{C} \\ t & \mapsto \Phi_X(t) = \mathbb{E}[e^{itX}] \end{cases}$$

Pour aller vite, la fonction caractéristique joue pour les variables aléatoires à densité le même rôle que la fonction génératrice des moments pour les variables discrètes, c'est-à-dire qu'il y a un lien entre les moments d'une variable aléatoire et les dérivées successives de sa fonction caractéristique.

Si X admet des moments de tout ordre, alors Φ_X est C^∞ sur \mathbb{R} et :

$$\Phi_X^{(n)}(0) = i^n \mathbb{E}[X^n].$$

Rappelons aussi que la loi d'une variable aléatoire est complètement caractérisée par sa fonction caractéristique (d'où son nom, la vie est bien faite...).

Proposition 7 (Fonction caractéristique d'une variable gaussienne)

Si $X \sim \mathcal{N}(m, \sigma^2)$, sa fonction caractéristique est donnée pour tout réel t par :

$$\Phi_X(t) = e^{imt - \frac{\sigma^2 t^2}{2}}.$$

Preuve. Soit $X \sim \mathcal{N}(0, 1)$, alors sa fonction caractéristique est définie par :

$$\Phi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

quantité complexe qu'on peut décomposer en parties réelle et imaginaire :

$$\Phi_X(t) = \int_{\mathbb{R}} \cos(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + i \int_{\mathbb{R}} \sin(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

qu'on écrit plus simplement :

$$\Phi_X(t) = \frac{1}{\sqrt{2\pi}} (F(t) + iG(t)).$$

Ainsi définie, la fonction

$$F : \begin{cases} \mathbb{R} & \rightarrow \mathbb{R} \\ t & \mapsto \int_{\mathbb{R}} \cos(tx) e^{-\frac{x^2}{2}} dx \end{cases}$$

est une intégrale dépendant d'un paramètre. On peut donc lui appliquer la théorie de Lebesgue, en commençant par s'assurer qu'elle est bien définie pour tout réel t puisque :

$$\left| \int_{\mathbb{R}} \cos(tx) e^{-\frac{x^2}{2}} dx \right| \leq \int_{\mathbb{R}} \left| \cos(tx) e^{-\frac{x^2}{2}} \right| dx \leq \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

On vérifie de même qu'elle est dérivable sur \mathbb{R} , sa dérivée s'obtenant tout simplement en dérivant par rapport à t sous le signe d'intégration :

$$\forall t \in \mathbb{R} \quad F'(t) = - \int_{\mathbb{R}} \sin(tx) x e^{-\frac{x^2}{2}} dx.$$

On effectue une intégration par parties :

$$F'(t) = \left[\sin(tx) e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} - t \int_{\mathbb{R}} \cos(tx) e^{-\frac{x^2}{2}} dx,$$

c'est-à-dire :

$$\forall t \in \mathbb{R} \quad F'(t) = -tF(t),$$

équation différentielle linéaire du premier ordre, qui s'intègre sans problème :

$$F(t) = \alpha e^{-\frac{t^2}{2}}.$$

Et puisqu'on a la condition initiale :

$$F(0) = \int_{\mathbb{R}} \cos(0x) e^{-\frac{x^2}{2}} dx = \sqrt{2\pi},$$

on en déduit que :

$$\forall t \in \mathbb{R} \quad F(t) = \sqrt{2\pi} e^{-\frac{t^2}{2}}.$$

Par ailleurs, la fonction G est identiquement nulle, puisque :

$$G(t) = \int_{\mathbb{R}} \sin(tx) e^{-\frac{x^2}{2}} dx$$

est l'intégrale d'une fonction impaire sur un domaine symétrique par rapport à 0, donc vaut 0. Ainsi, lorsque $X \sim \mathcal{N}(0, 1)$, sa fonction caractéristique est :

$$\forall t \in \mathbb{R} \quad \Phi_X(t) = e^{-\frac{t^2}{2}}.$$

Si maintenant on considère $Y = \sigma X + m$, alors $Y \sim \mathcal{N}(m, \sigma^2)$ et sa fonction caractéristique est :

$$\Phi_Y(t) = \mathbb{E}[e^{it(\sigma X + m)}] = e^{imt} \mathbb{E}[e^{i(t\sigma)X}] = e^{imt} \Phi_X(\sigma t),$$

et on peut se servir de ce qu'on vient de voir pour en déduire :

$$\forall t \in \mathbb{R} \quad \Phi_Y(t) = e^{imt - \frac{\sigma^2 t^2}{2}}. \quad \blacksquare$$

Exercice. A partir du développement en série entière de l'exponentielle ($e^u = \sum_{n=0}^{+\infty} \frac{u^n}{n!}$), montrer que si $X \sim \mathcal{N}(0, 1)$, alors ses moments sont donnés par :

$$\begin{cases} \mathbb{E}[X^{2n+1}] &= 0 \\ \mathbb{E}[X^{2n}] &= \frac{(2n)!}{2^n n!} \end{cases}$$

Via le théorème de Paul Lévy, les fonctions caractéristiques sont un outil efficace pour montrer la convergence en loi d'une suite de variables aléatoires : il suffit de prouver la convergence simple de la suite des fonctions caractéristiques.

C'est d'ailleurs ainsi qu'on montre le résultat qui fait toute l'importance de la loi normale, à savoir le théorème central limite. En voici la version la plus simple : si $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes et identiquement distribuées (en abrégé i.i.d.) de carré intégrable, alors en notant $S_n = X_1 + \dots + X_n$, on a la convergence en loi vers la loi normale centrée réduite :

$$\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n \operatorname{Var} X_1}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

c'est-à-dire que pour tout intervalle (a, b) de $\overline{\mathbb{R}}$, on a :

$$\mathbb{P} \left(a \leq \frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n \operatorname{Var} X_1}} \leq b \right) \xrightarrow[n \rightarrow +\infty]{} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Autrement dit, la somme d'un grand nombre de variables aléatoires i.i.d. se comporte comme une loi normale. L'aspect remarquable de ce résultat tient bien sûr au fait que la loi commune des X_n peut être n'importe quoi ! Celle-ci peut aussi bien être discrète qu'absolument continue, mixte ou singulière. La seule chose requise est l'existence du moment d'ordre 2.

2.1.2 Vecteurs gaussiens

La définition d'un vecteur gaussien est a priori un peu tordue.

Définition 14 (Vecteur gaussien)

On dit que le vecteur aléatoire¹ $X = [X_1, \dots, X_d]'$ est un vecteur gaussien si pour tout d -uplet $(\alpha_1, \dots, \alpha_d)$ de réels, la variable aléatoire $\alpha_1 X_1 + \dots + \alpha_d X_d$ est gaussienne.

En particulier, une variable aléatoire gaussienne est un vecteur gaussien de dimension 1. Par ailleurs, il découle de la définition le résultat suivant.

Proposition 8 (Vecteur gaussien \Rightarrow Composantes gaussiennes)

Si le vecteur aléatoire $X = [X_1, \dots, X_d]'$ est un vecteur gaussien, alors chaque variable aléatoire X_i est gaussienne.

Preuve. Si $X = [X_1, \dots, X_d]'$ est gaussien, alors en prenant $\alpha_1 = 1$ et $\alpha_i = 0$ pour tout $i \geq 2$, on en déduit que :

$$X_1 = \sum_{i=1}^d \alpha_i X_i$$

est gaussienne. Idem pour X_2, \dots, X_d .

■

La réciproque n'est pas vraie, comme le montre la situation suivante.

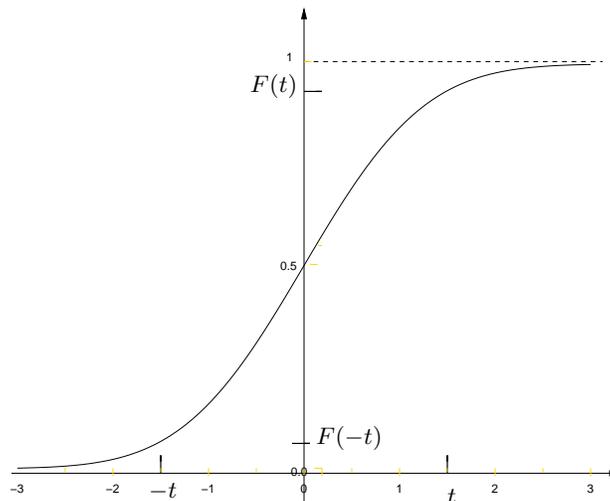


FIGURE 2.3 – Fonction de répartition F d'une loi normale $\mathcal{N}(0, 1)$ et relation : $F(-t) = 1 - F(t)$.

Remarque : Composantes gaussiennes \nRightarrow Vecteur gaussien.

Soit $X \sim \mathcal{N}(0, 1)$ et ε une variable aléatoire indépendante de X et suivant une loi de Rademacher : elle prend les valeurs $+1$ et -1 de façon équiprobable. Considérons la nouvelle variable $Y = \varepsilon X$

1. Dans tout le polycopié, le symbole ' correspond à la transposition.

et le vecteur aléatoire $V = [X, Y]'$. La variable aléatoire Y est gaussienne, comme le montre sa fonction de répartition :

$$F_Y(u) = \mathbb{P}(Y \leq u) = \mathbb{P}(\varepsilon X \leq u) = \mathbb{P}(-X \leq u | \varepsilon = -1) \mathbb{P}(\varepsilon = -1) + \mathbb{P}(X \leq u | \varepsilon = 1) \mathbb{P}(\varepsilon = 1),$$

expression qu'on peut simplifier grâce à l'indépendance de X et ε :

$$F_Y(u) = \frac{1}{2}(\mathbb{P}(-X \leq u) + \mathbb{P}(X \leq u)) = \frac{1}{2}(\mathbb{P}(X \geq -u) + \mathbb{P}(X \leq u)),$$

et en notant F_X la fonction de répartition d'une loi normale centrée réduite, c'est-à-dire :

$$F_X(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

on a donc :

$$F_Y(u) = \frac{1}{2}(1 - F_X(-u) + F_X(u)) = F_X(u),$$

la dernière égalité venant de la symétrie d'une loi normale centrée réduite par rapport à l'origine (voir figure 2.3) :

$$\forall u \in \mathbb{R} \quad F_X(-u) = 1 - F_X(u).$$

Ainsi Y suit une loi normale $\mathcal{N}(0, 1)$, tout comme X . Mais le vecteur $V = [X, Y]'$ n'est pas gaussien, puisque si on considère la variable aléatoire $Z = X + Y = (1 + \varepsilon)X$, on a :

$$\mathbb{P}(Z = 0) = \mathbb{P}(1 + \varepsilon = 0) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2},$$

ce qui est impossible pour une variable gaussienne ! En effet, cette probabilité vaut 0 pour toute loi gaussienne $\mathcal{N}(m, \sigma^2)$, sauf si $m = \sigma^2 = 0$, auquel cas elle vaut 1. A titre indicatif la fonction de répartition de Z est donnée figure 2.4 : c'est un exemple de loi mixte.

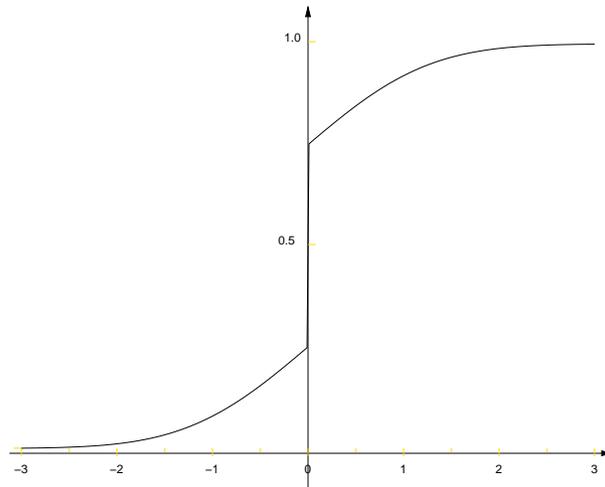


FIGURE 2.4 – Fonction de répartition de la variable aléatoire Z .

Il y a cependant une situation où les choses se passent bien.

Proposition 9 (Composantes gaussiennes indépendantes \Rightarrow Vecteur gaussien)

Soit (X_1, \dots, X_d) une suite indépendante de variables aléatoires. Le vecteur $X = [X_1, \dots, X_d]'$ est gaussien si et seulement si pour tout $i \in \{1, \dots, d\}$, la variable aléatoire X_i est gaussienne.

Preuve. Si les variables aléatoires gaussiennes $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$ sont gaussiennes et indépendantes, alors la variable aléatoire :

$$Y = \alpha_1 X_1 + \dots + \alpha_d X_d \sim \mathcal{N}\left(\sum_{i=1}^d \alpha_i m_i, \sum_{i=1}^d \alpha_i^2 \sigma_i^2\right).$$

Ceci se vérifie par exemple sans problème sur la fonction caractéristique de Y . Ainsi toute combinaison linéaire des composantes X_i est une variable gaussienne et par suite $X = [X_1, \dots, X_d]'$ est un vecteur gaussien.

Comme on l'a vu en proposition 8, la réciproque est toujours vraie, que les composantes soient indépendantes ou non. ■

Prenons un vecteur aléatoire $X = [X_1, \dots, X_d]'$, non nécessairement gaussien, mais dont toutes les composantes X_i admettent un moment d'ordre 2, ce qu'on note de façon naturelle $X \in L^2(\Omega)$. On peut alors définir la moyenne m de ce vecteur par :

$$m = \mathbb{E}[X] = \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_d] \end{bmatrix},$$

et sa matrice de covariance :

$$\Gamma = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])'],$$

encore appelée matrice de dispersion, de taille $d \times d$, avec pour terme générique :

$$\Gamma_{i,j} = \text{Cov}(X_i, X_j).$$

On peut donner une propriété générale sur ces matrices de dispersion.

Proposition 10 (Matrice de covariance)

Si elle existe, la matrice de covariance d'un vecteur aléatoire est symétrique réelle positive. Elle est donc diagonalisable en base orthonormée :

$$\Gamma = P' \Delta P,$$

avec $P' = P^{-1}$ et $\Delta = \text{diag}\{\lambda_1, \dots, \lambda_d\}$, les λ_i étant tous positifs ou nuls.

Achtung! Quand on parle d'une matrice symétrique réelle S , dire qu'elle est positive ne signifie pas que ses coefficients sont positifs! On entend par là que :

$$\forall u \in \mathbb{R}^d, \quad u' S u \geq 0.$$

On dit aussi que la forme quadratique associée est positive.

Preuve. L'aspect symétrique réel est clair par définition de la matrice de covariance. Il faut prouver que pour tout vecteur réel $u = [u_1, \dots, u_d]'$, on a $u' \Gamma u \geq 0$. Or cette quantité vaut :

$$u' \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])'] u = \mathbb{E}[(u'(X - \mathbb{E}[X]))((X - \mathbb{E}[X])' u)] = \mathbb{E}[(u'(X - \mathbb{E}[X]))^2] \geq 0.$$



Au passage, on a établi le résultat suivant, utile dans les applications.

Proposition 11 (Variance et matrice de dispersion)

Soit $X = [X_1, \dots, X_d]'$ un vecteur aléatoire de matrice de dispersion Γ . La variable aléatoire $Z = \alpha_1 X_1 + \dots + \alpha_d X_d = \alpha' X$ a pour variance :

$$\text{Var}(Z) = \alpha' \Gamma \alpha = [\alpha_1, \dots, \alpha_d] \Gamma \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_d \end{bmatrix}.$$

Remarque. La matrice Γ n'est pas nécessairement définie positive. Par exemple, si X_1 est une variable aléatoire de variance 1, le vecteur $X = [X_1, 1 + X_1]'$ a pour matrice de dispersion :

$$\Gamma = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

qui est clairement de rang 1. On voit que le vecteur aléatoire X , a priori à valeurs dans \mathbb{R}^2 , ne prend en fait ses valeurs que sur la droite d'équation $y = 1 + x$. Ceci est vrai de façon générale : Γ est de rang strictement inférieur à d si et seulement si le vecteur aléatoire X ne prend ses valeurs que dans un sous-espace affine de \mathbb{R}^d .

Rappelons que si X est un vecteur aléatoire de dimension d , on peut définir sa fonction caractéristique comme suit

$$\Phi_X : \begin{cases} \mathbb{R}^d & \rightarrow \mathbb{C} \\ u = [u_1, \dots, u_d]' & \mapsto \Phi_X(u) = \mathbb{E}[e^{i\langle u, X \rangle}] = \mathbb{E}[e^{i \sum_{j=1}^d u_j X_j}] \end{cases}$$

Sans énumérer toutes les propriétés de la fonction caractéristique d'un vecteur aléatoire, disons simplement que :

- Comme en dimension 1, elle sert à démontrer la convergence en loi d'une suite de vecteurs aléatoires (cf. infra la version vectorielle du théorème central limite).
- Les variables aléatoires X_1, \dots, X_d sont indépendantes si et seulement si :

$$\forall u \in \mathbb{R}^d \quad \Phi_X(u) = \prod_{j=1}^d \Phi_{X_j}(u_j).$$

De plus, tout comme en dimension 1, une loi gaussienne multidimensionnelle est complètement caractérisée par la fonction caractéristique, laquelle ne fait intervenir que le vecteur moyenne et la matrice de dispersion.

Proposition 12 (Fonction caractéristique d'un vecteur gaussien)

Soit X un vecteur aléatoire de dimension d , de vecteur moyenne m et de matrice de covariance Γ , alors X est gaussien si et seulement si sa fonction caractéristique s'écrit pour tout $u \in \mathbb{R}^d$:

$$\Phi_X(u) = e^{iu'm - \frac{1}{2}u'\Gamma u}$$

On note alors $X \sim \mathcal{N}_d(m, \Gamma)$.

Preuve. Supposons le vecteur $X = [X_1, \dots, X_d]'$ gaussien, de moyenne m et de matrice de covariance Γ . Alors la variable aléatoire :

$$Y = u_1 X_1 + \dots + u_d X_d = u'X$$

est gaussienne, de moyenne :

$$\mu = u_1 \mathbb{E}[X_1] + \dots + u_d \mathbb{E}[X_d] = u'm,$$

et de variance (cf. proposition 11) :

$$\sigma^2 = u'\Gamma u.$$

On peut alors appliquer la Proposition 7 :

$$\Phi_Y(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}} = e^{iu'mt - \frac{1}{2}u'\Gamma u t^2}.$$

Et on conclut en remarquant que :

$$\Phi_X(u) = \mathbb{E}[e^{iu'X}] = \mathbb{E}[e^{iY}] = \Phi_Y(1) = e^{iu'm - \frac{1}{2}u'\Gamma u}.$$

Réciproquement, supposons $X = [X_1, \dots, X_d]'$ vecteur aléatoire de moyenne m , de matrice de covariance Γ et de fonction caractéristique :

$$\Phi_X(u) = e^{iu'm - \frac{1}{2}u'\Gamma u}$$

La variable aléatoire :

$$Y = \alpha_1 X_1 + \dots + \alpha_d X_d = \alpha'X$$

a pour fonction caractéristique :

$$\Phi_Y(t) = \mathbb{E}[e^{itY}] = \mathbb{E}[e^{i(t\alpha)'X}] = \Phi_X(t\alpha),$$

c'est-à-dire :

$$\Phi_Y(t) = e^{i(t\alpha)'m - \frac{1}{2}(t\alpha)'\Gamma(t\alpha)} = e^{i(\alpha'm)t - \frac{1}{2}(\alpha'\Gamma\alpha)t^2}.$$

La Proposition 7 assure donc que Y est gaussienne et plus précisément :

$$Y \sim \mathcal{N}(\alpha'm, \alpha'\Gamma\alpha).$$

Ainsi X est bien un vecteur aléatoire gaussien. ■

Les lois normales sont stables par transformation affine. Le résultat suivant, sur lequel on serait tenté de jeter un coup d'œil distrait, est d'utilité constante dans la manipulation des vecteurs gaussiens.

Proposition 13 (Transformation affine)

Si X est un vecteur gaussien d -dimensionnel, avec $X \sim \mathcal{N}_d(m, \Gamma)$, si $A \in M_{k,d}(\mathbb{R})$ et si $B \in M_{k,1}(\mathbb{R})$, alors le vecteur $Y = AX + B$ est gaussien avec :

$$Y \sim \mathcal{N}_k(Am + B, A\Gamma A').$$

Preuve. Il suffit d'utiliser la caractérisation par la fonction caractéristique ci-dessus. On a en effet :

$$\forall u \in \mathbb{R}^k \quad \Phi_Y(u) = \mathbb{E}[e^{iu'Y}] = \mathbb{E}[e^{iu'(AX+B)}] = e^{iu'B} \mathbb{E}[e^{i(u'A)X}],$$

c'est-à-dire :

$$\Phi_Y(u) = e^{iu'B} \Phi_X(A'u) = e^{iu'(Am+B) - \frac{1}{2}u'(A\Gamma A')u},$$

ce qui exactement dire que :

$$Y \sim \mathcal{N}_k(Am + B, A\Gamma A').$$

■

Remarque. Il arrive souvent qu'un vecteur Y soit construit à partir d'un autre vecteur X par une transformation affine. Si X est un vecteur gaussien, par exemple lorsqu'il est composé de variables aléatoires gaussiennes indépendantes, ce résultat permet d'en déduire automatiquement le fait que Y est aussi un vecteur gaussien. Voir les exercices de fin de chapitre : *Processus autorégressif*, *Moyenne mobile*.

Rappel. Soit X et Y deux variables aléatoires de carrés intégrables. On dit qu'elles sont non corrélées si :

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

ce qui équivaut à dire que : $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, ou encore que la matrice de covariance du vecteur $[X, Y]'$ est diagonale. Ceci est bien sûr toujours vrai lorsqu'elles sont indépendantes, puisqu'alors on a plus généralement pour toutes "bonnes" fonctions f et g :

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

La réciproque est fautive en général, comme le montre l'exemple suivant.

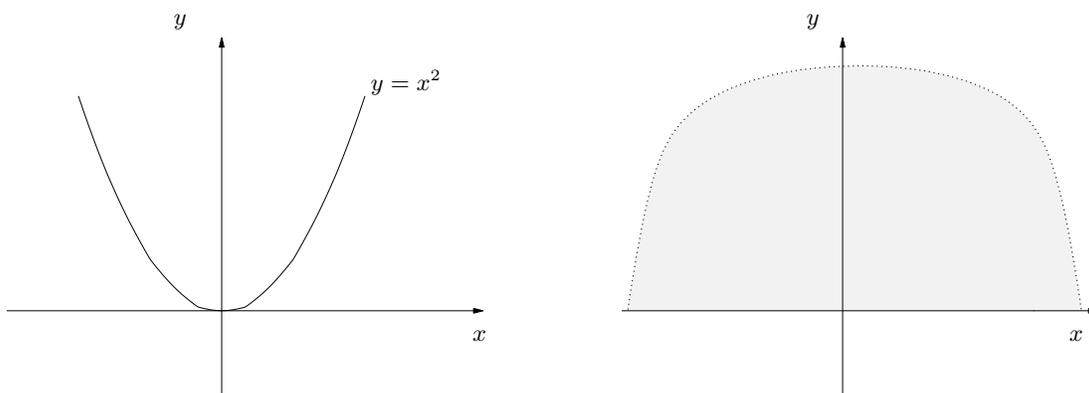


FIGURE 2.5 – $\text{Supp}(X, Y)$ (à gauche) $\neq \text{Supp}(X) \times \text{Supp}(Y)$ (à droite).

Exemple : Décorrélations $\not\Rightarrow$ Indépendance

Soit $X \sim \mathcal{N}(0, 1)$ et $Y = X^2$, donc $\mathbb{E}[Y] = \mathbb{E}[X^2] = \text{Var}(X) = 1$. X et Y sont bien décorrélées puisque $\mathbb{E}[X]\mathbb{E}[Y] = 0 \times 1 = 0$ et :

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = 0,$$

une loi gaussienne ayant tous ses moments d'ordres impairs nuls. Cependant X et Y ne sont pas indépendantes. Ceci est clair intuitivement puisque Y est une fonction déterministe de X . On peut aussi le justifier par l'espérance conditionnelle :

$$\mathbb{E}[Y|X] = \mathbb{E}[X^2|X] = X^2 \neq \mathbb{E}[Y] = 1.$$

Une dernière façon de le voir est de remarquer que le support du vecteur aléatoire $[X, Y]'$ est la parabole $y = x^2$ du plan et non le produit cartésien $\mathbb{R} \times \mathbb{R}^+$ des supports des variables (voir figure 2.5).

Dans le cas de vecteurs gaussiens, cependant, la décorrélation est équivalente à l'indépendance.

Proposition 14 (Indépendance \Leftrightarrow Décorrélation)

Soit $X = [X_1, \dots, X_d]'$ un vecteur aléatoire gaussien. Les variables aléatoires (X_1, \dots, X_d) sont indépendantes si et seulement si elles sont non corrélées, c'est-à-dire si et seulement si la matrice de dispersion Γ est diagonale.

Preuve. Supposons X gaussien et de composantes indépendantes. Alors ces composantes sont a fortiori non corrélées, c'est-à-dire :

$$\forall (i, j) \in \{1, \dots, d\}^2 \quad \text{Cov}(X_i, X_j) = 0,$$

et la matrice Γ est diagonale. Ceci est d'ailleurs toujours vrai, l'aspect gaussien de X n'est pas nécessaire.

Réciproquement, supposons X gaussien et de matrice de covariance Γ diagonale :

$$\Gamma = \text{diag}(\sigma_1^2, \dots, \sigma_d^2).$$

Si on note $m = [m_1, \dots, m_d]'$ la moyenne de X , celui-ci admet pour fonction caractéristique :

$$\Phi_X(u) = e^{iu'm - \frac{1}{2}u'\Gamma u},$$

qu'on peut factoriser en :

$$\Phi_X(u) = \prod_{j=1}^d \Phi_{X_j}(u_j),$$

où Φ_{X_j} est tout bonnement la fonction caractéristique de X_j :

$$\forall j \in \{1, \dots, d\} \quad \Phi_{X_j}(u_j) = e^{im_j u_j - \frac{\sigma_j^2 u_j^2}{2}}.$$

Ainsi la fonction caractéristique du vecteur $X = [X_1, \dots, X_d]'$ est le produit des fonctions caractéristiques de ses composantes X_j : c'est une caractérisation de l'indépendance des X_j . ■

Remarque. Pour pouvoir appliquer le critère d'indépendance ci-dessus, il faut que le **vecteur** soit gaussien : le fait que les **composantes** le soient n'est pas suffisant. Pour s'en convaincre, il suffit de revenir à l'exemple vu précédemment : $X \sim \mathcal{N}(0, 1)$ et $Y = \varepsilon X$, avec ε variable de Rademacher indépendante de X . On a vu que Y suit elle aussi une loi normale centrée réduite, donc les deux variables X et Y sont gaussiennes. De plus, puisque X et ε sont indépendantes, on a :

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[\varepsilon X^2] = \mathbb{E}[\varepsilon]\mathbb{E}[X^2] = 0,$$

la dernière égalité venant du fait que ε est centrée. Ainsi X et Y sont gaussiennes et décorrélées. Pourtant elles ne sont pas indépendantes : il appert qu'une fois connue la réalisation de X , Y ne

peut valoir que la même chose ou l'opposé. Plus rigoureusement, on peut à nouveau le justifier par les supports : le support du couple (X, Y) est l'union des deux droites $y = x$ et $y = -x$, tandis que le produit cartésien des supports de X et Y est le plan \mathbb{R}^2 . Le problème, déjà constaté, vient de ce que le vecteur $[X, Y]'$ n'est pas gaussien, bien que ses composantes le soient.

Etant donné un vecteur gaussien $X = [X_1, \dots, X_d]'$ de moyenne m et de matrice de covariance Γ données, on peut toujours se ramener à un vecteur aléatoire dont les composantes sont indépendantes.

Proposition 15 (Changement de repère orthonormal)

Soit $X = [X_1, \dots, X_d]'$ un vecteur gaussien de moyenne m et de matrice de covariance Γ . Il existe P orthogonale telle que $P\Gamma P' = \Delta = \text{diag}(\lambda_1, \dots, \lambda_d)$, avec les $\lambda_j \geq 0$. Alors les composantes Y_j du vecteur aléatoire $Y = P(X - m)$ sont des variables aléatoires gaussiennes indépendantes centrées de variances respectives λ_j .

Preuve. Puisque Γ est symétrique réelle positive, elle est diagonalisable en base orthonormée : $\Gamma = P'\Delta P$, avec :

$$\Delta = \text{diag}(\lambda_1, \dots, \lambda_d),$$

où les λ_j sont les valeurs propres positives de Γ et P une matrice orthogonale. Si on considère maintenant le nouveau vecteur aléatoire

$$Y = [Y_1, \dots, Y_d]' = P(X - m) = PX - Pm,$$

c'est encore un vecteur gaussien, en tant que transformée affine d'un vecteur gaussien (Proposition 13). Plus précisément, on sait que :

$$Y \sim \mathcal{N}_d(Pm - Pm, P\Gamma P') = \mathcal{N}_d(0, \Delta).$$

Ainsi le vecteur gaussien Y est centré et ses composantes sont indépendantes, puisque sa matrice de dispersion est diagonale (Proposition 14). ■

Remarques :

1. La représentation de la densité de Y est bien plus simple que celle de X puisqu'on s'est ramené à un produit de densités gaussiennes indépendantes : en dimension 2, on obtient donc une surface en cloche plus ou moins aplatie suivant la direction (cf. exercice *Changement de base*). La surface en cloche standard correspond à une loi centrée et de matrice de covariance identité (cf. figure 2.6).
2. Si $\lambda_j = 0$ pour un indice j , la loi du vecteur X est dégénérée. Et on a alors $Y_j = 0$ presque sûrement (cf. exercice *Problème de dégénérescence*).

Applications.

1. Simulation : supposons qu'on ait à simuler un vecteur gaussien $X = [X_1, \dots, X_d]'$ de moyenne m et de matrice de covariance Γ données. On dispose simplement d'un générateur de variables normales centrées réduites indépendantes (par exemple `rnorm` en R). On peut écrire comme ci-dessus $\Gamma = P'\Delta P = A'A$, avec $A = P'\sqrt{\Delta}P$ et :

$$\sqrt{\Delta} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d}).$$

Il suffit alors de simuler d variables i.i.d. $U_j \sim \mathcal{N}(0, 1)$ et de considérer : $X = AU + m$. Le résultat de transformation affine et le fait que la matrice de covariance de U soit l'identité assurent que X a les propriétés requises.

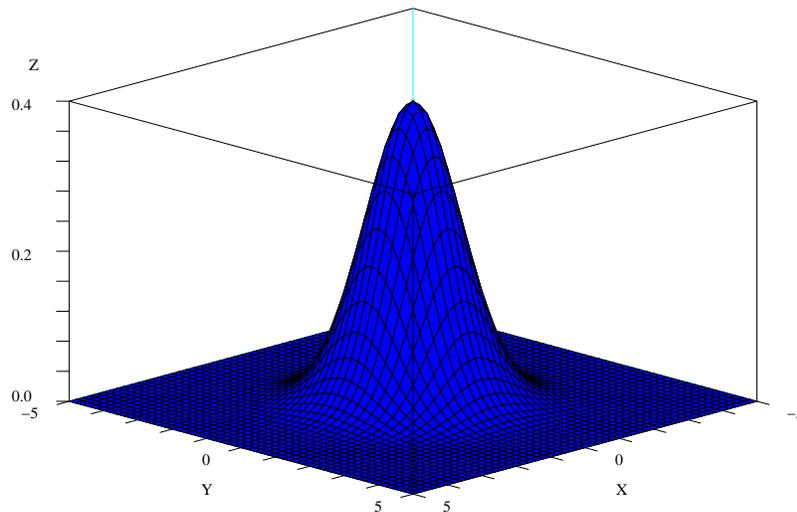


FIGURE 2.6 – Densité d’un vecteur gaussien centré de matrice de dispersion identité.

2. Composantes principales : dans les applications, on a souvent à traiter des vecteurs gaussiens $X = [X_1, \dots, X_d]'$ de très grande dimension d , mais dont les coordonnées présentent de très fortes corrélations entre elles. Dans ce cas, après transformation comme ci-dessus, le nombre de valeurs propres λ_j qui ne sont pas voisines de 0 est bien plus petit que d . Supposons que les valeurs propres soient rangées par ordre décroissant dans Δ . On se fixe alors un seuil $\varepsilon > 0$: il y a d_ε valeurs propres $\lambda_j > \varepsilon$. On considère le nouveau vecteur aléatoire $\tilde{Y} = [\tilde{Y}_1, \dots, \tilde{Y}_d]'$ défini par

$$\tilde{Y}_j = \begin{cases} Y_j & \text{si } j \leq d_\varepsilon \\ 0 & \text{sinon} \end{cases}$$

et le vecteur $\tilde{X} = P'\tilde{Y} + m$ associé. Les vecteurs aléatoires \tilde{X} et \tilde{Y} sont les versions tronquées de X et Y . L’erreur quadratique moyenne faite en remplaçant X par \tilde{X} est :

$$\|X - \tilde{X}\|^2 = \mathbb{E}[(X - \tilde{X})'(X - \tilde{X})] = \mathbb{E}[(Y - \tilde{Y})'(Y - \tilde{Y})] = \|Y - \tilde{Y}\|^2 = \sum_{j=d_\varepsilon+1}^d \lambda_j.$$

Il suffit de comparer cette erreur à la dispersion totale de X :

$$\|X - m\|^2 = \mathbb{E}[(X - m)'(X - m)] = \sum_{j=1}^d \lambda_j,$$

pour avoir une idée de la qualité de l’approximation, à savoir : \tilde{X} approche bien X si tant est que le rapport :

$$\frac{\sum_{j=d_\varepsilon+1}^d \lambda_j}{\sum_{j=1}^d \lambda_j}$$

est petit. On dit alors qu’on a réduit X à ses composantes principales.

La situation agréable pour un vecteur gaussien est celle où la matrice de covariance est inversible (i.e. **définie** positive). Il admet alors une densité dans \mathbb{R}^d , que l'on peut facilement expliciter.

Proposition 16 (Densité d'un vecteur gaussien)

Si $X \sim \mathcal{N}_d(m, \Gamma)$, avec Γ inversible, alors X admet pour densité :

$$f(x) = f(x_1, \dots, x_d) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma}} e^{-\frac{1}{2}(x-m)' \Gamma^{-1}(x-m)}.$$

Remarque. En dimension 1, la formule exprimant la densité d'une variable gaussienne en fonction de la moyenne m et de la variance σ^2 n'est valable que si σ^2 est strictement positive, c'est-à-dire inversible. Il n'est donc pas étonnant qu'on retrouve la même condition en dimension supérieure.

Preuve. On utilise la transformation affine du résultat précédent : $Y = P(X - m)$, avec :

$$P\Gamma P' = \Delta = \text{diag}(\lambda_1, \dots, \lambda_d).$$

Dire que Γ est inversible équivaut à dire que les valeurs propres λ_j sont toutes strictement positives. Les composantes Y_1, \dots, Y_j sont indépendantes, avec $Y_j \sim \mathcal{N}(0, \lambda_j)$, donc Y admet pour densité :

$$f_Y(y) = \prod_{j=1}^d f_j(y_j) = \prod_{j=1}^d \frac{1}{\sqrt{2\pi\lambda_j}} e^{-\frac{y_j^2}{2\lambda_j}},$$

qu'on peut encore écrire :

$$f_Y(y) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma}} e^{-\frac{1}{2}y' \Delta^{-1}y}.$$

Pour retrouver la densité de X , il suffit alors d'appliquer la formule de changement de variable pour le \mathcal{C}^1 -difféomorphisme :

$$\phi : \begin{cases} \mathbb{R}^d & \rightarrow \mathbb{R}^d \\ x & \mapsto y = P(x - m) \end{cases}$$

Ce qui donne :

$$f_X(x) = f_Y(P(x - m)) |\det J_\phi(x)|.$$

Or ϕ est une transformation affine, donc :

$$\forall x \in \mathbb{R}^d \quad J_\phi(x) = P,$$

et puisque P est orthogonale :

$$\forall x \in \mathbb{R}^d \quad |\det J_\phi(x)| = 1.$$

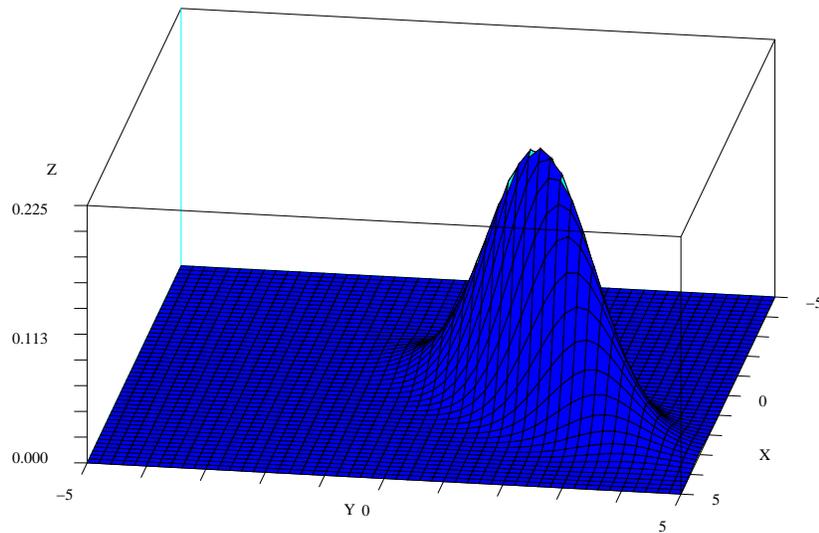
On en déduit la densité du vecteur X :

$$f(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma}} e^{-\frac{1}{2}(x-m)' \Gamma^{-1}(x-m)}.$$

■

Représentation. La figure 2.7 correspond à un vecteur gaussien $[X, Y]'$ de moyenne $m = [1, 2]'$ et de matrice de dispersion :

$$\Gamma = \begin{bmatrix} 1/\sqrt{2} & 1 \\ 1 & 1/\sqrt{2} \end{bmatrix}.$$

FIGURE 2.7 – Densité d'un vecteur gaussien $\mathcal{N}_2(m, \Gamma)$.

La densité correspondante est donc :

$$f(x, y) = \frac{1}{\pi\sqrt{2}} e^{-((x-1)^2 - \sqrt{2}(x-1)(y-2) + (y-2)^2)}$$

Si Γ n'est pas inversible, on dit que la loi de X est dégénérée. Comme mentionné plus haut, ceci signifie que le vecteur aléatoire ne prend ses valeurs que dans un sous-espace affine de \mathbb{R}^d (cf. exercice *Problème de dégénérescence*).

Exemple. Supposons que $X_1 \sim \mathcal{N}(0, 1)$, et considérons le vecteur gaussien $X = [X_1, 1 + X_1]'$. Sa matrice de dispersion est de rang 1 :

$$\Gamma = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

On voit que le vecteur aléatoire X ne prend en fait ses valeurs que sur la droite d'équation $y = 1 + x$, et non dans \mathbb{R}^2 tout entier. Ce vecteur n'admet donc pas de densité par rapport à la mesure de Lebesgue de \mathbb{R}^2 .

Enfin, on peut généraliser le théorème de la limite centrale aux vecteurs aléatoires. Le principe est rigoureusement le même qu'en dimension 1. On dit qu'une suite $(X_n)_{n \geq 0}$ de vecteurs aléatoires de \mathbb{R}^d converge en loi vers un vecteur gaussien $X \sim \mathcal{N}_d(m, \Gamma)$ si pour tout borélien $B \in \mathcal{B}_d$, on a :

$$\mathbb{P}(X_n \in B) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(X \in B).$$

Comme en dimension 1, la convergence simple sur \mathbb{R}^d de la suite des fonctions caractéristiques (Φ_n) vers la fonction caractéristique Φ_X est souvent le moyen le plus simple de montrer la convergence en loi.

Théorème 5 (Théorème central limite)

Soit $(X_n)_{n \geq 1} \in L^2(\Omega)$ une suite de vecteurs aléatoires de \mathbb{R}^d , indépendants et identiquement distribués, avec pour moyenne $m = \mathbb{E}[X_1]$ et matrice de dispersion $\Gamma = \mathbb{E}[(X_1 - m)(X_1 - m)']$. Alors la suite de vecteurs aléatoires $(\frac{S_n - nm}{\sqrt{n}})$, où $S_n = X_1 + \dots + X_n$, converge en loi vers un vecteur gaussien centré de matrice de dispersion Γ . On note :

$$\frac{S_n - nm}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_d(0, \Gamma).$$

Remarques.

- Rappelons le point remarquable du théorème central limite : quelle que soit la loi des vecteurs aléatoires X_n , on a convergence vers une loi gaussienne. C'est-à-dire que si n est assez grand, on a en gros :

$$\mathcal{L}(S_n) \approx \mathcal{N}(nm, n\Gamma).$$

- Avec les hypothèses ci-dessus, on s'intéresse à la suite (S_n) des sommes partielles. La loi forte des grands nombres dit que :

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} m.$$

C'est-à-dire qu'en première approximation, la somme de n vecteurs aléatoires i.i.d. de moyenne m se comporte comme une suite déterministe de vecteurs : $S_n \approx nm$. Question qui en découle naturellement : quel est l'ordre de grandeur de l'erreur effectuée en remplaçant S_n par nm ? C'est à cette question que répond le théorème central limite : l'erreur faite en remplaçant la somme partielle S_n par nm est de l'ordre de \sqrt{n} . On peut donc voir la loi des grands nombres et le théorème central limite comme le début du développement asymptotique de la somme S_n .

- L'exercice *Sauts de puce* donne un exemple d'application de ce théorème.

2.2 Conditionnement des vecteurs gaussiens

2.2.1 Conditionnement pour un couple gaussien

Rappel du premier chapitre : soit (X, Y) un couple aléatoire dont on connaît la loi jointe. Supposons qu'on observe $X = x$ et qu'on veuille en déduire une estimation de Y . Ceci signifie qu'on définit une fonction de \mathbb{R} dans \mathbb{R} qui à tout point x associe un point $u(x)$. Si la qualité de l'approximation est mesurée par l'erreur quadratique moyenne, i.e. par la quantité :

$$\mathbb{E}[(Y - u(X))^2],$$

on a vu que le mieux à faire est de prendre pour u la fonction de régression de Y sur X , c'est-à-dire la fonction qui à x associe $\mathbb{E}[Y|X = x]$. D'après le théorème de projection, la variable aléatoire $\mathbb{E}[Y|X]$ est la fonction $u(X)$ caractérisée par la double propriété² :

$$\begin{cases} u(X) \in L^2(X) \\ Y - u(X) \perp L^2(X) \end{cases}$$

avec comme en chapitre 1 :

$$L^2(X) \triangleq \{u(X) \text{ avec } u : \mathbb{R} \mapsto \mathbb{R} \text{ borélienne telle que } \mathbb{E}[u^2(X)] < +\infty\}.$$

Néanmoins, l'espérance conditionnelle n'est pas nécessairement un objet très simple à calculer ou à estimer. C'est pourquoi, faute de mieux, on se contente parfois de chercher à approcher au mieux

2. On rappelle que dans $L^2(\Omega) : X \perp Y$ signifie que $\mathbb{E}[XY] = 0$.

Y par une fonction affine de X : c'est ce qu'on appelle faire de la régression linéaire. Ce qu'on perd en qualité d'approximation, on le gagne en commodité de résolution. La droite de régression est donnée par : $f(X) = aX + b$, avec

$$\begin{cases} a &= \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \\ b &= \mathbb{E}[Y] - a\mathbb{E}[X] \end{cases}$$

c'est-à-dire :

$$f(X) = \mathbb{E}[Y] + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - \mathbb{E}[X]).$$

Nous allons voir que, dans le cas gaussien, ce procédé est complètement légitime : la fonction de régression est exactement la droite de régression !

Théorème 6 (Espérance conditionnelle \Leftrightarrow droite de régression)

Si $[X, Y]'$ est un vecteur gaussien, alors :

$$\mathbb{E}[Y|X] = \mathbb{E}[Y] + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - \mathbb{E}[X]).$$

Autrement dit, courbe de régression et droite de régression coïncident.

Preuve. Il suffit de prouver que la fonction u définie par :

$$u(X) = \mathbb{E}[Y] + \frac{\text{Cov}(X,Y)}{\text{Var}(X)}(X - \mathbb{E}[X]),$$

vérifie bien la double propriété de caractérisation de l'espérance conditionnelle. Puisque X est gaussienne, elle est dans $L^2(\Omega)$, et par suite $u(X) = aX + b$ est dans $L^2(X)$. Il reste à prouver que la variable aléatoire $(Y - u(X))$ est orthogonale au sous-espace $L^2(X)$, c'est-à-dire orthogonale à toute variable aléatoire $f(X)$ fonction de X .

On commence par montrer que $(Y - u(X))$ est indépendante de X . Puisque le vecteur $[X, Y]'$ est gaussien et que :

$$\begin{bmatrix} X \\ Y - u(X) \end{bmatrix} = \begin{bmatrix} X \\ Y - (aX + b) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -a & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} 0 \\ -b \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix} + B,$$

le vecteur $[X, Y - u(X)]'$ est gaussien aussi comme transformée affine d'un vecteur gaussien, donc montrer l'indépendance de ses composantes revient à montrer leur décorrélation. Or :

$$\text{Cov}(X, Y - u(X)) = \text{Cov}(X, Y) - \text{Cov}(X, u(X)) = \text{Cov}(X, Y) - \text{Cov}(X, aX + b),$$

et par définition de a , on a :

$$\text{Cov}(X, aX + b) = \text{Cov}(X, aX) = a\text{Var}(X) = \text{Cov}(X, Y),$$

de sorte qu'on a bien $\text{Cov}(X, Y - u(X)) = 0$, c'est-à-dire que X et $(Y - u(X))$ sont indépendantes. Mais si $(Y - u(X))$ est une variable aléatoire indépendante de X , elle est aussi indépendante de toute fonction $f(X)$ de la variable X . Par suite :

$$\langle f(X), Y - u(X) \rangle = \mathbb{E}[f(X)(Y - u(X))] = \mathbb{E}[f(X)]\mathbb{E}[Y - u(X)] = 0,$$

car $(Y - u(X))$ est centrée. Ainsi la seconde propriété de caractérisation de l'espérance conditionnelle est vérifiée et le théorème est prouvé. ■

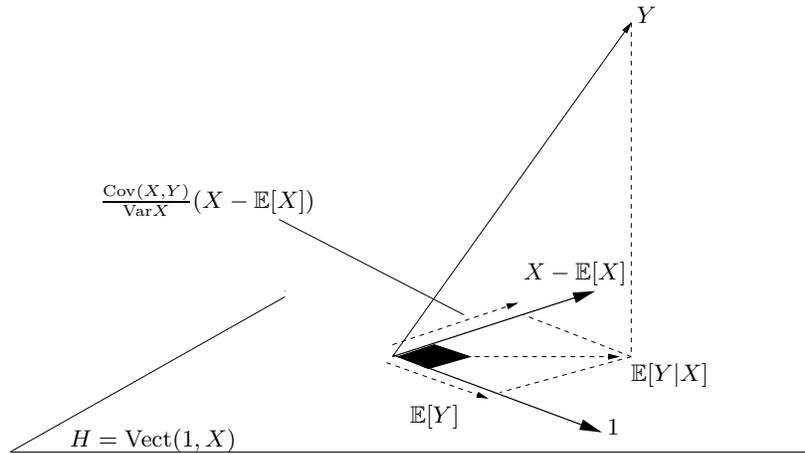


FIGURE 2.8 – $\mathbb{E}[Y|X]$ projeté orthogonal de Y sur la base orthogonale $(1, X - \mathbb{E}[X])$.

Rappel. Dans un espace de Hilbert, le projeté orthogonal de Y sur un sous-espace vectoriel $H = \text{Vect}(e_1, \dots, e_d)$, avec les e_i orthogonaux, est :

$$\pi_H(Y) = \sum_{i=1}^d \left\langle Y, \frac{e_i}{\|e_i\|} \right\rangle \frac{e_i}{\|e_i\|} = \sum_{i=1}^d \frac{\langle Y, e_i \rangle}{\|e_i\|^2} e_i.$$

Interprétation géométrique. Dans le cas général, la variable aléatoire :

$$\mathbb{E}[Y] + \frac{\text{Cov}(X, Y)}{\text{Var}(X)}(X - \mathbb{E}[X])$$

peut se voir comme la projection orthogonale de la variable aléatoire Y sur l'espace engendré par la variable aléatoire constante 1 (que nous appellerons Michel) et la variable aléatoire X . Notons $H = \text{Vect}(1, X)$ ce sous-espace vectoriel de $L^2(\Omega)$. Une famille orthogonale est formée des deux vecteurs 1 et $X - \mathbb{E}[X]$. Appliquons le rappel ci-dessus (voir aussi figure 2.8). La projection de Y sur le premier vecteur donne :

$$\frac{\mathbb{E}[Y \cdot 1]}{\mathbb{E}[1^2]} = \mathbb{E}[Y],$$

et la projection sur le second vecteur donne :

$$\frac{\mathbb{E}[Y \cdot (X - \mathbb{E}[X])]}{\mathbb{E}[(X - \mathbb{E}[X])^2]} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}.$$

Au total, on retrouve bien la droite de régression.

L'interprétation géométrique ci-dessus est toujours vraie. Dans le cas gaussien, on vient de plus de voir que la projection orthogonale de Y sur $L^2(X)$ est exactement la projection orthogonale sur $H = \text{Vect}(1, X)$.

2.2.2 Hyperplan de régression

On ne fait dans ce paragraphe aucune hypothèse de gaussianité. On suppose observer n variables aléatoires X_1, \dots, X_n et on veut connaître la fonction affine des X_i , donc de la forme :

$$f(X_1, \dots, X_n) = b + a_1 X_1 + \dots + a_n X_n,$$

qui approche le mieux la variable aléatoire Y au sens des moindres carrés, c'est-à-dire telle que l'erreur quadratique moyenne :

$$\mathbb{E}[(Y - (b + a_1X_1 + \dots + a_nX_n))^2]$$

soit minimale. Autrement dit, au lieu de chercher la droite de régression, on cherche l'hyperplan de régression (voir figure 2.9 pour $n = 2$). Ceci revient à déterminer la projection $\pi_H(Y)$ de Y sur le sous-espace :

$$H = \text{Vect}(1, X_1, \dots, X_n),$$

engendré par la constante 1 et les variables aléatoires X_i .

Hypothèses :

- Notons $X = [X_1, \dots, X_n]'$ le vecteur formé des variables X_i . On suppose dans toute la suite que la matrice de dispersion $\Gamma_X = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])']$ est inversible.
- Puisqu'on parle de projections et d'erreurs quadratiques, on suppose aussi que toutes les variables aléatoires sont de carrés intégrables.

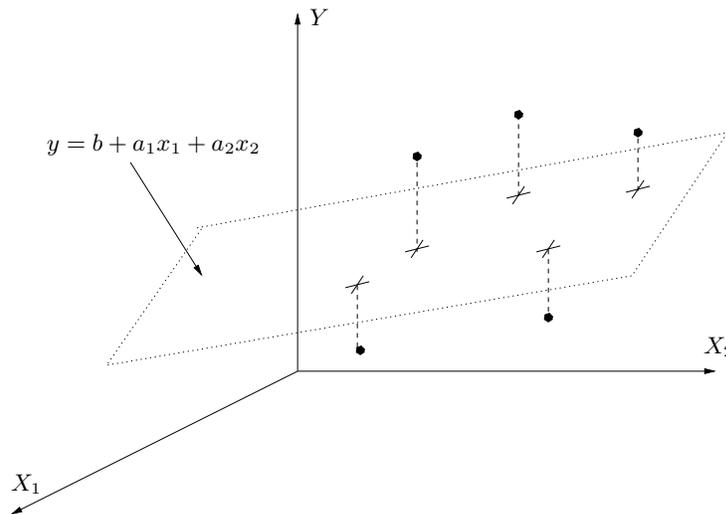


FIGURE 2.9 – Interprétation graphique de l'hyperplan de régression.

Théorème 7 (Hyperplan de régression)

La projection orthogonale de Y sur H est :

$$\pi_H(Y) = b + \sum_{i=1}^n a_i(X_i - \mathbb{E}[X_i]) = \mathbb{E}[Y] + \Gamma_{Y,X} \Gamma_X^{-1}(X - \mathbb{E}[X]),$$

avec :

$$\Gamma_{Y,X} = \mathbb{E}[(Y - \mathbb{E}[Y])(X - \mathbb{E}[X])'] = [\text{Cov}(Y, X_1), \dots, \text{Cov}(Y, X_n)],$$

matrice ligne de covariance de la variable aléatoire Y et du vecteur aléatoire X .

Preuve. La projection orthogonale de Y sur H est de la forme :

$$\pi_H(Y) = b + \sum_{i=1}^n a_i X_i.$$

Dire que $Y - \pi_H(Y)$ est orthogonal à H est équivalent à dire que $Y - \pi_H(Y)$ est orthogonal à chacun des vecteurs qui engendrent H , c'est-à-dire : $1, X_1, \dots, X_n$.

L'orthogonalité à 1 donne :

$$\left\langle Y - b - \sum_{i=1}^n a_i X_i, 1 \right\rangle = \mathbb{E}[Y] - b - \sum_{i=1}^n a_i \mathbb{E}[X_i] = 0,$$

c'est-à-dire :

$$b = \mathbb{E}[Y] - \sum_{i=1}^n a_i \mathbb{E}[X_i].$$

L'orthogonalité aux X_j donne les n équations :

$$\left\langle Y - b - \sum_{i=1}^n a_i X_i, X_j \right\rangle = 0 \quad 1 \leq j \leq n,$$

ce qui s'écrit encore :

$$\left\langle Y - \mathbb{E}[Y] - \sum_{i=1}^n a_i (X_i - \mathbb{E}[X_i]), X_j \right\rangle = 0 \quad 1 \leq j \leq n,$$

ou encore :

$$\left\langle Y - \mathbb{E}[Y] - \sum_{i=1}^n a_i (X_i - \mathbb{E}[X_i]), X_j - \mathbb{E}[X_j] \right\rangle = 0 \quad 1 \leq j \leq n.$$

Avec les notations de l'énoncé, ces n équations se résument sous forme matricielle à :

$$\Gamma_{Y,X} = [a_1, \dots, a_n] \Gamma_X,$$

c'est-à-dire :

$$[a_1, \dots, a_n] = \Gamma_{Y,X} \Gamma_X^{-1}.$$

En revenant à $\pi_H(Y)$, ceci donne :

$$\pi_H(Y) = b + \sum_{i=1}^n a_i X_i = \mathbb{E}[Y] - \Gamma_{Y,X} \Gamma_X^{-1} \mathbb{E}[X] + \Gamma_{Y,X} \Gamma_X^{-1} X,$$

c'est-à-dire :

$$\pi_H(Y) = \mathbb{E}[Y] + \Gamma_{Y,X} \Gamma_X^{-1} (X - \mathbb{E}[X]).$$

■

Remarque. En prenant $X = X_1$, on retrouve bien la droite de régression puisque $\Gamma_{Y,X} = \text{Cov}(X, Y)$ et $\Gamma_X = \text{Var}(X)$.

Corollaire 1 (Erreur quadratique moyenne)

L'erreur quadratique moyenne dans l'approximation par l'hyperplan de régression, encore appelée variance résiduelle ou résidu, est :

$$\mathbb{E}[(Y - \pi_H(Y))^2] = \Gamma_Y - \Gamma_{Y,X} \Gamma_X^{-1} \Gamma_{X,Y},$$

avec $\Gamma_Y = \text{Var}(Y)$ et $\Gamma_{X,Y} = (\Gamma_{Y,X})'$.

Preuve. Il suffit de l'écrire :

$$\mathbb{E}[(Y - \pi_H(Y))^2] = \mathbb{E}[(Y - \mathbb{E}[Y]) - \Gamma_{Y,X}\Gamma_X^{-1}(X - \mathbb{E}[X])]^2,$$

ce qui donne une combinaison de 3 termes. Le premier est simple :

$$\mathbb{E}[(Y - \mathbb{E}[Y])^2] = \text{Var}(Y).$$

Le deuxième l'est un peu moins :

$$\mathbb{E}[(Y - \mathbb{E}[Y])\Gamma_{Y,X}\Gamma_X^{-1}(X - \mathbb{E}[X])] = \Gamma_{Y,X}\Gamma_X^{-1}\mathbb{E}[(Y - \mathbb{E}[Y])(X - \mathbb{E}[X])] = \Gamma_{Y,X}\Gamma_X^{-1}\Gamma_{X,Y}.$$

Et le troisième encore moins :

$$\mathbb{E}[(\Gamma_{Y,X}\Gamma_X^{-1}(X - \mathbb{E}[X]))^2] = \mathbb{E}[(\Gamma_{Y,X}\Gamma_X^{-1}(X - \mathbb{E}[X]))(\Gamma_{Y,X}\Gamma_X^{-1}(X - \mathbb{E}[X]))'],$$

ce qui aboutit à :

$$\mathbb{E}[(\Gamma_{Y,X}\Gamma_X^{-1}(X - \mathbb{E}[X]))^2] = \Gamma_{Y,X}\Gamma_X^{-1}\Gamma_{X,Y}.$$

On remet tout bout à bout :

$$\mathbb{E}[(Y - \pi_H(Y))^2] = \text{Var}(Y) - 2\Gamma_{Y,X}\Gamma_X^{-1}\Gamma_{X,Y} + \Gamma_{Y,X}\Gamma_X^{-1}\Gamma_{X,Y} = \text{Var}(Y) - \Gamma_{Y,X}\Gamma_X^{-1}\Gamma_{X,Y}. \quad \blacksquare$$

$$\begin{bmatrix} X_1 \\ \vdots \\ X_n \\ Y \end{bmatrix} \implies \Gamma = \begin{bmatrix} \Gamma_X & \Gamma_{X,Y} \\ \Gamma_{Y,X} & \Gamma_Y \end{bmatrix}$$

FIGURE 2.10 – Les sous-matrices mises en jeu dans le calcul de $\mathbb{E}[Y|X]$.

Les matrices de covariance Γ_X , $\Gamma_{Y,X}$, $\Gamma_{X,Y}$ et Γ_Y peuvent se voir comme des sous-matrices de la matrice de covariance globale Γ du vecteur $[X, Y]' = [X_1, \dots, X_n, Y]'$. C'est ce que représente la figure 2.10.

Remarque. On adopte dès à présent la notation Γ_Y pour la variance de la variable aléatoire Y en vue de la généralisation de fin de chapitre : lorsque Y ne sera plus une variable aléatoire, mais un vecteur aléatoire, cette formule sera encore valide.

2.2.3 Espérance conditionnelle gaussienne

On suppose maintenant le vecteur $[X_1, \dots, X_n, Y]$ gaussien. L'espérance conditionnelle de Y sachant $X = [X_1, \dots, X_n]'$ est la projection orthogonale de Y sur l'espace des fonctions $u(X) = u(X_1, \dots, X_n)$, avec $u : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que $\mathbb{E}[u^2(X)] < +\infty$. C'est la fonction qui minimise

$\mathbb{E}[(Y - u(X))^2]$. On a vu que pour un vecteur gaussien bidimensionnel $[X, Y]'$, la droite de régression coïncide avec la courbe de régression. Plus généralement, on montre que pour un vecteur gaussien $[X_1, \dots, X_n, Y]'$, l'espérance conditionnelle coïncide avec la projection sur l'hyperplan de régression.

Théorème 8 (Espérance conditionnelle \Leftrightarrow Hyperplan de régression)

Si $[X_1, \dots, X_n, Y]'$ est un vecteur gaussien, alors :

$$\mathbb{E}[Y|X] = \mathbb{E}[Y|X_1, \dots, X_n] = \mathbb{E}[Y] + \Gamma_{Y,X} \Gamma_X^{-1} (X - \mathbb{E}[X]),$$

Et la variance résiduelle vaut :

$$\sigma^2 = \mathbb{E}[(Y - \mathbb{E}[Y|X])^2] = \Gamma_Y - \Gamma_{Y,X} \Gamma_X^{-1} \Gamma_{X,Y}.$$

Preuve. On pourrait la calquer sur celle vue pour la droite de régression. On adopte ici une autre technique, basée sur la propriété de distance minimale vérifiée par l'espérance conditionnelle. Notons comme ci-dessus $\pi_H(Y)$ la projection orthogonale de Y sur $H = \text{Vect}(1, X_1, \dots, X_n)$, c'est-à-dire :

$$\pi_H(Y) = \mathbb{E}[Y] + \Gamma_{Y,X} \Gamma_X^{-1} (X - \mathbb{E}[X]).$$

On a bien sûr $\pi_H(Y)$ de la forme $u(X)$ ci-dessus. De plus, dire que $(Y - \pi_H(Y))$ est orthogonale au sous-espace H signifie que $(Y - \pi_H(Y))$ est décorrélée des variables X_i (puisque $(Y - \pi_H(Y))$ est centrée). Mais puisque tout est gaussien, c'est exactement dire que $(Y - \pi_H(Y))$ est indépendante du vecteur X . Pour toute fonction u , on a donc :

$$\mathbb{E}[(Y - u(X))^2] = \mathbb{E}[(Y - \pi_H(Y)) + (\pi_H(Y) - u(X))]^2,$$

ce qui donne :

$$\mathbb{E}[(Y - u(X))^2] = \mathbb{E}[(Y - \pi_H(Y))^2] + 2\mathbb{E}[(Y - \pi_H(Y))(\pi_H(Y) - u(X))] + \mathbb{E}[(\pi_H(Y) - u(X))^2].$$

Or on vient de voir que :

$$\mathbb{E}[(Y - \pi_H(Y))(\pi_H(Y) - u(X))] = 0,$$

et le troisième terme est positif, donc pour toute fonction u , on a

$$\mathbb{E}[(Y - u(X))^2] \geq \mathbb{E}[(Y - \pi_H(Y))^2].$$

■

Remarque. Le terme $\Gamma_{Y,X} \Gamma_X^{-1} \Gamma_{X,Y}$ correspond à la variance de la variable aléatoire $\mathbb{E}[Y|X]$: il est donc positif et par suite $\mathbb{E}[(Y - \mathbb{E}[Y|X])^2] \leq \Gamma_Y$. Ce qui est logique : il y a moins d'incertitude sur la variable Y lorsqu'on connaît X et qu'on prend cette information en compte.

On a obtenu la décomposition orthogonale (voir figure 2.11) :

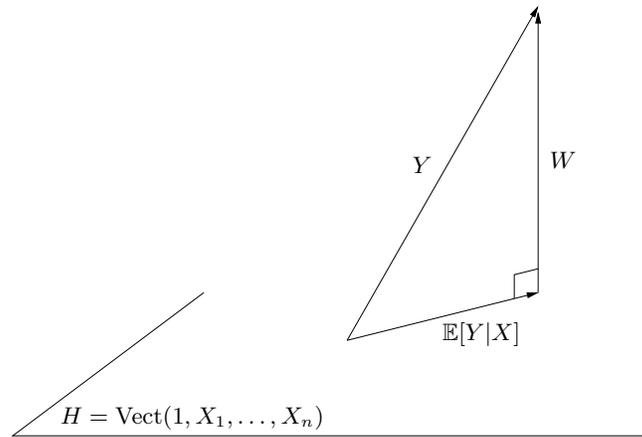
$$Y = \mathbb{E}[Y|X] + W = (\mathbb{E}[Y] + \Gamma_{Y,X} \Gamma_X^{-1} (X - \mathbb{E}[X])) + W,$$

c'est-à-dire que $W = Y - \mathbb{E}[Y|X]$ est une variable aléatoire gaussienne indépendante des X_i . W est centrée puisque $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$ et, par le théorème de Pythagore, sa variance est la variance résiduelle :

$$\sigma^2 = \Gamma_Y - \Gamma_{Y,X} \Gamma_X^{-1} \Gamma_{X,Y}.$$

En bref, on a :

$$\begin{cases} W \sim \mathcal{N}(0, \sigma^2) \\ W \perp X \end{cases}$$

FIGURE 2.11 – Décomposition orthogonale : $Y = \mathbb{E}[Y|X] + W$.

Application. Revenons à notre problème initial : on observe une réalisation $X = x = (x_1, \dots, x_n)$ du vecteur aléatoire X et on voudrait en déduire une idée de Y . Puisque W est indépendante des X_i , l'observation précédente n'a aucune conséquence sur la loi de W . Par contre, l'espérance conditionnelle de Y sachant X est désormais fixée à :

$$\mathbb{E}[Y|X = x] = \mathbb{E}[Y] + \Gamma_{Y,X}\Gamma_X^{-1}(x - \mathbb{E}[X]),$$

qui est un réel et non plus une variable aléatoire. Mais alors la décomposition orthogonale de Y devient :

$$Y = \mathbb{E}[Y|X = x] + W$$

et on en déduit que, sachant $X = x$:

$$Y \sim \mathcal{N}(\mathbb{E}[Y|X = x], \sigma^2).$$

Si on se souvient qu'initialement $Y \sim \mathcal{N}(\mathbb{E}[Y], \Gamma_Y)$, on voit que la nouvelle loi de Y a une moyenne dépendant de l'observation $X = x$ (voir figure 2.12). La variance résiduelle σ^2 est plus petite que $\text{Var}(Y)$, mais ne dépend que de la covariance des deux variables X et Y , pas de l'observation x proprement dite.

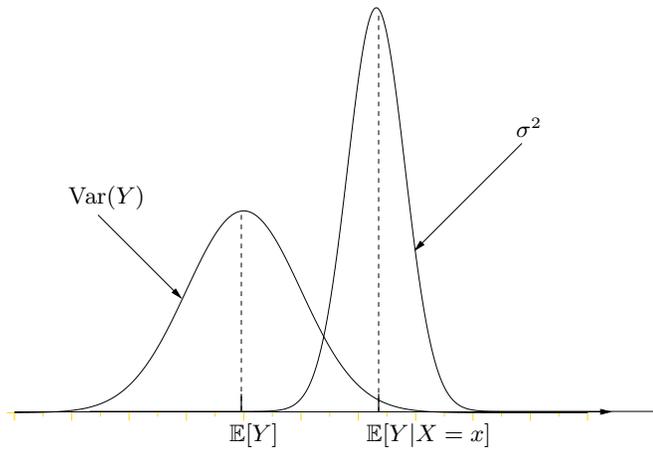
Cas particulier. Disons un mot du cas pathologique où la matrice de covariance Γ_X des X_i n'est pas inversible. Concrètement, ceci signifie que l'une au moins des composantes X_i peut s'exprimer linéairement en fonction des autres. Si on diagonalise la matrice de covariance en base orthonormée, i.e. $\Gamma_X = P'\Delta P$, certains termes de la diagonale sont nuls. On note Δ^+ la matrice diagonale obtenue en inversant les éléments non nuls de Δ (voir figure 2.13), alors :

$$\Gamma_X^+ = P\Delta^+P'$$

est appelée pseudo-inverse de Moore-Penrose de Γ_X (cf. cours d'algèbre linéaire). On montre alors que tous les résultats obtenus pour l'espérance conditionnelle sont conservés en remplaçant Γ_X^{-1} par Γ_X^+ . Par exemple :

$$\mathbb{E}[Y|X] = \mathbb{E}[Y] + \Gamma_{Y,X}\Gamma_X^+(X - \mathbb{E}[X]).$$

Généralisation. Si on suppose que le vecteur $[X, Y]' = [X_1, \dots, X_n, Y_1, \dots, Y_m]'$ est gaussien, toutes les formules et propriétés vues précédemment pour l'espérance conditionnelle passent sans

FIGURE 2.12 – Loi de Y et loi de Y sachant $X = x$.

problème. L'espérance conditionnelle de Y sachant X est un vecteur aléatoire de taille m donné par la formule :

$$\mathbb{E}[Y|X] = \mathbb{E}[Y] + \Gamma_{Y,X} \Gamma_X^+ (X - \mathbb{E}[X]).$$

C'est un vecteur gaussien, en tant que transformée affine du vecteur gaussien X , dont la moyenne vaut $\mathbb{E}[Y]$ et la matrice de covariance : $\Gamma_{Y,X} \Gamma_X^+ \Gamma_{X,Y}$. Le vecteur

$$W = Y - \mathbb{E}[Y|X]$$

est un vecteur gaussien centré de matrice de covariance : $\Gamma_Y - \Gamma_{Y,X} \Gamma_X^+ \Gamma_{X,Y}$. Qui plus est, W est indépendant de X .

$$\Delta = \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \lambda_r & \\ & & & & 0 \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix} \implies \Delta^+ = \begin{bmatrix} \frac{1}{\lambda_1} & & & & \\ & \frac{1}{\lambda_2} & & & \\ & & \ddots & & \\ & & & \frac{1}{\lambda_r} & \\ & & & & 0 \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$$

FIGURE 2.13 – Matrice diagonale Δ et sa pseudo-inverse de Moore-Penrose Δ^+ .

2.3 Exercices

“Tough ain’t enough.” Clint Eastwood, *Million Dollar Baby*.

Exercice 2.1 (Moments d'une loi normale)

Pour tout $n \in \mathbb{N}$, on note :

$$I_n = \int_{-\infty}^{+\infty} x^n e^{-\frac{x^2}{2}} dx.$$

1. Déterminer I_0 et I_1 .
2. Montrer que, pour tout $n \in \mathbb{N}$, on a : $I_{n+2} = (n+1)I_n$.
3. Donner alors I_{2n+1} pour tout $n \in \mathbb{N}$. Pouvait-on prévoir ce résultat sans calculs ?
4. Déterminer I_{2n} pour tout $n \in \mathbb{N}$.
5. Soit X une variable aléatoire gaussienne de moyenne 1 et de variance unité. Déterminer $\mathbb{E}[X^4]$.

Corrigé

1. $I_0 = \sqrt{2\pi}$ puisqu'on reconnaît la densité d'une loi normale centrée réduite. Pour I_1 , on a :

$$I_1 = \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx = \left[-e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} = 0.$$

2. Pour tout $n \in \mathbb{N}$, on peut écrire :

$$I_{n+2} = \int_{-\infty}^{+\infty} x^{n+2} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{+\infty} (x^{n+1})(x e^{-\frac{x^2}{2}}) dx,$$

et on effectue une intégration par parties :

$$I_{n+2} = \left[-x^{n+1} e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} (n+1)x^n e^{-\frac{x^2}{2}} dx = (n+1)I_n,$$

la dernière égalité venant du fait que l'exponentielle l'emporte sur la puissance :

$$\lim_{x \rightarrow +\infty} x^{n+1} e^{-\frac{x^2}{2}} = \lim_{x \rightarrow -\infty} x^{n+1} e^{-\frac{x^2}{2}} = 0.$$

3. Puisque $I_1 = 0$, on en déduit que $I_3 = 0$, puis que $I_5 = 0$, et de proche en proche il est clair que $I_{2n+1} = 0$ pour tout $n \in \mathbb{N}$. Ce résultat était d'ailleurs clair sans calculs puisqu'on intègre une fonction impaire sur un domaine symétrique par rapport à 0.
4. Pour les indices pairs, on a $I_2 = 1 \times I_0 = \sqrt{2\pi}$, puis $I_4 = 3 \times I_2 = 3 \times 1 \times I_0 = 3\sqrt{2\pi}$, et de proche en proche :

$$I_{2n} = (2n-1) \times (2n-3) \times \cdots \times 3 \times 1 \times I_0 = \frac{(2n)!}{2^n n!} \sqrt{2\pi}.$$

5. Pour déterminer $\mathbb{E}[X^4]$, il y a deux méthodes équivalentes.
 - Méthode analytique : on écrit l'espérance sous forme d'intégrale :

$$\mathbb{E}[X^4] = \int_{-\infty}^{+\infty} \frac{x^4}{\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2}} dx,$$

et on effectue le changement de variable $u = x - 1$, ce qui donne :

$$\mathbb{E}[X^4] = \int_{-\infty}^{+\infty} \frac{(u+1)^4}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du.$$

On utilise la formule du binôme : $(u+1)^4 = u^4 + 4u^3 + 6u^2 + 4u + 1$, et on peut alors tout exprimer en fonction des I_n :

$$\mathbb{E}[X^4] = \frac{1}{\sqrt{2\pi}} (I_4 + 4I_3 + 6I_2 + 4I_1 + I_0) = 10.$$

- Méthode probabiliste : l'idée est la même, puisqu'on sait que si $X \sim \mathcal{N}(1, 1)$, alors $Y = X - 1 \sim \mathcal{N}(0, 1)$. Donc, par les calculs faits avant, on sait que $\mathbb{E}[Y] = \mathbb{E}[Y^3] = 0$, $\mathbb{E}[Y^2] = 1$ et $\mathbb{E}[Y^4] = 3$. Or on a : $\mathbb{E}[X^4] = \mathbb{E}[(Y + 1)^4] = \mathbb{E}[Y^4] + 4\mathbb{E}[Y^3] + 6\mathbb{E}[Y^2] + 4\mathbb{E}[Y] + 1 = 3 + 6 + 1 = 10$.

Exercice 2.2 (Queue de la gaussienne)

On appelle fonction de Marcum, ou queue de la gaussienne, la fonction définie pour tout réel x par :

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-\frac{t^2}{2}} dt.$$

1. Soit X une variable aléatoire qui suit une loi normale centrée réduite $\mathcal{N}(0, 1)$. Représenter la densité de X , puis $Q(x)$ sur ce même dessin. Soit F la fonction de répartition de X : donner la relation entre $F(x)$ et $Q(x)$.
2. Soit $x > 0$ fixé. Dans l'intégrale définissant $Q(x)$, effectuer le changement de variable $t = x + u$ et, tenant compte de $e^{-ux} \leq 1$, montrer qu'on a :

$$Q(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}}.$$

3. Pour $t \geq x > 0$, montrer que :

$$\frac{1 + \frac{1}{t^2}}{1 + \frac{1}{x^2}} \leq 1 \leq \frac{t}{x}.$$

4. En déduire que :

$$\frac{1}{\left(1 + \frac{1}{x^2}\right)\sqrt{2\pi}} \int_x^{+\infty} \left(1 + \frac{1}{t^2}\right) e^{-\frac{t^2}{2}} dt \leq Q(x) \leq \frac{1}{x\sqrt{2\pi}} \int_x^{+\infty} t e^{-\frac{t^2}{2}} dt.$$

5. Calculer la dérivée de $\frac{1}{t} e^{-\frac{t^2}{2}}$. En déduire que, pour tout $x > 0$, on a :

$$\frac{1}{\left(1 + \frac{1}{x^2}\right)x\sqrt{2\pi}} e^{-\frac{x^2}{2}} \leq Q(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

6. En déduire un équivalent de $Q(x)$ en $+\infty$.
7. Application : en communications numériques, pour une modulation binaire, les symboles transmis valent $\pm\sqrt{E_b}$, où E_b est appelée énergie moyenne par bit. Quand il transite par un canal à bruit gaussien, le signal reçu en sortie Y est égal à la somme du symbole d'entrée et d'une variable aléatoire indépendante $B \sim \mathcal{N}(0, \frac{N_0}{2})$, où N_0 est appelé puissance moyenne du bruit.
 - (a) Supposons que le symbole d'entrée soit $+\sqrt{E_b}$. Donner la loi de Y en fonction de E_b et N_0 . Même question si le symbole d'entrée est $-\sqrt{E_b}$.
 - (b) On reçoit $y \in \mathbb{R}$ en sortie de canal, mais on ignore ce qu'était le symbole d'entrée : quelle règle simple proposez-vous pour décider si en entrée le symbole émis était a priori équiprobablement $+\sqrt{E_b}$ ou $-\sqrt{E_b}$?
 - (c) Montrer que la probabilité d'erreur P_e faite avec cette règle de décision est :

$$P_e = Q\left(\sqrt{\frac{2E_b}{N_0}}\right).$$

La quantité $\frac{E_b}{N_0}$ est appelée rapport signal à bruit et intervient très souvent en communications numériques (on l'exprime usuellement en décibels).

Corrigé

1. Pour tout réel x , on a $F(x) = 1 - Q(x)$.
2. Soit $x > 0$ fixé. Le changement de variable $t = x + u$ et le fait que $e^{-ux} \leq 1$ pour x et u positifs donne

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{(x+u)^2}{2}} du = e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{-ux} e^{-\frac{u^2}{2}} du \leq e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{u^2}{2}} du$$

et on aura reconnu la densité de la gaussienne standard

$$\frac{1}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{u^2}{2}} du = \frac{1}{2} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du \right) = \frac{1}{2}$$

ce qui donne bien pour tout x positif

$$Q(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}}.$$

3. Pour $t \geq x > 0$, on a

$$1 + \frac{1}{t^2} \leq 1 + \frac{1}{x^2} \Rightarrow \frac{1 + \frac{1}{t^2}}{1 + \frac{1}{x^2}} \leq 1$$

L'inégalité de droite est encore plus évidente.

4. On en déduit alors

$$\frac{1}{(1 + \frac{1}{x^2})\sqrt{2\pi}} \int_x^{+\infty} \left(1 + \frac{1}{t^2}\right) e^{-\frac{t^2}{2}} dt \leq \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} 1 \times e^{-\frac{t^2}{2}} dt \leq \frac{1}{x\sqrt{2\pi}} \int_x^{+\infty} t e^{-\frac{t^2}{2}} dt$$

5. Pour tout réel non nul t

$$\left(\frac{1}{t} e^{-\frac{t^2}{2}}\right)' = -\left(1 + \frac{1}{t^2}\right) e^{-\frac{t^2}{2}}$$

Ainsi

$$\int_x^{+\infty} \left(1 + \frac{1}{t^2}\right) e^{-\frac{t^2}{2}} dt = \left[-\frac{1}{t} e^{-\frac{t^2}{2}}\right]_x^{+\infty} = \frac{1}{x} e^{-\frac{x^2}{2}}$$

et l'inégalité de gauche est acquise. Celle de droite est encore plus simple puisque

$$\int_x^{+\infty} t e^{-\frac{t^2}{2}} dt = \left[-e^{-\frac{t^2}{2}}\right]_x^{+\infty} = e^{-\frac{x^2}{2}}$$

Au total, on a bien montré que pour tout $x > 0$

$$\frac{1}{(1 + \frac{1}{x^2})x\sqrt{2\pi}} e^{-\frac{x^2}{2}} \leq Q(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

6. Cet encadrement permet de voir que

$$\frac{Q(x)}{\frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}} \xrightarrow{x \rightarrow +\infty} 1$$

d'où un équivalent très simple de $Q(x)$ lorsque x tend vers $+\infty$:

$$Q(x) \sim \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

7. Application

- (a) Si le symbole d'entrée est $+\sqrt{E_b}$ (respectivement $-\sqrt{E_b}$), alors $Y \sim \mathcal{N}(+\sqrt{E_b}, \frac{N_0}{2})$ (respectivement $Y \sim \mathcal{N}(-\sqrt{E_b}, \frac{N_0}{2})$). De façon générale, $Y = X + B$ où B est le bruit additif, supposé gaussien centré de variance $N_0/2$ et indépendant de X , variable aléatoire binaire correspondant au symbole d'entrée.
- (b) Intuitivement, on se dit que le symbole d'entrée était plus vraisemblablement $+\sqrt{E_b}$ (respectivement $-\sqrt{E_b}$) si la sortie y est positive (respectivement négative). Cette règle est en effet la bonne si les symboles d'entrée sont équiprobables, c'est-à-dire si $\mathbb{P}(X = +\sqrt{E_b}) = \mathbb{P}(X = -\sqrt{E_b}) = 1/2$. Il suffit de comparer les probabilités conditionnelles pour s'en convaincre. Il convient juste d'adapter la formule de Bayes et celle des probabilités totales au cas d'un cocktail entre loi discrète et loi à densité, ce qui donne ici :

$$\mathbb{P}(X = +\sqrt{E_b}|y) = \frac{f(y|X = +\sqrt{E_b})\mathbb{P}(X = +\sqrt{E_b})}{f(y)}$$

d'où

$$\mathbb{P}(X = +\sqrt{E_b}|y) = \frac{f(y|X = +\sqrt{E_b})\mathbb{P}(X = +\sqrt{E_b})}{f(y|X = +\sqrt{E_b})\mathbb{P}(X = +\sqrt{E_b}) + f(y|X = -\sqrt{E_b})\mathbb{P}(X = -\sqrt{E_b})}$$

Il reste à tenir compte du fait que les symboles d'entrée sont équiprobables et des densités respectives de la réponse Y connaissant X pour obtenir

$$\mathbb{P}(X = +\sqrt{E_b}|y) = \frac{e^{-\frac{(y-\sqrt{E_b})^2}{N_0}}}{e^{-\frac{(y-\sqrt{E_b})^2}{N_0}} + e^{-\frac{(y+\sqrt{E_b})^2}{N_0}}} = \frac{1}{1 + e^{-4\frac{\sqrt{E_b}}{N_0}y}} \quad (2.1)$$

On en déduit automatiquement :

$$\mathbb{P}(X = -\sqrt{E_b}|y) = 1 - \mathbb{P}(X = +\sqrt{E_b}|y) = \frac{e^{-4\frac{\sqrt{E_b}}{N_0}y}}{1 + e^{-4\frac{\sqrt{E_b}}{N_0}y}}$$

et par suite

$$\frac{\mathbb{P}(X = +\sqrt{E_b}|y)}{\mathbb{P}(X = -\sqrt{E_b}|y)} = e^{4\frac{\sqrt{E_b}}{N_0}y}$$

de sorte que ce rapport est supérieur à 1 si et seulement si y est positif, et la règle de décision au maximum de vraisemblance correspond bien à la règle intuitive donnée ci-dessus.

Remarque : si les symboles d'entrée ne sont pas équiprobables, il faut en tenir compte dans la règle de décision. Supposons par exemple que $\mathbb{P}(X = +\sqrt{E_b}) = 3/4$, alors l'équation (2.1) devient

$$\mathbb{P}(X = +\sqrt{E_b}|y) = \frac{3}{3 + e^{-4\frac{\sqrt{E_b}}{N_0}y}}$$

et

$$\frac{\mathbb{P}(X = +\sqrt{E_b}|y)}{\mathbb{P}(X = -\sqrt{E_b}|y)} = 3 e^{4\frac{\sqrt{E_b}}{N_0}y}$$

Ainsi on décide que le symbole d'entrée était $X = +\sqrt{E_b}$ si

$$3 e^{4\frac{\sqrt{E_b}}{N_0}y} > 1 \Leftrightarrow y > \tau = \frac{-\ln 3}{4} \times \frac{N_0}{\sqrt{E_b}}$$

Ces résultats admettent une interprétation graphique très simple : les points d'abscisses 0 et τ sont les points d'intersection des fonctions $y \mapsto f(y|X = +\sqrt{E_b})\mathbb{P}(X = +\sqrt{E_b})$ et $y \mapsto f(y|X = -\sqrt{E_b})\mathbb{P}(X = -\sqrt{E_b})$ respectivement lorsque $\mathbb{P}(X = +\sqrt{E_b}) = 1/2$ et $\mathbb{P}(X = +\sqrt{E_b}) = 3/4$ (voir figure 2.14 dans le cas où $N_0 = \sqrt{E_b} = 1$, d'où en particulier $\tau = -\ln 3/4 \approx -0.27$).

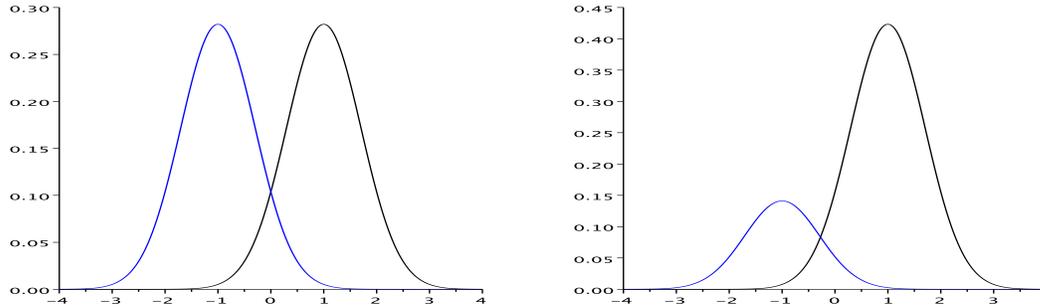


FIGURE 2.14 – Fonctions $y \mapsto f(y|X = +\sqrt{E_b})\mathbb{P}(X = +\sqrt{E_b})$ et $y \mapsto f(y|X = -\sqrt{E_b})\mathbb{P}(X = -\sqrt{E_b})$ lorsque $\mathbb{P}(X = +\sqrt{E_b}) = 1/2$ (à gauche) et $\mathbb{P}(X = +\sqrt{E_b}) = 3/4$ (à droite).

- (c) Dans le cas où les symboles d'entrée sont équiprobables, la probabilité d'erreur P_e est égale à la somme de la probabilité de décider $+\sqrt{E_b}$ alors que le symbole d'entrée était $-\sqrt{E_b}$ et vice-versa :

$$P_e = \mathbb{P}(Y > 0|X = -\sqrt{E_b})\mathbb{P}(X = -\sqrt{E_b}) + \mathbb{P}(Y < 0|X = +\sqrt{E_b})\mathbb{P}(X = +\sqrt{E_b})$$

et par symétrie des rôles, en notant toujours B le bruit additif :

$$P_e = \mathbb{P}(B > \sqrt{E_b}) = \mathbb{P}\left(\frac{B}{\sqrt{N_0/2}} > \frac{\sqrt{E_b}}{\sqrt{N_0/2}}\right) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right).$$

Exercice 2.3 (Entropie d'une variable aléatoire)

Si X est une variable aléatoire réelle admettant une densité f , on appelle entropie de X la quantité (si elle est définie) :

$$h(X) = \mathbb{E}[-\ln f(X)] = -\int_{-\infty}^{+\infty} f(x) \ln f(x) dx.$$

Grosso modo, l'entropie d'une variable aléatoire mesure le degré d'incertitude qu'on a sur l'issue d'un tirage de cette variable aléatoire.

- Supposons que $X \sim \mathcal{N}(0, 1)$, loi normale centrée réduite. Montrer qu'elle a pour entropie :

$$h(X) = \frac{1}{2}(1 + \ln(2\pi)).$$

- Supposons que $X \sim \mathcal{N}(m, \sigma^2)$. Montrer qu'elle a pour entropie : $h(X) = \frac{1}{2}(1 + \ln(2\pi\sigma^2))$. Ainsi, au moins pour les lois normales, l'entropie est d'autant plus grande que la variance est grande. On va montrer dans la suite que, parmi les variables aléatoires de variance donnée, celles qui ont la plus grande entropie sont celles qui suivent une loi normale.

- Soit donc $X_1 \sim \mathcal{N}(0, \sigma^2)$, dont la densité est notée φ , et X_2 une variable aléatoire centrée de densité f et de variance σ^2 , c'est-à-dire que :

$$\int_{-\infty}^{+\infty} x^2 f(x) dx = \sigma^2.$$

On suppose pour simplifier que f est strictement positive sur \mathbb{R} .

(a) Vérifier que (sous réserve d'existence des intégrales) :

$$h(X_2) = \int_{-\infty}^{+\infty} f(x) \ln \frac{\varphi(x)}{f(x)} dx - \int_{-\infty}^{+\infty} f(x) \ln \varphi(x) dx.$$

(b) Montrer que pour tout $x > 0$, $\log x \leq x - 1$. En déduire que :

$$\int_{-\infty}^{+\infty} f(x) \ln \frac{\varphi(x)}{f(x)} dx \leq 0.$$

(c) Montrer que :

$$- \int_{-\infty}^{+\infty} f(x) \ln \varphi(x) dx = \frac{1}{2}(1 + \ln(2\pi\sigma^2)).$$

(d) En déduire que $h(X_2) \leq h(X_1)$.

Corrigé

1. Si $X \sim \mathcal{N}(0, 1)$, alors son entropie s'écrit

$$h(X) = - \int_{-\infty}^{+\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \ln \left(\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \right) dx = \frac{\ln(2\pi)}{2} \int_{-\infty}^{+\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx + \frac{1}{2} \int_{-\infty}^{+\infty} x^2 \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

or

$$\int_{-\infty}^{+\infty} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = 1 \quad \text{et} \quad \int_{-\infty}^{+\infty} x^2 \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \mathbb{E}[X^2] = \text{Var}(X) = 1$$

d'où en effet

$$h(X) = \frac{1}{2}(1 + \ln(2\pi)).$$

2. Le même calcul que ci-dessus montre que si $X \sim \mathcal{N}(m, \sigma^2)$, alors elle a pour entropie : $h(X) = \frac{1}{2}(1 + \log(2\pi\sigma^2))$.

3. Soit donc $X_1 \sim \mathcal{N}(0, \sigma^2)$, dont la densité est notée φ , et X_2 une variable aléatoire centrée de densité f et de variance σ^2 , c'est-à-dire que :

$$\int_{-\infty}^{+\infty} x^2 f(x) dx = \sigma^2.$$

On suppose pour simplifier que f est strictement positive sur \mathbb{R} .

(a) Sous réserve d'existence des intégrales, par définition de l'entropie

$$h(X_2) = - \int_{-\infty}^{+\infty} f(x) \ln f(x) dx = \int_{-\infty}^{+\infty} f(x) \left(\ln \frac{\varphi(x)}{f(x)} - \ln \varphi(x) \right) dx$$

ce qui donne bien

$$h(X_2) = \int_{-\infty}^{+\infty} f(x) \ln \frac{\varphi(x)}{f(x)} dx - \int_{-\infty}^{+\infty} f(x) \ln \varphi(x) dx.$$

(b) Pour montrer que pour tout $x > 0$, $\ln x \leq x - 1$, il suffit par exemple d'étudier la fonction $g : x \mapsto x - 1 - \ln x$ sur $]0, +\infty[$. Sa dérivée est $g'(x) = 1 - 1/x$, qui est négative sur $]0, 1]$ et positive sur $[1, +\infty[$. Son minimum est donc $g(1) = 0$, autrement dit g est bien positive sur son domaine de définition. On en déduit que :

$$\int_{-\infty}^{+\infty} f(x) \log \frac{\varphi(x)}{f(x)} dx \leq \int_{-\infty}^{+\infty} f(x) \left(\frac{\varphi(x)}{f(x)} - 1 \right) dx = \int_{-\infty}^{+\infty} \varphi(x) dx - \int_{-\infty}^{+\infty} f(x) dx$$

or f et φ étant toutes deux des densités, elles intègrent à 1 et le majorant vaut bien 0.

(c) On a alors

$$-\int_{-\infty}^{+\infty} f(x) \ln \varphi(x) dx = -\int_{-\infty}^{+\infty} f(x) \ln \left(\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \right) dx$$

qui se calcule sans difficultés

$$-\int_{-\infty}^{+\infty} f(x) \ln \varphi(x) dx = \frac{\ln(2\pi)}{2} \int_{-\infty}^{+\infty} f(x) dx + \frac{1}{2} \int_{-\infty}^{+\infty} x^2 f(x) dx$$

en ayant en tête que

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad \text{et} \quad \int_{-\infty}^{+\infty} x^2 f(x) dx = \mathbb{E}[X_2^2] = \text{Var}(X_2) = \sigma^2$$

Au total on a bien

$$-\int_{-\infty}^{+\infty} f(x) \log \varphi(x) dx = \frac{1}{2}(1 + \ln(2\pi\sigma^2)).$$

(d) Des trois questions précédentes et du calcul de l'entropie pour une variable gaussienne $X_1 \sim \mathcal{N}(0, \sigma^2)$, on déduit que

$$h(X_2) \leq \frac{1}{2}(1 + \ln(2\pi\sigma^2)) = h(X_1),$$

c'est-à-dire que, à variance donnée, c'est la loi normale qui réalise le maximum de l'entropie.

Exercice 2.4 (Changement de base)

Soit $[X, Y]'$ un vecteur gaussien de moyenne $m = [1, 2]'$ et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

1. Vérifier que Γ est bien une matrice de covariance.
2. Diagonaliser Γ en base orthonormée, c'est-à-dire écrire une décomposition : $\Gamma = P'\Delta P$, avec P matrice orthogonale et Δ diagonale.
3. Représenter la fonction

$$f : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (u, v) & \mapsto f(u, v) = \frac{1}{4\pi\sqrt{2}} e^{-\frac{1}{2}(\frac{u^2}{4} + \frac{v^2}{2})} \end{cases}$$

4. En déduire la représentation de la densité du vecteur aléatoire $[X, Y]'$.

Exercice 2.5 (Problème de dégénérescence)

Soit $[X, Y]'$ un vecteur gaussien de moyenne $m = [0, 1]'$ et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}$$

1. Vérifier que Γ est bien une matrice de covariance.
2. Diagonaliser Γ .

3. Montrer que le vecteur $[X, Y]'$ ne prend ses valeurs que sur une droite de \mathbb{R}^2 , c'est-à-dire qu'il existe deux réels α et β tels que : $Y = \alpha X + \beta$ (on pourra chercher α et β tels que $\text{Var}(Y - (\alpha X + \beta)) = 0$). Comment le vecteur se distribue-t-il sur cette droite ?
4. Généralisation : soit $X = [X_1, \dots, X_d]'$ un vecteur aléatoire, non nécessairement gaussien, de moyenne m et de matrice de covariance Γ . Supposons Γ non inversible. Montrer que le vecteur aléatoire X ne prend ses valeurs que dans un sous-espace affine de \mathbb{R}^d . Quelle est, en fonction du spectre de Γ , la dimension de ce sous-espace affine ?
5. Exemple : montrer que le vecteur aléatoire X de moyenne $m = [7, 0, 1]'$ et de matrice de dispersion :

$$\Gamma = \begin{bmatrix} 7 & 2 & 5 \\ 2 & 4 & -2 \\ 5 & -2 & 7 \end{bmatrix}$$

appartient à un plan de \mathbb{R}^3 que l'on précisera.

Exercice 2.6 (Vecteur gaussien tridimensionnel)

Soit $V = [X, Y, Z]'$ un vecteur gaussien de moyenne $m = [1, 0, 3]'$ et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 4 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 6 \end{bmatrix}$$

1. Quelle est la fonction caractéristique de V ?
2. Quelles sont les lois marginales ?
3. Déterminer les lois des couples (X, Y) , (X, Z) et (Y, Z) .
4. Est-ce que deux composantes de V sont indépendantes ?
5. V admet-il une densité ? Si oui, la calculer.
6. Mêmes questions avec V de moyenne m comme ci-dessus et matrice de covariance :

$$\Gamma = \begin{bmatrix} 4 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Exercice 2.7 (Exercice de maniement)

Soit U_1 , U_2 et U_3 variables aléatoires indépendantes gaussiennes centrées de variances non nulles égales à v_1 , v_2 et v_3 . On considère le vecteur aléatoire $[X, Y]'$ défini par :

$$\begin{cases} X &= U_1 + U_2 \\ Y &= U_2 + U_3 \end{cases}$$

1. Quelles sont les lois de X et Y ?
2. Montrer qu'on peut trouver une combinaison linéaire non nulle de U_1 , U_2 et U_3 qui soit indépendante de X et de Y .
3. Donner le coefficient de corrélation linéaire ρ du couple (X, Y) , ainsi que la matrice de dispersion Γ .
4. On note σ_X^2 et σ_Y^2 les variances respectives de X et Y . Montrer que $[X, Y]'$ admet une densité et la préciser en fonction de ρ , σ_X et σ_Y .
5. Mêmes questions si on pose $Y = U_2 - U_3$.

Exercice 2.8 (Probabilité d'un quadrant)

On considère un vecteur gaussien centré $[X, Y]'$ de matrice de dispersion :

$$\Gamma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

On suppose dans un premier temps que $-1 < \rho < 1$.

1. Diagonaliser la matrice de dispersion sous la forme $\Gamma = P\Delta P'$. Préciser $\sqrt{\Delta^{-1}}$.
2. On veut calculer $\mathbb{P}(X \geq 0, Y \geq 0)$. Exprimer cette probabilité via une intégrale double.
3. Effectuer le changement de variables

$$\begin{bmatrix} u \\ v \end{bmatrix} = \sqrt{\Delta^{-1}}P' \begin{bmatrix} x \\ y \end{bmatrix},$$

puis un changement en coordonnées polaires pour montrer que :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \frac{1}{\pi} \arctan \sqrt{\frac{1+\rho}{1-\rho}}.$$

4. En déduire l'expression équivalente :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho.$$

5. Vérifier que cette formule est encore valable pour $\rho = \pm 1$.

Corrigé

1. On obtient $\Gamma = P\Delta P'$, avec :

$$P = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \Delta = \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}.$$

Une racine carrée de l'inverse de Δ est donc :

$$\sqrt{\Delta^{-1}} = \begin{bmatrix} \frac{1}{\sqrt{1+\rho}} & 0 \\ 0 & \frac{1}{\sqrt{1-\rho}} \end{bmatrix}.$$

2. La probabilité $\mathbb{P}(X \geq 0, Y \geq 0)$ est la probabilité que le point aléatoire (X, Y) tombe dans le quadrant $\mathbb{R}^+ \times \mathbb{R}^+$, c'est-à-dire :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \iint_{\mathbb{R}^+ \times \mathbb{R}^+} f(x, y) dx dy = \iint_{\mathbb{R}^+ \times \mathbb{R}^+} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}[x, y]\Gamma^{-1}[x, y]'} dx dy.$$

On peut encore l'écrire :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \iint_{\mathbb{R}^+ \times \mathbb{R}^+} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}} dx dy,$$

mais c'est la forme matricielle de la densité qui va nous servir dans la suite.

3. Le changement de variables $\varphi : (x, y) \mapsto (u, v)$ défini par :

$$\begin{bmatrix} u \\ v \end{bmatrix} = \sqrt{\Delta^{-1}}P' \begin{bmatrix} x \\ y \end{bmatrix} \Leftrightarrow \begin{cases} u = (x+y)/\sqrt{2(1+\rho)} \\ v = (x-y)/\sqrt{2(1-\rho)} \end{cases}$$

a pour réciproque :

$$\begin{cases} x &= (u\sqrt{1+\rho} + v\sqrt{1-\rho})/\sqrt{2} \\ y &= (u\sqrt{1+\rho} - v\sqrt{1-\rho})/\sqrt{2} \end{cases}$$

donc :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \iint_{\mathcal{D}} f(\varphi^{-1}(u, v)) |J_{\varphi^{-1}}(u, v)| \, du \, dv.$$

Or $[x, y]' = P\sqrt{\Delta}[u, v]'$ et par ailleurs $\Gamma^{-1} = P\Delta^{-1}P'$, donc :

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}[x, y]\Gamma^{-1}[x, y]'} = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}[u, v]\sqrt{\Delta}P'\Gamma^{-1}P\sqrt{\Delta}[u, v]'},$$

et le produit matriciel se simplifie, pour donner :

$$f(x, y) = f(\varphi^{-1}(u, v)) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{u^2+v^2}{2}}.$$

Par ailleurs, l'application φ^{-1} étant linéaire, son jacobien est tout simplement la matrice $J_{\varphi^{-1}}(u, v) = P\sqrt{\Delta}$, donc :

$$|J_{\varphi^{-1}}(u, v)| = |\det\sqrt{\Delta}| = \sqrt{1-\rho^2}.$$

Enfin, pour les nouvelles coordonnées u et v , le domaine d'intérêt s'écrit :

$$\mathcal{D} = \left\{ (u, v) \in \mathbb{R}^2 : u \geq 0, -u\sqrt{\frac{1+\rho}{1-\rho}} \leq v \leq u\sqrt{\frac{1+\rho}{1-\rho}} \right\}.$$

Finalement, on a obtenu :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \iint_{\mathcal{D}} \frac{1}{2\pi} e^{-\frac{u^2+v^2}{2}} \, du \, dv.$$

Le changement en coordonnées polaires est alors complètement naturel :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \iint_{\mathcal{D}'} \frac{1}{2\pi} e^{-\frac{r^2}{2}} r \, dr \, d\theta,$$

avec pour domaine d'intégration (voir aussi figure 2.15) :

$$\mathcal{D}' = \left\{ (r, \theta) : r \geq 0, -\arctan\sqrt{\frac{1+\rho}{1-\rho}} \leq \theta \leq \arctan\sqrt{\frac{1+\rho}{1-\rho}} \right\}.$$

Ceci donne finalement :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \frac{1}{\pi} \arctan\sqrt{\frac{1+\rho}{1-\rho}} \left[-e^{-\frac{r^2}{2}} \right]_0^{+\infty} = \frac{1}{\pi} \arctan\sqrt{\frac{1+\rho}{1-\rho}}.$$

4. Pour vérifier qu'on a de façon équivalente $\mathbb{P}(X \geq 0, Y \geq 0) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho$, il suffit de montrer que :

$$\forall \rho \in]-1, 1[\quad \frac{1}{\pi} \arctan\sqrt{\frac{1+\rho}{1-\rho}} = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho.$$

Ceci peut se faire en montrant tout d'abord que les deux fonctions ont même dérivée :

$$\left(\frac{1}{\pi} \arctan\sqrt{\frac{1+\rho}{1-\rho}} \right)' = \frac{1}{2\pi\sqrt{1-\rho^2}} = \left(\frac{1}{4} + \frac{1}{2\pi} \arcsin \rho \right)',$$

et en vérifiant ensuite que leurs valeurs en un point de l'intervalle $] -1, 1[$ coïncident, par exemple en 0 :

$$\frac{1}{\pi} \arctan\sqrt{\frac{1+0}{1-0}} = \frac{1}{4} = \frac{1}{4} + \frac{1}{2\pi} \arcsin 0.$$

5. Si $\rho = 1$, alors Γ n'est pas inversible et on a plus précisément :

$$\text{Var}(X - Y) = [1, -1]\Gamma \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0,$$

c'est-à-dire que $Y = X$ presque sûrement. Dans ce cas, puisqu'on sait de plus que X suit une loi normale centrée réduite, la probabilité cherchée est très simple :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \mathbb{P}(X \geq 0) = \frac{1}{2},$$

qui est bien égal à $\frac{1}{4} + \frac{1}{2\pi} \arcsin 1$. Si $\rho = -1$, alors on montre cette fois que $Y = -X$ presque sûrement, donc :

$$\mathbb{P}(X \geq 0, Y \geq 0) = \mathbb{P}(X \geq 0, X \leq 0) = \mathbb{P}(X = 0) = 0,$$

qui est bien égal à $\frac{1}{4} + \frac{1}{2\pi} \arcsin(-1)$. Donc la formule est encore vraie en ± 1 .

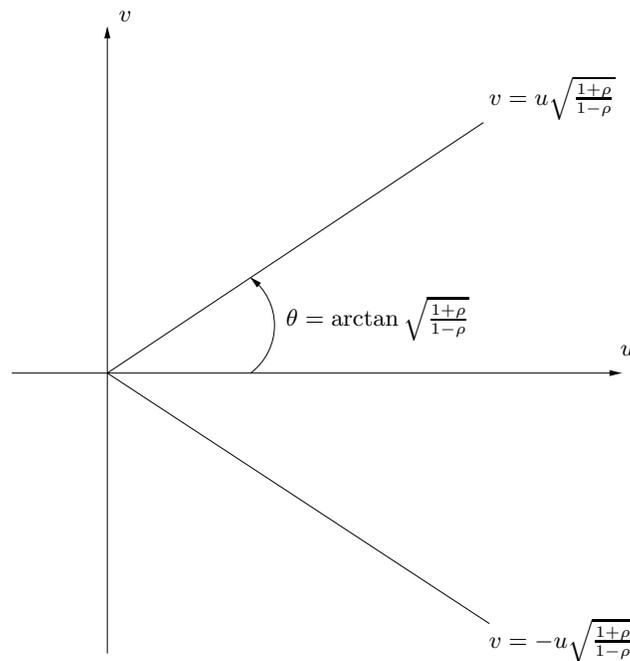


FIGURE 2.15 – Changement de variables en polaires.

Exercice 2.9 (Variations sur la sphère unité)

Soit $[X, Y, Z]'$ un vecteur gaussien centré de matrice de dispersion :

$$\Gamma = \begin{bmatrix} 1 & 0 & \sqrt{\frac{2}{3}} \\ 0 & 1 & \frac{1}{\sqrt{3}} \\ \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{3}} & 1 \end{bmatrix}$$

1. Pour tout triplet (α, β, γ) de \mathbb{R}^3 , déterminer la loi de la variable aléatoire $\alpha X + \beta Y + \gamma Z$.
2. Déterminer le noyau de Γ . Existe-t-il (α, β, γ) de $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$ tel que $\text{Var}(\alpha X + \beta Y + \gamma Z) = 0$?

3. La loi du vecteur aléatoire $[X, Y, Z]'$ admet-elle une densité ?
4. Entre quelles valeurs varie $\text{Var}(\alpha X + \beta Y + \gamma Z) = 0$ lorsque (α, β, γ) décrit la sphère unité, i.e. $\alpha^2 + \beta^2 + \gamma^2 = 1$? Indication : on pourra diagonaliser Γ .
5. On note $U = \frac{1}{\sqrt{3}}X - \sqrt{\frac{2}{3}}Y$ et $V = \frac{1}{\sqrt{3}}X + \frac{1}{\sqrt{6}}Y + \frac{1}{\sqrt{2}}Z$. Quelle est la loi du vecteur aléatoire $[U, V]'$?

Exercice 2.10 (Processus autorégressif)

Rappel : on peut montrer qu'une suite de variables aléatoires gaussiennes $X_n \sim \mathcal{N}(m_n, \sigma_n^2)$ converge en loi si et seulement si les suites (m_n) et (σ_n^2) sont toutes deux convergentes. Si m et σ^2 sont les limites respectives, alors la loi limite est la loi normale $\mathcal{N}(m, \sigma^2)$.

Soit $(\varepsilon_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d. de loi $\mathcal{N}(0, \sigma^2)$, avec $\sigma^2 > 0$ et θ un paramètre réel. On considère la suite $(X_n)_{n \geq 1}$ définie par :

$$\begin{cases} X_1 &= \varepsilon_1 \\ X_{n+1} &= \theta X_n + \varepsilon_{n+1} \end{cases}$$

1. Quelle est la loi de X_n ? Pour quelles valeurs de θ a-t-on convergence en loi ? Préciser alors la loi limite.
2. Calculer la covariance de X_n et X_{n+k} .
3. Quelle est la loi du vecteur aléatoire $[X_1, \dots, X_n]'$?
4. Exprimer la densité du vecteur $[\varepsilon_1, \dots, \varepsilon_n]'$. Grâce à la formule de changement de variable, en déduire celle du vecteur $[X_1, \dots, X_n]'$.
5. Déterminer l'estimateur du maximum de vraisemblance de (θ, σ^2) basé sur l'observation de l'échantillon (X_1, \dots, X_n) .

Corrigé

1. On a pour tout $n \geq 1$:

$$X_n = \varepsilon_n + \theta \varepsilon_{n-1} + \dots + \theta^{n-1} \varepsilon_1 = \sum_{k=1}^n \theta^{n-k} \varepsilon_k.$$

La variable X_n est la somme de $(n-1)$ variables gaussiennes indépendantes, donc elle suit une loi normale $\mathcal{N}(m_n, \sigma_n^2)$, avec bien sûr $m_n = 0$ et :

$$\sigma_n^2 = \sum_{k=1}^n \theta^{2(n-k)} \sigma^2 = \begin{cases} n\sigma^2 & \text{si } |\theta| = 1 \\ \frac{1-\theta^{2n}}{1-\theta^2} \sigma^2 & \text{si } |\theta| \neq 1 \end{cases}$$

D'après le rappel de l'énoncé, il y a donc convergence en loi de la suite (X_n) si et seulement si $|\theta| < 1$, auquel cas on a :

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{1-\theta^2}\right).$$

2. On a pour tout $k \geq 0$:

$$X_{n+k} = \varepsilon_{n+k} + \theta \varepsilon_{n+k-1} + \dots + \theta^{k-1} \varepsilon_{n+1} + \theta^k X_n = W_k + \theta^k X_n,$$

avec W_k indépendante de X_n . On en déduit la covariance entre X_n et X_{n+k} :

$$\text{Cov}(X_n, X_{n+k}) = \text{Cov}(X_n, W_k + \theta^k X_n) = \theta^k \text{Var}(X_n) = \begin{cases} n\theta^k \sigma^2 & \text{si } |\theta| = 1 \\ \frac{1-\theta^{2n}}{1-\theta^2} \theta^k \sigma^2 & \text{si } |\theta| \neq 1 \end{cases}$$

3. On peut écrire :

$$\begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} = A \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix},$$

avec A matrice triangulaire inférieure formée de 1 sur la diagonale, de θ sur la première sous-diagonale, de θ^2 sur la deuxième sous-diagonale, ..., le coefficient $A_{n,1}$ valant θ^{n-1} .

Puisque $[\varepsilon_1, \dots, \varepsilon_n]'$ est un vecteur gaussien, $[X_1, \dots, X_n]'$ l'est aussi. La matrice de dispersion de $[\varepsilon_1, \dots, \varepsilon_n]'$ étant égale à $\sigma^2 I_n$, on a :

$$\begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} \sim \mathcal{N}(0, \sigma^2 AA').$$

4. Puisque ses composantes sont indépendantes, le vecteur $[\varepsilon_1, \dots, \varepsilon_n]'$ a pour densité :

$$g(\varepsilon_1, \dots, \varepsilon_n) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(\varepsilon_1^2 + \dots + \varepsilon_n^2)}$$

et on peut appliquer la formule de changement de variable à la bijection :

$$\begin{cases} \varepsilon_1 = x_1 \\ \varepsilon_2 = x_2 - \theta x_1 \\ \dots = \dots \\ \varepsilon_n = x_n - \theta x_{n-1} \end{cases}$$

La densité du vecteur $[X_1, \dots, X_n]'$ est donc :

$$f(x_1, \dots, x_n) = g(x_1, x_2 - \theta x_1, \dots, x_n - \theta x_{n-1}) |\det(A^{-1})|.$$

Mais puisque $\det A = 1$, il en va de même pour A^{-1} et :

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}(x_1^2 + (x_2 - \theta x_1)^2 + \dots + (x_n - \theta x_{n-1})^2)}.$$

5. On suppose donc observer un échantillon (x_1, \dots, x_n) du processus (X_1, \dots, X_n) et on veut en déduire une estimation au maximum de vraisemblance du couple (θ, σ^2) . On admet que cette solution existe et est unique. Il suffit alors pour la trouver d'écrire la log-vraisemblance du couple (θ, σ^2) :

$$L(\theta, \sigma^2) = \ln f(x_1, \dots, x_n) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (x_1^2 + (x_2 - \theta x_1)^2 + \dots + (x_n - \theta x_{n-1})^2),$$

dont on cherche le point critique. On obtient pour dérivées partielles :

$$\begin{cases} \frac{\partial L}{\partial \theta} = \frac{1}{\sigma^2} \sum_{k=1}^{n-1} x_k (x_{k+1} - \theta x_k) \\ \frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (x_1^2 + (x_2 - \theta x_1)^2 + \dots + (x_n - \theta x_{n-1})^2) \end{cases}$$

L'annulation de la première dérivée partielle donne l'estimateur de θ au maximum de vraisemblance :

$$\hat{\theta} = \frac{\sum_{k=1}^{n-1} x_k x_{k+1}}{\sum_{k=1}^{n-1} x_k^2}.$$

D'où l'on déduit celui de σ^2 , via l'annulation de la seconde dérivée partielle :

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{k=0}^{n-1} (x_{k+1} - \hat{\theta} x_k)^2,$$

avec la convention $x_0 = 0$.

Exercice 2.11 (Moyenne mobile)

Soit $(\varepsilon_n)_{n \geq 1}$ une suite de variables aléatoires i.i.d. de loi $\mathcal{N}(0, \sigma^2)$, et θ un paramètre réel. On considère la suite $(X_n)_{n \geq 1}$ définie par

$$\begin{cases} X_1 &= \varepsilon_1 \\ X_{n+1} &= \theta \varepsilon_n + \varepsilon_{n+1} \end{cases}$$

1. Soit $X = [X_1, \dots, X_n]'$. Pourquoi X est-il un vecteur gaussien ?
2. Déterminer la moyenne et la matrice de dispersion de X .
3. Justifier le fait que X admet une densité et la préciser.

Exercice 2.12 (Sauts de puce)

Dans le plan, une puce part de l'origine $O = A_0$ et saute à chaque instant. De l'instant 1 à l'instant n , ses points de chute successifs sont notés A_1, \dots, A_n . Chaque saut a la même amplitude unité : $A_j A_{j+1} = 1$ pour tout indice j . La direction de chaque saut suit une loi uniforme sur $[0, 2\pi]$.

1. Soit $V = [\cos U, \sin U]'$ un vecteur aléatoire, avec U qui suit une loi uniforme sur $[0, 2\pi]$. Déterminer espérance et matrice de covariance de V .
2. Ecrire $S_n = \overrightarrow{OA_n}$ comme une somme de vecteurs aléatoires indépendants et identiquement distribués. Que dire de la convergence en loi de la suite (S_n/\sqrt{n}) ?
3. Soit maintenant $[X, Y]'$ un vecteur gaussien centré de matrice de covariance identité. Considérons le vecteur aléatoire $[R, \Theta]'$ correspondant en coordonnées polaires. Déterminer sa densité jointe grâce à la formule de changement de variables. En déduire que R et Θ sont indépendantes.
4. En calculant par exemple sa fonction de répartition, montrer que la variable aléatoire R^2 , carré de la distance à l'origine, suit une loi exponentielle : $R^2 \sim \mathcal{E}(\frac{1}{2})$.
5. Pour notre puce, en déduire qu'au bout de n sauts le carré de sa distance à l'origine suit "à peu près" une loi exponentielle de moyenne n .

Corrigé

Des exemples de trajectoires de la puce sont donnés figure 2.16.

1. Pour le vecteur moyenne, on a :

$$\mathbb{E}[\cos U] = \frac{1}{2\pi} \int_0^{2\pi} \cos u \, du = \frac{1}{2\pi} [\sin u]_0^{2\pi} = 0.$$

Et on trouve de la même façon $\mathbb{E}[\sin U] = 0$. Ainsi V est centré :

$$\mathbb{E}[V] = \begin{bmatrix} \mathbb{E}[\cos U] \\ \mathbb{E}[\sin U] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

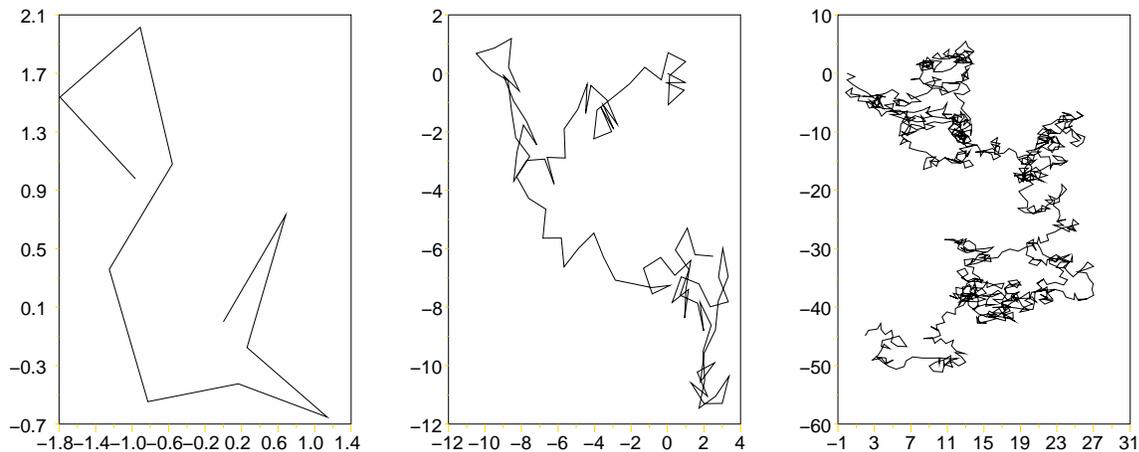


FIGURE 2.16 – Exemples de trajectoires de la puce pour 10, 100 et 1000 sauts.

ce qui est évident puisque la puce ne privilégie aucune direction dans ses sauts. Pour la matrice de dispersion, on a :

$$\text{Var}(\cos U) = \mathbb{E}[\cos^2 U] = \frac{1}{2\pi} \int_0^{2\pi} \cos^2 u \, du = \frac{1}{4\pi} \left[u + \frac{1}{2} \sin(2u) \right]_0^{2\pi} = \frac{1}{2}.$$

De même :

$$\text{Var}(\sin U) = \mathbb{E}[\sin^2 U] = \mathbb{E}[1 - \cos^2 U] = 1 - \mathbb{E}[\cos^2 U] = 1 - \text{Var}(\cos U) = \frac{1}{2}.$$

Pour la covariance entre abscisse et ordonnée :

$$\text{Cov}(\cos U, \sin U) = \mathbb{E}[\cos U \sin U] = \frac{1}{2\pi} \int_0^{2\pi} \cos u \sin u \, du = -\frac{1}{4} [\cos 2u]_0^{2\pi} = 0,$$

c'est-à-dire qu'abscisse et ordonnée de V sont décorrélées (mais clairement pas indépendantes, puisque la valeur de l'abscisse ne laisse plus le choix qu'entre deux valeurs possibles pour l'ordonnée). Ainsi le vecteur aléatoire centré V a pour matrice de covariance :

$$\Gamma = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}.$$

2. La relation de Chasles donne la décomposition :

$$S_n = \overrightarrow{OA_n} = \overrightarrow{OA_1} + \overrightarrow{A_1A_2} + \cdots + \overrightarrow{A_{n-1}A_n} = V_1 + \cdots + V_n,$$

avec les V_k i.i.d. de même loi que le vecteur V de la question précédente. Le théorème central limite nous apprend donc que :

$$\frac{S_n - n\mathbb{E}[V]}{\sqrt{n}} = \frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\right).$$

3. Le changement de variables en polaires s'écrit :

$$\begin{cases} x &= r \cos \theta \\ y &= r \sin \theta \end{cases}$$

Puisque le vecteur $[X, Y]'$ a pour densité :

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}},$$

on en déduit que $[R, \Theta]'$ a pour densité :

$$g(r, \theta) = \frac{r}{2\pi} e^{-\frac{r^2}{2}} \mathbb{1}_{[0, +\infty[}(r) \mathbb{1}_{[0, 2\pi]}(\theta).$$

La densité de R est alors :

$$g(r) = \int_0^{2\pi} g(r, \theta) d\theta = r e^{-\frac{r^2}{2}} \mathbb{1}_{[0, +\infty[}(r).$$

Quant à celle de Θ , on obtient :

$$g(\theta) = \int_0^{+\infty} g(r, \theta) dr = \frac{1}{2\pi} \mathbb{1}_{[0, 2\pi]}(\theta) \left[e^{-\frac{r^2}{2}} \right]_0^{+\infty} = \frac{1}{2\pi} \mathbb{1}_{[0, 2\pi]}(\theta),$$

c'est-à-dire que Θ suit une loi uniforme sur $[0, 2\pi]$. Ainsi on a pour tout couple $(r, \theta) \in [0, +\infty[\times [0, 2\pi]$:

$$g(r, \theta) = g(r) \times g(\theta),$$

donc les variables aléatoires R et Θ sont indépendantes.

4. Soit F la fonction de répartition de R^2 . Bien sûr $F(t) = 0$ si $t < 0$, et pour tout $t \geq 0$ on a :

$$F(t) = \mathbb{P}(R^2 \leq t) = \mathbb{P}(R \leq \sqrt{t}) = \int_0^{\sqrt{t}} g(r) dr = \left[e^{-\frac{r^2}{2}} \right]_0^{\sqrt{t}} = 1 - e^{-\frac{t}{2}},$$

où l'on reconnaît la fonction de répartition d'une loi exponentielle : $R^2 \sim \mathcal{E}(\frac{1}{2})$.

5. Notons X_n et Y_n les coordonnées aléatoires du point terminal A_n . Le carré de la distance à l'origine de la puce au bout de n sauts est donc $OA_n^2 = X_n^2 + Y_n^2$, avec $\overrightarrow{OA_n} = S_n$ qui suit à peu près une loi normale $\mathcal{N}(0, n\Gamma)$. C'est encore dire que, à peu de choses près :

$$\sqrt{\frac{2}{n}} S_n = \sqrt{\frac{2}{n}} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

Mais alors d'après la question précédente on a, toujours à peu de choses près :

$$\frac{2}{n} (X_n^2 + Y_n^2) \sim \mathcal{E}(\frac{1}{2}),$$

c'est-à-dire que le carré de la distance à l'origine suit à peu près une loi exponentielle de paramètre $\frac{1}{n}$.

Exercice 2.13 (Vecteur gaussien bidimensionnel)

On considère un vecteur gaussien $[X, Y]'$ de moyenne $m = [1, -1]'$ et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$$

1. Ecrire la densité du vecteur $[X, Y]'$.
2. Quelle est la loi de X ? de Y ? de $X + Y$?
3. Déterminer l'espérance conditionnelle $\mathbb{E}[X|Y]$. Quelle est sa loi ?

Corrigé

Cf. sujet de juin 2005 en annexe.

Exercice 2.14 (Conditionnement d'une variable par un couple)

Soit $[X, Y, Z]'$ un vecteur gaussien centré de matrice de covariance :

$$\Gamma = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 4 \end{bmatrix}$$

1. Calculer $\mathbb{E}[X|Y, Z]$, l'espérance conditionnelle de X sachant le couple (Y, Z) .
2. Vérifier sur cet exemple que $X - \mathbb{E}[X|Y, Z]$ est indépendant de (Y, Z) .
3. Calculer $\mathbb{E}[(X - \mathbb{E}[X|Y, Z])^2|Y, Z]$.
4. Quelle est la loi de X sachant (Y, Z) ?

Corrigé

1. La formule de conditionnement pour un vecteur gaussien centré donne :

$$\mathbb{E}[X|Y, Z] = \Gamma_{X,(Y,Z)} \Gamma_{(Y,Z)}^{-1} \begin{bmatrix} Y \\ Z \end{bmatrix} = [1, 2] \begin{bmatrix} 9 & -3 \\ -3 & 4 \end{bmatrix}^{-1} \begin{bmatrix} Y \\ Z \end{bmatrix},$$

c'est-à-dire :

$$\mathbb{E}[X|Y, Z] = \frac{1}{27}(10Y + 21Z).$$

2. On a donc $X - \mathbb{E}[X|Y, Z] = X - \frac{1}{27}(10Y + 21Z)$. Le vecteur $[X - \mathbb{E}[X|Y, Z], Y, Z]'$ est gaussien car obtenu par transformation linéaire d'un vecteur gaussien. Pour montrer que sa première composante est indépendante des deux dernières, il suffit donc de vérifier que les covariances sont nulles :

$$\text{Cov}(X - \mathbb{E}[X|Y, Z], Y) = \mathbb{E}[(X - \mathbb{E}[X|Y, Z])Y] = \mathbb{E} \left[\left(X - \frac{1}{27}(10Y + 21Z) \right) Y \right],$$

c'est-à-dire en développant :

$$\text{Cov}(X - \mathbb{E}[X|Y, Z], Y) = \mathbb{E}[XY] - \frac{10}{27}\mathbb{E}[Y^2] - \frac{7}{9}\mathbb{E}[YZ] = 1 - \frac{10}{27} \times 9 - \frac{7}{9} \times (-3) = 0.$$

On vérifie de la même façon que $\text{Cov}(X - \mathbb{E}[X|Y, Z], Z) = 0$.

3. Première méthode (brutale, pour ceux qui ne connaissent pas leur cours) : puisque la variable aléatoire $(X - \mathbb{E}[X|Y, Z])$ est indépendante du couple (Y, Z) , il en va de même pour la variable aléatoire $(X - \mathbb{E}[X|Y, Z])^2$. De ce fait, on a :

$$\mathbb{E}[(X - \mathbb{E}[X|Y, Z])^2|Y, Z] = \mathbb{E}[(X - \mathbb{E}[X|Y, Z])^2] = \mathbb{E} \left[\left(X - \frac{1}{27}(10Y + 21Z) \right)^2 \right],$$

et il suffit alors de développer le carré :

$$\mathbb{E}[(X - \mathbb{E}[X|Y, Z])^2|Y, Z] = \mathbb{E}[X^2] + \frac{100}{27^2}\mathbb{E}[Y^2] + \frac{49}{81}\mathbb{E}[Z^2] - \frac{20}{27}\mathbb{E}[XY] + \frac{420}{27^2}\mathbb{E}[YZ] - \frac{42}{27}\mathbb{E}[XZ],$$

ce qui donne au total :

$$\mathbb{E}[(X - \mathbb{E}[X|Y, Z])^2 | Y, Z] = \frac{56}{27}.$$

Seconde méthode (érudite, pour ceux qui connaissent leur cours) : la quantité cherchée est exactement la variance résiduelle, ou résidu quadratique, ou résidu, donné par la formule :

$$\sigma^2 = \Gamma_X - \Gamma_{X,(Y,Z)} \Gamma_{(Y,Z)}^{-1} \Gamma_{(Y,Z),X} = 4 - [1, 2] \begin{bmatrix} 9 & -3 \\ -3 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{56}{27}.$$

Etonnov, niet ?

4. On sait qu'on a la décomposition en somme de variables indépendantes :

$$X = \mathbb{E}[X|Y, Z] + (X - \mathbb{E}[X|Y, Z]) = \mathbb{E}[X|Y, Z] + T = \frac{1}{27}(10Y + 21Z) + T,$$

avec T indépendante du couple (Y, Z) et de loi $\mathcal{N}(0, \sigma^2)$. Donc sachant $Y = y$ et $Z = z$, on a simplement $X = \frac{1}{27}(10y + 21z) + T$, c'est-à-dire que :

$$X \sim \mathcal{N}\left(\frac{1}{27}(10y + 21z), \frac{56}{27}\right).$$

Exercice 2.15 (Mai 2007)

Soit $V = [X, Y]'$ un vecteur gaussien centré de matrice de covariance :

$$\Gamma = \begin{bmatrix} \sigma_X^2 & r\sigma_X\sigma_Y \\ r\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

- Rappeler à quelle condition V admet une densité $f(x, y)$ et dans ce cas donner sa formule en fonction de σ_X , σ_Y et r . On suppose cette condition vérifiée dans toute la suite.
- Donner l'expression de l'espérance conditionnelle $\mathbb{E}[Y|X]$ en fonction de σ_X , σ_Y et r .
- On considère un vecteur $V = [X, Y]'$ ayant pour densité :

$$\forall (x, y) \in \mathbb{R}^2 \quad f(x, y) = \frac{1}{4\pi} e^{-\frac{x^2}{8} - \frac{5y^2}{8} + \frac{xy}{4}}$$

- Déterminer sa matrice de covariance Γ .
- Sachant $x = 1$, quelle est la loi de Y ?
- Sachant $x = 1$, en déduire un intervalle dans lequel se situe Y avec 95% de chances.

Corrigé

Cf. corrigé du sujet de mai 2007 en annexe.

Exercice 2.16 (Covariances identiques)

Soit $[X, Y, Z]'$ un vecteur gaussien centré tel que les variances soient égales à 1 et les covariances égales à $1/2$.

- Calculer $\mathbb{E}[Z|X, Y]$.
- Calculer $\mathbb{E}[Z|X + Y]$. Faire un dessin pour expliquer le lien entre ce résultat et le résultat précédent.
- Sachant $X = x$, quelle est la loi de Y ?
- En déduire $\mathbb{E}[Y^2|X = x]$, puis $\mathbb{E}[Y^2|X]$.
- Grâce à la méthode de conditionnement, calculer alors $\mathbb{E}[X^2 Y^2]$.

6. Déterminer de même $\mathbb{E}[X^2Y^4]$.
7. Généralisation : on suppose que les covariances valent $a \in]-1/2, 1[$. Vérifier qu'on définit bien ainsi une matrice de covariance. Que deviennent les résultats précédents ?

Corrigé

1. Pour calculer $\mathbb{E}[Z|X, Y]$, on applique la formule de conditionnement :

$$\mathbb{E}[Z|X, Y] = \Gamma_{Z,(X,Y)} \Gamma_{(X,Y)}^{-1} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \frac{1}{3}(X + Y).$$

2. Puisque le couple $(T, Z) = (X + Y, Z)$ est gaussien, on peut calculer $\mathbb{E}[Z|X + Y]$ grâce à la formule de régression classique :

$$\mathbb{E}[Z|X + Y] = \mathbb{E}[Z|T] = \frac{\text{Cov}(Z, T)}{\text{Var}(T)} T.$$

Or :

$$\text{Cov}(Z, T) = \text{Cov}(Z, X + Y) = \mathbb{E}[Z(X + Y)] = \mathbb{E}[ZX] + \mathbb{E}[ZY] = 1,$$

et :

$$\text{Var}(T) = \text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y) = 3.$$

Finalement :

$$\mathbb{E}[Z|X + Y] = \frac{1}{3}(X + Y) = \mathbb{E}[Z|X, Y].$$

On pouvait s'en douter sans calculs : la valeur moyenne prise par la variable aléatoire Z lorsqu'on connaît à la fois X et Y ne dépend que de la somme $(X + Y)$. De ce fait, la valeur moyenne prise par la variable aléatoire Z lorsqu'on connaît simplement la somme $(X + Y)$ va être égale à celle trouvée précédemment. Ceci vient du fait que Z a la même covariance avec X qu'avec Y . Faire un dessin pour expliquer le lien entre ce résultat et le résultat précédent.

3. On a la décomposition orthogonale :

$$Y = \mathbb{E}[Y|X] + (Y - \mathbb{E}[Y|X]) = \mathbb{E}[Y|X] + W,$$

avec :

$$\mathbb{E}[Y|X] = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} X = \frac{X}{2},$$

et la variable résiduelle W qui est indépendante de X qui suit une loi $\mathcal{N}(0, \sigma^2)$, avec le résidu :

$$\sigma^2 = \text{Var}(Y) - \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)} = \frac{3}{4}.$$

Ainsi, lorsque $X = x$, on a :

$$Y \sim \mathcal{N}\left(\frac{x}{2}, \frac{3}{4}\right).$$

4. Sachant $X = x$, on a vu que Y suit une loi normale. Or, de façon générale, lorsque $V \sim \mathcal{N}(m, \sigma^2)$, on a : $\mathbb{E}[V^2] = m^2 + \sigma^2$, donc ici :

$$\mathbb{E}[Y^2|X = x] = \frac{x^2 + 3}{4},$$

et plus généralement : $\mathbb{E}[Y^2|X] = \frac{X^2 + 3}{4}$.

5. Par les propriétés classiques de l'espérance conditionnelle, on a alors :

$$\mathbb{E}[X^2Y^2] = \mathbb{E}[\mathbb{E}[X^2Y^2|X]] = \mathbb{E}[X^2\mathbb{E}[Y^2|X]] = \frac{1}{4}\mathbb{E}[X^4 + 3X^2].$$

Or X suit une loi normale centrée réduite, loi dont on connaît tous les moments, en particulier $\mathbb{E}[X^4] = 3$. Ainsi $\mathbb{E}[X^2Y^2] = \frac{3}{2}$.

6. On applique la même technique, mais en inversant les rôles joués par X et Y : ceci évite d'avoir à calculer $\mathbb{E}[Y^4|X]$. Puisque X et Y sont interchangeable, il est clair que $\mathbb{E}[X^2|Y] = \frac{Y^2+3}{4}$. Ainsi :

$$\mathbb{E}[X^2Y^4] = \mathbb{E}[\mathbb{E}[X^2Y^4|Y]] = \mathbb{E}\left[Y^4\left(\frac{Y^2+3}{4}\right)\right] = \frac{1}{4}\mathbb{E}[Y^6 + 3Y^4].$$

Or le moment d'ordre 6 d'une loi normale centrée réduite vaut 15, donc $\mathbb{E}[X^2Y^4] = 6$.

7. Généralisation : on suppose que les covariances valent $a \in]-1/2, 1[$. Vérifier qu'on définit bien ainsi une matrice de covariance. Que deviennent les résultats précédents ? Ils se généralisent...

Exercice 2.17 (Vecteur gaussien tridimensionnel)

On considère un vecteur gaussien $[X, Y, Z]'$ centré et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

1. Donner l'espérance conditionnelle $\mathbb{E}[X|Y, Z]$.
2. Quelle est la loi de Z sachant $X = 1$ et $Y = 2$?
3. Déterminer l'espérance conditionnelle du vecteur $[X, Y]'$ sachant Z , notée $\mathbb{E}[X, Y|Z]$.
4. Quelle est la loi du vecteur $[X, Y]'$ sachant $Z = 1$?
5. On pose alors :

$$\begin{cases} U = -X + Y + Z \\ V = X - Y + Z \\ W = X + Y - Z \end{cases}$$

Pourquoi le vecteur aléatoire $[U, V, W]'$ est-il gaussien ?

6. Quelles sont les lois de U , de V , de W ?
7. Déterminer la moyenne et la matrice de covariance de $[U, V, W]'$. Qu'en déduire sur la corrélation de U , V et W ?

Corrigé

Le corrigé de cet exercice est donné en annexe, sujet de juin 2005.

Exercice 2.18 (Variance conditionnelle)

Soit $[X_1, X_2]'$ un vecteur gaussien centré de matrice de covariance :

$$\Gamma = \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}$$

1. Déterminer l'espérance conditionnelle $\mathbb{E}[X_1|X_2]$.

2. Si $[X, Y]'$ est un vecteur aléatoire (non nécessairement gaussien), avec Y de carré intégrable, on définit la variance conditionnelle de Y sachant X , notée $\text{Var}[Y|X]$, comme la variable aléatoire :

$$\text{Var}[Y|X] = \mathbb{E}[(Y - \mathbb{E}[Y|X])^2|X].$$

Montrer que dans notre situation, on a :

$$\text{Var}[X_1|X_2] = \mathbb{E}[X_1^2] - \mathbb{E}[(\mathbb{E}[X_1|X_2])^2].$$

3. En déduire la variance conditionnelle $\text{Var}[X_1|X_2]$.
 4. Vérifier que : $\text{Var}[X_1|X_2 = 0] \neq \mathbb{E}[(X_1 - \mathbb{E}[X_1|X_2 = 0])^2]$.

Exercice 2.19 (Un été 2006)

On considère un vecteur gaussien $V = [X, Y, Z]^T$ de moyenne $m = [1, -1, 1]'$ et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

1. Quelle est la loi du vecteur $[X, Y]'$? Donner sa densité.
2. Déterminer l'équation de la droite de régression de Y en X . Préciser le résidu quadratique.
3. Donner une variable aléatoire $U = \alpha X + \beta Y + \gamma$ qui soit indépendante de X .
4. Quelle est la fonction caractéristique du vecteur V ?
5. Donner l'expression de $\mathbb{E}[Z|X, Y]$, espérance conditionnelle de Z sachant (X, Y) ?
6. En déduire une variable aléatoire $T = aX + bY + cZ + d$ qui soit indépendante de X et Y . Préciser la variance de T .
7. On observe $Y = 1$ et $Z = 2$. Quelle est la loi de la variable aléatoire X sachant ces données?

Corrigé

Cf. sujet de juin 2006 en annexe.

Exercice 2.20 (Statistique bayésienne)

Soit (X, M) un couple aléatoire dont la loi jointe est telle que : M suit une loi normale $\mathcal{N}(0, \tau^2)$. Sachant $M = m$, X suit une loi normale $\mathcal{N}(m, \sigma^2)$.

1. Calculer $\mathbb{E}[X|M]$.
2. Montrer que la densité conditionnelle de M sachant $X = x$ peut s'écrire :

$$f(m|x) = c_x \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \frac{e^{-\frac{m^2}{2\theta^2}}}{\sqrt{2\pi\theta^2}},$$

où c_x est une constante ne dépendant que de x .

3. En déduire que, sachant $X = x$, $M \sim \mathcal{N}(\frac{\sigma_0^2}{\sigma^2}x, \sigma_0^2)$, avec $\sigma_0^2 = \frac{\sigma^2\theta^2}{\sigma^2 + \theta^2}$.
4. Comparer σ_0^2 à σ^2 . Calculer $\mathbb{E}[M|X]$.
5. Par la même méthode que ci-dessus, montrer que $X \sim \mathcal{N}(0, \sigma^2 + \theta^2)$.
6. Via le calcul de la densité jointe, montrer que (X, M) est un vecteur aléatoire gaussien, dont on précisera moyenne et matrice de dispersion.

Corrigé succinct

1. D'après l'énoncé, on a : $\mathbb{E}[X|M = m] = m$, donc de façon générale $\mathbb{E}[X|M] = M$.
2. On cherche à déterminer la densité conditionnelle $f(m|x)$. La formule de Bayes donne :

$$f(m|x) = \frac{f(m, x)}{f(x)} = \frac{f(x|m)f(m)}{f(x)}.$$

Or x est fixé, donc on peut considérer $f(x)$ comme une constante c_x . La connaissance des densités du numérateur permet d'écrire :

$$f(m|x) = c_x \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \frac{e^{-\frac{m^2}{2\theta^2}}}{\sqrt{2\pi\theta^2}}.$$

3. Après calculs, on trouve que ceci est la densité d'une variable aléatoire gaussienne $\mathcal{N}(\frac{\sigma_0^2}{\sigma^2}x, \sigma_0^2)$, avec :

$$\sigma_0^2 = \frac{\sigma^2\theta^2}{\sigma^2 + \theta^2} \leq \sigma^2.$$

4. On a :

$$\sigma_0^2 = \frac{\sigma^2}{1 + \frac{\theta^2}{\sigma^2}} \leq \sigma^2.$$

Pour tout x , on a donc : $\mathbb{E}[M|X = x] = \frac{\sigma_0^2}{\sigma^2}x$. On en déduit que :

$$\mathbb{E}[M|X] = \frac{\sigma_0^2}{\sigma^2}X.$$

5. On cherche la densité f de X . Le même type de calcul que précédemment montre que X suit une loi gaussienne $\mathcal{N}(0, \sigma^2 + \theta^2)$.
6. On cherche la densité $f_{X,M}$ du couple (X, M) . On trouve que c'est une densité gaussienne de moyenne $(0, 0)$ et de matrice de covariance Γ , avec :

$$\Gamma = \begin{bmatrix} \sigma^2 + \theta^2 & \theta^2 \\ \theta^2 & \theta^2 \end{bmatrix}.$$

Exercice 2.21 (Espérance conditionnelle et indépendance)

On considère trois variables aléatoires U , V et W indépendantes et suivant la même loi normale centrée réduite $\mathcal{N}(0, 1)$.

1. On définit le vecteur aléatoire $[X_1, X_2, X_3]'$ comme suit :

$$\begin{cases} X_1 = U \\ X_2 = U + V \\ X_3 = U + V + W \end{cases}$$

- (a) Quelle est la loi du vecteur $[X_1, X_2, X_3]'$?
- (b) Déterminer $\mathbb{E}[X_3|X_1, X_2]$. Comparer à $\mathbb{E}[X_3|X_2]$.
- (c) Expliquer le résultat de la question précédente par un dessin.
- (d) Soit $[X, Y, Z]$ un vecteur aléatoire. Commenter l'affirmation :

$$\mathbb{E}[Z|X, Y] = \mathbb{E}[Z|Y] \Rightarrow Z \text{ est indépendante de } X.$$

2. On définit cette fois le vecteur aléatoire $[X_1, X_2, X_3]'$ comme suit :

$$\begin{cases} X_1 = U \\ X_2 = U + V \\ X_3 = V \end{cases}$$

- Quelle est la loi du vecteur $[X_1, X_2, X_3]'$?
- Déterminer $\mathbb{E}[X_3|X_1, X_2]$. Comparer à $\mathbb{E}[X_3|X_2]$.
- Expliquer le résultat de la question précédente par un dessin.
- Soit $[X, Y, Z]$ un vecteur aléatoire. Commenter l'affirmation :

$$Z \text{ est indépendante de } X \Rightarrow \mathbb{E}[Z|X, Y] = \mathbb{E}[Z|Y].$$

Corrigé

Cet exercice est corrigé en annexe, sujet de mai 2007.

Exercice 2.22 (Régression linéaire vs. Espérance conditionnelle)

Soit $[X, Y, Z]'$ un vecteur gaussien centré de matrice de covariance :

$$\Gamma = \begin{bmatrix} 1 & 0.4 & 0.8 \\ 0.4 & 1 & 0.6 \\ 0.8 & 0.6 & 1 \end{bmatrix}$$

- Quelle est la loi du vecteur $[Y, Z]'$ sachant $X = x$?
- Déterminer la variable aléatoire $\mathbb{E}[Z|X, Y]$ et la loi de Z sachant $(X, Y) = (x, y)$.
- Etablir qu'il existe une variable aléatoire U combinaison linéaire de X et Z qui n'est pas corrélée avec Z .
- Que dire de $\mathbb{E}[UZ^2]$? En déduire que $\mathbb{E}[XZ^2] = 0$.
- Prouver de même que $\mathbb{E}[YZ^2] = 0$.
- Soit $T = 3X + 5Y + 2Z$. Quelle est la régression linéaire \hat{T} de T sur (X, Y) ? Quel est l'écart quadratique entre \hat{T} et T , c'est-à-dire $\mathbb{E}[(T - \hat{T})^2]$?
- Quelle est l'espérance conditionnelle $\tilde{T} = \mathbb{E}[T|X, Y]$? Ecart quadratique entre \tilde{T} et T ?
- Soit $S = 3X + 5Y + 2Z^2$. Quelle est la régression linéaire \hat{S} de S sur (X, Y) ? Ecart quadratique entre \hat{S} et S ?
- Quelle est l'espérance conditionnelle $\tilde{S} = \mathbb{E}[S|X, Y]$? Ecart quadratique entre \tilde{S} et S ?
- Comparer les écarts quadratiques des deux questions précédentes.

Exercice 2.23 (Un hiver 2001)

Soit $[X_1, X_2, X_3]'$ un vecteur gaussien centré de matrice de covariance :

$$\Gamma = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 3 & 0 \\ -1 & 0 & 5 \end{bmatrix}$$

- Quelle est la loi du vecteur aléatoire $[X_1, X_3]'$ sachant $X_2 = 1$?
- Trouver les variables aléatoires $\mathbb{E}[X_2|X_1, X_3]$ et $\mathbb{E}[X_1|X_2, X_3]$.
- Quelle est la loi de X_1 conditionnée par $X_2 = 7$ et $X_3 = 11$?

4. Déterminer $\mathbb{E}[X_1^2|X_2, X_3]$ et vérifier que $\mathbb{E}[X_1^2|X_2, X_3] = \mathbb{E}[X_1^2|X_3]$.
5. Soit $Y = 3X_1 + 2X_2 + X_3$. Que vaut $\mathbb{E}[Y|X_2, X_3]$? Quel est l'écart quadratique entre Y et $\mathbb{E}[Y|X_2, X_3]$, c'est-à-dire $\mathbb{E}[(Y - \mathbb{E}[Y|X_2, X_3])^2]$?
6. Mêmes questions avec $Z = 3X_1^2 + 2X_2 + X_3$.

Exercice 2.24 (Un hiver 2002)

Soit $[X, Y, Z]'$ un vecteur gaussien centré de matrice de covariance :

$$\Gamma = \begin{bmatrix} 4 & -1 & -2 \\ -1 & 3 & 3 \\ -2 & 3 & 4 \end{bmatrix}$$

1. Déterminer $f(Y, Z) = \mathbb{E}[X|Y, Z]$.
2. Calculer $\mathbb{E}[(X - f(Y, Z))^2|Y, Z]$ et $\mathbb{E}[X^2|Y, Z]$.
3. Quelle est la loi conditionnelle de X sachant $(Y, Z) = (y, z)$.
4. Déterminer $\mathbb{E}[3X + YZ|Y, Z]$, $\mathbb{E}[YX^2 - \exp(Z)|Y, Z]$, $\mathbb{E}[X/(1 + Y^2)|Y, Z]$, ainsi que $\mathbb{E}[(X + Y)^3|Y, Z]$.
5. Déterminer $\mathbb{E}[Y|Z]$ et $\mathbb{E}[Z|Y]$. En déduire $\mathbb{E}[XYZ]$.
6. Déterminer $\mathbb{E}[Z^2|Y]$. En déduire $\mathbb{E}[XYZ^2]$.

Exercice 2.25 (Un printemps 2008)

Soit $X = [X_1, X_2, X_3]'$ un vecteur gaussien centré et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 4 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

1. Quelles sont les lois marginales de X ?
2. Certaines composantes de X sont-elles indépendantes deux à deux? Si oui, lesquelles? Certaines composantes de X sont-elles orthogonales deux à deux? Si oui, lesquelles?
3. Quelle est la loi de (X_1, X_2) ?
4. Sans calcul, déterminer $\mathbb{E}[X_2|X_3]$ et $\mathbb{E}[(X_2 - \mathbb{E}[X_2|X_3])^2]$.
5. Calculer $\mathbb{E}[X_1|X_3]$ et $\mathbb{E}[(X_1 - \mathbb{E}[X_1|X_3])^2]$.
6. Calculer $\mathbb{E}[X_1|X_2, X_3]$ et $\mathbb{E}[(X_1 - \mathbb{E}[X_1|X_2, X_3])^2]$.
7. Quelle est la loi de X_1 sachant $(X_2 = x_2, X_3 = x_3)$?
8. On pose $S = \mathbb{E}[X_1^2|X_2, X_3]$ et \hat{S} le résultat de la régression linéaire de X_1^2 sur (X_2, X_3) . Justifier que $\mathbb{E}[(X_1^2 - S)^2] \leq \mathbb{E}[(X_1^2 - \hat{S})^2]$.

Corrigé

Cete exercice est corrigé en annexe, sujet de mai 2008.

Exercice 2.26 (Observations bruitées)

On considère une variable aléatoire $X \sim \mathcal{N}(0, \sigma^2)$, de variance σ^2 connue. On dispose pour X de n observations Y_1, \dots, Y_n bruitées, c'est-à-dire que :

$$\forall i \in \{1, \dots, n\} \quad Y_i = X + \varepsilon_i.$$

Les variables aléatoires ε_i sont indépendantes, identiquement distribuées suivant une loi $\mathcal{N}(0, \tau^2)$, et indépendantes de X . On suppose de plus la variance τ^2 du bruit d'observation connue. On veut déduire de Y_1, \dots, Y_n une estimation de X .

1. Expliquer pourquoi le vecteur $[X, Y]' = [X, Y_1, \dots, Y_n]'$ est gaussien.
2. Calculer $\text{Cov}(X, Y_i)$, $\text{Var}(Y_i)$ et $\text{Cov}(Y_i, Y_j)$.
3. Déterminer la matrice de covariance Γ_Y de $[Y_1, \dots, Y_n]'$. Déterminer son inverse (on pourra : soit inverser un système linéaire ; soit la chercher sous la même forme que Γ_Y , c'est-à-dire termes diagonaux égaux, termes hors-diagonal égaux).
4. On note $\hat{X}_n = \mathbb{E}[X|Y]$ l'estimateur des moindres carrés de X sachant Y . Déduire des questions précédentes que :

$$\hat{X}_n = \frac{\sigma^2}{n\sigma^2 + \tau^2}(Y_1 + \dots + Y_n).$$

5. Déterminer la variance résiduelle

$$\sigma_n^2 = \mathbb{E}[(X - \hat{X}_n)^2].$$

La comparer à la variance initiale σ^2 de la variable aléatoire X .

6. En déduire que :

$$\hat{X}_n \xrightarrow[n \rightarrow +\infty]{L^2} X.$$

7. Supposons $X = x$ fixé. On dispose donc des observations Y_1, \dots, Y_n , avec $Y_i = x + \varepsilon_i$. Quelle est la loi de \hat{X}_n sachant $X = x$?
8. Justifier le fait que :

$$\hat{X}_n \xrightarrow[n \rightarrow +\infty]{p.s.} x.$$

“Vous savez, les gens normaux n'ont rien d'exceptionnel...”

Jackie Berroyer, dans un film de Laurence Ferreira-Barbosa (1993).

Chapitre 3

Chaînes de Markov

Introduction

Un modèle dynamique pour lequel le futur dépend de l'état présent et du hasard est appelé une chaîne de Markov : c'est un modèle simple pour représenter un phénomène aléatoire évoluant au cours du temps. Il est néanmoins efficace dans de nombreuses applications : sciences de la vie (génétique, modèles d'épidémie), finance (les cours de la bourse), théorie du signal (problèmes de filtrage, de prédiction), traitement d'image, traitement de la parole, informatique (files d'attente dans les réseaux), etc. On se contente ici d'étudier les chaînes de Markov en temps discret et à espace d'états fini. Le cours est en grande partie tiré du chapitre 1 du livre de Rick Durrett, *Essentials of Stochastic Processes* [3].

3.1 Définition d'une chaîne de Markov

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires à valeurs dans un ensemble E supposé fini, typiquement $E = \{1, 2, \dots, M\}$. E est appelé l'espace d'états.

Définition 15 (Chaîne de Markov)

$(X_n)_{n \geq 0}$ est une chaîne de Markov si pour tout $n \geq 1$ et toute suite $(i_0, i_1, \dots, i_{n-1}, i, j)$ de E , on a l'égalité suivante :

$$\mathbb{P}(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = \mathbb{P}(X_{n+1} = j | X_n = i).$$

Remarque. Ceci suppose bien sûr que le membre de gauche a un sens, c'est-à-dire que :

$$\mathbb{P}(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) > 0.$$

Autrement dit, sachant le présent, le futur est indépendant du passé. Ou encore : étant donné l'état présent, toute information sur le passé est inutile pour prévoir l'état futur. A priori, rien ne dit que la probabilité précédente est indépendante de l'instant n . C'est pourtant une situation fréquente et c'est l'hypothèse, dite d'homogénéité, que nous faisons désormais.

Définition 16 (Homogénéité)

Une chaîne de Markov est dite homogène dans le temps si la probabilité précédente ne dépend pas de n . On appelle alors probabilité de transition de l'état i vers l'état j la quantité :

$$p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i).$$

On appelle matrice de transition de la chaîne la matrice $P = [p_{ij}]_{1 \leq i, j \leq M}$ de taille $M \times M$.

La connaissance de la loi initiale, c'est-à-dire des $\mathbb{P}(X_0 = i)$ pour tout $i \in \{1, \dots, M\}$, et des probabilités de transition permet d'écrire très simplement la loi jointe du vecteur aléatoire (X_0, \dots, X_n) , puisque :

$$\begin{aligned}\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) &= \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1|X_0 = i_0)\dots\mathbb{P}(X_n = i_n|X_{n-1} = i_{n-1}) \\ &= \mathbb{P}(X_0 = i_0)p_{i_0 i_1} \dots p_{i_{n-1} i_n}\end{aligned}$$

On peut énoncer quelques résultats très simples sur les matrices de transition.

Propriétés 8 (Matrice de transition)

Toute matrice de transition vérifie les propriétés suivantes.

– Encadrement des coefficients :

$$\forall (i, j) \in \{1, \dots, M\}^2, 0 \leq p_{ij} \leq 1.$$

– Somme par ligne : pour tout $i \in \{1, \dots, M\}$, on a :

$$\sum_{j=1}^M p_{ij} = 1.$$

– Spectre : P admet la valeur propre 1, le vecteur $e = [1, \dots, 1]'$ étant un vecteur propre associé.

Preuve. Soit $P = [p_{ij}]_{1 \leq i, j \leq M}$ une matrice de transition.

– Pour tout couple (i, j) , la quantité $p_{i,j}$ est une probabilité, donc elle est bien comprise entre 0 et 1.

– Pour tout indice $i \in \{1, \dots, M\}$:

$$\sum_{j=1}^M p_{ij} = \frac{1}{\mathbb{P}(X_n = i)} \sum_{j=1}^M \mathbb{P}(X_{n+1} = j, X_n = i) = \frac{\mathbb{P}(X_{n+1} \in \{1, \dots, M\}, X_n = i)}{\mathbb{P}(X_n = i)},$$

c'est-à-dire :

$$\sum_{j=1}^M p_{ij} = \frac{\mathbb{P}(X_n = i)}{\mathbb{P}(X_n = i)} = 1.$$

– En notant $e = [1, \dots, 1]'$, on a donc :

$$Pe = \begin{bmatrix} \sum_{j=1}^M p_{1j} \\ \vdots \\ \sum_{j=1}^M p_{Mj} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = e,$$

donc e est un vecteur propre pour la valeur propre 1. ■

On appelle encore matrice stochastique, ou matrice markovienne, toute matrice vérifiant les deux premières propriétés (elle vérifie alors aussi la dernière).

Remarque. Le sous-espace propre associé à la valeur propre 1 n'est pas nécessairement de dimension égale à 1. Pour preuve l'exemple trivial de la matrice identité : noter que cet exemple

correspondrait à une chaîne qui ne change jamais d'état, il ne présente donc pas un grand intérêt... Exemple moins trivial, prendre pour matrice de transition :

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

C'est bien une matrice de transition donc e est vecteur propre pour la valeur propre 1. Mais on voit que le vecteur $v = [1, 1, 0, 0]'$ est un autre vecteur propre associé à 1. Noter que cette chaîne a elle aussi un comportement déterministe : si on connaît l'état à un instant, on le connaît à tous les instants suivants.

A toute chaîne de Markov peut être associé un graphe de transition de la façon suivante : les sommets du graphe sont les états $1, \dots, M$ de la chaîne et il existe un arc, étiqueté p_{ij} , de i vers j si $p_{ij} > 0$. Cette construction est commode lorsque la matrice P est très creuse. Autrement dit lorsque d'un état on ne peut transiter que vers un petit nombre d'états.

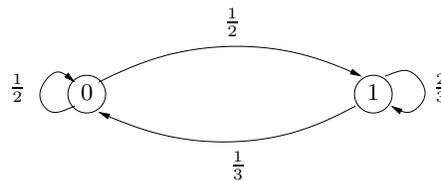


FIGURE 3.1 – Graphe de transition de la ligne téléphonique.

Exemple : la ligne téléphonique

On considère une ligne de téléphone. L'état X_n de cette ligne à l'étape n est 0 si elle est libre et 1 si elle est occupée. Entre deux instants successifs, il y a une probabilité $1/2$ pour qu'un appel arrive. Si la ligne est occupée et qu'un appel arrive, cet appel est perdu. La probabilité pour que la ligne se libère entre l'instant n et l'instant $(n + 1)$ est $1/3$. Le graphe de transition de cette chaîne de Markov est donné figure 3.1. La matrice de transition est la suivante :

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{bmatrix}.$$

3.2 Equations de Chapman-Kolmogorov

Les probabilités de transition en n étapes sont en fait complètement déterminées par les probabilités de transition en un coup, c'est-à-dire par la matrice de transition. Ceci est explicité par les équations de Chapman-Kolmogorov, que nous allons voir maintenant.

Notation. La probabilité d'aller de l'état i à l'état j en n coups est notée :

$$p_{ij}^{(n)} = \mathbb{P}(X_n = j | X_0 = i),$$

et la matrice de transition en n coups est notée :

$$P^{(n)} = \left[p_{ij}^{(n)} \right]_{1 \leq i, j \leq M}.$$

On adopte aussi la convention $P^{(0)} = I_M$, matrice identité de taille M .

Proposition 17 (Equations de Chapman-Kolmogorov)

Pour tout $n \geq 0$, la matrice de transition en n coups est la puissance $n^{\text{ème}}$ de la matrice de transition de la chaîne, c'est-à-dire :

$$P^{(n)} = P^n.$$

Preuve. Par récurrence sur n . $P^{(0)} = I_M = P^0$. Pour $n \geq 0$, on suppose que $P^{(n)} = P^n$ et on veut montrer que $P^{(n+1)} = P^{n+1}$, c'est-à-dire que :

$$\forall 1 \leq i, j \leq M \quad p_{ij}^{(n+1)} = (P^{n+1})(i, j).$$

Or on écrit :

$$\begin{aligned} p_{ij}^{(n+1)} &= \mathbb{P}(X_{n+1} = j | X_0 = i) = \sum_{k=1}^M \mathbb{P}(X_{n+1} = j, X_n = k | X_0 = i) \\ &= \sum_{k=1}^M \mathbb{P}(X_{n+1} = j | X_n = k, X_0 = i) \cdot \mathbb{P}(X_n = k | X_0 = i), \end{aligned}$$

mais par la propriété de Markov, on a :

$$\mathbb{P}(X_{n+1} = j | X_n = k, X_0 = i) = \mathbb{P}(X_{n+1} = j | X_n = k),$$

donc :

$$\begin{aligned} p_{ij}^{(n+1)} &= \sum_{k=1}^M \mathbb{P}(X_{n+1} = j | X_n = k) \cdot \mathbb{P}(X_n = k | X_0 = i) \\ &= \sum_{k=1}^M p_{ik}^{(n)} \cdot p_{kj}, \end{aligned}$$

ce qui exactement dire que d'un point de vue matriciel :

$$P^{(n+1)} = P^{(n)} \cdot P.$$

Or par hypothèse, on sait que $P^{(n)} = P^n$, donc la récurrence passe. ■

Remarque. On en déduit que pour tout couple d'entiers naturels (m, n) :

$$P^{(m+n)} = P^{m+n} = P^m \cdot P^n = P^{(m)} \cdot P^{(n)}.$$

C'est plutôt cette équation qu'on appelle relation de Chapman-Kolmogorov. Ce qu'on traduit comme suit : aller de i à j en $(m+n)$ pas, c'est aller de i à un certain k en m pas et de k à j en n pas.

Notation. Tout comme les transitions de la chaîne, la position initiale X_0 peut être aléatoire. On convient de noter la loi de X_0 comme un vecteur **ligne** de taille M :

$$\mu = [\mu_1, \dots, \mu_M] = [\mathbb{P}(X_0 = 1), \dots, \mathbb{P}(X_0 = M)].$$

De même, on notera en vecteur ligne la loi de X_n :

$$\mathbb{P}(X_n) = [\mathbb{P}(X_n = 1), \dots, \mathbb{P}(X_n = M)].$$

Corollaire 2 (Loi marginale de la chaîne)

Soit (X_n) une chaîne de Markov de loi initiale μ et de matrice de transition P , alors pour tout entier naturel n , la loi de X_n est :

$$\mathbb{P}(X_n) = \mu P^n.$$

Preuve. Pour tout $j \in \{1, \dots, M\}$, on a :

$$\mathbb{P}(X_n = j) = \sum_{i=1}^M \mathbb{P}(X_n = j, X_0 = i) = \sum_{i=1}^M \mathbb{P}(X_0 = i) \cdot \mathbb{P}(X_n = j | X_0 = i),$$

donc avec les notations adoptées :

$$\mathbb{P}(X_n = j) = \sum_{i=1}^M \mu_i P_{ij}^{(n)} = \sum_{i=1}^M \mu_i P_{ij}^n,$$

c'est-à-dire : $\mathbb{P}(X_n) = \mu P^n$. ■

Rappel. Pour une suite de variables aléatoires $(X_n)_{n \geq 0}$ à valeurs dans un ensemble fini $E = \{1, \dots, M\}$, la loi de X_n est simplement le vecteur ligne $\mathbb{P}(X_n) = [\mathbb{P}(X_n = 1), \dots, \mathbb{P}(X_n = M)]$. La convergence en loi correspond donc simplement à la convergence de ce vecteur de taille M , c'est-à-dire à la convergence de chacune de ses M composantes. Puisque $\mathbb{P}(X_n) = \mu P^n$, une condition suffisante pour la convergence en loi de (X_n) est donc la convergence de la suite $(P^n)_{n \geq 0}$ des puissances de la matrice P .

Exemple : La ligne téléphonique

On note encore $\mu = [\mu_0, \mu_1]$ la loi initiale, probabilités que la ligne soit initialement libre ou occupée. La matrice de transition est :

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{bmatrix}.$$

Pour étudier les puissances successives de P , l'idée naturelle est de la diagonaliser. On obtient pour valeurs propres 1 (on le savait déjà) et $1/6$, vecteurs propres associés $[1, 1]'$ et $[-1/2, 1/3]'$. On en déduit la matrice de passage :

$$Q = \begin{bmatrix} 1 & -1/2 \\ 1 & 1/3 \end{bmatrix},$$

ainsi que son inverse :

$$Q^{-1} = \begin{bmatrix} 2/5 & 3/5 \\ -6/5 & 6/5 \end{bmatrix}.$$

En notant Δ la matrice diagonale de coefficients 1 et $1/6$, on a donc :

$$\Delta^n \xrightarrow[n \rightarrow \infty]{} \Delta_\infty = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

d'où on déduit :

$$P^n \xrightarrow[n \rightarrow \infty]{} P_\infty = \begin{bmatrix} 2/5 & 3/5 \\ 2/5 & 3/5 \end{bmatrix}.$$

Ainsi, quelle que soit la loi initiale μ , on a convergence en loi :

$$\mathbb{P}(X_n) = \mu P^n \xrightarrow[n \rightarrow \infty]{} \mu_\infty = \left[\frac{2}{5}, \frac{3}{5} \right].$$

Interprétation : au bout d'un certain temps (histoire d'oublier la condition initiale), on regarde l'état de la ligne. Il y a deux chances sur cinq pour qu'elle soit libre et trois chances sur cinq pour qu'elle soit occupée.

Remarques.

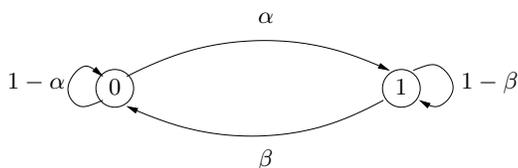


FIGURE 3.2 – Graphe de transition pour une chaîne à deux états.

1. Une autre méthode pour étudier les puissances de P sur cet exemple très simple est de noter que $P = I - A$, avec $A^2 = \frac{5}{6}A$, et d'appliquer la formule du binôme.
2. On peut généraliser le résultat précédent (voir figure 3.2). Si on considère la matrice de transition :

$$P = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix},$$

avec $0 < \alpha + \beta < 2$, on a :

$$P^n \xrightarrow{n \rightarrow \infty} P_\infty = \begin{bmatrix} \frac{\beta}{\alpha + \beta} & \frac{\alpha}{\alpha + \beta} \\ \frac{\beta}{\alpha + \beta} & \frac{\alpha}{\alpha + \beta} \end{bmatrix}.$$

En particulier, quelle que soit loi initiale μ , on a convergence en loi de (X_n) vers le vecteur de probabilité $\left[\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right]$.

3. Si on se place dans le cas particulier $\alpha = \beta = 1$, on n'a pas convergence de la suite des puissances de la matrice de transition P , puisque pour tout $n \geq 0$:

$$\begin{cases} P^{2n} &= I_2 \\ P^{2n+1} &= P \end{cases}$$

Et, à part dans le cas particulier où $\mu = [1/2, 1/2]$, on n'a pas convergence en loi de (X_n) . Ceci vient du phénomène de périodicité de la chaîne, nous y reviendrons plus tard.

Il est important de voir dès maintenant que, même lorsque la suite (P^n) converge, la convergence de la loi de X_n vers une loi **indépendante** de la condition initiale n'est pas vraie pour toutes les chaînes de Markov.

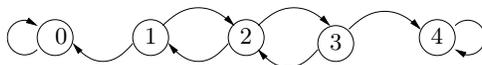


FIGURE 3.3 – Graphe de transition de la ruine du joueur.

Exemple : la ruine du joueur

A joue contre B une suite de pile ou face non biaisés et indépendants. La somme de leurs fortunes est de 4€. A chaque partie, le joueur qui gagne reçoit 1€. Le jeu s'arrête lorsque l'un des deux joueurs est ruiné. L'état X_n de la chaîne est la fortune de A à l'étape n , donc $X_n \in \{0, 1, 2, 3, 4\}$. Son graphe de transition est représenté figure 3.3 et la matrice de transition est :

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

On vérifie que :

$$P^n \xrightarrow[n \rightarrow \infty]{} P_\infty = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 3/4 & 0 & 0 & 0 & 1/4 \\ 1/2 & 0 & 0 & 0 & 1/2 \\ 1/4 & 0 & 0 & 0 & 3/4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Et on voit que si la loi initiale est $\mu = [\mu_0, \dots, \mu_4]$, alors la loi asymptotique est :

$$\mu_\infty = \left[\mu_0 + \frac{3}{4}\mu_1 + \frac{1}{2}\mu_2 + \frac{1}{4}\mu_3, 0, 0, 0, \frac{1}{4}\mu_1 + \frac{1}{2}\mu_2 + \frac{3}{4}\mu_3 + \mu_4 \right].$$

Au final, l'un des deux joueurs sera ruiné, et ce avec une probabilité qui dépend de la répartition initiale des 4€. Par exemple si A part avec 1€, alors il a trois chances sur quatre de finir ruiné, tandis que s'il part avec 2€, il n'a qu'une chance sur deux de finir ruiné. Le graphe de transition pour P_∞ est donné figure 3.4. Le problème ici vient de l'existence de plusieurs classes de récurrence, sujet sur lequel nous reviendrons par la suite.

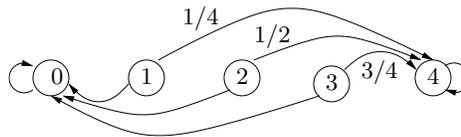


FIGURE 3.4 – Transitions en temps infini pour la ruine du joueur.

Définition 17 (Etat absorbant)

On appelle état absorbant tout état tel que : une fois dans cet état, la chaîne y reste. Autrement dit, l'état i est absorbant si $p_{ii} = 1$.

Exemple. Dans le modèle de la ruine du joueur, les états 0 et 4 sont absorbants.

3.3 Classification des états

Beaucoup de chaînes ont la propriété suivante : d'une part, la loi de X_n tend vers une limite; d'autre part, celle-ci est indépendante de la loi initiale μ . C'est par exemple le cas pour la ligne téléphonique, mais pas pour la ruine du joueur. On dit qu'elles sont ergodiques. Pour pouvoir les caractériser, il faut commencer par différencier les états que la chaîne peut visiter une infinité de fois (dits récurrents) de ceux qui ne peuvent l'être qu'un nombre fini de fois (dits transitoires).

Définition 18 (Etats communicants)

On dit que i communique avec j et on note $i \rightarrow j$ s'il existe une suite d'indices $i_0 = i, i_1, \dots, i_n = j$ telle que :

$$\forall k \in \{0, \dots, n-1\} \quad p_{i_k, i_{k+1}} > 0.$$

Autrement dit, sur le graphe de transition, on peut aller de i à j en un certain nombre d'étapes.

Remarques.

- $i \rightarrow j$ ssi il existe $n > 0$ tel que : $(P^n)_{ij} = \mathbb{P}(X_n = j | X_0 = i) > 0$.
- La communication entre états est une relation transitive : si $i \rightarrow j$ et $j \rightarrow k$, alors $i \rightarrow k$.
- Si i ne communique pas avec j , on note $i \nrightarrow j$.

Supposons que la chaîne soit initialement dans l'état i et notons :

$$T_i = \min\{n \geq 1 : X_n = i\},$$

l'instant de premier retour dans cet état. T_i est une variable aléatoire à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$, car il se peut qu'on ne revienne jamais, auquel cas $T_i = +\infty$. On introduit donc :

$$r_i = \mathbb{P}(T_i < +\infty | X_0 = i).$$

Définition 19 (Etats transitoires, états récurrents)

Avec les notations précédentes, on dit que l'état i est :

- transitoire si $r_i < 1$;
- récurrent si $r_i = 1$.

Problème de cette définition : en général, on ne sait pas calculer r_i . Heureusement, il y a une façon commode de reconnaître un état transitoire sur le graphe de transition d'une chaîne de Markov.

Proposition 18 (Transience et communication)

Soit $i \in \{1, \dots, M\}$. S'il existe j tel que $i \rightarrow j$, mais $j \not\rightarrow i$, alors i est transitoire.

Preuve. Supposons i pour lequel il existe un état j tel que $i \rightarrow j$, mais $j \not\rightarrow i$. Notons $m = \min\{n \in \mathbb{N}, (P^n)_{ij} > 0\}$. Ainsi il existe des états k_1, \dots, k_{m-1} tels que

$$P_{ik_1} \dots P_{k_{m-1}j} > 0,$$

ces états étant tous distincts de i , sans quoi m ne serait pas minimal. Dès lors, puisque la probabilité de revenir en i une fois l'état j atteint est nulle, on a :

$$r_i \leq 1 - P_{ik_1} \dots P_{k_{m-1}j} < 1. \quad \blacksquare$$

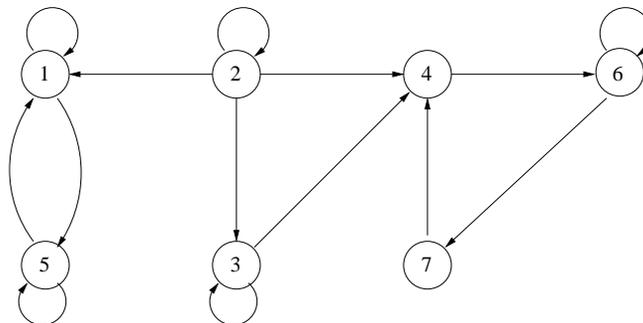


FIGURE 3.5 – Une chaîne à 7 états.

Exemple. On considère la chaîne à 7 états de la figure 3.5. $2 \rightarrow 1$, mais $1 \not\rightarrow 2$, donc 2 est transitoire. $3 \rightarrow 4$, mais $4 \not\rightarrow 3$, donc 3 est transitoire. Il reste deux ensembles : $\{1, 5\}$ et $\{4, 6, 7\}$, à l'intérieur desquels il y a communication.

Définition 20 (Ensemble clos, ensemble irréductible)

- Un sous-ensemble A de E est clos, ou fermé, s'il est impossible d'en sortir :

$$\forall i \in A, \forall j \in \bar{A} \quad p_{ij} = 0$$

– Un sous-ensemble B de E est irréductible si tous ses états communiquent :

$$\forall (i, j) \in B \times B \quad i \rightarrow j$$

Exemple. Dans l'exemple de chaîne à 7 états ci-dessus, l'ensemble $\{1, 5\}$ est fermé, l'ensemble $\{4, 6, 7\}$ aussi. Mais alors $\{1, 4, 5, 6, 7\}$ aussi, ainsi que $\{1, 2, 3, 4, 5, 6, 7\}$. Ces deux derniers exemples ne semblent pas bien pertinents, d'où la notion d'irréductibilité pour préciser les choses : les ensembles $\{1\}$, $\{2\}$, $\{3\}$, $\{5\}$, $\{6\}$, $\{1, 5\}$ et $\{4, 6, 7\}$ sont irréductibles. Cette fois, ce sont les cinq premiers ensembles qui ne semblent pas pertinents. La proposition suivante précise cette intuition.

Proposition 19 (Fermé + Irréductible \Rightarrow Récurrenents)

Si un ensemble est fermé et irréductible, tous ses états sont récurrents.

Exemple. Les seuls ensembles à la fois fermés et irréductibles sont $\{1, 5\}$ et $\{4, 6, 7\}$. De ce fait, les états 1, 4, 5, 6 et 7 sont récurrents.

Preuve (heuristique). Elle se fait en deux étapes :

- La première consiste à remarquer que des états communicants sont de même nature : si i est récurrent et si $i \rightarrow j$, alors j est récurrent. Intuitivement : partons de j , alors on est sûr de passer par i au bout d'un temps fini, sinon on aurait une contradiction avec le fait que i est récurrent et communique avec j . Or, chaque fois que la chaîne passe par i , elle a une probabilité non nulle d'aller vers j . Mais puisqu'elle passe une infinité de fois par i (état récurrent), il est certain qu'elle finira bien par passer par j . En clair, partant de j , on est sûr d'y revenir en un temps fini.
- Dans un second temps, il reste à voir que dans un ensemble fermé, il y a au moins un état récurrent. En effet, la chaîne va passer un temps infini dans cet ensemble, qui est de taille finie. Donc si elle passait un temps fini en chaque sommet, on aboutirait à une contradiction. ■

On peut alors classer tous les états d'une chaîne de Markov.

Théorème 9 (Partition de l'ensemble des états)

Soit E l'espace d'états, alors on peut partitionner E comme suit :

$$E = T \cup R_1 \cup \dots \cup R_k,$$

où T est l'ensemble des états transients et les R_i sont des classes de récurrence, c'est-à-dire des ensembles fermés et irréductibles d'états récurrents.

Preuve. Soit l'ensemble :

$$T = \{i \in E : \exists j \in E, i \rightarrow j, j \nrightarrow i\}.$$

Alors tout état i de E est transient d'après ce qui a été vu plus haut. Montrons maintenant que tout état i de $E \setminus T$ est récurrent : soit $i \in E \setminus T$ et

$$C_i = \{j \in E : i \rightarrow j\}$$

C_i est clos puisque, pour tout $j \in C_i$, si $j \rightarrow k$, alors par transitivité $i \rightarrow k$ donc $k \in C_i$; ainsi, partant de C_i , on ne peut sortir. Par ailleurs, i n'étant pas transitoire, pour tout $j \in C_i$, on a aussi $j \rightarrow i$. Ainsi, si j et k appartiennent à C_i , $j \rightarrow i$ et $i \rightarrow k$, donc $j \rightarrow k$: tous les états de C_i communiquent et C_i est un ensemble clos et irréductible. C_i est donc une classe de récurrence. Notons $R_1 = C_i$. Si $E = T \cup R_1$, la messe est dite. Sinon, on prend $i' \in E \setminus (T \cup R_1)$ et on itère le raisonnement. Puisque E est fini, on aboutira bien à un nombre fini de classes de récurrence.



Exemple. Toujours dans le même exemple, on a ainsi la partition (cf. figure 3.6) :

$$E = T \cup R_1 \cup R_2 = \{2, 3\} \cup \{1, 5\} \cup \{4, 6, 7\}.$$

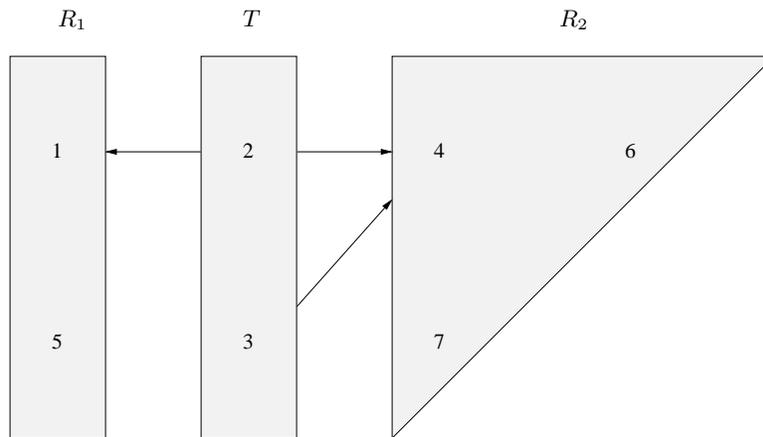


FIGURE 3.6 – Partition de E pour la chaîne à 7 états.

Le cas confortable, et très fréquent, est celui où l'espace d'états n'est formé que d'une classe de récurrence.

Définition 21 (Chaîne irréductible)

Si $E = R_1$ n'est formé que d'une classe de récurrence, on dit que la chaîne est irréductible.

Sur le graphe de transition, ceci signifie qu'on peut aller de n'importe quel sommet à n'importe quel autre en un certain nombre d'étapes. En théorie des graphes, c'est ce qu'on appelle un graphe orienté fortement connexe.

Exemple. Si α et β sont strictement positifs, la chaîne à deux états est irréductible.

Un cadre un peu plus général pour lequel les théorèmes asymptotiques du paragraphe suivant s'appliqueront est celui où on a éventuellement, en plus, des états transitoires.

Définition 22 (Chaîne indécomposable)

Si $E = T \cup R_1$ est formé d'états transitoires et d'une seule classe de récurrence, on dit que la chaîne est indécomposable.

Exemples.

- Si $\alpha > 0$, mais $\beta = 0$, la chaîne à deux états est indécomposable.
- La chaîne de la ruine du joueur n'est ni irréductible, ni indécomposable, puisqu'elle a deux classes de récurrence :

$$E = T \cup R_1 \cup R_2 = \{1, 2, 3\} \cup \{0\} \cup \{4\}.$$

3.4 Comportement asymptotique

Le but de ce paragraphe est de faire le distinguo entre les différentes situations possibles pour la convergence en loi d'une chaîne de Markov. On applique la notion d'irréductibilité du paragraphe précédent. Il faut de plus préciser le phénomène de périodicité que l'on peut rencontrer dans certains modèles.

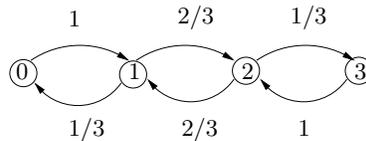


FIGURE 3.7 – Graphe de transition pour le modèle d'Ehrenfest.

Exemples.

1. La ligne téléphonique : si $0 < \alpha, \beta < 1$, on a convergence de la suite des puissances de matrices de transition (P^n) vers une matrice P_∞ dont toutes les lignes sont les mêmes. Ceci assure la convergence en loi de la chaîne (X_n) . De plus, la loi de (X_n) converge vers une loi de probabilité indépendante de la loi initiale de la chaîne. C'est la situation la plus confortable.
2. La ruine du joueur : on a encore convergence de la suite (P^n) vers une matrice P_∞ . Ceci implique qu'on a ici encore convergence en loi de la chaîne (X_n) . Néanmoins, puisque toutes les lignes de P_∞ ne sont pas égales, la loi limite de (X_n) dépend de la loi initiale de la chaîne.
3. Modèle d'Ehrenfest : on considère deux urnes A et B , contenant N boules à elles deux, numérotées de 1 à N . A chaque instant, on choisit un numéro $i \in \{1, \dots, N\}$ de façon équiprobable et on change d'urne à la boule numéro i . L'état X_n de la chaîne est le nombre de boules à l'instant n dans l'urne A . Par exemple, avec $N = 3$ boules (voir figure 3.7), on vérifie qu'on n'a pas convergence de la suite (P^n) , mais qu'on a convergence des deux sous-suites (P^{2n}) et (P^{2n+1}) . Plus précisément, si on note Q_∞ et R_∞ les limites respectives de ces sous-suites, on a (voir aussi la figure 3.8) :

$$P^{2n} \xrightarrow{n \rightarrow +\infty} Q_\infty = \begin{bmatrix} 1/4 & 0 & 3/4 & 0 \\ 0 & 3/4 & 0 & 1/4 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 3/4 & 0 & 1/4 \end{bmatrix}$$

et :

$$P^{2n+1} \xrightarrow{n \rightarrow +\infty} R_\infty = \begin{bmatrix} 0 & 3/4 & 0 & 1/4 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 3/4 & 0 & 1/4 \\ 1/4 & 0 & 3/4 & 0 \end{bmatrix}$$

Ceci n'est pas étonnant : le nombre de boules dans une urne change de parité à chaque opération, donc on ne peut avoir convergence de (P^n) . De même, la loi de X_n dépend de la parité de n , donc il n'y aura pas convergence en loi dans le cas général (i.e. sauf pour des lois initiales particulières).

Définition 23 (Période d'un état)

La période d_i de l'état i est le plus grand entier naturel divisant tout nombre n tel que $P^n(i, i) > 0$. C'est-à-dire en notant $Z_i = \{n \geq 1 : P^n(i, i) > 0\}$, on a :

$$d_i = \text{pgcd}(Z_i).$$

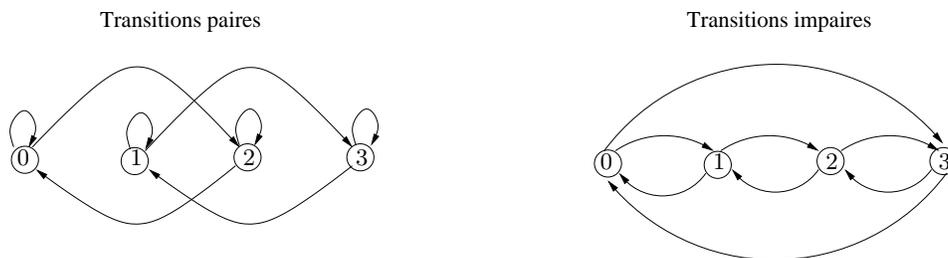


FIGURE 3.8 – Graphes de transition de P^{2n} et de P^{2n+1} pour le modèle d'Ehrenfest.

Si $d_i = 1$, on dit que i est apériodique. Si tous les états sont apériodiques, on dit que la chaîne est apériodique.

Exemples.

1. La ligne téléphonique : si $0 < \alpha, \beta < 1$, les deux états sont apériodiques.
2. La ruine du joueur : les états 0 et N sont apériodiques, tous les autres sont de période 2.
3. Modèle d'Ehrenfest : tous les états sont de période 2.

Remarque. La périodicité est un phénomène théoriquement possible, mais très rare en pratique.

Propriétés 9 (Autour de la périodicité)

- Si $p_{ii} > 0$, alors i est apériodique.
- Si i est apériodique, alors il existe $n_0 \in \mathbb{N}$ tel que :

$$\forall n \geq n_0 \quad n \in Z_i.$$

- Si i et j communiquent, alors i et j ont même période.
- Si la chaîne est irréductible, tous les états ont même période.

Preuve.

- Il est clair que si on peut boucler sur l'état i , alors :

$$Z_i = \{n \geq 1 : P^n(i, i) > 0\} = \mathbb{N}^*,$$

donc $d_i = \text{pgcd}(Z_i) = 1$.

- Puisque i est apériodique, il existe des entiers n_1, \dots, n_k dans Z_i dont le plus grand commun diviseur est 1. Mais alors le théorème de Bezout assure qu'il existe des entiers relatifs $\alpha_1, \dots, \alpha_k$ tels que :

$$\alpha_1 n_1 + \dots + \alpha_k n_k = 1.$$

En notant $\alpha^+ = \sup(\alpha, 0)$ et $\alpha^- = \sup(-\alpha, 0)$, on a donc :

$$\sum_{i=1}^k \alpha_i^+ n_i = 1 + \sum_{i=1}^k \alpha_i^- n_i.$$

Les deux sommes ci-dessus sont dans Z_i comme sommes d'éléments de Z_i et puisque Z_i est stable pour l'addition. Si on note :

$$N = \sum_{i=1}^k \alpha_i^- n_i,$$

on en déduit que les deux entiers positifs consécutifs N et $(N + 1)$ sont dans Z_i . Montrons alors que tout entier n tel que $n \geq N^2 - 1$ est dans Z_i . On effectue la division euclidienne de n par N :

$$n = qN + r \quad \text{avec } r < N.$$

Puisque $n \geq N^2 - 1$, on a $q \geq r$ et on peut écrire $q = r + r'$, d'où :

$$n = r'N + r(N + 1) \Rightarrow n \in Z_i,$$

puisque N et $(N + 1)$ sont dans Z_i .

- Supposons que i et j communiquent, que i a pour période d et j pour période $d' < d$. Puisque i et j communiquent, il existe deux entiers k et m tels que :

$$(P^k)_{ij} > 0 \quad \text{et} \quad (P^m)_{ji} > 0.$$

Puisque la probabilité de revenir en i en $(k + m)$ étapes vérifie l'inégalité :

$$(P^{k+m})_{ii} \geq (P^k)_{ij}(P^m)_{ji} > 0,$$

on en déduit que $(k + m) \in Z_i$. Mais puisque i a pour période d , il s'ensuit que d divise $(k + m)$.

Par ailleurs, pour tout entier l de Z_j , c'est-à-dire tel que $(P^l)_{jj} > 0$, on a :

$$(P^{k+l+m})_{ii} \geq (P^k)_{ij}(P^l)_{jj}(P^m)_{ji} > 0,$$

de sorte que d divise $(k + l + m)$. Mais puisqu'on vient de dire que d divise $(k + m)$, on en déduit que d divise aussi l . Or ceci est vrai pour tout entier l de Z_j , donc par définition du plus grand commun diviseur, on en déduit que d divise d' . Ceci contredit le fait que $d' < d$.

- Ce point découle directement du précédent. ■

Remarque. S'il existe $n_0 > 0$ tel que $\mathbb{P}(X_{n_0} = i | X_0 = i) > 0$ et $\mathbb{P}(X_{n_0+1} = i | X_0 = i) > 0$, l'état i est apériodique.

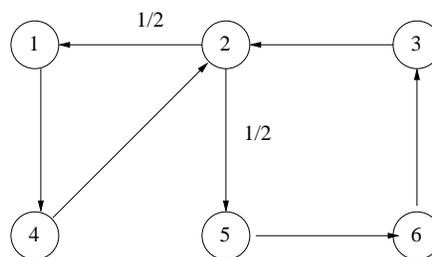


FIGURE 3.9 – Triangle et carré.

Exemple : triangle et carré.

On considère l'exemple de la figure 3.9. Pour l'état 2, on a $Z_2 = \{3, 4, 6, 7, \dots\}$ donc $d_2 = 1$. La chaîne étant irréductible, on en déduit que tous les états sont apériodiques : cette chaîne est apériodique.

Si on revient aux exemples du début de paragraphe, on voit que la situation où tout se passe bien est celle de la ligne téléphonique, à la fois irréductible et apériodique. C'est ce que nous allons préciser maintenant.

Définition 24 (Loi stationnaire)

Un vecteur de probabilité ligne π est une loi stationnaire, ou invariante, ou d'équilibre, d'une chaîne de Markov de matrice de transition P si : $\pi P = \pi$.

Le vecteur de probabilité π est stationnaire si c'est un vecteur propre "à gauche" de la matrice de transition P pour la valeur propre 1.

Exemple. Pour la chaîne d'Ehrenfest à 3 boules, on cherche donc $\pi = [\pi_0, \dots, \pi_3]$ tel que :

$$\begin{cases} \pi P = \pi \\ \sum_{i=0}^3 \pi_i = 1 \end{cases}$$

Après calculs, on obtient :

$$\pi = \left[\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8} \right],$$

c'est-à-dire que la loi stationnaire correspond à une loi binomiale $\mathcal{B}(3, \frac{1}{2})$. De façon générale, pour le modèle d'Ehrenfest à M boules, on montre qu'il y a une unique loi stationnaire $\pi \sim \mathcal{B}(M, \frac{1}{2})$.

Lemme 2

Si $X_0 \sim \pi$, alors pour tout $n \geq 0$: $X_n \sim \pi$.

Preuve. On sait que si $\mathbb{P}(X_0) = \mu$, alors $\mathbb{P}(X_n) = \mu P^n$. Si $X_0 \sim \pi$, on a donc :

$$\mathbb{P}(X_n) = \pi P^n = (\pi P) P^{n-1} = \pi P^{n-1} = \dots = \pi.$$

■

En physique, la loi stationnaire correspond à l'état d'équilibre pour le système.

Propriétés 10 (Existence d'une loi stationnaire)

Pour toute matrice de transition P , il existe (au moins) une loi stationnaire π .

Preuve. Soit p_0 un vecteur de probabilité ligne de taille M . On considère la suite (p_n) définie par :

$$p_n = \frac{1}{n+1} \sum_{i=0}^n p_0 P^i.$$

L'ensemble \mathcal{P} des lois de probabilités sur un ensemble à M éléments s'écrit :

$$\mathcal{P} = \{p \in \mathbb{R}_+^M : p_1 + \dots + p_M = 1\}.$$

C'est un compact de \mathbb{R}^M (fermé et borné), donc de toute suite d'éléments de \mathcal{P} on peut extraire une sous-suite convergente dans \mathcal{P} . Notons donc $(p_{n_k})_{k \geq 0}$ une telle sous-suite et π sa probabilité limite. On veut montrer que π est une loi stationnaire pour P . Il suffit de remarquer que :

$$p_{n_k} P = p_{n_k} + \frac{p_0 P^{n_k+1} - p_0}{n_k + 1},$$

et de passer à la limite lorsque k tend vers l'infini pour obtenir :

$$\pi P = \pi.$$

■

Remarque. On a vu au début du chapitre que 1 est valeur propre de P , donc c'est aussi une valeur propre de P' . Ainsi il existe un vecteur colonne v non nul tel que $P'v = v$, ce qui équivaut à dire que $v'P = v'$. Ce qui est moins évident, c'est de montrer qu'il existe un vecteur propre qui soit un vecteur de probabilité.

Proposition 20 (Loi stationnaire et états transitoires)

Si π est une loi stationnaire d'une chaîne de Markov, alors pour tout état i transitoire, on a $\pi_i = 0$.

Preuve. On montre la contraposée, à savoir : $\pi_j > 0$ implique j récurrent. Soit donc j un état tel que $\pi_j > 0$. Soit N_j le nombre de visites à l'état j , c'est-à-dire :

$$N_j = \sum_{n=1}^{+\infty} \mathbb{1}_{\{X_n=j\}},$$

de sorte que :

$$\mathbb{E}[N_j|X_0 = i] = \sum_{n=1}^{+\infty} \mathbb{P}(X_n = j|X_0 = i) = \sum_{n=1}^{+\infty} (P^n)_{ij}.$$

On en déduit que :

$$\sum_{i=1}^M \pi_i \mathbb{E}[N_j|X_0 = i] = \sum_{i=1}^M \pi_i \left(\sum_{n=1}^{+\infty} (P^n)_{ij} \right),$$

et tout étant positif, on peut intervertir les ordres de sommation :

$$\sum_{i=1}^M \pi_i \mathbb{E}[N_j|X_0 = i] = \sum_{n=1}^{+\infty} \left(\sum_{i=1}^M \pi_i (P^n)_{ij} \right),$$

mais puisque $\pi P^n = \pi$, cette dernière expression se simplifie :

$$\sum_{i=1}^M \pi_i \mathbb{E}[N_j|X_0 = i] = \sum_{n=1}^{+\infty} \pi_j = +\infty.$$

Soit maintenant T_j^k la date de k -ème visite à l'état j :

$$T_j^k = \min\{n > T_j^{k-1} : X_n = j\},$$

avec bien sûr pour initialisation $T_j^0 = 0$. Alors de la propriété de Markov on déduit que pour effectuer k visites à l'état j partant à l'instant 0 de l'état i , il faut commencer par aller de l'état i à l'état j , ce qui arrive avec probabilité r_{ij} , puis boucler $(k-1)$ fois sur l'état j , ce qui arrive avec probabilité r_j^{k-1} . Autrement dit :

$$\mathbb{P}(T_j^k < +\infty | X_0 = i) = r_{ij} r_j^{k-1}.$$

Rappelons que si une variable aléatoire X est à valeurs dans \mathbb{N} , son espérance peut s'écrire sous la forme :

$$\mathbb{E}[X] = \sum_{k=1}^{+\infty} \mathbb{P}(X \geq k),$$

résultat que l'on applique à la variable aléatoire N_j :

$$\mathbb{E}[N_j|X_0 = i] = \sum_{k=1}^{+\infty} \mathbb{P}(N_j \geq k | X_0 = i).$$

Or $\mathbb{P}(N_j \geq k | X_0 = i) = \mathbb{P}(T_j^k < +\infty | X_0 = i)$, donc :

$$\mathbb{E}[N_j | X_0 = i] = \sum_{k=1}^{+\infty} r_{ij} r_j^{k-1} = \frac{r_{ij}}{1 - r_j},$$

avec la convention $1/(1 - r_j) = +\infty$ si $r_j = 1$. On en déduit :

$$\sum_{i=1}^M \pi_i \mathbb{E}[N_j | X_0 = i] = \sum_{i=1}^M \pi_i \frac{r_{ij}}{1 - r_j} \leq \sum_{i=1}^M \pi_i \frac{1}{1 - r_j} = \frac{1}{1 - r_j}.$$

Si on rapproche les deux résultats obtenus, on a donc :

$$\frac{1}{1 - r_j} \geq +\infty,$$

c'est-à-dire que $r_j = 1$, et l'état j est récurrent. ■

La loi stationnaire π n'est pas nécessairement unique, comme le montre l'exemple d'une chaîne de Markov à deux états de matrice de transition $P = I$. Pour avoir unicité de la loi stationnaire, il ne faut avoir qu'une classe de récurrence.

Proposition 21 (Irréductibilité \Rightarrow Unicité de la loi stationnaire)

Si la chaîne est irréductible, alors la loi stationnaire π est unique. De plus $\pi_i > 0$ pour tout état i de E .

Preuve. Soit π une loi stationnaire d'une chaîne de Markov irréductible. On commence par prouver que toutes les coordonnées de π sont strictement positives. Puisque π est un vecteur de probabilité, il est clair qu'il existe un état i tel que $\pi_i > 0$. Soit alors j un autre état : étant donné que la chaîne est irréductible, il existe k tel que $(P^k)_{ij} > 0$. On utilise alors la stationnarité de π , qui implique en particulier $\pi P^k = \pi$:

$$\pi_j = \sum_{l=1}^M \pi_l (P^k)_{lj} \geq \pi_i (P^k)_{ij} > 0.$$

Passons à l'unicité de π . On sait qu'il existe une loi stationnaire par la proposition 10. Pour montrer qu'elle est unique, il suffit de prouver que le sous-espace propre associé à la valeur propre 1 est de dimension 1 pour P' . Mais puisqu'une matrice et sa transposée ont mêmes valeurs propres avec mêmes ordres de multiplicité, et qu'on sait que le vecteur constant $e = [1, \dots, 1]'$ vérifie $Pe = e$, il nous suffit de montrer que tout vecteur $f = [f_1, \dots, f_M]'$ vérifiant $Pf = f$ est constant, c'est-à-dire brièvement :

$$Pf = f \Rightarrow f_1 = \dots = f_M.$$

Si u et v sont deux vecteurs (colonnes), on convient de noter uv le vecteur défini par :

$$\forall i \in \{1, \dots, M\} \quad (uv)_i = u_i v_i,$$

et u^2 le vecteur colonne uu . Introduisons l'opérateur Γ :

$$\Gamma : \begin{cases} \mathbb{R}^M & \rightarrow \mathbb{R}^M \\ f & \mapsto \Gamma(f) = f^2 - 2f(Pf) + P(f^2) \end{cases}$$

On vérifie que :

$$\forall i \in \{1, \dots, M\} \quad \Gamma(f)_i = \sum_{j=1}^M P_{ij} (f_i - f_j)^2 \geq 0 \quad (+)$$

Par ailleurs, puisque π est stationnaire, on a $\pi P(f^2) = \pi(f^2)$, donc :

$$\pi\Gamma(f) = 2(\pi(f^2) - \pi f(Pf)) = 2\pi(f(f - Pf)).$$

Donc si f est un vecteur propre de P pour la valeur propre 1, on déduit de l'équation ci-dessus que :

$$\pi\Gamma(f) = 2\pi(f(f - Pf)) = 0,$$

mais on a aussi simplement :

$$\pi\Gamma(f) = \sum_{i=1}^M \pi_i \Gamma(f)_i,$$

avec $\Gamma(f)_i \geq 0$ pour tout i d'après ci-dessus, et $\pi_i > 0$ puisque la chaîne est irréductible. On déduit des deux dernières équations que :

$$\forall i \in \{1, \dots, M\} \quad \Gamma(f)_i = 0$$

Et, à nouveau grâce à (+), on conclut que si $P_{ij} > 0$, alors $f_i = f_j$. Et puisque la chaîne est irréductible, on en déduit de proche en proche que :

$$\forall (i, j) \in \{1, \dots, M\}^2 \quad f_i = f_j,$$

ce qui finit la démonstration. ■

Exemple. Pour le modèle de ruine du joueur, si on cherche à résoudre le système d'équations $\pi P = \pi$, on obtient une infinité de vecteurs de probabilités solutions, tous ceux de la forme :

$$\pi = [p, 0, 0, 0, 1 - p],$$

avec p loisible entre 0 et 1. Ceci est bien sûr dû au fait que la chaîne n'est pas irréductible : il y a deux classes de récurrence correspondant aux deux états absorbants. Remarquons néanmoins que tout vecteur de probabilité solution est nul sur les états récurrents 1, 2 et 3, comme annoncé en proposition 20.

Remarque. En fait, ce résultat est encore vrai si on suppose seulement que la chaîne est indécomposable. Dans ce cas, on a vu que l'unique loi stationnaire π vaut 0 sur tous les états transients. Ceci permet de gagner du temps dans la résolution du système linéaire $\pi P = \pi$.

Même si la loi stationnaire π est unique, on n'est pas assuré de la convergence de la loi de X_n vers π : les phénomènes de périodicité peuvent poser problème, comme on l'a vu dans le modèle d'Ehrenfest.

Théorème 10 (Convergence en loi)

Si la chaîne est irréductible et aperiodique, de loi stationnaire π , alors :

$$\forall (i, j) \in \{1, \dots, M\} \quad p_{ij}^{(n)} \xrightarrow[n \rightarrow \infty]{} \pi_j$$

En particulier, pour toute loi initiale μ , la loi de X_n converge vers π :

$$\mathbb{P}(X_n) = \mu P^n \xrightarrow[n \rightarrow \infty]{} \pi$$

Remarque. Une autre façon d'exprimer le résultat ci-dessus est de dire que la suite de matrices (P^n) converge vers une matrice Π dont toutes les lignes sont égales à π :

$$P^n \xrightarrow[n \rightarrow +\infty]{} \Pi = \begin{bmatrix} \pi \\ \cdot \\ \cdot \\ \cdot \\ \pi \end{bmatrix} = \begin{bmatrix} \pi_1 \dots \pi_M \\ \cdot \\ \cdot \\ \cdot \\ \pi_1 \dots \pi_M \end{bmatrix}.$$

Preuve. Soit $S = E \times E$ et la probabilité de transition \bar{p} sur S comme suit :

$$\bar{p}((i_1, i_2), (j_1, j_2)) = p_{i_1 j_1} p_{i_2 j_2}.$$

En d'autres termes, on introduit une chaîne de Markov à deux coordonnées, chacune se déplaçant indépendamment suivant la dynamique donnée par P . On commence par vérifier que cette nouvelle chaîne est irréductible. Puisque P l'est, il existe deux entiers k et l tels que :

$$(P^k)_{i_1 j_1} > 0 \quad \text{et} \quad (P^l)_{i_2 j_2} > 0.$$

Puisque j_1 et j_2 sont apériodiques, alors par le deuxième point des propriétés 9, on sait que pour m assez grand on a :

$$(P^{l+m})_{j_1 j_1} > 0 \quad \text{et} \quad (P^{k+m})_{j_2 j_2} > 0,$$

On écrit alors :

$$\bar{p}^{(k+l+m)}((i_1, i_2), (j_1, j_2)) = (P^{k+l+m})_{i_1 j_1} (P^{k+l+m})_{i_2 j_2}.$$

de sorte que :

$$\bar{p}^{(k+l+m)}((i_1, i_2), (j_1, j_2)) = (P^k)_{i_1 j_1} (P^{l+m})_{j_1 j_1} (P^l)_{i_2 j_2} (P^{k+m})_{j_2 j_2} > 0,$$

et la chaîne \bar{p} est donc irréductible.

On vérifie alors facilement que l'unique loi stationnaire de \bar{p} est $\pi \times \pi$. Soit (X_n, Y_n) la chaîne sur S , soit T le premier instant auquel les deux coordonnées sont égales :

$$T = \min\{n \geq 0 : X_n = Y_n\},$$

et soit V_{ii} l'instant de première visite de l'état (i, i) :

$$V_{ii} = \min\{n \geq 0 : X_n = Y_n = i\}.$$

Puisque P est irréductible, on a $V_{ii} < +\infty$ presque sûrement et puisque $T \leq V_{ii}$, on en déduit que T est fini presque sûrement.

On veut maintenant prouver que sur l'événement $\{T \leq n\}$, X_n et Y_n ont même loi. Pour cela on écrit :

$$\mathbb{P}(X_n = j, T \leq n) = \sum_{m=1}^n \sum_{i=1}^M \mathbb{P}(T = m, X_m = i, X_n = j)$$

que l'on simplifie grâce à la propriété de Markov :

$$\mathbb{P}(X_n = j, T \leq n) = \sum_{m=1}^n \sum_{i=1}^M \mathbb{P}(X_n = j | X_m = i) \mathbb{P}(T = m, X_m = i).$$

Par définition de T et puisque X_n et Y_n ont même dynamique :

$$\mathbb{P}(X_n = j, T \leq n) = \sum_{m=1}^n \sum_{i=1}^M \mathbb{P}(Y_n = j | Y_m = i) \mathbb{P}(T = m, Y_m = i),$$

et en remontant les équations avec Y au lieu de X , on arrive donc à :

$$\mathbb{P}(X_n = j, T \leq n) = \mathbb{P}(Y_n = j, T \leq n).$$

On a alors :

$$\mathbb{P}(X_n = j) = \mathbb{P}(X_n = j, T \leq n) + \mathbb{P}(X_n = j, T > n) = \mathbb{P}(Y_n = j, T \leq n) + \mathbb{P}(X_n = j, T > n),$$

ce qui se majore facilement :

$$\mathbb{P}(X_n = j) \leq \mathbb{P}(Y_n = j) + \mathbb{P}(X_n = j, T > n).$$

De façon symétrique on a :

$$\mathbb{P}(Y_n = j) \leq \mathbb{P}(X_n = j) + \mathbb{P}(Y_n = j, T > n).$$

Mais puisque le maximum de deux nombres positifs est inférieur à leur somme, on a alors :

$$|\mathbb{P}(X_n = j) - \mathbb{P}(Y_n = j)| \leq \mathbb{P}(X_n = j, T > n) + \mathbb{P}(Y_n = j, T > n),$$

et en sommant sur j :

$$\sum_{j=1}^M |\mathbb{P}(X_n = j) - \mathbb{P}(Y_n = j)| \leq 2\mathbb{P}(T > n).$$

Si on considère comme conditions initiales $X_0 = x$ et $Y_0 \sim \pi$, il vient :

$$\sum_{j=1}^M |\mathbb{P}(X_n = j | X_0 = x) - \pi_j| \leq 2\mathbb{P}(T > n).$$

Mais puisque T est presque sûrement finie, on a :

$$\mathbb{P}(T > n) \xrightarrow[n \rightarrow +\infty]{} 0,$$

d'où on déduit :

$$\sum_{j=1}^M |\mathbb{P}(X_n = j | X_0 = x) - \pi_j| \xrightarrow[n \rightarrow +\infty]{} 0,$$

donc a fortiori :

$$\forall (i, j) \in \{1, \dots, M\} \quad p_{ij}^{(n)} \xrightarrow[n \rightarrow \infty]{} \pi_j,$$

et la convergence en loi est prouvée. ■

Rappel. Soit (X_n) une suite de variables aléatoires i.i.d. à valeurs dans l'espace d'états fini $\{1, \dots, M\}$ et de loi commune π , alors pour toute fonction $f : \{1, \dots, M\} \rightarrow \mathbb{R}$, on a :

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow[n \rightarrow \infty]{p.s.} \sum_{i=1}^M f_i \pi_i$$

C'est ce qu'on appelle la loi forte des grands nombres. On montre (mais nous l'admettrons...) que ceci est encore vrai pour des variables qui ne sont plus indépendantes, mais à dépendance markovienne, si tant est que la loi π soit définie de façon unique!

Théorème 11 (Loi forte des grands nombres)

Soit (X_n) irréductible de loi stationnaire π , soit $f : E \rightarrow \mathbb{R}$ une fonction, alors :

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow[n \rightarrow \infty]{p.s.} \sum_{i=1}^M f_i \pi_i$$

Remarque. Ainsi, la loi des grands nombres est vérifiée sous des hypothèses moins fortes que la convergence en loi. Ceci n'est pas étonnant car il y a un phénomène de moyennisation sur l'ensemble d'une trajectoire, qui fait que l'éventuelle périodicité de la chaîne n'est pas gênante.

La loi des grands nombres nous renseigne en particulier sur le temps relatif passé par une trajectoire de la chaîne dans chacun des états et donne une interprétation simple de la loi d'équilibre π .

Corollaire 3 (Temps moyen dans chaque état)

Soit (X_n) irréductible de loi stationnaire π , alors le temps relatif passé par une trajectoire de la chaîne dans l'état i converge presque sûrement vers π_i :

$$\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X_k=i\}} \xrightarrow[n \rightarrow \infty]{p.s.} \pi_i$$

Preuve. Il suffit d'appliquer la loi forte des grands nombres à la fonction

$$f : \begin{cases} E & \rightarrow \mathbb{R} \\ x & \mapsto \mathbb{1}_{\{x=i\}} \end{cases}$$

Sa moyenne sous π vaut bien sûr π_i et la somme :

$$\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X_k=i\}}$$

compte le nombre de passages de la chaîne par l'état i sur le nombre total d'étapes : quantité que l'on peut interpréter comme le temps relatif passé dans l'état i entre les dates 1 et n . ■

Exemples.

1. Dans le modèle de la ligne téléphonique, sur un intervalle de temps assez long, celle-ci est donc occupée les trois cinquièmes du temps.
2. Dans le modèle d'Ehrenfest à 3 boules, il y a 1 ou 2 boules dans l'urne A les trois quarts du temps.
3. Pour le modèle d'Ehrenfest à M boules, on a vu que la loi stationnaire π est une loi binomiale $\mathcal{B}(M, \frac{1}{2})$. Donc, si M est grand, π se comporte à peu de choses près comme un loi normale :

$$\pi \approx \mathcal{N}\left(\frac{M}{2}, \frac{M}{4}\right).$$

La moyenne du nombre de boules dans l'urne A est $M/2$, c'était intuitivement évident puisqu'il n'y aucune dissymétrie entre les deux urnes. De plus, la proportion du temps passée avec un nombre de boules compris entre $M/2 - \sqrt{M}$ et $M/2 + \sqrt{M}$ est de l'ordre de 95%.

Application. Si on considère $M = 10000$ dans le modèle d'Ehrenfest et qu'on observe le comportement de la chaîne sur une longue période de temps, il y aura en gros entre 4900 et 5100 boules dans les deux urnes pendant 95% du temps. Considérer de grandes valeurs pour M n'est pas farfelu puisqu'Ehrenfest a développé son modèle pour la physique statistique, les boules correspondant à des molécules, donc en très grand nombre.

Remarque : Chaîne indécomposable.

Les résultats asymptotiques (convergence en loi et loi des grands nombres) sont encore vérifiés si on suppose seulement la chaîne indécomposable. Pour la convergence en loi, aucune hypothèse de périodicité n'est requise sur les états transients, car de toute façon la chaîne va les quitter au bout d'un certain temps. Résultat que l'on retrouve dans la loi des grands nombres : puisque $\pi_i = 0$ pour tout état transitoire, le temps relatif passé par la chaîne dans l'état i tend vers zéro avec n .

“En résumé, j'aimerais avoir un message un peu positif à vous transmettre...
Je n'en ai pas. Est-ce que deux messages négatifs, ça vous irait ?” Woody Allen.

3.5 Exercices

Exercice 3.1 (Lancers de pièce)

On lance une pièce équilibrée : les résultats des lancers sont des variables aléatoires indépendantes Y_0, Y_1, \dots à valeurs 0 ou 1. Pour tout $n \geq 1$, on note $X_n = Y_n + Y_{n-1}$.

1. Calculer $\mathbb{P}(X_3 = 0 | X_1 = 0, X_2 = 1)$ et $\mathbb{P}(X_3 = 0 | X_2 = 1)$.
2. Est-ce que (X_n) est une chaîne de Markov ?

Corrigé

1. Si $X_1 = 0$ et $X_2 = 1$, ceci signifie que $Y_0 = Y_1 = 0$ et $Y_2 = 1$, auquel cas $X_3 \in \{1, 2\}$, donc $\mathbb{P}(X_3 = 0 | X_1 = 0, X_2 = 1) = 0$. Par contre

$$\mathbb{P}(X_3 = 0 | X_2 = 1) = \frac{\mathbb{P}(Y_3 = 0, Y_2 = 0, Y_1 = 1)}{\mathbb{P}(X_2 = 1)} = \frac{(1/2)^3}{1/2} = \frac{1}{4}.$$

2. L'exemple ci-dessus montre que (X_n) n'est pas une chaîne de Markov.

Les situations des exercices suivants sont modélisées par des chaînes de Markov. Donner dans chaque cas la matrice et le graphe de transition associés.

Exercice 3.2 (Le rat et le labyrinthe)

Un rat se déplace dans un labyrinthe qui comporte neuf compartiments (voir figure 3.10). A chaque étape, il change de compartiment. Lorsqu'il est dans un compartiment à k portes, il choisit l'une de ces k portes de façon équiprobable. X_n est le numéro du compartiment dans lequel se trouve le rat à l'étape n .

Exercice 3.3 (La roulette de la lose)

On joue à la roulette : il y a 18 numéros rouges, 18 numéros noirs et 1 vert, le numéro zéro. On joue rouge pour 1€ à chaque fois. On commence avec 50€ et on s'arrête si on a 100€ ou si on est ruiné. X_n est notre fortune après n coups.

1	2	3
4	5	6
7	8	9

FIGURE 3.10 – Labyrinthe

Exercice 3.4 (Gare au gorille)

Un zoo a reçu six gorilles, trois mâles et trois femelles répartis au hasard en deux cages de trois singes. Le directeur presbyte, incapable de discerner les sexes, décide de favoriser leur reproduction en permutant chaque semaine deux pensionnaires pris au hasard, un dans chaque cage. X_n est le nombre de guenons présentes la semaine n dans la première cage.

Exercice 3.5 (Modèle de diffusion d'Ehrenfest)

On considère deux urnes A et B , contenant N boules à elles deux, numérotées de 1 à N . A chaque instant, on choisit un numéro $i \in \{1, \dots, N\}$ de façon équiprobable et on change d'urne à la boule numéro i . L'état X_n de la chaîne est le nombre de boules à l'instant n dans l'urne A .

Exercice 3.6 (Singin' in the rain)

Un employé se rend chaque matin à pied de son appartement à son bureau et fait le contraire le soir. Il dispose en tout de 3 parapluies, certains chez lui, les autres au bureau. A Rennes, ville peu ensoleillée, il pleut 2 fois sur 3 lorsqu'il fait le trajet, et ce indépendamment du passé. X_n est le nombre de parapluies à son domicile lorsqu'il le quitte le matin.

Exercice 3.7 (La Suisse (son chocolat, ses banques, son temps mesuré))

Un magasin suisse vend des horloges : pour des raisons de place, il ne peut pas en stocker plus de 3. Le gérant a constaté que, en une journée, les probabilités de demande de 0, 1, 2 ou au moins 3 horloges sont respectivement :

$$p = [p_0, p_1, p_2, p_{\geq 3}] = [0.3, 0.4, 0.2, 0.1].$$

Chaque soir, il peut en commander à nouveau, qui seront disponibles en magasin le lendemain matin. On dit que l'Helvète applique une méthode $(i, j)_{0 \leq i < j \leq 3}$ s'il passe commande lorsqu'à la fermeture il lui reste un nombre inférieur ou égal à i en stock afin d'en avoir j en magasin le matin suivant. X_n est le nombre d'horloges dans le magasin le soir à la fermeture.

1. Méthode $(2, 3)$: donner la matrice de transition.
2. Mêmes questions avec les méthodes $(1, 3)$ et $(0, 3)$.

Corrigé

1. La méthode $(2, 3)$ signifie que chaque matin, à l'ouverture, le magasin a refait son stock, donc il y a 3 horloges au début de la journée. Le soir, il peut en rester 0, 1, 2 ou 3 suivant les demandes de la journée et conformément aux probabilités de l'énoncé. Autrement dit $X_n \in \{0, 1, 2, 3\}$ et toutes les lignes de la matrice de transition P sont identiques dans ce cas, à savoir :

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 & 0.3 \end{bmatrix}$$

2. Traitons maintenant le cas des deux autres méthodes.

- (a) Méthode (1, 3) : dans cette éventualité, le gérant repasse commande le soir dès lors qu'il lui reste strictement moins de 2 horloges en stock le soir. Nous avons donc toujours $X_n \in \{0, 1, 2, 3\}$, mais cette fois les lignes de la matrice de transition P ne sont pas toutes identiques :

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.3 & 0 \\ 0.1 & 0.2 & 0.4 & 0.3 \end{bmatrix}$$

- (b) Méthode (0, 3) : cette fois, le gérant ne repasse commande le soir que s'il ne lui reste plus rien en magasin. La matrice de transition P s'écrit donc :

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.4 & 0.3 \\ 0.7 & 0.3 & 0 & 0 \\ 0.3 & 0.4 & 0.3 & 0 \\ 0.1 & 0.2 & 0.4 & 0.3 \end{bmatrix}$$

Exercice 3.8 (Un jeudi soir rue Saint-Michel)

Un étudiant rennais, passablement enivré après moult boissons frelatées, part d'un point d'origine et à chaque instant avance d'une unité avec probabilité p et recule avec probabilité $(1 - p)$. L'état X_n de la chaîne est l'abscisse de l'ivrogne à l'étape n . Remarque : ce modèle est appelé marche aléatoire sur \mathbb{Z} . Par rapport aux situations précédentes, l'espace d'états est infini. L'étude théorique des chaînes de Markov à espace d'états infini est plus délicate, nous ne la verrons pas en cours.

Exercice 3.9 (Collections et problèmes de caries)

1. Soit T une variable aléatoire géométrique de paramètre $p \in]0, 1[$, c'est-à-dire que T est à valeurs dans \mathbb{N}^* et pour tout $k \in \mathbb{N}^*$: $\mathbb{P}(T = k) = p(1 - p)^{k-1}$. Déterminer l'espérance de T .
2. Un enfant collectionne des images. Son album comporte N images. Chaque jour, il achète une tablette de chocolat, dans laquelle il y a une image. Soit X_n le nombre d'images distinctes dont dispose l'enfant au soir du jour n , avec la convention $X_0 = 0$. Donner matrice et graphe de transition de (X_n) .
3. Pour $i \in \{1, \dots, N\}$, soit T_i la variable aléatoire définie par :

$$T_i = \min\{n \geq 1 \mid X_n = i\}$$

Que signifie concrètement T_i ? Et $(T_{i+1} - T_i)$? Donner la loi de $(T_{i+1} - T_i)$, c'est-à-dire :

$$\forall k \in \mathbb{N}^* \quad \mathbb{P}(T_{i+1} - T_i = k).$$

4. En déduire $\mathbb{E}[T_{i+1} - T_i]$, puis $\mathbb{E}[T_N]$ et enfin un équivalent de $\mathbb{E}[T_N]$ (rappel : $1 + 1/2 + \dots + 1/N \sim \ln N$). Interpréter. Déterminer approximativement le nombre de tablettes de chocolat qu'il devra manger s'il veut compléter son album de 100 images¹.
5. Application : on lance un dé à six faces jusqu'à ce qu'on ait vu les six numéros sortir. Combien de fois en moyenne va-t-il falloir lancer le dé ?

Corrigé

1. On trouvera plus d'informations sur ce thème dans le livre de Gilles Pagès et Claude Bouzitat [6].

1. On a :

$$\mathbb{E}[T] = \sum_{k=1}^{+\infty} k \mathbb{P}(T = k) = \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = p \sum_{k=1}^{+\infty} k(1-p)^{k-1}.$$

On reconnaît une série entière de type $\sum kx^{k-1}$, dérivée terme à terme de la série entière $\sum x^k$. Or la dérivée de $\frac{1}{1-x}$ est $\frac{1}{(1-x)^2}$, donc :

$$\mathbb{E}[T] = p \cdot \frac{1}{(1 - (1-p))^2} = \frac{1}{p}.$$

2. Pour tout $n \geq 1$, X_n est donc compris entre 1 et N et on a bien sûr $X_1 = 1$. Dans la matrice de transition, de taille N , seules la diagonale et la surdiagonale ne sont pas nulles, avec

$$\begin{cases} P(i, i) &= \frac{i}{N} \\ P(i, i+1) &= \frac{N-i}{N} \end{cases}$$

Tous les états sont transitoires, sauf l'état $\{N\}$, qui est absorbant. La seule loi stationnaire

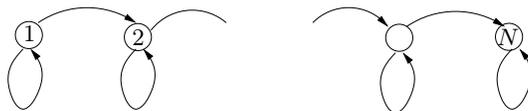


FIGURE 3.11 – Graphe de transition pour la collection d'images.

est donc déterministe : c'est $\pi = [0, \dots, 0, 1]$, ce qu'on vérifie sans problème par la résolution du système $\pi P = \pi$.

3. T_i correspond au nombre de jours nécessaires pour avoir i images distinctes. $(T_{i+1} - T_i)$ est le nombre de jours nécessaire pour obtenir une $(i+1)$ -ème nouvelle image. Si au jour n on a i images, la probabilité de ne pas en avoir de nouvelle le jour suivant est $\frac{i}{N}$, celle d'en avoir une nouvelle étant $\frac{N-i}{N}$. La probabilité $\mathbb{P}(T_{i+1} - T_i = k)$ est la probabilité de ne pas avoir de nouvelle image pendant $(k-1)$ jours et d'en avoir une nouvelle le k -ème jour. On a donc :

$$\mathbb{P}(T_{i+1} - T_i = k) = \frac{N-i}{N} \left(\frac{i}{N} \right)^{k-1}.$$

Autrement dit $(T_{i+1} - T_i)$ suit une loi géométrique de paramètre $\frac{N-i}{N}$.

4. La première question donne alors : $\mathbb{E}[T_{i+1} - T_i] = \frac{N}{N-i}$. Par linéarité de l'espérance :

$$\mathbb{E}[T_N] = \mathbb{E}[T_N - T_{N-1}] + \dots + \mathbb{E}[T_2 - T_1] + \mathbb{E}[T_1],$$

or il est clair que $T_1 = 1$, donc $\mathbb{E}[T_1] = 1$. Ainsi :

$$\mathbb{E}[T_N] = \sum_{i=0}^{N-1} \frac{N}{N-i} = N \sum_{k=1}^N \frac{1}{k} \sim N \ln N.$$

En moyenne, il faut donc environ $N \ln N$ jours à l'enfant pour compléter son album. S'il y a 100 images dans l'album, il faut environ $100 \ln(100) \approx 460$ jours pour le compléter².

2. Le résultat exact est : $100 \left(1 + \frac{1}{2} + \dots + \frac{1}{100} \right) \approx 519$. L'approximation par l'équivalent donne donc une erreur relative d'environ 10%.

5. Application : on lance un dé à six faces jusqu'à ce qu'on ait vu les six numéros sortir. Le problème est rigoureusement le même. Le nombre moyen de lancers nécessaires est donc :

$$\mathbb{E}[T] = \mathbb{E}[T_6] = 6 \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} \right) \approx 14,7.$$

Exercice 3.10 (Flash-back)

On revient sur certains exercices déjà vus : le labyrinthe, la roulette, les gorilles priapiques, le modèle d'Ehrenfest, les parapluies, les horloges, la collection.

1. La chaîne est-elle irréductible ? indécomposable ?
2. Déterminer la (ou les) loi(s) stationnaire(s).
3. Pour les chaînes irréductibles, préciser la périodicité.

Exercice 3.11 (Classification d'états)

On considère la matrice de transition suivante :

$$P = \begin{bmatrix} .4 & .3 & .3 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 \\ .5 & 0 & .5 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 \\ 0 & .3 & 0 & .3 & .4 \end{bmatrix}$$

1. Quels sont les états récurrents, quels sont les états transitoires ?
2. Déterminer la (ou les) loi(s) stationnaire(s).

Corrigé

1. On a $1 \rightarrow 2$, mais $2 \not\rightarrow 1$, donc 1 est transient. De même $3 \rightarrow 2$, mais $2 \not\rightarrow 3$. De même $5 \rightarrow 2$, mais $2 \not\rightarrow 5$. Par contre, on a $2 \rightarrow 4$ et $4 \rightarrow 2$. Ainsi, il y a trois états transitoires, $T = \{1, 3, 5\}$, et une unique classe de récurrence, $R_1 = \{2, 4\}$.
2. Il n'y a qu'une classe de récurrence donc la chaîne est irréductible et il y a unicité de la loi stationnaire $\pi = [\pi_1, \pi_2, \pi_3, \pi_4, \pi_5]$. On a alors sans calcul : $\pi_1 = \pi_3 = \pi_5 = 0$. Il reste à résoudre le système $\pi P = \pi$ avec les deux équations restantes : ceci donne $\pi_2 = \pi_4 = 1/2$.

Exercice 3.12 (Trafic routier)

Sur une route, en moyenne, trois camions sur quatre sont suivis par une voiture, tandis que seule une voiture sur cinq est suivie par un camion. Déterminer les proportions de voitures et de camions sur cette route.

Corrigé

Si on se place à un endroit donné de la route, on note (X_n) la chaîne de Markov à deux valeurs : $X_n = 1$ si le n -ème véhicule qu'on voit passer est une voiture, $X_n = 2$ si c'est un camion. Les hypothèses donnent alors la matrice de transition suivante :

$$P = \begin{bmatrix} 4/5 & 1/5 \\ 3/4 & 1/4 \end{bmatrix}$$

Cette chaîne est clairement irréductible. On détermine sa loi stationnaire en résolvant le système $\pi P = \pi$, ce qui donne :

$$\pi = \left[\frac{15}{19}, \frac{4}{19} \right].$$

Ceci correspond aux proportions respectives de voitures et de camions sur cette route.

Exercice 3.13 (Mujeres al borde de un ataque de nervios)

Un jour donné, Aude est soit énervée, soit grognon, soit de bonne humeur. Si elle est énervée aujourd'hui, elle sera énervée ou grognon ou de bonne humeur demain avec probabilités 0.7, 0.2 et 0.1 respectivement. Si elle est grognon aujourd'hui, elle sera énervée ou grognon ou de bonne humeur demain avec probabilités 0.4, 0.3 et 0.3 respectivement. Si elle est de bonne humeur aujourd'hui, elle sera énervée ou grognon ou de bonne humeur demain avec probabilités 0.2, 0.4 et 0.4 respectivement. En moyenne, combien de jours par an Aude est-elle énervée? Connaissant Aude, ce modèle vous semble-t-il réaliste?

Exercice 3.14 (Boules et urnes)

On considère 4 boules numérotées de 1 à 4, réparties en deux urnes A et B . A chaque instant, on tire un nombre k au hasard entre 1 et 4, on enlève la boule numéro k de l'urne dans laquelle elle se trouve et on la remet au hasard dans l'une des deux urnes. On note X_n le nombre de boules dans l'urne A à l'instant n .

1. Donner la matrice et le graphe de transition de (X_n) .
2. La chaîne est-elle irréductible? apériodique?
3. Loi(s) stationnaire(s)?
4. On commence avec l'urne A vide. Au bout d'un temps supposé assez grand, on observe le nombre de boules dans l'urne A . Quelle est (à peu de choses près) la probabilité que ce nombre soit pair?
5. On commence avec l'urne A pleine. On observe une réalisation $(X_0, X_1, \dots, X_n, \dots)$ de la chaîne. Quelle est la proportion du temps où il y a strictement moins de boules dans A que dans B ?
6. Généraliser l'étude précédente avec M boules numérotées de 1 à M .

Corrigé

Cet exercice est corrigé en annexe, sujet de juin 2005.

Exercice 3.15 (Changements de signes)

On considère la chaîne de Markov (X_n) à valeurs dans $\{-N, \dots, N\} \setminus \{0\}$, définie pour tout $n \geq 1$ par : $X_n = -\text{signe}(X_{n-1})Z_n$, où les Z_n sont indépendantes et identiquement distribuées de loi uniforme sur l'ensemble $\{1, \dots, N\}$ et avec la convention : $\text{signe}(x) = 1$ si $x > 0$, et $\text{signe}(x) = -1$ si $x < 0$.

1. Donner la matrice de transition P de cette chaîne. Déterminer P^n selon la parité de n .
2. La chaîne est-elle irréductible? apériodique?
3. Déterminer la loi stationnaire π de la chaîne. Comparer à $\lim_{n \rightarrow \infty} P^n$.

Corrigé

1. P est une matrice de taille $2N$. Soit J la matrice carrée de taille N ne comportant que des 1, $[0]$ la matrice carrée de taille N ne comportant que des 0. Alors P se décrit très simplement par blocs :

$$P = \frac{1}{N} \begin{bmatrix} [0] & J \\ J & [0] \end{bmatrix}.$$

Puisque $J^2 = N \cdot J$, on a pour tout $n \geq 1$: $J^n = N^{n-1}J$. Le calcul des puissances de P se fait très simplement par blocs et on obtient pour tout $n \geq 1$: $P^{2n} = P$ et

$$P^{2n+1} = \frac{1}{N} \begin{bmatrix} J & [0] \\ [0] & J \end{bmatrix}.$$

- Tous les états communiquent donc la chaîne est irréductible. Tous les états sont clairement de période 2.
- La chaîne étant irréductible, elle admet une unique loi stationnaire π , solution de $\pi P = \pi$. La résolution du système donne la loi uniforme sur les $2N$ états :

$$\pi = \left[\frac{1}{2N}, \dots, \frac{1}{2N} \right].$$

On n'a pas $\lim_{n \rightarrow \infty} P^n = \Pi$, où Π serait la matrice aux lignes identiques, toutes égales à π . Ceci vient du fait que la chaîne n'est pas apériodique.

Exercice 3.16 (Bistochasticité et Monopoly)

- On dit qu'une matrice de transition (ou matrice stochastique) P est bistochastique si la somme de chaque colonne est aussi égale à 1. Soit (X_n) une chaîne de Markov ayant une telle matrice comme matrice de transition : vérifier que la loi uniforme est une loi stationnaire de cette chaîne.
- Un jeu du genre Monopoly a dix cases (voir figure 3.12 à droite). On part de la case 0 et on lance un dé équilibré à six faces pour avancer le pion. X_n est la position du pion après le n -ème lancer.
 - Déterminer la matrice de transition de la chaîne de Markov (X_n) .
 - La chaîne est-elle irréductible ? apériodique ?
 - Déterminer la (ou les) loi(s) stationnaire(s).

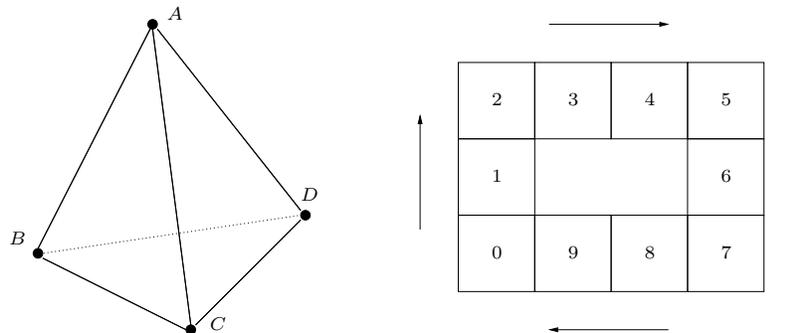


FIGURE 3.12 – Tétraèdre et Monopoly

Exercice 3.17 (Le scarabée)

Un scarabée se déplace sur les arêtes d'un tétraèdre régulier (voir figure 3.12 à gauche). Quel que soit le sommet où il se trouve à un instant donné, il choisit au hasard et de façon équiprobable le sommet vers lequel il va se diriger. Il lui faut une unité de temps pour l'atteindre. On suppose de plus que le scarabée se déplace en continu, c'est-à-dire qu'il ne s'arrête jamais en un sommet. X_n est la position du scarabée à l'instant n .

- Déterminer la matrice de transition de la chaîne de Markov (X_n) . Loi(s) stationnaire(s) ?
- A-t-on convergence en loi de (X_n) ?
- Le scarabée paye 1€ chaque fois qu'il passe au sommet A , 2€ chaque fois qu'il passe au sommet B , 3€ chaque fois qu'il passe au sommet C , 4€ chaque fois qu'il passe au sommet D . Soit C_N le coût de sa trajectoire jusqu'à l'instant N . Que dire de la convergence de $\frac{C_N}{N}$?

- Supposons maintenant qu'en chaque sommet, le scarabée reste sur place avec probabilité $7/10$ et parte vers chacun des autres sommets avec probabilité $1/10$. Que deviennent les résultats précédents ?

Corrigé

- La matrice de transition P de la chaîne de Markov (X_n) a des coefficients $1/3$ partout sauf sur sa diagonale. Cette chaîne est une marche aléatoire sur un graphe connexe, et l'unique loi stationnaire est la loi uniforme $\pi = [0.25, 0.25, 0.25, 0.25]$. Ceci n'est pas étonnant puisqu'on peut tout aussi bien remarquer que la matrice P est bistochastique.
- Partant du sommet A , on peut y revenir en 2 coups, 3 coups, 4 coups, etc. Donc A est apériodique. Puisque tous les sommets communiquent, ils sont tous apériodiques et la chaîne est apériodique. On a donc convergence de la loi de (X_n) vers π .
- Par contre, on a le théorème ergodique, i.e. la convergence presque sûre de la suite de variables aléatoires $(\frac{C_N}{N})$ vers le coût moyen, à savoir $\frac{1+2+3+4}{4} = 2.5$ €.
- La chaîne est bistochastique donc la loi stationnaire est toujours la loi uniforme π . On n'a plus de périodicité, donc il y a convergence en loi : la loi de X_n tend vers π . A fortiori, on a la loi des grands nombres.

Exercice 3.18 (Le coup du parapluie)

Un employé lambda, appelons le Franz Kafka, se rend chaque matin de son appartement à son bureau et fait le contraire le soir. Il dispose en tout de 3 parapluies, certains chez lui, les autres au bureau. A Prague, ville peu ensoleillée au delà du raisonnable, il pleut 2 fois sur 3 lorsqu'il fait le trajet, et ce indépendamment du passé. Soit X_n le nombre de parapluies à son domicile lorsqu'il le quitte le matin.

- Déterminer la matrice de transition de la chaîne de Markov associée.
- Quelle est la proportion du temps où Kafka est mouillé ?
- Généraliser avec n parapluies.

Corrigé

- La chaîne de Markov associée a pour matrice de transition :

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 2/9 & 5/9 & 2/9 & 0 \\ 0 & 2/9 & 5/9 & 2/9 \\ 0 & 0 & 2/9 & 7/9 \end{bmatrix}$$

- Après calculs, l'unique loi stationnaire est :

$$\pi = \left[\frac{1}{10}, \frac{3}{10}, \frac{3}{10}, \frac{3}{10} \right].$$

Sur un trajet du matin, Kafka est mouillé s'il n'a aucun parapluie chez lui et qu'il pleut : ceci arrive avec probabilité : $\frac{1}{10} \times \frac{2}{3} = \frac{1}{15}$. Sur un trajet du soir, il est mouillé s'il n'a aucun parapluie au bureau (i.e. ils étaient tous au domicile le matin et il faisait beau le matin) et qu'il pleut : ceci arrive avec probabilité : $\frac{3}{10} \times \frac{1}{3} \times \frac{2}{3} = \frac{1}{15}$. Que ce soit le matin ou le soir, Kafka est mouillé une fois sur quinze.

- La généralisation avec n parapluies ne pose pas problème : les lignes de la matrice de transition sont les mêmes que ci-dessus. La loi stationnaire est cette fois

$$\pi = \left[\frac{1}{3n+1}, \frac{3}{3n+1}, \dots, \frac{3}{3n+1} \right].$$

Et que ce soit à l'aller ou au retour, Kafka est mouillé avec probabilité $\frac{2}{9n+3}$.

Exercice 3.19 (Modèle épidémiologique)

On considère l'évolution d'un virus dans une population de taille N . Au jour n , il y a X_n individus infectés et $S_n = (N - X_n)$ individus sains. Le jour suivant, les X_n individus précédemment infectés sont sains, mais chacun des S_n individus précédemment sains a eu une probabilité p de rencontrer chacun des I_n infectés de la veille et donc de contracter l'infection, toutes ces rencontres étant indépendantes l'une de l'autre. Il est clair que si personne n'est infecté, il en sera de même le lendemain. Le paramètre p est appelé taux de contact infectieux.

- On suppose que la population ne compte que 3 individus et que le taux de contact infectieux est $p = 1/3$.
 - Donner matrice et graphe de transition de la chaîne de Markov (X_n) .
 - La chaîne est-elle irréductible ? indécomposable ?
 - Déterminer la (ou les) loi(s) stationnaire(s). Interpréter.
- On suppose toujours avoir une population de 3 individus, mais le taux de contact infectieux est le paramètre $p \in]0, 1[$. En notant $q = (1 - p)$, donner le graphe de transition. Loi(s) stationnaire(s) ?
- On suppose maintenant une population de N individus, avec taux de contact infectieux $p \in]0, 1[$. Justifier le fait que (X_n) admet pour probabilités de transition :

$$p_{ij} = \binom{N-i}{j} (1-q^i) q^{i(N-i-j)} \mathbb{1}_{\{i+j \leq N\}}.$$

Corrigé

Le corrigé est donné en annexe, sujet de juin 2006.

Exercice 3.20 (Pile ou Face)

On joue une suite infinie de Pile ou Face non biaisés : ceci fournit une suite de variables aléatoires $(X_n)_{n \geq 0}$ indépendantes et identiquement distribuées avec $\mathbb{P}(X_n = P) = \mathbb{P}(X_n = F) = 1/2$. A partir de cette suite on considère la chaîne de Markov $(Y_n)_{n \geq 1}$ définie par : $Y_1 = (X_0, X_1)$, $Y_2 = (X_1, X_2)$, et de façon générale $Y_n = (X_{n-1}, X_n)$ pour tout $n \geq 1$. Cette chaîne est donc à valeurs dans l'espace d'états :

$$E = \{E_1, E_2, E_3, E_4\} = \{(P, F), (F, F), (F, P), (P, P)\}.$$

- Donner la matrice et le graphe de transition de (Y_n) .
- La chaîne est-elle irréductible, apériodique ?
- Loi(s) stationnaire(s) ?
- Retrouver le résultat de la question précédente en calculant directement la loi de Y_n :

$$\mathbb{P}(Y_n) = [\mathbb{P}(Y_n = E_1), \mathbb{P}(Y_n = E_2), \mathbb{P}(Y_n = E_3), \mathbb{P}(Y_n = E_4)].$$

Corrigé

Le corrigé est donné en annexe, sujet de mai 2007.

Exercice 3.21 (Mai 2008)

Soit l'espace d'états $E = \{1, 2, 3, 4\}$ d'une chaîne de Markov homogène.

1. Compléter la matrice suivante pour qu'elle soit une matrice de transition :

$$P = \begin{bmatrix} \cdot & \frac{1}{2} & 0 & 0 \\ \cdot & \frac{2}{3} & 0 & 0 \\ 0 & \cdot & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \cdot & 0 & \frac{1}{2} \end{bmatrix}$$

2. Représenter le graphe associé.
3. La chaîne est-elle irréductible ? Indécomposable ?
4. Déterminer la (ou les) loi(s) stationnaire(s).
5. On considère qu'au temps 0, on est dans l'état 3. Pour un grand nombre d'unité de temps n , quelles sont les probabilités qu'on soit dans chacun des quatre états ?

Corrigé

Le corrigé est donné en annexe, sujet de mai 2008.

Exercice 3.22 (Et les Athéniens s'atteignent)

On reprend le problème de transmission bruitée vu au premier chapitre (exercice 1.7) : un message doit être transmis d'un point successivement à travers N canaux. Ce message peut prendre deux valeurs, 0 ou 1. Durant le passage par un canal, le message a la probabilité $p \in]0, 1[$ d'être bruité, i.e. d'être transformé en son contraire, et $(1-p)$ d'être transmis fidèlement. Les canaux se comportent indépendamment les uns des autres. Pour tout $n \in \{1, \dots, N\}$, notons p_n la probabilité qu'en sortie de n^e canal, le message soit le même que celui transmis initialement.

1. Modéliser le problème par une chaîne de Markov à deux états, de matrice de transition P .
2. Montrer que (P^n) converge vers une matrice P_∞ que l'on précisera.
3. En déduire $\lim_{n \rightarrow +\infty} p_n$.

Exercice 3.23 (Simulation d'une chaîne de Markov)

On veut simuler les déplacements d'un scarabée sur les sommets d'un tétraèdre $ABCD$ et étudier la chaîne de Markov associée (voir exercice 3.17). On rappelle que X_n correspond à la position du scarabée à l'étape n . On suppose que le scarabée part du sommet A . Quel que soit le sommet où il se trouve à un instant donné, il choisit au hasard de façon équiprobable de se diriger vers l'un des trois autres sommets.

1. Simuler une trajectoire X_0, \dots, X_{1000} du scarabée à l'aide de la fonction `sample`. Vérifier sur cet exemple la loi des grands nombres, c'est-à-dire que le scarabée passe à peu près un quart du temps en chaque sommet.
2. On veut observer la convergence en loi, c'est-à-dire vérifier que :

$$\mathbb{P}(X_n) = [\mathbb{P}(X_n = A), \mathbb{P}(X_n = B), \mathbb{P}(X_n = C), \mathbb{P}(X_n = D)] \xrightarrow[n \rightarrow +\infty]{} \pi = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right].$$

Etudier la position du scarabée pour $n = 50$.

3. On considère maintenant que les déplacements du scarabée sont régis par la matrice de transition :

$$P = \begin{bmatrix} 0 & 2/3 & 0 & 1/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 2/3 & 0 & 1/3 & 0 \end{bmatrix}.$$

- (a) Vérifier la loi des grands nombres.
 (b) Que dire de la convergence en loi?
4. Tirer au hasard une matrice de transition P à l'aide de la fonction `runif`. Vérifier que la loi des grands nombres et la convergence en loi permettent de trouver un même vecteur probabilité ligne π . Retrouver précisément cette loi d'équilibre grâce à la fonction `eigen` et en utilisant sa propriété caractéristique : π est un vecteur propre à gauche de P associé à la valeur propre 1.

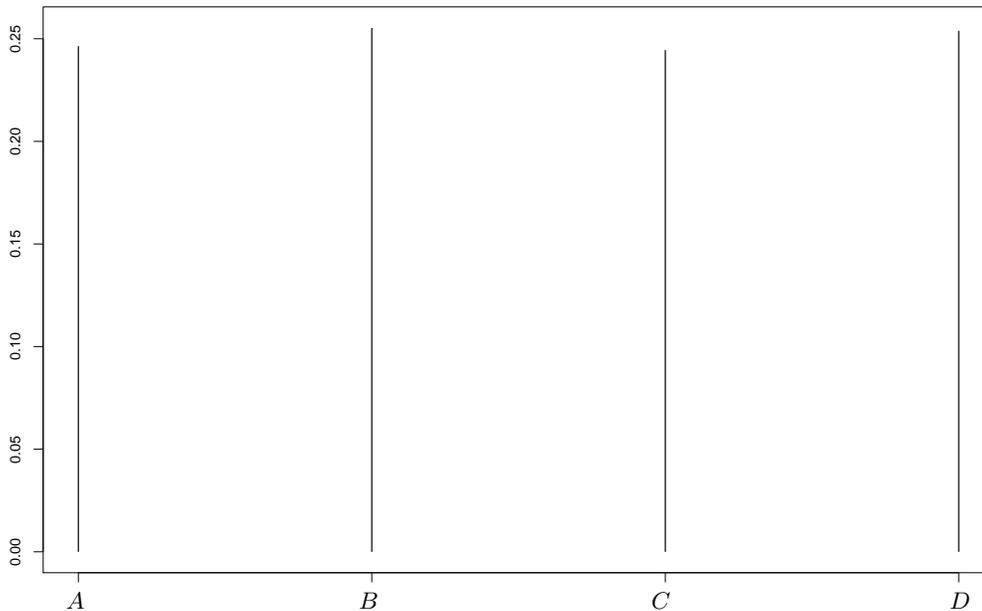


FIGURE 3.13 – Fréquences empiriques pour une seule trajectoire de longueur 1000.

Corrigé

Les simulations sont effectuées à l'aide du logiciel R.

1. Le résultat est représenté figure 3.13.

```
P=matrix(1/3,4,4)-diag(1/3,4) # P est la matrice de transition.
mu0=c(1,0,0,0) # le scarabée part du sommet A.
n0=1001 # le scarabée marche de l'instant 0 à l'instant n0-1.
x=numeric(n0)
x[1]=sample(1:4,1,prob=mu0)
for (i in 1:(n0-1)){x[i+1]=sample(1:4,1,prob=P[x[i],])}
plot(table(x)/n0,type='h',xlab='position du scarabee',ylab='frequences empiriques')
```

2. Le résultat est représenté figure 3.14.

```
P=matrix(1/3,4,4)-diag(1/3,4) # P est la matrice de transition.
mu0=c(1,0,0,0) # le scarabée part du sommet A.
n0=51 # le scarabée marche de l'instant 0 à l'instant n0-1.
ns=1000 # on simule ns trajectoires du scarabée.
S=matrix(0,nrow=ns,ncol=n0)
for (j in (1:ns)){
  S[j,1]=sample(1:4,1,prob=mu0)
```

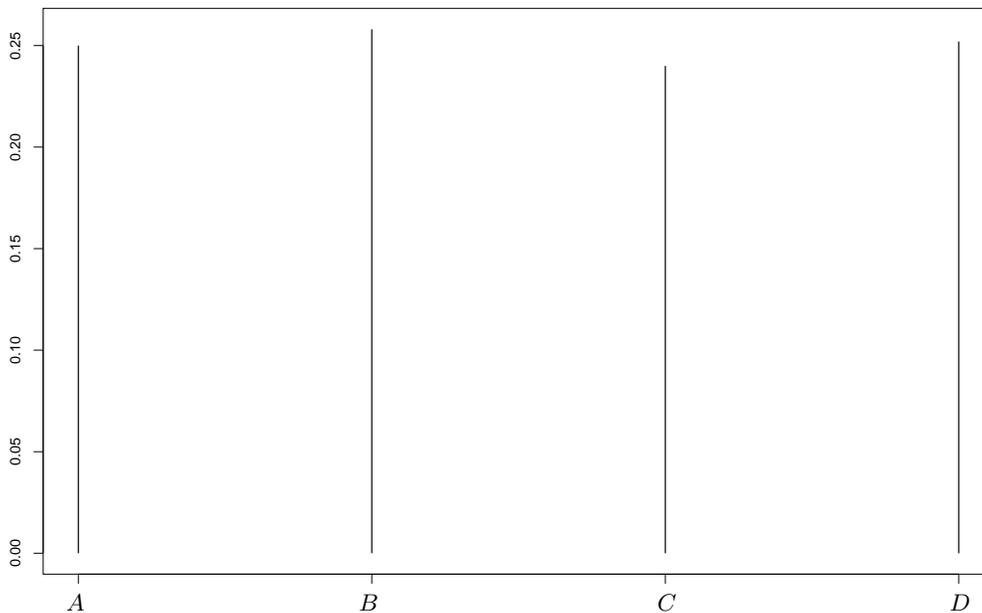


FIGURE 3.14 – Fréquences empiriques à la date 50 pour 1000 trajectoires.

```
for (i in 1:(n0-1)){S[j,i+1]=sample(1:4,1,prob=P[S[j,i],])}
plot(table(S[,n0])/ns,type='h',xlab='position du scarabee',ylab='frequences
empiriques')
```

3. (a) Le résultat est représenté figure 3.15. La chaîne étant irréductible et la matrice de transition bistochastique, la seule loi d'équilibre est à nouveau la loi π uniforme sur les quatre états.

```
P=matrix(c(0,2/3,0,1/3,1/3,0,2/3,0,0,1/3,0,2/3,2/3,0,1/3,0),4,4,byrow=TRUE)
mu0=c(1,0,0,0)
n0=1001
x=numeric(n0)
x[1]=sample(1:4,1,prob=mu0)
for (i in 1:(n0-1)){x[i+1]=sample(1:4,1,prob=P[x[i],])}
plot(table(x)/n0,type='h',xlab='position du scarabee',ylab='frequences
empiriques')
```

- (b) Le résultat est représenté figure 3.16. La chaîne étant périodique, de période 2, le scarabée ne peut être qu'au sommet A ou C après un nombre pair de déplacements. On voit donc que la loi des grands nombres s'applique, mais pas la convergence en loi.

```
P=matrix(c(0,2/3,0,1/3,1/3,0,2/3,0,0,1/3,0,2/3,2/3,0,1/3,0),4,4,byrow=TRUE)
mu0=c(1,0,0,0)
n0=51
ns=1000
S=matrix(0,nrow=ns,ncol=n0)
for (j in (1:ns)){
S[j,1]=sample(1:4,1,prob=mu0)
for (i in 1:(n0-1)){S[j,i+1]=sample(1:4,1,prob=P[S[j,i],])}
plot(table(S[,n0])/ns,type='h',xlab='position du scarabee',ylab='frequences
empiriques')
```

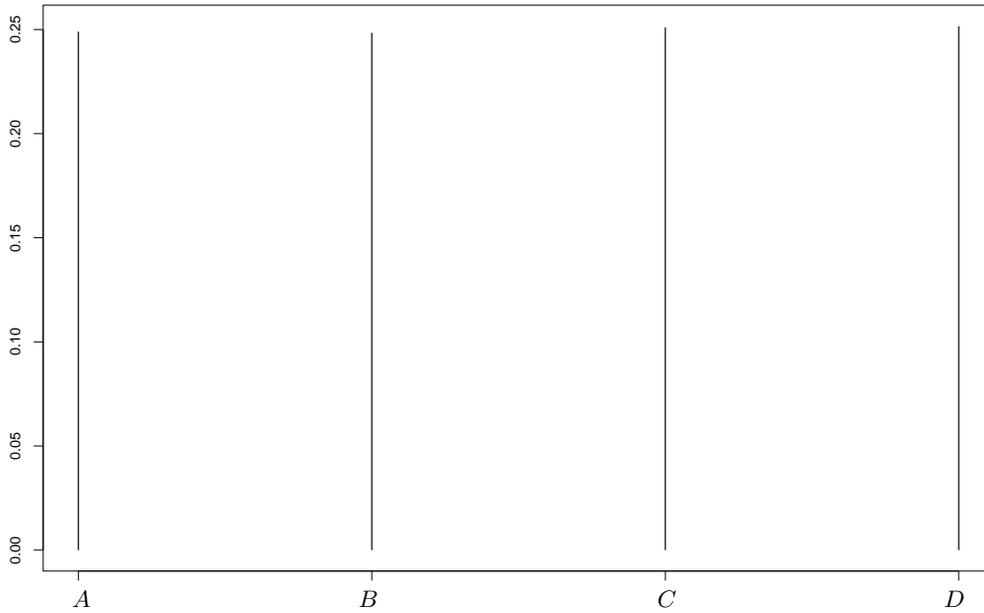


FIGURE 3.15 – Fréquences empiriques pour une seule trajectoire de longueur 1000.

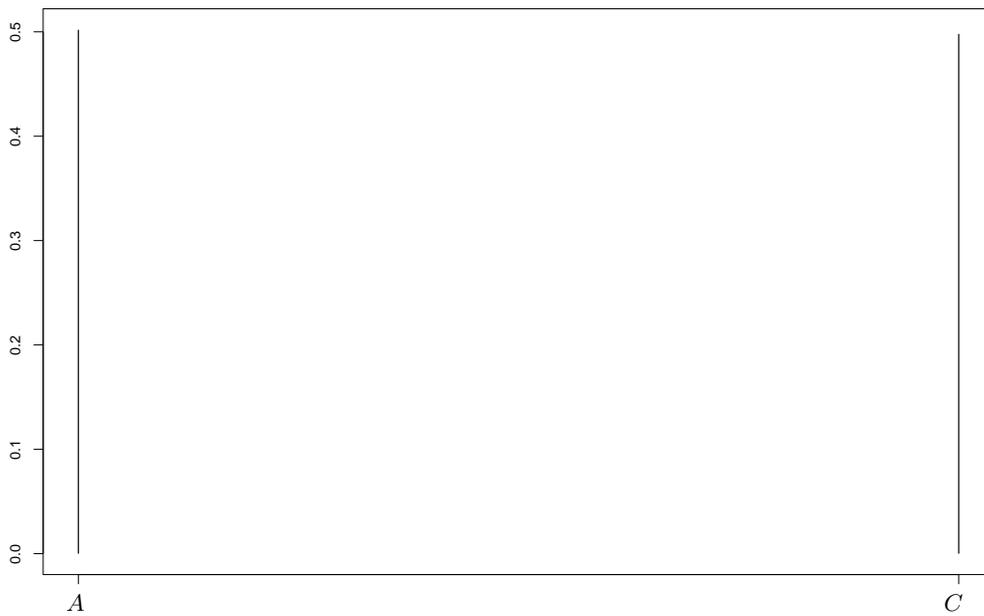


FIGURE 3.16 – Fréquences empiriques à la date 50 pour 1000 trajectoires.

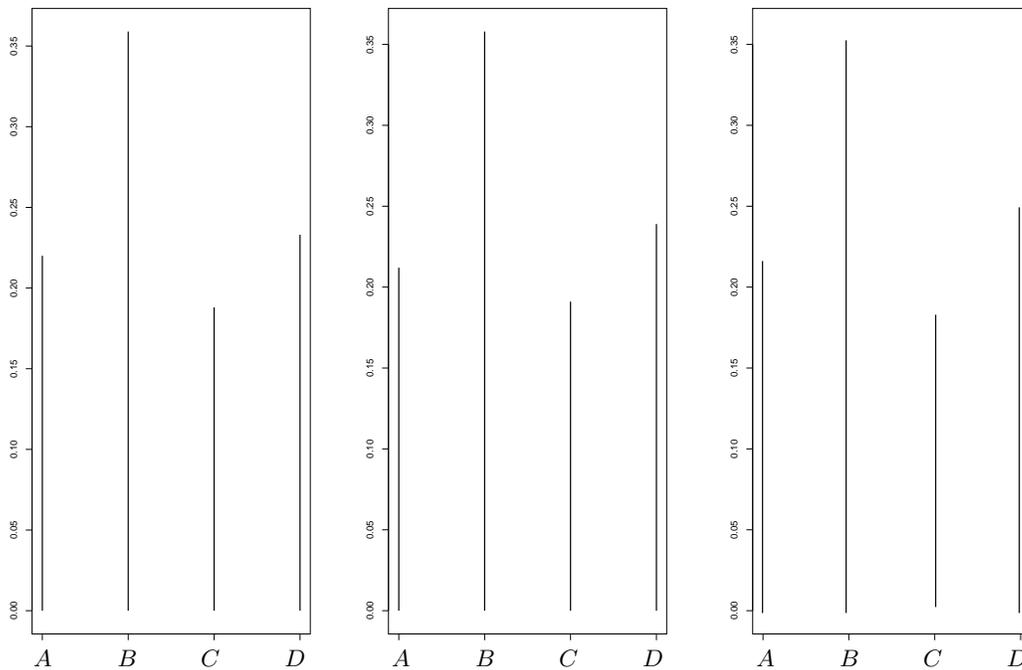


FIGURE 3.17 – Loi stationnaire obtenue de 3 façons : convergence en loi (à gauche), loi des grands nombres (au centre), vecteur propre principal (à droite).

4. Le résultat est représenté figure 3.17. Si on tire une matrice au hasard, on est sûr d'obtenir une chaîne irréductible et apériodique. La convergence en loi comme la loi des grands nombres permettent donc de retrouver la loi stationnaire théorique.

```
P=matrix(runif(16),4,4)
P=P/(rowSums(P)%*%t(rep(1,4)))
vp=as.numeric(eigen(t(P))$vectors[,1])
vp=vp/(sum(vp))
mu0=c(1,0,0,0)
n0=1001
ns=1000
S=matrix(0,nrow=ns,ncol=n0)
for (j in (1:ns)){
S[j,1]=sample(1:4,1,prob=mu0)
for (i in 1:(n0-1)){S[j,i+1]=sample(1:4,1,prob=P[S[j,i],])}}
m=matrix(1:3,nrow=1,ncol=3)
layout(m)
plot(table(S[,n0])/ns,type='h',xlab='position du scarabee',ylab='frequences empiriques')
plot(table(S[1,])/n0,type='h',xlab='position du scarabee',ylab='frequences empiriques')
plot(vp,type='h',xlab='position du scarabee',ylab='probabilité théorique')
```

Annexe A

Annales

Université de Rennes 2
Licence MASS 3
Arnaud Guyader

Mercredi 8 Juin 2005
Durée : 2 heures
Calculatrice autorisée

Examen de Probabilités

I. Couple aléatoire (6 points)

On considère la fonction suivante :

$$f(x, y) = e^{-y} \mathbb{1}_{\{0 < x < y\}}.$$

1. Vérifier que $f(x, y)$ définit une densité de probabilité sur \mathbb{R}^2 .
2. Calculer les densités marginales $f(x)$ et $f(y)$ de X et Y . Les variables X et Y sont-elles indépendantes ?
3. Calculer $\mathbb{E}[X]$, $\mathbb{E}[Y]$, $\text{Cov}(X, Y)$.
4. Déterminer la densité conditionnelle $f(y|x)$ de Y sachant $X = x$. En déduire $\mathbb{E}[Y|X = x]$, puis l'espérance conditionnelle $\mathbb{E}[Y|X]$.
5. Déterminer de même $\mathbb{E}[X|Y]$.
6. On considère le couple aléatoire (Z, T) défini par

$$\begin{cases} Z &= X + Y \\ T &= Y - X \end{cases}$$

Déterminer la loi jointe $f_{Z,T}(z, t)$ du couple (Z, T) .

7. En déduire les densités marginales de Z et T .

II. Vecteur gaussien bidimensionnel (4 points)

On considère un vecteur gaussien $[X, Y]'$ de moyenne $m = [1, -1]'$ et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$$

1. Ecrire la densité du vecteur $[X, Y]'$.
2. Quelle est la loi de X ? de Y ? de $X + Y$?
3. Déterminer l'espérance conditionnelle $\mathbb{E}[X|Y]$. Quelle est sa loi ?

III. Vecteur gaussien tridimensionnel (6 points)

On considère un vecteur gaussien $[X, Y, Z]'$ centré et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

1. Donner l'espérance conditionnelle $\mathbb{E}[X|Y, Z]$.
2. Quelle est la loi de Z sachant $X = 1$ et $Y = 2$?
3. Déterminer l'espérance conditionnelle du vecteur $[X, Y]'$ sachant Z , notée $\mathbb{E}[X, Y|Z]$.
4. Quelle est la loi du vecteur $[X, Y]'$ sachant $Z = 1$?
5. On pose alors :

$$\begin{cases} U = -X + Y + Z \\ V = X - Y + Z \\ W = X + Y - Z \end{cases}$$

Pourquoi le vecteur aléatoire $[U, V, W]'$ est-il gaussien ?

6. Quelles sont les lois de U , de V , de W ?
7. Déterminer la moyenne et la matrice de covariance de $[U, V, W]'$. Qu'en déduire sur la corrélation de U , V et W ?

IV. Chaîne de Markov (4 points)

On considère 4 boules numérotées de 1 à 4, réparties en deux urnes A et B . A chaque instant, on tire un nombre k au hasard entre 1 et 4, on enlève la boule numéro k de l'urne dans laquelle elle se trouve et on la remet au hasard dans l'une des deux urnes. On note X_n le nombre de boules dans l'urne A à l'instant n .

1. Justifier par une phrase le fait que (X_n) est une chaîne de Markov.
2. Donner la matrice et le graphe de transition de (X_n) .
3. La chaîne est-elle irréductible ? apériodique ?
4. Loi(s) stationnaire(s) ?
5. On commence avec l'urne A vide. Au bout d'un temps supposé assez grand, on observe le nombre de boules dans l'urne A . Quelle est (à peu de choses près) la probabilité que ce nombre soit pair ?
6. On commence avec l'urne A pleine. On observe une réalisation $(X_0, X_1, \dots, X_n, \dots)$ de la chaîne. Quelle est la proportion du temps où il y a strictement moins de boules dans l'urne A que dans l'urne B ?
7. Bonus : Généraliser l'étude précédente avec M boules numérotées de 1 à M .

Examen de Probabilités

Corrigé

I. Couple aléatoire

On considère la fonction suivante :

$$f(x, y) = e^{-y} \mathbb{1}_{\{0 < x < y\}}.$$

1. $f(x, y)$ est mesurable positive et on vérifie sans problème que son intégrale sur \mathbb{R}^2 vaut 1. Donc $f(x, y)$ définit bien une densité de probabilité sur \mathbb{R}^2 .
2. Pour les densités marginales, on obtient $f(x) = e^{-x} \mathbb{1}_{]0, +\infty[}(x)$, c'est-à-dire que X suit une loi exponentielle $\mathcal{E}(1)$. De même, on trouve :

$$f(y) = ye^{-y} \mathbb{1}_{]0, +\infty[}(y).$$

Les variables X et Y ne sont pas indépendantes puisque :

$$f(x, y) \neq f(x)f(y).$$

On pouvait l'affirmer dès le début puisque le support de la loi jointe n'est pas un produit d'intervalles.

3. Rappelons que si $V \sim \mathcal{E}(1)$, alors :

$$\mathbb{E}[V^n] = n!$$

On a donc : $\mathbb{E}[X] = 1$. De même : $\mathbb{E}[Y] = \mathbb{E}[V^2] = 2! = 2$. Pour la covariance :

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] - 2.$$

Or :

$$\mathbb{E}[XY] = \iint_{\mathbb{R}^2} xyf(x, y) dx dy = \frac{1}{2} \int_0^{+\infty} y^3 e^{-y} dy = \frac{1}{2} \mathbb{E}[V^3] = 3,$$

d'où l'on déduit : $\text{Cov}(X, Y) = 1$.

4. La densité conditionnelle $f(y|x)$ de Y sachant $X = x$ vaut par définition pour tout $x > 0$:

$$f(y|x) = \frac{f(x, y)}{f(x)} = e^{x-y} \mathbb{1}_{\{x < y\}}.$$

C'est donc une loi exponentielle translatée. Son espérance vaut :

$$\mathbb{E}[Y|X = x] = \int_x^{+\infty} ye^{x-y} dy = x + 1.$$

De façon générale, on a donc :

$$\mathbb{E}[Y|X] = X + 1.$$

5. De même, on a successivement :

$$f(x|y) = \frac{1}{y} \mathbb{1}_{\{0 < x < y\}}.$$

C'est donc une loi uniforme sur $[0, y]$. Sa moyenne vaut :

$$\mathbb{E}[X|Y = y] = \frac{y}{2},$$

et de façon générale :

$$\mathbb{E}[X|Y] = \frac{Y}{2}.$$

6. On définit naturellement le \mathcal{C}^1 -difféomorphisme entre ouverts de \mathbb{R}^2

$$\varphi : \begin{cases} U & \rightarrow V \\ (x, y) & \mapsto (z, t) = (x + y, y - x) \end{cases}$$

avec $U = \{(x, y) \in \mathbb{R}^2 : 0 < x < y\}$ et $V = \{(z, t) \in \mathbb{R}^2 : 0 < t < z\}$. Le théorème de changement de variable donne alors pour la densité du couple aléatoire (Z, T) :

$$f_{Z,T}(z, t) = f_{X,Y}(\varphi^{-1}(z, t)) \cdot |J_{\varphi^{-1}}(z, t)|.$$

Quelques calculs donnent :

$$f_{Z,T}(z, t) = \frac{1}{2} e^{-\frac{1}{2}(z+t)} \mathbb{1}_{\{0 < t < z\}}.$$

7. Les densités marginales sont :

$$f_Z(z) = e^{-\frac{z}{2}} \mathbb{1}_{\{z > 0\}} - e^{-z} \mathbb{1}_{\{z > 0\}}.$$

et :

$$f_T(t) = e^{-t} \mathbb{1}_{\{t > 0\}},$$

c'est-à-dire que $T \sim \mathcal{E}(1)$.

II. Vecteur gaussien bidimensionnel

On considère un vecteur gaussien $[X, Y]'$ de moyenne $m = [1, -1]'$ et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$$

1. La densité du vecteur $[X, Y]'$ est :

$$f(x, y) = \frac{1}{2\pi\sqrt{3}} e^{-\frac{1}{6}(4(x-1)^2 - 2(x-1)(y+1) + (y+1)^2)}$$

2. $X \sim \mathcal{N}(1, 1)$, $Y \sim \mathcal{N}(-1, 4)$. Enfin $X + Y \sim \mathcal{N}(0, 7)$, car $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ et :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

3. Pour déterminer l'espérance conditionnelle $\mathbb{E}[X|Y]$, on utilise la formule générale du conditionnement gaussien :

$$\mathbb{E}[X|Y] = \mathbb{E}[X] + \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}(Y - \mathbb{E}[Y]),$$

ce qui donne ici : $\mathbb{E}[X|Y] = 1 + \frac{1}{4}(Y + 1)$. Puisque $Y \sim \mathcal{N}(-1, 4)$, on a :

$$\mathbb{E}[X|Y] \sim \mathcal{N}\left(1, \frac{1}{4}\right).$$

III. Vecteur gaussien tridimensionnel

On considère un vecteur gaussien $[X, Y, Z]'$ centré et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

1. L'espérance conditionnelle de X sachant (Y, Z) est :

$$\mathbb{E}[X|Y, Z] = \Gamma_{X,(Y,Z)} \Gamma_{Y,Z}^{-1} [Y, Z]' = \frac{1}{3}(Y + Z).$$

2. On a la décomposition orthogonale :

$$Z = \mathbb{E}[Z|X, Y] + (Z - \mathbb{E}[Z|X, Y]) = \mathbb{E}[Z|X, Y] + W = \frac{1}{3}(X + Y) + W,$$

avec W indépendante du couple (X, Y) . Quant à sa loi : $W \sim \mathcal{N}(0, \sigma^2)$, avec σ^2 variance résiduelle donnée par :

$$\sigma^2 = \Gamma_Z - \Gamma_{Z,(X,Y)} \Gamma_{X,Y}^{-1} \Gamma_{(X,Y),Z} = \frac{4}{3}$$

Il suit que, sachant $X = 1$ et $Y = 2$, $Z \sim \mathcal{N}\left(1, \frac{4}{3}\right)$

3. L'espérance conditionnelle $\mathbb{E}[X, Y|Z]$ est donnée par la formule de projection :

$$\mathbb{E}[X, Y|Z] = \Gamma_{(X,Y),Z} \Gamma_Z^{-1} Z = [Z/2, Z/2]'$$

Autrement dit, c'est tout simplement le vecteur $[\mathbb{E}[X|Z], \mathbb{E}[Y|Z]]'$.

4. On a comme ci-dessus la décomposition orthogonale :

$$[X, Y]' = \mathbb{E}[X, Y|Z] + ([X, Y]' - \mathbb{E}[X, Y|Z]) = [Z/2, Z/2]' + W,$$

avec W vecteur gaussien indépendant de Z , et plus précisément $W \sim \mathcal{N}([0, 0]', \Gamma_W)$, avec :

$$\Gamma_W = \Gamma_{X,Y} - \Gamma_{(X,Y),Z} \Gamma_Z^{-1} \Gamma_{Z,(X,Y)} = \begin{bmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{bmatrix}$$

5. On pose alors

$$\begin{cases} U &= -X + Y + Z \\ V &= X - Y + Z \\ W &= X + Y - Z \end{cases}$$

Le vecteur aléatoire $[U, V, W]'$ est gaussien, puisque c'est l'image du vecteur gaussien $[X, Y, Z]'$ par une application linéaire. Si on note :

$$A = \begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix},$$

on a $[U, V, W]' = A[X, Y, Z]'$

6. Les variables U , V et W sont donc gaussiennes. Par symétrie, elles suivent la même loi $\mathcal{N}(0, 4)$. La variance s'obtient par exemple via la formule :

$$\text{Var}(U) = \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z) - 2\text{Cov}(X, Y) - 2\text{Cov}(X, Z) + 2\text{Cov}(Y, Z).$$

7. Le vecteur gaussien $[U, V, W]'$ est centré et sa matrice de covariance est :

$$\Gamma_2 = A\Gamma A' = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Cette matrice est diagonale, ce qui est dans le cas gaussien une condition nécessaire et suffisante d'indépendance des variables U , V et W .

IV. Chaîne de Markov

On considère 4 boules numérotées de 1 à 4, réparties en deux urnes A et B . A chaque instant, on tire un nombre k au hasard entre 1 et 4, on enlève la boule numéro k de l'urne dans laquelle elle se trouve et on la remet au hasard dans l'une des deux urnes. On note X_n le nombre de boules dans l'urne A à l'instant n .

1. L'opération effectuée à chaque étape est aléatoire, mais ne dépend que de la composition présente des urnes, indépendamment de ce qui s'est passé au préalable. Ceci fait de (X_n) une chaîne de Markov.
2. La variable aléatoire X_n est à valeurs dans $\{0, 1, 2, 3, 4\}$. Sa matrice de transition est :

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/8 & 1/2 & 3/8 & 0 & 0 \\ 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 3/8 & 1/2 & 1/8 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}.$$

Le graphe de transition est donné figure A.1.

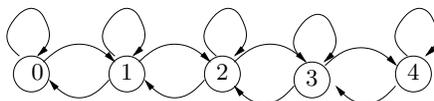


FIGURE A.1 – Graphe de transition de la chaîne de Markov (X_n) .

3. On peut passer de tout état à tout autre donc la chaîne est irréductible. Par ailleurs on peut boucler sur chaque état, donc elle est apériodique.
4. L'irréductibilité de la chaîne entraîne l'existence d'une unique loi stationnaire π , c'est-à-dire un vecteur ligne :

$$\pi = [\pi_0, \pi_1, \pi_2, \pi_3, \pi_4],$$

avec les π_i compris entre 0 et 1 et sommant à 1. On la détermine en résolvant le système d'équations $\pi P = \pi$. Après quelques calculs, on obtient :

$$\pi = \left[\frac{1}{16}, \frac{4}{16}, \frac{6}{16}, \frac{4}{16}, \frac{1}{16} \right].$$

On peut noter que de façon générale :

$$\forall i \in \{0, 1, 2, 3, 4\} \quad \pi_i = \binom{4}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{4-i},$$

c'est-à-dire que le nombre de boules dans l'urne A suit une loi binômiale $\mathcal{B}(4, \frac{1}{2})$.

5. Le théorème de convergence en loi pour les chaînes de Markov irréductibles et apériodiques assure que la loi de X_n tend vers la loi π , indépendamment de la répartition initiale entre les deux urnes. La probabilité que le nombre de boules dans l'urne A soit pair à l'instant n est donc :

$$\mathbb{P}(X_n \in \{0, 2, 4\}) \xrightarrow[n \rightarrow \infty]{} \pi_0 + \pi_2 + \pi_4 = \frac{1}{2}.$$

6. La loi forte des grands nombres pour les chaînes de Markov irréductibles assure que la proportion du temps où il y a moins de boules dans l'urne A que dans l'urne B tend vers la probabilité que ceci arrive pour la loi π . C'est-à-dire qu'on a la convergence presque sûre :

$$\frac{1}{n+1} \sum_{k=0}^n \mathbb{1}_{\{0,1\}}(X_k) \xrightarrow[n \rightarrow \infty]{p.s.} \pi_0 + \pi_1 = \frac{5}{16}.$$

7. La généralisation avec M boules est directe : la chaîne est encore irréductible et apériodique, d'unique loi stationnaire $\pi \sim \mathcal{B}(M, \frac{1}{2})$.

Université de Rennes 2
Licence MASS 3
Arnaud Guyader

Jeudi 8 Juin 2006
Durée : 2 heures
Aucun document autorisé

Examen de Probabilités

I. Variable Y définie à partir de X (3 points)

Soit X une variable aléatoire de densité :

$$\frac{2}{(\ln 2)^2} \frac{\ln(1+x)}{1+x} \mathbb{1}_{[0,1]}(x)$$

Soit Y une variable aléatoire telle que la loi conditionnelle de Y sachant $X = x$ est :

$$\frac{1}{\ln(1+x)} \frac{1}{1+y} \mathbb{1}_{[0,x]}(y)$$

1. Donner la densité jointe du couple (X, Y) .
2. Quelle est la densité de Y ? Densité conditionnelle de X sachant Y ?
3. Les variables X et Y sont-elles indépendantes?
4. Déterminer l'espérance conditionnelle $\mathbb{E}[X|Y]$.

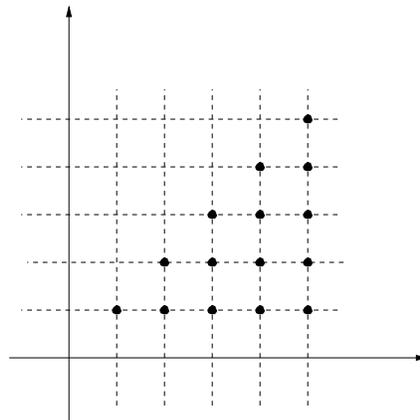


FIGURE A.2 – Tirage uniforme dans un triangle.

II. Couple aléatoire discret (6 points)

On tire un point de façon uniforme parmi ceux de la figure A.3. Ceci donne un couple aléatoire (X, Y) dont la loi jointe est : $p_{ij} = \mathbb{P}(X = i, Y = j) = \frac{1}{15}$, $1 \leq j \leq i \leq 5$.

1. Donner les lois marginales de X et de Y .

2. Soit $j \in \{1, \dots, 5\}$ fixé : déterminer la loi conditionnelle de X sachant $Y = j$, c'est-à-dire $\mathbb{P}(X = i|Y = j)$.
3. Calculer $\mathbb{E}[X|Y = j]$, en déduire $\mathbb{E}[X|Y]$, puis $\mathbb{E}[X]$ en fonction de $\mathbb{E}[Y]$.
4. Déterminer de même la loi conditionnelle de Y sachant $X = i$.
5. Calculer $\mathbb{E}[Y|X = i]$, en déduire $\mathbb{E}[Y|X]$, puis $\mathbb{E}[Y]$ en fonction de $\mathbb{E}[X]$.
6. Déduire des questions précédentes $\mathbb{E}[X]$ et $\mathbb{E}[Y]$.
7. Généralisation : soit $N \in \mathbb{N}^*$ fixé, reprendre les questions précédentes en remplaçant 5 par N (on ne demande pas le détail des calculs, uniquement les résultats).

III. Vecteur gaussien (6 points)

On considère un vecteur gaussien $V = [X, Y, Z]^T$ de moyenne $m = [1, -1, 1]'$ et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

1. Quelle est la loi du vecteur $[X, Y]'$? Donner sa densité.
2. Déterminer l'équation de la droite de régression de Y en X . Préciser le résidu quadratique.
3. Donner une variable aléatoire $U = \alpha X + \beta Y + \gamma$ qui soit indépendante de X .
4. Quelle est la fonction caractéristique du vecteur V ?
5. Donner l'expression de $\mathbb{E}[Z|X, Y]$, espérance conditionnelle de Z sachant (X, Y) ?
6. En déduire une variable aléatoire $T = aX + bY + cZ + d$ qui soit indépendante de X et Y . Préciser la variance de T .
7. On observe $Y = 1$ et $Z = 2$. Quelle est la loi de la variable aléatoire X sachant ces données ?

IV. Modèle épidémiologique (5 points)

On considère l'évolution d'un virus dans une population de taille N . Au jour n , il y a X_n individus infectés et $S_n = (N - X_n)$ individus sains. Le jour suivant, les X_n individus précédemment infectés sont sains, mais chacun des S_n individus précédemment sains a eu une probabilité p de rencontrer chacun des I_n infectés de la veille et donc de contracter l'infection, toutes ces rencontres étant indépendantes l'une de l'autre. Il est clair que si personne n'est infecté, il en sera de même le lendemain. Le paramètre p est appelé taux de contact infectieux.

1. On suppose que la population ne compte que 3 individus et que le taux de contact infectieux est $p = 1/3$.
 - (a) Donner matrice et graphe de transition de la chaîne de Markov (X_n) .
 - (b) La chaîne est-elle irréductible ? indécomposable ?
 - (c) Déterminer la (ou les) loi(s) stationnaire(s). Interpréter.
2. On suppose toujours avoir une population de 3 individus, mais le taux de contact infectieux est le paramètre $p \in]0, 1[$. En notant $q = (1 - p)$, donner la matrice de transition. Loi(s) stationnaire(s) ?
3. On suppose maintenant une population de N individus, avec taux de contact infectieux $p \in]0, 1[$. Justifier le fait que (X_n) admet pour probabilités de transition :

$$p_{ij} = \binom{N-i}{j} (1-q)^j q^{i(N-i-j)} \mathbb{1}_{\{i+j \leq N\}}.$$

Université de Rennes 2
Licence MASS 3
Arnaud Guyader

Jeudi 8 Juin 2006
Durée : 2 heures
Aucun document autorisé

Examen de Probabilités

Corrigé

I. Variable Y définie à partir de X

Soit X une variable aléatoire de densité :

$$\frac{2}{(\ln 2)^2} \frac{\ln(1+x)}{1+x} \mathbb{1}_{[0,1]}(x)$$

Soit Y une variable aléatoire telle que la loi conditionnelle de Y sachant $X = x$ est :

$$\frac{1}{\ln(1+x)} \frac{1}{1+y} \mathbb{1}_{[0,x]}(y)$$

1. Puisqu'on connaît $f(x)$ et $f(y|x)$, on en déduit la densité jointe du couple (X, Y) :

$$f(x, y) = f(y|x)f(x) = \frac{2}{(\ln 2)^2} \frac{1}{(1+x)(1+y)} \mathbb{1}_{\{0 \leq y \leq x \leq 1\}}.$$

2. La densité de Y s'obtient en marginalisant la loi jointe par rapport à x . Pour tout y dans $[0, 1]$:

$$f(y) = \int_y^1 \frac{2}{(\ln 2)^2} \frac{1}{(1+x)(1+y)} dx = \frac{2}{(\ln 2)^2(1+y)} [\ln(1+x)]_y^1,$$

ce qui donne :

$$f(y) = \frac{2}{(\ln 2)^2(1+y)} (\ln 2 - \ln(1+y)) \mathbb{1}_{[0,1]}(y).$$

On en déduit la densité conditionnelle de X sachant Y :

$$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{1}{(1+x)(\ln 2 - \ln(1+y))} \mathbb{1}_{\{0 \leq y \leq x \leq 1\}}.$$

En toute rigueur, cette formule n'est valable que pour $y \in [0, 1[$, mais pour $y = 1$ on a forcément $x = 1$. Autrement dit, la loi de X sachant $Y = 1$ n'admet pas de densité, c'est un Dirac au point 1.

3. Le support de la loi jointe $f(x, y)$ est le domaine :

$$\mathcal{D} = \{(x, y) : 0 \leq y \leq x \leq 1\},$$

qui n'est pas un pavé, donc X et Y ne peuvent être indépendantes. Une autre façon de le vérifier est de remarquer que le produit des marginales n'est pas égal à la loi jointe : $f(x, y) \neq f(x)f(y)$. Une autre méthode est de voir que la densité conditionnelle de X sachant Y n'est pas égale à la densité marginale de X : $f(x|y) \neq f(x)$. Dans la même idée, on vérifie aussi que : $f(y|x) \neq f(y)$.

4. Soit $y \in [0, 1]$ fixé. On a par définition :

$$\mathbb{E}[X|Y = y] = \int_y^1 x f(x|y) dx = \frac{1}{\ln 2 - \ln(1+y)} \int_y^1 \frac{x}{1+x} dx,$$

et il suffit d'écrire :

$$\frac{x}{1+x} = 1 - \frac{1}{1+x},$$

pour en déduire :

$$\mathbb{E}[X|Y = y] = \frac{1}{\ln 2 - \ln(1+y)} [x - \ln(1+x)]_y^1 = \frac{1-y}{\ln 2 - \ln(1+y)} - 1.$$

Noter que cette formule est encore valable pour $y = 1$, puisque le développement limité de $\ln(1+y)$ au voisinage de $y = 1$ donne :

$$\ln(1+y) = \ln 2 + \frac{1}{2}(1-y) + o(1-y),$$

d'où l'on déduit :

$$\frac{1-y}{\ln 2 - \ln(1+y)} - 1 \xrightarrow{y \rightarrow 1} 1,$$

ce qui est bien la moyenne de la variable X lorsque $Y = 1$.

De façon générale, on a donc :

$$\mathbb{E}[X|Y] = \frac{1-Y}{\ln 2 - \ln(1+Y)} - 1.$$

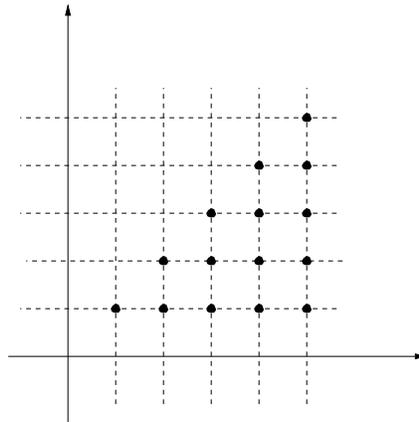


FIGURE A.3 – Tirage uniforme dans un triangle

II. Couple aléatoire discret

On tire un point de façon uniforme parmi ceux de la figure A.3. Ceci donne un couple aléatoire (X, Y) dont la loi jointe est : $p_{ij} = \mathbb{P}(X = i, Y = j) = \frac{1}{15}$, $1 \leq j \leq i \leq 5$.

1. La variable aléatoire X est à valeurs dans $\{1, \dots, 5\}$ et pour tout i dans cet ensemble, on a :

$$p_{i.} = \mathbb{P}(X = i) = \sum_{j=1}^i p_{ij} = \frac{i}{15}.$$

De même, Y est à valeurs dans $\{1, \dots, 5\}$ et pour tout j dans cet ensemble, on a :

$$p_{.j} = \mathbb{P}(Y = j) = \sum_{i=j}^5 p_{ij} = \frac{6-j}{15}.$$

2. Soit $j \in \{1, \dots, 5\}$ fixé, la loi conditionnelle de X sachant $Y = j$ est :

$$p_{i|j} = \frac{p_{ij}}{p_{.j}} = \frac{1}{6-j} \quad 1 \leq j \leq i \leq 5.$$

3. On en déduit :

$$\mathbb{E}[X|Y = j] = \sum_{i=j}^5 i p_{i|j} = \frac{1}{6-j} \sum_{i=j}^5 i,$$

et on reconnaît la somme des termes d'une suite arithmétique, d'où :

$$\mathbb{E}[X|Y = j] = \frac{j+5}{2}.$$

On en déduit l'espérance conditionnelle de X sachant Y :

$$\mathbb{E}[X|Y] = \frac{1}{2}Y + \frac{5}{2}.$$

Puisqu'on sait que $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$, on en déduit une première relation entre les espérances de X et Y :

$$\mathbb{E}[X] = \frac{1}{2}\mathbb{E}[Y] + \frac{5}{2}.$$

4. Soit $i \in \{1, \dots, 5\}$ fixé, la loi conditionnelle de Y sachant $X = i$ est :

$$p_{j|i} = \frac{p_{ij}}{p_{.i}} = \frac{1}{i} \quad 1 \leq j \leq i \leq 5.$$

5. On en déduit :

$$\mathbb{E}[Y|X = i] = \sum_{j=1}^i j p_{j|i} = \frac{1}{i} \sum_{j=1}^i j,$$

et on reconnaît à nouveau une somme arithmétique :

$$\mathbb{E}[Y|X = i] = \frac{i+1}{2},$$

d'où l'espérance conditionnelle :

$$\mathbb{E}[Y|X] = \frac{1}{2}X + \frac{1}{2},$$

et une seconde relation entre les espérances de X et Y :

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

6. Les questions précédentes donnent un système linéaire de deux équations à deux inconnues :

$$\begin{cases} \mathbb{E}[X] &= \frac{1}{2}\mathbb{E}[Y] + \frac{5}{2} \\ \mathbb{E}[Y] &= \frac{1}{2}\mathbb{E}[X] + \frac{1}{2} \end{cases}$$

que l'on résout sans problème pour arriver à $\mathbb{E}[X] = \frac{11}{3}$ et $\mathbb{E}[Y] = \frac{7}{3}$.

7. Généralisation : pour $N \in \mathbb{N}^*$ fixé, les calculs précédents se généralisent sans difficulté. On commence par remarquer qu'il y a maintenant $\frac{N(N+1)}{2}$ points dans le triangle. On obtient donc comme loi jointe :

$$p_{ij} = \mathbb{P}(X = i, Y = j) = \frac{2}{N(N+1)} \quad 1 \leq j \leq i \leq N.$$

La loi marginale de X est :

$$p_{i.} = \frac{2i}{N(N+1)},$$

tandis que celle de Y s'écrit :

$$p_{.j} = \frac{2(N+1-j)}{N(N+1)}.$$

La loi conditionnelle de X sachant $Y = j$ est donc :

$$p_{i|j} = \frac{1}{N+1-j} \quad 1 \leq j \leq i \leq N.$$

On en déduit :

$$\mathbb{E}[X|Y = j] = \frac{j+N}{2} \Rightarrow \mathbb{E}[X|Y] = \frac{1}{2}Y + \frac{N}{2},$$

d'où une première relation entre les espérances :

$$\mathbb{E}[X] = \frac{1}{2}\mathbb{E}[Y] + \frac{N}{2}.$$

De même, on trouve :

$$p_{j|i} = \frac{1}{i} \quad 1 \leq j \leq i \leq N.$$

On en déduit :

$$\mathbb{E}[Y|X = i] = \frac{i+1}{2} \Rightarrow \mathbb{E}[Y|X] = \frac{1}{2}X + \frac{1}{2},$$

d'où une seconde relation entre les espérances :

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

Les deux relations obtenues permettent d'en déduire $\mathbb{E}[X] = \frac{2N+1}{3}$ et $\mathbb{E}[Y] = \frac{N+2}{3}$.

III. Vecteur gaussien

On considère un vecteur gaussien $V = [X, Y, Z]^T$ de moyenne $m = [1, -1, 1]^T$ et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

1. Le vecteur $[X, Y]^T$ est gaussien comme image du vecteur gaussien V par une application linéaire :

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

Plus précisément, il est de moyenne $\mu = [1, -1]^T$ et de matrice de covariance :

$$\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

Puisque Σ est inversible, il admet une densité donnée par la formule :

$$f(x, y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}([x, y] - \mu')\Sigma^{-1}([x, y]' - \mu)\right).$$

Après calculs de $\det \Sigma$ et Σ^{-1} , on obtient :

$$f(x, y) = \frac{1}{2\pi} \exp\left(-x^2 + 3x + xy - \frac{1}{2}y^2 - 2y - \frac{5}{2}\right).$$

2. La droite de régression de Y en X correspond à l'espérance conditionnelle de Y sachant X :

$$\mathbb{E}[Y|X] = \mathbb{E}[Y] + \frac{\text{Cov}(Y, X)}{\text{Var}(X)}(X - \mathbb{E}[X]) = X - 2,$$

c'est-à-dire la droite d'équation $y = x - 2$. Le résidu quadratique est alors :

$$\sigma^2 = \text{Var}(Y) - \frac{\text{Cov}^2(Y, X)}{\text{Var}(X)} = 1.$$

3. Puisque $\mathbb{E}[Y|X]$ est la projection orthogonale de Y sur $H = \text{Vect}(1, X)$, on en déduit que la variable aléatoire $U = Y - \mathbb{E}[Y|X] = Y - X + 2$ est orthogonale à H , donc en particulier à la variable aléatoire $X - \mathbb{E}[X]$ de H . Ceci signifie que :

$$0 = \langle U, X - \mathbb{E}[X] \rangle = \mathbb{E}[U(X - \mathbb{E}[X])] = \text{Cov}(U, X).$$

Donc les variables aléatoires U et X sont décorréliées et puisque le vecteur $[X, U]'$ est gaussien comme image du vecteur gaussien V par une application affine, ceci est équivalent à dire que X et U sont indépendantes.

4. Soit $u = [u_1, u_2, u_3]' \in \mathbb{R}^3$, alors la fonction caractéristique du vecteur V est :

$$\Phi_V(u) = e^{iu'm - \frac{1}{2}u'\Gamma u}.$$

En développant, on a donc :

$$\Phi_V(u) = \exp\left(i(u_1 - u_2 + u_3) - \frac{1}{2}(u_1^2 + 2u_1u_2 + 2u_1u_3 + 2u_2^2 + 4u_2u_3 + 3u_3^2)\right).$$

5. L'espérance conditionnelle de Z sachant (X, Y) est :

$$\mathbb{E}[Z|X, Y] = \mathbb{E}[Z] + \Gamma_{Z, (X, Y)}\Gamma_{(X, Y)}^{-1}([X, Y]' - \mu).$$

Or $\Gamma_{(X, Y)} = \Sigma$, donc son inverse a déjà été calculée. On obtient alors :

$$\mathbb{E}[Z|X, Y] = Y + 2.$$

6. Par le même raisonnement que ci-dessus, on en déduit que la variable aléatoire $T = Z - \mathbb{E}[Z|X, Y] = Z - Y - 2$ est indépendante de X et de Y . La variance de T est le résidu quadratique, c'est-à-dire :

$$s^2 = \text{Var}(Z) - \Gamma_{Z, (X, Y)}\Gamma_{(X, Y)}^{-1}\Gamma_{(X, Y), Z} = 1.$$

7. On sait qu'on a la décomposition :

$$X = \mathbb{E}[X|Y, Z] + W,$$

avec $W \sim \mathcal{N}(0, v)$ indépendante du couple (Y, Z) . Or le même calcul que ci-dessus donne :

$$\mathbb{E}[X|Y, Z] = \frac{1}{2}Y + \frac{3}{2},$$

et la variance v de W est le résidu quadratique : $v = \frac{1}{2}$. Sachant que $Y = 1$ et $Z = 2$, la variable aléatoire X suit donc une loi $\mathcal{N}(2, 1/2)$.

IV. Modèle épidémiologique

1. On suppose que la population ne compte que 3 individus et que le taux de contact infectieux est $p = 1/3$.

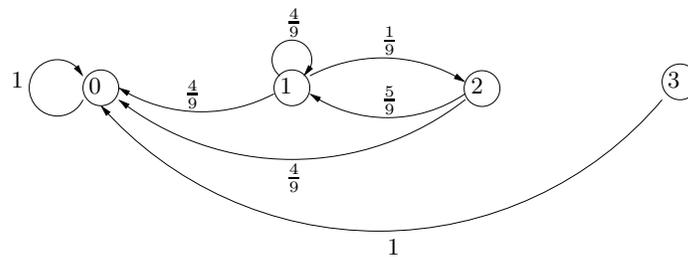


FIGURE A.4 – Graphe de transition pour le modèle épidémiologique

- (a) – Si $X_n = 0$, personne n'est infecté le jour n , donc le jour suivant non plus. Ainsi $p_{00} = 1$ et $p_{0j} = 0$ pour tout $j \in \{1, 2, 3\}$. En d'autres termes, l'état $\{0\}$ est absorbant.
- Sachant $X_n = 1$: puisque la personne infectée le jour n est guérie le lendemain, on a nécessairement $X_{n+1} \in \{0, 1, 2\}$. Plus précisément, la probabilité que $X_{n+1} = 0$ est la probabilité que la personne infectée n'ait rencontré aucune des deux autres personnes, ce qui arrive avec probabilité $p_{10} = 2/3 \times 2/3 = 4/9$; la probabilité que $X_{n+1} = 2$ est la probabilité que la personne infectée ait rencontré chacune des deux autres personnes, ce qui arrive avec probabilité $p_{12} = 1/3 \times 1/3 = 1/9$; par suite, la probabilité que $X_{n+1} = 1$ est égale à :

$$p_{11} = 1 - (p_{10} + p_{12}) = \frac{4}{9}.$$

- Sachant $X_n = 2$, puisque les personnes infectées le jour n sont guéries le lendemain, on a nécessairement $X_{n+1} \in \{0, 1\}$. Plus précisément, la probabilité que $X_{n+1} = 0$ est la probabilité que la personne saine ne rencontre aucune des deux personnes infectées, ce qui arrive avec probabilité $p_{20} = 2/3 \times 2/3 = 4/9$; la probabilité que $X_{n+1} = 1$ est donc $p_{21} = 1 - p_{20} = 5/9$.
- Sachant $X_n = 3$, on a forcément $X_{n+1} = 0$.

On en déduit la matrice de transition :

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{4}{9} & \frac{4}{9} & \frac{1}{9} & 0 \\ \frac{4}{9} & \frac{5}{9} & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Le graphe de transition est donné figure [A.4](#).

- (b) La chaîne est indécomposable : il y a un état récurrent $\{0\}$ et trois états transitoires $\{1, 2, 3\}$.
- (c) Puisque la chaîne est indécomposable, il y a unicité de la loi stationnaire π . De plus, celle-ci est nulle pour les états transitoires, donc :

$$\pi = [1, 0, 0, 0].$$

On pouvait aussi obtenir ce résultat en résolvant le système linéaire $\pi P = \pi$. Ceci signifie que, quelque soit le nombre de personnes initialement infectées, on est presque sûr qu'au bout d'un certain temps, toute la population sera saine.

2. On suppose toujours avoir une population de 3 individus, mais le taux de contact infectieux est le paramètre $p \in]0, 1[$. En notant $q = (1 - p)$, on obtient la matrice de transition :

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ q^2 & 2pq & p^2 & 0 \\ q^2 & 1 - q^2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Le raisonnement fait ci-dessus tient toujours donc l'unique loi stationnaire est encore $\pi = [1, 0, 0, 0]$.

3. On suppose maintenant une population de N individus, avec taux de contact infectieux $p \in]0, 1[$. Sachant que $X_n = i$, on a forcément $X_{n+1} \in \{0, \dots, N - i\}$, ce qui explique l'indicatrice dans la formule.

On remarque aussi que la probabilité qu'une personne saine un jour le soit encore le lendemain est égale à la probabilité qu'elle ne rencontre aucune des i personnes infectées, ce qui arrive avec probabilité q^i , puisque les rencontres sont indépendantes.

Soit alors $j \in \{0, \dots, N - i\}$: la probabilité de transition p_{ij} est la probabilité que parmi les $(N - i)$ personnes saines, j soient infectées. Puisque les rencontres sont indépendantes et que l'infection arrive avec probabilité $1 - q^i$, c'est exactement la probabilité qu'une loi binômiale $\mathcal{B}(N - i, 1 - q^i)$ prenne la valeur j . Ceci donne bien :

$$p_{ij} = \binom{N - i}{j} (1 - q^i)^j q^{i(N - i - j)} \mathbb{1}_{\{i + j \leq N\}}.$$

Examen de Probabilités

I. Couple aléatoire (5 points)

Soit (X, Y) un couple aléatoire de densité jointe :

$$f(x, y) = cx(y - x)e^{-y} \mathbb{1}_{\{0 < x \leq y\}}.$$

1. Soit V une variable aléatoire qui suit une loi exponentielle de paramètre λ . Rappeler son moment d'ordre n , c'est-à-dire $\mathbb{E}[V^n]$.
2. Déterminer c pour que f soit effectivement une densité.
3. Calculer $f(x|y)$, densité conditionnelle de X sachant $Y = y$.
4. En déduire que $\mathbb{E}[X|Y] = Y/2$.
5. Calculer $f(y|x)$, densité conditionnelle de Y sachant $X = x$.
6. En déduire que $\mathbb{E}[Y|X] = X + 2$.
7. Déduire des questions 4 et 6 les quantités $\mathbb{E}[X]$ et $\mathbb{E}[Y]$.

II. Couple gaussien (5 points)

Soit $V = [X, Y]'$ un vecteur gaussien centré de matrice de covariance :

$$\Gamma = \begin{bmatrix} \sigma_X^2 & r\sigma_X\sigma_Y \\ r\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

1. Rappeler à quelle condition V admet une densité $f(x, y)$ et dans ce cas donner sa formule en fonction de σ_X , σ_Y et r . On suppose cette condition vérifiée dans toute la suite.
2. Donner l'expression de l'espérance conditionnelle $\mathbb{E}[Y|X]$ en fonction de σ_X , σ_Y et r .
3. On considère un vecteur $V = [X, Y]'$ ayant pour densité :

$$\forall (x, y) \in \mathbb{R}^2 \quad f(x, y) = \frac{1}{4\pi} e^{-\frac{x^2}{8} - \frac{5y^2}{8} + \frac{xy}{4}}$$

- (a) Déterminer sa matrice de covariance Γ .
- (b) Sachant $x = 1$, quelle est la loi de Y ?
- (c) Sachant $x = 1$, en déduire un intervalle dans lequel se situe Y avec 95% de chances.

III. Espérance conditionnelle et indépendance (6 points)

On considère trois variables aléatoires U , V et W indépendantes et suivant la même loi normale centrée réduite $\mathcal{N}(0, 1)$.

1. On définit le vecteur aléatoire $[X_1, X_2, X_3]'$ comme suit :

$$\begin{cases} X_1 = U \\ X_2 = U + V \\ X_3 = U + V + W \end{cases}$$

- Quelle est la loi du vecteur $[X_1, X_2, X_3]'$?
- Déterminer $\mathbb{E}[X_3|X_1, X_2]$. Comparer à $\mathbb{E}[X_3|X_2]$.
- Expliquer le résultat de la question précédente par un dessin.
- Soit $[X, Y, Z]$ un vecteur aléatoire. Commenter l'affirmation :

$$\mathbb{E}[Z|X, Y] = \mathbb{E}[Z|Y] \Rightarrow Z \text{ est indépendante de } X.$$

2. On définit cette fois le vecteur aléatoire $[X_1, X_2, X_3]'$ comme suit :

$$\begin{cases} X_1 = U \\ X_2 = U + V \\ X_3 = V \end{cases}$$

- Quelle est la loi du vecteur $[X_1, X_2, X_3]'$?
- Déterminer $\mathbb{E}[X_3|X_1, X_2]$. Comparer à $\mathbb{E}[X_3|X_2]$.
- Expliquer le résultat de la question précédente par un dessin.
- Soit $[X, Y, Z]$ un vecteur aléatoire. Commenter l'affirmation :

$$Z \text{ est indépendante de } X \Rightarrow \mathbb{E}[Z|X, Y] = \mathbb{E}[Z|Y].$$

IV. Pile ou Face (4 points)

On joue une suite infinie de Pile ou Face non biaisés : ceci fournit une suite de variables aléatoires $(X_n)_{n \geq 0}$ indépendantes et identiquement distribuées avec $\mathbb{P}(X_n = P) = \mathbb{P}(X_n = F) = 1/2$. A partir de cette suite on considère la chaîne de Markov $(Y_n)_{n \geq 1}$ définie par : $Y_1 = (X_0, X_1)$, $Y_2 = (X_1, X_2)$, et de façon générale $Y_n = (X_{n-1}, X_n)$ pour tout $n \geq 1$. Cette chaîne est donc à valeurs dans l'espace d'états :

$$E = \{E_1, E_2, E_3, E_4\} = \{(P, F), (F, F), (F, P), (P, P)\}.$$

- Donner la matrice et le graphe de transition de (Y_n) .
- La chaîne est-elle irréductible, apériodique ?
- Loi(s) stationnaire(s) ?
- Retrouver le résultat de la question précédente en calculant directement la loi de Y_n :

$$\mathbb{P}(Y_n) = [\mathbb{P}(Y_n = E_1), \mathbb{P}(Y_n = E_2), \mathbb{P}(Y_n = E_3), \mathbb{P}(Y_n = E_4)].$$

Examen de Probabilités

Corrigé

I. Couple aléatoire

Soit (X, Y) un couple aléatoire de densité jointe :

$$f(x, y) = cx(y - x)e^{-y} \mathbb{1}_{\{0 < x \leq y\}}.$$

1. Si $V \sim \mathcal{E}(\lambda)$, alors $\mathbb{E}[V^n] = n!/\lambda^n$.
2. La constante c doit être positive pour que f soit positive. Plus précisément, il faut que l'intégrale double de f sur \mathbb{R}^2 soit égale à 1. Or le théorème de Fubini-Tonelli permet d'écrire :

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = \int_0^{+\infty} \left(\int_0^y cx(y - x)e^{-y} dx \right) dy,$$

ce qui donne :

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = c \int_0^{+\infty} \left[y \frac{x^2}{2} - \frac{x^3}{3} \right]_0^y e^{-y} dy = \frac{c}{6} \int_0^{+\infty} y^3 e^{-y} dy,$$

et on retrouve le moment d'ordre 3 d'une loi exponentielle de paramètre 1 :

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = \frac{c}{6} \mathbb{E}[V^3] = c.$$

Il faut donc que c soit égale à 1 pour que f soit une densité de probabilité.

3. Pour déterminer $f(x|y)$, on commence par calculer la loi marginale de Y :

$$f(y) = \int_{\mathbb{R}} f(x, y) dx = e^{-y} \mathbb{1}_{\{y > 0\}} \int_0^y x(y - x) dx = \frac{y^3}{6} e^{-y} \mathbb{1}_{\{y > 0\}},$$

c'est-à-dire que $Y \sim \Gamma(4, 1)$. On a alors :

$$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{6x(y - x)}{y^3} \mathbb{1}_{\{0 < x \leq y\}}.$$

4. Ceci permet de calculer l'espérance de X sachant $Y = y > 0$:

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}} xf(x|y) dx = \frac{6}{y^3} \int_0^y x^2(y - x) dx = \frac{y}{2},$$

d'où on déduit de façon plus générale que $\mathbb{E}[X|Y] = Y/2$.

5. La densité de X est :

$$f(x) = \int_{\mathbb{R}} f(x, y) dy = x \mathbb{1}_{\{x>0\}} \int_x^{+\infty} (y-x)e^{-y} dy = x \mathbb{1}_{\{x>0\}} \int_0^{+\infty} ue^{-u-x} du,$$

grâce au changement de variable $u = y - x$, d'où :

$$f(x) = xe^{-x} \mathbb{1}_{\{x>0\}} \int_0^{+\infty} ue^{-u} du = xe^{-x} \mathbb{1}_{\{x>0\}} \mathbb{E}[V] = xe^{-x} \mathbb{1}_{\{x>0\}},$$

c'est-à-dire que $X \sim \Gamma(2, 1)$. On a donc :

$$f(y|x) = \frac{f(x, y)}{f(x)} = (y-x)e^{-(y-x)} \mathbb{1}_{\{0 < x \leq y\}}.$$

Autrement dit, sachant $X = x$, Y suit une loi $\Gamma(2, 1)$ translatée sur l'intervalle $[x, +\infty[$.

6. Pour tout $x > 0$, on a donc $\mathbb{E}[Y|X = x] = x + \mathbb{E}[\Gamma(2, 1)] = x + 2$, résultat que l'on peut retrouver par le calcul :

$$\mathbb{E}[Y|X = x] = \int_x^{+\infty} y(y-x)e^{-(y-x)} dy = \int_0^{+\infty} (u+x)ue^{-u} du,$$

qu'on sépare en deux intégrales :

$$\mathbb{E}[Y|X = x] = \int_0^{+\infty} u^2 e^{-u} du + x \int_0^{+\infty} ue^{-u} du,$$

et on reconnaît les moments d'une loi exponentielle de paramètre 1 :

$$\mathbb{E}[Y|X = x] = \mathbb{E}[V^2] + x\mathbb{E}[V] = 2 + x.$$

On en déduit que $\mathbb{E}[Y|X] = X + 2$.

7. Des questions 4 et 6, en prenant les espérances, on tire le système d'équations linéaires :

$$\begin{cases} \mathbb{E}[X] = \mathbb{E}[Y]/2 \\ \mathbb{E}[Y] = \mathbb{E}[X] + 2 \end{cases} \Leftrightarrow \begin{cases} \mathbb{E}[X] = 2 \\ \mathbb{E}[Y] = 4 \end{cases}$$

Vérification : rappelons que si $T \sim \Gamma(n, \lambda)$, alors $\mathbb{E}[T] = n\lambda$. Or on a vu plus haut que $X \sim \Gamma(2, 1)$ et $Y \sim \Gamma(4, 1)$, donc on retrouve bien $\mathbb{E}[X] = 2$ et $\mathbb{E}[Y] = 4$.

II. Couple gaussien

Soit $V = [X, Y]'$ un vecteur gaussien centré de matrice de covariance :

$$\Gamma = \begin{bmatrix} \sigma_X^2 & r\sigma_X\sigma_Y \\ r\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

1. Le vecteur gaussien V admet une densité sur \mathbb{R}^2 si et seulement si sa matrice de dispersion est inversible, c'est-à-dire ssi :

$$\det\Gamma = (1 - r^2)\sigma_X^2\sigma_Y^2 \neq 0.$$

Dans ce cas, on a :

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-r^2}} e^{-\frac{\sigma_Y^2 x^2 - 2r\sigma_X\sigma_Y xy + \sigma_X^2 y^2}{2\sigma_X^2\sigma_Y^2(1-r^2)}}$$

2. L'espérance conditionnelle de Y sachant X est :

$$\mathbb{E}[Y|X] = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}X = \frac{r\sigma_X\sigma_Y}{\sigma_X^2}X = \frac{r\sigma_Y}{\sigma_X}X.$$

3. On considère un vecteur gaussien centré $V = [X, Y]'$ ayant pour densité :

$$f(x, y) = \frac{1}{4\pi}e^{-\frac{x^2}{4} + \frac{5y^2}{4} - \frac{xy}{2}}$$

(a) Par identification des coefficients avec ceux de la question 1, on obtient :

$$\Gamma = \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}.$$

Autrement dit le coefficient de corrélation linéaire r vaut $1/\sqrt{5}$.

(b) On a la décomposition : $Y = \mathbb{E}[Y|X] + W$, avec W normale centrée indépendante de X et de variance :

$$\sigma^2 = \sigma_Y^2(1 - r^2) = \frac{4}{5}.$$

Par ailleurs on a $\mathbb{E}[Y|X] = \frac{X}{5}$. Donc sachant $x = 1$, Y suit une loi normale $\mathcal{N}(\frac{1}{5}, \frac{4}{5})$.

(c) On sait que, avec probabilité 0.95, une variable aléatoire suivant une loi normale tombe à une distance inférieure à deux fois l'écart-type par rapport à sa moyenne. On en déduit que, sachant $x = 1$, on a avec 95% de chances :

$$Y \in \left[\frac{1}{5} - \frac{4}{\sqrt{5}}, \frac{1}{5} + \frac{4}{\sqrt{5}} \right].$$

III. Espérance conditionnelle et indépendance

On considère trois variables aléatoires U , V et W indépendantes et suivant la même loi normale centrée réduite $\mathcal{N}(0, 1)$.

1. On définit le vecteur aléatoire $[X_1, X_2, X_3]'$ comme suit :

$$\begin{cases} X_1 = U \\ X_2 = U + V \\ X_3 = U + V + W \end{cases}$$

(a) Le vecteur $[X_1, X_2, X_3]'$ est gaussien en tant que transformée linéaire d'un vecteur gaussien :

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} U \\ V \\ W \end{bmatrix} \Rightarrow \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} \right).$$

(b) On en déduit l'espérance conditionnelle de X_3 sachant X_1 et X_2 :

$$\mathbb{E}[X_3|X_1, X_2] = \Gamma_{X_3, (X_1, X_2)} \Gamma_{X_1, X_2}^{-1} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = X_2.$$

On trouve aussi :

$$\mathbb{E}[X_3|X_2] = \frac{\text{Cov}(X_2, X_3)}{\text{Var}(X_2)}X_2 = X_2,$$

c'est-à-dire que : $\mathbb{E}[X_3|X_1, X_2] = \mathbb{E}[X_3|X_2]$. Noter qu'on peut retrouver ces résultats en considérant directement U, V et W :

$$\mathbb{E}[X_3|X_1, X_2] = \mathbb{E}[U+V+W|U, U+V] = U+V+\mathbb{E}[W|U, U+V] = U+V+\mathbb{E}[W] = U+V,$$

puisque W est indépendante de U et V . De même :

$$\mathbb{E}[X_3|X_2] = \mathbb{E}[U+V+W|U+V] = U+V+\mathbb{E}[W|U+V] = U+V.$$

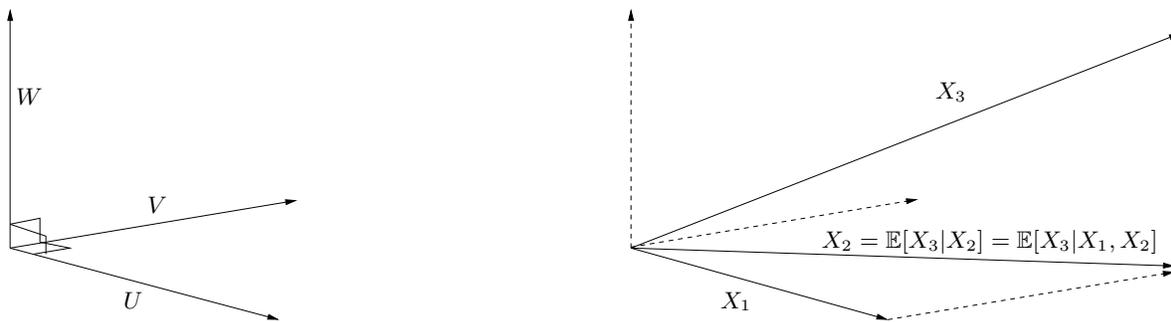


FIGURE A.5 – Interprétation géométrique de l'égalité $\mathbb{E}[X_3|X_1, X_2] = \mathbb{E}[X_3|X_2]$.

- (c) Puisque les variables sont centrées et de carrés intégrables, l'espérance conditionnelle de X_3 sachant (X_1, X_2) correspond à la projection orthogonale de X_3 sur le plan vectoriel engendré par X_1 et X_2 . La figure A.5 explique alors pourquoi dans notre cas particulier on a l'égalité : $\mathbb{E}[X_3|X_1, X_2] = \mathbb{E}[X_3|X_2]$.
- (d) Cette affirmation est fausse puisque si on prend $[X, Y, Z] = [X_1, X_2, X_3]$, X_1 et X_3 ne sont pas indépendantes (cf. $\text{Cov}(X_1, X_3) = 1$) mais $\mathbb{E}[X_3|X_1, X_2] = \mathbb{E}[X_3|X_2]$.
2. On définit cette fois le vecteur aléatoire $[X_1, X_2, X_3]'$ comme suit :

$$\begin{cases} X_1 = U \\ X_2 = U + V \\ X_3 = V \end{cases}$$

- (a) Le vecteur $[X_1, X_2, X_3]'$ est gaussien lui aussi :

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \right).$$

- (b) On a cette fois :

$$\mathbb{E}[X_3|X_1, X_2] = X_2 - X_1 = X_3 \neq \mathbb{E}[X_3|X_2] = \frac{X_2}{2}.$$

- (c) On voit que $X_3 = (U + V) - U = X_2 - X_1$, donc X_3 appartient au plan vectoriel engendré par X_1 et X_2 . Par conséquent sa projection orthogonale sur ce plan est lui-même : ceci explique géométriquement l'égalité $\mathbb{E}[X_3|X_1, X_2] = X_3$. Pour la relation $\mathbb{E}[X_3|X_2] = \frac{X_2}{2}$, voir figure A.6.
- (d) Cette affirmation est fausse aussi puisque si on prend $[X, Y, Z] = [X_1, X_2, X_3]$, $X_1 = U$ et $X_3 = V$ sont indépendantes mais $\mathbb{E}[X_3|X_1, X_2] \neq \mathbb{E}[X_3|X_2]$.

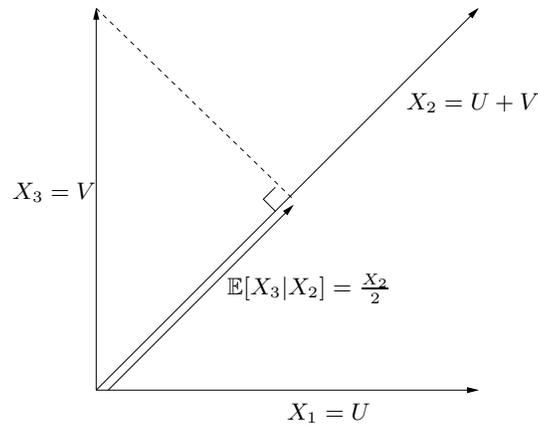


FIGURE A.6 – Interprétation géométrique de l'inégalité $\mathbb{E}[X_3|X_1, X_2] = X_3 \neq \mathbb{E}[X_3|X_2] = \frac{X_2}{2}$.

IV. Pile ou Face

La chaîne de Markov $(Y_n)_{n \geq 1}$ est à valeurs dans l'espace d'états :

$$E = \{E_1, E_2, E_3, E_4\} = \{(P, F), (F, F), (F, P), (P, P)\}.$$

1. La matrice de transition de la chaîne est :

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix},$$

et le graphe de transition est représenté figure A.7.

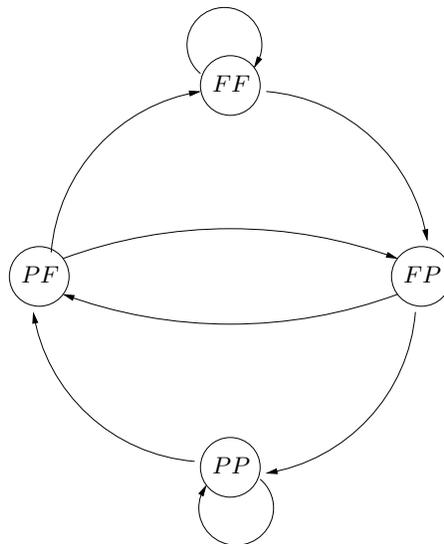


FIGURE A.7 – Graphe de transition pour le jeu de Pile ou Face.

2. La chaîne est irréductible puisque tous les états communiquent entre eux. Elle est aussi apériodique, puisque $E_2 = (F, F)$ l'est (on peut boucler sur cet état) et que la chaîne est irréductible.

3. On trouve pour unique loi stationnaire la loi uniforme sur les quatre états :

$$\pi = \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right],$$

ou bien en résolvant le système d'équations $\pi P = \pi$, ou bien en remarquant que la matrice P est bistochastique.

4. Puisque $Y_n = (X_{n-1}, X_n)$, avec X_{n-1} et X_n indépendantes, on a par exemple :

$$\mathbb{P}(Y_n = E_1) = \mathbb{P}((X_{n-1}, X_n) = (P, F)) = \mathbb{P}(X_{n-1} = P)\mathbb{P}(X_n = F) = \frac{1}{4},$$

et même chose pour E_2 , E_3 et E_4 . Ainsi, pour tout $n \geq 1$, on a : $\mathbb{P}(Y_n) = \pi$. En particulier $\mathbb{P}(Y_1) = \mathbb{P}(Y_2) = \pi$. Mais par définition de la matrice de transition P , on sait qu'on a aussi $P(Y_2) = \mathbb{P}(Y_1)P$, c'est-à-dire que $\pi P = \pi$, et π est donc une loi d'équilibre de la chaîne (Y_n) . Puisque cette chaîne est irréductible, c'est même l'unique loi d'équilibre. De plus, le calcul de la loi de Y_n montre que cette chaîne est tout le temps à l'équilibre.

Examen de Probabilités

I. Couple aléatoire

On considère le couple aléatoire (X, Y) de densité la fonction f définie sur \mathbb{R}^2 par :

$$f(x, y) = \begin{cases} 2 & \text{si } 0 \leq x \leq y \leq 1 \\ 0 & \text{sinon} \end{cases}$$

1. Représenter f et vérifier qu'il s'agit bien d'une fonction de densité.
2. Les variables X et Y sont-elles indépendantes ?
3. Déterminer les lois marginales puis calculer $\mathbb{E}[X]$ et $\mathbb{E}[Y]$.
4. Calculer les lois conditionnelles $f(y|x)$ et $f(x|y)$.
5. Calculer $\mathbb{E}[X|Y]$.
6. En utilisant le résultat précédent et la valeur de $\mathbb{E}[Y]$ de la question 3, retrouver la valeur de $\mathbb{E}[X]$.

II. Lois exponentielles

Soit X et Y deux variables aléatoires indépendantes suivant une loi exponentielle de même paramètre $\lambda > 0$.

1. Quelle est la loi jointe $f_{X,Y}$ du couple (X, Y) ?
2. Déterminer la loi jointe $f_{V,W}$ du couple (V, W) défini par :

$$\begin{cases} V & = & X + Y \\ W & = & X \end{cases}$$

3. En déduire la densité de V .
4. Calculer $f(w|v)$. Quelle loi reconnaît-on ?

III. Vecteur gaussien

Soit $X = [X_1, X_2, X_3]'$ un vecteur gaussien centré et de matrice de covariance :

$$\Gamma = \begin{bmatrix} 4 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

1. Quelles sont les lois marginales de X ?

2. Certaines composantes de X sont-elles indépendantes deux à deux ? Si oui, lesquelles ?
Certaines composantes de X sont-elles orthogonales deux à deux ? Si oui, lesquelles ?
3. Quelle est la loi de (X_1, X_2) ?
4. Sans calcul, déterminer $\mathbb{E}[X_2|X_3]$ et $\mathbb{E}[(X_2 - \mathbb{E}[X_2|X_3])^2]$.
5. Calculer $\mathbb{E}[X_1|X_3]$ et $\mathbb{E}[(X_1 - \mathbb{E}[X_1|X_3])^2]$.
6. Calculer $\mathbb{E}[X_1|X_2, X_3]$ et $\mathbb{E}[(X_1 - \mathbb{E}[X_1|X_2, X_3])^2]$.
7. Quelle est la loi de X_1 sachant $(X_2 = x_2, X_3 = x_3)$?
8. On pose $S = \mathbb{E}[X_1^2|X_2, X_3]$ et \hat{S} le résultat de la régression linéaire de X_1^2 sur (X_2, X_3) .
Justifier que $\mathbb{E}[(X_1^2 - S)^2] \leq \mathbb{E}[(X_1^2 - \hat{S})^2]$.

IV. Chaîne de Markov

Soit l'espace d'états $E = \{1, 2, 3, 4\}$ d'une chaîne de Markov homogène.

1. Compléter la matrice suivante pour qu'elle soit une matrice de transition :

$$P = \begin{bmatrix} \cdot & \frac{1}{2} & 0 & 0 \\ \cdot & \frac{2}{3} & 0 & 0 \\ 0 & \cdot & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \cdot & 0 & \frac{1}{2} \end{bmatrix}$$

2. Représenter le graphe associé.
3. La chaîne est-elle irréductible ? Indécomposable ?
4. Déterminer la (ou les) loi(s) stationnaire(s).
5. On considère qu'au temps 0, on est dans l'état 3. Pour un grand nombre d'unité de temps n , quelles sont les probabilités qu'on soit dans chacun des quatre états ?

Examen de Probabilités

Corrigé

I. Couple aléatoire

1. La densité est représentée en gras sur la figure suivante :

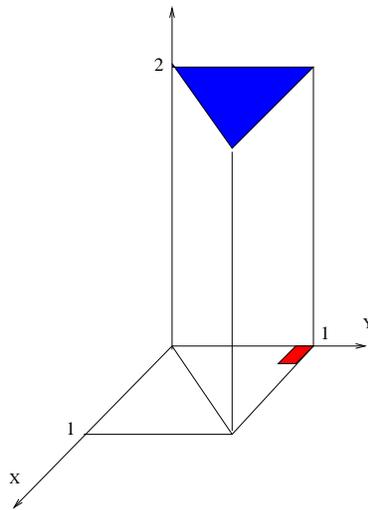


FIGURE A.8 – Représentation de la densité.

f est bien une densité car $f \geq 0$ et $\int_{\mathbb{R}^2} f(x, y) dx dy$, qui correspond au volume représenté, vaut 1.

2. Les variables X et Y ne sont pas indépendantes. En effet, le support de (X, Y) est triangulaire alors que le produit cartésien des supports de X et Y est $[0, 1]^2$.
3. Soit $0 \leq x \leq 1$, on a :

$$f(x) = \int_{\mathbb{R}} f(x, y) dy = \int_x^1 2 dy = 2(1 - x).$$

On calcule ainsi :

$$\mathbb{E}[X] = \int_0^1 2x(1 - x) dx = \frac{1}{3}.$$

Soit maintenant $0 \leq y \leq 1$, on a :

$$f(y) = \int_{\mathbb{R}} f(x, y) dx = \int_0^y 2 dx = 2y.$$

Ainsi :

$$\mathbb{E}[Y] = \int_0^1 2y^2 dy = \frac{2}{3}.$$

4. Soit $x \in [0, 1]$ fixé. Pour $x \leq y \leq 1$, on a :

$$f(y|x) = \frac{f(x, y)}{f(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x}.$$

Finalement :

$$f(y|x) = \frac{1}{1-x} \mathbb{1}_{\{x \leq y \leq 1\}}(y),$$

c'est-à-dire que, sachant $X = x$, Y suit une loi uniforme sur $[x, 1]$.

Soit maintenant $y \in [0, 1]$ fixé. Pour tout $0 \leq x \leq y$, on a :

$$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{2}{2y} = \frac{1}{y}.$$

Autrement dit, sachant $Y = y$, X suit une loi uniforme sur $[0, y]$.

5. Pour y fixé dans $[0, 1]$, calculons d'abord $\mathbb{E}[X|Y = y]$:

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}} x \frac{1}{y} \mathbb{1}_{\{0 \leq x \leq y\}}(x) dx = \frac{1}{y} \int_0^y x dx = \frac{y}{2}.$$

Ainsi $\mathbb{E}[X|Y] = \frac{Y}{2}$. Remarquons qu'on avait directement ce résultat en utilisant la question précédente (moyenne d'une loi uniforme).

6. On sait que $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$. Or la question précédente donne $\mathbb{E}[X] = \frac{\mathbb{E}[Y]}{2}$ et on retrouve :

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

II. Lois exponentielles

1. X et Y sont indépendantes, la densité du couple (X, Y) est donc le produit des densités de X et de Y :

$$f_{X,Y}(x, y) = \lambda e^{-\lambda x} \mathbb{1}_{[0, +\infty[}(x) \lambda e^{-\lambda y} \mathbb{1}_{[0, +\infty[}(y) = \lambda^2 e^{-\lambda(x+y)} \mathbb{1}_{[0, +\infty[^2}(x, y).$$

2. Le changement de variable proposé est linéaire et bijectif avec comme bijection réciproque :

$$\begin{cases} X &= W \\ Y &= V - W \end{cases}$$

Le support de (X, Y) est $D = \mathbb{R}_+^2$ ce qui s'écrit pour $(V, W) : \Delta = \{(v, w) \in \mathbb{R}^2 : 0 \leq w \leq v\}$. On introduit le C^1 -difféomorphisme :

$$\phi : \begin{cases} \Delta & \longrightarrow D \\ (v, w) & \longmapsto \begin{cases} x &= w \\ y &= v - w \end{cases} \end{cases}$$

Alors :

$$f_{V,W}(v, w) = f_{X,Y}(\phi(v, w)) |\det J_\phi(v, w)|.$$

On en déduit :

$$f_{V,W}(v, w) = \lambda^2 e^{-\lambda v} \mathbb{1}_{\{0 \leq w \leq v\}}(v, w).$$

3. Soit $v > 0$ fixé, on a :

$$f(v) = \int_{\mathbb{R}} f(v, w) dw = \int_0^v \lambda^2 e^{-\lambda w} dw = \lambda^2 v e^{-\lambda v}.$$

4. On en déduit :

$$f(w|v) = \frac{f(v, w)}{f(v)} = \frac{\lambda^2 e^{-\lambda w} \mathbb{1}_{\{0 \leq w \leq v\}}(v, w)}{\lambda^2 v e^{-\lambda v} \mathbb{1}_{\{0 \leq v\}}(v)} = \frac{1}{v} \mathbb{1}_{\{0 \leq w \leq v\}}(w).$$

Ainsi, sachant $v > 0$ fixé, W suit une loi uniforme sur $[0, v]$.

III. Vecteur gaussien

1. X est gaussien donc toutes ses composantes sont gaussiennes. Elles sont par ailleurs toutes centrées et les variances se lisent sur la diagonale de Γ . On a donc :

$$X_1 \sim \mathcal{N}(0, 4) \quad X_2 \sim \mathcal{N}(0, 2) \quad X_3 \sim \mathcal{N}(0, 1).$$

2. X étant gaussien, l'indépendance de deux variables équivaut à leur décorrélation. Nous avons donc X_2 et X_3 indépendantes. Les composantes étant centrées, la non-corrélation équivaut à l'orthogonalité. Les deux variables précédentes sont donc aussi orthogonales.

3. On peut poser $[X_1, X_2]' = AX$ avec

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Ainsi $[X_1, X_2]'$ est gaussien comme transformation affine d'un vecteur gaussien. Il est bien sûr centré et les éléments de sa matrice de covariance sont donnés dans Γ :

$$[X_1, X_2] \sim \mathcal{N}\left([0, 0]', \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}\right).$$

4. Calculer $\mathbb{E}[X_2|X_3]$ revient à projeter orthogonalement X_2 sur $H = \text{Vect}(1, X_3)$ or X_2 est orthogonale à 1 car c'est une variable centrée et on a vu que X_2 est aussi orthogonale à X_3 . On a donc $\mathbb{E}[X_2|X_3] = 0$ et la variance résiduelle $\mathbb{E}[(X_2 - \mathbb{E}[X_2|X_3])^2]$ est égale à la variance de X_2 , soit $\mathbb{E}[(X_2 - \mathbb{E}[X_2|X_3])^2] = 2$.

5. On obtient :

$$\mathbb{E}[X_1|X_3] = \mathbb{E}[X_1] + \frac{\text{Cov}(X_1, X_3)}{\text{Var}(X_3)}(X_3 - \mathbb{E}(X_3)) = -X_3$$

et

$$\mathbb{E}[(X_1 - \mathbb{E}[X_1|X_3])^2] = \text{Var}(X_1) - \frac{\text{Cov}^2(X_1, X_3)}{\text{Var}(X_3)} = 3.$$

6. On obtient :

$$\begin{aligned} \mathbb{E}[X_1|X_2, X_3] &= \mathbb{E}[X_1] + \Gamma_{X_1, (X_2, X_3)} \Gamma_{X_2, X_3}^{-1} \begin{bmatrix} X_2 - \mathbb{E}[X_2] \\ X_3 - \mathbb{E}[X_3] \end{bmatrix} \\ &= [1, -1] \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} \\ &= [1, -1] \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} \\ &= \frac{1}{2} X_2 - X_3. \end{aligned}$$

De plus :

$$\begin{aligned}\mathbb{E}[(X_1 - \mathbb{E}[X_1|X_2, X_3])^2] &= \Gamma_{X_1} - \Gamma_{X_1, (X_2, X_3)} \Gamma_{X_2, X_3}^{-1} (\Gamma_{X_1, (X_2, X_3)})' \\ &= 4 - [1, -1] \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= 2.5.\end{aligned}$$

7. On a la décomposition orthogonale $X_1 = \mathbb{E}[X_1|X_2, X_3] + W$. La variable W est centrée et de variance :

$$\text{Var}(W) = \mathbb{E}[(X_1 - \mathbb{E}[X_1|X_2, X_3])^2] = 2.5.$$

Par ailleurs, comme W est indépendante de (X_2, X_3) , sachant $(X_2, X_3) = (x_2, x_3)$, on a :

$$X_1 \sim \mathcal{N}(\mathbb{E}[X_1|x_2, x_3], \text{Var}(W)) = \mathcal{N}\left(\frac{1}{2}x_2 - x_3, 2.5\right).$$

8. S est le projeté orthogonal de X_1^2 sur $L^2(X_2, X_3)$, espace des variables aléatoires fonctions de X_2 et X_3 et de carré intégrable. \hat{S} est le projeté orthogonal de X_1^2 sur $H = \text{Vect}(X_2, X_3)$, qui est un sous-espace vectoriel de $L^2(X_2, X_3)$. Dans le cas gaussien, ces deux projetés coïncident mais X_1^2 n'étant pas gaussienne, ce n'est pas le cas ici. On a donc :

$$\mathbb{E}[(X_1^2 - S)^2] \leq \mathbb{E}[(X_1^2 - \hat{S})^2].$$

IV. Chaîne de Markov

1. On complète la matrice de sorte que la somme des lignes soit égale à 1 :

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix}$$

2. L'ensemble $\{1, 2\}$ est clos et irréductible alors que l'ensemble $\{3, 4\}$ est transitoire. La chaîne est donc indécomposable avec :

$$R \cup T = \{1, 2\} \cup \{3, 4\}.$$

3. La chaîne admet une loi stationnaire unique μ , solution de $\mu P = \mu$. Après calcul, on obtient :

$$\mu = \left[\frac{2}{5}, \frac{3}{5}, 0, 0 \right].$$

4. Quelle que soit la loi initiale, on a convergence de la loi de (X_n) vers μ . Ainsi, les probabilités qu'on soit pour n grand dans les états 3 et 4 sont nulles alors que celles qu'on soit dans les états 1 et 2 sont respectivement $2/5$ et $3/5$.

“Si j'étais pas tellement contraint, obligé pour gagner ma vie, je vous le dis tout de suite, je supprimerais tout. Je laisserais pas passer plus une ligne.”

Louis-Ferdinand Céline, *Voyage au bout de la nuit*, préface à la 2ème édition, 1949.

Bibliographie

- [1] Nicolas Bouleau. *Probabilités de l'ingénieur*. Hermann, 2002.
- [2] Yves Ducel. *Introduction à la théorie mathématique des probabilités*. Ellipses, 1998.
- [3] Rick Durrett. *Essentials of Stochastic Processes*. Springer Texts in Statistics. Springer-Verlag, New York, 1999.
- [4] Dominique Foata et Aimé Fuchs. *Calcul des probabilités*. Dunod, 1998.
- [5] Dominique Foata et Aimé Fuchs. *Processus stochastiques*. Dunod, 2002.
- [6] Gilles Pagès et Claude Bouzitat. *En passant par hasard... Les probabilités de tous les jours*. Vuibert, 2000.
- [7] Eva Cantoni, Philippe Huber et Elvezio Ronchetti. *Maîtriser l'aléatoire (Exercices résolus de probabilités et statistique)*. Springer, 2006.
- [8] Pierre-André Cornillon et Eric Matzner-Lober. *Régression avec R*. Springer, Paris, 2010.
- [9] Didier Dacunha-Castelle et Marie Duflo. *Probabilités et statistiques 2. Problèmes à temps mobile*. Masson, 1993.
- [10] Didier Dacunha-Castelle et Marie Duflo. *Probabilités et statistiques 1. Problèmes à temps fixe*. Masson, 1994.
- [11] Philippe Barbe et Michel Ledoux. *Probabilités*. Belin, 1998.
- [12] Guy Auliac, Christiane Coccozza-Thivent, Sophie Mercier et Michel Roussignol. *Exercices de probabilités*. Cassini, 1999.
- [13] Michel Benaïm et Nicole El Karoui. *Promenade aléatoire*. Editions de l'Ecole Polytechnique, 2004.
- [14] Valérie Girardin et Nikolaos Limnios. *Probabilités*. Vuibert, 2001.
- [15] Jean Jacod et Philip Protter. *L'essentiel en théorie des probabilités*. Cassini, 2003.
- [16] Dominique Bakry, Laure Coutin et Thierry Delmotte. *Chaînes de Markov finies*. Format électronique, 2004.
- [17] Marie Cottrell, Valentine Genon-Catalot, Christian Duhamel et Thierry Meyre. *Exercices de probabilités*. Cassini, 1999.
- [18] Jean-Pascal Ansel et Yves Ducel. *Exercices corrigés en théorie des probabilités*. Ellipses, 1996.
- [19] Geoffrey R. Grimmett and David R. Stirzaker. *One Thousand Exercises in Probability*. Oxford University Press, New York, 2001.
- [20] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes*. Oxford University Press, New York, 2001.
- [21] Michel Métivier. *Probabilités : dix leçons d'introduction*. Ellipses, 1987.
- [22] Jacques Neveu. *Probabilités*. Editions de l'Ecole Polytechnique, 1996.
- [23] James R. Norris. *Markov Chains*. Cambridge University Press, 1997.

- [24] Jean-Yves Oувrard. *Probabilités 1*. Cassini, 1998.
- [25] Sheldon M. Ross. *Initiation aux probabilités*. Presses polytechniques et universitaires romandes, 1987.
- [26] Bernard Ycart. *Chaînes de Markov*. Cahiers de Mathématiques Appliquées, CMA 11, 2004.
- [27] Bernard Ycart. *Vecteurs et suites aléatoires*. Cahiers de Mathématiques Appliquées, 2004.