
UNIVERSITÉ PARIS DESCARTES

Licence Informatique 2e année

Notes du cours

Probabilités et Statistiques pour l'Informatique.

2ème année de Licence Informatique, Université Paris Descartes
2018-2019

Raphaël Lachieze-Rey

Table des matières

1	Probabilités, événements	1
1.1	Statistiques et informatique	1
1.2	Rappels sur les notations mathématiques	1
1.2.1	Ensembles	1
1.2.2	Sous-ensembles	3
1.2.3	Couples, triplets, n -uplets et suites	3
1.3	Événements et probabilité	4
1.3.1	Exemple introductif	5
1.3.2	Univers, événements, et réalisations	6
1.3.3	Intersections d'événements	7
1.3.4	Unions d'événements et probabilité d'une union	8
1.3.5	Partition	10
1.3.6	Événement complémentaire	12
1.3.7	Exclusion d'événements	13
2	Probabilités conditionnelles et indépendance d'évènements	15
2.1	Probabilités conditionnelles	15
2.2	Formule des probabilités totales	19
2.3	Formule de Bayes	21
2.4	Indépendance	22
2.4.1	Indépendance de deux événements	22
2.4.2	Indépendance et événements complémentaires	24
2.4.3	Indépendance de 3 événements ou plus	24
2.5	Indépendance conditionnelle	25
3	Variables aléatoires et lois	27
3.1	Définitions générales	27
3.1.1	Variables aléatoires	27
3.1.2	Support d'une variable aléatoire	28
3.1.3	Fonction de répartition d'une variable aléatoire	29
3.1.4	Variables discrètes	30
3.1.5	Les événements élémentaires $[X = x]$	36
3.2	Variables aléatoires continues, ou à densité	36

3.2.1	variable uniforme	36
3.2.2	Définition	37
3.2.3	Variable gaussienne	39
3.2.4	Loi exponentielle	41
3.2.5	Fonction de répartition d'une variable aléatoire à densité	42
3.2.6	Autre approche de la densité	43
3.3	Espérance et variance d'une variable aléatoire	46
3.3.1	Espérance	46
3.3.2	Variance	48
3.3.3	Exemples de calculs	50
3.3.4	Moments	53
4	Vecteurs aléatoires	55
4.1	Variables aléatoires indépendantes	55
4.1.1	Quelques propriétés	56
4.2	Exemples de calculs de lois utilisant l'indépendance	58
4.2.1	Maximum ou minimum de variables indépendantes	59
4.3	Covariance et corrélation	60
4.3.1	Conditionnement par rapport à un événement	66
5	Théorèmes limites et estimation	69
5.1	Jeu de données et échantillon	69
5.2	Modélisation	70
5.2.1	La loi des grands nombres	71
5.3	Estimation ponctuelle de l'espérance	73
5.4	Intervalles de confiance et Théorème Central Limite	74
5.4.1	Précision et intervalle de confiance	74
5.4.2	Renormalisation	76
5.4.3	Théorème central limite	76
5.4.4	Intervalle de confiance grâce au théorème central limite	79
5.4.5	Estimation ponctuelle de la variance	81

Chapitre 1

Probabilités, événements

1.1 Statistiques et informatique

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 informatique :

- 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
- 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 $0, 1, 2, \dots$. 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 apparaît au plus une fois dans un ensemble : $\{1, 3, 5\} = \{1, 3, 5, 3\} = \{1, 1, 1, 3, 5\} = \dots$

• $\mathbf{N} = \{0, 1, 2, 3, \dots\}$: \mathbf{N} est l'ensemble des nombres 0, 1, 2, etc.

• $B = \{0, 1, 4, 9, \dots\}$: 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 \emptyset . On a par exemple $\{x \in \mathbf{R} : x \leq 2, x \geq 3\} = \emptyset$.

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

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

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 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, \emptyset 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 :

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

Associativité :

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

Distributivité :

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

Passage au complémentaire :

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

1.2.3 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.
- $(1, 3, 5, \dots, 19, 21)$ 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 $\text{Card}(\{1, 3, 5\}^2) = 9 = 3^2 = (\text{Card}\{1, 3, 5\})^2$.

- $\mathbf{N} \times \mathbf{N}$ 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 $(1, 3, 5, \dots, 19, 21) \in \mathbf{N}^{10}$.
- $\mathbf{R} \times \mathbf{N} = \{(x, n), x \in \mathbf{R}, n \in \mathbf{N}\}$. Par exemple, $(2.45, 3) \in \mathbf{R} \times \mathbf{N}$.
- $\mathbf{R}^n = \{(x_1, x_2, \dots, x_n), x_1 \in \mathbf{R}, \dots, x_n \in \mathbf{R}\}$: 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 $(1, 1/2, 1/3, 1/4, \dots)$ est la suite $(u_n)_{n \in \mathbf{N}}$ avec $u_n = 1/n$.

Il est indispensable de maîtriser ces notions pour la suite du cours. Pour le vérifier, vous devez savoir faire par vous-même sans problème les exercices 1 et 2 du TD1, et vous entraîner sur le Quiz 1 sur Moodle.

1.3 Événements et probabilité

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 de l'expérience (qui est souvent vous-même)** ne connaît pas son résultat. La notion **d'aléatoire** est plus liée à la notion **d'incertitude** qu'à celle de **hasard**.

Lors du lancer d'un dé à 6 faces, si l'on s'intéresse à la probabilité de l'évènement A = "le dé tombe sur 6", on a que A est un évènement aléatoire et que $\mathbb{P}(A) = 1/6$ dans toutes les situations suivantes :

- 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 exact d'être humains vivants sur la planète à minuit était pair (aléatoire pour tout le monde, bien que déterminé).
- 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 résultat d'un nombre aléatoire tiré par un ordinateur. Il faut savoir que les ordinateurs ne peuvent pas générer spontanément de nombres aléatoires. Lorsqu'on lui demande un nombre aléatoire, il effectue en réalité un grand nombre d'opérations très compliquées sur un nombre fixé à l'avance ("seed" ou "graine"), ce qui donne un résultat aléatoire pour l'utilisateur, alors que c'est le résultats d'opérations déterministes. Il prend souvent l'heure comme graine. S'il doit tirer un autre nombre aléatoire dans le même programme, pour éviter de donner le même résultat, il utilisera comme graine le dernier résultat obtenu.

Le degré d'incertitude lié à un évènement 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.

1.3.1 Exemple introductif

Expérience aléatoire 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.

Question : Quelle est la probabilité qu'elles soient toutes rouges ?

On pose l'évènement $A = \text{"les trois billes piochées sont rouges"}$. Analysons le problème pour le simplifier.

- Une *issue* est une manière de choisir trois billes parmi les 6 (sans voir les couleurs). Autrement dit, c'est le nombre d'ensembles à trois éléments que l'on trouve dans un ensemble à 6 éléments, indépendamment du fait que ce soit des billes, des points, des carrés, ...
- Il n'y a en fait qu'une issue qui nous intéresse, celle où cet ensemble est constitué des trois billes rouges.

On a donc

$$\begin{aligned} \mathbb{P}(A) &= \frac{\text{nombre d'issues favorables}}{\text{nombre total d'issues}} \\ &= \frac{1}{\#\{\text{sous-ensembles à 3 éléments parmi un ensemble de 6 éléments}\}} \end{aligned}$$

Le nombre d'ensembles à 3 éléments parmi un ensemble à 6 éléments est

$$C_6^3 = \frac{6!}{3!3!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 3 \times 2 \times 1} = 6 \times 4 \times 2 = 24.$$

Donc

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

Attention : cette formule ne marche que parce que tous les événements ont la même probabilité.

Autre méthode : On considère qu'on pioche les billes les unes à la suite des autres, donc **sans remise**. Dans ce cas là l'ordre est important. L'ensemble de toutes les issues possibles est l'ensemble de manière de choisir 1 boules parmi 8, puis 1 boule parmi les 7 restantes, puis 1 boule parmi les 6 restantes. Il y a donc $8 \times 7 \times 6 = 336$ possibilités. Le nombre d'issues qui nous intéresse est toutes les issues où les trois billes qu'on a piochées sont rouges. Imaginons que les billes sont numérotées de 1 à 8, et que les rouges sont les billes 1,2,3 (le raisonnement est le même si on choisit d'autres numéros). Cette fois les issues favorables sont :

issues favorables : $(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)$.

On a donc

$$\mathbb{P}(A) = \frac{\#\{\text{issues favorables}\}}{\#\{\text{issues possibles}\}} = \frac{6}{336} = \frac{1}{56}.$$

1.3.2 Univers, événements, et réalisations

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

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

Il est inutile de distinguer des "réalités" qui donnent les mêmes issues à l'expérience. Dans cet exemple,

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

(il y a 36 éventualités possibles, donc 36 éléments dans Ω . On les appelle $\omega_1, \omega_2, \dots, \omega_{36}$). Une "réalité" s'appelle une **réalisation** (parfois appelé **événement élémentaire**), et Ω l'ensemble des réalisations possibles s'appelle **l'univers**. On note souvent une réalisation par la lettre grecque minuscule ω .

On peut choisir plusieurs univers possibles pour la même expérience. Dans l'exemple des trois billes rouges, on pouvait prendre

$$\Omega = \{\text{ensembles de 3 billes}\}$$

ou

$$\Omega = \{\text{triplets (=ensembles ordonnés) de 3 billes}\}$$

Chacun des deux choix nous donne une méthode qui marche, mais le premier choix est beaucoup plus simple. D'une manière générale, il faut choisir Ω suffisamment grand pour pouvoir traiter le problème, mais pas trop grand pour ne pas trop augmenter la complexité. Par contre pour une question du type : "Quelle est la probabilité que la première bille soit verte, la deuxième rouge et la troisième verte", il faut se placer dans le second cas.

Les **événements** 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'élément $\omega_1 = \text{"Le dé rouge tombe sur 1 et le vert aussi"}$ mais pas l'élément $\omega_2 = \text{"le dé rouge tombe sur 1 et le vert sur 2"}$. On peut également considérer l'évènement Ω lui-même, qui contient toutes les réalisations, où l'évènement vide \emptyset , qui ne contient aucune réalisation.

Remarquons que l'univers peut être infini. Prenons par exemple l'expérience suivante : **Expérience aléatoire** : On jette un dé à 6 faces plusieurs fois, 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 $n \geq 1$. L'univers Ω est dans ce cas

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

Même si il n'y aura en pratique qu'un nombre fini de lancers, le nombre d'éventualités possibles est infini. Remarquons qu'on aurait pu choisir un univers beaucoup plus complexe.

Autre exemple :

Expérience aléatoire : j'attends le bus, et je m'intéresse au temps aléatoire qu'il va mettre à arriver, sachant que ça ne peut pas être plus de 10 minutes.

Cette fois l'univers est $\Omega = [0, 10]$, c'est à nouveau un ensemble infini.

1.3.3 Intersections d'événements

Imaginons une expérience où on lance deux dès, un rouge et un vert. L'évènement $A = \text{"le dé rouge tombe sur un nombre pair plus grand que 3"}$ se décompose naturellement de la manière suivante :

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

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

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

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 $\omega \in B \text{ ET } C$.



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

On visualise les événements comme des ensembles, et l'événement que deux événements se réalisent ensemble correspond à l'intersection d'ensembles :

Plutôt que $A = B \text{ et } C$, on écrira en fait :

$$A = B \cap C.$$

$$A = B \text{ ET } C.$$

Pour se familiariser avec ces notions, on utilise en parallèle les notations \cap et **ET** dans ce chapitre, mais on n'utilisera plus que \cap par la suite.

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 .

1.3.4 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 :

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

Par conséquent A est composé de deux événements élémentaires, que l'on peut noter D_4, D_6 , avec $D_i = \text{"le résultat du dé est } i\text{"}$. 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 $\omega \in D_4 \text{ OU } D_6$. On écrira donc en fait :

$$A = D_4 \cup D_6.$$

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

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

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

$$\mathbb{P}(E \text{ OU } F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \text{ ET } F).$$

Remarque On utilise la notation ensembliste car elle répond aux mêmes lois que celles du calcul logique ou du calcul booléen : Exemples

$$\begin{aligned} (A \text{ OU } B) \text{ ET } C &= (A \text{ ET } C) \text{ OU } (B \text{ ET } C) & (A \text{ ET } B) \text{ OU } C &= (A \text{ OU } B) \text{ ET } (A \text{ OU } C) & \text{etc...} \\ (A \cup B) \cap C &= (A \cap C) \cup (B \cap C) & (A \cap B) \cup C &= (A \cup B) \cap (A \cup C) \end{aligned}$$

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 $E \cap F = \emptyset$) alors

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F).$$

$$\mathbb{P}(E \text{ OU } F) = \mathbb{P}(E) + \mathbb{P}(F).$$

C'est logique puisque $\mathbb{P}(E \cap F) = 0$...

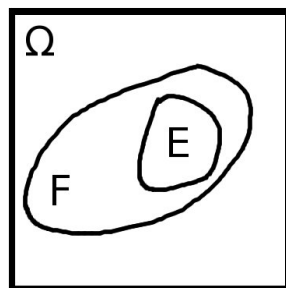
- Soient E_1, \dots, E_n plusieurs événements. S'ils sont **deux à deux disjoints** ($E_i \text{ ET } E_j = \emptyset$ pour tous $i \neq j$) alors

$$\mathbb{P}(E_1 \cup E_2 \cup \dots \cup E_n) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \dots + \mathbb{P}(E_n).$$

$$\mathbb{P}(E_1 \text{ OU } E_2 \text{ OU } \dots \text{ OU } E_n) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \dots + \mathbb{P}(E_n).$$

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

$$\mathbb{P}(A) = \mathbb{P}(D_4) + \mathbb{P}(D_6).$$

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

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

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 peut être représenté par un carré de côté 1).

On dit qu'un événement A est **inclus dans** un événement B , noté $A \subset B$, si chaque fois que A est réalisé, B est également réalisé. Exemple

- “Le dé tombe sur 2” \subset “le dé est un nombre pair”, mais pas l'inverse
- “Il pleuvra demain” \subset “il va pleuvoir pendant les 15 prochains jours”
- “Je possède plus de 1000 euros” \subset “Je possède plus de 100 euros”

Voici une autre relation importante découlant des formules précédentes :

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

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

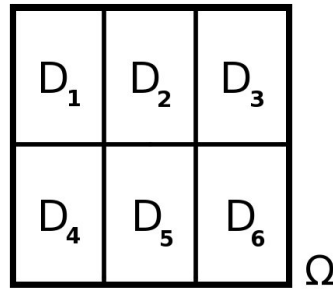
1.3.5 Partition

Des événements E_1, \dots, E_n forment une **partition** s'ils sont deux à deux incompatibles et qu'il y a toujours l'un d'entre eux qui est vrai. Autrement dit : un et un seul de ces événements se réalise. Alors on a la relation :

$$\mathbb{P}(E_1) + \dots + \mathbb{P}(E_n) = 1.$$

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 événements 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 $\mathbb{P}(\Omega) = 1$:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(E_1 \cup E_2 \cup \dots \cup E_n) = \mathbb{P}(E_1) + \dots + \mathbb{P}(E_n).$$

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(E_1 \text{ OU } E_2 \text{ OU } \dots \text{ OU } E_n) = \mathbb{P}(E_1) + \dots + \mathbb{P}(E_n).$$

Dans l'exemple du lancer de dé, les événements D_1, \dots, D_6 forment une partition de Ω . Puisque chacun de ces événements a pour probabilité $\frac{1}{6}$, on vérifie bien la formule : $\frac{1}{6} + \dots + \frac{1}{6} = 6 \times \frac{1}{6} = 1$ (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, \dots, E_n s'ils forment une partition de Ω et que leurs probabilités sont toutes égales : $\mathbb{P}(E_1) = \mathbb{P}(E_2) = \dots = \mathbb{P}(E_n) = \frac{1}{n}$.

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 :

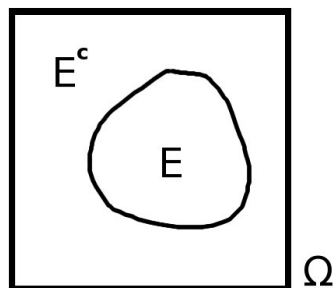
$$\begin{aligned} \Omega = & (\text{"on pioche les billes 1, 2, 3"}) \text{ OU } (\text{"on pioche les billes 1, 2, 4"}) \text{ OU } \dots \\ & \text{OU } (\text{"on pioche les billes 6, 7, 8"}) \end{aligned}$$

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

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

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

ou encore

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

- E_1 = “les 3 billes sont paires”
- E_2 = “les 3 billes sont impaires”
- E_3 = “2 billes sont paires et 1 impaire”
- E_4 = “2 billes sont impaires et 1 paire”

On a $\mathbb{P}(E_1) = \mathbb{P}(E_2) = \frac{1}{8} \frac{3}{7} \frac{2}{6} = \frac{1}{14}$. On a aussi $\mathbb{P}(E_3) = \mathbb{P}(E_4)$ et

$$\begin{aligned} \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3) + \mathbb{P}(E_4) &= 1 \\ \frac{1}{14} + \frac{1}{14} + 2\mathbb{P}(E_3) &= 1. \end{aligned}$$

On en déduit

$$\mathbb{P}(E_3) = \frac{1 - \frac{2}{14}}{2} = \frac{3}{7}.$$

Donc cette partition n’est pas équiprobable.

1.3.6 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

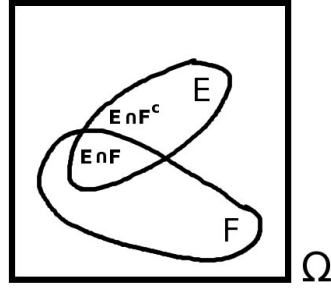
$$\boxed{\mathbb{P}(E^c) = 1 - \mathbb{P}(E),}$$

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

La relation suivante est très utile :

Soient E et F deux événements. Alors

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

FIGURE 1.5 – $\mathbb{P}(E) = \mathbb{P}(E \text{ ET } F) + \mathbb{P}(E \text{ ET } F^c)$.

$$\boxed{\mathbb{P}(E) = \mathbb{P}(E \text{ ET } F) + \mathbb{P}(E \text{ ET } F^c).}$$

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 :

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

$$A = A \text{ ET } (B \text{ OU } B^c) = (A \text{ ET } B) \text{ OU } (A \text{ ET } B^c).$$

On vérifie que cette union est disjointe :

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

$$(A \text{ ET } B) \text{ ET } (A \text{ ET } B^c) = A \text{ ET } B \text{ ET } A \text{ ET } B^c = \emptyset$$

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

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

$$\mathbb{P}(A) = \mathbb{P}(A \text{ ET } B) + \mathbb{P}(A \text{ ET } B^c).$$

□

1.3.7 Exclusion d'évènements

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

$$\mathbb{P}(B \text{ ET } A^c) = \mathbb{P}(B) - \mathbb{P}(A)$$

On note aussi $B \text{ ET } A^c = B \setminus A$ (diagramme).

Devinette : 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 ?

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 conditionnelles 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 (puisque'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 évènements. On note $\mathbb{P}(E|F)$ (ce qui se lit "probabilité de E sachant F ") la probabilité que l'évènement E se réalise sachant que l'évènement 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), au lieu de considérer toutes les issues possibles, ce changement de situation probabiliste consiste à considérer uniquement les issues de F , et donc à ne plus considérer ce qui est en dehors de F . Autrement dit, la probabilité d'un évènement 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 $\mathbb{P}(F)$, afin que dans la nouvelle situation, la probabilité de toutes les issues possibles soit bien égale à 1.

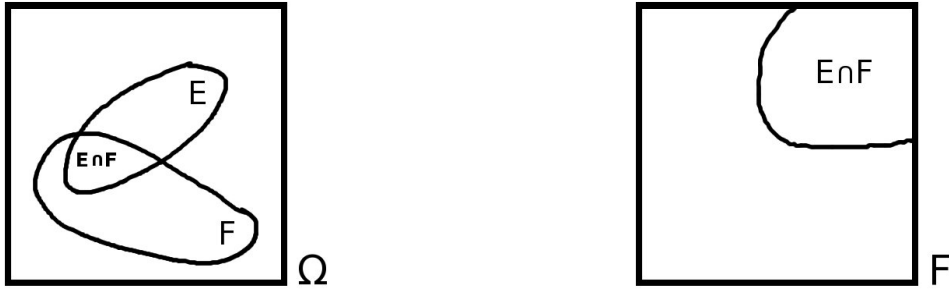


FIGURE 2.1 – Schéma du changement de situation probabiliste correspondant au conditionnement par rapport à un événement F . À gauche, situation initiale dans l'univers Ω ; à droite, situation conditionnée à F : l'événement F devient le nouvel univers, et seule la partie de E comprise dans F est considérée.

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

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

On voit par cette formule que $\mathbb{P}(E|F)$ n'est défini que si $\mathbb{P}(F) \neq 0$; 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. On utilisera toujours ce type de raisonnement en 3 étapes :

1. Définir l'expérience aléatoire : On tire 2 cartes à la suite dans un paquet de 32 cartes.

2. Introduction des événements 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".

3. Calculs et conclusion On a donc $C = A \cap B$. D'après la formule des probabilités conditionnelles, on a

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

La probabilité de tirer un pique en premier est de $\mathbb{P}(A) = \frac{8}{32}$. 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 $\mathbb{P}(B|A) = \frac{7}{31}$. Finalement on obtient

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

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 $\mathbb{P}(E|F)$ a un sens. En fait la probabilité conditionnelle $\mathbb{P}(E|F)$ se note parfois $\mathbb{P}_F(E)$, notation moins usuelle mais plus juste car elle traduit précisément le changement de situation probabiliste évoqué plus haut : la mesure \mathbb{P}_F est la mesure de probabilité de la nouvelle situation, en remplacement de la mesure \mathbb{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 l'univers est réduit à G . Cette nouvelle situation change les probabilités des événements... Les probabilités calculées conditionnellement à G obéissent aux mêmes règles que les probabilités calculées dans la configuration initiale. Pour cette raison on note parfois $\mathbb{P}(E|G) = P_G(E)$ 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 mêmes règles.

Voici quelques exemples de notions et formules découlant de ce principe. On indique en rouge ce qui change par rapport aux formules classiques (sans conditionnement) :

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

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

Ces formules se retiennent mieux avec la notation P_G :

$$P_G(E \cup F) = P_G(E) + P_G(F) - P_G(E \cap F).$$

Ça marche aussi avec le complémentaire :

$$\mathbb{P}(E^c|G) = 1 - \mathbb{P}(E|G)$$

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

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

$$P_{F \cap G}(E) = \frac{P_G(E \cap F)}{P_G(F)}.$$

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.

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

□

Formule des probabilités composées On peut généraliser cette formule à une suite d'événements A_1, A_2, \dots, A_n :

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1|A_2 \cap \dots \cap A_n)\mathbb{P}(A_2|A_3 \cap \dots \cap A_n)\dots\mathbb{P}(A_n)$$

Exemple On a dans un sac trois billes vertes et deux boules bleues. On enlève les boules du sac au hasard les unes après les autres. Quelle est la probabilité de l'événement A = "la première boule bleue tirée est la 4ème" ?

1. Expérience aléatoire : On tire successivement 4 billes du sac.

2. Introduction des événements : Pour $1 \leq i \leq 4$, on pose A_i = "la i -ème boule tirée est verte".

3. Calculs. On a

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A_4^c \cap A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_4^c|A_1 \cap A_2 \cap A_3)\mathbb{P}(A_3|A_1 \cap A_2)\mathbb{P}(A_2|A_1)\mathbb{P}(A_1) \\ &= 1 \cdot \frac{1}{3} \cdot \frac{2}{4} \cdot \frac{3}{5} = \frac{1}{10}.\end{aligned}$$

Devinette : Fils et filles Une personne affirme avoir 2 enfants, dont un fils. Quelle est la probabilité que l'autre enfant soit une fille ? La réponse est-elle modifiée si cette personne nous apprend en plus que le fils dont elle parle est né un mardi ?

Le plus grand nombre d'erreurs viennent de la confusion entre $\mathbb{P}(A|B)$ et $\mathbb{P}(B|A)$.

Voyons quelques exemples :

— Extrait du rattrapage de l'année dernière : Un journal mensuel offre en janvier 2017 un mois d'abonnement à 1000 personnes. Il estime que chacune d'entre elles a 60% de chances de rester abonné au moins un an.

400 clients ont moins de 20 ans, et 600 ont plus de 20 ans. On observe que parmi les clients qui sont restés jusqu'en janvier 2018, 61% ont moins de 20 ans. Déduisez-en la probabilité qu'un client reste abonné au moins un an sachant qu'il a plus de 20 ans.

Expérience aléatoire : On considère un client pris au hasard. L'expérience aléatoire concerne son âge, et le fait qu'il se réabonne ou non.

Événements :

- J : il a moins de 20 ans
- J^c : il a plus de 20 ans
- A : il se réabonne
- A^c : il se désabonne.

Données :

- $\mathbb{P}(A) = 0,6$ ($\mathbb{P}(A^c) = 0,4$).
- $P(J) = \frac{400}{400+600} = 0,4$.
- $\mathbb{P}(\dots|\dots) = 0,61$.

Comment remplir les pointillés ?

Exemple 1 (Probabilités conditionnelles et confusions). *Un test de grossesse annonce 99,9% d'efficacité. Pourtant, sur 10 femmes pour lequel le test est positif, seules 9 sont réellement enceintes. Qu'en pensez-vous ?*

Le test a deux manières d'être efficace :

- *En indiquant aux femmes enceintes qu'elles le sont ($\mathbb{P}(E|P)$)*
- *En indiquant aux femmes non-enceintes qu'elles ne le sont pas ($\mathbb{P}(E^c|P^c)$).*

Il est plus grave pour le test de se tromper dans le 1er cas, car il est important pour une femme enceinte de le savoir rapidement. Dans le second cas, la femme ira simplement faire des analyses complémentaires et s'apercevra que la prédiction était fausse.

Exemple 2. *Un politicien observe lorsqu'il arrive au pouvoir que parmi les diplômés, 10% viennent des milieux défavorisés. Lors de sa campagne de réélection, quatre ans plus tard, il clame que sa politique a été efficace car désormais 30% des jeunes défavorisés ont obtenu un diplôme.*

Soit l'expérience aléatoire : On prend un individu au hasard dans la population. On s'intéresse aux événements D = "il est issu d'un milieu défavorisé" et E = "il est diplômé". Les données sont $\mathbb{P}(D|E) = 0,1$ et $\mathbb{P}(E|D) = 0,3$, ce qui ne veut rien dire si on n'en connaît pas plus.

Dans ces deux exemples, la formule de Bayes nous permettra de statuer si l'on a des données complémentaires.

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 $\mathbb{P}(B_2)$. La formule suivante, vue au premier chapitre, est le point de départ du calcul :

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

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 $\mathbb{P}(B_2 \cap B_1)$ et $\mathbb{P}(B_2 \cap B_1^c)$, ce qui donne :

$$\mathbb{P}(B_2) = \mathbb{P}(B_2|B_1)\mathbb{P}(B_1) + \mathbb{P}(B_2|B_1^c)\mathbb{P}(B_1^c).$$

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

- Pour déterminer $\mathbb{P}(B_2|B_1)$ 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 $\mathbb{P}(B_2|B_1) = \frac{1}{2}$.
- Pour déterminer $\mathbb{P}(B_2|B_1^c)$ 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 $\mathbb{P}(B_2|B_1^c) = \frac{2}{6} = \frac{1}{3}$.
- $\mathbb{P}(B_1)$ est la probabilité que la première boule tirée soit blanche : $\mathbb{P}(B_1) = \frac{1}{3}$.
- Enfin $\mathbb{P}(B_1^c) = 1 - \mathbb{P}(B_1) = \frac{2}{3}$.

Finalement on trouve donc

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

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 $\mathbb{P}(F)$ et $\mathbb{P}(F^c)$ non nuls. Alors

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

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

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

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

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 reçoit le médicament MA , alors on guérit dans 98% des cas". On a donc $\mathbb{P}(E^c|A) = 0.98$, et de même $\mathbb{P}(E^c|B) = 0.96$ et $\mathbb{P}(E^c|C) = 0.95$. 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) :

$$\mathbb{P}(E^c) = \mathbb{P}(E^c|A)\mathbb{P}(A) + \mathbb{P}(E^c|B)\mathbb{P}(B) + \mathbb{P}(E^c|C)\mathbb{P}(C).$$

1er cas : $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{3}$. Donc $\mathbb{P}(E^c) = 0.98 \times \frac{1}{3} + 0.96 \times \frac{1}{3} + 0.95 \times \frac{1}{3} = 0.963$. Finalement $\mathbb{P}(E) = 1 - \mathbb{P}(E^c) = 0.037$.

2e cas : $\mathbb{P}(A) = 0.5$, $\mathbb{P}(B) = 0.3$ et $\mathbb{P}(C) = 0.2$. Donc $\mathbb{P}(E^c) = 0.98 \times 0.5 + 0.96 \times 0.3 + 0.95 \times 0.2 = 0.968$, et $\mathbb{P}(E) = 0.032$.

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

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

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 $\mathbb{P}(B_1|B_2)$. Cette probabilité conditionnelle ne peut pas être trouvée directement comme c'est le cas pour $\mathbb{P}(B_2|B_1)$. 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 :

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

A présent on peut effectuer le calcul :

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

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

Formule de Bayes . Soient E et F deux événements avec $\mathbb{P}(E)$, $\mathbb{P}(F)$ et $\mathbb{P}(F^c)$ non nuls. Alors

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

Cette formule permet d’une certaine manière de “retourner” la probabilité conditionnelle. Il est plus facile de savoir le retrouver que de l’apprendre par coeur.

Exemple 3. *Un mammogramme est un test qui doit détecter si une femme a le cancer du sein. On estime que ce cancer touche environ 0,8% des femmes. Une étude montre que si une femme n’a pas le cancer, le mammogramme va quand même être positif dans 10% des cas. Si elle a le cancer, le mammogramme le détectera 75% du temps. La question est : si une femme a un mammogramme positif, faut-il s’affoler ?*

Traduisons les données :

- C : “femme atteinte du cancer du sein”
- P : “mammogramme positif”
- $\mathbb{P}(C) = 0,008$
- $\mathbb{P}(P|C^c) = 0,1$
- $\mathbb{P}(P|C) = 0,75$
- $\mathbb{P}(C|P) = ??$

La réponse est qu’il y environ 6% de chances d’avoir le cancer du sein. En effet, Bayes nous dit

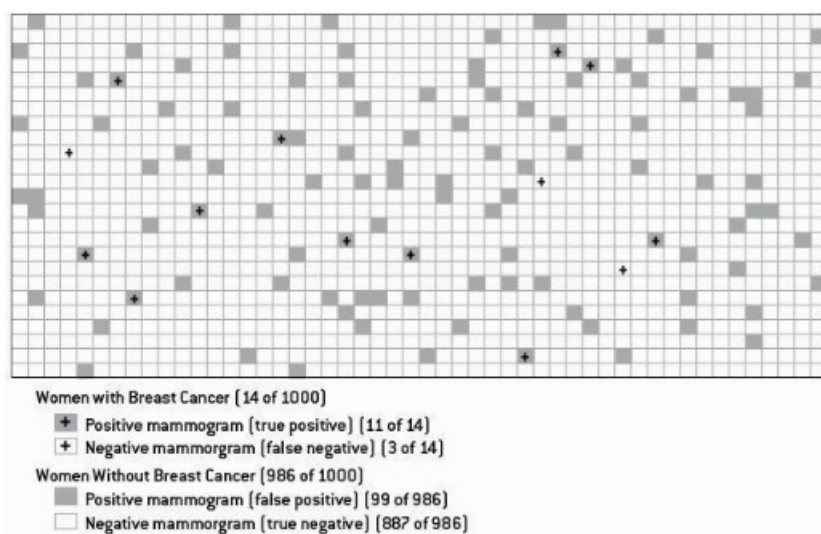
$$\begin{aligned} \mathbb{P}(C|P) &= \frac{\mathbb{P}(P|C)\mathbb{P}(C)}{\mathbb{P}(P)} = \frac{0,75 \cdot \mathbb{P}(C)}{\mathbb{P}(P|C)\mathbb{P}(C) + \mathbb{P}(P|C^c)\mathbb{P}(C^c)} \\ &= \frac{0,75 \cdot 0,008}{0,75 \cdot 0,008 + 0,1 \cdot (1 - 0,008)}. \end{aligned}$$

Ce qui donne $\mathbb{P}(C|P) = 5,7\%$. On peut visualiser la population avec cette grille pour mieux se rendre compte :

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. 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 sont à Brest mais ne se sont jamais rencontrés, les évènements E et F sont-ils indépendants ?

Autre exemple Je jette un dé. L'évènement "le dé tombe sur un nombre pair" est-il indépendant de l'évènement "le dé fait un score ≥ 3 " ?

Réponse (1er exemple) : 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 :

$$\mathbb{P}(F|E) > \mathbb{P}(F).$$

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

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

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

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

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

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F).$$

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

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 $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$ est alors utilisée pour faire des calculs.

réponse au 2ème exemple B ="le résultat est un nombre pair", C ="le résultat est ≥ 3 ". En principe B et C n'ont aucune raison d'être indépendants puisqu'ils concernent le même lancer de dé. Pourtant on a $\mathbb{P}(B) = \frac{1}{2}$, $\mathbb{P}(C) = \frac{2}{3}$, $\mathbb{P}(B \cap C) = \frac{1}{3}$. On a donc $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$, et donc B et C sont mathématiquement indépendants.

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 $\mathbb{P}(B)\mathbb{P}(D) = \frac{1}{2} \times \frac{1}{2} \neq \mathbb{P}(B \cap D) = \frac{1}{3}$; c'est-à-dire que B et D ne sont pas indépendants.

Pour résumer, l'analyse du contexte peut nous dire si des événements sont indépendants (comme les résultats de deux dés), 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 ou plus

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 $\mathbb{P}(C) = \frac{1}{2}$; 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 $\mathbb{P}(C|A) = \mathbb{P}(C) = \frac{1}{2}$ 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)

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

$$\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C),$$

$$\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C).$$

Cependant on voit bien que les 3 événements ne sont pas indépendants, car si par exemple A et B sont réalisés, alors C est sûr de ne pas être réalisé. Avoir de l'information sur A et B nous apporte bel et bien de l'information sur C .

Définition. On dit que des événements A_1, \dots, A_n sont indépendants si pour tout sous-ensemble I de $\{1, 2, \dots, n\}$ tel que $\#I \geq 2$,

$$\mathbb{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i).$$

Pour $n = 2$ il n'y a que l'ensemble $I = \{1, 2\}$, et tout ce qu'il faut vérifier est $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$, comme on a déjà vu.

Par exemple pour $n = 3$ il faut que toutes les relations suivantes soient vraies :

- $I = \{1, 2\} : \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$
- $I = \{1, 3\} : \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_3)$
- $I = \{2, 3\} : \mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2)\mathbb{P}(A_3)$
- $I = \{1, 2, 3\} : \mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$

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

Pour $n = 4$ c'est plus compliqué...

2.5 Indépendance conditionnelle

La notion d'indépendance conditionnelle est souvent utile : On introduit les événements

- P = "il pleut à Brest"
- A = "Mon ami qui habite à Brest prend son parapluie"
- G = "Ma grand-mère qui habite à Brest prend son parapluie"

et on suppose que mon ami et ma grand-mère ne se connaissent pas, que mon ami prend son parapluie 2 fois sur 3 quand il pleut, et ma grand-mère 4 fois sur 5.

Comme ils ne se connaissent pas, on peut dire que A et G sont indépendants **conditionnellement** à P , alors que A et G ne sont pas indépendants, ce qui se traduit par $\mathbb{P}(A \cap G|P) = \mathbb{P}(A|P)\mathbb{P}(G|P)$. En effet, si on suppose qu'il pleut $1/4$ du temps à Brest et que personne ne prend son parapluie quand il fait beau, on a

$$\mathbb{P}(A \cap G) = \mathbb{P}(A \cap G|P)\mathbb{P}(P) + 0 = \mathbb{P}(A|P)\mathbb{P}(G|P)\mathbb{P}(P) = \frac{2}{3} \frac{4}{5} \frac{1}{4} = \frac{2}{15}$$

alors que

$$\mathbb{P}(A)\mathbb{P}(G) = \mathbb{P}(A|P)\mathbb{P}(P)\mathbb{P}(G|P)\mathbb{P}(P) = \frac{2}{60}.$$

Indépendance conditionnelle : Soient E, F, G trois événements, avec $\mathbb{P}(G) \neq 0$. 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,

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

Des événements conditionnellement indépendants n'ont aucune raison d'être 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 " $X = 1$ ", " $X = 2$ ", ..., " $X = 6$ " : on a $\mathbb{P}("X = 1") = \mathbb{P}("X = 2") = \dots = \mathbb{P}("X = 6") = \frac{1}{6}$.

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

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 $\mathbb{P}("X = 1")$ ou $\mathbb{P}([X = 1])$ on écrira simplement $\mathbb{P}(X = 1)$ 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 $X \in \{0, 1, 2, \dots\}$, $\mathcal{S}(X) = \mathbf{N}$.

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 $[0, \pi]$.

Intéressons-nous dans les 3 cas à la probabilité $\mathbb{P}(X = x)$ pour un certain x du support.

1. Si je prends $x = 4$, j'ai $\mathbb{P}(X = 1) = 1/6$.
2. Si je prends $x = 2$, j'ai $\mathbb{P}(X = 2) = \frac{5}{6} \cdot \frac{1}{6} = \frac{5}{36}$.
3. Si je prends $x = \pi/4$, quelle est la chance que le stylo fasse **exactement** un angle de $\pi/4$? La réponse est 0. On pourra lancer le style autant de fois qu'on veut, il ne tombera jamais exactement sur $\pi/4$, même s'il se rapprochera beaucoup. On aboutit donc au paradoxe suivante :

$$\forall x \in [0, \pi], \mathbb{P}(X = x) = 0 \text{ mais pourtant } \mathbb{P}(X \in [0, \pi]) = 1.$$

Cela vient du fait qu'on ne peut pas écrire

$$\mathbb{P}(X \in [0, \pi]) = \sum_{x \in [0, \pi]} \mathbb{P}(X = x)$$

alors qu'on pouvait écrire dans le deuxième cas

$$\mathbb{P}(X \in \mathbf{N}) = \sum_{n \in \mathbf{N}} \mathbb{P}(X = n).$$

Cela est du à la nature de l'ensemble $[0, \pi]$, qui est un intervalle continu de \mathbb{R} , alors que \mathbf{N} est l'ensemble des entiers. On dit que $[0, \pi]$ est infini indénombrable alors que \mathbf{N} est infini dénombrable.

En pratique, les variables discrètes auront comme support un ensemble fini (ex : $\{1, 2, 3, 4, 5, 6\}$), ou un ensemble infini dénombrable comme \mathbf{N} ou \mathbf{Z} .

aa

Définition.

Il existe également des variables qui ne sont ni discrètes ni continues.

Exemple 4. Je joue au golf, et j'appelle X la distance entre ma balle et le trou après un coup. J'ai $\mathcal{S}(X) = [0, +\infty[$. J'ai $\mathbb{P}(X = 0) > 0$ car il est possible que je mette la balle dans le trou, et $\mathbb{P}(X = x) = 0$ pour $x > 0$. Donc X n'est ni discrète ni continue.

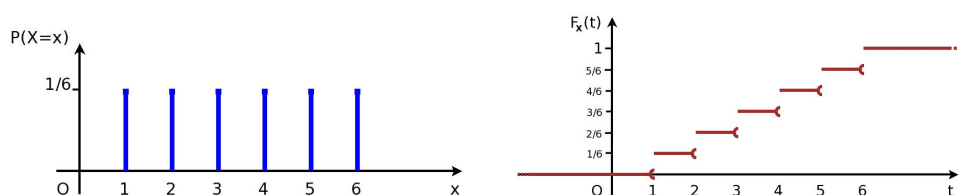


FIGURE 3.1 – Loi et fonction de répartition d'un lancer de dé

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

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

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 $\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \dots = \mathbb{P}(X = 6) = \frac{1}{6}$: c'est la loi de X . On peut la présenter sous forme de tableau :

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

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.

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 $\mathbb{P}(X \leq t) \leq \mathbb{P}(X \leq u)$, c'est-à-dire $F_X(t) \leq F_X(u)$. Donc F_X est croissante. \square

Grâce à la fonction de distribution, on peut calculer sa probabilité de tomber dans un intervalle quelconque : Pour $t, s \in \mathbf{R}$ tel que $t \leq s$,

$$\mathbb{P}(X \in]t, s]) = \mathbb{P}((X \leq s) \setminus (X \leq t)) = \mathbb{P}(X \leq s) - \mathbb{P}(X \leq t) = F_X(s) - F_X(t).$$

3.1.4 Variables discrètes

Loi d'une variable discrète. Donner la loi d'une variable aléatoire discrète X , c'est calculer les probabilités $\mathbb{P}(X = x)$ pour toutes les valeurs x possibles prises par X (autrement dit pour tous les x appartenant au support de X).

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

Exemple 5. 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 $E = \{\text{"pile"}, \text{"face"}\}$.

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 "Pile" obtenu. Quelle est la loi de X ?

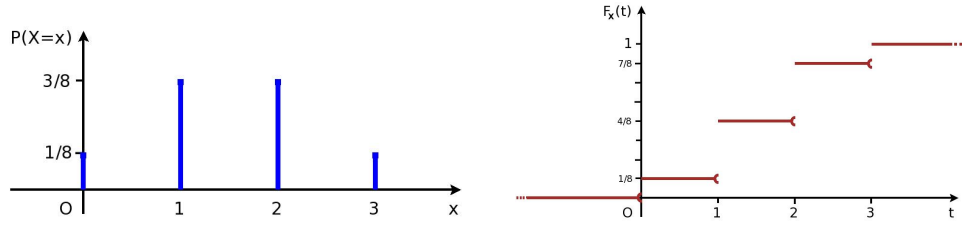
Le support de X est ici $\{0, 1, 2, 3\}$. On doit donc calculer $\mathbb{P}(X = 0)$, $\mathbb{P}(X = 1)$, $\mathbb{P}(X = 2)$ et $\mathbb{P}(X = 3)$ pour décrire complètement la loi de X .

Définissons $P_1 = \text{"la première pièce tombe sur Face"}$; et de même P_2 et P_3 . On peut clairement supposer que les trois lancers de pièce sont indépendants ici; et donc que P_1, P_2, P_3 sont indépendants.

- $[X = 0] = P_1^c \cap P_2^c \cap P_3^c$ donc $\mathbb{P}(X = 0) = \mathbb{P}(P_1^c)\mathbb{P}(P_2^c)\mathbb{P}(P_3^c)$ grâce à l'indépendance. Ainsi $\mathbb{P}(X = 0) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$.
- $[X = 1] = (P_1 \cap P_2^c \cap P_3^c) \cup (P_1^c \cap P_2 \cap P_3^c) \cup (P_1^c \cap P_2^c \cap P_3)$. Cette union est disjointe, donc on peut additionner les probabilités :

$$\begin{aligned} \mathbb{P}(X = 1) &= \mathbb{P}(P_1 \cap P_2^c \cap P_3^c) + \mathbb{P}(P_1^c \cap P_2 \cap P_3^c) + \mathbb{P}(P_1^c \cap P_2^c \cap P_3), \\ &= \mathbb{P}(P_1)\mathbb{P}(P_2^c)\mathbb{P}(P_3^c) + \mathbb{P}(P_1^c)\mathbb{P}(P_2)\mathbb{P}(P_3^c) + \mathbb{P}(P_1^c)\mathbb{P}(P_2^c)\mathbb{P}(P_3), \\ &\quad \text{(par indépendance)} \\ &= \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) \times 3 = \frac{3}{8}. \end{aligned}$$

- Pour calculer intelligemment $\mathbb{P}(X = 2)$, 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 $\mathbb{P}(X = 2) = \mathbb{P}(Y = 1) = \mathbb{P}(X = 1) = 3/8$.
- Avec le même raisonnement, $\mathbb{P}(X = 3) = \mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = 1/8$.

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

La loi de X est en fait un exemple de loi binomiale. La figure 3.2 montre le graphique de cette loi ainsi que la fonction de répartition.

Supposons désormais que la pièce soit biaisée, c'est-à-dire qu'elle a une probabilité p de tomber sur pile, où $p \in [0, 1]$ est un paramètre inconnu. Alors on a avec le même raisonnement

$$\begin{aligned}\mathbb{P}(X = 0) &= (1 - p)^3 \\ \mathbb{P}(X = 1) &= 3p^2(1 - p) \\ \mathbb{P}(X = 2) &= 3p(1 - p)^2 \\ \mathbb{P}(X = 3) &= p^3.\end{aligned}$$

Finalement, supposons qu'on ne lance plus 3 pièces mais n pièces, où $n \in \mathbb{N}^*$. Cette fois $\mathcal{S}(X) = \{0, 1, 2, \dots, n\}$. Pour donner la loi de X , il faut calculer pour tout $k \in \{0, 1, 2, \dots, n\}$ $\mathbb{P}(X = k)$. L'évènement $X = k$ survient si exactement k pièces tombent sur Pile, et $n - k$ tombent sur Face. On numérote les pièces de 1 à n . L'évènement $X = k$ survient donc si il y a un sous-ensemble P à k éléments tel que P_i est vrai pour $i \in P$, et P_i^c est vrai pour $i \notin P$. En sommant sur tous les sous-ensembles P possibles on a

$$\begin{aligned}\mathbb{P}(X = k) &= \sum_{\substack{\text{sous-ensembles } P \\ \text{à } k \text{ éléments}}} \mathbb{P}(P_i \text{ est vrai pour } i \in P, P_i^c \text{ est vrai pour } i \notin P) \\ &= \sum_{\substack{\text{sous-ensembles } P \\ \text{à } k \text{ éléments}}} \underbrace{\prod_{i \in P} \mathbb{P}(P_i)}_{=p^k \text{ car il y a } k \text{ elements dans } P} \underbrace{\prod_{i \notin P} \mathbb{P}(P_i^c)}_{=(1-p)^{n-k}} \text{ car les évènements sont indépendants} \\ &= \sum_{\substack{\text{sous-ensembles } P \\ \text{à } k \text{ éléments}}} p^k (1 - p)^{n-k} \\ &= \#\{\text{sous-ensembles a } k \text{ elements}\} p^k (1 - p)^{n-k} \\ &= C_n^k p^k (1 - p)^{n-k}.\end{aligned}$$

On remarque que la somme de toutes les probabilités est bien égale à 1 car par la

formule du binôme

$$\sum_{k=0}^n \mathbb{P}(X = k) = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

Définition. Soit $n \in \mathbb{N}^*$ et $p \in [0, 1]$. On appelle **loi binomiale** de paramètre n et p , notée $\mathcal{B}(n, p)$ la loi d'une variable X telle que $\mathcal{S}(X) = \{0, 1, 2, \dots, n\}$ et pour $0 \leq k \leq n$,

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

Une variable binomiale représente le nombre de succès lors de n répétitions indépendantes d'expériences aléatoires qui ont chacune une probabilité p de réussir.

Exemple 6. — 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 à 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.

On appelle **expérience de Bernoulli** une expérience aléatoire qui n'a que deux issues possibles ("Pile" ou "Face", "Noir" ou "Blanc", "succès" ou "échec", 0 ou 1, etc...).

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

Exemple 7. — 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 $X \sim \mathcal{B}(1/6)$.

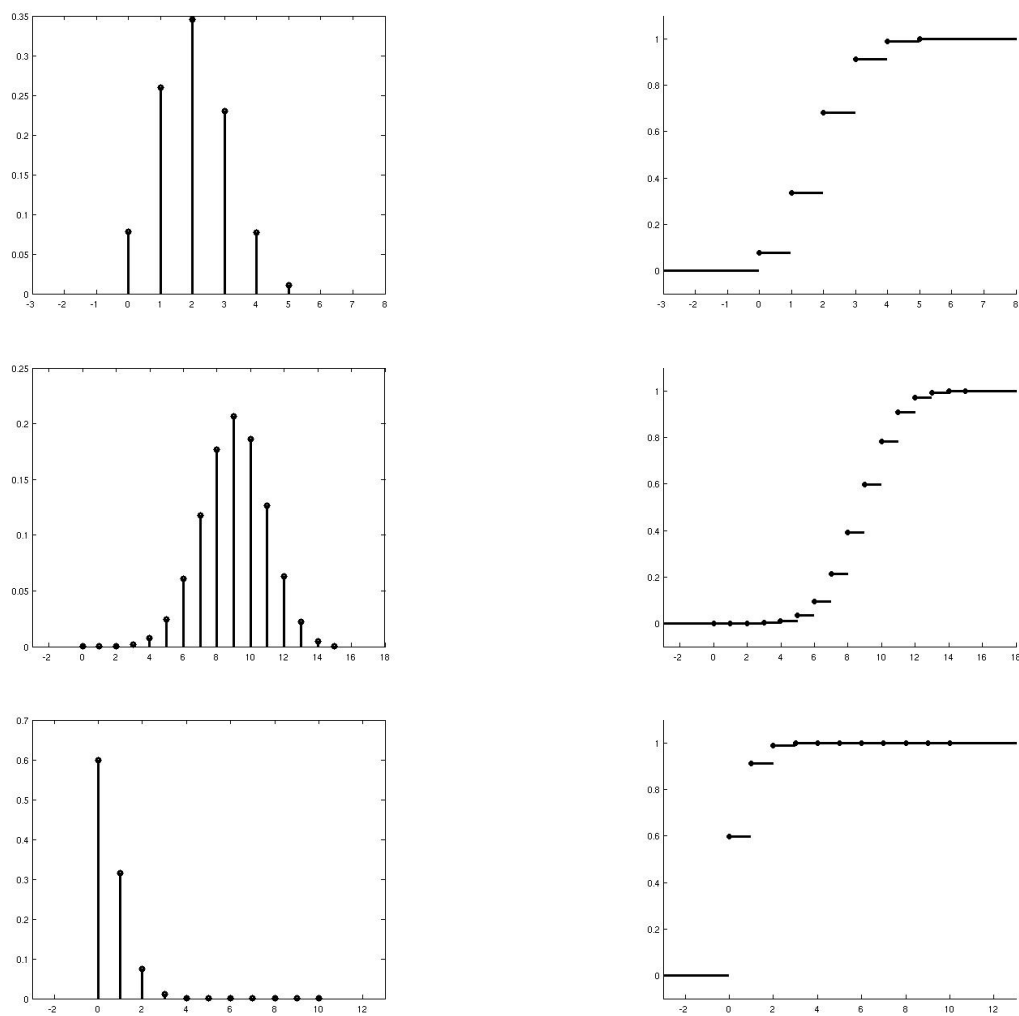


FIGURE 3.3 – lois binomiales (à gauche) et fonctions de répartition F_X correspondantes (à droite) pour différents paramètres : 1e ligne : $n = 5, p = 0.4$; 2e ligne : $n = 15, p = 0.6$; 3e ligne : $n = 10, p = 0.05$.

Exemple 8. Soit $n \geq 1$ et $p \in [0, 1]$. Soit $X_i, i = 1, \dots, n$ des variables de Bernoulli indépendantes de paramètre p . Soit $X = \sum_{i=1}^n X_i$. Alors $X \sim \mathcal{B}(n, p)$. En effet, en réfléchissant un peu, on voit que

$$X = \#\{i : X_i = 1\}.$$

Donc X est le nombre d'expériences de Bernoulli qui réussissent, où chaque expérience a une probabilité p de réussir, indépendamment des autres.

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

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

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

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 .

Une autre manière de définir la loi géométrique est via sa fonction de répartition :

$$\begin{aligned} \mathbb{P}(X \leq n) &= 1 - \mathbb{P}(X > n) = 1 - \mathbb{P}(A_1^c \cap \dots \cap A_n^c) \\ &= 1 - \mathbb{P}(A_1^c) \dots \mathbb{P}(A_n^c) \\ &= 1 - (1 - p)^n. \end{aligned}$$

Pour une loi géométrique il est souvent plus simple de considérer la fonction "quantile" :

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

Loi de Poisson. On considère un ensemble d'évènements qui peuvent intervenir dans le temps à n'importe quel moment, indépendamment les uns des autres, et tel qu'aucune période ne soit privilégiée par rapport à une autre par rapport à l'arrivée de ces évènements.

Exemples :

- Clients qui rentrent dans un bureau de poste
- Poissons qui mordent à un hameçon (en supposant qu'on peut les décrocher et remettre la ligne instantanément)
- Naissances dans une population donnée
- Avions qui passent dans le ciel
- Tremblements de terre dans le monde

Alors le nombre d'évènements qui surviennent dans une période donnée (1 heure, 1 journée, ...) suit une loi de Poisson. On voit bien que le nombre d'avions qui passent dans le ciel et le nombre de tremblements de terre ne suivent pas la même loi. Il s'agit en fait d'une famille de lois qui dépendent d'un paramètre $\lambda > 0$ (comme la famille des géométriques, des exponentielles, ...)

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

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

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

λ nous donne (ou est défini par) le nombre moyen d'évènements qui surviennent sur la période donnée.

Exercice 1. Expliquer pourquoi les VA suivantes ne suivent pas une loi de Poisson :

- Nombre de fois que le record du 100m sera battu dans le prochain siècle
- Nombre d'étudiants qui vont valider la matière dans un groupe de 30 étudiants
- Nombre de gagnants au loto dans une population de 10 millions d'habitants.

Lien Poisson - binomiale

La loi de Poisson est la version en temps continu de la loi binomiale (mais ça reste une variable discrète). En effet, considérons l'exemple du pêcheur, et du nombre X de poissons (sans majuscule) attrapés en une heure. On note $\lambda = \mathbb{E}X$ le nombre moyen de poissons attrapés en une heure : $X \sim \mathcal{P}(\lambda)$. On découpe l'heure en 60 minutes et on introduit Y le nombre de minutes où il a attrapé 1 ou plusieurs poissons. Comme ce qui se passe lors d'une minute donnée (disons la 10ème minute) est indépendant de ce qui se passe à une autre (la 27ème minute par exemple), alors Y est une loi binomiale de paramètre $n = 60$, et $p = \mathbb{P}(\text{il attrape un poisson ou plus sur une durée d'une minute})$. Donc Y s'obtient à partir de X en "découpant" le temps. Pour résumer :

- Le nombre de poissons attrapés suit une loi de Poisson
- Le nombre de minutes où il a attrapé un ou des poisson(s) suit une loi binomiale.

3.1.5 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 $[X = x] \cap [X = y] = \emptyset$ si $x \neq y$. Par conséquent la somme totale des $\mathbb{P}(X = x)$ doit toujours être égale à 1, ce qui s'écrit, en notant \mathcal{S} le support de X ,

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

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 la loi géométrique, on a une série géométrique

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

- Pour la loi binomiale : si $X \sim \mathcal{B}(n, p)$,

$$\sum_{x \in \mathcal{S}(X)} \mathbb{P}(X = x) = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

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 $\mathbb{P}(X = x)$ 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 :

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

3.2 Variables aléatoires continues, ou à densité

3.2.1 variable uniforme

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 $\mathbb{P}(X = \alpha)$ pour $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. En fait on verra que $\mathbb{P}(X = \alpha)$ 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** (et c'est le mot important ici) un angle α avec la table, de la même manière qu'aucun être humain sur terre ne mesure **exactement** 1m70, au milliardième de nanomètre près. 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 $\mathbb{P}(0 \leq X \leq \frac{\pi}{2}) = \frac{1}{2}$, ou que $\mathbb{P}(\frac{\pi}{4} \leq X \leq \frac{\pi}{2}) = \frac{1}{4}$. Plus généralement, on peut 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 pour $0 \leq \alpha \leq \beta \leq \pi$,

$$\mathbb{P}(\alpha \leq X \leq \beta) = c(\beta - \alpha)$$

où c est une certaine constante de proportionnalité que l'on va déterminer.

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.

3.2.2 Définition

Pour une variable aléatoire continue X , il existe une fonction $f_X : \mathbf{R} \rightarrow \mathbf{R}^+$ telle que

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

*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

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

En effet, pour $[a, b] \subset [0, \pi]$, la probabilité que $X \in [a, b]$ est proportionnel à la longueur de l'intervalle $[a, b]$:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx = c(b - a).$$

Pour déterminer la valeur de c , il suffit de remarquer que

$$1 = \mathbb{P}(X \in [-\pi/2, \pi/2]) = c \left(\frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right) = c\pi,$$

d'où $c = \frac{1}{\pi}$.

Si par exemple $a = -1, b \in [0, \pi]$,

$$\mathbb{P}(a \leq X \leq b) = \frac{b}{\pi} = \int_{-1}^b f(x)dx = \int_0^b \frac{1}{\pi}dx = \frac{b}{\pi}.$$

La formule est bien vérifiée dans ce cas de figure. Elle 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

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

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

Interprétation graphique : La probabilité $\mathbb{P}(a \leq X \leq b)$ 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 $\mathbb{P}(X = a) = \mathbb{P}(a \leq X \leq a) = \int_a^a f_X(x)dx = 0$. Du coup,
 $\mathbb{P}(a \leq X \leq b) = \mathbb{P}("a < X \leq b" \cup "X = a") = \mathbb{P}(a < X \leq b) + \mathbb{P}(X = a) = \mathbb{P}(a < X \leq b)$.
 Pour des raisons similaires,

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b).$$

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 à 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 $\mathbb{P}(X = x)$, pour $x \in \mathcal{S}(X)$).

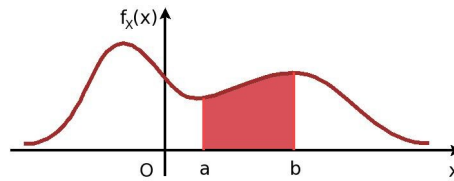


FIGURE 3.4 – Graphe d’une densité de probabilité f_X . La partie colorée correspond à $\mathbb{P}(a \leq X \leq b)$.

3.2.3 Variable gaussienne

Je relève chaque jour la différence X entre la température extérieure et la moyenne saisonnière.

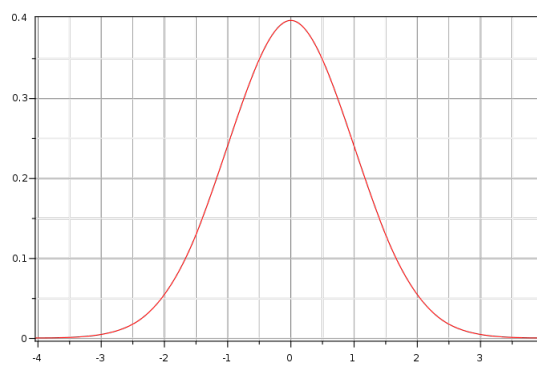
L’ensemble des valeurs possibles pour cette variable aléatoire est \mathbb{R} . En suivant le même raisonnement qu’auparavant, $\mathbb{P}(X = t) = 0$ pour $t \in \mathbb{R}$, et il faut ici aussi considérer des intervalles, et une fonction de densité pour étudier cette variable. La distribution gaussienne est souvent utilisée pour ce type de variable qui a en général des petites variations, positives et négatives, et très rarement de grandes variations.

On appelle **variable gaussienne de déviation standard σ^2** la variable continue X avec la densité

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

- Cette densité est souvent utilisée par les mathématiciens car elle a des bonnes propriétés mathématiques et correspond à ce que l’on observe pour de nombreux phénomènes naturels.
- On divise par $\sqrt{2\pi\sigma^2}$ car $\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) = \sqrt{2\pi\sigma^2} \neq 1$, donc il faut diviser pour avoir une vraie densité de probabilité (i.e. une fonction positive dont l’intégrale vaut 1)

On note cette loi $\mathcal{N}(0, \sigma^2)$. Voici la courbe de la fonction f_X pour $\sigma = \sqrt{2} \approx 1,41$:



On l’appelle courbe de Gauss, ou courbe en cloche. Elle présente la particularité d’avoir une grosse bosse au milieu, et de décroître ensuite très rapidement vers 0. Cela implique

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad \Phi(-x) = 1 - \Phi(x), \quad 0 \leq x < 1.99$$

X\0.0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670

FIGURE 3.5 – X : variable gaussienne standard. Cette table permet de déterminer $\mathbb{P}(X \geq x)$ pour x entre 0 et 2

que la variable aléatoire qui a cette densité a de plus fortes chances de prendre des valeurs proches de 0, ou plus précisément dans l'intervalle $[-\sigma, \sigma]$.

Pour savoir quelle est la probabilité que ma température reste sous un seuil donné, on peut utiliser la fonction de répartition :

$$\mathbb{P}(X \leq t) = \int_{-\infty}^t f_X(x) dx = \int_{-\infty}^t \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma^2} dx$$

Voilà une table donnant ses valeurs pour $\sigma = 1$.

En supposant qu'on mesure la température en degrés celsius, on choisit $\sigma = 5$ (on verra plus tard comment choisir une valeur pertinente pour σ), ce qui veut dire que la température varie en moyenne de $\pm 5^\circ$ de la moyenne. Remarquons que si on définit $Y = \frac{X}{5}$, alors Y suit des variations d'amplitude 1. On choisit donc comme loi $Y \sim \mathcal{N}(0, 1)$. On a **renormalisé** X pour obtenir une loi gaussienne de “variabilité” 1 (on appellera en fait ce nombre la “variance”).

On appelle cette loi la **loi Gaussienne standard**, ou **loi normale**, ou loi **gaussienne centrée réduite**.

Si maintenant on appelle Z la température “tout court”, on a $Z = m + X$, où $m = 17$ est la moyenne saisonnière (par exemple...). La valeur moyenne de Z est m et on note la loi de Z par $\mathcal{N}(m, \sigma^2)$.

En conclusion :

— $Y \sim \mathcal{N}(0, 1)$ signifie que la densité de Y est $f_Y(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$

- Pour $m \in \mathbb{R}, \sigma > 0$, on note $\mathcal{N}(m, \sigma^2)$ la loi de la variable $Z = m + \sigma Y$. Elle a pour densité

$$f_Z(x) = \frac{\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}$$

Proposition. Soit $X \sim \mathcal{N}(0, 1)$, et $\sigma > 0$. Alors $\sigma X \sim \mathcal{N}(0, \sigma^2)$. De plus pour $m \in \mathbb{R}$,

$$m + \sigma X \sim \mathcal{N}(m, \sigma^2).$$

Exemple 9. Un joueur joue aux fléchettes. Soit X la distance entre sa fléchette et le centre de la cible. Alors il est naturel de modéliser X par une loi $\mathcal{N}(0, \sigma^2)$, cela signifie que la fléchette va être typiquement à une distance plus ou moins σ du centre de la cible, et plus rarement à de plus larges distances. Si le joueur est mauvais, on peut mettre $\sigma = 50$ (on compte en centimètres), et s'il est très bon, on peut mettre $\sigma = 2$ (toujours en cm).

3.2.4 Loi exponentielle

Un pêcheur met son hameçon à l'eau. Il appelle X le temps (aléatoire) qu'il devra attendre avant qu'un poisson ne s'attaque à son appât. On voit bien que X est continue car la probabilité $\mathbb{P}(X = t) = 0$ pour un temps donné $t \geq 0$. On peut aussi dire que X est une variable aléatoire **sans mémoire** car elle satisfait la condition suivante : sachant qu'aucun poisson n'a mordu au temps t , la probabilité qu'il morde dans les prochaines minutes ne dépend pas de t . Autrement dit, ça n'est pas parce que vous n'avez attrapé aucun poisson en une heure dans un lac rempli de poissons que vous avez plus de chances d'y arriver que vous n'en aviez au départ. Cette propriété ressemble à celle des variables discrètes de loi géométriques : si on lance le dé 100 fois sans faire de 6, la probabilité qu'on fasse 6 au 101ème coup n'a pas changé, elle est toujours de $1/6$.

Cette loi est en fait l'analogue continu de la loi géométrique. C'est la loi du temps que mettra un évènement à survenir sachant que la probabilité qu'il survienne ne change pas avec le temps.

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

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

Si on divisait le temps en minutes au lieu de considérer tous les temps possibles, la loi Y de la première minute où un Poisson mord est pour $k \in \mathbb{N}$

$$\mathbb{P}(Y = k) = (1 - p)^{k-1}p$$

où p est la probabilité qu'un Poisson morde dans une minute donnée.

Autrement dit :

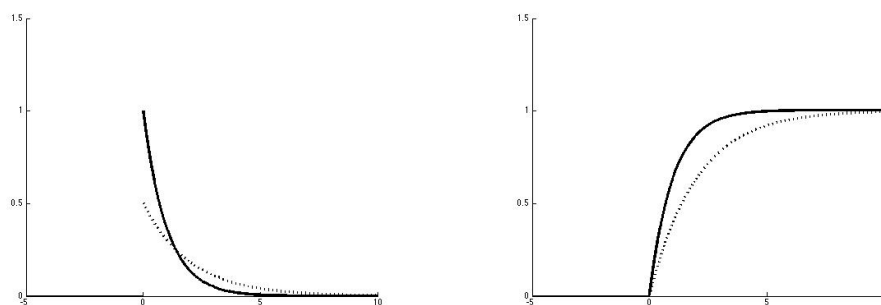


FIGURE 3.6 – Graphes de la densité et de la fonction de répartition de la loi exponentielle pour $a = 1$ (traits pleins) et $a = 0.5$ (pointillés). La densité est la dérivée de la fonction de répartition.

- Le premier temps où l'on va attraper un poisson suit une loi exponentielle
- Le nombre de minutes (entières) à attendre avant d'attraper un Poisson suit une loi géométrique

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

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

Il est plus simple de retenir que pour une variable exponentielle,

$$\mathbb{P}(X > x) = e^{-ax}.$$

On peut en déduire facilement $F_X(x) = 1 - \mathbb{P}(X > x)$, et la densité $f_X(x) = F'_X(x)$.

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

3.2.5 Fonction de répartition d'une variable aléatoire à densité

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

Ainsi F_X est une primitive de la fonction densité f_X , et f_X est donc la dérivée de F_X .

On répète que vouloir donner une valeur (autres que 0) à la quantité $P(Z = x)$ n'a pas de sens, et en particulier **il ne faut surtout pas écrire**

$$\mathbb{P}(Z = x) = f_Z(x).$$

On va néanmoins ici pouvoir donner une valeur de probabilité associée à une valeur x donné en considérant un petit intervalle $[x, x + \varepsilon]$, où $\varepsilon > 0$: on a

$$\mathbb{P}(x \leq Z \leq x + \varepsilon)$$

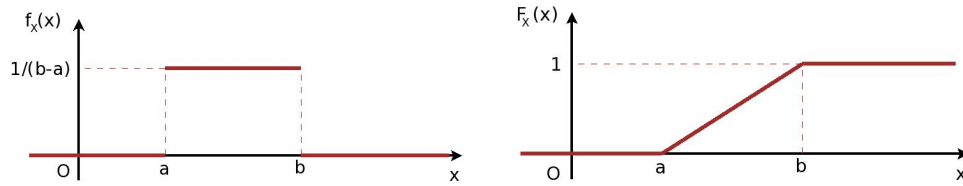


FIGURE 3.7 – Graphes de la densité et de la fonction de répartition de la loi uniforme sur un intervalle $[a, b]$.

tend vers 0 lorsque $\varepsilon \rightarrow 0$, ce qui confirme l'intuition " $\mathbb{P}(Z = x) = 0$ ". On obtient par contre une valeur non-nulle en prenant la limite pour ε tendant vers 0 de

$$\begin{aligned} \frac{\mathbb{P}(x \leq Z \leq x + \varepsilon)}{\text{longueur}([x, x + \varepsilon])} &= \frac{\mathbb{P}(x \leq Z \leq x + \varepsilon)}{\varepsilon} = \frac{\mathbb{P}(Z \leq x + \varepsilon, Z \geq x)}{\varepsilon} = \frac{\mathbb{P}((Z \leq x + \varepsilon) \cap (Z \geq x)^c)}{\varepsilon} \\ &= \frac{\mathbb{P}(Z \leq x + \varepsilon) - \mathbb{P}(Z \leq x)}{\varepsilon} \\ &= \frac{F(x + \varepsilon) - F(x)}{\varepsilon} \xrightarrow{\varepsilon \rightarrow 0} F'_Z(x) = f_Z(x) \end{aligned}$$

par définition de la dérivée. On peut donc interpréter la densité par

$$f(x) \sim \mathbb{P}(x \leq X \leq x + \varepsilon) \cdot \varepsilon$$

pour les petites valeurs de ε .

Exercice 2. Parmi les variables aléatoires suivantes, lesquelles sont des variables exponentielles ? géométriques ? Aucune des deux ? Pour ces dernières, essayez de représenter le graphe de leur densité au jugé.

- nombre de jours avant la prochaine pluie
- temps avant l'arrêt de la musique lors d'une partie de chaises musicales
- temps avant qu'un dé qu'on relance à chaque fois tombe sur 6
- temps avant que l'eau ne bouille
- temps d'attente entre 2 bus à un arrêt de bus donné
- résultat de la somme de deux dés

3.2.6 Autre approche de la densité

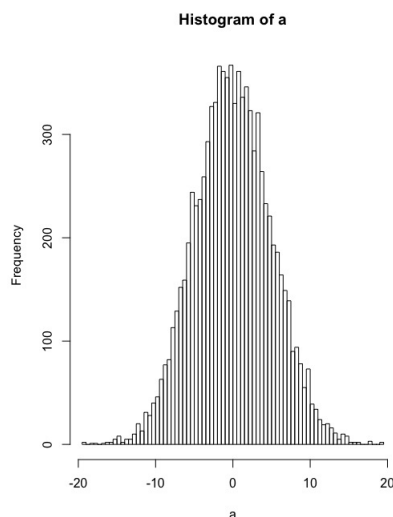
Revenons sur l'exemple des températures. Supposons qu'on ne sait pas quelle est la densité de la variable X , et on souhaite la connaître pour mieux prévoir l'évolution de la température. Le mieux est de faire des relevés chaque jour. On appelle X_1 la température le premier jour, X_2 le second jour, etc...Admettons qu'on le fasse pendant 3 mois, on obtient les valeurs :

```

[1]  5.893082936  0.476590629 -4.975856986  6.353986209 -2.819646778
[6] -2.600312296  5.322012089 -5.889846645  5.036474997  2.627002258
[11] -0.238853606  2.131592012 -5.787344516 -0.652443975 -2.773824848
[16]  4.512755084 -1.493351485 -5.653128866  2.222518088  1.624409237
[21]  0.670897402 -5.993308593  4.028839230 -1.571186934 -3.115128315
[26] -1.988530432 -0.658501570 -1.936193138 -4.772928122  3.201032186
[31]  0.088281595 -5.142267471 -6.714603068  1.085774250 -3.069822995
[36] -3.307266133 -6.997900383 -7.649775268  2.100272838  0.412602276
[41]  3.889824918 -4.572778827 -3.661292918 -5.327222859  1.251417963
[46]  0.131689785 -2.657205465  3.792682674 -5.737088094 -6.832552595
etc...

```

et on répartit les valeurs selon un histogramme :



On observe que la courbe de l’histogramme “ressemble” à la densité gaussienne. En effet, la probabilité que X tombe dans un intervalle $[\alpha, \beta]$ est censée être égale à l’intégrale de la densité sur cet intervalle :

$$\frac{\#\{i : X_i \in [\alpha, \beta]\}}{90} \sim \int_{\alpha}^{\beta} f_X(x) dx.$$

De même si l’on répète l’expérience du stylo n fois et qu’on classe les valeurs dans un histogramme :

```

[1] 1.129833677 2.748030738 0.150411771 0.970978099 0.572609356 1.210333499
[7] 2.076127195 1.020104539 2.287847331 2.986691481 0.839017433 0.986278694

```

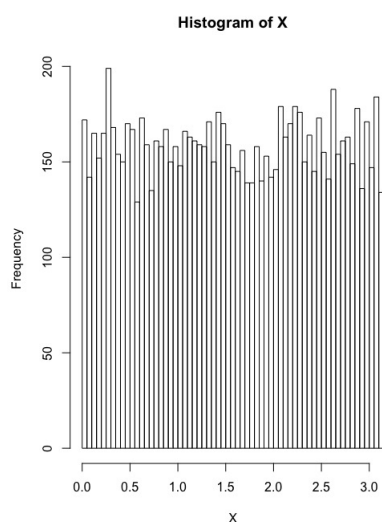


```

[13] 1.622985846 0.773610484 1.542120046 1.896705853 0.954943314 1.833749310
[19] 0.677580081 0.882258916 0.210909201 2.163946811 2.820300840 2.109249091
[25] 0.394468161 2.864241053 1.538403147 0.086120333 2.049142621 2.872915020
[31] 0.133887250 0.218906602 3.081498382 1.184213041 1.661214751 0.423515700
[37] 1.009790752 1.223642619 3.017891116 0.557445226 2.797671587 1.372617729
[43] 2.996023937 2.939949668 2.238954675 1.170315801 0.221124011 1.366222654
[49] 2.934229681 0.897024649 0.924580764 0.177186087 1.147346948 1.568169366
[55] 2.321640294 1.719135550 0.433922807 2.687428924 1.759458449 1.445761843
[61] 1.942275293 1.632271497 2.151737066 0.989490716 1.093898171 2.347600735
[67] 0.237344615 2.041322132 2.618540052 2.505556260 1.566771590 1.205401710
[73] 2.659631827 2.095410736 2.057543751 2.343286032 1.145112098 0.805917172
[79] 0.210801148 2.160158256 1.086770576 2.329035630 2.563283049 2.406078872
[85] 1.153179146 2.815855775 0.900861471 1.739431687 2.947073995 2.224538203
[91] 3.014341481 3.122373180 2.764589830 1.845570366 1.055203157 3.062053226
[97] 0.795077802 1.001609375 0.128736049 2.922929433 2.603936005 1.546293493
[103] 2.251352524 0.640966122 0.576557076 1.592871548 2.382338400 3.005989913
[109] 1.390574522 1.337582687 0.300019803 2.092564153 1.603383568 2.963757437
...
[985] 1.345273774 0.318922373 2.850842153 0.114303518 1.746244447 2.831837422
[991] 1.111246687 1.918784305 1.299183408 0.679878024 2.736802059 2.245113772
[997] 2.423832218 2.061692067 1.258624131 0.075497973

```

Sur un histogramme :



Généré avec les commandes

```

> X=runif(100,0,pi)
> hist(X,100)

```

3.3 Espérance et variance d'une variable aléatoire

3.3.1 Espérance

L'espérance d'une variable aléatoire X est la moyenne des valeurs prises par la variable, pondérées par leurs probabilités. On la note $\mathbb{E}(X)$.

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

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

lorsque cette somme est bien définie.

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 :

- $\mathbb{P}(X = 1) = \frac{18}{2 \cdot 18 + 1} = \frac{18}{37} < 1/2$.
- $\mathbb{P}(X = -1) = \frac{18+1}{2 \cdot 18 + 1} = \frac{19}{37} > 1/2$.

L'espérance de gain est donc

$$\mathbb{E}(X) = 1 \times \frac{18}{37} + (-1) \times \frac{19}{37} = -\frac{1}{37}.$$

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 (donc beaucoup d'argent si n est grand !). Le casino, en revanche, gagnera $n/37$.

Remarque Si X n'est pas supposée positive, on a affaire à une somme infinie de termes qui peuvent être positifs et négatifs, et il faut faire appel à la théorie des séries pour savoir si cette somme est réellement calculable. Cette somme est bien définie uniquement si la série des termes est **absolument convergente**, c'est-à-dire si

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

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

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

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

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

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

On a avec une intégration par parties

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

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.

- Si $a \in \mathbf{R}$, $\mathbb{E}(a) = a$ puisque a est non aléatoire : autrement dit, a s'interprète comme une variable aléatoire prenant la valeur a avec probabilité 1, et les autres valeurs avec la probabilité 0.
- L'espérance est une fonction linéaire : si a, b sont des réels, et X, Y des variables aléatoires,

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y), \mathbb{E}(aX) = a\mathbb{E}(X),$$

$$\mathbb{E}(X - Y) = \mathbb{E}(X) + \mathbb{E}(-Y) = \mathbb{E}(X) - \mathbb{E}(Y), \dots$$

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).

Exemple 11. 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 2 euros par minute de retard d'Aurore".

Aurore a en moyenne 2 minutes de retard à ses rendez-vous. Quelle est l'espérance de gain de Julien ?

Appelons R le retard d'Aurore. Le gain de Julien est donc

$$G = -3 + 2R.$$

On a donc

$$\mathbb{E}(G) = \mathbb{E}(-3 + 2R) = \mathbb{E}(-3) + \mathbb{E}(2R) = -3 + 2\mathbb{E}(R) = -3 + 4 = 1.$$

Julien a donc une espérance de gain positive égale à 1 euro.

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

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

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

Par exemple

$$\mathbb{E}(X^2) = \sum_{x \in \mathcal{S}(X)} x^2 \mathbb{P}(X = x) \quad (\text{cas discret})$$

$$\mathbb{E}(X^2) = \int_{\mathbf{R}} x^2 f_X(x) dx \quad (\text{cas continu}).$$

Ce calcul servira pour calculer la variance. Ici encore ces sommes ou ces intégrales peuvent ne pas être absolument convergentes.

3.3.2 Variance

Connaître la moyenne d'une variable aléatoire n'est qu'une information très partielle. Par exemple, supposons que je doive choisir entre deux bus A et B qui ont tous les deux des temps d'attente moyens de 5 min. Il se peut très bien que l'on soit dans la situation suivante :

- Le bus A a une chance sur 5 d'arriver en 15 min, et 4 chances sur 5 d'arriver en 2min30s.
- Le bus B a une chance sur 2 d'arriver en 4 min, et une chance sur 2 d'arriver en 6 min.

$$\mathbb{E}(A) = 2,5 \cdot \mathbb{P}(A = 2,5) + 15 \mathbb{P}(A = 15) = 2,5 \frac{4}{5} + 15 \frac{1}{5} = 5$$

$$\mathbb{E}(B) = 4 \mathbb{P}(A = 4) + 6 \mathbb{P}(A = 6) = \frac{4}{2} + \frac{6}{2} = 5.$$

Ces deux bus ont bien un temps d'attente moyen de 5 minutes, mais le premier est plus imprévisible. On dit qu'il a plus de "variabilité", dû aux écarts possibles entre les temps d'attente probables, ou que sa "variance" est plus grande.

La variance permet de mesurer l'écart des valeurs d'une variable aléatoire par rapport à son espérance :

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

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

Montrons que cette dernière égalité est vraie : on a

$$\mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 + \mathbb{E}(X)^2 - 2X\mathbb{E}(X)) = \mathbb{E}(X^2) + \mathbb{E}(\mathbb{E}(X)) - 2\mathbb{E}(X\mathbb{E}(X)),$$

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

$$\mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) + \mathbb{E}(X) - 2\mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

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

Retour sur l'exemple des bus :

$$\begin{aligned} V(A) &= \mathbb{E}(A^2) - (\mathbb{E}(A))^2 \\ &= (2,5)^2 \mathbb{P}(A = 2,5) + 15^2 \mathbb{P}(A = 15) - 5^2 = 6,25 \cdot \frac{4}{5} + 225 \cdot \frac{1}{5} = 50 - 25 = 25 \\ V(B) &= 4^2 \mathbb{P}(B = 4) + 6^2 \mathbb{P}(B = 6) - 5^2 = 1. \end{aligned}$$

En effet, $V(B) < V(A)$.

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

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

C'est une interprétation de la valeur de laquelle on va dévier de la moyenne, "en moyenne". En d'autre termes, on a $X = \mathbb{E}(X) \pm \sigma(X)$.

Pour l'exemple des bus : $\sigma(A) = \sqrt{25} = 5$, et $\sigma(B) = \sqrt{1} = 1$. Donc $A = 5 \pm 5$ en moyenne, et $B = 5 \pm 1$.

Exercice 3. On lance une fois un dé non pipé.

1. On suppose qu'on reçoit 15 euros si on obtient 1, rien si on obtient 2,3 ou 4, et 6 euros si on obtient 5 ou 6. Soit G la variable aléatoire égale au gain de ce jeu. Que vaut le gain moyen ? Quelle est la variance ?
2. Mêmes questions en supposant qu'on gagne 27 euros pour un 1 et rien sinon. Préférez vous jouer au jeu du 1) ou à celui-ci ?

1. La loi est $\mathbb{P}(G = 15) = 1/6, \mathbb{P}(G = 0) = 1/2, \mathbb{P}(G = 6) = 1/3$.

$$\mathbb{E}(G) = \frac{15}{6} + 0 + \frac{6}{3} = \frac{27}{6}.$$

Pour calculer la variance on a besoin de

$$\mathbb{E}(G^2) = \sum_{x \in \mathcal{S}(X)} x^2 \mathbb{P}(X = x) = \frac{(15)^2}{6} + \frac{36}{3} = \frac{369}{6} = 61,5.$$

La variance est donc

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 61,5 - \left(\frac{27}{6}\right)^2 = 29,25,$$

donc l'écart-type vaut $\sigma(G) = \sqrt{29,25} \approx 5,4$.

2. On a $\mathbb{P}(G = 27) = 1/6, \mathbb{P}(G = 0) = 5/6$. On a

$$\mathbb{E}(G) = \frac{27}{6} + 0.$$

La variance est

$$V(G) = \mathbb{E}(G^2) - \left(\frac{27}{6}\right)^2 = \frac{(27)^2}{6} - \left(\frac{27}{6}\right)^2 = 101,25.$$

On a cette fois $\sigma(X) \approx 10$, le gain a plus de “variabilité” dans le second cas.

3.3.3 Exemples de calculs

Espérance et variance d'une Bernoulli

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

$$\mathbb{E}(X) = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = p.$$

et

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 0^2 \cdot \mathbb{P}(X = 0) + 1^2 \cdot \mathbb{P}(X = 1) - p^2 = p - p^2 = p(1 - p).$$

Espérance et variance d'une variable géométrique

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

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

Calculons $\mathbb{E}(X)$:

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

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

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

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

Calculons $V(X)$:

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

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

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

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 $V(X)$:

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

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

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

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

Variance d'une variable de loi Gaussienne

Soit $X \sim \mathcal{N}(m, \sigma^2)$, où $m \in \mathbb{R}$ et $\sigma > 0$. Alors on admet les faits suivants :

$$\begin{aligned} \mathbb{E}(X) &= m \\ V(X) &= \sigma^2. \end{aligned}$$

$\sigma = \sqrt{V(X)} = \sigma(X)$ est donc l'écart-type de X .

3.3.4 Moments

Soit $k \in \mathbb{N}$, et X une variable aléatoire. On dit que X admet un moment d'ordre k si $\mathbb{E}[|X|^k] < \infty$.

Donc X admet une espérance veut dire que X admet un moment d'ordre 1.

Proposition. *Soit $k \leq q$. Alors si X admet un moment d'ordre q , elle admet aussi un moment d'ordre k . La réciproque est fausse en général.*

Donc si X admet un moment d'ordre 2, elle admet aussi un moment d'ordre 1. C'est équivalent au fait d'admettre une variance.

Chapitre 4

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 .

4.1 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, $\forall I, J \subset \mathbf{R}$,

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

On note parfois $X \perp Y$. (" 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.

Exemple 12. Soit X, Y les résultats des lancers de deux dés. Soit Z la variable aléatoire produit : $Z = XY$. En énumérant toutes les possibilités, on a

$$\mathcal{S}(Z) = \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 25, 30, 36\}.$$

Pour calculer la loi de Z on utilise l'indépendance. Par exemple

$$\begin{aligned}\mathbb{P}(Z = 4) &= \sum_{x \in \mathcal{S}(X)} \sum_{y \in \mathcal{S}(Y): xy=4} \mathbb{P}(X = x, Y = y) \\ &= \sum_{x \in \{1, \dots, 6\}} \sum_{y \in \{1, \dots, 6\}: xy=4} \mathbb{P}(X = x) \mathbb{P}(Y = y) \\ &= \mathbb{P}(X = 1) \mathbb{P}(Y = 4) + \mathbb{P}(X = 2) \mathbb{P}(Y = 2) + \mathbb{P}(Y = 4) \mathbb{P}(X = 2) = \frac{3}{36} = \frac{1}{12}.\end{aligned}$$

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

$$\mathbb{P}(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = \mathbb{P}(X_1 \in I_1) \mathbb{P}(X_2 \in I_2) \cdots \mathbb{P}(X_n \in I_n).$$

Exemple 13. Soit X, Y les résultats de deux lancers de dés, et Z la variable de Bernoulli définie par

$$Z = \begin{cases} 1 & \text{si } X \text{ et } Y \text{ ont même parité} \\ 0 & \text{sinon.} \end{cases}$$

Alors $(X, Y), (X, Z), (Y, Z)$ sont des couples de variables aléatoires indépendantes, mais (X, Y, Z) n'est pas un triplet de variables aléatoires indépendantes. En effet,

$$\mathbb{P}(X = 1, Y = 1, Z = 1) \neq \mathbb{P}(X = 1) \mathbb{P}(Y = 1) \mathbb{P}(Z = 1).$$

4.1.1 Quelques propriétés

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

Preuve. Pour des variables discrètes positives,

$$\mathbb{E}(XY) = \sum_{z \in \mathcal{S}(XY)} z \mathbb{P}(XY = z)$$

Pour calculer $\mathbb{P}(XY = z)$, on fait comme dans l'exemple où $Z = XY$ est le produit du lancer de deux dés :

$$\mathbb{P}(Z = z) = \sum_{x \in \mathcal{S}(X)} \sum_{y \in \mathcal{S}(Y): xy=z} \mathbb{P}(X = x, Y = y) = \sum_{x \in \mathcal{S}(X)} \sum_{y \in \mathcal{S}(Y)} \mathbf{1}_{\{z=xy\}} \mathbb{P}(X = x) \mathbb{P}(Y = y).$$

Donc

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

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 sont positives.

On admet le résultat pour des variables à densité. \square

Proposition. — Soit X, Y deux variables indépendantes qui admettent une variance (ou un moment d'ordre 2). Alors

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

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

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

— Une variable aléatoire telle que $V(X) = 0$ est dite déterministe, c'est-à-dire qu'il existe un nombre réel $x \in \mathbb{R}$ tel que $\mathbb{P}(X = x) = 1$. En effet, si $\mathbb{P}(|X - \mathbb{E}(X)| > a) > 0$ pour un $a > 0$, alors $\mathbb{E}(X - \mathbb{E}(X))^2 \geq a^2 \mathbb{P}(|X - \mathbb{E}(X)| \geq a) > 0$, ce qui contredit l'hypothèse.

C'est faux si les variables ne sont pas indépendantes. Par exemple si $Y = -X$,

$$V(X + Y) = V(X - X) = V(0) = 0.$$

Preuve.

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

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

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

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

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

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

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

□

4.2 Exemples de calculs de lois utilisant l'indépendance

Somme de binomiales indépendantes

Soit $n \in \mathbb{N}, p \in [0, 1]$. Soit X le nombre de succès lors de n expériences de Bernoulli de paramètre p . On sait que $X \sim \mathcal{B}(n, p)$. Soit maintenant $m \in \mathbb{N}$, et $Y \sim \mathcal{B}(m, p)$ indépendante de X . Alors $X + Y$ est le nombre de succès lors de $m + n$ expériences de Bernoulli indépendantes. Donc $X + Y \sim \mathcal{B}(n + m, p)$.

Exercice 4. Montrer le même résultat par le calcul (c'est compliqué!).

Somme de Poisson indépendantes

Soit $\lambda, \mu > 0$. Soit $X \sim \mathcal{P}(\lambda)$, et $Y \sim \mathcal{P}(\mu)$ indépendante de X . On sait que X représente le nombre d'événements aléatoires qui arrivent sur une période λ , où chaque événement est indépendant des autres et il y a en moyenne 1 événements par unité de temps. Y est le nombre d'événements qui arrivent sur une période de temps μ . Donc $X + Y$ est le nombre d'événements sur une durée $\lambda + \mu$. Donc $X + Y \sim \mathcal{P}(\lambda + \mu)$.

Somme de Gaussiennes indépendantes

Soit $X \sim \mathcal{N}(m, \sigma^2)$ et $Y \sim \mathcal{N}(m', s^2)$ deux variables Gaussiennes indépendantes. Alors $X + Y$ est encore une loi Gaussienne (dur à démontrer). On a

$$\begin{aligned}\mathbb{E}(X + Y) &= \mathbb{E}(X) + \mathbb{E}(Y) = m + m' \\ V(X + Y) &= V(X) + V(Y) = \sigma^2 + s^2 \text{ car } X \text{ et } Y \text{ sont indépendantes.}\end{aligned}$$

Ainsi, $X + Y \sim \mathcal{N}(m + m', \sigma^2 + s^2)$.

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 $1 \leq i \leq n$

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

Les Y_i sont en fait des variables de Bernoulli de paramètre p . On rappelle que $\mathbb{E}(Y_i) = p$, $V(Y_i) = p(1 - p)$. 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

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{E}(Y_i) = n\mathbb{E}(Y_1) = np.$$

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

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

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.

4.2.1 Maximum ou minimum de variables indépendantes

Théorème 1. Soit X_1, X_2 2 variables aléatoires indépendantes, et $X = \min(X_1, X_2)$. Alors leurs fonctions de répartition vérifient, pour $x \in \mathbb{R}$,

$$F_X(x) = F_{X_1}(x)F_{X_2}(x).$$

Preuve.

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X_1 \leq x, X_2 \leq x) = \mathbb{P}(X_1 \leq x)\mathbb{P}(X_2 \leq x) = F_{X_1}(x)F_{X_2}(x).$$

□

Exemple 14. On lance deux dés plusieurs fois de suite. Dès qu'un dé fait 6, on le laisse sur le côté et on lance l'autre jusqu'à ce qu'il fasse 6 également. Soit X le nombre de lancers nécessaires. Que vaut $F_X(x)$, pour $x \in \mathbb{R}$?

Soit X_i le nombre de lancers nécessaires au i -ème dé pour faire 6. Alors $X = \max(X_1, X_2)$.
Donc

$$F_X(x) = F_{X_1}(x)F_{X_2}(x).$$

Comme X_1 et X_2 ont la même loi (géométrique de paramètre $1/6$), pour $k \in \mathbb{N}$, $F_{X_1}(k) = F_{X_2}(k) = 1 - (5/6)^k$. Donc

$$F_X(k) = (1 - 5/6)^{2k}.$$

Exercice 5. Si $X = \min(X_1, X_2)$, alors on a la relation

$$(1 - F_X(x)) = (1 - F_{X_1}(x))(1 - F_{X_2}(x)).$$

4.3 Covariance et corrélation

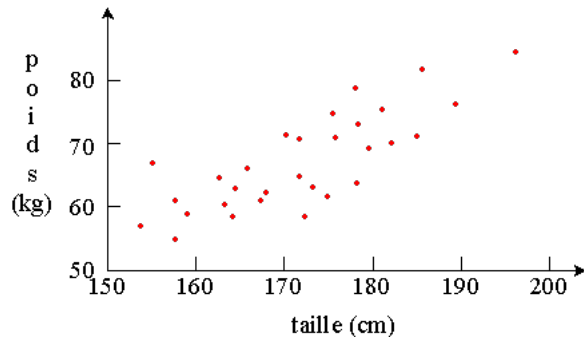
Important : Deux variables aléatoires ne sont pas toujours indépendantes !

Etant donné deux variables aléatoires X et Y , il peut être intéressant de savoir si X et Y sont effectivement indépendantes. Par exemple :

1. X est la quantité de gaz à effet de serre émise une année (en tonnes), et Y est le réchauffement de l'atmosphère (en degrés)
2. X est le nombre d'années d'études d'un français tiré au hasard, et Y est son salaire
3. X est le sexe d'un français tiré au hasard ("homme" ou "femme"), et Y est son salaire
4. X est le choix d'une publicité par un annonceur en ligne ("pub A" ou "pub B"), et Y est le trafic généré par cette publicité.
5. X est le fait qu'une personne tirée au hasard dans le monde parle français ("oui" ou "non"), et Y est la quantité de camembert qu'il mange chaque année (en grammes).
6. X est le nombre d'années d'exposition d'une personne à un téléphone portable, Y est le fait qu'il ait contracté un cancer dans sa vie ("oui" ou "non")

Répondre à ces questions peut être un enjeu crucial pour de nombreux problèmes sociétaux, écologiques, ou économiques.

Exemple 15. On considère un nuage de points qui représentent 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

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

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

Dans le langage courant, on dit souvent corrélées au lieu de positivement corrélées. Cela signifie que X a plus tendance à être grande (au-dessus de sa moyenne) quand Y est grande. On dit que X et Y sont **négativement corrélées** si $\text{Cov}(X, Y) < 0$. Cela signifie au contraire que X a tendance à être petite si Y est grande, et vice-versa, et ça revient aussi à dire que X et $(-Y)$ sont positivement corrélées.

Remarque 2. Pour toute variable aléatoire X , $V(X) = \text{Cov}(X, X)$.

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 $\text{Cov}(X, Y) = 0$.

Preuve. On a vu que comme X, Y sont indépendantes, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Donc $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$. \square

Attention : La réciproque de la proposition est fausse en général : $\text{cov}(X, Y) = 0$ 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 X son salaire annuel et Y son nombre d'années d'études, on espère que $\text{Cov}(X, Y) > 0$...
- On pose $X = 1$ si cette personne est une femme, et 0 sinon. Soit Y la taille de cette personne arrondie au cm (pour avoir une VA discrète). On suppose que la taille

moyenne des français est de 1m65. On observe que la taille moyenne des femmes est 1m60. Qu'en déduisez vous sur la covariance ?

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

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 :

$$\begin{aligned}
 X &= \begin{cases} 1 & \text{si cette personne parle français} \\ 0 & \text{sinon} \end{cases} \\
 Y &= \begin{cases} 1 & \text{si cette personne mange du camembert} \\ 0 & \text{sinon} \end{cases}
 \end{aligned}$$

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

Proposition. *La covariance est symétrique et linéaire en chacune de ces coordonnées (ou bilinéaire symétrique) :*

— Soit X, Y des variables aléatoires. Alors

$$\text{Cov}(X, Y) = \text{Cov}(Y, X).$$

— Soit X, Y, Z des variables aléatoires, $a, b \in \mathbb{R}$:

$$\text{Cov}(aX + bY, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z)$$

Preuve.

- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(YX) - \mathbb{E}(Y)\mathbb{E}(X)$.
- Pour $a = b = 1$:

$$\begin{aligned}\text{Cov}(X + Y, Z) &= \mathbb{E}((X + Y)Z) - \mathbb{E}(X + Y)\mathbb{E}(Z) = \mathbb{E}(XZ + YZ) - (\mathbb{E}(X) + \mathbb{E}(Y))\mathbb{E}(Z) \\ &= \mathbb{E}(XZ) + \mathbb{E}(YZ) - \mathbb{E}(X)\mathbb{E}(Z) - \mathbb{E}(Y)\mathbb{E}(Z) \\ &= \underbrace{(\mathbb{E}(XZ) - \mathbb{E}(X)\mathbb{E}(Z))}_{\text{Cov}(X, Z)} + \underbrace{(\mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z))}_{\text{Cov}(Y, Z)}\end{aligned}$$

Pour $Y = 0$,

$$\text{Cov}(aX, Z) = \mathbb{E}(aXZ) - \mathbb{E}(aX)\mathbb{E}(Z) = a\mathbb{E}(XZ) - a\mathbb{E}(X)\mathbb{E}(Z) = a\text{Cov}(X, Z).$$

Cas général :

$$\text{Cov}(aX + bY, Z) = \text{Cov}(aX, Z) + \text{Cov}(bY, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z).$$

□

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 ,

$$\begin{aligned}\text{Cov}(X', Y) &= \mathbb{E}(X'Y) - \mathbb{E}(X')\mathbb{E}(Y) = \mathbb{E}(100 \cdot X \cdot Y) - \mathbb{E}(100 \cdot X)\mathbb{E}(Y) \\ &= 100\mathbb{E}(XY) - 100\mathbb{E}(X)\mathbb{E}(Y) = 100\text{Cov}(X, Y).\end{aligned}$$

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

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}.$$

Cette quantité n'est pas sensible à un changement d'unité de mesure :

$$\text{Corr}(X', Y) = \frac{\text{Cov}(100X, Y)}{\sqrt{V(100X)}\sqrt{V(Y)}} = \frac{100\text{Cov}(X, Y)}{\sqrt{100^2 V(X)}\sqrt{V(Y)}} = \frac{100\text{Cov}(X, Y)}{100\sqrt{V(X)}\sqrt{V(Y)}} = \text{Corr}(X, Y).$$

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

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

Remarque 3. Pour toute variable aléatoire X ,

$$\text{Corr}(X, X) = \frac{\text{Cov}(X, X)}{\sqrt{V(X)}\sqrt{V(X)}} = \frac{V(X)}{V(X)} = 1.$$

Une variable aléatoire est complètement corrélée avec elle-même. $\text{Corr}(X, X) = 1$. De plus, si $\text{Corr}(X, Y) = 1$, cela veut dire qu'il existe un nombre $a > 0$ tel que $\mathbb{P}(X = aY) = 1$, autrement dit X et Y sont proportionnelles (avec probabilité 1). Si $\text{Corr}(X, Y) = -1$, alors $X = aY$ pour un $a < 0$.

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

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

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

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

Ainsi $f(t) = at^2 + bt + c$, avec $a = \mathbb{E}(Y^2)$, $b = 2\mathbb{E}(XY)$, et $c = \mathbb{E}(X^2)$. Or $f(t)$ est toujours positif, puisque $f(t) = \mathbb{E}((X + tY)^2)$ (espérance d'une variable positive). Donc le polynôme $at^2 + bt + c$ 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 remplaçant a, b, c par leurs valeurs, on obtient $\Delta = 4\mathbb{E}(XY)^2 - 4\mathbb{E}(Y^2)\mathbb{E}(X^2) \leq 0$, et donc $\mathbb{E}(XY)^2 \leq \mathbb{E}(Y^2)\mathbb{E}(X^2)$, soit encore $\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$. \square

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

Exemple 16. Soit X la taille en mètres, X' la taille en cm, P le poids, Y la longueur du nez, et Z la pointure des chaussures. En relevant ces quantités sur un vaste échantillon de française on calcule les corrélations suivantes :

- $\text{Corr}(X, P) = 0,7$. La taille et le poids sont corrélés
- $\text{Corr}(X, Y) \approx 0$. La longueur du nez n'est pas corrélée à la taille.
- $\text{Corr}(X, Z) = 0,3$. La longueur des pieds est faiblement corrélée à la taille.
- $X, X' = 1$. La taille en mètres est proportionnelle à la taille en cm.

interprétation visuelle On peut visualiser la dépendance de deux variables aléatoires en observant la répartition spatiale de nuages de points de couples de variables tirées suivant ces lois.

Exemple 17. Une corrélation nulle ne veut pas forcément dire des variables indépendantes. Considérons l'exemple où

- X : Age d'un individu tiré au hasard dans la population

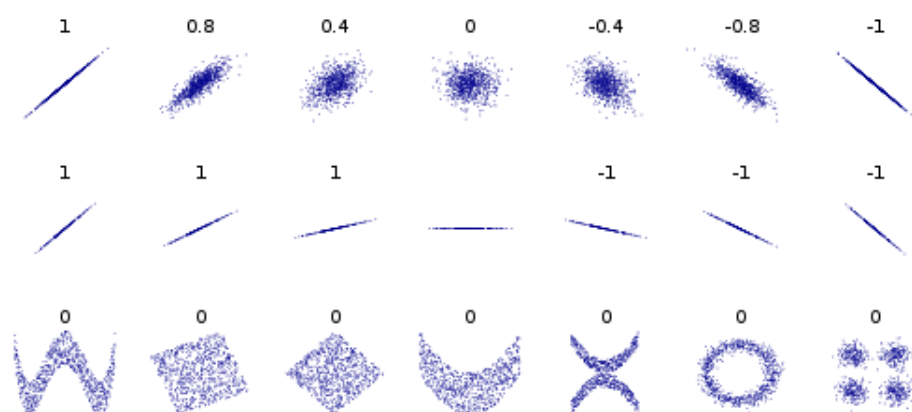


FIGURE 4.1 – Les deux rangées du haut montre des exemples de couples où le coefficient de corrélation indique bien la dépendance ou l’indépendance. La rangée du bas montre des couples où les variables aléatoires sont dépendantes mais ont tout de même une corrélation nulle.

— Y : Son temps au 100 mètres.

On peut imaginer que le nuage de répartition des points aura une forme de “croissant” comme sur la rangée du bas de la figure précédente. Les deux extrémités du croissant représentent l’enfance et la vieillesse. Il est donc possible d’avoir une corrélation égale à 0, alors que X et Y ne sont pas du tout indépendantes (une personne de 25 ans a plus de chance de faire un meilleur temps qu’une personne de 80 ans).

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, $\text{Corr}(X, Y) \sim 0,8$.

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.

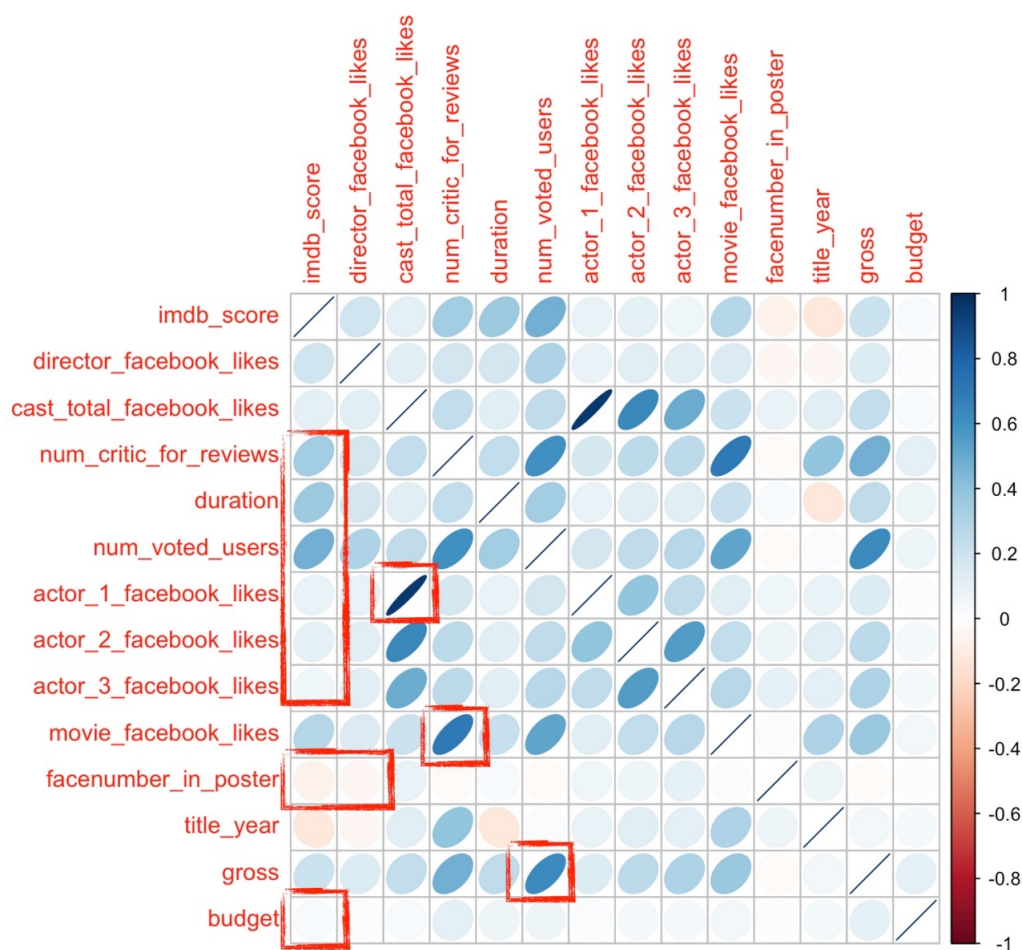


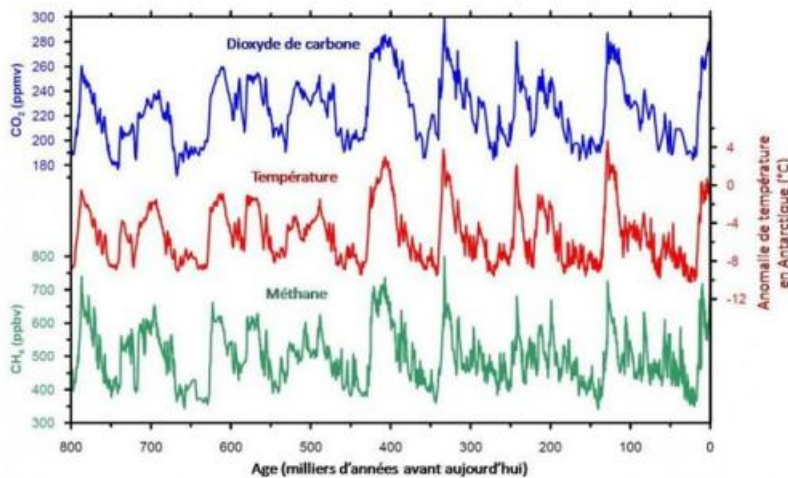
FIGURE 4.2 – Etude de la corrélation pour diverses variables liées à la distributions de films. <https://blog.nydatascience.com/student-works/machine-learning/movie-rating-prediction/> Ici les ellipses ne sont pas des vrais nuages de points mais juste des manières schématiques de représenter la corrélation.

4.3.1 Conditionnement par rapport à un événement

Définition. Soit X une variable et A un événement tel que $\mathbb{P}(A) \neq 0$. 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 $\mathbb{P}(X = x|A)$ 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 $\mathbb{E}(X|A)$ Dans le cas où X est une variable discrète, on a

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



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

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

Exercice 6. (cf. TD4, Exo 9) Un pêcheur pêche dans un lac rempli de poissons (plusieurs milliers...). On fait les hypothèses suivantes :

- Il attrappe en moyenne 2 poissons par heure
- 60% des poissons sont des carpes
- Les carpes sont aussi faciles à attraper que les autres poissons
- Si le pêcheur attrappe un poisson à un instant donné, ça n'a aucune influence sur ses chances d'attraper des poissons par la suite.

1. Quelle est la loi du nombre X total de poissons attrappés sur une heure ?
2. Quelle est la loi du nombre Y de carpes attrappées sur une heure ?
3. On pose l'évènement $A = "X = n"$, où $n \in \mathbb{N}^*$. Quelle est la loi conditionnelle de Y sachant A ? Que vaut $\mathbb{E}(Y|A)$?
4. Utiliser (en justifiant) la formule

$$\mathbb{E}(XY) = \sum_{n=0}^{\infty} \mathbb{E}(XY|X = n)\mathbb{P}(X = n)$$

pour calculer $\text{Cov}(X, Y)$ et $\text{Corr}(X, Y)$.

Chapitre 5

Théorèmes limites et estimation

5.1 Jeu de données et échantillon

Un jeu de données est un ensemble de chiffres recueillis lors d'un phénomène aléatoire :

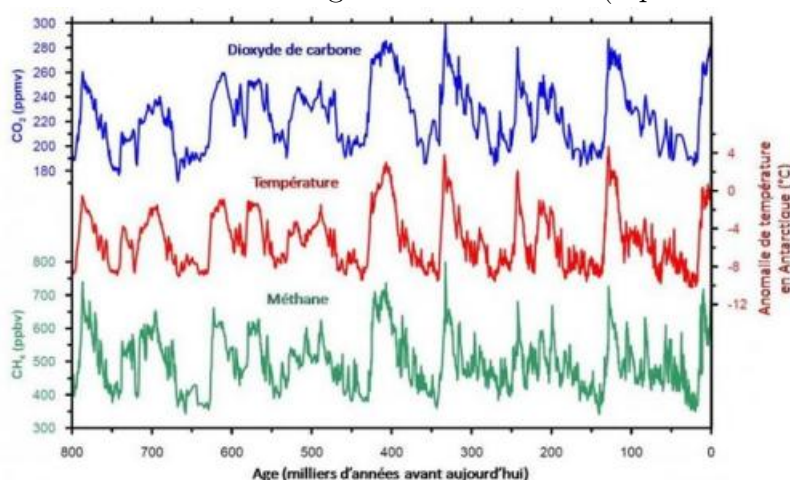
- Intervalles entre les temps d'arrivées de bus à un arrêt donné, en minutes :

5,6	3,2	4	1,3	8	4,5	7,2	3
-----	-----	---	-----	---	-----	-----	---

- Résultats des votes à une élection :

Nombre de votants pour le candidat A	12 453 122
Nombre de votants pour le candidat B	5 233 021 0
Nombre de votants pour le candidat C	14 669 821

- Température et concentration en gaz à effet de serre (représenté ici sous forme de



graphe) :

Un jeu de donnée est donc un ensemble de chiffres que l'on note de manière abstraite $\{x_1, \dots, x_n\}$, où n est un entier à priori grand.

5.2 Modélisation

Lorsque l'on étudie un jeu de données, il faut se fixer un modèle, à l'aide de variables aléatoires.

Exemple Les intervalles de temps entre les arrivées de bus sont des variables aléatoires indépendantes et de même loi. On note alors

- X_1 : temps entre le 1er et le second bus
- X_2 : temps entre le 2d et le troisième bus
- ...
- X_n : temps entre le n -ème et le $n + 1$ -ème bus

On dit que les X_i sont “iid” (indépendantes et identiquement distribuées), ou que c'est un “échantillon”.

Lorsque l'on étudie un jeu de données $\{x_1, \dots, x_n\}$, on cherche à connaître une information sur le phénomène aléatoire qui a généré ce jeu de données. On part donc d'un “modèle” $\{X_1, \dots, X_n\}$, c'est-à-dire un objet mathématique censé reproduire la réalité le plus fidèlement possible, en supposant que les x_i sont les valeurs qu'ont prises les X_i : $X_i = x_i$. La réalité est en général plus compliquée que le modèle. On doit en général On note μ l'espérance de X_1 (ou X_2 ou n'importe quel autre X_i vu qu'ils ont tous la même loi, et donc la même moyenne). On note $\sigma^2 = V(X_1) = V(X_2) = \dots$

Exemple : On observe un bus pendant plusieurs journées. Comment estimer le temps d'attente moyen ?

On appelle μ le temps moyen d'attente. La manière la plus naturelle d'estimer μ à partir des x_i est de faire la moyenne :

$$\mu \sim \frac{1}{n}(x_1 + \dots + x_n)$$

Comment exprimer ça rigoureusement ? Que dire du “ \sim ” ?

Définition. Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires. On appelle **moyenne empirique** (des n premiers termes de la suite) la variable aléatoire

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Une approximation \sim se traduit mathématiquement par une convergence : On voudrait donc que pour une suite de variables X_i iid,

$$\underbrace{\frac{1}{n}(X_1 + \dots + X_n)}_{\overline{X}_n} \rightarrow \mu$$

Remarque 4. Y peut être un nombre fixé (on dit aussi “déterministe”) dans la définition précédente, car un nombre est une variable aléatoire particulière qui prend toujours la même valeur.

5.2.1 La loi des grands nombres

Evènement “presque sûr”

Définition. Un évènement “presque sûr” est un évènement dont la probabilité est 1.

Les exemples triviaux d'évènements presque sûrs sont les évènements qui sont toujours réalisés. Par exemple l'évènement “Je tire un dé et je fais un nombre entre 1 et 6”.

Si il n'y avait que ceux-là cette notion ne serait pas très intéressante. Il y a en fait des situations plus compliquées. Un exemple que l'on a déjà vu est celui des variables continues : Soit X une loi exponentielle de paramètre 1 et A l'évènement $X \neq 1$. L'évènement A^c est théoriquement possible, mais il se trouve que sa probabilité est nulle :

$$\mathbb{P}(A) = 1 - \mathbb{P}(X = 1) = 1 - 0 = 1.$$

Définition. Soit $Y_n, n \geq 1$ Une suite de variable aléatoire, et Y une autre variable aléatoire. On dit que $(Y_n)_{n \geq 1}$ converge **presque sûrement** vers Y si l'évènement $A = \text{“la quantité } Y_n \text{ converge vers } Y \text{ lorsque } n \rightarrow \infty \text{”}$ a probabilité 1 :

$$\mathbb{P}(Y_n \rightarrow Y \text{ quand } n \rightarrow \infty) = 1.$$

On a les outils pour conclure sur la convergence de la moyenne dont on parlait avant :

Théorème 2. Soit $(X_n)_{n \geq 0}$ une suite de variables indépendantes et toutes de même loi. On suppose que l'espérance $\mu = \mathbb{E}(X_1)$ est bien définie, c'est-à-dire $\mathbb{E}(|X_1|) < \infty$. Alors $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$ converge vers μ “presque sûrement”.

La preuve de ce théorème est difficile ; il n'est pas question de la faire dans ce cours.

On dit “presque sûre” car il est possible que la suite ne converge pas. Si par exemple on est dans l'évènement $B = “X_1 = X_2 = X_3 = \dots = 0”$, alors $A = “\overline{X_n} \rightarrow \mathbb{E}(X)”$ n'est en général pas vérifié ($A \cap B = \emptyset$). Mais comme $\mathbb{P}(B) = 0$, cela n'empêche pas $\mathbb{P}(A) = 1$ (sauf dans le cas particulier où $\mathcal{S}(X) = \{0\}$).

Une manière d'appréhender ce théorème est le résultat suivant :

Proposition. Soit $(X_n)_{n \geq 0}$ une suite de variables indépendantes et toutes de même loi. On suppose que l'espérance et la variance communes à tous les X_n : $\mu = \mathbb{E}(X_1)$, $\sigma^2 = V(X_1)$, sont bien définies (c'est le cas si $\mathbb{E}(X_1^2) < \infty$). Alors pour tout $n \geq 1$,

$$\mathbb{E}(\overline{X_n}) = \mu \quad \text{et} \quad V(\overline{X_n}) = \frac{\sigma^2}{n}.$$

Preuve.

- $\mathbb{E}(\overline{X_n}) = \mathbb{E}(\frac{1}{n} \sum_{i=1}^n X_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \times n\mu$ puisque tous les X_i ont même espérance. Donc $\mathbb{E}(\overline{X_n}) = \mu$.

- $V(\overline{X_n}) = V(\frac{1}{n} \sum_{i=1}^n X_i) = \frac{1}{n^2} V(\sum_{i=1}^n X_i) = \frac{1}{n^2} \sum_{i=1}^n V(X_i)$ car les X_i sont des variables indépendantes. Donc $V(\overline{X_n}) = \frac{1}{n^2} \times n\sigma^2 = \frac{\sigma^2}{n}$ puisque tous les X_i ont même variance.

□

Remarque 5. Une VA X qui a une variance égale à 0 n'est en fait pas aléatoire :

$$V(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = 0 = \begin{cases} \sum_{x \in \mathcal{S}(X)} (x - \mathbb{E}(X))^2 \mathbb{P}(X = x) & \text{si } X \text{ est discrète} \\ \int_{\mathbb{R}} (x - \mathbb{E}(X))^2 f_X(x) dx & \text{si } X \text{ est à densité} \end{cases}$$

donc $\mathbb{P}(X \neq \mathbb{E}(X)) = 0$. On dit que X est “déterministe.”

Ce résultat est très important car il montre que lorsque n tend vers $+\infty$, la variance de $\overline{X_n}$ tend vers 0, ce qui laisse à penser que $\overline{X_n}$ converge vers une variable “déterministe” (c'est-à-dire un nombre), égal à son espérance $\mathbb{E}(\overline{X_n}) = \mu$. C'est précisément la loi des grands nombres.

Voici quelques exemples d'application de la loi des grands nombres :

Exemple 18. : On considère un dé truqué telle que $\mathbb{P}(X = 6) = p > 1/6$, et toutes les autres probabilités sont égales : $\mathbb{P}(X = 1) = \dots = \mathbb{P}(X = 5) = (1 - p)/5$. Si l'on fait n lancers X_i iid et que l'on observe la moyenne

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i,$$

alors $\overline{X_n} \rightarrow \mathbb{E}(X)$ presque sûrement, et $\mathbb{E}(X) = 6p + (1 + 2 + 3 + 4 + 5)(1 - p)/5 = 6p + 3(1 - p) = 3(p + 1)$.

Exemple 19. Sur un sondage de 1000 électeurs, 521 prétendent voter pour le candidat A d'une élection. Chaque électeur est “modélisé” par une variable de Bernoulli de paramètre p , où p est le paramètre inconnu, à savoir la proportion de français qui vont voter A à l'élection. On suppose que les choix des personnes sondées sont indépendants, et les X_i sont donc iid. D'après la LGN,

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mathbb{E}(X_1) \text{ p.s.}$$

Comme $X_1 \sim \mathcal{B}(p)$, $\mathbb{E}(X_1) = p$. Notre jeu de données nous donne

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{1000} \#\{i : X_i = 1\} = \frac{521}{1000} = 0,521.$$

Donc d'après la LGN, p est proche de 521, et quand le nombre de personnes sondées n grandit, $\overline{X_n}$ se rapproche de p .

Remarque 6. *Il existe des variables aléatoires qui peuvent prendre de très grandes valeurs avec faible probabilité (on n'en a pas vu en cours), mais qui fait rater la loi des grands nombres car elles n'ont pas d'espérance finie, c'est-à-dire*

$$\sum_{x \in \mathcal{S}(X)} |x| \mathbb{P}(X = x) = \infty \text{ ou } \int_{\mathbb{R}} |x| f_X(x) dx = \infty.$$

Ce genre de phénomène doit être étudié dans la théorie du risque, par exemple en assurances pour prévoir les conséquences de catastrophes naturelles, ou en finance pour évaluer les risques liés aux crashes boursiers (TD7, exo 5).

Exercice 7. *Soit X une variable de Cauchy, c'est-à-dire dont la densité est*

$$f_X(x) = c \frac{1}{1 + x^2}, x \in \mathbb{R}.$$

1. Déterminer c .
2. Montrer que $\mathbb{E}(|X|) = +\infty$.

5.3 Estimation ponctuelle de l'espérance

Supposons à présent que X est le résultat d'une expérience aléatoire, et que l'espérance $\mathbb{E}(X) = \mu$ est inconnue, comme dans l'exemple du dé truqué. La loi des grands nombres montre que l'on peut obtenir une valeur approchée de μ en réalisant un grand nombre n de fois la même expérience, puis en calculant la moyenne $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, où X_i correspond au résultat de la i^e expérience. Il s'agit ici d'un problème d'estimation, et nous allons introduire un peu de vocabulaire relatif à ce contexte :

Définition. *Un paramètre inconnu λ est un nombre réel, fixé mais inconnu, dont on voudrait connaître la valeur.*

Par exemple si l'espérance $\mu = \mathbb{E}(X)$ d'une variable est inconnue, μ est un paramètre inconnu. Dans l'exemple du dé truqué, la probabilité p que le dé tombe sur 6 est un paramètre inconnu.

Définition. *Un échantillon (X_1, X_2, \dots, X_n) de même loi que X est un vecteur aléatoire tel que toutes les variables X_i sont indépendantes et ont toutes même loi que X .*

Définition. *Un estimateur du paramètre λ est une variable aléatoire E_n fonction de l'échantillon, c'est-à-dire fonction des X_i pour $1 \leq i \leq n$.*

Cette notion seule est donc vide de sens, puisqu'on ne demande pas a priori que l'estimateur L_n soit proche de λ .

Définition. *On dit qu'un estimateur L_n du paramètre λ est **convergent** ou **consistant** si L_n converge vers λ presque sûrement.*

Définition. On dit qu'un estimateur L_n de λ est **sans biais** si $\mathbb{E}(L_n) = \lambda$.

En utilisant ce vocabulaire de l'estimation, la loi des grands nombres se traduit donc ainsi : la moyenne $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur du paramètre $\mu = \mathbb{E}(X_1)$ convergent. C'est aussi un estimateur sans biais car $\mathbb{E}(\overline{X_n}) = \mu$.

Dans l'exemple du dé truqué, la moyenne $\overline{X_n}$ est donc un estimateur de $\mu = \mathbb{E}(X_1)$ convergent. Cependant dans cet exemple on cherchait un estimateur du paramètre p et non pas de μ . On avait vu que $\mu = 3(p+1)$, donc l'estimateur de p est en fait $E_n = \frac{\overline{X_n}}{3} - 1$. On vérifie facilement qu'il est convergent et sans biais :

$$\begin{aligned} \mathbb{P}\left(E_n \xrightarrow[n \rightarrow \infty]{} p\right) &= \mathbb{P}\left(\frac{\overline{X_n}}{3} - 1 \xrightarrow[n \rightarrow \infty]{} p\right) = \mathbb{P}\left(\overline{X_n} \xrightarrow[n \rightarrow \infty]{} 3(p+1)\right) = \mathbb{P}\left(\overline{X_n} \xrightarrow[n \rightarrow \infty]{} \mu\right) = 1, \\ \mathbb{E}(E_n) &= E\left(\frac{\overline{X_n}}{3} - 1\right) = \frac{\mathbb{E}(\overline{X_n})}{3} - 1 = \frac{\mu}{3} - 1 = p. \end{aligned}$$

Exemple 20. La moyenne n'est pas le seul estimateur possible. Considérons par exemple le problème suivant. On appelle t le temps minimal qu'il faut à un être humain pour courir le 100 mètres (en supposant que t existe. Certaines études scientifiques situent t autour des 9,40 sec). On cherche à estimer t .

On numérote les êtres humains (du passé, du présent et du futur), et on appelle $X_i, i \geq 1$ le temps réalisé au 100m par le i -ème être humain. t est donc la plus petite valeur prise par les X_i . On suppose qu'on a accès aux valeurs de X_1, \dots, X_n pour un $n \in \mathbb{N}$. Il y a des chances que la moyenne ait plutôt une valeur de l'ordre de 15/20 sec., ce qui est un mauvais estimateur car on sait déjà que $t < 10$ sec, et il ne semble pas possible de déduire ce minimum de la moyenne.

On va plutôt définir l'estimateur suivant, plus naturel dans ce cas-là :

$$E_n = \min_{i=1, \dots, n} (X_i).$$

Si on prend le coureur le plus rapide de cet échantillon, on a plus de chance de se rapprocher de t , et si en effet $n \rightarrow \infty$, on peut imaginer que E_n converge vers t , c'est-à-dire que E_n est consistant.

Par contre, à moins d'avoir sous la main l'homme le plus rapide de l'histoire (passée et future) de l'humanité, on a probablement $E_n > t$, ce qui veut dire que $\mathbb{E}(E_n) > t$, et E_n n'est pas un estimateur sans biais, il est donc "biaisé".

5.4 Intervalles de confiance et Théorème Central Limite

5.4.1 Précision et intervalle de confiance

Soit $X_i, i \geq 1$ des variables IID admettant un moment d'ordre 1, soit $\lambda = \mathbb{E}(X_1)$. Soit $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$ leur moyenne empirique. On sait d'après la LGN que

$$\overline{X_n} \xrightarrow[p.s.]{n \rightarrow \infty} \lambda$$

c'est-à-dire \overline{X}_n est un estimateur consistant. On voudrait savoir plus précisément, à n fixé, à quelle "distance" \overline{X}_n est de λ . On cherche un intervalle de la forme $I = [\overline{X}_n - \varepsilon, \overline{X}_n + \varepsilon]$ tel que l'on puisse dire

"Il y a de grandes chances que $\lambda \in I$ ".

Définition. Soit $\alpha \in [0, 1]$. On appelle *Intervalle de confiance (IdC)* au niveau de confiance α un intervalle aléatoire $I = [A, B]$ tel que

$\mathbb{P}(\lambda \in I) \geq \alpha$ et non pas $1 - \alpha$ comme indiqué dans la version imprimée du poly.

En général, I dépend de n le nombre de personnes interrogées, et plus n est grand, plus l'intervalle est petit, et donc plus l'estimation est précise.

Exemple 21. Lors d'une élection où les citoyens doivent choisir entre 2 alternatives, A et B , on modélise les citoyens par des variables de Bernoulli X_i de paramètre p , où p est la probabilité de voter pour A . p est donc le paramètre inconnu, car il détermine le résultat de l'élection. Si après un sondage de n personnes on a $\overline{X}_n = 53\%$, la question est de savoir si $p \in I := [\overline{X}_n - 0,03, \overline{X}_n + 0,03]$, car il suffit de savoir si $p > 1/2$ ou si $p < 1/2$.

Dans ce cas, on cherche à déterminer

$$\mathbb{P}(A \text{ est choisi}) = \mathbb{P}(p \in I),$$

c'est-à-dire le niveau de confiance de l'intervalle aléatoire I

Dans cet exemple, plus n est grand, plus la confiance dans l'intervalle sera grande. Dans d'autres cas, on effectue le chemin inverse : on part d'un niveau de confiance α , et on récolte suffisamment de données pour obtenir un niveau de confiance au niveau α .

Exemple 22. Un médicament est suspecté d'avoir des effets secondaires graves, et en particulier de provoquer des crises cardiaques. On appelle $p \in [0, 1]$ la probabilité que ce médicament provoque une crise cardiaque sur un patient pris au hasard. p est donc le paramètre inconnu que l'on cherche à estimer.

Avant d'autoriser la mise sur le marché du médicament, les autorités demandent un intervalle de confiance au niveau 99%. Le laboratoire pharmaceutique doit donc effectuer des tests dont les résultats sont notés $X_i, i = 1, \dots, n$ ($X_i = 1$ si le médicament a provoqué une crise au i -ème test, $X_i = 0$ sinon). Ils obtiendront un résultat du type

$$\mathbb{P}(p \in [\overline{X}_n - \varepsilon ; \overline{X}_n + \varepsilon]) = \alpha_n$$

où plus n est grand, plus α_n est proche de 1 ($\lim_{n \rightarrow \infty} \alpha_n = 1$).

Le travail du laboratoire sera donc de déterminer les valeurs de ε et n pour que $\alpha_n \geq 0,99$, puis de réaliser n tests. Si alors $\overline{X}_n + \varepsilon$ est suffisamment petit, le médicament pourra être utilisé. S'il est légèrement trop grand, on pourra augmenter le nombre de données n pour réduire ε_n , tout en gardant le niveau α_n constant.

Les tests étant coûteux, il est important pour le laboratoire de prendre n le plus petit possible tout en respectant cette condition.

5.4.2 Renormalisation

Soit X une VA admettant un moment d'ordre 2. La renormalisation, ou standardisation, de X , consiste à lui retrancher son espérance et la diviser par son écart-type :

$$X' = \frac{X - \mathbb{E}(X)}{\sigma(X)},$$

on rappelle que $\sigma(X) = \sqrt{V(X)}$. Cette version renormalisée de X satisfait

$$\begin{aligned}\mathbb{E}(X') &= \mathbb{E}\left(\frac{X - \mathbb{E}(X)}{\sigma(X)}\right) = \frac{\mathbb{E}(X) - \mathbb{E}(X)}{\sigma(X)} = 0 \\ V(X') &= V\left(\frac{X - \mathbb{E}(X)}{\sigma(X)}\right) = \left(\frac{1}{\sigma(X)}\right)^2 V(X - \mathbb{E}(X)) = \frac{1}{\sigma(X)^2} V(X) = \frac{1}{V(X)} V(X) = 1.\end{aligned}$$

Cette renormalisation permet de comparer X à une autre variable équitablement.

Par exemple, la corrélation de deux VA X et Y s'écrit

$$\text{Corr}(X, Y) = \mathbb{E}(X'Y'),$$

où X' et Y' sont les versions renormalisées de X et Y .

Soit $X_i, i = 1, \dots, n$. Pour les problèmes précédents (élections, tests de médicament, ...), on cherche à savoir comment se comporte \overline{X}_n lorsque n est grand. Pour avoir une variable à la bonne échelle, on considère donc la version renormalisée

$$\overline{X}_n' = \frac{X_n - \mathbb{E}(X_n)}{\sqrt{V(\overline{X}_n)}} = \frac{X_n - n\mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}}{\sigma}(\overline{X}_n - n\mu)$$

où $\mu = \mathbb{E}(X), \sigma^2 = V(X)$.

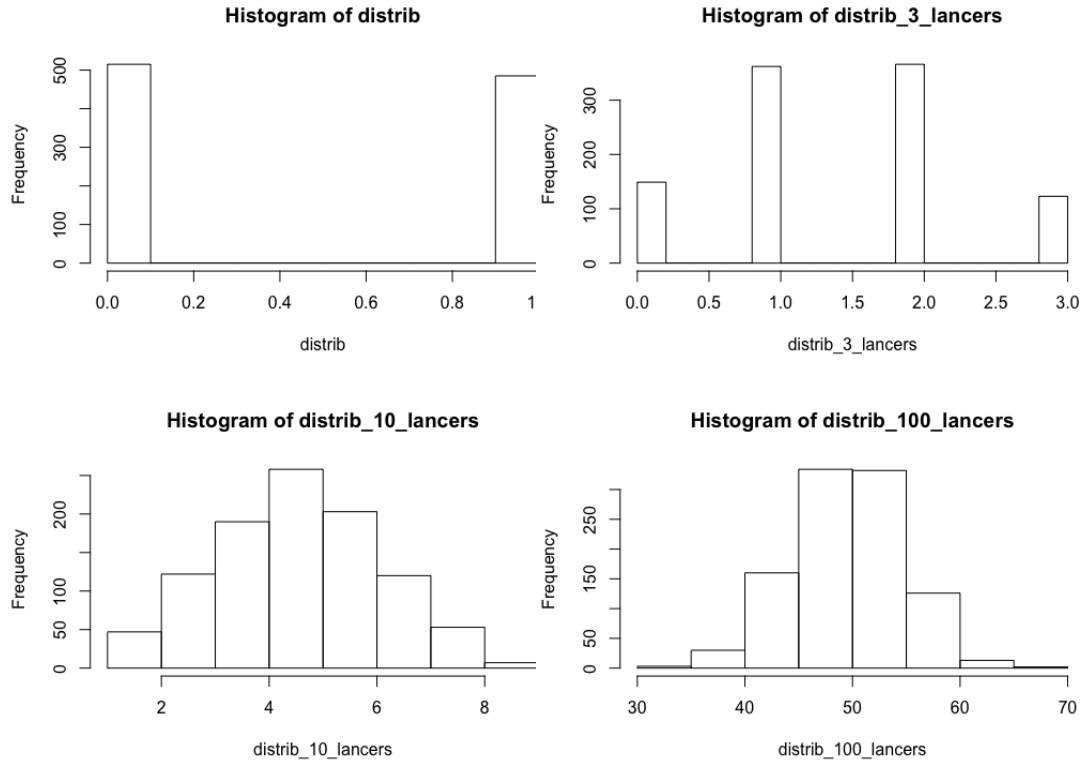
5.4.3 Théorème central limite

Phénomène de recentrage

Un ami vous propose de choisir entre les deux jeux suivants :

1. Vous jouez 15 euros à pile ou face.
2. Vous jouez trois fois 5 euros à pile ou face.

Dans le premier cas, vous pouvez gagner soit -15 , soit 15 . Dans le second cas, vous gagnez un nombre mieux réparti dans l'ensemble $\{-15, -5, 5, 15\}$. On obtient les histogrammes suivants :



Les deux histogrammes suivants sont obtenus avec 10, puis 100 lancers.

Cette évolution correspond à, plutôt que de faire une fois une expérience aléatoire, faire la moyenne sur n fois. Voir aussi le TP3.

On observe le même phénomène de “recentration”. Plus le nombre n de variables IID est grand, plus la courbe ressemble à une montagne, qui en fait la densité gaussienne. On verra que ce phénomène est “universel”, dans le sens qu’il survient quelle que soit la loi des variables iid (dans ce cas précis une loi de Bernoulli).

Théorème 3. Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires indépendantes et identiquement distribuées, et X une variable de loi $\mathcal{N}(0, 1)$. On suppose que l’espérance $\mu = \mathbb{E}(X_1)$ et la variance $\sigma^2 = V(X_1)$ sont bien définies. On pose

$$\overline{X}_n' = \sqrt{n} \frac{X_n - \mu}{\sigma}.$$

Alors pour tous a, b réels tels que $a < b$,

$$\mathbb{P}\left(a \leq \overline{X}_n' \leq b\right) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Remarque 7. On peut prendre $a = -\infty$ ou $b = +\infty$ dans ce résultat.

Définition. Ce type de convergence s’appelle “convergence en loi”. Soit $Y_n, n \geq 1$ une suite de VA, et Y une autre VA. On dit que $Y_n \rightarrow Y$ “en loi” si pour tout intervalle I de \mathbb{R} ,

$$\mathbb{P}(Y_n \in I) \rightarrow \mathbb{P}(Y \in I).$$

Cette convergence est différente de la convergence “presque sûre” : On ne peut pas savoir ce qui arrive à Y_n pour n grand de manière sûre, on ne peut que connaître la tendance.

Autrement dit, ce théorème dit que la loi de

$$\overline{X_n}' = \frac{\sqrt{n}}{\sigma}(\overline{X_n} - \mu)$$

converge vers celle de X , c'est-à-dire que tout calcul de probabilité effectué sur cette variable peut être approché par un calcul de probabilité sur X .

remarque : l'expression $\frac{\sqrt{n}}{\sigma}(\overline{X_n} - \mu)$ peut s'écrire de différentes façons : souvent on note $S_n = \sum_{i=1}^n X_i$, de sorte que $\overline{X_n} = \frac{1}{n}S_n$, et donc

$$\overline{X_n}' = \frac{\sqrt{n}}{\sigma}(\overline{X_n} - \mu) = \frac{\sqrt{n}}{\sigma} \left(\frac{1}{n}S_n - \mu \right) = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Quantiles gaussiens

La limite dans le TCL est la quantité $\mathbb{P}(a \leq X \leq b)$, où $a, b \in \mathbb{R} \cup \{\pm\infty\}$. Pour établir un intervalle de confiance à un niveau α , il faut trouver les bonnes valeurs de a, b pour que

$$\mathbb{P}(a \leq X \leq b) = \alpha.$$

Nous allons voir comment déterminer ces valeurs dans deux cas à l'aide de la fonction de répartition F_X de la loi gaussienne :

$$\text{1er cas : } a = -\infty \Leftrightarrow \alpha = \mathbb{P}(X \leq b) = F_X(b).$$

Pour déterminer b tel que $F_X(b) = \alpha$, on utilise la notion de **quantile** :

Définition. Soit X une VA. Soit $\alpha \in [0, 1]$. On appelle **quantile de niveau α** le plus petit nombre b tel que $F_X(b) \geq \alpha$. On le note $b = F^{-1}(\alpha)$.

Remarque 8. Malgré la notation, F^{-1} n'est pas toujours l'application inverse de F , car F n'est pas forcément une bijection, on l'appelle “pseudo-inverse”. Si F est bijective, F^{-1} est bien son inverse.

Pour connaître les quantiles gaussiens, on peut se référer à un tableau donnant les valeurs de F_X . Pour déterminer le quantile $F_X^{-1}(\alpha)$, il faut identifier le plus grand nombre $\alpha' \geq \alpha$ dans la table (en général, α' est très proche de α), et identifier avec la ligne et la colonne le nombre b tel que $F_X(b) = \alpha'$. On utilise alors $b \approx F^{-1}(\alpha)$ comme quantile.

Exemple 23. Prenons $\alpha = 95\%$. On repère $\alpha' = 0,95053$ à la ligne 1,6 et la colonne 0,05. On a donc $F_X(b) = \alpha'$ avec $b = 1,65$. On prend donc b comme quantile de niveau 95% car $F_X(b) \geq 95\%$. On calcule (approximativement) de la même manière $F_X^{-1}(90\%)$, $F_X^{-1}(80\%)$. On a

$$\begin{aligned} F_X^{-1}(95\%) &\approx 1,65 \\ F_X^{-1}(90\%) &\approx 1,29 \\ F_X^{-1}(80\%) &\approx 0,85. \end{aligned}$$

Le cas que l'on rencontre le plus souvent est le suivant :

$$2^{\text{ème}} \text{ cas : } b > 0 \text{ et } a = -b \Leftrightarrow \alpha = \mathbb{P}(-b \leq X \leq b) = \mathbb{P}(|X| < b).$$

Etant donné $\alpha \in [0, 1]$, pour trouver b tel que $\mathbb{P}(|X| \leq b) \geq \alpha$, on utilise l'astuce suivante, à connaître :

$$\mathbb{P}(|X| \leq b) = 1 - \mathbb{P}(|X| > b) = 1 - \mathbb{P}((X > b) \cup (X < -b)) = 1 - \mathbb{P}(X > b) - \mathbb{P}(X < -b).$$

On utilise le fait que X est symétrique, c'est-à-dire que sa densité est paire : $f_X(x) = f_X(-x)$. Donc

$$\mathbb{P}(X < -b) = \int_{-\infty}^{-b} f_X(x) dx = \int_b^{\infty} f_X(x) dx = \mathbb{P}(X > b) = 1 - F_X(b).$$

On a donc pour $b > 0$

$$\begin{aligned} \mathbb{P}(|X| < b) = \alpha &\Leftrightarrow \alpha = 1 - 2(1 - F_X(b)) = 2F_X(b) - 1 \\ &\Leftrightarrow F_X(b) = \frac{1 + \alpha}{2} \\ &\Leftrightarrow b = F_X^{-1}\left(\frac{1 + \alpha}{2}\right). \end{aligned}$$

Exemple 24. On prend $\alpha = 95\%$, donc $\frac{1+\alpha}{2} = 97,5\%$. On lit sur le tableau

$$b = F_X^{-1}(97,5\%) \approx 1,96.$$

On résume ces deux cas dans le tableau suivant (à connaître ou savoir retrouver) :

a	α	b	$\alpha = 95\%$	$\alpha = 90\%$	$\alpha = 80\%$
$-\infty$	$\mathbb{P}(X \leq b)$	$F_X^{-1}(\alpha)$	1,65	1,29	0,85
$-b$	$\mathbb{P}(X \leq b)$	$F_X^{-1}\left(\frac{1+\alpha}{2}\right)$	1,96	1,65	1,29

5.4.4 Intervalle de confiance grâce au théorème central limite

On reprend les notations précédentes : les X_i sont des VA iid, et $\overline{X}_n' \rightarrow X \sim \mathcal{N}(0, 1)$ en loi. D'après le théorème central limite, pour n "grand", $b > 0$,

$$\mathbb{P}\left(-b \leq \frac{\sqrt{n}}{\sigma} (\overline{X}_n - \mu) \leq b\right) \simeq \mathbb{P}(-b \leq X \leq b).$$

Voilà comment en déduire un intervalle de confiance au niveau α .

1. Tout d'abord, on cherche b tel que $\mathbb{P}(-b \leq X \leq b) \approx \alpha$, en utilisant les quantiles gaussiens. On voit dans le tableau qu'il faut prendre $b = F_X^{-1}\left(\frac{1+\alpha}{2}\right)$

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad \Phi(-x) = 1 - \Phi(x), \quad 0 \leq x < 1.99$$

X\0.0	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670

FIGURE 5.1 – X : variable gaussienne standard. Cette table permet de déterminer $\mathbb{P}(X \leq x)$ pour x entre 0 et 2

- Si l'on connaît σ , on pose $\hat{\sigma} = \sigma$ et on passe à l'étape suivante. Sinon, il faut trouver une valeur $\hat{\sigma}$ telle que $\hat{\sigma} \approx \sigma$ ou $\hat{\sigma} \geq \sigma$. On pose alors plutôt

$$I' = \left[\overline{X}_n - \frac{\hat{\sigma}b}{\sqrt{n}}, \overline{X}_n + \frac{\hat{\sigma}b}{\sqrt{n}} \right].$$

Voir ci-dessous pour une manière de trouver $\hat{\sigma}$.

Explication : Comme $\hat{\sigma} > \sigma$, $I \subset I'$. Comme $\mathbb{P}(\mu \in I) \geq \alpha$, on a aussi $\mathbb{P}(\mu \in I') \geq \alpha$, donc I' est aussi un IdC au niveau α , mais il est moins précis.

- On effectue ensuite le calcul suivant :

$$\begin{aligned} \mathbb{P}\left(-b \leq \frac{\sqrt{n}}{\sigma} (\overline{X}_n - \mu) \leq b\right) &= \mathbb{P}\left(-\frac{\sigma}{\sqrt{n}}b \leq \overline{X}_n - \mu \leq \frac{\sigma}{\sqrt{n}}b\right) \\ &= \mathbb{P}\left(\overline{X}_n - \frac{\sigma}{\sqrt{n}}b \leq \mu \leq \overline{X}_n + \frac{\sigma}{\sqrt{n}}b\right) \\ &= \mathbb{P}(\mu \in I) \text{ où } I = \left[\overline{X}_n - \frac{\sigma b}{\sqrt{n}}, \overline{X}_n + \frac{\sigma b}{\sqrt{n}} \right]. \end{aligned}$$

- On conclut : comme $\mathbb{P}(\mu \in I) \approx \mathbb{P}(-b \leq X \leq b) \geq \alpha$, I est bien un intervalle de confiance au niveau α .

On peut utiliser sans démonstration le fait suivant :

Proposition. Pour $\alpha \in [0, 1]$ et $b = F_X^{-1}\left(\frac{1+\alpha}{2}\right)$,

$$I = \left[\overline{X}_n - \frac{\hat{\sigma}b}{\sqrt{n}}, \overline{X}_n + \frac{\hat{\sigma}b}{\sqrt{n}} \right]$$

est un IdC au niveau α , pour $\hat{\sigma} \geq \sigma$. Si l'on connaît σ , il faut remplacer $\hat{\sigma}$ par σ .

Exemple 25. Supposons que les X_i aient des lois de Bernoulli, comme dans le cas de l'élection. Alors $\sigma = \sqrt{V(X_i)} = \sqrt{p(1-p)} \leq 1$. Donc au point 2 on prend $\hat{\sigma} = 1$,

$$I' = \left[\overline{X_n} - \frac{b}{\sqrt{n}}, \overline{X_n} + \frac{b}{\sqrt{n}} \right].$$

Admettons pour l'application numérique que $\overline{X_n} = 0,53$, $n = 10000$, $\alpha = 90\%$. On a $b = F_X((1+\alpha)/2) \approx 1,96$, $\hat{\sigma} = 1$.

$$I = [0,53 \pm \frac{1,96}{\sqrt{n}}] \approx [0,51; 0,55].$$

Dans le cas des variables de Bernoulli, on peut même faire mieux. On a en effet le résultat suivant :

$$\forall p \in [0,1], p(1-p) \leq 1/4.$$

Donc $\sigma \leq \frac{1}{2}$ et on peut prendre plutôt $I' = \left[\overline{X_n} - \frac{b}{2\sqrt{n}}, \overline{X_n} + \frac{b}{2\sqrt{n}} \right]$.

Pour prouver $p(1-p) \leq \frac{1}{4}$, on fait l'étude de la fonction $f(p) = p(1-p) : f(0) = 0, f(1) = 0$, et $f'(p) = 1 - 2p$. Donc $f'(p) = 0$ uniquement si $p = 1/2$, ce qui veut dire que f admet son maximum à cet endroit. Donc $\max_{p \in [0,1]} f(p) = f(1/2) = \frac{1}{4}$.

Nous allons voir maintenant une manière générale d'estimer la variance.

5.4.5 Estimation ponctuelle de la variance

On explique ici comment déterminer $\hat{\sigma}$ dans le cas général. On a

$$\sigma^2 = V(X_1) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mu^2.$$

Les VA $X_i^2 - \mu, i \in \mathbb{N}$, sont IID (car les X_k le sont), donc d'après la LGN,

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \rightarrow \mathbb{E}(X_1^2)$$

presque sûrement. Donc

$$\hat{\sigma}_n = \frac{1}{n} \sum_{i=1}^n X_i^2 - \mu^2 \rightarrow \sigma^2.$$

C'est donc un estimateur consistant de la variance. Le nouveau problème est qu'on ne connaît pas μ (le but du problème initial est de l'estimer). La difficulté est donc qu'on estime deux quantités à la fois.

On va devoir trouver un estimateur de la variance qui ne dépend pas de μ .

Pour estimer σ à partir de la donnée des X_i , on a le résultat suivant.

Proposition.

$$V'_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$$

est un estimateur sans biais et convergent.

Preuve.

On va d'abord étudier l'estimateur

$$\begin{aligned} V'_n &:= \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i^2 + \overline{X}_n^2 - 2X_i \overline{X}_n) \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 + \frac{1}{n} \sum_{i=1}^n \overline{X}_n^2 - 2\overline{X}_n \frac{1}{n} \sum_{i=1}^n X_i \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 + \frac{1}{n} n \overline{X}_n^2 - 2\overline{X}_n^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X}_n^2. \end{aligned}$$

On va regarder la convergence de $\frac{1}{n} \sum_{i=1}^n X_i^2$, puis celle de \overline{X}_n^2 :

- D'après la loi des grands nombres $\frac{1}{n} \sum_{i=1}^n X_i^2$ converge vers $\mathbb{E}(X_1^2)$, qui est égal à $V(X_1) + \mathbb{E}(X_1)^2 = \sigma^2 + \mu^2$.
- D'après la loi des grands nombres, \overline{X}_n converge vers μ , et donc \overline{X}_n^2 converge vers μ^2 : en effet,

$$\mathbb{P}(\overline{X}_n^2 \xrightarrow[n \rightarrow \infty]{} \mu^2) = \mathbb{P}(\overline{X}_n \xrightarrow[n \rightarrow \infty]{} \mu) = 1.$$

Par conséquent V_n converge vers $\sigma^2 + \mu^2 - \mu^2 = \sigma^2$, c'est donc bien un estimateur de σ^2 , et $\sqrt{V_n}$ est un estimateur de σ .

V_n a-t-il un biais ? Un calcul donne

$$\mathbb{E}(V_n) = \frac{n}{n-1} \sigma^2 \neq \sigma^2,$$

il y a donc un léger biais. Par contre si l'on pose

$$V'_n = \frac{n-1}{n} V_n,$$

on a par linéarité de l'espérance

$$\mathbb{E}(V'_n) = \frac{n-1}{n} \mathbb{E}(V_n) = \frac{n-1}{n} \frac{n}{n-1} \sigma^2 = \sigma^2,$$

c'est donc un estimateur consistant et sans biais. □