UNIVERSITÉ PARIS DESCARTES
Licence Informatique 2e année

Notes du cours

Probablilités et Statistiques pour l'Informatique.

Version du 21 novembre 2016.

La version la plus récente de ces notes de cours se trouve sur la page moodle
`https://moodle.mi.parisdescartes.fr/course/view.php?id=461`

Table des matières

Chapitre 1

Probabilités, événements

(1 cours)

1.1 Introduction

L'aléatoire existe naturellement dans beaucoup de phénomènes naturels, ou artificiels. L'étude rigoureuse des phénomènes aléatoires repose sur une théorie mathématique, la "théorie des probabilités". Le but de ce cours est de maîtriser les fondements de la théorie des probabilités, et de commencer à comprendre comment l'appliquer rigoureusement aux statistiques.

L'aléatoire est utilisée dans plein de domaines en informatiques :

— Gestion des serveurs où arrivent des requêtes avec des poids aléatoires et à des temps aléatoires / Calibrage des antennes-relais et des serveurs
— Génération de nombres aléatoires, de processus aléatoires (graphisme, jeux vidéos, ...)
— Dilemmes exploitation/exploration (bandits-manchots), Optimisation de l'affichage d'un contenu en fonction des utilisateurs (aléatoires)
— Moteurs de recherche, Web crawlers
— Gestion des interférences entre dispositifs sans fil évoluant au hasard dans l'espace
— Informatique quantique
— Evaluation par chaînes de Markov (Méthode de Monte Carlo)
— Intelligence artificielle, deep learning (Alpha Go, ...)
— Data Mining, Big data, exploration et représentation de données, ...
— ...

1.2 Rappels sur les notations mathématiques

1.2.1 Ensembles

Un ensemble est un regroupement d'éléments. Par exemple, on note $\mathbf{N}$ l'ensemble de tous les entiers $0, 1, 2, \ldots$. Le symbole $\in$ permet de noter l'appartenance à un ensemble : par exemple, "$3 \in \mathbf{N}$" signifie "3 est un élément de $\mathbf{N}$", ou encore "3 appartient à $\mathbf{N}$".

Pour décrire un ensemble à partir de ses éléments, on utilise les accolades. Par exemple,
•$A = \{1, 5, 3\}$: A est l'ensemble formé par les nombres 1, 3, et 5. Bien noter que dans cette notation **l'ordre ne compte pas** : $\{1, 5, 3\} = \{1, 3, 5\} = \{3, 1, 5\}$.
• De plus, un élément apparait au plus une fois dans un ensemble : $\{1, 3, 5\} = \{1, 3, 5, 3\} = \{1, 1, 1, 3, 5\} = ...$
•$\mathbf{N} = \{0, 1, 2, 3, \ldots\}$: $\mathbf{N}$ est l'ensemble des nombres 0, 1, 2, etc.
•$B = \{0, 1, 4, 9, \ldots\}$: B est l'ensemble des carrés des entiers.

Souvent on utilise des expressions logiques entre les accolades pour décrire un ensemble plus précisément. Par exemple, pour le dernier ensemble B, on écrira plutôt :

$$B = \{n^2, n \in \mathbf{N}\},$$

ce qui se lit : "B est l'ensemble des n^2 **tels que** n appartient à $\mathbf{N}$".
Ou encore :

$$B = \{n \in \mathbf{N}, n = k^2, k \in \mathbf{N}\},$$

c'est-à -dire : "B est l'ensemble des entiers n **tels que** $n = k^2$, où k peut être n'importe quel entier". Dans une telle notation la première virgule fait office de "tel que", tandis que les suivantes se lisent "et" ou "avec". On utilise parfois deux points " :", ou point-virgule " ;",

$$B = \{n \in \mathbf{N}, n = k^2, k \in \mathbf{N}\} = \{n \in \mathbf{N} : n = k^2, k \in \mathbf{N}\} = \{n \in \mathbf{N}; n = k^2, k \in \mathbf{N}\}.$$

A noter aussi que les lettres utilisées (n, k) dans les expressions sont complètement interchangeables. On aurait pu aussi bien écrire

$$B = \{k \in \mathbf{N}, k = n^2, n \in \mathbf{N}\},$$

ou encore

$$B = \{i \in \mathbf{N}, i = j^2, j \in \mathbf{N}\}.$$

Autre exemple : $C = \{x \in \mathbf{R}, x \leq 3, x \geq 2\}$: ensemble des nombres réels supérieurs à 2 et inférieurs à 3 : c'est l'intervalle $[2, 3]$.

L'**ensemble vide** ne contient aucun élément. Il est noté $\varnothing$. On a par exemple $\{x \in \mathbf{R}, x \leq 2, x \geq 3\} = \varnothing$.

Le **cardinal** d'un ensemble est le nombre de ses éléments. Il est noté "Card" ou "$\sharp$". Par exemple,

- $\text{Card}\{1,5,3\} = \sharp\{1,5,3\} = 3$,
- $\text{Card}\{1,5,3,1\} = 3$,
- $\sharp(\varnothing) = 0$.

Lorsqu'un ensemble n'a qu'un seul élément, comme par exemple l'ensemble $\{4\}$, on l'appelle **singleton**. Attention, il est essentiel de distinguer éléments et ensembles : on écrit $4 \in \{4\}$ (le nombre 4 appartient à l'ensemble $\{4\}$) mais surtout pas $\{4\} = 4$!

1.2.2 Ensembles finis, dénombrables, non dénombrables

Un ensemble est dit **fini** lorsqu'il possède un nombre fini d'éléments. Par exemple $\{1,5,3\}$ (3 éléments), ou encore, si n est n'importe quel entier, $\{1,2,\ldots,n\}$ (n éléments). Sinon c'est un ensemble infini.

On sait que les ensembles $\mathbb{R}$ et $\mathbb{N}$ sont infinis, mais $\mathbb{R}$ est "plus infini". Ainsi, on peut affecter chaque élément de $\mathbb{N}$ à un élément de $\mathbb{R}$ (lui-même...), mais on ne peut pas faire le contraire.

Un ensemble est dit **dénombrable** s'il est infini mais que l'on peut énumérer ses éléments, c'est-à -dire attribuer un numéro unique à chacun de ses éléments, ou encore en termes mathématiques : trouver une bijection de $\mathbf{N}$ vers cet ensemble. Par exemple,

- $\mathbf{N}$ est dénombrable (puisque $\mathbf{N} = \{0,1,2,3,4,\ldots\}$)
- $\mathbf{Z}$ est dénombrable ($\mathbf{Z} = \{0,1,-1,2,-2,3,-3,4,-4,\ldots\}$.)
- $\mathbf{Q}$ est dénombrable (mais la numérotation n'est pas évidente).

On ne vous demandera pas de savoir que d'autres ensembles sont dénombrables. A noter : **Un sous-ensemble d'un ensemble dénombrable est fini ou dénombrable !** Par exemple, l'ensemble de tous les nombres pairs est dénombrable car c'est un sous-ensemble de $\mathbb{Z}$ infini.

Enfin, quelques exemples d'ensembles **infinis non dénombrables** : $\mathbf{R}$, $[0,1]$, $\mathbf{R} \setminus \mathbf{Q}$. Cela veut dire qu'on ne peut pas trouver une manière d'"énumérer" $\mathbb{R}$:

$$\mathbb{R} = \{0,1,2,\ldots,1/2,1/3,1/4,\ldots,2/3,2/5,\ldots,\pi,e,\sqrt{2},\sqrt{3},\ldots\}$$

On n'aura jamais tous les éléments qui "rentrent" dans une telle écriture (dur à montrer !).

1.2.3 Sous-ensembles

On dit qu'un ensemble E est une **partie** d'un autre ensemble F lorsque tous les éléments de E appartiennent aussi à F. On dit encore que E est un **sous-ensemble** de F ou que E est **inclus** dans F, et on note $E \subset F$. Par exemple, $\{3,5\} \subset \{1,5,3\}$, ou bien $\{5\} \subset \{1,5,3\}$. (Mais surtout pas $5 \subset \{1,5,3\}$: 5 est un élément, pas un ensemble).

Par convention, $\varnothing$ est toujours sous-ensemble de n'importe quel ensemble.

L'**union** de E et F est l'ensemble des éléments appartenant à E **ou** à F, noté $E \cup F$. On a ainsi $E \cup F = \{x, x \in E \text{ ou } x \in F\}$.

L'**intersection** de E et F est l'ensemble des éléments appartenant à E **et** à F, noté $E \cap F$. On a ainsi $E \cap F = \{x, x \in E, x \in F\}$.

Si l'on ne considère que des sous-ensembles d'un même ensemble Ω (ce sera le cas par la suite), on peut parler d'ensemble **complémentaire** d'un sous-ensemble E de Ω : il regroupe tous les éléments de Ω ne faisant pas partie de E, et on le note E^c. Par exemple si $\Omega = \mathbf{R}$, $[2,4]^c =]-\infty, 2[\cup]4, \infty[$.

Quelques propriétés : (rappelez-vous du calcul booléen et du calcul logique !)

Commutativité :

$$E \cup F = F \cup E \qquad \text{et} \qquad E \cap F = F \cap E.$$

Associativité :

$$(E \cup F) \cup G = E \cup (F \cup G) \qquad \text{et} \qquad (E \cap F) \cap G = E \cap (F \cap G).$$

Distributivité :

$$(E \cup F) \cap G = (E \cap G) \cup (F \cap G) \qquad \text{et} \qquad (E \cap F) \cup G = (E \cup G) \cap (F \cup G).$$

Passage au complémentaire :

$$(E \cup F)^c = E^c \cap F^c \qquad \text{et} \qquad (E \cap F)^c = E^c \cup F^c.$$

1.2.4 Ensemble des parties d'un ensemble

On peut regrouper tous les sous-ensembles possibles d'un ensemble E : c'est l'ensemble des parties de E, noté $\mathcal{P}(E)$. Par exemple,

$$\mathcal{P}(\{1,5,3\}) = \{\varnothing, \{1\}, \{3\}, \{5\}, \{1,3\}, \{5,3\}, \{1,5\}, \{1,3,5\}\}.$$

Attention, les sous-ensembles de E sont des éléments de $\mathcal{P}(E)$: écrire $F \subset E$ est équivalent à écrire $F \in \mathcal{P}(E)$.

1.2.5 Couples, triplets, n-uplets et suites

Pour désigner des suites ordonnées d'éléments, on utilise les parenthèses :

- $(1,5,3)$ est le **triplet** formé des nombres 1, 5 puis 3. Bien noter que dans ce cas l'ordre compte : $(1,5,3) \neq (1,3,5)$.

- $(2.45, 3)$ est le **couple** formé du nombre réel 2.45 et de l'entier 3.
- $(1, 3, 5, \ldots, 19, 21)$ est le 10-uplet des 10 premiers entiers impairs.

Pour désigner les ensembles de couples, de triplets, etc, on utilise la notation $\times$ (prononcer "croix") :
- $\{1, 3, 5\} \times \{1, 3, 5\}$ (noté aussi $\{1, 3, 5\}^2$) est l'ensemble des couples d'éléments de $\{1, 3, 5\}$. On a ainsi

$$\{1, 3, 5\}^2 = \{(1, 1), (1, 3), (1, 5), (3, 1), (3, 3), (3, 5), (5, 1), (5, 3), (5, 5)\}$$

  On peut remarquer que $Card(\{1, 3, 5\}^2) = 9 = 3^2 = (Card\{1, 3, 5\})^2$.
- $\mathbf{N} \times \mathbf{N}$ est l'ensemble des couples de nombres entiers, ce qui peut s'écrire $\mathbf{N} \times \mathbf{N} = \{(a, b), a \in \mathbf{N}, b \in \mathbf{N}\}$. On écrit aussi $\mathbf{N} \times \mathbf{N} = \mathbf{N}^2$.
- $\mathbf{N} \times \mathbf{N} \times \mathbf{N} = \mathbf{N}^3$: ensemble des triplets d'entiers.
- $\mathbf{N}^{10}$: ensemble des 10-uplets d'entiers. Par exemple $(1, 3, 5, \ldots, 19, 21) \in \mathbf{N}^{10}$.
- $\mathbf{R} \times \mathbf{N} = \{(x, n), x \in \mathbf{R}, n \in \mathbf{N}\}$. Par exemple, $(2.45, 3) \in \mathbf{R} \times \mathbf{N}$.
- $\mathbf{R}^n = \{(x_1, x_2, \ldots, x_n), x_1 \in \mathbf{R}, \ldots, x_n \in \mathbf{R}\}$: n-uplets de nombres réels (si n est un entier fixé).

Lorsque la suite a une infinité d'éléments, on parle de **suite** tout court, et on note avec des indices de position : par exemple $(1, 1/2, 1/3, 1/4, \ldots)$ est la suite $(u_n)_{n\in\mathbf{N}}$ avec $u_n = 1/n$.

1.3 Combinatoire

1.3.1 Cardinal de l'ensemble des couples ou des n-uplets

Si $\mathrm{Card}(A) < \infty$ et $\mathrm{Card}(B) < \infty$,

$$\mathrm{Card}(A \times B) = \mathrm{Card}(A)\mathrm{Card}(B).$$

Plus généralement si $\mathrm{Card}(A_1) < \infty, \mathrm{Card}(A_2) < \infty, \ldots, \mathrm{Card}(A_n) < \infty$, alors

$$\mathrm{Card}(A_1 \times A_2 \times \cdots \times A_n) = \mathrm{Card}(A_1)\mathrm{Card}(A_2) \cdots \mathrm{Card}(A_n).$$

En particulier si $\mathrm{Card}(A) < \infty$,

$$\mathrm{Card}(A^n) = (\mathrm{Card}(A))^n.$$

1.3.2 Cardinal de l'ensemble des parties

Si $\mathrm{Card}(A) < \infty$,

$$\mathrm{Card}(\mathcal{P}(A)) = 2^{\mathrm{Card}(A)}.$$

1.3.3 Ensemble des combinaisons de p éléments parmi n

Soit A un ensemble contenant n éléments, et p un entier. On note $\mathcal{P}_p(A)$ l'ensemble des sous-ensembles de A contenant p éléments. Par exemple si $A = \{1, 3, 5\}$,

$$\begin{aligned}
\mathcal{P}_0(A) &= \{\varnothing\}, \\
\mathcal{P}_1(A) &= \{\{1\}, \{3\}, \{5\}\}, \\
\mathcal{P}_2(A) &= \{\{1, 3\}, \{3, 5\}, \{1, 5\}\}, \\
\mathcal{P}_3(A) &= \{\{1, 3, 5\}\}.
\end{aligned}$$

On remarque ici que $\mathrm{Card}(\mathcal{P}_0(A)) = 1$, $\mathrm{Card}(\mathcal{P}_1(A)) = 3$, $\mathrm{Card}(\mathcal{P}_2(A)) = 3$, et $\mathrm{Card}(\mathcal{P}_3(A)) = 1$. De manière générale le cardinal de $\mathcal{P}_p(A)$ lorsque A contient n éléments est noté $\binom{n}{p}$ et vaut

$$\binom{n}{p} = \frac{n!}{(n-p)!p!}.$$

Ces nombres sont les coefficients binomiaux et peuvent se retrouver aussi grâce au triangle de Pascal.

1.3.4 Ensemble des permutations d'un ensemble à n éléments

Une permutation d'un ensemble A à n éléments est un n-uplet formé des éléments de A sans répétition. L'ensemble de ces permutations est noté $\mathcal{S}(A)$. Autrement dit on forme $\mathcal{S}(A)$ en considérant tous les ordres possibles pour les éléments de A. Par exemple,

$$\mathcal{S}(\{1, 3, 5\}) = \{(1, 3, 5), (1, 5, 3), (3, 1, 5), (3, 5, 1), (5, 1, 3), (5, 3, 1)\}.$$

Le cardinal de A, c'est-à -dire le nombre de permutations de n éléments, est égal à

$$\mathrm{Card}(\mathcal{S}(A)) = n!.$$

1.3.5 Ensemble des arrangements de p éléments parmi n

Soit toujours A un ensemble à n éléments, et p un entier. On note $\mathcal{S}_p(A)$ l'ensemble des p-uplets formés d'éléments de A, sans répétition. Par exemple si $A = \{1, 3, 5\}$,

$$\begin{aligned}
\mathcal{S}_1(A) &= \{1, 3, 5\}, \\
\mathcal{S}_2(A) &= \{(1, 3), (3, 1), (3, 5), (5, 3), (1, 5), (5, 1)\}, \\
\mathcal{S}_3(A) &= \{(1, 3, 5), (1, 5, 3), (3, 1, 5), (3, 5, 1), (5, 1, 3), (5, 3, 1)\} = \mathcal{S}(A).
\end{aligned}$$

On voit que pour construire $S_p(A)$, on forme, à partir de chaque élément de $\mathcal{P}_p(A)$, les p-uplets obtenus en permutant ses éléments. Comme il y a $p!$ telles permutations, le nombre total d'éléments de $S_p(A)$ est égal à

$$\mathrm{Card}(\mathcal{S}_p(A)) = p!\ \mathrm{Card}(\mathcal{P}_p(A)) = p!\binom{n}{p} = \frac{n!}{(n-p)!}.$$

1.4 Événements et probabilité

1.4.1 Exemple introductif

On lance un dé. Quelle est la probabilité que le résultat soit un nombre pair supérieur ou égal à trois ?

Ce problème est très simple et peut se résoudre de tête : le dé doit tomber sur 4 ou 6 pour que le résultat soit à la fois pair et supérieur à 3 ; on peut donc répondre directement qu'il y a 2 chance sur 6 pour que ça arrive.

Nous allons voir comment introduire des notations mathématiques rigoureuses générales permettant de résoudre des problèmes tels que celui-ci, ou d'autres plus complexes pour lesquels l'intuition ne suffit pas. Cependant pour simplifier la compréhension, nous allons suivre ce premier exemple.

Il s'agit de trouver la probabilité qu'un certain **événement** se produise, en l'occurrence l'événement

$$A = \text{"le résultat du dé est un nombre pair supérieur ou égal à 3"}.$$

On note cette probabilité $P(A)$. Par convention la **probabilité** d'un événement est un nombre réel compris entre 0 et 1 (c'est-à -dire appartenant à l'intervalle $[0,1]$), 0 signifiant que l'événement n'a aucune chance de se réaliser, et 1 qu'il est certain de se réaliser.

Comme le nombre d'issues possibles est fini, on peut calculer la probabilité d'un évènement A de la manière suivante :

$$P(A) = \frac{\#\ \text{d'issues qui réalisent } A}{\#\ \text{d'issues possibles}}$$

Ici,

$$P(A) = \frac{2}{6} = \frac{1}{3}.$$

Autre exemple : Il y a trois billes rouges, trois billes vertes, et deux billes jaunes dans un sac. On en pioche trois au hasard, d'un seul coup. Quelle est la probabilité qu'elles soient toutes rouges ?

On pose l'évènement $A =$ “les trois billes piochées sont rouges″. Le nombre d'issues possibles est le nombre d'ensembles de trois billes dans l'ensemble des 8 billes, c'est donc

$$C_8^3 = \frac{8!}{3!5!} = \frac{8\times 7\times 6\times 5\times 4\times 3\times 2}{3\times 2\times 5\times 4\times 3\times 2} = 8\times 7 = 56.$$

Le nombre d'issues qui réalisent A est 1 car il n'y a que l'issue où on pioche les 3 billes rouges. Donc

$$P(A) = \frac{1}{56}.$$

Le fait qu'un évènement est aléatoire ne veut pas forcément dire qu'il ne s'est pas encore produit, ou que personne ne connaît son résultat, mais simplement que **le sujet** ne connaît pas son résultat. L'évènement précédent aurait pu être énoncé

— Un croupier lance un dé et le cache sous un gobelet, pour ne pas voir son résultat.
— Un croupier lance un dé, le cache sous un gobelet, et regarde le résultat sans vous le montrer.

Formulé comme ça, c'est assez évident que la réponse à la question sera toujours la même. Pour ces raisons là, les évènements suivants peuvent également être des évènements aléatoires, même si leur résultat n'est pas (ou plus) le fruit du hasard :

— Il a plu le 12 janvier 1132 (aléatoire pour tout le monde, aujourd'hui).
— Le nombre d'atomes de carbone sur terre est pair (aléatoire pour tout le monde).
— La couleur préférée du président de la république est le jaune (aléatoire pour presque tout le monde).
— Vous avez un poisson d'avril sur le dos (aléatoire uniquement pour vous).
— Un malfaiteur se cache à une certaine adresse (aléatoire pour la police mais pas pour le malfaiteur)

Le degré d'incertitude lié à un évènment aléatoire est également lié à des connaissances partielles du sujet. Par exemple, le cours du dollar demain pour nous est "plus aléatoire" que pour un expert du marché américain, qui en plus connaît le cours aujourd'hui.

1.4.2 Univers, évènements, et réalisations

Pour comprendre les notations utilisées en calcul des probabilités, imaginons une expérience où on lance deux dès, un rouge et un vert. On ne sait pas quelle "réalité" va se produire, et on appelle

$$\Omega = \{\text{ Ensemble de toutes les réalités possibles}\}.$$

Dans l'exemple précédent,

$$\begin{aligned}\Omega = \{&\text{"Le dé rouge tombe sur 1 et le vert aussi"},\\ &\text{"le dé rouge tombe sur 1 et le vert sur 2"},\\ &\dots,\\ &\text{"Les deux dés tombent sur 6"}\}\end{aligned}$$

(il y a 36 éventualités possibles, donc 36 éléments dans Ω. On les appelle $\omega_1, \omega_2, \ldots, \omega_{36}$). Une "réalité" s'appelle une **réalisation**, et Ω l'ensemble des réalisations possibles s'appelle **l'univers**.

Les **évènements** sont des sous-ensembles de Ω qui ne contiennent que certaines réalisations. Par exemple, l'évènement $A = \{\text{le dé vert tombe sur un nombre impair}\}$ contient l'élèment $\omega_1 =$"Le dé rouge tombe sur 1 et le vert aussi" mais pas l'élèment $\omega_2 =$"le dé rouge tombe sur 1 et le vert sur 2". On peut également considérer l'évènement Ω lui-même, qui contient toutes les réalisations, où l'évènement vide $\emptyset$, qui ne contient aucune réalisation. Il se peut qu'Ω contienne une infinité de réalisations !

FIGURE 1.1 – L'événement A représenté comme intersection des événements B et C puis comme union des événements D_4 et D_6.

Remarquons que l'univers peut être infini. Prenons par exemple l'expérience suivante : On jette un dé à 6 faces, et on s'arrête lorsque l'on a obtenu un 6. On appelle ω_n la réalisation "on obtient 6 au n-ème lancer", pour $n \geqslant 1$. L'univers Ω est dans ce cas

$$\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_n, \ldots\}.$$

Même si il n'y aura en pratique qu'un nombre fini de lancers, le nombre d'éventualités possibles est infini.

1.4.3 Intersections d'événements

L'événement A ="le dé rouge tombe sur un nombre pair plus grand que 3" se décompose naturellement de la manière suivante :

$A =$ "le résultat du dé est un nombre pair" et "le résultat du dé est supérieur ou égal à 3",

ou encore, si l'on a défini $B =$ "le résultat du dé est un nombre pair" puis
$C =$ "le résultat du dé est supérieur ou égal à 3", on écrira simplement

$$A = (B \text{ et } C) = (B \text{ se réalise}) \text{ et } (C \text{ se réalise}).$$

En termes de sous-ensembles de Ω, le "et" de l'égalité précédente se traduit par une intersection : Une réalisation ω appartient à A ssi elle appartient à B et C à la fois, c'est-à-dire si $\omega \in B \cap C$. Plutôt que $A = B$ et C, on écrira en fait :

$$A = B \cap C.$$

Malheureusement cette décomposition ne permet pas directement de calculer la probabilité de l'événement A car il n'existe pas de formule générale reliant cette probabilité à celles de B et de C.

1.4.4 Unions d'événements et probabilité d'une union

Un dé a six faces, numérotées de 1 à 6, et par conséquent pour que le résultat soit pair, il faut qu'il tombe sur 2, 4, ou 6. Donc pour qu'il soit pair et supérieur à 3, il faut qu'il tombe sur 4 ou 6. On peut donc formuler différemment l'événement A :

$$A = \text{"le résultat du dé est 4 ou 6"}$$

ou encore :

$$A = \text{"le résultat du dé est 4" ou "le résultat du dé est 6"}$$

Par conséquent A est composé de deux événements élémentaires, que l'on peut noter D_4, D_6, avec $D_i =$ "le résultat du dé est i". On a donc

$$A = D_4 \text{ ou } D_6.$$

Les "ou" dans l'égalité précédente se traduisent alors par des unions de sous-ensembles : $\omega \in A$ ssi $\omega \in D_4$ ou $\omega \in D_6$, c'est-à-dire ssi $\omega \in D_4 \cup D_6$. On écrira donc en fait :

$$A = D_4 \cup D_6.$$

Pour calculer la probabilité d'une union d'événements on dispose des formules suivantes :

Probabilité de l'union de deux événements - formule générale. *Soient E et F deux événements. Alors*

$$\boxed{P(E \cup F) = P(E) + P(F) - P(E \cap F).}$$

Probabilité d'une union disjointe.

- *Soient E et F deux événements. On dit que E et F sont disjoints s'il est impossible qu'ils soient réalisés simultanément, c'est-à-dire si aucune réalisation n'appartient aux 2. Si E et F sont* **disjoints** *(c'est-à -dire disjoints : $E \cap F = \varnothing$) alors*

$$P(E \cup F) = P(E) + P(F).$$

  C'est logique puisque $P(E \cap F) = 0$...
- *Soient $E_1, \dots, E_n$ plusieurs événements. S'ils sont* **deux à deux disjoints** *($E_i \cap E_j = \varnothing$ pour tous $i \neq j$) alors*

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = P(E_1) + P(E_2) + \cdots + P(E_n).$$

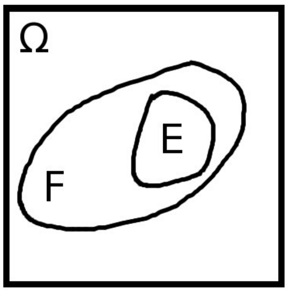


FIGURE 1.2 – Si $E \subset F$ alors $P(E) \leq P(F)$.

Dans notre exemple les événements D_4 et D_6 sont disjoints. En effet l'événement $D_4 \cap D_6$ correspond à : "Le dé tombe sur 4 et sur 6", ce qui est impossible. Ainsi $D_4 \cap D_6 = \varnothing$. On peut donc écrire

$$P(A) = P(D_4) + P(D_6).$$

Finalement on sait que la probabilité de tomber sur chaque face est de $\frac{1}{6}$. Ainsi $P(A) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Interprétation graphique : Les formules précédentes se retrouvent facilement en dessinant un diagramme tel que celui de la figure 1.1, et en imaginant que les probabilités des événements sont les aires des sous-ensembles correspondants (en supposant que Ω est un carré de côté 1 par exemple).

Voici une autre relation importante découlant des formules précédentes :

$$\boxed{\textit{Si } E \subset F \textit{ alors } P(E) \leq P(F).}$$

Encore une fois, ceci se comprend très facilement sur un schéma (cf. figure 1.2)

1.4.5 Partition de Ω

Des événements $E_1, \ldots, E_n$ forment une **partition** de l'univers Ω s'ils sont deux à deux incompatibles et que leur réunion est égale à Ω. Autrement dit : un et un seul de ces événements se réalise. Alors on a la relation :

$$P(E_1) + \cdots + P(E_n) = 1.$$

Par exemple, si on considère les événements :
— E_1 : Le résultat du dé est une puissance de 2 $(2, 4, 8, ...)$
— E_2 : Le résultat du dé est impair
— E_3 : Le résultat du dé est 6,

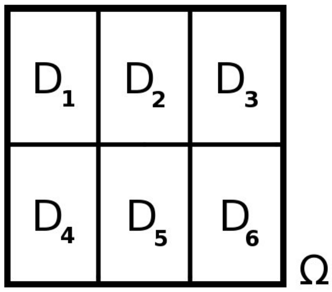


FIGURE 1.3 – Les événements D_1 à D_6 forment une partition de Ω.

on a bien une partition de toutes les possibilités en 3 évènements qui ne peuvent pas survenir simultanément.

Cette relation s'obtient en appliquant simplement la formule de probabilité d'une union disjointe, en se rappelant que $P(\Omega) = 1$:

$$1 = P(\Omega) = P(E_1 \cup E_2 \cup \cdots \cup E_n) = P(E_1) + \cdots + P(E_n).$$

Dans l'exemple du lancer de dé, les événements $D_1, \ldots, D_6$ forment une partition de Ω. Puisque chacun de ces événements a pour probabilité $\frac{1}{6}$, on vérifie bien la formule : $\frac{1}{6} + \cdots + \frac{1}{6} = 6 \times \frac{1}{6} = 1$ (voir figure 1.3). Dans ce cas on voit que non seulement ces ensembles forment une partition de Ω, mais qu'aussi ils ont tous la même probabilité. C'est ce qu'on appelle l'équiprobabilité :

Définition. *On dit qu'il y a* **équiprobabilité** *entre des événements $E_1, \ldots, E_n$ s'ils forment une partition de Ω et que leurs probabilités sont toutes égales : $P(E_1) = P(E_2) = \cdots = P(E_n) = \frac{1}{n}$.*

Dans l'exemple où l'on pioche trois billes dans un sac de 8 billes, en supposant que les billes sont numérotées de 1 à 8, on a une partition de Ω en considérant tous les triplets possibles :

$$\Omega = (\text{“on pioche les billes } 1, 2, 3'') \cup (\text{“on pioche les billes } 1, 2, 4'') \cup ... \cup (\text{“on pioche les billes } 6, 7, 8'')$$

Comme il y a $C_8^3 = 56$ possibilités, chaque évènement a une probabilité de $1/56$.

Il y a d'autres possibilités de partitionner Ω, par exemple

$$\begin{aligned}\Omega = &(\text{“Les trois billes sont de trois couleurs différentes''})\\ &\cup (\text{“Deux billes sont d'une couleur, et la troisième d'une autre''})\\ &\cup (\text{“Les trois billes sont de la même couleur''})\end{aligned}$$

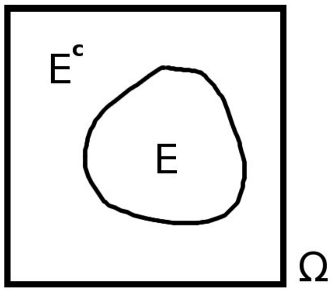


FIGURE 1.4 – E et E^c forment une partition de Ω.

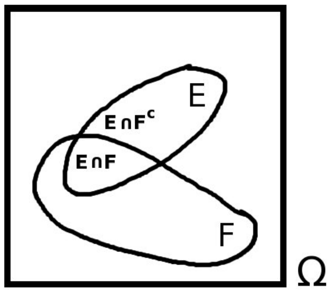


FIGURE 1.5 – $P(E) = P(E \cap F) + P(E \cap F^c)$.

1.4.6 Événement complémentaire

L'**événement complémentaire** d'un événement E est l'événement correspondant à l'ensemble complémentaire E^c. Il s'agit de la négation de l'événement E : E^c se réalise si et seulement si E ne se réalise pas. On a la formule

$$\boxed{P(E^c) = 1 - P(E),}$$

qui provient simplement du fait que E et E^c forment une partition de Ω.

La relation suivante est très utile :

Soient E et F deux événements. Alors

$$\boxed{P(E) = P(E \cap F) + P(E \cap F^c).}$$

On comprend facilement cette relation en dessinant des patates (cf. figure 1.5). La preuve mathématique est assez simple aussi ; la voici :

Preuve. On va utiliser la formule de probabilité d'une union disjointe. On écrit d'abord A comme une union :

$$A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

On vérifie que cette union est disjointe :

$$(A \cap B) \cap (A \cap B^c) = A \cap B \cap A \cap B^c = \varnothing$$

On applique donc la formule de probabilité d'une union disjointe :

$$P(A) = P(A \cap B) + P(A \cap B^c).$$

□

1.4.7 Exclusion d'évènements

On a vu que quand A et B sont deux évènements tels que $A \subset B$, alors $P(A) \leqslant P(B)$. On peut même être plus précis en écrivant

$$P(B \cap A^c) = P(B) - P(A)$$

On note aussi $B \cap A^c = B \setminus A$ (diagramme).

Application : Jeu télévisé

Lors d'un jeu télévisé, un candidat a le choix entre trois rideaux. Derrière l'un des rideaux se cache un cadeau, et derrière les deux autres rien du tout. Le jeu se passe en deux temps :

- Tout d'abord, le candidat choisit un rideau. Le présentateur, qui sait où est le cadeau, ouvre un des deux autres rideaux **derrière lequel il n'y a rien** (il y en a au moins un derrière lequel il n'y a rien, voire les deux)
- Ensuite il demande au candidat s'il veut garder son choix, ou s'il veut choisir l'autre rideau.

Question : Vaut-il mieux changer de rideau ou garder le même ? **Correction :** Comme souvent, le tout est de bien poser le problème. On introduit deux stratégies : La stratégie "GARDE", où on garde le même rideau, et la stratégie "CHANGE", où on change. On appelle G l'évènement "Le candidat opte pour la stratégie "GARDE"", et C l'évènement "le candidat" opte pour la stratégie "CHANGE".

On peut dire très facilement deux choses :

- La probabilité de l'évènement G est la probabilité de bien choisir dès le début, elle est donc de $1/3$.
- L'évènement C est le complémentaire de l'évènement G, car le candidat ne peut choisir que l'une ou l'autre des stratégies. On a $C = G^c$.

On en déduit que $P(C) = 1 - P(G) = 1 - 1/3 = 2/3$. Donc, en changeant son choix, le candidat a deux fois plus de chances de remporter le cadeau !

Chapitre 2

Probabilités conditionnelles et indépendance d'évènements

2.1 Probabilités conditionnelles

Parfois, on dispose d'informations partielles concernant un évènement aléatoire. Le but des probabilités conditionelles est de prendre en compte cette information partielle pour le calcul de la probabilité.

On tire deux cartes successivement et sans remise d'un jeu de 32 cartes. Quelle est la probabilité que ce soient deux piques ?

La réponse "intuitive" est la suivante : il y a 8 chances sur 32 de tirer un pique pour la première carte, puis 7 chances sur 31 pour la deuxième (puisqu'il reste 31 cartes dont 7 piques), donc la probabilité demandée est égale à $\frac{8}{32} \times \frac{7}{31}$.

La notion de probabilité conditionnelle permet de formuler rigoureusement une telle réponse.

Définition (Probabilité conditionnelle)**.** *Soient E et F deux événements. On note $P(E|F)$ (ce qui se lit "probabilité de E sachant F") la probabilité que l'événement E se réalise sachant que l'événement F est réalisé.*

Autrement dit, à partir d'une expérience probabiliste où E et F sont aléatoires, on change de situation : on suppose que F n'est plus aléatoire mais réalisé, et on cherche à calculer la probabilité de E dans cette nouvelle situation. Sur un schéma ensembliste (figure 2.1), ce changement de situation probabiliste consiste à considérer l'ensemble F comme le nouvel univers, à la place de Ω, et donc à ne plus considérer ce qui est en dehors de F. Autrement dit, la probabilité d'un événement E sera obtenue en ne retenant que la probabilité de sa partie contenue dans le nouvel univers, c'est-à -dire l'intersection $E \cap F$, et en divisant par $P(F)$, afin que dans la nouvelle situation, la probabilité de l'univers F soit égale à 1.

FIGURE 2.1 – Schéma du changement de situation probabiliste correspondant au conditionnement par rapport à un événement F. A gauche, situation initiale dans l'univers Ω ; à droite, situation conditionnée à F : l'événement F devient le nouvel univers, et seule la partie de E comprise dans F est considérée.

Ceci permet de justifier la formule fondamentale suivante, qui est en fait la définition mathématique de la probabilité conditionnelle :

$$\boxed{P(E|F) = \frac{P(E \cap F)}{P(F)}.}$$

On voit par cette formule que $P(E|F)$ n'est défini que si $P(F) \neq 0$; ce qui est cohérent : on ne peut pas supposer que F est réalisé s'il n'a aucune chance de se réaliser.

Voici maintenant comment exprimer rigoureusement la réponse à l'exemple précédent : soient A l'événement "La première carte est un pique", B l'événement "La deuxième carte est un pique", et C l'événement "Les deux cartes tirées sont des piques". On a donc $C = A \cap B$. D'après la formule des probabilités conditionnelles, on a

$$P(C) = P(A \cap B) = P(B|A)P(A).$$

La probabilité de tirer un pique en premier est de $P(A) = \frac{8}{32}$. Maintenant si A est réalisé, alors il reste 31 cartes dont 7 piques, et donc la probabilité de tirer à nouveau un pique est de $\frac{7}{31}$: ainsi $P(B|A) = \frac{7}{31}$. Finalement on obtient

$$P(C) = P(B|A)P(A) = \frac{7}{31} \times \frac{8}{32} = \frac{7}{124}.$$

remarque : Toujours penser à vérifier que la probabilité calculée est bien un nombre compris entre 0 et 1.

remarque : Attention, $E|F$ ne signifie rien en soi, ce n'est pas un événement. Seule l'expression $P(E|F)$ a un sens. En fait la probabilité conditionnelle $P(E|F)$ se note parfois $P_F(E)$, notation moins usuelle mais bien plus juste car elle traduit précisément le changement de situation probabiliste évoqué plus haut : la fonction P_F est la fonction de probabilité de la nouvelle situation, en remplacement de la fonction P initiale.

Comme on l'a vu dès le début de ce chapitre, conditionner par rapport à un événement G signifie se placer dans une nouvelle situation probabiliste dans laquelle G est réalisé. Cette nouvelle situation change les probabilités des événements... Les probabilités calculées conditionnellement à G obéissent aux même règles que les probabilités calculées dans la configuration initiale. Pour cette raison on note parfois $P(E|G) = P_G(E)$ la probabilité d'un événement E sachant G. En termes mathématiques, on dit que P_G est une **probabilité**, au même titre que la probabilité de référence P, et obéit aux même règles.

Voici quelques exemples de notions et formules découlant de ce principe :

Probabilité conditionnelle d'une union *Soient E, F, G trois événements, avec $P(G) \neq 0$. Alors*

$$P(E \cup F|G) = P(E|G) + P(F|G) - P(E \cap F|G).$$

Double conditionnement *Soient E, F, G trois événements, avec $P(F \cap G) \neq 0$. Alors*

$$P(E|F \cap G) = \frac{P(E \cap F|G)}{P(F|G)}.$$

Cette formule devient claire si on la comprend comme la définition, sous l'hypothèse que G est réalisé, de la probabilité conditionnelle de E sachant F. Autrement, elle se démontre très facilement à partir des définitions :

Preuve.

$$P(E|F \cap G) = \frac{P(E \cap F \cap G)}{P(F \cap G)} = \frac{P(E \cap F|G)P(G)}{P(F|G)P(G)} = \frac{P(E \cap F|G)}{P(F|G)}.$$

□

Formule des probabilités totales conditionnelle *: Soient E, F, G trois événements avec $P(F \cap G) \neq 0$. Alors*

$$P(E|G) = P(E|F \cap G)P(F|G) + P(E|F^c \cap G))P(F^c|G).$$

2.2 Formule des probabilités totales

On dispose de deux urnes. La première contient deux boules noires et une boule blanche ; et la deuxième contient trois boules noires et deux boules blanches. On tire au hasard une boule de la première urne et on la place dans la deuxième urne. Puis on tire au hasard une boule de la deuxième urne. Quelle est la probabilité que cette deuxième boule soit blanche ?

On note B_1 l'événement "la première boule tirée est blanche", et B_2 l'événement "la deuxième boule tirée est blanche". On cherche donc à calculer $P(B_2)$. La formule suivante, vue au premier chapitre, est le point de départ du calcul :

$$P(B_2) = P(B_2 \cap B_1) + P(B_2 \cap B_1^c).$$

Autrement dit on décompose l'événement "la deuxième boule tirée est blanche" en deux : "la deuxième boule tirée est blanche et la première était blanche", et "la deuxième boule tirée est blanche et la première n'était pas blanche (donc était noire)". On utilise alors la définition mathématique des probabilités conditionnelles pour calculer $P(B_2 \cap B_1)$ et $P(B_2 \cap B_1^c)$, ce qui donne :

$$P(B_2) = P(B_2|B_1)P(B_1) + P(B_2|B_1^c)P(B_1^c).$$

C'est la formule des probabilités totales. A présent toutes les probabilités à droite de l'égalité peuvent se calculer :

- Pour déterminer $P(B_2|B_1)$ on se place dans la situation où B_1 est réalisé, c'est-à - dire que la première boule tirée est blanche. Dans ce cas la deuxième urne contiendra trois boules blanches et trois boules noires, et donc la probabilité de tirer une boule blanche sera de $\frac{3}{6} = \frac{1}{2}$. Ainsi on a montré que $P(B_2|B_1) = \frac{1}{2}$.
- Pour déterminer $P(B_2|B_1^c)$ on se place dans la situation inverse : la première boule tirée est noire, donc la deuxième urne contient deux boules blanches et quatre boules noires. Ainsi $P(B_2|B_1^c) = \frac{2}{6} = \frac{1}{3}$.
- $P(B_1)$ est la probabilité que la première boule tirée soit blanche : $P(B_1) = \frac{1}{3}$.
- Enfin $P(B_1^c) = 1 - P(B_1) = \frac{2}{3}$.

Finalement on trouve donc

$$P(B_2) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3} = \frac{7}{18}.$$

On a démontré et utilisé dans ce calcul la formule des probabilités totales :

Formule des probabilités totales, cas simple. *Soient E et F deux événements, avec $P(F)$ et $P(F^c)$ non nuls. Alors*

$$\boxed{P(E) = P(E|F)P(F) + P(E|F^c)P(F^c).}$$

La version générale de la formule consiste à conditionner non plus par seulement deux événements (F et F^c) pour le calcul de $P(E)$, mais par un nombre n :

Formule des probabilités totales, cas général. *Soient E un événement et $F_1, F_2, \ldots, F_n$ des événements formant une partition de Ω, avec $P(F_i) \neq 0$ pour tout i. Alors*

$$\boxed{P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) + \cdots + P(E|F_n)P(F_n).}$$

En voici un exemple d'utilisation (cas $n = 3$) :

Pour traiter une maladie, les médecins disposent de trois nouveaux médicaments MA, MB, MC. Dans un premier cas, les médecins prescrivent indifféremment l'un des trois médicaments pour chaque traitement. Dans un deuxième cas, ils commencent à connaître mieux ces médicaments, et prescrivent MA dans 50% des cas, MB dans 30% des cas, et MC dans 20% des cas. En fait les taux de réussite de ces médicaments sont respectivement de 98%, 96% et 95%. Calculer la probabilité d'échec du traitement dans chaque cas.

On va noter A l'événement "recevoir le médicament MA", et de même B et C. On cherche la probabilité de E="échec du traitement". Il faut bien comprendre les données du problème : les taux de réussite correspondent en fait à des probabilités conditionnelles : "si on reÃğoit le médicament MA, alors on guérit dans 98% des cas". On a donc $P(E^c|A) = 0.98$, et de même $P(E^c|B) = 0.96$ et $P(E^c|C) = 0.95$. On va donc utiliser la formule des probabilités totales en remarquant que A, B et C forment une partition de Ω (car un et un seul des trois médicaments est administré lors d'un traitement) :

$$P(E^c) = P(E^c|A)P(A) + P(E^c|B)P(B) + P(E^c|C)P(C).$$

<u>1er cas</u> : $P(A) = P(B) = P(C) = \frac{1}{3}$. Donc $P(E^c) = 0.98 \times \frac{1}{3} + 0.96 \times \frac{1}{3} + 0.95 \times \frac{1}{3} = 0.963$. Finalement $P(E) = 1 - P(E^c) = 0.037$.
<u>2e cas</u> : $P(A) = 0.5$, $P(B) = 0.3$ et $P(C) = 0.2$. Donc $P(E^c) = 0.98 \times 0.5 + 0.96 \times 0.3 + 0.95 \times 0.2 = 0.968$, et $P(E) = 0.032$.

2.3 Formule de Bayes

Revenons au problème des deux urnes, et cherchons à répondre à cette deuxième question :

Si la deuxième boule tirée est blanche, quelle est la probabilité que la première boule tirée ait été blanche aussi ?

Ici on demande de calculer $P(B_1|B_2)$. Cette probabilité conditionnelle ne peut pas être trouvée directement comme c'est le cas pour $P(B_2|B_1)$. En effet ici on cherche la probabilité d'un événement en conditionnant par rapport à un événement qui en découle, ce qui va contre l'intuition. Il faut donc faire une manipulation pour "retourner" la probabilité conditionnelle :

$$P(B_1|B_2) \quad = \quad \frac{P(B_1 \cap B_2)}{P(B_2)} = \frac{P(B_2|B_1)P(B_1)}{P(B_2)}.$$

A présent on peut effectuer le calcul :

$$P(B_1|B_2) = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{7}{18}} = \frac{3}{7}.$$

Nous avons donc utilisé la formule suivante, qui est la base de la formule de Bayes :

Soient E et F deux événements avec $P(E)$ et $P(F)$ non nuls. Alors

$$\boxed{P(F|E) = \frac{P(E|F)P(F)}{P(E)}}.$$

La formule de Bayes est simplement la combinaison de cette formule et de la formule des probabilités totales pour le calcul de $P(E)$:

Formule de Bayes, cas simple. *Soient E et F deux événements avec $P(E)$, $P(F)$ et $P(F^c)$ non nuls. Alors*

$$\boxed{P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|F^c)P(F^c)}}.$$

Cette formule permet d'une certaine manière de "retourner" la probabilité conditionnelle.

2.4 Indépendance

2.4.1 Indépendance de deux événements

La définition intuitive de l'indépendance est la suivante : deux événements sont indépendants lorsque le résultat de l'un n'influence pas le résultat de l'autre.

Attention ! Cette définition est à prendre au sens large, l'influence n'a pas besoin d'être directe. Exemples : Je cherche à savoir s'il pleut à Brest, car je viens de là-bas. Je considère l'évènement E="Ma grand-mère a pris son parapluie" et l'évènement F="Mon ami Pierre a pris son parapluie". Sachant que Pierre et ma grand-mère ne se sont jamais rencontrés, les évènements E et F sont-ils indépendants ? **Réponse :** Non car ils dépendent tous les deux du même évènement G="Il pleut à Brest". Ainsi, si je sais que ma grand-mère a pris son parapluie, j'en déduis qu'il y a des chances qu'il pleuve, et donc que Pierre a pris son parapluie. Autrement dit :

$$P(F|E) > P(F).$$

Une définition intuitive précise de l'indépendance est : "Deux évènements sont indépendants" si le fait de savoir des informations sur l'un n'apporte pas d'informations, ou autrement dit le fait de savoir que l'un d'entre eux s'est réalisé n'influence pas la probabilité que l'autre se soit réalisé.

Autrement dit si E et F sont ces deux événements, le fait de supposer que F est réalisé ne change pas la probabilité de réalisation de E et inversement. On a donc :

$$P(E|F) = P(E) \quad \text{et} \quad P(F|E) = P(F).$$

Or $P(E|F) = \frac{P(E \cap F)}{P(F)}$ et $P(F|E) = \frac{P(E \cap F)}{P(F)}$, et par conséquent on voit que les deux conditions ci-dessus se résument en une seule : $P(E \cap F) = P(E)P(F)$. C'est la définition mathématique de l'indépendance :

Définition. *Deux événements E et F sont* **indépendants** *si et seulement si*

$$P(E \cap F) = P(E)P(F).$$

remarque : Cette définition a l'avantage de ne pas supposer $P(E)$ et $P(F)$ non nuls, à la différence de la double condition précédente. En fait si $P(E) = 0$ ou si $P(F) = 0$, la formule est toujours vérifiée. Autrement dit : un événement de probabilité nulle est toujours indépendant de tout autre événement.

Pour revenir à l'exemple précédent, en admettant qu'il y ait 1 chance sur 4 qu'il pleuve à Brest ce jour-là, et que Pierre et ma grand-mère prennent systématiquement leur parapluie quand il pleut, on a $P(E) = P(F) = 1/4$. On a aussi,

$$P(E \cap F) = P(G) = 1/4.$$

Donc $P(E)P(F) = 1/16 \neq P(E \cap F)$, il est clair que ces évènements ne sont pas indépendants.

En pratique c'est souvent l'intuition qui permet de décider si deux événements sont indépendants ou pas, en faisant bien attention aux pièges d'interprétation pour les influences non-directes.

La formule $P(E \cap F) = P(E)P(F)$ est alors utilisée pour faire des calculs.

exemple 1 : *On lance une pièce de monnaie deux fois de suite sur une table. Quelle est la probabilité d'obtenir deux fois face ?*

On note $A =$"on obtient face au premier lancer" et $B =$"on obtient face au deuxième lancer". Ici ces deux événements sont clairement indépendants : le fait d'obtenir face en premier ne change pas la probabilité d'obtenir face au deuxième. Par conséquent la probabilité demandée ("obtenir deux fois face") sera : $P(A \cap B) = P(A)P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

exemple 2 : *On lance un dé. Quelle est la probabilité que le résultat soit un nombre pair supérieur ou égal à trois ?*

Il s'agit ici du premier exemple du chapitre 1. On avait noté B="le résultat est un nombre pair", C="le résultat est ≥ 3", et la probabilité cherchée est $P(A)$ avec $A = B \cap C$. En principe B et C n'ont aucune raison d'être indépendants puisqu'ils concernent le même lancer de dé. Pourtant on a $P(B) = \frac{1}{2}$, $P(C) = \frac{2}{3}$, et on avait trouvé $P(A) = P(B \cap C) = \frac{1}{3}$. On a donc $P(B \cap C) = P(B)P(C)$, et donc B et C sont mathématiquement indépendants. C'est une sorte d'indépendance fortuite, qui va contre l'intuition.

Modifions à présent légèrement l'énoncé : *On lance un dé. Quelle est la probabilité que le résultat soit un nombre pair supérieur ou égal à quatre ?*
Ici l'analyse est la même, à savoir que les événements B et D="le résultat est ≥ 4" n'ont pas de raison d'être indépendants ; et le calcul montre en effet que $P(B)P(D) = \frac{1}{2} \times \frac{1}{2} \neq P(B \cap D) = \frac{1}{3}$; c'est-à -dire que B et D ne sont pas indépendants.

Pour résumer, l'intuition peut nous dire si des événements sont indépendants, mais à l'inverse on ne peut jamais être certain que des événements ne sont pas indépendants sans faire le calcul.

2.4.2 Indépendance et événements complémentaires

Il est facile de montrer que

$$\begin{aligned} E \text{ et } F \text{ indépendants} &\Leftrightarrow E \text{ et } F^c \text{ indépendants,} \\ &\Leftrightarrow E^c \text{ et } F \text{ indépendants,} \\ &\Leftrightarrow E^c \text{ et } F^c \text{ indépendant.} \end{aligned}$$

Intuitivement, on est simplement en train de dire que le fait que E se réalise/ne se réalise pas est indépendant du fait que F se réalise/ne se réalise pas.

2.4.3 Indépendance de 3 événements

On généralise la notion d'indépendance pour plus de 2 événements. Voici d'abord la version à 3 événements :

Définition. *Trois événements E, F, G sont dits* **indépendants** *(on dit aussi* **mutuellement indépendants***) lorsque les quatre relations suivantes sont vérifiées :*

$$P(E \cap F) = P(E)P(F),$$

$$P(E \cap G) = P(E)P(G),$$
$$P(F \cap G) = P(F)P(G),$$
$$P(E \cap F \cap G) = P(E)P(F)P(G).$$

Pourquoi a-t-on besoin de toutes ces relations ? On pourrait penser que les deux premières ou les trois premières suffisent et entraînent les autres. Mais ceci est faux, comme le montre le contre-exemple suivant :

Exemple : *On tire deux fois un dé à six faces. Les événements suivants sont-ils indépendants ?*
$A =$ *"le premier dé tombe sur un nombre pair",*
$B =$ *"le deuxième dé tombe sur un nombre impair",*
$C =$ *"les deux dés ont même parité".*

Il est clair que A et B sont indépendants. A et C sont aussi indépendants : en effet, d'une part la probabilité de C vaut $P(C) = \frac{1}{2}$; d'autre part si l'on suppose que A est réalisé (le premier dé est pair), alors les deux dés auront même parité si le deuxième est aussi pair, donc avec probabilité $\frac{1}{2}$. Ainsi $P(C|A) = P(C) = \frac{1}{2}$ et donc A et C sont indépendants. Par le même raisonnement on peut voir que B et C sont aussi indépendants. Pour résumer on a les trois relations (indépendances deux-à -deux)

$$P(A \cap B) = P(A)P(B),$$
$$P(A \cap C) = P(A)P(C),$$
$$P(B \cap C) = P(B)P(C).$$

Cependant il est facile de voir que $P(A \cap B \cap C) = 0$ (les trois événements ne peuvent se réaliser ensemble) ; et donc que $P(A \cap B \cap C)$ est différent de $P(A)P(B)P(C)$. Ainsi les trois événements ne sont pas indépendants puisqu'il manque la dernière relation, alors qu'ils sont indépendants deux-à -deux.

2.5 Indépendance conditionnelle

La notion d'indépendance conditionnelle est souvent utile :

Indépendance conditionnelle *: Soient E, F, G trois événements, avec $P(G) \neq 0$. On dit que E et F sont* **indépendants conditionnellement** *à G lorsque, sous l'hypothèse que G est réalisé, E et F sont indépendants. Autrement dit,*

$$P(E \cap F|G) = P(E|G)P(F|G).$$

Des événements conditionnellement indépendants n'ont aucune raison d'être indépendants. Voici un exemple de cette situation.

Exemple *: Un tribunal doit traiter d'affaires de meurtres. A la fin du procès deux jurés doivent se prononcer et l'accusé sera condamné si les deux jurés se prononcent pour la condamnation. On suppose que lorsque l'accusé a réellement commis un meurtre, le procès lui est logiquement défavorable, si bien que chaque juré le déclarera coupable avec une probabilité de* 0.7 *et leurs deux avis sont alors indépendants. Lorsqu'il n'est pas coupable, leurs deux avis sont aussi indépendants et la probabilité qu'il soit déclaré coupable est alors de* 0.2 *pour chacun des deux jurés.*

On peut donc dire que les évènements A*="Le premier juré déclare l'accusé coupable" et l'évènement* B*="Le second juré le déclare coupable" sont indépendants* **conditionnellement à l'évènement** C*="L'accusé est coupable".*

Pourtant, on va montrer que ces deux évènements ne sont pas indépendants ("tout court") car la culpabilité de l'accusé va influence l'avis final des jurés simultanéments.

On suppose que 60% *des accusés sont effectivement coupables. Un accusé est jugé.*

1. *Calculer la probabilité qu'il soit condamné.*
2. *Les événements* A *et* B *sont-ils indépendants ?*

1. On doit calculer $P(A \cap B)$. Traduisons d'abord les hypothèses. Notons M ="l'accusé est un meurtrier". On a :

$$P(A|M) = P(B|M) = 0.7, \text{ et } P(A|M^c) = P(B|M^c) = 0.2.$$

Les hypothèses d'indépendance de l'énoncé sont ici des indépendances conditionnelles : conditionnellement au fait que l'accusé est un meurtrier (c'est-à -dire que M est réalisé), les deux avis sont indépendants, et donc A et B sont indépendants. On a donc

$$P(A \cap B|M) = P(A|M)P(B|M).$$

De même A et B sont indépendants conditionnellement à M^c :

$$P(A \cap B|M^c) = P(A|M^c)P(B|M^c).$$

Enfin on a $P(M) = 0.6$. A présent on peut calculer :

$$\begin{aligned}
P(A \cap B) &= P(A \cap B|M)P(M) + P(A \cap B|M^c)P(M^c) \quad \text{(formule des probabilités totales)}, \\
&= P(A|M)P(B|M)P(M) + P(A|M^c)P(B|M^c)P(M^c), \\
&= 0.7 \times 0.7 \times 0.6 + 0.2 \times 0.2 \times 0.4 = \frac{31}{100}.
\end{aligned}$$

2. On doit comparer $P(A \cap B)$ à $P(A)P(B)$. On a :

$$P(A) = P(A|M)P(M) + P(A|M^c)P(M^c) = 0.7 \times 0.6 + 0.2 \times 0.4 = \frac{1}{2},$$

et de même $P(B) = \frac{1}{2}$. Ainsi $P(A)P(B) = \frac{1}{4} \neq P(A \cap B)$. A et B ne sont donc pas indépendants.

Chapitre 3

Variables aléatoires et lois

3.1 Définitions générales

3.1.1 Variables aléatoires

On appelle **variable aléatoire** tout nombre réel aléatoire, c'est-à-dire dont la valeur dépend du résultat d'une expérience probabiliste. Par exemple :

On lance un dé. Soit X le résultat obtenu.

Ici X est une variable aléatoire et les valeurs possibles de X sont $1, 2, 3, 4, 5, 6$. Pour chacune de ces valeurs, X a une certaine probabilité de lui être égal. Ici en fait on peut donner directement les probabilités des événements $"X = 1", "X = 2", \dots, "X = 6"$: on a $P("X = 1") = P("X = 2") = \cdots = P("X = 6") = \frac{1}{6}$.

Remarque : Un nombre réel fixé (c'est-à-dire non aléatoire) peut être vu comme une variable aléatoire ayant une probabilité 1 d'être égale à la valeur considérée. Par exemple le nombre $x = 2$ sera identifié à une variable X telle que $P(X = 2) = 1$.

Remarques sur les notations :

- Par convention, les variables aléatoires sont en général notées avec des lettres capitales (X,Y,T,etc.) pour les différencier des nombres réels non aléatoires.
- Pour noter les événements relatifs à une variable aléatoire X, comme par exemple $"X = 1", "X \leq 2", "0 \leq X \leq 4"$, on utilise souvent plutôt les crochets : $[X = 1], [X \leq 2], [0 \leq X \leq 4]$.
- De même, plutôt que $P("X = 1")$ ou $P([X = 1])$ on écrira simplement $P(X = 1)$ ou $P[X = 1]$.

3.1.2 Support d'une variable aléatoire

Le **support** d'une variable aléatoire est l'ensemble des ses valeurs possibles. C'est la première chose à préciser lorsqu'on considère une variable aléatoire. On notera $\mathcal{S}(X)$ le support d'une variable aléatoire X.

Exemple 1 : On reprend l'exemple précédent : X est le résultat d'un lancer de dé. Le support de X est alors $\{1, 2, 3, 4, 5, 6\}$.

Exemple 2 : On lance un dé à 6 faces, et l'on recommence jusqu'à obtenir un 6. On note X le nombre de lancer nécessaires. Alors $X \in \{0, 1, 2, \dots\} = \mathbf{N}$.

Exemple 3 : On lance un stylo sur un bureau, et on note X l'angle que fait le stylo avec l'axe du bureau. Alors le support de X est $[0, 2\pi]$.

On vient en fait de voir trois types de support différents avec ces trois exemples : support fini pour le premier, support infini dénombrable pour le second, support infini non dénombrable pour le troisième. Cette distinction est essentielle en probabilités, car les calculs de probabilités vont s'effectuer de façon complètement différentes suivant les cas.

Définition. *Une* **variable aléatoire discrète** *est une variable aléatoire dont le support est un ensemble fini ou dénombrable.*

3.1.3 Fonction de répartition d'une variable aléatoire

Définition. *La* **fonction de répartition** *d'une variable aléatoire X est la fonction définie pour tout $t \in \mathbf{R}$ par*

$$\boxed{F_X(t) = P(X \leq t).}$$

Autrement dit, $F_X(t)$ est la probabilité de l'événement "la valeur de X est inférieure ou égale à t".

Exemple 1 : X est le résultat du lancer d'un dé. On a vu que $P(X = 1) = P(X = 2) = \cdots = P(X = 6) = \frac{1}{6}$: c'est la loi de X. On peut la présenter sous forme de tableau :

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

On peut aussi présenter ce résultat sous la forme d'un graphique, ou diagramme en bâtons (figure 3.1 à gauche). Enfin on peut aussi tracer le graphe de la fonction de répartition F_X (figure 3.1 à droite). On voit ici que la fonction de répartition est constante par morceaux : elle présente des sauts pour les valeurs du support de X, mais reste constante entre deux de ces valeurs. Ce sera toujours le cas pour des variables discrètes.

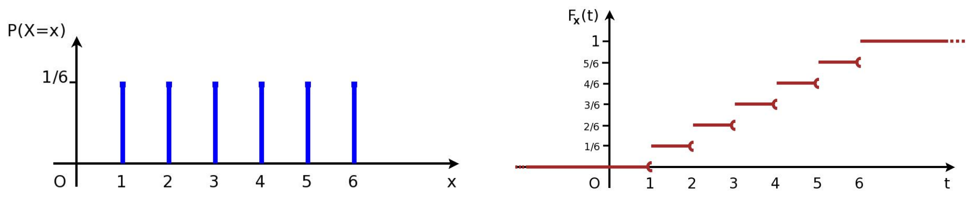


FIGURE 3.1 – Loi et fonction de répartition d'un lancer de dé

Proposition. *On a $F_X(t) \in [0,1]$ pour tout $t \in \mathbf{R}$; et F_X est une fonction croissante.*

Preuve. $F_X(t) \in [0,1]$ car c'est une probabilité. De plus, si $t \leq u$, on a $[X \leq t] \subset [X \leq u]$ et donc $P(X \leq t) \leq P(X \leq u)$, c'est-à-dire $F_X(t) \leq F_X(u)$. Donc F_X est croissante. □

Grâce à la fonction de distribution, on peut calculer sa probabilité de tomber dans un intervalle quelconque : Pour $a, b \in \mathbb{R}$ tel que $a \leqslant b$,

$$P(X \in]a,b]) = P((X \leqslant b) \setminus (X \leqslant a)) = P(X \leqslant b) - P(X \leqslant a) = F_X(b) - F_X(a).$$

3.1.4 Egalité en loi

On appelle $X \in \{0,1\}$ le résultat d'une pièce lancée à Pile-ou-face (0=Pile, 1=Face). On appelle Y la variable inverse, c'est-à-dire,

$$Y = \begin{cases} 0 \text{ si } X = 1 \\ 1 \text{ si } X = 0. \end{cases}$$

Il est clair que $Y \neq X$. Pourtant, on peut dire que X et Y sont **égales en loi**, ou **ont la même loi**. Cela veut dire que pour toute valeur x,

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

En effet, il suffit de vérifier que

$$P(X = 0) = 1/2 = P(Y = 0),$$
$$P(X = 1) = 1/2 = P(Y = 1).$$

Ainsi, on peut avoir deux variables X et Y qui sont égales en loi mais qui sont différentes. Pour décrire la phénomène de pile-ou-face à une autre personne, la variable X ou la variable Y peuvent aussi bien convenir l'une que l'autre, même si Y ne vient pas d'un rééI tirage. Il suffit de savoir qu'une éventualité à une probabilité $1/2$, et l'autre également.

Définition. *On dit que deux variables aléatoires X et Y ont la même loi si elles ont la même fonction de répartition. On le note $Y \sim X$.*

Attention : $X \sim Y$ n'implique pas forcément que $X = Y$. Par contre $X = Y$ implique bien $X \sim Y$.

Dans l'exemple précédent, on observe que pour $x \in \mathbb{R}$.

$$F_X(x) = P(X \leqslant x) = \begin{cases} 0 \text{ si } x < 0 \\ 1/2 \text{ si } 0 \leqslant x < 1 \\ 1 \text{ sinon} \end{cases} = P(Y \leqslant x)$$

Cette définition fonctionnera également avec les variables continues.

Exemple 1. *Soit X le résultat d'un lancer de dé. La variable aléatoire $Y = 7 - X \in \{6, 5, 4, 3, 2, 1\}$ a la même loi que X. En effet,*

$$P(Y = 1) = P(X = 6) = 1/6 = P(X = 1).$$

3.1.5 Variables discrètes

Loi d'une variable discrète. *Donner la loi d'une variable aléatoire discrète X, c'est calculer les probabilités $P(X = x)$ pour toutes les valeurs x possibles prises par X (autrement dit pour tous les x appartenant au support de X). Ainsi, deux variables discrètes X et Y ont la même loi si $P(X = x) = P(Y = y)$ pour tout x, y dans leur support.* **Attention ! Ceci est faux pour les variables aléatoires dites continues (ou "à densité"), que l'on verra plus tard.**

Lorsque toutes les probabilités formant la loi de X sont égales, comme dans l'exemple du dé, on parle de loi uniforme. C'est l'exemple le plus simple de variable aléatoire.

Définition. *Soit un ensemble fini $E = \{x_1, x_2, \ldots, x_n\}$. On dit que la variable X suit la* **loi uniforme** *sur E lorsque $\mathcal{S}(X) = E$ et $P(X = x_i) = \frac{1}{n}$ pour tout $1 \leq i \leq n$. On note $X \sim \mathcal{U}_E$.*

Exemple 2. *Si X est le résultat d'un lancer de dés, $X \sim \mathcal{U}_E$ avec $n = 6$ et $E = \{1, 2, 3, 4, 5, 6\}$.*

Si X est le résultat d'un pile-ou-face, $X \sim \mathcal{U}_E$ avec $n = 2$ et $E = \{$"pile", "face"$\}$.

Voici un deuxième exemple élémentaire de variable aléatoire discrète :

Exemple 2 *(Loi binomiale pour $n = 3$) : On lance trois pièces de monnaie. Soit X le nombre de "Face" obtenu. Quelle est la loi de X ?*

Le support de X est ici $\{0, 1, 2, 3\}$. On doit donc calculer $P(X = 0), P(X = 1), P(X = 2)$ et $P(X = 3)$.

Définissons F_1="la première pièce tombe sur Face" ; et de même F_2 et F_3. On peut clairement supposer que les trois lancers de pièce sont indépendants ici ; et donc que F_1, F_2, F_3 sont indépendants.

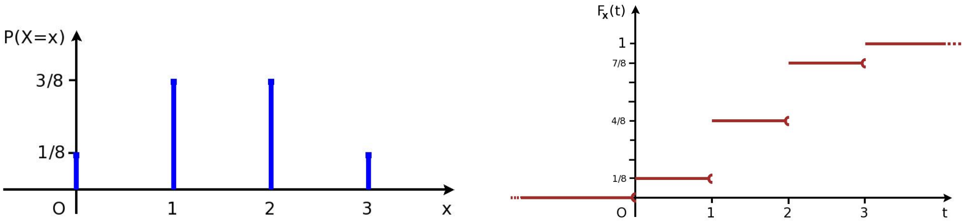


FIGURE 3.2 – Loi et fonction de répartition de X pour l'exemple 2

- $[X=0]=F_1^c\cap F_2^c\cap F_3^c$ donc $P(X=0)=P(F_1^c)P(F_2^c)P(F_3^c)$ grâce à l'indépendance. Ainsi $P(X=0)=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{8}$.
- $[X=1]=(F_1\cap F_2^c\cap F_3^c)\cup(F_1^c\cap F_2\cap F_3^c)\cup(F_1^c\cap F_2^c\cap F_3)$. Cette union est disjointe, donc on peut additionner les probabilités :

$$\begin{aligned}
P(X=1) &= P(F_1\cap F_2^c\cap F_3^c)+P(F_1^c\cap F_2\cap F_3^c)+P(F_1^c\cap F_2^c\cap F_3),\\
&= P(F_1)P(F_2^c)P(F_3^c)+P(F_1^c)P(F_2)P(F_3^c)+P(F_1^c)P(F_2^c)P(F_3),\\
&\qquad\qquad\text{(par indépendance)}\\
&= (\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2})\times 3=\frac{3}{8}.
\end{aligned}$$

- Pour calculer intelligemment $P(X=2)$, on introduit Y le nombre de "pile" obtenus. Il est clair que le nombre de "pile" suit la même loi que le nombre de "face", donc $Y\sim X$. On a de plus $Y=X-3$. Donc $P(X=2)=P(Y=1)=P(X=1)=3/8$.
- Avec le même raisonnement, $P(X=3)=P(Y=0)=P(X=0)=1/8$.

La loi de X est en fait un exemple de loi binomiale. Le cas général sera vu un peu plus loin. La figure 3.2 montre le graphique de cette loi ainsi que la fonction de répartition.

Exemple 3 *(Loi géométrique) : On lance un dé et on recommence jusqu'à l'apparition du premier* 6 *(par exemple* $[X=5]=$*"on obtient* 6 *pour la première fois au* 5^e *lancer"). Déterminer la loi de* X*.*

Le support de X est ici $\{1,2,3,\ldots\}=\mathbf{N}^*$. En toute rigueur il faudrait ici remarquer que X n'est pas nécessairement défini en envisageant le cas où le 6 n'apparaît jamais ; ou bien rajouter la valeur $+\infty$ au support en définissant l'événement $[X=+\infty]=$"le 6 n'apparaît jamais". Cependant on peut montrer que la probabilité de cet événement est nulle, et donc qu'il n'y a pas lieu de le prendre en compte.

Pour tout $n\geq 1$, notons A_n l'événement "le n^e lancer vaut 6". A_n ne dépend que du résultat du n^e lancer, et il est clair que les lancers sont des expériences indépendantes. Par conséquent les A_n sont des événements indépendants.

- $[X=1]=A_1$, donc $P(X=1)=P(A_1)=\frac{1}{6}$,
- $[X=2]=A_1^c\cap A_2$, donc $P(X=2)=P(A_1^c)P(A_2)$ par indépendance. Donc $P(X=2)=\frac{5}{6}\times\frac{1}{6}$.

- $[X=3] = A_1^c \cap A_2^c \cap A_3$ donc $P(X=3) = P(A_1^c)P(A_2^c)P(A_3)$ grâce à l'indépendance. Donc $P(X=3) = (\frac{5}{6})^2 \times \frac{1}{6}$.
- Plus généralement, on aura $[X=n] = A_1^c \cap A_2^c \cap \cdots \cap A_{n-1}^c \cap A_n$, donc $P(X=n) = P(A_1^c)P(A_2^c)\dots P(A_{n-1}^c)P(A_n)$, toujours grâce à l'indépendance. Ainsi $P(X=n) = (\frac{5}{6})^{n-1}\frac{1}{6}$.

Finalement cette dernière formule est valable pour tout $n \geq 1$; on a donc ainsi bien déterminé la loi de X. Cette loi s'appelle la loi géométrique de paramètre $\frac{1}{6}$.

Définition. *On dit qu'une variable aléatoire suit la* **loi géométrique** *de paramètre p, où p est un nombre $p \in [0,1]$ fixé, si le support de X est égal à $\mathbf{N}^*$ et que*

$$P(X=n) = (1-p)^{n-1}p.$$

On note $\mathcal{G}(p)$ la loi géométrique de paramètre p, et on note $X \sim \mathcal{G}(p)$ pour dire que X suit une loi géométrique de paramètre p.

3.1.6 Quelques lois discrètes usuelles

On a déjà défini les lois **uniformes** et **géométriques**. Voici quelques autres lois classiques :

Loi de Bernoulli. C'est la loi la plus élémentaire : 0 ou 1.

Définition. *Une variable aléatoire X suit la* **loi de Bernoulli** *de paramètre $p \in [0,1]$ si $X \in \{0,1\}$ et $P(X=1) = p$, $P(X=0) = 1-p$. On note $X \sim \mathcal{B}(p)$.*

Exemple 3. — *On jette une pièce en l'air. Soit X la variable*

$$X = \begin{cases} 1 \text{ si la pièce tombe sur pile} \\ 0 \text{ sinon.} \end{cases}$$

Alors $X \sim \mathcal{B}(1/2)$.

— *On jette un dé. Soit la variable*

$$X = \begin{cases} 1 \text{ si le dé fait 6} \\ 0 \text{ sinon.} \end{cases}$$

Alors $X \sim \mathcal{B}(1/6)$.

Loi binomiale. C'est la loi du nombre de succès lors de n essais indépendants d'une même expérience probabiliste, qui a une probabilité p d'être réussie (donc p variables binomiales indépendantes). Nous allons la calculer sur un exemple simple.

exemple : *On lance n dés et on note X le nombre de fois que l'on obtient* 6. *Quelle est la loi de X ?*

Cet exemple est simplement la généralisation de l'exemple 2 de la section précédente. Tout d'abord le support de X est $\{0, 1, 2, \dots, n\}$. Notons A_i l'événement "On obtient 6 au i-ème lancer". Tous ces événements sont indépendants. On a :

$[X = 0] = A_1^c \cap A_2^c \cap \cdots \cap A_n^c$ donc $P(X = 0) = P(A_1^c)P(A_2^c)\cdots P(A_n^c) = (\frac{5}{6})^n$.

Pour l'événement $[X = 1]$ on peut le décomposer ainsi : $[X = 1] = (A_1 \cap A_2^c \cap \cdots \cap A_n^c) \cup (A_1^c \cap A_2 \cap \cdots \cap A_n^c) \cup \cdots \cup (A_1^c \cap A_2^c \cap \cdots \cap A_n)$, donc

$$\begin{aligned} P(X = 1) &= P(A_1 \cap A_2^c \cap \cdots \cap A_n^c) + P(A_1^c \cap A_2 \cap \cdots \cap A_n^c) + \cdots + P(A_1^c \cap A_2^c \cap \cdots \cap A_n), \\ &= P(A_1)P(A_2^c)\cdots P(A_n^c) + P(A_1^c)P(A_2)\cdots P(A_n^c) + \cdots + P(A_1^c)P(A_2^c)\cdots P(A_n), \\ &= \frac{1}{6}(\frac{5}{6})^{n-1} + \frac{1}{6}(\frac{5}{6})^{n-1} + \cdots + \frac{1}{6}(\frac{5}{6})^{n-1} = n\frac{1}{6}(\frac{5}{6})^{n-1}. \end{aligned}$$

Pour $[X = 2]$ et les suivants, on procède de la même manière, mais il devient compliqué de l'écrire in-extenso. On voit qu'à chaque fois on décompose l'événement $[X = k]$ en considérant tous les cas possibles pour les k succès envisagés. Pour chacun de ces cas la probabilité sera la même : $(\frac{1}{6})^k(\frac{5}{6})^{n-k}$; et il reste donc à compter le nombre de ces cas. Il s'agit de compter le nombre de choix possibles pour les positions des k succès parmi les n essais. Il y en a $\binom{n}{k}$. Ainsi $P(X = k) = \binom{n}{k}(\frac{1}{6})^k(\frac{5}{6})^{n-k}$.

Définition. *Une variable aléatoire suit la loi binomiale de paramètres $n \in \mathbf{N}$ et $p \in [0, 1]$ lorsque $X \in \{0, 1, 2, \dots, n\}$ et*

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

La loi binomiale de paramètres n et p se note $\mathcal{B}(n, p)$ On note alors $X \sim \mathcal{B}(n, p)$.

Exemple 4.
- *Je choisis au hasard* 6 *étudiants dans la promo, et j'appelle X le nombre de ces étudiants qui auront plus de* 15 *en note finale. Comme j'ai choisi les étudiants au hasard, les succès des uns et des autres sont des variables aléatoires* **indépendantes** *les unes des autres. En effet, savoir qu'un des étudiants a eu 15 ne me donne aucun renseignement sur les résultats des autres. Donc, en supposant que* $2/3$ *des étudiants de la promo auront plus de* 15*, $X \sim \mathcal{B}(6, 2/3)$.*
- *Je choisis cette fois un étudiant au hasard dans la promo, et je considère les notes qu'il aura aux* 6 *matières du semestre. Je note X le nombre de ces notes qui seront supérieures à* 15. *Cette fois, X ne suit pas une loi binomiale car les résultats de cet étudiant aux différentes matières dépendent de certains facteurs : Le niveau de l'étudiant, la quantité de travail qu'il va fournir, etc...Par exemple, si je sais que l'étudiant a eu 15 a la 1re matière, c'est probablement un bon étudiant, et il aura probablement de bonnes notes dans les autres matières...*

La figure 3.3 montre des exemples de lois binomiales pour différents paramètres.

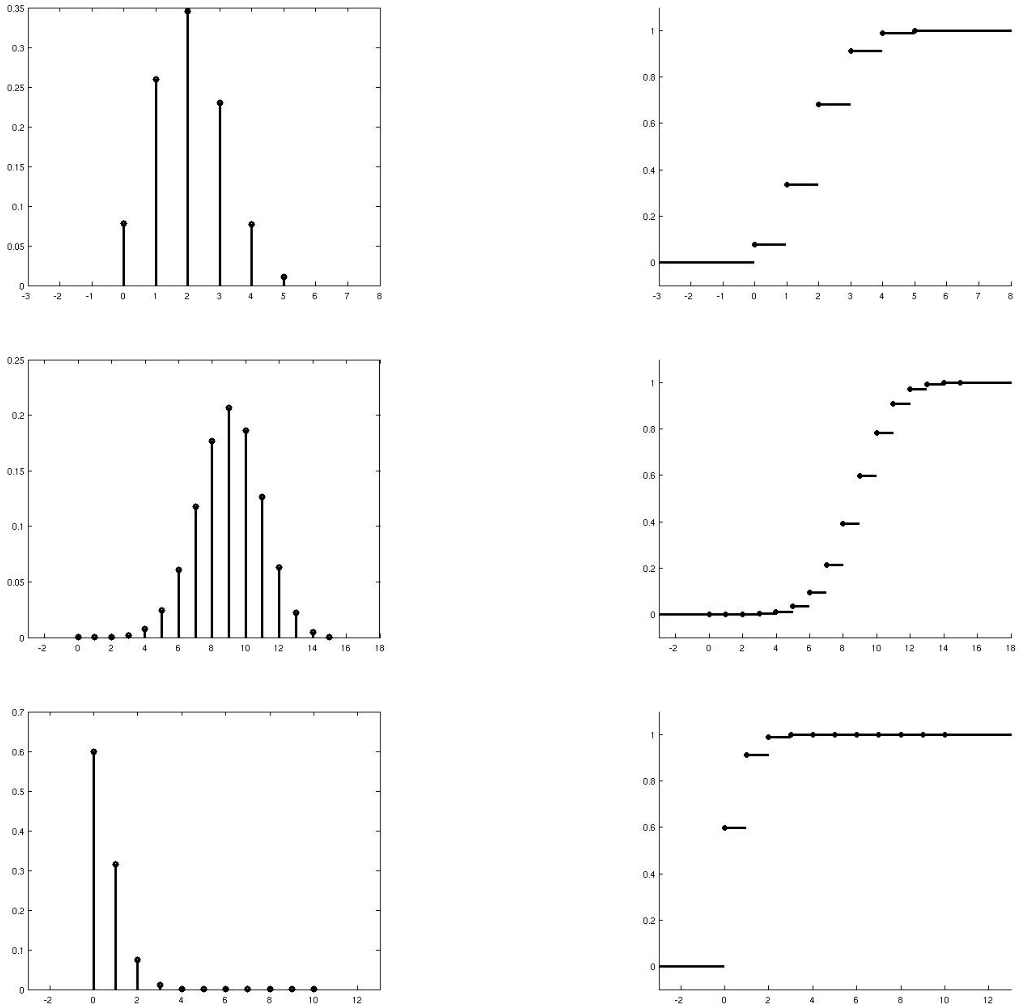


FIGURE 3.3 – lois binomiales (à gauche) et fonctions de répartition F_X correspondantes (à droite) pour différents paramètres : 1e ligne : $n = 5, p = 0.4$; 2e ligne : $n = 15, p = 0.6$; 3e ligne : $n = 10, p = 0.05$.

Loi de Poisson. La loi de Poisson sera vu plus en détail par la suite. Nous donnons simplement la définition ici :

Définition. *Soit* $\lambda \geq 0$. *Une variable* X *suit la* **loi de Poisson** *de paramètre* λ *si pour tout entier* $k \geq 0$*,*

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

Le support de X *est ici* $\mathbf{N}$. *La loi de Poisson de paramètre* λ *est notée* $\mathcal{P}(\lambda)$ *(et donc* $X \sim \mathcal{P}(\lambda)$*).*

3.1.7 Les événements élémentaires $[X = x]$

1) Les événements $[X = x]$, considérés pour tous les x appartenant au support, forment une partition de Ω. En effet, X prend nécessairement une et une seule de ces valeurs, ce qui prouve bien que la réunion des $[X = x]$ est égale à Ω, et que $[X = x] \cap [X = y] = \varnothing$ si $x \neq y$. Par conséquent la somme totale des $P(X = x)$ doit toujours être égale à 1, ce qui s'écrit, en notant $\mathcal{S}$ le support de X,

$$\sum_{x \in \mathcal{S}} P(X = x) = 1.$$

Il faut toujours penser à le vérifier lorsqu'on calcule une loi. Pour les exemples précédents, on a :

- pour le dé : $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$,
- pour l'exemple 2 : $\frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$,
- pour la loi géométrique :

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

2) De plus ces événements $[X = x]$ sont les événements élémentaires pour la variable X, au sens où tout événement relatif à X s'exprime comme une union de ces événements, et sa probabilité est la somme des $P(X = x)$ correspondants.

exemple : L'événement "le dé tombe sur un nombre pair supérieur à 3" est égal à $[X = 4] \cup [X = 6]$ si X est le résultat du dé. Sa probabilité est donc de $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

exemple : Si X est une variable discrète quelconque, l'événement $[X \leq t]$ est l'union des $[X = x]$ pour tous les x appartenant au support de X tels que $x \leq t$. On peut ainsi donner une formule pour la fonction de répartition d'une variable discrète :

$$F_X(t) = \sum_{x \in \mathcal{S}, x \leq t} P(X = x).$$

3.2 Variables aléatoires à densité

3.2.1 Exemple introductif

On jette un stylo sur une table, et on note X l'angle entre 0 et π qu'il forme avec le bord de la table. Quelle est la loi de X ?

L'ensemble des valeurs possibles pour cette variable aléatoire est $[0,\pi]$. En suivant l'idée vue auparavant, on voudrait donc chercher à calculer tous les $P(X=\alpha)$ pour $\alpha \in [0,\pi]$. En fait on verra que $P(X=\alpha)$ sera toujours égal à 0, ce qui signifie que quelle que soit la valeur de α, le stylo n'a aucune chance de former exactement l'angle α avec la table. On est obligé pour donner un sens aux probabilités ici, de considérer des intervalles et non des valeurs uniques, et de regarder les probabilités que X soit dans ces intervalles. Par exemple on peut raisonnablement penser ici que $P(0 \leq X \leq \frac{\pi}{2}) = \frac{1}{2}$, ou que $P(\frac{\pi}{4} \leq X \leq \frac{\pi}{2}) = \frac{1}{4}$. Plus généralement, on peut raisonnablement supposer que la probabilité que X appartienne à un intervalle $[\alpha,\beta] \subset [0,\pi]$ correspond à la proportion d'angles compris dans cet intervalle, c'est-à-dire que

$$P(\alpha \leq X \leq \beta) = \frac{\beta-\alpha}{\pi}.$$

Ceci permet de caractériser entièrement la variable X car la probabilité de tout événement lié à X peut se calculer à partir de cette formule.

On va néanmoins ici pouvoir donner une valeur de probabilité associée à un angle α donné en considérant un petit intervalle $[\alpha-\varepsilon,\alpha+\varepsilon]$: on a

$$P(\alpha-\varepsilon \leq X \leq \alpha+\varepsilon) = \frac{(\alpha+\varepsilon)-(\alpha+\varepsilon)}{\pi} = \frac{2\varepsilon}{\pi}.$$

On obtient ce qu'on souhaite en prenant la limite pour ε tendant vers 0 de

$$\frac{P(\alpha-\varepsilon \leq X \leq \alpha+\varepsilon)}{2\varepsilon},$$

c'est-à-dire du rapport entre la probabilité de l'intervalle et la longueur de l'intervalle. Ici ce rapport vaut $\frac{1}{\pi}$, donc sa limite aussi. Cette valeur $\frac{1}{\pi}$ est appelée densité de X en α. Inversement, si l'on connaît la densité de X en tout point $x \in [0,\pi]$ on obtiendra une probabilité du type $P(\alpha \leq X \leq \beta)$ en calculant l'intégrale de cette densité sur l'intervalle $[\alpha,\beta]$.

3.2.2 Définition

Une variable aléatoire X est dite **à densité** *lorsqu'il existe une fonction $f_X : \mathbf{R} \to \mathbf{R}^+$ telle que*

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx \qquad \textit{pour tous } a,b \in \mathbf{R}, \quad a \leq b.$$

Cette fonction f_X est appelée **densité** *de X.*

Reprenons l'exemple du stylo qui tombe sur le bureau. C'est en fait une variable de densité

$$f(x) = \begin{cases} \frac{1}{\pi} \text{ si } x \in [0, \pi], \\ 0 \text{ sinon.} \end{cases}$$

En effet, pour $[a, b] \subset [0, \pi]$,

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx = \frac{b-a}{\pi}.$$

Si par exemple $a < 0, b \in [0, \pi]$,

$$P(a \leqslant X \leqslant b) = P(0 \leqslant X \leqslant b) = \frac{b}{\pi} = \int_0^b f(x)dx = \int_a^b f(x)dx$$

car $f = 0$ sur $[a, 0]$. La formule est bien vérifiée dans ce cas de figure, et idem si $b > \pi$. Donc la formule marche dans tous les cas de figure. On aura donc par exemple :

$$P\left(X \leq \frac{\pi}{2}\right) = \int_0^{\frac{\pi}{2}} \frac{1}{\pi} dx = \frac{1}{2},$$

ou encore :

$$P\left(\frac{\pi}{4} \leq X \leq \frac{\pi}{2}\right) = \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} \frac{1}{\pi} dx = \frac{1}{4}.$$

C'est en fait un cas particulier de la loi uniforme.

Définition. *La loi uniforme sur un intervalle $[\alpha, \beta]$ est la loi de densité*

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} \text{ si } x \in [\alpha, \beta], \\ 0 \text{ sinon.} \end{cases}$$

On note $X \sim \mathcal{U}([\alpha, \beta])$ ("X suit la loi uniforme sur $[\alpha, \beta]$").

Remarque 1. *Ne pas confondre loi uniforme pour une variable discrète et pour une variable continue !*

remarque : on peut prendre $a = -\infty$ ou $b = +\infty$ dans cette formule. On a par exemple

$$P(-\infty \leqslant X \leqslant +\infty) = \int_{-\infty}^{+\infty} f_X(x)dx$$

et bien évidemment $P(-\infty \leqslant X \leqslant +\infty) = 1$, donc toute fonction de densité doit avoir une intégrale entre $-\infty$ et $+\infty$ égale à 1. Il faut bien penser à le vérifier.

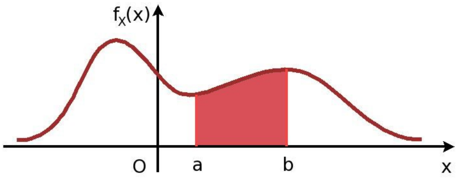


FIGURE 3.4 – Graphe d'une densité de probabilité f_X. La partie colorée correspond à $P(a \leq X \leq b)$.

Interprétation graphique : La probabilité $P(a \leq X \leq b)$ correspond à l'aire du domaine situé sous le graphe de f_X entre les abscisses a et b (voir figure 3.4).

remarque 1 : on a bien $P(X = a) = P(a \leq X \leq a) = \int_a^a f_X(x)dx = 0$. Du coup,

$$P(a \leqslant X \leqslant b) = P(\text{“}a < X \leqslant b\text{''} \cup \text{“}X = a\text{''}) = P(a < X \leqslant b) + P(X = a) = P(a < X \leqslant b).$$

Pour des raisons similaires,

$$P(a \leqslant X \leqslant b) = P(a < X \leqslant b) = P(a \leqslant X < b) = P(a < X < b).$$

Il n'est pas important de préciser si les frontières de l'intervalle sont incluses ou non dans ce genre de calcul, **pour des variables a densité**.

remarque 3 : Calculer la loi d'une variable à densité, c'est calculer sa densité (de la même manière que pour calculer la loi d'une variable discrète, il fallait calculer les $P(X = x)$, pour $x \in \mathcal{S}(X)$).

Il existe des variables aléatoires qui ne sont pas discrètes, et qui ne sont pas non plus à densité, mais on ne les évoquera pas dans ce cours.

3.2.3 Loi exponentielle

Soit $a > 0$ un réel. On dit que X suit la loi exponentielle de paramètre a si elle admet la densité

$$f_X(x) = ae^{-ax}1_{\mathbf{R}^+}(x) = \begin{cases} ae^{-ax} \text{ si } x \geq 0, \\ 0 \text{ sinon.} \end{cases}$$

On note $X \sim \mathcal{E}(a)$. On vérifie que

$$\int_{-\infty}^{+\infty} f_X(x) = a\int_0^{+\infty} e^{-ax}dx = a\left[\frac{e^{-ax}}{-a}\right]_0^{+\infty} = a(0 - (-1/a)) = 1.$$

Remarque 2. *Il existe des variables aléatoires qui ne sont ni discrètes, ni à densité.*

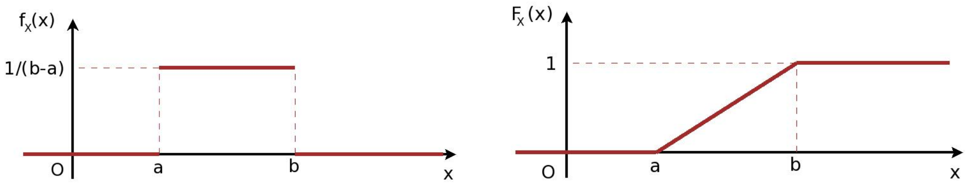


FIGURE 3.5 – Graphes de la densité et de la fonction de répartition de la loi uniforme sur un intervalle $[a, b]$.

3.2.4 Fonction de répartition d'une variable aléatoire à densité

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.$$

Ainsi F_X est une primitive de la fonction densité f_X.

exemple : Si X suit la loi uniforme sur $[\alpha, \beta]$, on a

$$f_X(x) = \begin{cases} \frac{1}{\beta-\alpha} \text{ si } x \in [\alpha, \beta], \\ 0 \text{ sinon,} \end{cases}$$

et donc

$$F_X(x) = \begin{cases} 0 \text{ si } x \leq \alpha \\ \frac{x-\alpha}{\beta-\alpha} \text{ si } x \in [\alpha, \beta], \\ 1 \text{ si } x \geq \beta \end{cases}$$

La densité et la fonction de répartition d'une loi uniforme sont représentées sur la figure 3.5.

Pour $X \sim \mathcal{E}(a)$, sa fonction de répartition est :

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 1 - e^{-ax} \text{ si } x \geq 0, \\ 0 \text{ sinon.} \end{cases}$$

La figure 3.6 montre densités et fonctions de répartition de la loi exponentielle pour deux valeurs différentes du paramètre a.

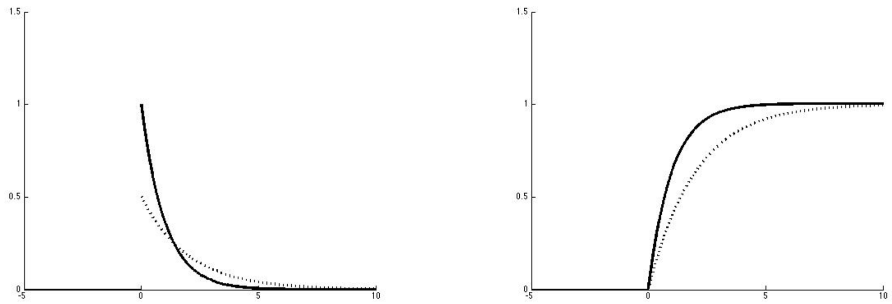


FIGURE 3.6 – Graphes de la densité et de la fonction de répartition de la loi exponentielle pour $a = 1$ (traits pleins) et $a = 0.5$ (pointillés).

3.3 Espérance et variance d'une variable aléatoire

3.3.1 Définitions

3.4 Espérance et variance d'une variable aléatoire

3.4.1 Quelques rappels sur les sommes, séries et intégrales

Convergence absolue.

Une **série** $\sum_{n=0}^{+\infty} a_n$ est dite **convergente** si les sommes $\sum_{n=0}^{N} a_n$ convergent lorsque N tend vers $+\infty$, et on note alors $\sum_{n=0}^{+\infty} a_n = \lim_{N\to+\infty} \sum_{n=0}^{N} a_n$.

Une **série** $\sum_{n=0}^{+\infty} a_n$ est dite **absolument convergente** lorsque $\sum_{n=0}^{+\infty} |a_n|$ est une série convergente. Dans ce cas $\sum_{n=0}^{+\infty} a_n$ est convergente, et de plus il est permis de sommer les termes a_n dans n'importe quel ordre : $\sum_{n=0}^{+\infty} a_n = a_0 + a_1 + a_2 + a_4 + \cdots = a_3 + a_{10} + a_8 + a_5 + a_4 + a_7 + \cdots$ (on somme tous les a_n mais dans un ordre différent). Pour bien mettre en valeur cette propriété, on pourra noter cette somme $\sum_{n\in\mathbf{N}} a_n$: somme des a_n pour tous les $n \in \mathbf{N}$, quel que soit l'ordre d'énumération.

De la même manière une intégrale $\int_a^b f(x)dx$ est dite **absolument convergente** lorsque $\int_a^b |f(x)|dx$ est une intégrale convergente.

En probabilités on ne s'intéressera qu'aux séries et intégrales absolument convergentes.

Quelques techniques de calcul.

- L'indice utilisé dans une notation $\sum$ est "muet", c'est-Ãă-dire qu'il n'a pas d'existence en dehors de cette somme. On peut donc le remplacer par n'importe quelle autre lettre :
$$\sum_{k=1}^{n} a_k = \sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j.$$
- Décalage d'indices :
$$\begin{aligned} \sum_{i=0}^{n-1} a_{i+1} &= \sum_{k=1}^{n} a_k \qquad \text{(en posant k=i+1)} \\ &= \sum_{i=1}^{n} a_i \qquad \text{(puisque l'indice est "muet")} \end{aligned}$$
Pour une série on aura de même
$$\sum_{i=0}^{+\infty} a_{i+1} = \sum_{i=1}^{+\infty} a_i.$$
- Factorisation : $\displaystyle\sum_{i=0}^{n} x a_i = x \sum_{i=0}^{n} a_i$
- Série géométrique :
$$\sum_{i=0}^{n} x^i = \frac{1-x^{n+1}}{1-x} \qquad \text{si } x \neq 1.$$
$$\sum_{n=0}^{+\infty} x^n = \frac{1}{1-x} \qquad \text{si } |x| < 1.$$
Par exemple,
$$\sum_{i=3}^{n} x^i = \sum_{i=0}^{n-3} x^{i+3} = x^3 \sum_{i=0}^{n-3} x^i = x^3 \frac{1-x^{n-2}}{1-x}.$$
- Séries entières : Dans certains cas on peut obtenir un développement en série entière d'une fonction $f(x)$:
$$f(x) = \sum_{n=0}^{+\infty} a_n x^n.$$
Par exemple
$$\frac{1}{1-x} = \sum_{n=0}^{+\infty} x^n \qquad \text{pour } x \in]-1,1[,$$

$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \qquad \text{pour tout } x \in \mathbf{R}.$$

On a alors le droit de dériver la série entière :

$$f'(x) = \sum_{n=1}^{+\infty} n a_n x^{n-1} = \sum_{n=0}^{+\infty} (n+1) a_{n+1} x^n.$$

Cette propriété est très utile pour les calculs d'espérance et de variance.

3.4.2 Espérance

L'espérance d'une variable aléatoire X est la moyenne des valeurs prises par la variable, pondérées par leurs probabilités. On la note $E(X)$. Avant de voir des définitions plus précises, voyons un exemple :

exemple : Un joueur joue à la roulette (18 numéros rouge, 18 numéros noirs, 1 numéro vert). Il parie 1 euro sur le noir. Quelle est son espérance de gain ?

Son gain X est une variable discrète dont la loi est la suivante :

— $P(X=1) = \frac{13}{2*13+1} = \frac{18}{37} < 1/2$.

— $P(X=-1) = \frac{13+1}{2*13+1} = \frac{19}{37} > 1/2$.

L'espérance de gain est donc

$$E(X) = 1 \times \frac{13}{37} + (-1) \times \frac{14}{37} = -\frac{1}{37}.$$

Elle est (légèrement) négative.

On verra plus tard que l'espérance correspond aussi à l'idée de valeur moyenne obtenue lorsqu'on répète un grand nombre de fois la même expérience : intuitivement, dans l'exemple précédent, si on joue à ce jeu un grand nombre n de fois, on perdra environ $-n/37$ euros.

Définition. *L'***espérance** *d'une variable discrète X de support $\mathcal{S}(X)$ est :*

$$\boxed{E(X) = \sum_{x \in \mathcal{S}(X)} x P(X=x),}$$

lorsque cette somme est bien définie.

Rappel Si $\mathcal{S}(X)$ est fini, ce calcul ne pose pas de problème. Sinon, on a affaire à une somme infinie de termes,et il faut faire appel à la théorie des séries. Cette somme est bien définie uniquement si la série des termes est absolument convergente, c'est-à-dire si

$$\sum_{x \in \mathcal{S}(X)} |x| P(X=x) < \infty.$$

Le fait que la série soit convergente ne suffit pas. Si la série n'est pas absolument convergence, l'espérance n'est tout simplement pas définie.

Définition. *L'***espérance** *d'une variable aléatoire X de densité f sur $\mathbf{R}$ est :*

$$\boxed{E(X) = \int_{\mathbf{R}} x f(x) dx,}$$

lorsque cette intégrale est bien définie.

Cette intégrale n'est bien définie que si elle est absolument convergente, c'est-à-dire

$$\int_{\mathbf{R}} |x| f(x) dx < \infty.$$

Ici encore cette intégrale peut ne pas être absolument convergente ; dans ce cas l'espérance n'est pas définie.

L'espérance ne dépend que de la loi. Autrement dit, si deux variables X et Y (discrètes ou continues) ont la même loi, et qu'elles ont une espérance finie, alors elles ont la même espérance.

La notion d'espérance rejoint la notion de moyenne.

Donnons un exemple avec une variable continue :

Exemple 5. Espérance de la loi exponentielle *On rappelle la densité d'une variable $X \sim \mathcal{E}(a)$, ou $a > 0$:*

$$f_X(x) = \begin{cases} ae^{-ax} si\ x > 0 \\ 0\ sinon. \end{cases}$$

On a avec une intégration par parties

$$E(X) = a \int_0^{\infty} x e^{-ax} dx = [\frac{axe^{-ax}}{-a}]_0^{\infty} - a \int_0^{\infty} \frac{e^{-ax}}{-a} dx = 0 - 0 + [\frac{e^{-ax}}{-a}]_0^{\infty} = \frac{1}{a}.$$

L'espérance vaut $1/a$. (Comme on intègre une fonction positive, l'intégrale existe ssi elle est absolument convergente, et le fait que le calcul précédent donne une valeur finie prouve qu'elle est bien convergente)

Quelques propriétés de l'espérance.

- L'espérance est une fonction linéaire : si a, b sont des réels, et X, Y des variables aléatoires,
$$E(X + Y) = E(X) + E(Y), E(aX) = aE(X),$$
$E(X - Y) = E(X) + E(-Y) = E(X) - E(Y), \ldots.$
Attention : La loi de $X + Y$ ne se calcule pas en faisant la somme des lois de X et de Y (ce qui n'aurait aucun sens).
- Si $a \in \mathbf{R}$, $E(a) = a$ puisque a est non aléatoire : autrement dit, a s'interprète comme une variable aléatoire prenant la valeur a avec probabilité 1.

- Soient X une variable aléatoire, et $g : \mathbf{R} \to \mathbf{R}$ une fonction quelconque. L'espérance de $g(X)$ se calcule ainsi :

$$E(g(X)) = \sum_{x \in \mathcal{S}(X)} g(x)P(X = x) \qquad \text{si } X \text{ est une variable discrète,}$$

$$E(g(X)) = \int_{\mathbf{R}} g(x)f_X(x)dx \qquad \text{si } X \text{ est une variable à densité.}$$

Ici encore ces sommes ou ces intégrales peuvent ne pas être absolument convergentes. Par exemple il se peut très bien que $E(g(X))$ n'existe pas alors que $E(X)$ existe.

Exemple 6. *Julien et Maxime arrivent à l'heure à un rendez-vous avec une troisième personne, Aurore. Ils décident de parier sur le retard qu'elle aura. Julien dit à Maxime : "Je te donne 3 euros tout de suite, mais toi tu me donnes 5 euros par minute de retard".*

On admet que le temps de retard de Aurore en minutes est une variable exponentielle de paramètre 2. Quelle est l'espérance de gain de Julien ?

Appelons R le retard d'Aurore. Comme $R \sim \mathcal{E}(2)$, $E(R) = 1/2$. Le gain de Julien est donc

$$G = -3 + 5 * R.$$

On a donc

$$E(G) = E(-3 + 5 * R) = E(-3) + E(5R) = -3 + 5E(R) = -3 + \frac{5}{2} = -1/2.$$

Julien a donc une espérance de gain négative.

3.4.3 Variance

La variance permet de mesurer l'écart des valeurs de la variable par rapport à l'espérance :

Définition. *La* **variance** *d'une variable aléatoire X est*

$$\boxed{V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2.}$$

Montrons que cette dernière égalité est vraie : on a

$$E((X - E(X))^2) = E(X^2 + E(X)^2 - 2XE(X)) = E(X^2) + E(E(X)) - 2E(XE(X)),$$

par la propriété de linéarité. A présent $E(X)$ est un nombre réel ; on peut donc écrire $E(E(X)) = E(X)$, et $E(XE(X)) = E(X)E(X) = E(X)^2$. Ainsi

$$E((X - E(X))^2) = E(X^2) + E(X) - 2E(X)^2 = E(X^2) - E(X)^2.$$

La variance est un nombre positif qui peut être infini, même lorsque l'espérance est définie.

Définition. *L'* **écart-type** *d'une variable aléatoire X est la racine carrée de sa variance :*

$$\boxed{\sigma(X) = \sqrt{V(X)}.}$$

C'est une interprétation de la valeur de laquelle on va dévier de la moyenne, "en moyenne".

Exemple 7. *Deux amis, Jean et Jacques jouent aux dés de deux manières différentes.*

- *Dans la première manière, Jacques gagne 5 euros si le dé fait 6, et autrement c'est Jean qui gagne 1 euro.*
- *Dans la seconde manière, ils misent la même somme, et le premier gagne si le dé fait 1,2,3, et donc le second gagne si le fait fait 4,5, ou 6.*

Calculer dans chaque cas l'espérance, la variance et l'écart-type du gain du premier joueur.

Appelons X l'espérance du gain de Jean. **1er cas** *L'espérance de X vaut*

$$E(X) = (+1)P(1 \leqslant X \leqslant 5) - 5P(X = 6) = 1 \times \frac{5}{6} - 5 \times (1/6) = 0.$$

Dans le second cas,

$$E(X) = (+1)P(1 \leqslant X \leqslant 3) + (-1)P(4 \leqslant 5 \leqslant 6) = +1 \times \frac{1}{2} - 1 \times \frac{1}{2} = 0.$$

Donc l'espérance est nulle dans les deux cas, ce qui veut dire qu'en moyenne le jeu est équitable, et aucun joueur de va s'enrichir aux dépens de l'autre.

Calculons la variance. **1er cas**

$$\begin{aligned} V(X) =& E(X^2 - (\underbrace{E(X)}_{=0})^2) = E(X^2) = \sum_{y \in \mathcal{S}(X^2)} yP(X^2 = y) \\ =& \sum_{x \in \mathcal{S}(X)} x^2 P(X^2 = x^2) = \sum_{x \in \mathcal{S}(X)} x^2 P(X = x) \\ =& P(X = 1) \times 1^2 + P(X = -5) \times (-5)^2 = \frac{5}{6} \times 1 + \frac{1}{6} \times 25 = \frac{30}{6} = 5. \end{aligned}$$

Donc l'écart-type vaut $\sigma(X) = \sqrt{V(X)} = \sqrt{5} \approx 2,23$. On dit que X est égale à 0 à environ plus ou moins 2,23 près, ce qui se note aussi $X \approx 0 \pm 2,23$.

Second cas *De la même manière*

$$V(X) = E(X - 0)^2 = \sum_{x \in \mathcal{S}(X)} x^2 P(X = x) = (+1)^2 P(X = 1) + (-1)^2 P(X = -1) = 1 \times \frac{1}{2} + 1\frac{1}{2} = 1.$$

L'écart-type est $\sigma(X) = \sqrt{1} = 1$, donc dans ce cas-là Jean gagne 0 à ± 1 près.

On voit que le gain comporte moins de variabilité dans le second cas, ce qui est conforme à l'intuition.

3.4.4 Exemples de calculs

Espérance et variance d'une Bernoulli

Soit $X \sim \mathcal{B}(p)$. Alors

$$E(X) = 0.P(X = 0) + 1P(X = 1) = p.$$

et

$$V(X) = E(X^2) - (E(X))^2 = 0^2.P(X = 0) + 1^2P(X = 1) - p^2 = p - p^2 = p(1 - p).$$

Espérance et variance d'une variable géométrique

Une variable X de loi géométrique de paramètre $p \in [0, 1]$ a pour loi :

$$P(X = n) = (1 - p)^{n-1}p \qquad \text{pour } n \geq 1.$$

Calculons $E(X)$:

$$\begin{aligned} E(X) &= \sum_{n\geq1} nP(X = n) \\ &= \sum_{n\geq1} n(1 - p)^{n-1}p \\ &= p\sum_{n\geq1} n(1 - p)^{n-1}. \end{aligned}$$

Notons $f(x) = \sum_{n\geq1} nx^{n-1}$. $f(x)$ est la dérivée de $\sum_{n\geq1} x^n = \sum_{n\geq0} x^n - x^0 = \frac{1}{1-x} - 1$, donc $f(x) = \left(\frac{1}{1-x} - 1\right)' = \frac{1}{(1-x)^2}$. Ainsi

$$E(X) = pf(1 - p) = p\frac{1}{p^2} = \frac{1}{p}.$$

$$\boxed{E(X) = \frac{1}{p}}$$

Calculons $V(X)$:

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 \\ &= \sum_{n\geq1} n^2P(X = n) - \frac{1}{p^2} \\ &= p\sum_{n\geq1} n^2(1 - p)^{n-1} - \frac{1}{p^2}. \end{aligned}$$

Notons $f(x) = \sum_{n \geq 1} n^2 x^{n-1}$. On a

$$\begin{aligned}
f(x) &= \sum_{n \geq 1} n(n+1)x^{n-1} - \sum_{n \geq 1} n x^{n-1} \\
&= \left(\sum_{n \geq 1} x^{n+1} \right)'' - \left(\sum_{n \geq 1} x^n \right)' \\
&= \left(\sum_{n \geq 2} x^n \right)'' - \left(\sum_{n \geq 1} x^n \right)' \\
&= \left(\frac{1}{1-x} - 1 - x \right)'' - \left(\frac{1}{1-x} - 1 \right)' \\
&= \frac{2}{(1-x)^3} - \frac{1}{(1-x)^2},
\end{aligned}$$

et donc

$$V(X) = pf(1-p) - \frac{1}{p^2} = p\left(\frac{2}{p^3} - \frac{1}{p^2} \right) - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p}.$$

$$\boxed{V(X) = \frac{1-p}{p^2}}$$

Variance d'une variable de loi exponentielle

On a déjà vu que l'espérance d'une loi exponentielle de paramètre a vaut $1/a$ (à savoir). Calculons sa variance.

Soit X une variable aléatoire de loi exponentielle de paramètre $a > 0$. X admet la densité

$$f_X(x) = ae^{-ax} 1_{\mathbf{R}^+} = \begin{cases} ae^{-ax} \text{ si } x \geq 0, \\ 0 \text{ sinon.} \end{cases}.$$

Calculons $V(X)$:

$$V(X) = E(X^2) - E(X)^2 = \int_{\mathbf{R}} x^2 f_X(x) dx - \frac{1}{a^2} = \int_0^{+\infty} x^2 a e^{-ax} dx - \frac{1}{a^2}.$$

On pose $u = x^2$, $u' = 2x$ et $v' = ae^{-ax}$, $v = -e^{-ax}$:

$$\begin{aligned} V(X) &= \left[x^2(-e^{-ax})\right]_0^{+\infty} - \int_0^{+\infty} 2x(-e^{-ax})dx - \frac{1}{a^2} \\ &= 0 + 2\int_0^{+\infty} xe^{-ax}dx - \frac{1}{a^2} \\ &= 2\left[x(-\frac{1}{a}e^{-ax})\right]_0^{+\infty} - 2\int_0^{+\infty} (-\frac{1}{a}e^{-ax})dx - \frac{1}{a^2} \\ &= 0 + \frac{2}{a}\int_0^{+\infty} e^{-ax}dx - \frac{1}{a^2} \\ &= \frac{2}{a}\left[-\frac{1}{a}e^{-ax}\right]_0^{+\infty} - \frac{1}{a^2} \\ &= \frac{2}{a}\frac{1}{a} - \frac{1}{a^2} = \frac{1}{a^2}. \end{aligned}$$

$$\boxed{V(X) = \frac{1}{a^2}}$$

3.5 Variables aléatoires indépendantes

C'est sans doute la notion la plus utilisée en probabilités et statistiques.

Définition. *Deux variables aléatoires X et Y sont dites* **indépendantes** *lorsque tout événement "$X \in I$" avec $I \subset \mathbf{R}$ est indépendant de tout événement "$Y \in J$" avec $J \subset \mathbf{R}$. Autrement dit , $\forall I, J \subset \mathbf{R}$,*

$$P(X \in I, Y \in J) = P(X \in I)P(Y \in J).$$

On note parfois $X \perp\!\!\!\perp Y$. ("X indépendant de Y").

En pratique, c'est souvent l'intuition qui permet de décider si deux variables sont indépendantes. On peut alors utiliser la définition comme une propriété : tout événement relatif à une des deux variables sera indépendant de tout événement relatif à l'autre.

exemple : On lance deux dés. Soit X le résultat du premier et Y le résultat du second. Alors X et Y sont indépendantes.

On généralise facilement cette définition à un nombre quelconque, voire une infinité, de variables aléatoires :

Définition. *Des variables aléatoires $X_1, X_2, X_3, \ldots$ sont* **indépendantes** *(ou mutuellement indépendantes) si pour tous sous-ensembles $I_1, I_2, I_3, \ldots$ de $\mathbf{R}$, les événements "$X_i \in I_i$" sont indépendants. C'est-à-dire : pour tous $I_1, I_2, I_3, \ldots \subset \mathbf{R}$,*

$$P(X_1 \in I_1, X_2 \in I_2, \ldots, X_n \in I_n) = P(X_1 \in I_1)P(X_2 \in I_2) \cdots P(X_n \in I_n).$$

3.6 Exemples de calculs de lois utilisant l'indépendance

3.6.1 Somme de variables indépendantes

exemple 1 : *On lance deux dés. Calculer la loi de la somme S des résultats obtenus.*

$S \in \{2, 3, 4, \ldots, 12\}$. Si on note X et Y les résultats des deux dés, X et Y sont des variables aléatoires indépendantes, de même loi uniforme sur $\{1, 2, \ldots, 6\}$: $P(X = i) = P(Y = i) = \frac{1}{6}$.

$[S = 2] = [X = 1] \cap [Y = 1]$ donc $P(S = 2) = P(X = 1)P(Y = 1) = \frac{1}{36}$.

$[S = 3] = ([X = 1] \cap [Y = 2]) \cup ([X = 2] \cap [Y = 1])$ donc $P(S = 3) = P(X = 1)P(Y = 2) + P(X = 2)P(Y = 1) = \frac{2}{36}$.

$[S = 4] = ([X = 1] \cap [Y = 3]) \cup ([X = 2] \cap [Y = 2]) \cup ([X = 3] \cap [Y = 1])$ donc $P(S = 4) = P(X = 1)P(Y = 3) + P(X = 2)P(Y = 2) + P(X = 3)P(Y = 1) = \frac{3}{36}$.

Ainsi de suite. On voit donc que pour chaque valeur possible de la somme S, la probabilité correspondante s'obtient en comptant le nombre de façons différentes d'obtenir cette somme avec les deux résultats. On aboutit ainsi à la loi suivante :

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Proposition. *Si X et Y sont deux variables aléatoires indépendantes admettant une espérance, $E(XY) = E(X)E(Y)$.*

Proof Pour des variables discrètes,

$$\begin{aligned}
E(XY) &= \sum_{z \in \mathcal{S}(XY)} zP(XY = z) = \sum_{z \in \mathcal{S}(XY)} \sum_{x \in \mathcal{S}(X)} \sum_{y \in \mathcal{S}(Y)} \mathbf{1}_{\{z=xy\}} xyP(X = x, Y = y) \\
&= \sum_{x \in \mathcal{S}(X)} \sum_{y \in \mathcal{S}(Y)} xyP(X = x)P(Y = y) \underbrace{\sum_{z \in \mathcal{S}(XY)} \mathbf{1}_{\{z=xy\}}}_{=1} \\
&= \left(\sum_{x \in \mathcal{S}(X)} xP(X = x) \right) \left(\sum_{y \in \mathcal{S}(Y)} yP(Y = y). \right)
\end{aligned}$$

Le fait d'échanger des sommations requiert normalement des hypothèses techniques qu'on passe sous silence ici, mais qui sont vérifiées car les deux variables admettent une espérance.

On admet le résultat pour des variables à densité.

□

Proposition. — *Soit X, Y deux variables indépendantes. Alors*

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

— *Soit $X_1, \dots, X_n$ n variables aléatoires indépendantes. Alors*

$$V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n).$$

Proof *Supposons que $E(X) = 0, V(X) = 0$. Alors $V(X) = E(X^2) - (E(X))^2 = E(X^2)$, et de même $V(Y) = E(Y^2)$. On a*

$$\begin{aligned} V(X+Y) &= E((X+Y)^2) = E(X^2 + 2XY + Y^2) = E(X^2) + 2E(XY) + E(Y^2) \\ &= \underbrace{E(X^2)}_{V(X)} + 2\underbrace{E(X)}_{=0}\underbrace{E(Y)}_{=0} + \underbrace{E(Y^2)}_{V(Y)} \end{aligned}$$

Pour n variables aléatoires, on sait que X_1 est indépendante de $X_2 + \cdots + X_n$, donc

$$V(\sum_{i=1}^n X_i) = V(X_1) + V(X_2 + \cdots + X_n).$$

De même, $V(X_2 + \cdots + X_n) = V(X_2) + V(X_3 + \cdots + X_n)$, et donc

$$V(\sum_{i=1}^n X_i) = V(X_1) + V(X_2) + V(X_3 + \cdots + X_n).$$

En raisonnant par récurrence, on en déduit le résultat, c'est-à-dire

$$V(\sum_{i=1}^n X_i) = \sum_{i=1}^n V(X_i).$$

□

Application au calcul de l'espérance et de la variance de la loi binomiale Soit $n \in \mathbb{N}^*, p \in [0,1]$. Soit X une variable de loi $\mathcal{B}(n,p)$. Alors on sait que X est le nombre de succès lors de la répétition de n expériences aléatoires identiques et indépendantes, où chaque expérience a une probabilité de succès égale à p.

On note pour $1 \leqslant i \leqslant n$

$$Y_i = \begin{cases} 1 \text{ si la } i\text{-ème expérience réussit} \\ 0 \text{ sinon.} \end{cases}$$

Let Y_i sont en fait des variables de Bernoulli de paramètre p. On rappelle que $E(Y_i) = p, V(Y_i) = p(1-p)$. X est donc le nombre de Y_i qui sont égaux à 1. On a donc

$$X = \sum_{i=1}^n Y_i.$$

On a par linéarité de l'espérance

$$E(X) = E(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} E(Y_i) = nE(Y_1) = p.$$

Comme les expériences sont indépendantes, les variables Y_i sont indépendantes, et donc

$$V(X) = V(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} V(Y_i) = nV(Y_1) = np(1-p).$$

On n'a même pas eu besoin de l'expression de la loi d'une binomiale. Ces valeurs sont beaucoup plus dures à obtenir par un calcul direct.

exemple 2 : *Soient X une variable de loi de Poisson $\mathcal{P}(\lambda)$ et Y de loi de Poisson $\mathcal{P}(\mu)$. On suppose X et Y indépendantes. Calculer la loi de $Z = X + Y$.*

Les trois variables X, Y, Z sont à valeurs dans $\mathbf{N}$. Calculons $P(Z = n)$ pour tout $n \in \mathbf{N}$: on va décomposer l'événement $[Z = n]$ ainsi :

$$[Z = n] = [Z = n, X = 0] \cup [Z = n, X = 1] \cup [Z = n, X = 2] \cup \cdots .$$

Cette union est disjointe, donc on peut additionner les probabilités :

$$\begin{aligned} P(Z = n) &= P(Z = n, X = 0) + P(Z = n, X = 1) + P(Z = n, X = 2) + \cdots \\ &= \sum_{k \geq 0} P(Z = n, X = k). \end{aligned}$$

Or

$$\begin{aligned} P(Z = n, X = k) &= P(X + Y = n, X = k) \\ &= P(k + Y = n, X = k) \\ &= P(Y = n - k, X = k) \\ &= P(Y = n - k)P(X = k), \end{aligned}$$

grâce à l'indépendance de X et Y.

Or $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ donc $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ si $k \geq 0$ (et $P(X = k) = 0$ sinon), et $P(Y = n - k) = e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}$ si $n - k \geq 0$, soit $k \geq n$ (et $P(Y = n - k) = 0$ sinon). Ainsi on voit que $P(X = k)P(Y = n - k)$ est non nul seulement si $0 \leq k \leq n$, et donc la somme sur k s'arrête en fait à $k = n$:

$$\begin{aligned}
P(Z=n) &= \sum_{k\geq 0} P(Z=n, X=k) = \sum_{k=0}^{n} e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!} \\
&= e^{-\lambda-\mu}\sum_{k=0}^{n}\frac{1}{k!(n-k)!}\lambda^k\mu^{n-k} \\
&= e^{-(\lambda+\mu)}\frac{1}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda^k\mu^{n-k} \\
&= e^{-(\lambda+\mu)}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda^k\mu^{n-k} \\
&= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!}.
\end{aligned}$$

Ceci correspond à la formule de la loi de Poisson de paramètre $\lambda+\mu$. Ainsi la loi de $X+Y$ est la loi de Poisson de paramètre $\lambda+\mu$.

Le principe de calcul qui vient d'être vu dans ces deux exemples se généralise : il s'agit de calculer la loi de la somme de deux variables discrètes indépendantes :

Proposition. *Soient X et Y deux variables discrètes indépendantes, et $Z = X+Y$. Alors la loi de Z se calcule à partir de celles de X et Y via la formule : pour tout $z \in \mathcal{S}(Z)$,*

$$P(Z=z) = \sum_{x\in\mathcal{S}(X)} P(X=x)P(Y=z-x).$$

Dans le cas de variables à densité, il existe une formule similaire, en remplaçant la somme par une intégrale :

Proposition. *Soient X et Y deux variables indépendantes admettant des densités f_X et f_Y, et $Z = X+Y$. Alors Z est une variable à densité et f_Z se calcule à partir de f_X et f_Y via la formule : pour tout $z \in \mathbf{R}$,*

$$f_Z(z) = \int_{\mathbf{R}} f_X(x)f_Y(z-x)dx.$$

On dit que f_Z est le **produit de convolution** de f_X et f_Y. On verra un exemple avec les lois gaussiennes plus tard dans le cours.

3.6.2 Maximum ou minimum de variables indépendantes

Exemple 8. *Soit un serveur qui reçoit des requêtes de deux machines, A et B. On appelle X_A le temps avant la prochaine requête venant de A, et X_B le temps avant la prochaine requête venant de B. Le temps moyen avant une requête venant de A est $t_A > 0$, et pour B c'est $t_B > 0$. Quel est le temps moyen avant une requête (de A ou B) ?*

Comme on n'a aucune information sur X_A et X_B, on les modélise par des exponentielles $X_A \sim \mathcal{E}(\lambda_A), X_B \sim \mathcal{E}(\lambda_B)$. Comme la moyenne de X_A est $1/\lambda_A$, on a $t_A = 1/\lambda_A$. De même $t_B = 1/\lambda_B$.

Soit X le temps avant la prochaine requête. On a donc $X = \min(X_A, X_B)$. Calculons la fonction de répartition. Pour calculer la fonction de répartition d'un minimum de variables aléatoires, il vaut mieux calculer le complémentaire :

Pour $x \geqslant 0$,

$$\begin{aligned}P(X \leqslant x) =&1 - P(X > x)\\ =&1 - P(X_A > x, X_B > x) = 1 - P(X_A > x)P(X_B > x)\\ =&1 - (1 - (1 - \exp(-\lambda_A x))(1 - (1 - \exp(-\lambda_B x))))\\ =&e^{-\lambda_A x}e^{-\lambda_B x} = e^{-(\lambda_A+\lambda_B)x} = 1 - (1 - \exp(-(\lambda_A + \lambda_B)x))\end{aligned}$$

Donc X a la fonction de répartition d'une loi exponentielle de paramètre $\lambda_A+\lambda_B = \frac{1}{t_A}+\frac{1}{t_B}$, c'est donc bien sa loi. Donc sa moyenne, c'est-à-dire le temps moyen avant une requête, est

$$E(X) = \frac{1}{\frac{1}{t_A} + \frac{1}{t_B}}$$

Exemple 9 (TD4, exercice 10)**.** *On lance trois dés. Certains de ces dés tombent sur un. On relance les autres, et on garde ceux qui tombent sur 1. On recommence jusqu'à ce que tous les dés soient tombés sur 1. On appelle T le nombre de lancer nécessaires. Quelle est la distribution de T ? Quelle est son espérance ?*

On appelle T_i le nombre de fois qu'il faut lancer le dé i pour obtenir 1. On a déjà vu que $T_i \sim \mathcal{G}(1/6)$. Alors $T = \max(T_1, T_2, T_3)$. On a pour $k \in \mathbb{N}$

$$\begin{aligned}P(T \leqslant k) =&P(T_1 \leqslant k, T_2 \leqslant k, T_3 \leqslant k) = P(T_1 \leqslant k)P(T_2 \leqslant k)P(T_3 \leqslant k)\\ =&(1 - e^{-k/6})^3\\ P(T = k) =&P(T \leqslant k) - P(T \leqslant k-1) = \begin{cases}(1 - e^{-k/6})^3 - (1 - e^{-(k-1)/6})^3 \text{ si } k > 0\\ 0 \text{ si } k = 0.\end{cases}\end{aligned}$$

Pour calculer le max (ou le min) de variables aléatoires, il est souvent intéressant de passer par la fonction de répartition (ou sa complémentaire).

Chapitre 4

Lois jointes de variables aléatoires

4.1 Couples, triplets, vecteurs aléatoires

Définition. *Un* **couple aléatoire** *est un couple* (X,Y) *où* X *et* Y *sont des variables aléatoires. De même on parle de triplet aléatoire* (X,Y,Z) *si on considère trois variables* X,Y,Z*, et plus généralement de vecteur aléatoire, ou* n*-uplet aléatoire* $(X_1,X_2,\dots,X_n)$*, en considérant* n *variables* $X_1,X_2,\dots,X_n$*.*

Définition. *Le* **support d'un couple** *aléatoire* (X,Y) *est l'ensemble des valeurs prises par* (X,Y)*, c'est-à-dire l'ensemble des couples de valeurs prises par* X *et* Y*. On le note* $\mathcal{S}(X,Y)$*, et il est donc égal à* $\mathcal{S}(X)\times\mathcal{S}(Y)$*.*

exemples :

- Soit $X\sim\mathcal{G}(p), Y\sim\mathcal{E}(\lambda)$. Alors $\mathcal{S}((X,Y))=\{(n,x):n\in\mathbb{N},x\in\mathbb{R}_+\}=\mathbb{N}\times\mathbb{R}_+$.
- Si X et Y correspondent à des lancers de dés, on a $\mathcal{S}(X)=\mathcal{S}(Y)=\{1,2,3,4,5,6\}$ et $\mathcal{S}(X,Y)=\{(x,y),x,y\in\{1,2,3,4,5,6\}\}=\{1,2,3,4,5,6\}\times\{1,2,3,4,5,6\}$, ce qui se note aussi $\{1,2,3,4,5,6\}^2$. Le cardinal de $\mathcal{S}(X,Y)$ (nombre de valeurs prises par (X,Y)) est donc $6^2=36$.
- Avec 10 lancers de dé, on obtient 10 variables $X_1,\dots,X_{10}$, et donc un vecteur aléatoire $(X_1,\dots,X_{10})$. Le support de ce vecteur est l'ensemble $\mathcal{S}((X_1,\dots,X_{10}))=\{1,2,3,4,5,6\}^{10}$, et donc $Card(\mathcal{S}((X_1,\dots,X_{10})))=6^{10}$, soit environ 60 millions de valeurs possibles.

4.2 Loi jointe et lois marginales d'un couple de variables discrètes

Soit (X,Y) un couple aléatoire. Alors

$$
\begin{aligned}
(X,Y) \text{ est une variable discrète } \Leftrightarrow & \mathcal{S}(X,Y) \text{ dénombrable } \Leftrightarrow \mathcal{S}(X)\times\mathcal{S}(Y) \text{ dénombrable}\\
\Leftrightarrow & \mathcal{S}(X) \text{ dénombrables et } \mathcal{S}(Y) \text{ dénombrable}\\
\Leftrightarrow & X \text{ et } Y \text{ sont des variables discrètes}
\end{aligned}
$$

Définition. *La* **loi d'un couple** (X,Y) *de variables discrètes est l'ensemble des probabilités* $P((X,Y)=(x,y)=P(X=x \text{ et } Y=y)$ *(souvent noté* $P(X=x,Y=y)$*) pour tous les couples* $(x,y)\in\mathcal{S}(X,Y)$*, c'est-à-dire pour tous les* $x\in\mathcal{S}(X)$ *et* $y\in\mathcal{S}(Y)$*.*

On peut présenter cette loi sous forme d'un tableau.

Exemple 10. *On lance deux dés et une pièce. On note* X,Y *et* Z *les variables suivantes :*

- X *est le résultat du premier dé.*
- Y *est le résultat du second dé.*
- $Z=X$ *si la pièce tombe sur pile, et* $Z=Y$ *autrement.*

On voit que X *et* Y *sont indépendentes. Représentons la loi de* (X,Y) *sous forme d'un tableau. On a par exemple* $P(X=1,Y=1)=P(X=1)P(Y=1)=1/36$ *par indépendance, et de même pour tous les couples de valeurs possibles.*

	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

Représentons la loi de (X,Z) *sous forme d'un tableau. On a avec la formule des probabilités totale*

$$
\begin{aligned}
P(X=1,Z=1)= & P(X=1,Z=1|\text{Pile})P(\text{Pile})+P(X=1,Z=1|\text{Face})P(\text{Face})\\
= & P(X=1|\text{Pile})P(\text{Pile})+P(X=1)P(Z=1|\text{Face})\frac{1}{2}=\frac{1}{12}+\frac{1}{6\cdot 12}=\frac{1}{12}+\frac{1}{72}\\
= & \frac{7}{72}
\end{aligned}
$$

Un calcul similaire nous donne les $P(X=k,Z=k)=7/72$ *pour* $k\in\{1,2,3,4,5,6\}$*. On a par contre*

$$
\begin{aligned}
P(X=1,Z=2)= & P(X=1,Z=2|\text{Pile})P(\text{Pile})+P(X=1,Z=2|\text{Face})\text{Face}\\
= & 0+\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{2}=\frac{1}{72}.
\end{aligned}
$$

On a le même résultat pour toutes les valeurs de $P(X = k, Y = l) = 1/72$ *pour* $k \neq l$. *On en déduit la loi du couple* (X, Z).

	1	2	3	4	5	6
1	7/72	1/72	1/72	1/72	1/72	1/72
2	1/72	7/72	1/72	1/72	1/72	1/72
3	1/72	1/72	7/72	1/72	1/72	1/72
4	1/72	1/72	1/72	7/72	1/72	1/72
5	1/72	1/72	1/72	1/72	7/72	1/72
6	1/72	1/72	1/72	1/72	1/72	7/72

On remarque que (Y, Z) *a la même loi car on peut l'obtenir par les mêmes calculs (alors que* $(X, Y) \neq (Y, Z)$*...).*

On peut en déduire la loi de Z, *en faisant la somme sur toute une colonne*

$$P(Z = 1) = \sum_{k=1}^{6} P(Z = 1, X = k) = \frac{12}{72} = \frac{1}{6},$$

et de même pour tous les $P(Z = k), k \in \{1, 2, \ldots, 6\}$. *Donc* Z *a la même loi que* X *et* Y.

On voit que X, Z *ne sont pas indépendantes car par exemple* $P(X = 1, Z = 2) = 7/72$ *alors que* $P(X = 1)P(Z = 2) = \frac{1}{36}$.

On appelle les lois de X et de Z les **lois marginales du couple** (X, Z) (puisqu'on les lit à la "marge" du tableau, **en faisant la somme sur une ligne ou une colonne**). Souvent on les appelle aussi lois marginales de X et de Z, ce qui n'est pas logique (mais usuel).

Remarquons que (X, Y) et (X, Z) ont les même lois marginales, alors que les couples n'ont pas la même loi. Il n'est en général pas possible de retrouver la loi du couple (X, Z) à partir uniquement des marginales, c'est-à-dire des lois de X et Z.

Une autre manière de dire ça est que **la loi du couple ne peut pas se retrouver à partir de couple des lois (marginales).** Sauf dans le cas où on sait que X et Z sont indépendantes, auquel cas la loi du couple est donnée par $P(X = x, Z = z) = P(X = x)P(Z = z)$ (voir le premier tableau).

4.2.1 Formule de calcul d'une espérance

Lorsque l'on connait la loi d'un couple de variables il est possible de calculer l'espérance d'une fonction réelle f de ces deux variables. La formule pour des variables discrètes est :

$$E(f(X, Y)) = \sum_{x \in \mathcal{S}(X)} \sum_{y \in S(Y)} f(x, y)P(X = x, Y = y).$$

4.3 Loi d'un vecteur de variables discrètes

La loi d'un vecteur $(X_1, X_2, \dots, X_n)$ de variables discrètes est donnée par les probabilités $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ pour tous les $x_1 \in \mathcal{S}(X_1)$, $x_2 \in \mathcal{S}(X_2)$,... $x_n \in \mathcal{S}(X_n)$. Ces valeurs pourraient se placer dans un tableau à n dimensions si on pouvait le dessiner !, de même type que celui obtenu pour un couple (X, Y). Il existe également une notion de densité pour les vecteurs aléatoires, mais plus compliquée à définir que pour une seule variable à la fois.

4.4 Loi conditionnelle d'une variable

4.4.1 Conditionnement par rapport à un événement

Définition. *Soit X une variable et A un événement tel que $P(A) \neq 0$. La* **loi conditionnelle de X sachant** *A est la loi de X dans la situation où A est réalisé. Si X est une variable discrète, la loi de X sachant A est donnée par les probabilités $P(X = x|A)$ pour tous les $x \in \mathcal{S}(X)$.*

Définition. *L'***espérance conditionnelle de X sachant** *A est l'espérance de X dans la situation où A est réalisé. On la note $E(X|A)$ Dans le cas où X est une variable discrète, on a*

$$E(X|A) = \sum_{x \in \mathcal{S}(X)} x P(X = x|A).$$

On peut également retrouver l'espérance de X en appliquant ces résultats à une partition (A_i) de l'univers :

$$E(X) = \sum_i E(X|A_i) P(A_i).$$

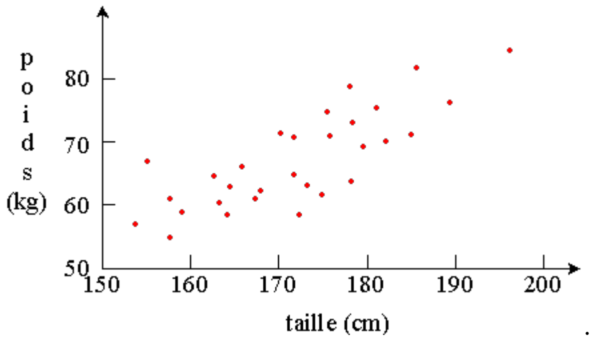
Exemple 11. *TD5, exo 1 :*

Dans la mémoire d'un ordinateur il se peut que certains "bit" enregistrés soient inexacts. Ce phénomène étant rare on suppose que le nombre de "bit" erronés suit une loi de Poisson. Cependant il est plus fréquent de voir apparaître un 0 à la place d'un 1 que le contraire. Soit X_1 le nombre de faux 0 et X_2 le nombre de faux 1. On suppose ces variables indépendantes et de loi de Poisson de paramètres respectifs λ_1 et λ_2, avec $\lambda_1 > \lambda_2$.

1. *Quelle est la loi du nombre total d'erreurs ? (Réponse : une loi de Poisson de paramètre $\lambda_1 + \lambda_2$, voir cours)*
2. *Soit $n \geqslant 1$. On suppose que n erreurs ont été commises. Soit X le nombre de faux 0. Quelle est la loi de X ?*

On appelle A l'évènement "$X_1 + X_2 = n$". Alors on montre que

$$P(X_1 = k|A) = C_n^k p^k (1-p)^{n-k}$$



avec $p = \lambda_1/(\lambda_1+\lambda_2)$. *On remarque que comme* $0 < \lambda_1 < \lambda_1+\lambda_2$, *on a*

Comme $X_1 \sim \mathcal{B}(n,p)$ *conditionnellement à* A,

$$E(X|A) = np = \frac{n\lambda_1}{\lambda_1+\lambda_2}.$$

hors-programme

On retrouve l'espérance de X *avec la partition* $(A_n; n\in\mathbb{N})$ *où* $A_n = (X_1+X_2=n)$ *de probabilité* $P(A_n) = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}$ *:*

$$\begin{aligned}E(X) =& \sum_{n=0}^{\infty} E(X|A_n)P(A_n) = \sum_{n=0}^{\infty} n\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}\\ =&\lambda_1 e^{-(\lambda_1+\lambda_2)}\sum_{n=1}^{\infty}\frac{1}{(n-1)!}(\lambda_1+\lambda_2)^{n-1} = \lambda_1 e^{-(\lambda_1+\lambda_2)}\sum_{n=0}^{\infty}\frac{(\lambda_1+\lambda_2)^n}{n!} = \lambda_1 e^{-(\lambda_1+\lambda_2)}e^{(\lambda_1+\lambda_2)}\\ =&\lambda_1.\end{aligned}$$

4.5 Covariance et corrélation

Exemple 12. *On considère la courbe suivante, qui représente la taille et le poids d'individus pris au hasard dans la population : On remarque qu'il y a une tendance au poids à être élevé quand la taille est élevée, et vice-versa (ce qui n'est pas une surprise). Comment traduire ça avec des variables aléatoires ? On tire au hasard un individu dans la population, et on appelle* X *son poids, et* Y *sa taille. Comment traduire sur la loi de* (X,Y) *cette relation ?*

Définition. *Soient* X *et* Y *deux variables aléatoires dont les variances sont finies. La* **covariance** *de* X *et* Y *est*

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = E((X-E(X))(Y-E(Y))).$$

Définition. *On dit que* X *et* Y *sont positivement corrélées si* $\mathrm{Cov}(X,Y) > 0$.

Dans le langage courant, on dit souvent corrélées au lieu de positivement corrélées. On dit que X et Y sont **négativement corrélées** si $\mathrm{Cov}(X, Y) < 0$. Cela signifie au contraire que X a tendance a être petite si Y est grande, et vice-versa.

La covariance est une mesure incomplète de l'indépendance de deux variables. En effet on a la résultat suivant :

Proposition. *Soient X et Y deux variables indépendantes. Alors* $\mathrm{Cov}(X, Y) = 0$.

Preuve. On a vu que comme X, Y sont indépendantes, $E(XY) = E(X)E(Y)$. Donc $Cov(X, Y) = E(XY) - E(X)E(Y) = 0$. □

Attention : La réciproque de la proposition est fausse en général : $cov(X, Y) = 0$ n'implique pas que X et Y sont indépendantes.

Exemple : On tire une personne au hasard dans la population française (si tant est que c'est possible...).

— Si on appelle P son poids et T sa taille, on a $\mathrm{Cov}(P, T) > 0$, ce qui veut dire qu'une personne de grande taille a des chances d'avoir un poids au dessus de la moyenne.
— Si on appelle X son salaire annuel et Y son nombre d'années d'études, on espère que $Cov(X, Y) > 0$...
— On pose $X = 1$ si cette personne est une femme, et 0 sinon. Soit Y la taille de cette personne. On suppose que la taille moyenne des franfais est de 1m65. On observe que la taille moyenne des femmes est $1m60$. Qu'en déduisez vous sur la covariance ?

$$\begin{aligned}
Cov(X, Y) =&E(XY) - E(X)E(Y) = \sum_{y \in \mathcal{S}(Y)} yP(Y = y|X)P(X) - \frac{1}{2} \cdot 1,65 \\
=& \underbrace{\left(\sum_{y \in \mathcal{S}(Y)} yP(Y = y|X = 1) \right)}_{\text{taille moyenne conditionnellement à } X = 1} \times \frac{1}{2} - 1,65 \cdot \frac{1}{2} \\
=&1,60 \cdot \frac{1}{2} - 1,65 \cdot \frac{1}{2} < 0
\end{aligned}$$

Le fait que deux variables aléatoires sont corrélées peut impliquer qu'elles soient influencées par un même phénomène.

Exemple : On tire au hasard une personne dans le monde. On introduit les variables de Bernoulli suivantes :

$$X = \begin{cases} 1 \text{ si cette personne parle français} \\ 0 \text{ sinon} \end{cases}$$

$$Y = \begin{cases} 1 \text{ si cette personne mange du camembert} \\ 0 \text{ sinon} \end{cases}$$

On observe que $E[X] = 1/10000$ et $E[Y] = 1/5000$. On observe que $E[XY] = 1/30000$. Peut-on en déduire que manger du camembert aide à mieux parler français ?

En général, il est rare d'avoir exactement 0 pour la covariance de deux variables, même si elles n'ont aucun rapport. Pour dire que X et Y sont corrélées, il faut avoir une mesure de la corrélation. Malheureusement, on voit par exemple que si X est la taille en m, X' est la taille en cm, et Y le poids en kg,

$$\begin{aligned} \mathrm{C}ov(X', Y) &= E(X'Y) - E(X')E(Y) = E(100 \cdot X \cdot Y) - E(100 \cdot X)E(Y) \\ &= 100E(XY) - 100E(X)E(Y) = 100\mathrm{C}ov(X, Y). \end{aligned}$$

Ce qui voudrait dire que la taille en cm est 100 fois plus corrélée que la taille en mètres. C'est absurde, car l'unité de mesure est une convention et n'influence pas le lien entre poids et taille. Il faut en fait renormaliser par la déviation standard. On définit la corrélation entre X et Y par

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{C}ov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}.$$

Exercice : avec poids et taille,

$$\mathrm{Corr}(X', Y) = \mathrm{Corr}(X, Y).$$

Cette quantité n'est pas modifiée par un changement d'unité de mesure.

Proposition. *Pour toutes variables aléatoires X et Y, la corrélation est un nombre de $[-1, 1]$. Autrement dit,*

$$|Cov(X, Y)| \leq \sigma(X)\sigma(Y).$$

Ce résultat est en fait une conséquence immédiate de l'inégalité de Cauchy-Schwartz :

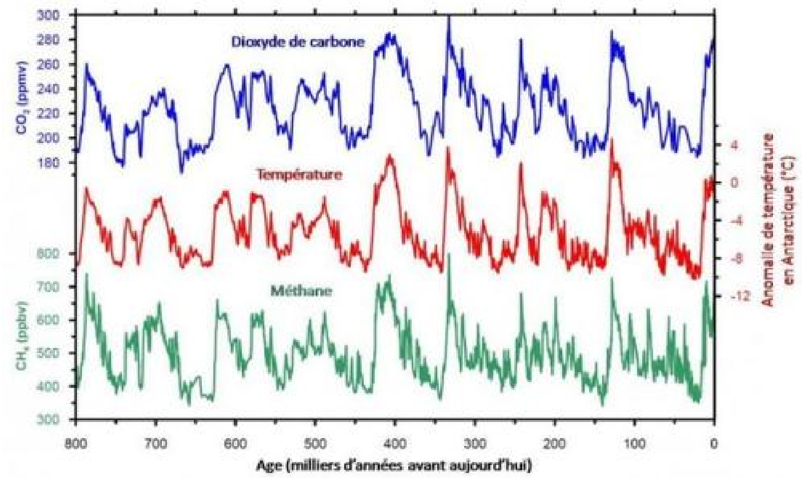
Inégalité de Cauchy-Schwarz Soient X, Y deux variables aléatoires dont les variances sont bien définies. Alors

$$E(XY) \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

Preuve. Soit $t \in \mathbf{R}$. On pose $f(t) = E((X + tY)^2)$. On a

$$\begin{aligned} f(t) &= E(X^2 + t^2Y^2 + 2tXY) = E(X^2) + t^2E(Y^2) + 2tE(XY), \\ &= E(Y^2)t^2 + 2E(XY)t + E(X^2). \end{aligned}$$

Ainsi $f(t) = at^2 + bt + c$, avec $a = E(Y^2)$, $b = 2E(XY)$, et $c = E(X^2)$. Or $f(t)$ est toujours positif, puisque $f(t) = E((X + tY)^2)$ (espérance d'une variable positive). Donc le polynàťme $at^2 + bt + c$ ne peut pas avoir deux racines distinctes réelles (sinon son signe change lorsque t parcourt $\mathbf{R}$), et donc son discriminant est négatif ou nul : $\Delta = b^2 - 4ac \leq 0$. En remplaàğant a, b, c par leurs valeurs, on obtient $\Delta = 4E(XY)^2 - 4E(Y^2)E(X^2) \leq 0$, et donc $E(XY)^2 \leq E(Y^2)E(X^2)$, soit encore $E(XY) \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$. $\square$



Pour prouver la proposition il suffit d'appliquer cette inégalité à $X-E(X)$ et $Y-E(Y)$.

Par exemple, on peut dire que dans la population française, la corrélation entre poids et taille est d'environ $0,7$, ce qui indique une forte corrélation. Par contre, la corrélation entre la longueur du nez et la taille est de $0,1$, ce qui est probablement fortuit. On dit en statistiques que cette corrélation n'est pas significative.

Voyons un autre exemple.

Exemple

On appelle X la température de la terre a une année prise au hasard, et Y la concentration en CO_2 (on pourrait aussi prendre le méthane). On observe que X et Y sont très corrélées, $\mathrm{Corr}(X,Y) \sim 0,8$.

Interprétation des résultats

Le fait que 2 variables aléatoires sont corrélées peut indiquer plusieurs choses :

— La variable X influence la variable Y. Exemple : Le nombre d'années d'études d'un individu est corrélé à son salaire.

— Les variables X et Y sont influencées par une troisième, "cachée". Exemple : le poids de fromage ingéré par un individu est proportionnel au nombres de mots de français qu'ils parlent. C'est parce qu'ils sont influencés par la variable de Bernoulli qui vaut 1 si l'individu est français et 0 sinon.

— Il peut n'y avoir aucune cause, ou on ne la comprend pas, ou notre estimation est fausse car on n'a pas récolté suffisamment de données.