UNIVERSITÉ DE FRANCHE-COMTÉ

Licence Informatique 2ème année

UFR Sciences et Techniques

2018–2019

Probabilités élémentaires

Jürgen ANGST

Notes de cours
http://www.angst.fr

Table des matières

Première partie

Éléments de théorie des probabilités

Chapitre 1

Le formalisme de la théorie des probabilités

1.1 Théorie des ensembles et dénombrement

Avant toute chose, nous commençons par énoncer quelques rappels élémentaires de théorie des ensembles ainsi que de combinatoire (appelée aussi dénombrement) qui nous seront indispensables dans la suite.

1.1.1 Rappels de théorie des ensembles

La notion d'ensemble est au coeur de l'axiomatique des mathématiques. Elle nous est familière puisqu'on l'utilise quotidiennement lorsque l'on parle de "l'ensemble des étudiants de la licence d'informatique", "l'ensemble des mains possibles au poker", qui sont deux ensembles finis, ou encore "l'ensemble des textes que l'on peut écrire en français" ou "l'ensemble de toutes les nuances de couleurs", qui eux sont des ensembles infinis. Cependant, la notion d'ensemble cache des subtilités / difficultés importantes, comme l'a montré Russell au début du vingtième siècle avec son contre-exemple célèbre : l'ensemble de tous les ensembles n'est pas un ensemble.

On définit informellement un ensemble comme une collection d'éléments :

$$\{0,1\}, \quad \{\text{rouge}, \text{noir}\}, \quad \{0,1,2,3,\ldots\} = \mathbb{N}.$$

Un ensemble joue un rôle particulier, l'ensemble vide, noté $\emptyset$, c'est l'ensemble qui ne contient aucun élément. On note $x \in E$ si x est un élément de E, et $x \notin E$ dans le cas contraire. On dit qu'un ensemble E est inclus dans un ensemble F et on note $E \subset F$ si tout élément de E est aussi un élément de F. On dit aussi que E est une partie de F. Deux ensembles E et F sont égaux, i.e. $E = F$ si et seulement si $E \subset F$ et $F \subset E$. On note $\mathcal{P}(E)$ l'ensemble des parties de E. Par exemple si $E = \{1,2,3\}$:

$$\mathcal{P}(\{1,2,3\}) = \big\{\varnothing, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\big\}.$$

Étant donnés deux ensembles A et B, on appelle l'*union* de A et B et on note $A \cup B$ l'ensemble formé des éléments qui appartiennent à l'ensemble A ou à l'ensemble B.

On appelle *intersection* de A et B et on note $A \cap B$ l'ensemble formé des éléments qui appartiennent à l'ensemble A et l'ensemble B. Si l'intersection de A et B est vide, *i.e.* $A \cap B = \emptyset$, on dit que les ensembles A et B sont *disjoints*. Dans ce cas, l'union de A et B est dite *union disjointe*, et l'on note $A \sqcup B$.

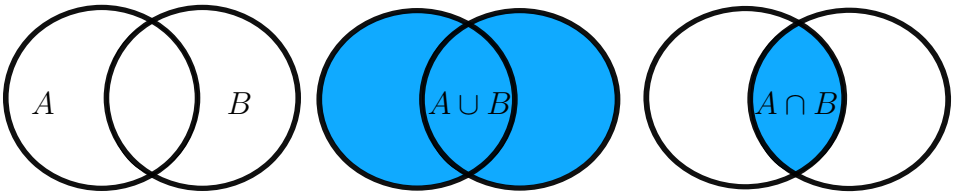


FIGURE 1.1 – Union et intersection de deux ensembles

Soient A, B, C des parties d'un ensemble E, on a les règles de calcul suivantes :

- $A \cap B = B \cap A$
- $A \cap (B \cap C) = (A \cap B) \cap C$ (on peut donc écrire $A \cap B \cap C$ sans ambigüité)
- $A \cap \varnothing = \varnothing, \quad A \cap A = A, \quad A \subset B \iff A \cap B = A$

et

- $A \cup B = B \cup A$
- $A \cup (B \cup C) = (A \cup B) \cup C$ (on peut donc écrire $A \cup B \cup C$ sans ambiguïté)
- $A \cup \varnothing = A, \quad A \cup A = A, \quad A \subset B \iff A \cup B = B.$

Par ailleurs, union et intersection se distribuent de la façon suivante (notez l'analogie avec la distributivité de $+$ et $\times$)

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$

Plus généralement, étant donnés des ensembles $(A_i)_{i \in I}$ indexés par un ensemble d'indice I, on note $\cup_{i \in I} A_i$ l'ensemble des éléments qui appartiennent à l'un des A_i et $\cap_{i \in I} A_i$ l'ensemble des éléments qui appartiennent à tous les A_i, de sorte que

- $x \in \bigcup_{i \in I} A_i$ signifie que x appartient à l'un des ensembles A_i ;
- $x \in \bigcap_{i \in I} A_i$ signifie que x appartient à tous les ensembles A_i.

Soient trois ensembles A, B et Ω tels que $A \subset \Omega$ et $B \subset \Omega$. On appelle *complémentaire* de A (dans Ω) et on note A^c l'ensemble des éléments de Ω qui ne sont pas dans A. On désigne par B *privé de* A et on note $B \backslash A$, l'ensemble des éléments de B qui ne sont pas dans A, c'est-à-dire $B \cap A^c$. On a les règles de calcul suivantes :

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

Plus généralement, si $(A_i)_{i \in I}$ est une famille d'ensembles inclus dans Ω, on a alors les relations :

$$\left(\bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c, \qquad \left(\bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c.$$

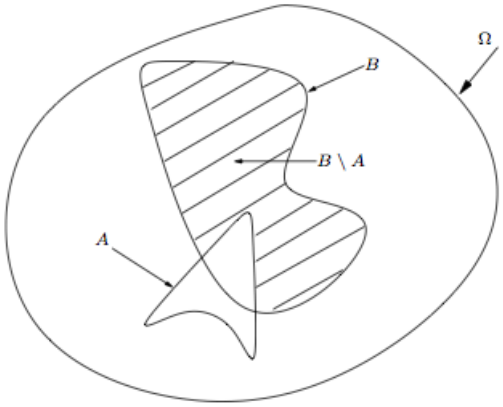


FIGURE 1.2 – Soutraction de deux ensembles.

Par exemple, si l'on considère les ensembles G et A des germanophones et des anglophones dans la population francaise, le complémentaire de $G \cap A$ est $G^c \cup A^c$, *i.e.* le contraire de "parler allemand et anglais" et "ne pas parler allemand ou ne pas parler anglais.

Si E et F sont deux ensembles, le produit cartesien de E et F, noté $E \times F$, est l'ensemble des couples (x, y) où $x \in E$ et $y \in F$. Dans le cas ou $E = F$, on note simplement $E^2 = E \times E$, et plus généralement $E^n = E \times \ldots \times E$ si le produit est réalisé n fois. Par exemple

$$\begin{aligned} \{0,1\}^2 &= \{(0,0),(0,1),(1,0),(1,1)\}, \\ \mathbb{R}^2 &= \mathbb{R} \times \mathbb{R} = \{(x,y),\, x,y \in \mathbb{R}\}, \\ [0,1]^3 &= \{(x,y,z),\, 0 \leq x \leq 1,\, 0 \leq y \leq 1,\, 0 \leq z \leq 1\}. \end{aligned}$$

Exercice 1 :
Soient $A = \{0,1,2\}$ et $B = \{0,1,2,3\}$. Décrire les ensembles $A \cap B$, $A \cup B$, $A \times B$.

Correction : On a l'inclusion $A \subset B$ donc $A \cap B = A$, $A \cup B = B$ et par définition $A \times B = \{(x,y),\, x \in A,\, y \in B\}$.

Exercice 2 :
Soient $A = [1,3]$ et $B = [2,4]$. Décrire les ensembles $A \cap B$, $A \cup B$, $B \backslash A$.

Correction : On a $A \cap B = [2,3]$, $A \cup B = [1,4]$ et $B \backslash A =]3,4]$.

Exercice 3 :
Si A, B et C sont trois ensembles, montrez que :

$$(A \backslash B) \backslash C = A \backslash (B \cup C), \quad (A \backslash B) \cap (C \backslash D) = (A \cap C) \backslash (B \cup D).$$

Correction : Par définition, on a $(A \backslash B) \backslash C = (A \cap B^c) \cap C^c = A \cap (B^c \cap C^c)$. Comme $B^c \cap C^c = (B \cup C)^c$, on a donc $(A \backslash B) \backslash C = A \cap (B \cup C)^c = A \backslash (B \cup C)$. De la même façon, on a $(A \backslash B) \cap (C \backslash D) = A \cap B^c \cap C \cap D^c = A \cap C \cap (B \cup D)^c = (A \cap C) \backslash (B \cup D)$.

Exercice 4 :
Soient $A =]-\infty, 3]$, $B =]-2, 7]$, $C =]-5, +\infty[$. Déterminez les ensembles $A \cap B$, $A \cup B$, $B \cap C$, $B \cup C$, $\mathbb{R}\backslash A$, $A\backslash B$, $(\mathbb{R}\backslash A) \cap (\mathbb{R}\backslash B)$, $(A \cap B) \cup (A \cap C)$.

Correction : On a $A \cap B =]-2, 3]$, $A \cup B =]-\infty, 7]$, $B \cap C =]-2, 7]$, $B \cup C =]-5, +\infty]$. D'autre part, on a $\mathbb{R}\backslash A =]3, +\infty[$ et $A\backslash B =]-\infty, -2]$. D'après l'exercice précédent, on a $(\mathbb{R}\backslash A) \cap (\mathbb{R}\backslash B) = \mathbb{R}\backslash(A \cup B) =]7, +\infty[$. Enfin $(A \cap B) \cup (A \cap C) = A \cap (B \cup C)$ et on a donc $(A \cap B) \cup (A \cap C) =]-5, 3]$.

1.1.2 Rappels de combinatoire

On appelle cardinal de A et on note $\mathrm{Card}(A)$ ou encore $\#A$ le nombre d'éléments qu'il contient. Si A et B sont des ensembles finis, on a la relation

$$\mathrm{Card}(A) + \mathrm{Card}(B) = \mathrm{Card}(A \cup B) + \mathrm{Card}(A \cap B).$$

Si Ω est un ensemble fini de cardinal n, alors on a $\mathrm{Card}(\mathcal{P}(\Omega)) = 2^n$. Par exemple, si on considère l'ensemble $\Omega = \{0, 1\}$, alors $\mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$ et l'on a bien $\mathrm{Card}(\Omega) = 2$ et $\mathrm{Card}(\mathcal{P}(\Omega)) = 2^2 = 4$.

On rappelle les notations usuelles concernant les sommes et les produits, si $a_1, a_2, \ldots, a_n$ sont des nombres réels :

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n, \quad \prod_{i=1}^{n} a_i = a_1 \times a_2 \times \ldots \times a_n.$$

Soit A un ensemble à n éléments. Le nombre de permutations des éléments de A est appelé *factorielle* n, que l'on note $n!$. Ce nombre est égal à

$$n! := n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1.$$

Par exemple, il y a $6 = 3!$ permutations possibles de 3 symboles a, b, c : (a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a).

Remarque 1.1.1. Tous les élements sont ici supposés distinguables et on tient compte de l'ordre des éléments.

On peut aussi définir la factorielle grâce à la fonction Γ : $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$ qui a les propriétés suivantes : $\Gamma(n+1) = n!$ pour n entier et $\Gamma(x+1) = x\Gamma(x)$. La formule de Stirling permet de construire une estimation asymptotique de la factorielle

$$n! \approx n^n e^{-n} \sqrt{2\pi n}(1 + \frac{1}{12n} + \frac{1}{288n^2} + \ldots).$$

Le nombre de façons de choisir p éléments de A parmi les n est appelé *arrangement de p objets parmi n*. Il est souvent noté A_n^p et vaut :

$$A_n^p := \frac{n!}{(n-p)!} = n \times (n-1) \times (n-2) \times \cdots \times (n-p+1).$$

Remarque 1.1.2. Ici encore, on tient compte de l'ordre des éléments.

Le nombre de façons de choisir p éléments de A parmi les n éléments sans tenir compte de l'ordre est appelé *combinaison de p objets parmi n*. Il est noté C_n^p ou encore $\binom{n}{p}$ et vaut :

$$C_n^p := \frac{n!}{p!(n-p)!} = \frac{A_n^p}{p!}.$$

Les nombres C_n^p sont appelés *coefficients binomiaux* et possèdent les propriétés suivantes :

$$C_n^0 = C_n^n = 1, \quad C_n^p = C_n^{n-p}, \quad C_n^p = C_{n-1}^{p-1} + C_{n-1}^p, \quad \sum_{p=1}^{n} C_n^p = 2^n.$$

La propriété $C_n^p = C_{n-1}^{p-1} + C_{n-1}^p$ permet de calculer les coefficients de proche en proche grâce au triangle de Pascal :

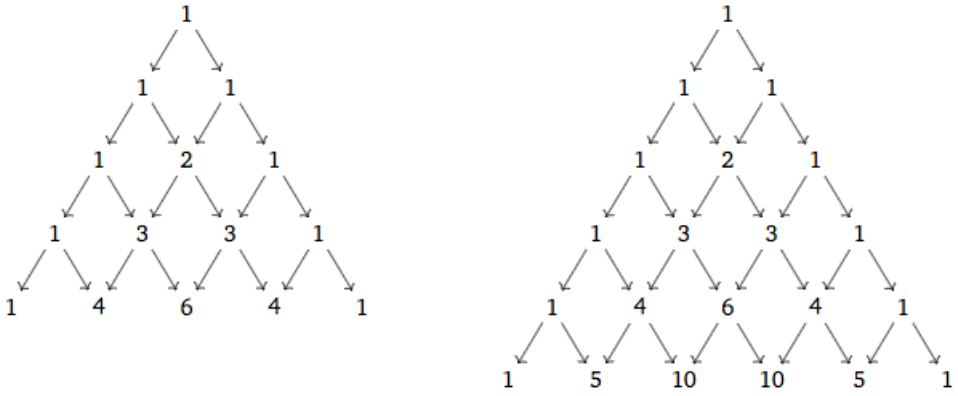


FIGURE 1.3 – Triangle de Pascal.

Les coefficients binomiaux sont les nombres qui apparaissent dans la formule du binôme de Newton qui généralise l'identité remarquable $(a+b)^2 = a^2 + 2ab + b^2$. On a ainsi, pour tout $a, b \in \mathbb{R}$ et pour tout entier positif n :

$$\begin{aligned}(a+b)^n &= a^n + C_1^n a^{n-1}b + C_2^n a^{n-2}b^2 + \ldots + C_p^n a^{n-p}b^p + \ldots + C_n^{n-1}ab^{n-1} + b^n, \\ &= \sum_{p=0}^{n} C_p^n a^{n-p}b^p.\end{aligned}$$

Par exemple, en lisant les coefficients binomiaux sur la 3ème ligne du triangle de Pascal, on a

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3.$$

De mème, de la ligne suivante, on déduit la formule

$$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4.$$

Exercice 1 :
Combien existe-t-il de plaques minéralogiques à 7 caractères (les 2 premiers étant des lettres et les 5 autres des chiffres) ? Même question si l'on impose que les répétitions de lettres ou de chiffres sont exclues.

Correction : Si on autorise les répétitions, on a 26×26 choix pour les lettres, et $10 \times 10 \times ... \times 10 = 10^5$ pour les chiffres, soit au total : $N = 26^2 \times 10^5$ possibilités. Si les répétitions sont proscrites, alors on a 26×25 choix pour les lettres et $10 \times 9 \times 8 \times 7 \times 6$ choix pour les chiffres, soit au total : $N = 26 \times 25 \times 10 \times 9 \times 8 \times 7 \times 6 = 19656000$ possibilités.

Exercice 2 :
On doit asseoir sur un même rang 4 allemands, 3 français, et 3 anglais ; les gens de même nationalité devant rester groupés. Combien de dispositions sont possibles ?

Correction : Les personnes de même nationalité devant rester groupées, on peut tout d'abord choisir l'ordre des 3 nationalités sur le rang : pour cela on $N = 3! = 6$ configurations possibles :

D	F	GB
D	GB	F
F	D	GB
F	GB	D
GB	D	F
GB	F	D

Ensuite, on peut permuter les personnes au sein d'une même nationalité, au total il y a donc $N = 6 \times 4! \times 3! \times 3!$ configurations.

Exercice 3 :
Combien existe-t-il d'arrangements différents avec les lettres des mots suivants : a) pinte ; b) proposition ; c) Mississipi ; d) arrangement ?

Correction : Dans le mot "pinte" chaque lettre apparaît une seule fois, le nombre d'arrangements de lettres distincts que l'on peut former est donc $5! = 120$. Dans le mot "proposition", il y a 11 lettres dont 2 "p", 3 "o", 2 "i". Pour ne pas compter plusieurs fois le même arrangement (par exemple, si on ne regarde que les "p", "pproosition" apparaît deux fois, si on ne regarde que les "o", "oooprpsitin" apparaît $3! = 6$ fois...) on est amené à diviser le nombre des permutations possibles des lettres par $2! \times 3! \times 2! = 24$. Le nombre d'arrangements distincts est donc

$$N = \frac{11!}{2! \times 3! \times 2!} = 1663200.$$

De même pour "Mississipi", il y a 10 lettres dont 4 "i" et 4 "s", le nombre de possibilités est alors $N = \frac{10!}{4! \times 4!} = 6300$. Pour "arrangement", on trouve

$$N = \frac{11!}{2! \times 2! \times 2! \times 2!} = 2494800.$$

Exercice 4 :
On veut former un comité de 7 personnes, constitué de 2 démocrates, 2 républicains, et 3 indépendants. On a le choix parmi 6 démocrates, 5 républicains, et 4 indépendants. Combien de choix sont possibles ?

Correction : On détermine le nombre de possibilités dans chacune des 3 obédiences, le nombre total de choix possibles est alors le produit de ces trois nombres. Pour les démocrates, on a C_6^2 choix, pour les républicains C_5^2, et pour les indépendants C_4^3. Le nombre comités distincts que l'ont peut ainsi former est :

$$N = C_6^2 \times C_5^2 \times C_4^3 = 600.$$

1.2 Espace de probabilités

L'objet de la théorie des probabilités est de modéliser des phénomènes complexes dont il n'est pas en général possible de prédire avec certitude leur évolution ou les conséquences qu'ils peuvent engendrer. L'archétype d'un tel phénomème est le lancer d'une pièce à pile ou face : les mécanismes physiques à prendre en compte pour décrire l'expérience du lancer sont d'une telle complexité qu'il n'est pas envisageable de répondre de façon déterministe à la question la pièce va-t-elle tomber coté pile, face, sur la tranche ? On dit alors que le résultat de l'expérience est *aléatoire* ou encore *stochastique*. Voici d'autres exemples d'expériences usuelles dont le résultat est de nature aléatoire :

Expérience	Résultat observable
Lancer d'un dé	Un entier $k \in \{1, \ldots, 6\}$
Lancer d'une fléchette sur une cible	Point d'impact
Sondage à la sortie des urnes au cours d'un référendum	Nombre de Oui et de Non dans l'échantillon
Saut en longueur dans une compétition d'athlétisme	Saut éventuellement mordu, sinon un nombre $\ell \geqslant 0$
Mouvement d'un grain de pollen dans un liquide	Une trajectoire continue dans l'espace à trois dimensions

Pour modéliser ce type d'expériences, la démarche du probabiliste consiste tout d'abord à en préciser tous les résultats possibles. Ensuite, chaque résultat possible se voit attribuer un certain poids, une probabilité. Dans l'exemple du lancer à pile ou face, l'ensemble des résultat possibles est $\{\text{pile}, \text{face}, \text{tranche}\}$ ou plus simplement, si l'on néglige la possibilité que la pièce tombe sur la tranche, $\{\text{pile}, \text{face}\}$. Si la pièce est équilibrée, il est alors naturel de choisir les probabilités que la pièce tombe sur pile ou face égales à un demi. À la question de quel coté va tomber la pièce, la réponse du probabiliste n'est alors pas déterministe mais statistique : la pièce a une chance sur deux de tomber sur pile ou sur face.

Dans les prochains paragraphes, nous précisons le formalisme général de la théorie des probabilités, c'est-à-dire le cadre mathématique rigoureux dans lequel se formule cette théorie. Ce formalisme a été introduit au début du vingtième siècle par le mathématicien russe A. Kolmogorov.

Définition 1.2.1. Un espace de probabilités est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, où Ω est un ensemble, $\mathcal{F}$ une tribu, et $\mathbb{P}$ une mesure de probabilité.

L'objet des prochains paragraphes et de donner la définition et le rôle de chacun des éléments de ce triplet.

1.2.1 Univers des possibles

Comme indiqué ci-dessus, le premier élément Ω d'un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ est un ensemble. Plus précisément, on a la définition suivante :

Définition 1.2.2. Étant donnée une expérience aléatoire, on appelle *univers des possibles*, et l'on note souvent Ω, l'ensemble des résultats possibles de l'expérience.

La description explicite de l'ensemble Ω est la première étape fondamentale dans la modélisation d'un phénomène aléatoire. Comme nous le verrons plus loin, le choix de Ω n'est pas toujours unique. Les pseudo-paradoxes qui apparaissent parfois entre deux protagonistes concernant une expérience où intervient le hasard relèvent le plus souvent de deux choix distincts d'ensembles des possibles. Aussi est-t-il important de bien choisir l'ensemble Ω avec lequel on travaille, et de se tenir à ce choix.

Exemple 1.2.3. Voici quelques expériences aléatoires et les ensembles des possibles correspondants :

1. On jette un dé. L'ensemble Ω est alors l'ensemble $\{1, 2, 3, 4, 5, 6\}$ à 6 éléments. Ici, l'élément $\omega = 2 \in \Omega$ signifie que la face visible du dé après le lancer est 2.
2. On jette deux dés. L'ensemble Ω est alors l'ensemble $\{1, 2, 3, 4, 5, 6\}^2$ c'est-à-dire $\Omega = \{(i, j),\ i, j \in \{1, 2, 3, 4, 5, 6\}\} = \{(1, 1), (2, 1), (3, 6), \ldots\}$. L'élément $\omega = (3, 5) \in \Omega$ correspond à un lancer où le premier dé donne 3 et le second dé donne 5 ;
3. On joue dix fois à pile ou face. On a alors $\Omega = \{\text{pile}, \text{face}\}^{10}$. On peut aussi choisir pour ensemble des possibles $\Omega' = \{\text{pile}, \text{face}, \text{tranche}\}^{10}$ si l'on veut tenir compte du fait que la pièce peut tomber sur la tranche ;
4. On fait un sondage auprès de 1000 personnes à la sortie d'un référendum. On a alors $\Omega = \{\text{oui}, \text{non}, \text{blanc}\}^{1000}$;
5. On distribue une main au poker. L'ensemble des possibles correspondant à cette expérience est alors $\Omega = \{\text{choix de 5 cartes parmi 52}\}$ qui a pour cardinal le coefficient binomial $\binom{52}{5}$.

Remarque 1.2.4. Il n'est pas toujours possible de décrire de façon rigoureuse l'univers des possibles. On peut penser par exemple à l'expérience aléatoire de la météo du lendemain ! Néanmoins, dans les cas simples que nous envisagerons dans la suite, on peut la plupart du temps décrire explicitement l'ensemble Ω.

1.2.2 Tribu et évènements

Dans la suite, on va vouloir calculer la probabilité de certaines parties de l'ensemble des possibles Ω. Par exemple, lorsque l'on jette deux dés, l'ensemble des possibles Ω est $\{1,2,3,4,5,6\}^2$, et l'on voudrait calculer la probabilité que le premier dé donne 2 et le second est impair, c'est-à-dire la probabilité de l'ensemble :

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

Définition 1.2.5. On appelle *tribu* et on note $\mathcal{F}$ l'ensemble des parties de Ω dont on pourra calculer la probabilité. Lorsque l'ensemble Ω est fini ou dénombrable, on choisira pour $\mathcal{F}$ l'ensemble de toutes les parties de Ω c'est-à-dire :

$$\mathcal{F} = \mathcal{P}(\Omega).$$

Définition 1.2.6. Les éléments de $\mathcal{F} = \mathcal{P}(\Omega)$ sont appelés des *évènements*. On dit encore que ce sont des ensembles *mesurables* par rapport à la tribu $\mathcal{F}$.

Le texte qui suit, en miniature, pourra être ommis par le lecteur, il concerne la "vraie" définition de la notion de tribu. En effet, sauf lorsque Ω est fini ou dénombrable, on ne peut pas s'intéresser à l'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω, celui-ci étant en quelque sorte "trop gros". On se restreindra donc à un sous-ensemble $\mathcal{F}$ de $\mathcal{P}(\Omega)$, qui constituera l'ensemble des parties dont on peut calculer la probabilité. Afin d'obtenir un modèle aussi cohérent que possible, il importe néanmoins d'imposer certaines conditions de stabilité à l'ensemble $\mathcal{F}$: par union, intersection, passage au complémentaire, etc. Aussi, voici la "vraie" notion de tribu.

Définition 1.2.7. Soit Ω un ensemble et $\mathcal{F}$ un sous-ensemble de parties de Ω, *i.e.* $\mathcal{F} \subset \mathcal{P}(\Omega)$. On dit que $\mathcal{F}$ est une tribu si elle vérifie les 3 conditions suivantes :

1. $\Omega \in \mathcal{F}$;
2. si A appartient à $\mathcal{F}$, alors son complémentaire A^c appartient aussi à $\mathcal{F}$;
3. si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{F}$, alors $\bigcup_{n=0}^{\infty} A_n$ appartient à $\mathcal{F}$.

On vérifie sans problème à partir des trois axiomes ci-dessus que toute tribu $\mathcal{F}$ contient l'ensemble vide $\emptyset$, est stable par union finie, intersection finie ou dénombrable. Ainsi, on retiendra qu'une tribu est stable par combinaisons au plus dénombrables d'opérations usuelles sur les ensembles, bref par toutes les manipulations classiques.

Exemple 1.2.8. Voici trois exemples classiques de tribus :
- La tribu triviale : $\mathcal{F} = \{\emptyset, \Omega\}$;
- La tribu engendreée par une partie A de Ω : $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$;
- La tribu pleine : $\mathcal{F} = \mathcal{P}(\Omega)$.

Exemple 1.2.9. On jette deux dés discernables. L'ensemble des résultats possibles est alors

$$\Omega = \{(i,j),\ i,j \in \{1,2,3,4,5,6\}\}.$$

La tribu engendrée par le singleton $\{(1,1)\}$ est composée des quatre évènements $\{\emptyset, (1,1), \Omega\backslash(1,1), \Omega\}$. Si on choisit la tribu pleine $\mathcal{F} = \mathcal{P}(\Omega)$, l'évènement "la somme des deux dés est supérieure ou égale à dix" correspond à l'ensemble $\{(5,5),(5,6),(6,5)\}$; si on introduit les deux ensembles

$$A = \{\text{les deux dés sont pairs}\},\ \text{ et }\ B = \{\text{les deux sont distincts}\},$$

alors $A \cap B$ correspond à l'évènement $\{(2,4),(4,2),(2,6),(6,2),(4,6),(6,4)\}$.

En pratique, lorsque Ω est fini ou dénombrable, on considère donc en général la tribu pleine $\mathcal{P}(\Omega)$. En revanche, si Ω n'est pas dénombrable, comme c'est le cas dans l'exemple d'une suite infinie de lancers ($\Omega = \{\text{pile}, \text{face}\}^{\mathbb{N}}$), on ne considérera pas la tribu $\mathcal{F} = \mathcal{P}(\Omega)$, mais une tribu plus petite.

Le couple $(\Omega, \mathcal{F})$ est appelée espace mesurable ou encore espace probabilisable. Pour compléter la description de la notion d'espace de probabilités, il nous reste à introduire la notion de mesure de probabilité. C'est l'objet du prochain paragraphe.

1.2.3 Probabilité

Une fois fixés un univers Ω et une tribu $\mathcal{F}$, on peut définir proprement ce qu'est une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F})$ et par suite un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$: à chaque évènement, on associe un nombre positif compris entre 0 et 1, sa probabilité.

Définition 1.2.10. On appelle probabilité sur $(\Omega, \mathcal{F})$ une application $\mathbb{P}$ de $\mathcal{F}$ dans l'intervalle $[0, 1]$ telle que :

1. $\mathbb{P}(\Omega) = 1$;
2. pour toute famille au plus dénombrable d'évènements deux à deux disjoints $(A_n)_{n\geqslant 0}$ on a $\mathbb{P}(\bigcup_{n\geqslant 0} A_n) = \sum_{n\geqslant 0} \mathbb{P}(A_n)$.

Le triplé $(\Omega, \mathcal{F}, P)$ est alors appelé *espace de probabilités* ou encore *espace probabilisé*.

De l'axiomatique de Kolmogorov, on déduit aisément les propriétés suivantes :

Proposition 1.2.11. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités. Alors on a*

1. $\mathbb{P}(\emptyset) = 0$;
2. *pour tout $A \in \mathcal{F}$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;*
3. *pour tout $A, B \in \mathcal{F}$ tels que $A \subset B$, $\mathbb{P}(A) \leqslant \mathbb{P}(B)$;*
4. *pour tout $A, B \in \mathcal{F}$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;*
5. *Si $(A_n)_{n\geqslant 0}$ est une suite d'évènements, alors $\mathbb{P}(\bigcup_{n\geqslant 0} A_n) \leqslant \sum_{n\geqslant 0} \mathbb{P}(A_n)$. Il n'y a égalité que si les évènements A_n sont deux à deux disjoints.*

Proposition 1.2.12 (Continuité monotone séquentielle). *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités et $(A_n)_{n\geqslant 0}$ une suite d'évènements.*

1. *Si la suite A_n est croissante, c'est-à-dire si $A_0 \subset A_1 \subset \ldots \subset A_n \subset \ldots$, alors*

$$\mathbb{P}\left(\bigcup_{n\geqslant 0} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) ;$$

2. *Si la suite A_n est décroissante, c'est-à-dire si $A_0 \supset A_1 \supset \ldots \supset A_n \supset \ldots$, alors*

$$\mathbb{P}\left(\bigcap_{n\geqslant 0} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

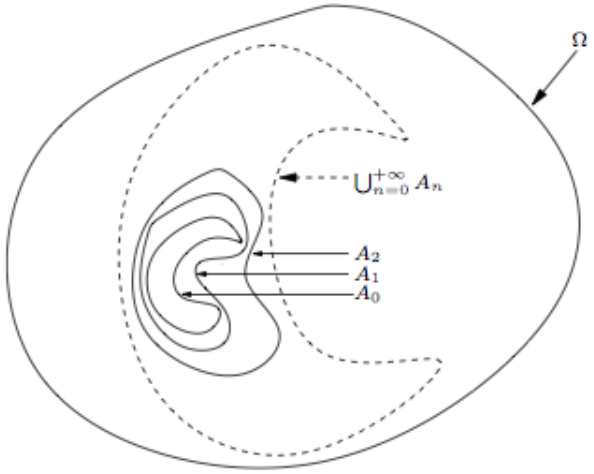


FIGURE 1.4 – Une famille croissante d'ensembles.

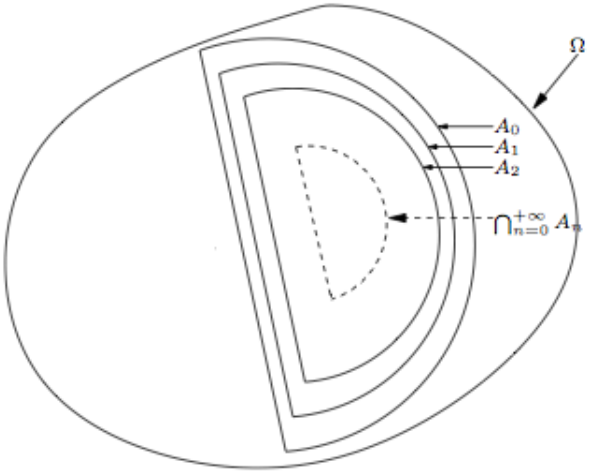


FIGURE 1.5 – Suite décroissante d'ensembles.

Quelques exemples de calculs de probabilités

Exercice 1 :
Dix athlètes participent à une course que chacun a la même chance d'emporter (pas d'ex aequo). Ils portent des dossards numérotés de 1 à 10. Quelle est la probabilité que l'un des coureurs portant les numéros 1, 2 ou 3 l'emporte ?

Correction : On note A_i l'évènement "le coureur au dossard i l'emporte" et A l'évènement "un des coureurs portant les numéros 1, 2 ou 3 l'emporte". L'évènement A s'écrit simplement comme l'union $A = A_1 \cup A_2 \cup A_3$ et les trois évènements sont disjoints donc

$$\mathbb{P}(A) = \mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) = \frac{1}{10} + \frac{1}{10} + \frac{1}{10} = \frac{3}{10}.$$

Exercice 2 :
Un sac contient des billes noires et rouges, portant une marque ou non. La probabilité d'observer une bille rouge et marquée est de 2/10, une bille marquée de 3/10 et une bille noire de 7/10. Quelle est la probabilité d'observer une bille rouge ou marquée ?

Correction : On note R pour rouge, N pour noire, M pour marquée et M^c pour non marquée. On cherche la probabilité de l'évènement $R \cup M$. On a

$$\mathbb{P}(R \cup M) = \mathbb{P}(R) + \mathbb{P}(M) - \mathbb{P}(R \cap M) = \frac{3}{10} + \frac{3}{10} - \frac{2}{10} = \frac{4}{10}.$$

Exercice 3 :
Lors d'une loterie de Noël, 300 billets sont vendus aux parents d'élèves d'une école ; 4 billets sont gagnants. J'achète 10 billets, quelle est la probabilité pour que je gagne au moins un lot ?

Correction : L'univers des possibles est ici l'ensemble des combinaisons de 10 billets parmi les 300 ; il y en a $\binom{300}{10}$. Je ne 10 gagne rien si les 10 billets achetés se trouvent parmi les 296 billets perdants, ceci arrive avec la probabilité :

$$q = \frac{\binom{296}{10}}{\binom{300}{10}}.$$

La probabilité p cherchée est celle de l'évèment complémentaire :

$$p = 1 - q = 1 - \frac{\binom{296}{10}}{\binom{300}{10}} \approx 0.127.$$

Proposition 1.2.13. *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'évènements qui constituent une partition de l'ensemble Ω c'est-à-dire $\Omega = \bigsqcup_{n\in\mathbb{N}} A_n$. Alors pour tout $B \in \mathcal{F}$, on a*

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

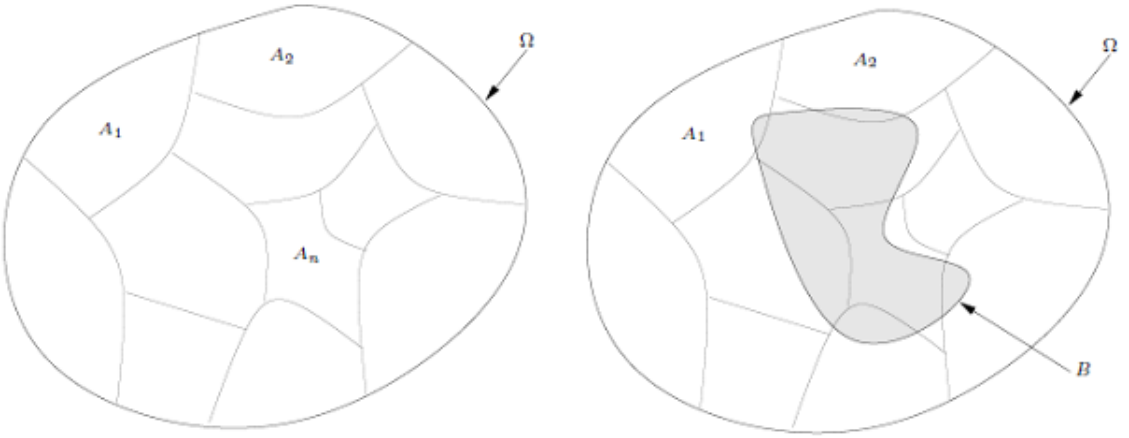


FIGURE 1.6 – Probabilité et partition.

Exemple 1.2.14. Un étudiant a les probabilités suivantes d'avoir la note i à un module, le module étant noté sur 10 *i.e.* $i = 1 \dots 10$. Quelle est la probabilté qu'il valide son module, c'est-à-dire qu'il obtienne une note supérieure ou égale à 5 ?

Note	0	1	2	3	4	5	6	7	8	9	10
Proba	1/11	0	0	1/11	1/11	2/11	2/11	2/11	1/11	1/11	0

On note B l'évèment "il valide son module" et A_i l'évèment "il obtient la note i. Les A_i forment une partition de l'ensemble des notes possibles et l'on a donc :

$$\mathbb{P}(B) = \sum_{i=0}^{10} \mathbb{P}(B \cap A_i) = 0 + \sum_{i=5}^{10} \mathbb{P}(A_i) = 8/11.$$

Remarque 1.2.15. Étant donné un espace probabilisable $(\Omega, \mathcal{F})$, le choix de la probabilité $\mathbb{P}$ n'est bien sûr pas unique. Ce choix doit se faire en accord avec l'expérience aléatoire que l'on souhaite modéliser. Par exemple, si on joue à pile ou face et que l'on précise que la pièce est équilibrée, on choisira naturellement $\mathbb{P}$ de sorte que

$$\mathbb{P}(\text{pile}) = \mathbb{P}(\text{face}) = 1/2.$$

En revanche, si l'on précise que la pièce est truquée, on préférera choisir $\mathbb{P}$ de sorte que $\mathbb{P}(\text{pile}) \neq \mathbb{P}(\text{face})$.

Remarque 1.2.16. Au risque de se répéter, insistons sur le fait que dans la modélisation d'une expérience aléatoire, l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ avec lequel on travaille n'est a priori pas unique. Il résulte d'un choix, et que ce choix doit pouvoir être justifié :

- le choix de l'ensemble des possibles Ω n'est pas unique, pensez au jeu de pile ou face avec $\Omega_1 = \{\text{pile}, \text{face}\}$ et $\Omega_2 = \{\text{pile}, \text{face}, \text{tranche}\}$;
- le choix de la tribu n'est pas unique, on peut choisir la tribu pleine, la tribu engendrée par un évènement, etc. ;
- le choix de la probabilité $\mathbb{P}$ n'est pas unique comme indiqué dans la remarque précédente.

L'exemple suivant est caractéristique. L'énoncé est n'est pas assez précis, de sorte que plusieurs choix de modélisations sont possibles et donc plusieurs réponses à la question posée sont envisageables. Il n'y a pas une réponse meilleure que l'autre : elles répondent à des questions différentes !

Exemple 1.2.17. On tire une corde au hasard dans un disque de rayon R. Quelle est la probabilité que la longueur ℓ de la corde soit supérieure à R ?

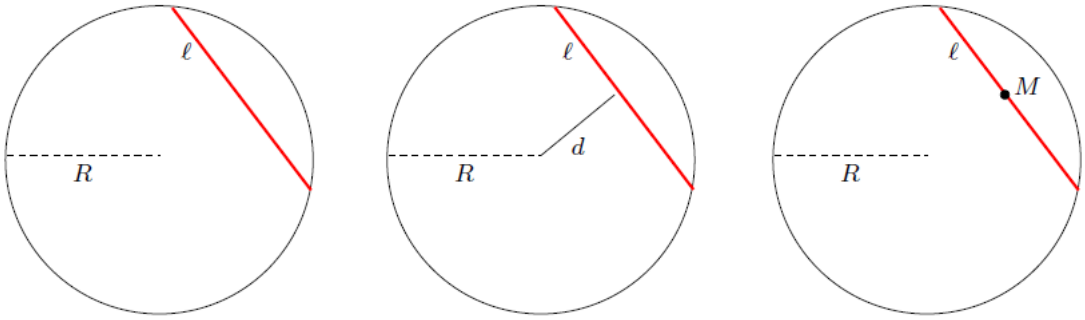


FIGURE 1.7 – Dans les trois exemples, on tire uniformément selon la longueur, la distance au centre, le milieu de la corde.

1. Dans ce premier cas, on choisit ici de modéliser le hasard en supposant que longueur de la corde est choisie uniformément parmi toutes les longueurs possibles. La longueur ℓ varie ici continûment dans $[0, 2R]$, de sorte que la probabilité cherchée vaut $1/2$;
2. On décide maintenant de modéliser le hasard en supposant que c'est la distance au centre de la corde qui est choisie uniformément au hasard. La longueur ℓ est donc déterminée par la distance d de la corde au centre du disque. Ici, d varie continûment dans $[0, R]$, et $\ell = \sqrt{R^2 - d^2} \geqslant R \Leftrightarrow d \leqslant \sqrt{3}/2R$, de sorte que la probabilité cherchée vaut $\sqrt{3}/2$;
3. Enfin, on décide de modéliser l'expérience en supposant que c'est la milieu M de la corde qui est choisi uniformément dans le disque. Dans ce cas, $\ell \geqslant R$ a lieu *ssi* M est dans le disque concentrique de rayon $\sqrt{3}/2$ de sorte que la probabilité cherchée vaut $3/4$.

1.3 Exemples d'espaces de probabilités

Afin de se familiariser avec les notions introduites ci-dessus, on donne maintenant des exemples d'expériences aléatoires et les espaces de probabilités correspondants.

1.3.1 Probabilité uniforme sur un ensemble fini

Loi uniforme pour un dé

Reprenons l'exemple du lancer de dé. On a vu que l'univers des possibles est est $\Omega = \{1, 2, 3, 4, 5, 6\}$ de cardinal 6. On munit Ω de la tribu des parties $\mathcal{F} = \mathcal{P}(\Omega)$. On vérifie alors que l'application

$$\mathbb{P} : \mathcal{F} \to [0, 1], \qquad A \mapsto \mathbb{P}(A) := \frac{\mathrm{Card}(A)}{6},$$

est bien une mesure de probabilité. Ainsi, dans cette modélisation, la probabilité d'obtenir un chiffre plus grand que 5 avec un lancer est

$$\mathbb{P}(\{5\} \cup \{6\}) = \mathbb{P}(\{5, 6\}) = \frac{2}{6} = \frac{1}{3}.$$

Loi uniforme sur un ensemble fini

Plus généralement, dès qu'on considère une expérience aléatoire où $\mathrm{Card}(\Omega)$, le nombre de résultats possibles, est fini, et que parmi ces résultats aucun n'est privilégié, on choisira naturellement la tribu des parties $\mathcal{F} = \mathcal{P}(\Omega)$ et la probabilité dite *uniforme* définie de la façon suivante :

$$\mathbb{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}, \quad \text{pour tout } A \in \mathcal{F}.$$

Par exemple, si on joue trois fois de suite à pile ou face (on note $\mathtt{p}, \mathtt{f}$ pour simplifier) avec une pièce équilibrée, l'ensemble des possibles est $\Omega = \{\mathtt{p}, \mathtt{f}\}^3$ qui a pour cardinal $2^3 = 8$. Notons A l'évènement le premier et le troisième lancer donnent pile, c'est-à-dire $A = \{(\mathtt{p}, \mathtt{p}, \mathtt{p}), (\mathtt{p}, \mathtt{f}, \mathtt{p})\}$. Alors

$$\mathbb{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} = \frac{2}{2^3} = \frac{1}{4}.$$

Remarque 1.3.1. La probabilité uniforme sur un ensemble fini est encore appelée *équiprobabilité*. On dit alors que tous les évènements élémentaires $\omega \in \Omega$ sont équiprobables.

1.3.2 Probabilité sur un ensemble au plus dénombrable

Loi générale sur un ensemble fini

On a vu que lorsqu'on a équiprobabilité sur un univers fini, la mesure de probabilité $\mathbb{P}$ est celle qui à tout évènement A associe le rapport de son cardinal au

cardinal de Ω. En d'autres termes $\Omega = \{\omega_1, \ldots, \omega_n\}$ et pour tout $i = 1, \ldots, n$: $p_i = \mathbb{P}(\{\omega_i\}) = 1/n$. Supposer que l'on n'a pas équiprobabilité des évènements élémentaires ω_i revient à considérer une suite $(p_1, \ldots, p_n)$ de nombres positifs et sommant à 1, mais dont tous les coefficients p_i ne sont pas égaux. On définit alors encore une mesure de probabilité sur $\mathcal{P}(\Omega)$ en considérant pour tout évènement $A \in \mathcal{P}(\Omega)$:

$$\mathbb{P}(A) = \sum_{i, \omega_i \in A} p_i$$

où la notation "$i, \omega_i \in A$" signifie que la somme est effectuée sur l'ensemble des indices i pour lesquels ω_i appartient à A.

Exemple 1.3.2. On lance 3 fois de suite une pièce équilibrée et on compte le nombre de fois où pile est apparu. On a donc $\Omega = \{0, 1, 2, 3\}$, mais il n'y a pas équiprobabilité puisque les probabilités élémentaires sont $(1/8, 3/8, 3/8, 1/8)$.

Exemple 1.3.3. On lance deux dés équilibrés et on note S la somme des deux lancers. L'ensemble des valeurs possibles pour S est $\Omega = \{2, 3, \ldots, 11, 12\}$. Les probabilités pour les valeurs possibles de S sont alors :

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(S = k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Loi sur un ensemble dénombrable

Si on veut construire une probabilité $\mathbb{P}$ sur un ensemble infini dénombrable, typiquement sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, on ne peut plus avoir équiprobabilité des évènements élémentaires $\{n\}$. Supposons en effet que pour tout $n \in \mathbb{N}$ on ait $\mathbb{P}(\{n\}) = p > 0$, alors l'additivité de $\mathbb{P}$ imposerait que :

$$\mathbb{P}(\mathbb{N}) = \sum_{n \geqslant 0} \mathbb{P}(\{n\}) = \sum_{n \geqslant 0} p = +\infty$$

ce qui est en contradiction avec la condition $\mathbb{P}(\mathbb{N}) = 1$. Une façon de construire une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ est de généraliser le procédé que l'on vient de voir pour les ensembles finis : considérer une suite $(p_n)_{n \geqslant 0}$ de nombres positifs telle que la série $\sum_{n \geqslant 0} p_n$ soit convergente et de somme 1. Comme précédemment, on définit alors pour tout événement $A \in \mathcal{P}(\mathbb{N})$:

$$\mathbb{P}(A) = \sum_{n, n \in A} p_n.$$

Exemple 1.3.4. On lance une pièce équilibrée jusqu' à ce que pile apparaisse (toujours en excluant le cas improbable où pile n'apparaît jamais). On a donc $\Omega = \{1, 2, \ldots\} = \mathbb{N}^*$. On a clairement $p_1 = \mathbb{P}(\{1\}) = 1/2$, $p_2 = \mathbb{P}(\{2\}) = 1/4$ et de façon générale $p_n = \mathbb{P}(\{n\}) = 1/2^n$. On reconnaît dans les p_n les termes d'une suite géométrique dont la somme vaut bien 1 :

$$\sum_{n \geqslant 1} 2^{-n} = 1.$$

1.3.3 Espace de probabilités continu

Donnons à présent quelques exemples de probabilités sur des espaces continus. Considérons ainsi un intervalle $\Omega =]a, b[\subset \mathbb{R}$.

On est ici dans un cas où l'ensemble Ω n'est pas dénombrable et où la tribu $\mathcal{P}(\Omega)$ est "trop grosse". Aussi, on considère une tribu $\mathcal{F}$ plus "petite", celle formée des intersections / unions dénombrables d'intervalles du type $[c, d[$ (on parle de tribu *borelienne*).

Supposons que l'on dispose d'une fonction positive f définie sur l'intervalle $[a, b]$ et telle que

$$\int_a^b f(x)dx = 1.$$

On peut alors définir une probabilité $\mathbb{P}$ sur $\mathcal{F}$ de la façon suivante : pour tout intervalle $A = [c, d[$ dans $[a, b]$

$$\mathbb{P}(A) = \int_A f(x)dx = \int_c^d f(x)dx.$$

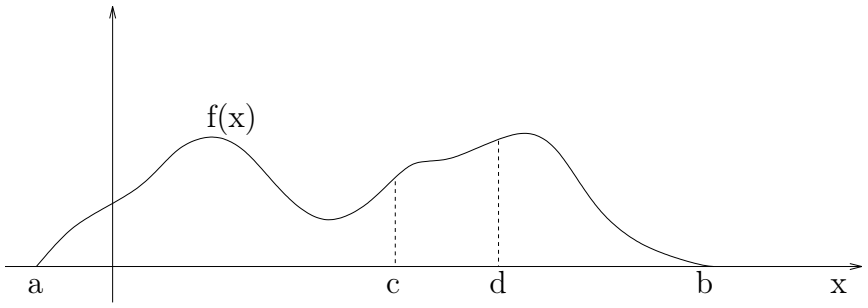


FIGURE 1.8 – Probabilité sur un intervalle via une densité.

Exemple 1.3.5 (Probabilité uniforme continue)**.** Un bus est censé passer toutes les dix minutes à République pour se rendre à Beaulieu. Un passager arrive à l'arrêt de bus. On cherche à modéliser son temps d'attente T. A priori, on peut supposer que ce temps d'attente est dans l'intervalle $\Omega = [0, 10]$. On munit cet ensemble de la tribu borélienne. N'ayant pas d'information sur l'heure théorique de passage du bus et l'heure d'arrivée du passager, on peut supposer que le temps d'attente est uniforme, *i.e.* pour tout $0 < c < d < 10$:

$$\mathbb{P}(T \in [c, d[) = \frac{1}{10}|d - c| = \int_c^d f(x)dx$$

où la fonction f est constante égale à $1/10$ sur l'intervalle $[0, 10]$ de sorte que $\int_0^{10} f(x)dx = 1$.

Exemple 1.3.6. On cherche à modéliser le temps de demi-vie d'un atome radioactif. Ce temps T est aléatoire et l'expérience montre qu'il peut être très grand. On supposera que T est à valeurs dans $\Omega = \mathbb{R}^+ = [0, +\infty[$. Là encore, on suppose Ω muni de sa tribu borélienne. Des considérations physiques montrent que la probabilité ci-dessous décrit bien le temps de demi-vie T :

$$\mathbb{P}(T \in [c, d[) = \int_c^d e^{-x} dx, \quad \text{pour } 0 < c < d < +\infty.$$

Chapitre 2

Indépendance et conditionnement

Nous introduisons à présent deux notions fondamentales en théorie des probabilités. La première, le conditionnement, permet de prendre en compte une information supplémentaire dans le calcul d'une probabilité. La seconde, l'indépendance, rend compte du fait que deux évènements n'ont aucune incidence l'un sur l'autre, et donc que l'on peut évaluer la probabilité du premier indépendamment du fait que le second ait lieu ou non.

2.1 Probabilité conditionnelle

La notion de conditionnement nous sera très utile dans la suite puisqu'elle permet par exemple de tenir compte de l'information dont on dispose déjà pour évaluer la probabilité d'un nouvel évènement. Même en l'absence de toute chronologie sur les évènements, un détour par un conditionnement astucieux nous permettra souvent d'arriver à nos fins.

2.1.1 Définition

Dans tout ce qui suit, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilités arbitraire et tous les ensembles considérés sont des évènements de la tribu $\mathcal{F}$. Nous commençons par définir la probabilité conditionnelle sachant un évènement.

Définition 2.1.1 (Probabilité conditionnelle)**.** Soit A un évènement tel que $\mathbb{P}(A) > 0$. Pour tout évènement B, on définit la probabilité de B sachant A par :

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

On définit ainsi une nouvelle probabilité sur $(\Omega, \mathcal{F})$, notée $\mathbb{P}(.|A)$ ou encore $\mathbb{P}_A(.)$, et appelée probabilité conditionnelle sachant A.

La vérification que $\mathbb{P}(.|A)$ est bien une probabilité, *i.e.* vérifie bien les critères de la définition 1.2.10 est laissée en exercice.

Concrètement, l'expression "probabilité de B sachant A" signifie "probabilité que B se réalise sachant que A s'est réalisé". La probabilité de B peut être faible alors que la probabilité de B sachant A est grande (et réciproquement).

Exemple 2.1.2. Une urne contient 90 boules noires, 9 boules blanches et 1 boule rouge. On tire une boule au hasard : quelle est la probabilité qu'elle soit blanche ? La réponse est bien sûr $\mathbb{P}(B) = 9/100$, donc une probabilité faible. On tire une boule au hasard : quelle est la probabilité qu'elle soit blanche, sachant que la boule tirée n'est pas noire ? Si on note A l'évènement "La boule tirée n'est pas noire", on a donc $\mathbb{P}(A) = 1/10$ et la réponse à la question est :

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = 9/10,$$

donc une grande probabilité.

On donne maintenant quelques propriétés relatives au conditionnement.

Proposition 2.1.3 (Inversement du conditionnement). *Soient A et B deux évènements tels que $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$. Alors on a la relation suivante :*

$$\mathbb{P}(B|A) = \mathbb{P}(A|B) \times \frac{\mathbb{P}(B)}{\mathbb{P}(A)}.$$

Démonstration. Il suffit d'appliquer deux fois la définition de la probabilité conditionnelle :

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

□

Proposition 2.1.4 (Formule des probabilités composées). *Soit A_0, A_1, ..., A_n une suite d'évènements ayant une intersection commune non nulle, i.e. $\bigcap_{k=0}^{n} A_k \neq \emptyset$, on a alors*

$$\mathbb{P}\left(\bigcap_{k=0}^{n} A_k\right) = \mathbb{P}(A_0)\mathbb{P}(A_1|A_0)\mathbb{P}(A_2|A_0 \cap A_1)\ldots\mathbb{P}(A_n|A_0 \cap A_1 \ldots \cap A_{n-1})$$

Démonstration. On commence par noter que tous les conditionnements sont justifiés puisque par monotonie :

$$0 < \mathbb{P}(A_0 \cap \ldots \cap A_{n-1}) \leqslant \mathbb{P}(A_0 \cap \ldots \cap A_{n-2}) \leqslant \ldots \leqslant (A_0 \cap A_1) \leqslant \mathbb{P}(A_0).$$

Il reste à remarquer qu'en développant les termes du produit via la définition de la probabilité conditionnelle $\mathbb{P}(B|A) = \mathbb{P}(B \cap A)/\mathbb{P}(A)$, tous se télescopent sauf le dernier. □

Remarque 2.1.5. On peut se servir de ce résultat comme d'une poupée russe : soit à calculer $\mathbb{P}(A_n)$, on introduit une suite croissante d'évènements $A_0 \subset A_1 \subset \ldots \subset A_n$ et la formule devient tout simplement :

$$\mathbb{P}(A_n) = \mathbb{P}(A_0)\mathbb{P}(A_1|A_0)\mathbb{P}(A_2|A_1)\ldots\mathbb{P}(A_n|A_{n-1}).$$

Sous les mêmes hypothèses que celles de la proposition 1.2.13, on a la proposition suivante :

Proposition 2.1.6 (Formule des probabilités totales). *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'évènements qui constituent une partition de l'ensemble Ω c'est-à-dire $\Omega = \bigsqcup_{n\in\mathbb{N}} A_n$. Alors pour tout $B \in \mathcal{F}$, on a*

$$\mathbb{P}(B) = \sum_n \mathbb{P}(B|A_n)\mathbb{P}(A_n).$$

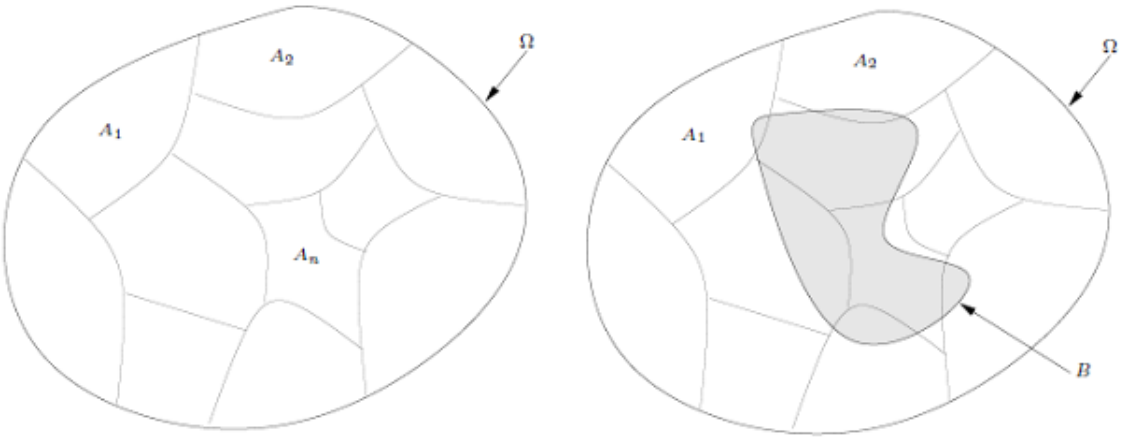


FIGURE 2.1 – Probabilité conditionnelle et partition.

Remarque 2.1.7. En pratique, on utilise souvent cette formule des probabilités totales en conditionnant successivement par un évènement et son contraire, c'est-à-dire en prenant tout simplement une partition de Ω du type $\Omega = A \sqcup A^c$, ce qui donne

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

Considérons l'exemple d'une urne qui contient des boules blanches et noires, marquées ou non. On suppose que parmi les boules marquées il y a 30% de boules blanche et parmi les non marquées 60%. Par ailleurs, on sait que 80% des boules sont marquées. Quelle est la probabilité de tirer une boule blanche ? On note B pour blanche et A pour marquée, alors

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c) = \frac{30}{100} \times \frac{80}{100} + \frac{60}{100} \times \frac{20}{100} = \frac{36}{100}.$$

2.1.2 Formule Bayes

De la fomule d'inversement du conditionnement et de la formule des probabilités totales, on déduit la formule de Bayes :

Proposition 2.1.8 (Formule de Bayes). *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'évènements qui constituent une partition de l'ensemble Ω c'est-à-dire $\Omega = \bigsqcup_{n\in\mathbb{N}} A_n$. Alors pour tout $B \in \mathcal{F}$, on a*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_n \mathbb{P}(B|A_n)\mathbb{P}(A_n)}.$$

Remarque 2.1.9. Lorsque la partition de Ω est du type $\Omega = A \sqcup A^c$, la formule de Bayes s'écrit simplement

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

Exemple 2.1.10. Deux machines M_1 et M_2 produisent respectivement 100 et 200 objets. M_1 produit 5% de pièces défectueuses et M_2 en produit 6%. Quelle est la probabilité pour qu'un objet défectueux ait été fabriqué par la machine M_1 ? L'évènement constaté, que l'on note A, est la présence d'une pièce défectueuse et les causes sont les machines M_1 et M_2. Compte tenu des productions de ces machines, on a $\mathbb{P}(M_1) = \frac{1}{3}$ et $\mathbb{P}(M_2) = \frac{2}{3}$. De plus, les probabilités conditionnelles de l'évènement A selon les machines sont $\mathbb{P}(A|M_1) = \frac{5}{100}$ et $\mathbb{P}(A|M_2) = \frac{6}{100}$. En reportant ces valeurs dans la formule générale, on obtient

$$\mathbb{P}(M_1|A) = \frac{\frac{1}{3}\times\frac{5}{100}}{\left(\frac{1}{3}\times\frac{5}{100} + \frac{2}{3}\times\frac{6}{100}\right)} = \frac{5}{17} \approx 0.29$$

Exemple 2.1.11. Le quart d'une population est vacciné contre le choléra. Au cours d'une épidémie, on constate qu'il y a parmi les malades un vacciné pour 4 non-vaccinés, et qu'il y a un malade sur 12 parmi les vaccinés. Quelle est la probabilité qu'un non-vacciné tombe malade ?
On note V pour vacciné, NV pour non vacciné, M pour malade, S pour sain. D'après les hypothèses,

$$\mathbb{P}(V) = \frac{1}{4}, \quad \mathbb{P}(V \mid M) = \frac{1}{5}, \quad \mathbb{P}(NV \mid M) = \frac{4}{5}, \quad \mathbb{P}(M \mid V) = \frac{1}{12}.$$

Par définition, on a

$$\mathbb{P}(M \mid NV) = \frac{\mathbb{P}(NV \cap M)}{\mathbb{P}(NV)} = \frac{\mathbb{P}(NV \mid M)\mathbb{P}(M)}{1 - \mathbb{P}(V)} = \frac{16}{15}\mathbb{P}(M).$$

Or

$$\mathbb{P}(M \mid V) = \frac{\mathbb{P}(V \cap M)}{\mathbb{P}(V)} = \frac{\mathbb{P}(V \cap M)}{1/4} = \frac{1}{12} \qquad \text{donc} \quad \mathbb{P}(V \cap M) = 1/48.$$

$$\mathbb{P}(V \mid M) = \frac{\mathbb{P}(V \cap M)}{\mathbb{P}(M)} = 1/5 \qquad \text{donc} \quad \mathbb{P}(M) = \frac{5}{48}$$

Finalement

$$\mathbb{P}(M \mid NV) = \frac{16}{15} \times \frac{5}{48} = \frac{1}{9}.$$

2.2 La notion d'indépendance

La notion d'indépendance intervient de façon constante en probabilités. Intuitivement, deux évènements sont indépendants si la réalisation de l'un "n'a aucune influence" sur la réalisation ou non de l'autre. Le but de cette section est de préciser ceci mathématiquement et de l'étendre cette notion à plus de deux évènements. Dans toute la suite, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé fixé.

2.2.1 Indépendance de deux évènements

Définition 2.2.1 (Indépendance de deux évènements)**.** On dit que deux évènements A et B sont indépendants, et on note $A \perp B$, si

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

Si A est tel que $\mathbb{P}(A) > 0$, l'indépendance de A et B s'écrit encore $\mathbb{P}(B|A) = \mathbb{P}(B)$ et on retrouve la notion intuitive d'indépendance : le fait que A se soit réalisé ne change rien quant à la probabilité que B se réalise.

Exemple 2.2.2. Voici quelques exemples d'évènements indépendants ou non :

1. On lance un dé deux fois de suite. Soit A l'évènement : "Le premier lancer donne un nombre pair" et B l'évènement : "Le second lancer donne un nombre pair". L'univers naturel est $\Omega = \{(i, j), 1 \leqslant i, j \leqslant 6\}$, ensemble à 36 éléments muni de la probabilité uniforme. Il est clair que $\mathbb{P}(A) = \mathbb{P}(B) = 18/36 = 1/2$ et que :
$$\mathbb{P}(A \cap B) = 9/36 = 1/4 = \mathbb{P}(A)\mathbb{P}(B),$$
donc A et B sont indépendants.
2. On tire une carte au hasard d'un jeu de 32 cartes. Soit A l'évènement : "La carte tirée est un 7" et B l'évènement : "La carte tirée est un pique". On a $\mathbb{P}(A) = 1/8$ et $\mathbb{P}(B) = 1/4$. L'évènement $A \cap B$ correspond au tirage du sept de pique $\mathbb{P}(A \cap B) = 1/32$. Ainsi on a
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$
et les évènements A et B sont indépendants.
3. On joue deux fois de suite à pile ou face, $\Omega = \{\text{pile}, \text{face}\}$. On désigne par A et B les évènements "on obtient deux fois pile" et "on obtient au moins une fois pile". Alors $\mathbb{P}(A) = 1/4$, $\mathbb{P}(B) = 3/4$, et $\mathbb{P}(A \cap B) = 1/4$. On a donc $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$ et les deux évènements ne sont pas indépendants.

Proposition 2.2.3. *Si A et B sont indépendants, alors il en va de même pour :*
- *les évènements A^c et B ;*
- *les évènements A et B^c ;*
- *les évènements A^c et B^c ;*

2.2.2 Indépendance de n évènements

Définition 2.2.4 (Indépendance 2 à 2, indépendance mutuelle)**.** Soit $(A_n)_{n\geqslant 1}$ une suite d'évènements. On dit qu'ils sont :

— 2 à 2 indépendants si pour tout couple (i, j) d'indices distincts, A_i et A_j sont indépendants ;

— mutuellement indépendants si pour tout ensemble fini d'indices $(i_1, \ldots, i_k)$ distincts, on a

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k}).$$

Exemple 2.2.5. Pour que 3 évènements (A, B, C) soient :

— 2 à 2 indépendants, il faut que $A \perp B$, $A \perp C$ et $B \perp C$, c'est-à-dire

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

— mutuellement indépendants, il faut que les 3 relations précédents soient vérifiées et de plus que

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

Exemple 2.2.6. On reprend l'exemple des deux lancers successifs d'un dé et on note C l'évènement : "La somme des deux lancers est paire". On a donc $\mathbb{P}(C) = 1/2$. On vérifie que les évènements (A, B, C) sont 2 à 2 indépendants, mais que :

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = 1/4 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 1/8.$$

Remarque 2.2.7. En pratique, ce sera l'indépendance mutuelle qui nous intéressera et c'est aussi celle que l'on rencontrera le plus souvent. Ainsi, quand on parlera d'une famille d'évènements indépendants (sans plus de précisions), il faudra désormais comprendre mutuellement indépendants.

Chapitre 3

Les variables aléatoires et leurs caractéristiques

Dans ce chapitre, nous introduisons la notion fondamentale de variable aléatoire (réelle) qui jouera un rôle important dans la suite, aussi bien en théorie des probabilités qu'en statistique. Nous donnons en particulier des exemples classiques de variables discrètes et continues et nous introduisons certaines de leurs caractéristiques : fonction de répartition, densité, moyenne, variance, et autres moments.

3.1 Variables aléatoires

Nous commençons par donner ici les définitions d'une variable aléatoire et de la loi d'une variable aléatoire. Dans les prochains paragraphes, nous donnerons de nombreux exemples de telles variables dans les cas dicret et continu.

Définition 3.1.1. Une variable aléatoire X (réelle) est une application "mesurable" d'un espace de probabilités $(\Omega,\mathcal{F},\mathbb{P})$ dans l'ensemble $\mathbb{R}$ des nombres réels.

Remarque 3.1.2. Le terme "mesurable" indique que la fonction considérée doit "bien" se comporter vis à vis de la tribu $\mathcal{F}$. Précisément, l'image réciproque par X de tout intervalle doit être un élément de la tribu $\mathcal{F}$.

Exemple 3.1.3. Considérons un jeu de pile ou face avec une pièce équilibrée, que l'on modélise par un triplet $(\Omega,\mathcal{F},\mathbb{P})$ où $\Omega = \{\texttt{pile}, \texttt{face}\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, et $\mathbb{P}$ uniforme. Si on tombe sur `pile`, on gagne 10 euros, si on tombe sur `face` on perd 10 euros. Le gain G est une variable aléatoire. En effet, c'est une fonction définie sur l'ensemble Ω et à valeurs dans l'ensemble $\{-10, 10\} \subset \mathbb{R}$, avec

$$G(\texttt{pile}) = 10, \quad G(\texttt{face}) = -10.$$

Exemple 3.1.4. Considérons le jet de deux dés, que l'on modélise par un triplet $(\Omega,\mathcal{F},\mathbb{P})$ où $\Omega = \{1,2,3,4,5,6\}^2 = \{\omega = (\omega_1,\omega_2),\ \omega_1,\omega_2 \in \{1,2,3,4,5,6\}\}$, $\mathcal{F}$ est la tribu des parties $\mathcal{F} = \mathcal{P}(\Omega)$, et $\mathbb{P}$ uniforme. On note S la somme des deux dés.

Alors S est une variable aléatoire. C'est une fonction définie sur l'ensemble Ω et à valeurs dans l'ensemble $\{2, 3, \ldots, 12\} \subset \mathbb{R}$, avec

$$S(\omega) = S(\omega_1, \omega_2) := \omega_1 + \omega_2.$$

Si l'on dispose d'un espace de probabilités (Ω,$\mathcal{F}$,$\mathbb{P}$) et d'une variable aléatoire $X : \Omega \to X(\Omega) \subset \mathbb{R}$, on peut construire de façon naturelle une probabilité sur $X(\Omega)$, l'ensemble des valeurs prises par la fonction X.

Proposition 3.1.5. *Pour tout sous-ensemble "mesurable" B de $X(\Omega)$, on définit :*

$$\mathbb{P}_X(B) = \mathbb{P}(\{\omega | X(\omega) \in B\}) = \mathbb{P}(\{X^{-1}(B)\}).$$

Ce faisant, on définit une probabilité sur $X(\Omega)$, appelée la loi de X.

Exemple 3.1.6. Considérons le jeu de pile ou face précédent où l'on gagne ou perd 10 euros selon que la pièce tombe sur `pile` ou `face`. Comme ci-dessus, on note G le gain après le lancer. La variable G définit une probabilité sur les gains possibles $G(\Omega) = \{-10, 10\} \subset \mathbb{R}$:

$$\mathbb{P}_G(\{-10\}) := \mathbb{P}(G = -10) = \mathbb{P}(\{\omega,\ G(\omega) = -10\}) = \mathbb{P}(\texttt{face}) = 1/2,$$

$$\mathbb{P}_G(\{10\}) := \mathbb{P}(G = -10) = \mathbb{P}(\{\omega,\ G(\omega) = 10\}) = \mathbb{P}(\texttt{pile}) = 1/2.$$

Exemple 3.1.7. Considérons maintenant l'exemple précédent de la somme S de deux dés. On obtient alors une probabilité $\mathbb{P}_S$ sur l'ensemble $S(\Omega) = \{2, 3, \ldots, 12\}$, avec

$$\begin{aligned}
&\mathbb{P}_S(\{2\}) := \mathbb{P}(\{\omega,\ S(\omega) = 2\}) = \mathbb{P}(\{(1,1)\}) = 1/36,\\
&\mathbb{P}_S(\{3\}) := \mathbb{P}(\{\omega,\ S(\omega) = 3\}) = \mathbb{P}(\{(1,2),(2,1)\}) = 2/36,\\
&\mathbb{P}_S(\{4\}) := \mathbb{P}(\{\omega,\ S(\omega) = 4\}) = \mathbb{P}(\{(1,3),(3,1),(2,2)\}) = 3/36,\\
&\mathbb{P}_S(\{5\}) := \mathbb{P}(\{\omega,\ S(\omega) = 5\}) = \mathbb{P}(\{(1,4),(4,1),(2,3),(3,2)\}) = 4/36,\\
&\mathbb{P}_S(\{6\}) := \mathbb{P}(\{\omega,\ S(\omega) = 6\}) = \mathbb{P}(\{(1,5),(5,1),(2,4),(4,2),(3,3)\}) = 5/36,\\
&\mathbb{P}_S(\{7\}) := \mathbb{P}(\{\omega,\ S(\omega) = 7\}) = \mathbb{P}(\{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\}) = 6/36,\\
&\mathbb{P}_S(\{8\}) := \mathbb{P}(\{\omega,\ S(\omega) = 8\}) = \mathbb{P}(\{(2,6),(6,2),(3,5),(5,3),(4,4)\}) = 5/36,\\
&\mathbb{P}_S(\{9\}) := \mathbb{P}(\{\omega,\ S(\omega) = 9\}) = \mathbb{P}(\{(3,6),(6,3),(4,5),(5,4)\}) = 4/36,\\
&\mathbb{P}_S(\{10\}) := \mathbb{P}(\{\omega,\ S(\omega) = 10\}) = \mathbb{P}(\{(4,6),(6,4),(5,5)\}) = 3/36,\\
&\mathbb{P}_S(\{11\}) := \mathbb{P}(\{\omega,\ S(\omega) = 11\}) = \mathbb{P}(\{(5,6),(6,5)\}) = 2/36,\\
&\mathbb{P}_S(\{12\}) := \mathbb{P}(\{\omega,\ S(\omega) = 12\}) = \mathbb{P}(\{(6,6)\}) = 1/36.
\end{aligned}$$

3.1.1 Variables aléatoires discrètes

On s'intéresse ici de plus près au cas de variables aléatoires discrètes qui constituent l'essentiel des variables que nous considérerons dans la suite.

Définition 3.1.8. On appelle variable aléatoire discrète une variable aléatoire $X : \Omega \to X(\Omega)$ dont l'ensemble d'arrivée $X(\Omega)$ est fini ou dénombrable.

Cela signifie de la fonction X ne peut prendre qu'un nombre fini ou dénombrable de valeurs $x_1, x_2, \ldots, x_n, \ldots$. Dans ce cas, la loi $\mathbb{P}_X$ de la variable X est caractérisée par la donnée des probabilités des singletons $\mathbb{P}_X(\{x_i\}) = \mathbb{P}(X = x_i) := p_i$, pour tout $i = 1, 2, \ldots$. Les nombres p_i verifient les propriétés du paragraphe 1.3.2 , à savoir : $p_i \geqslant 0$ pour tout i et $\sum_i p_i = 1$. La probabilité d'une partie mesurable A de $X(\Omega)$ est alors donnée par

$$\mathbb{P}(A) = \sum_{i, x_i \in A} p_i.$$

Exemples de variables discrètes

Voici quelques exemples classiques de loi discrètes.

Loi de Bernoulli : on dit qu'une variable aléatoire X suit une loi de Bernoulli de paramètre $p \in [0, 1]$ et on note $X \sim \mathcal{B}(p)$, si X est à valeurs dans l'ensemble $\{0, 1\}$ et

$$\mathbb{P}_X(\{0\}) = \mathbb{P}(X = 0) = 1 - p, \qquad \mathbb{P}_X(\{1\}) = \mathbb{P}(X = 1) = p.$$

C'est la loi d'un jet de pile ou face ou de n'importe quelle expérience aléatoire qui n'a que deux issues possibles.

Loi uniforme : on dit qu'une variable aléatoire X suit une loi uniforme sur un ensemble fini $E = \{x_1, \ldots, x_n\}$ et on note $X \sim U_E$, si X est à valeurs dans l'ensemble E et

$$\mathbb{P}(X = x_i) = \frac{1}{n}, \ i = 1, \ldots, n.$$

La loi uniforme est utilisée lorsque qu'aucun point de l'ensemble d'arrivée n'est privilégie : chacun a le même poids, ici $1/n$.

Loi binomiale : on dit qu'une variable aléatoire X suit une loi binomiale de paramètres (n, p) et on note $X \sim \mathcal{B}(n, p)$, si X est à valeurs dans l'ensemble $\{0, 1, \ldots, n\}$ et

$$\mathbb{P}_X(\{k\}) = \mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \ k \in \{0, 1, \ldots, n\}.$$

Si on joue a pile ou face n fois de suite, la loi binomiale est la loi du nombre de pile au cours des n lancers.

Loi géométrique : on dit qu'une variable aléatoire X suit une loi géométrique de paramètres p et on note $X \sim \mathcal{G}(p)$, si X est à valeurs dans l'ensemble $\{1, 2, \ldots\}$ et

$$\mathbb{P}_X(\{k\}) = \mathbb{P}(X = k) = p(1 - p)^{k-1}, \ k \in \{1, 2, \ldots\}.$$

Si l'on répète un jeu de pile ou face, la loi géométrique est la loi du temps d'apparition du premier pile.

Loi géométrique bis : on dit qu'une variable aléatoire X suit une loi géométrique de paramètres p et on note $X \sim \mathcal{G}(p)$, si X est à valeurs dans l'ensemble $\{0, 1, 2, \ldots\}$ et

$$\mathbb{P}_X(\{k\}) = \mathbb{P}(X = k) = p(1 - p)^k, \ k \in \{0, 1, 2, \ldots\}.$$

Loi de Poisson : on dit qu'une variable aléatoire X suit une loi de Poisson de paramètres λ et on note $X \sim \mathcal{P}(\lambda)$, si X est à valeurs dans l'ensemble $\{0, 1, 2, \ldots\}$ et

$$\mathbb{P}_X(\{k\}) = \mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!},\ k \in \{0, 1, 2, \ldots\}.$$

La loi de Poisson peut être vue comme un cas limite de loi binomiale. En effet, on montre qu'une loi de Poisson est la limite d'une $\mathcal{B}(n, p)$ pour laquelle on a $n \to \infty$ et $p \to 0$ et $np \to \lambda \neq \infty$.

3.1.2 Variables aléatoires continues

Nous introduisons à présent les variables aléatoires continues : ce sont des variables aléatoires qui peuvent prendre un nombre infini (non dénombrable) de valeurs, typiquement ce sont les variables à valeurs dans un intervalle de la droite réelle.

Définition 3.1.9. On dit qu'une variable aléatoire $X : \Omega \to X(\Omega)$ est continue si l'ensemble de ses valeurs $X(\Omega)$ est un intervalle de $\mathbb{R}$. On dit qu'une variable continue admet une densité $f(x)$ si pour tout intervalle $[a, b] \subset X(\Omega)$:

$$\mathbb{P}_X([a, b]) = \mathbb{P}(X \in [a, b]) = \int_a^b f(x)dx,$$

où f est une fonction continue, positive sur $X(\Omega)$ telle que $\int_{X(\Omega)} f(x)dx = 1$.

Remarque 3.1.10. Si X est une variable continue et admet une densité f, alors pour tout $x_0 \in X(\Omega)$, on a $\mathbb{P}_X(\{x_0\}) = \mathbb{P}(X = x_0) = 0$. Autrement, la variable X a une probabilité nulle de tomber sur un point donné de l'intervalle $X(\Omega)$. En revanche, on a une chance non nulle de tomber dans un petit intervalle autour de x_0 :

$$\mathbb{P}([x_0 - h, x_0 + h]) = \mathbb{P}(X = [x_0 - h, x_0 + h]) = \int_{x_0-h}^{x_0+h} f(x)dx > 0.$$

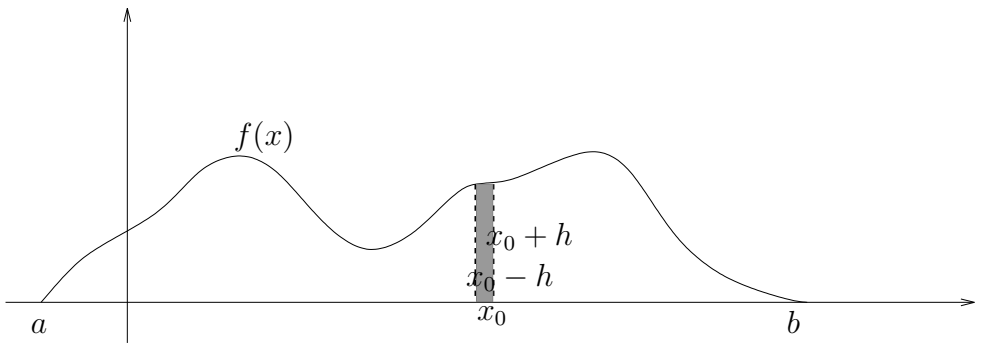


FIGURE 3.1 – Probabilité sur un intervalle via une densité.

Exemples de variables continues

Nous donnons à présent des exemples usuels de loi de probabilité sur des intervalles de $\mathbb{R}$. La loi gaussienne, appelée encore loi normale jouera en particulier un rôle fondamental dans la suite du cours.

Loi uniforme : on dit qu'une variable aléatoire X suit une loi uniforme sur un intervalle $[a, b]$ si X est à valeurs dans l'ensemble $[a, b]$ et pour tout$[c, d[\subset [a, b]$:

$$\mathbb{P}_X([c, d[) = \mathbb{P}(X \in [c, d[) = \frac{d-c}{b-a} = \frac{1}{b-a}\int_c^d 1dx,$$

autrement dit, X a la densité $f(x) \equiv 1/(b-a)$ sur l'intervalle $[a, b]$.

Loi normale ou gaussienne : on dit qu'une variable aléatoire X suit une loi normale de paramètres (μ, σ^2) et on note $X \sim \mathcal{N}(\mu, \sigma^2)$ si X est à valeurs dans $\mathbb{R}$ et pour tout intervalle $[a, b] \subset \mathbb{R}$:

$$\mathbb{P}_X([a, b]) = \mathbb{P}(X \in [a, b]) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx,$$

autrement dit, X a pour densité $f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ sur $\mathbb{R}$.

Loi exponentielle : on dit qu'une variable aléatoire X suit une loi exponentielle de paramètre λ et on note $X \sim \mathcal{E}(\lambda)$ si X est à valeurs dans $[0, +\infty[$ et pour tout intervalle $[a, b] \subset \mathbb{R}$:

$$\mathbb{P}_X([a, b]) = \mathbb{P}(X \in [a, b]) = e^{-\lambda a} - e^{-\lambda b} = \int_a^b \lambda e^{-\lambda x}dx,$$

autrement dit, X a pour densité la fonction $f(x) = \lambda e^{-\lambda x}$ sur $\mathbb{R}^+$.

Loi gamma : on dit qu'une variable aléatoire X suit une loi gamma de paramètres (a, b) et on note $X \sim \Gamma(a, b)$ si X est à valeurs dans $[0, +\infty[$ et pour tout intervalle $[c, d] \subset \mathbb{R}$:

$$\mathbb{P}_X([c, d]) = \mathbb{P}(X \in [c, d]) = \frac{b^a}{\Gamma(a)} \int_c^d x^{a-1} e^{-bx} dx,$$

autrement dit, X a pour densité $f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$ sur $\mathbb{R}^+$.

3.2 Fonction de répartition

Nous introduisons dans ce paragraphe la notion de fonction de répartition d'une variable aléatoire (discrète ou continue). Cette fonction caractérise la loi d'une variable aléatoire, et nous sera utile dans la suite pour dire qu'une suite de variables aléatoires converge vers une variable limite.

Définition 3.2.1. Soit $X : \Omega \to X(\Omega)$ une variable aléatoire. On appelle fonction de répartition de répartition de X, et on note F_X, la fonction de $\mathbb{R}$ dans l'intervalle $[0, 1]$ définie par

$$F_X(x) = \mathbb{P}(X \leqslant x).$$

Proposition 3.2.2. *Soit $X : \Omega \to X(\Omega)$ une variable aléatoire. Alors sa fonction de répartition F_X vérifie les propriétés suivantes :*

1. *F_X est croissante.*
2. *F_X est continue à droite.*
3. *$F_X(-\infty) = 0$ et $F_X(+\infty) = 1$*

Être continu à droite signifie que si la fonction "saute", sa valeur au point de saut est la valeur à droite de celui-ci, *i.e.* les points gris sur la figure ci-après.

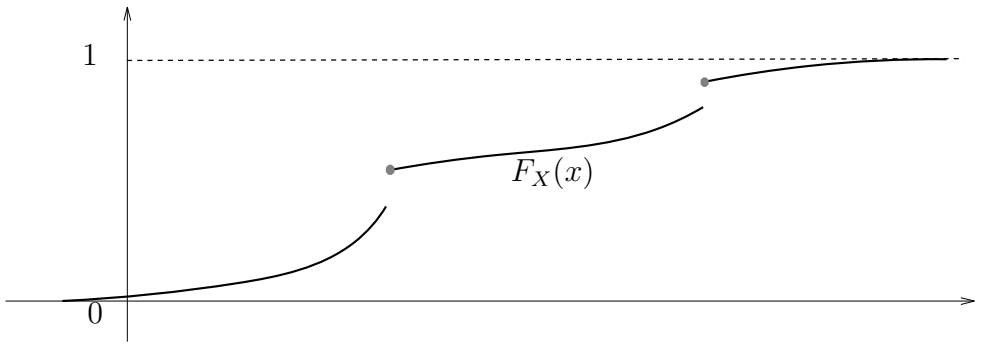


FIGURE 3.2 – Fonction de répartition générique.

Remarque 3.2.3. D'après la définiton de la fonction de répartition, pour tous réels a et b, avec $a < b$ on a : $\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a)$. En particulier, pour tout $x \in \mathbb{R}$: $\mathbb{P}(X > x) = 1 - F_X(x)$.

3.2.1 Fonction de répartition d'une variable discrète

Soit X une variable aléatoire discrète pouvant prendre les valeurs $x_1, x_2, \ldots, x_n, \ldots$ de probabilités respectivement $p_1, p_2, \ldots, p_n, \ldots$ avec $x_1 < x_2 < \ldots < x_n < \ldots$. Alors la fonction de répartition de X est donnée par la formule :

$$F_X(x) = \sum_{i=1}^{i=k} p_i,$$

où k est l'indice tel que $x_k \leqslant x < x_{k+1}$. La fonction $x \mapsto F_X(x)$ est alors une fonction constante par morceaux, dont le graphe a l'allure ci-dessous.

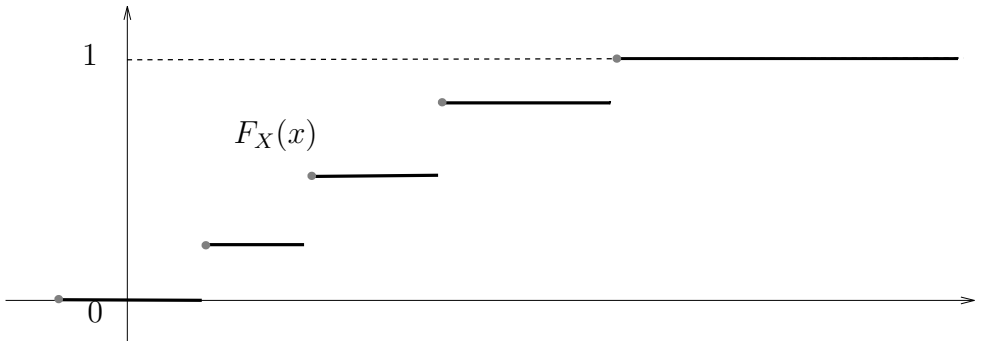


FIGURE 3.3 – Fonction de répartition d'une variable discrète.

Exemple 3.2.4. Ci-dessous, la fonction F_X lorsque $X \sim \mathcal{B}(p)$. La fonction fait un "saut" d'une hauteur p en zéro, et d'une hauteur de $(1-p)$ en un.

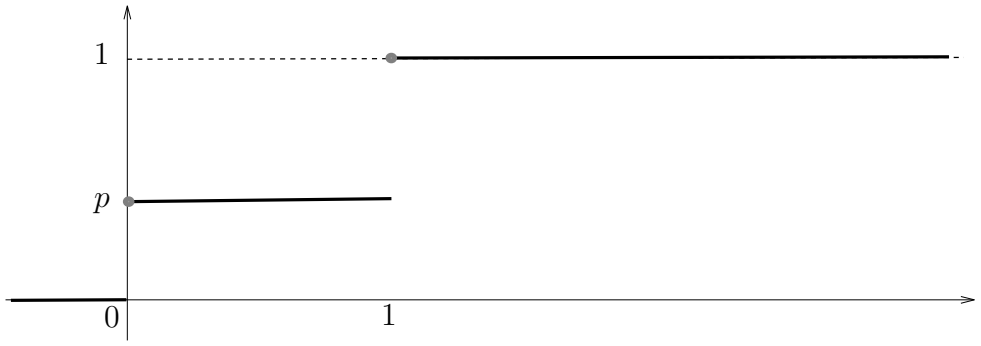


FIGURE 3.4 – Fonction de répartition d'une variable de Bernoulli $\mathcal{B}(p)$.

Exemple 3.2.5. Ci-après, la fonction de répartition d'une variable S de l'exemple 3.1.7, la somme de deux dés. La fonction F_S fait un "saut" d'une hauteur $1/36$ en zéro, d'une hauteur de $2/36$ en un, d'une hauteur $3/36$ en deux etc.

Exemple 3.2.6. Soit X une variable de loi géométrique sur $\{1, 2, \ldots\}$, *i.e.* telle que

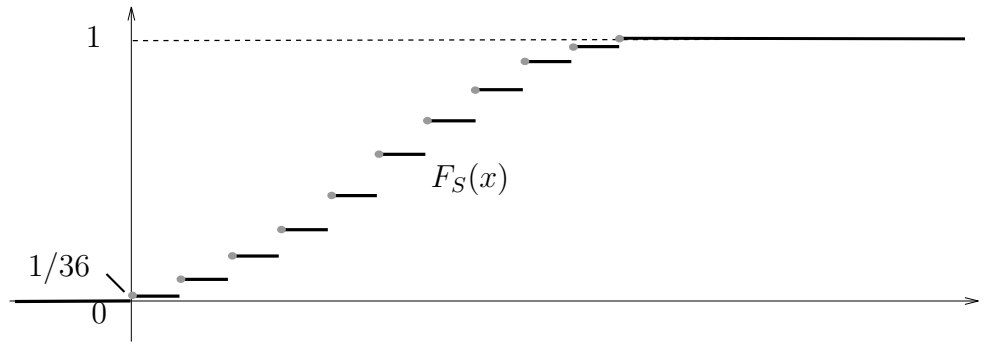


FIGURE 3.5 – Fonction de répartition de la variable S (somme de deux dés).

$\mathbb{P}(X = k) = p(1-p)^{k-1}$. Alors, pour tout entier $m \geqslant 1$, on a

$$\mathbb{P}(X \leqslant m) = 1 - \mathbb{P}(X > m) = 1 - \sum_{k=m+1}^{+\infty} p(1-p)^{k-1} = 1 - (1-p)^m.$$

3.2.2 Fonction de répartition d'une variable continue

Soit X une variable aléatoire continue de densité $f(x)$. Alors, la fonction de répartition de X est la primitive de f : $F_X(x) = \int_{-\infty}^{x} f(u)du$. Dans ce cas, la fonction F_X est une fonction continue à gauche et à droite : on peut la tracer sans lever le stylo.

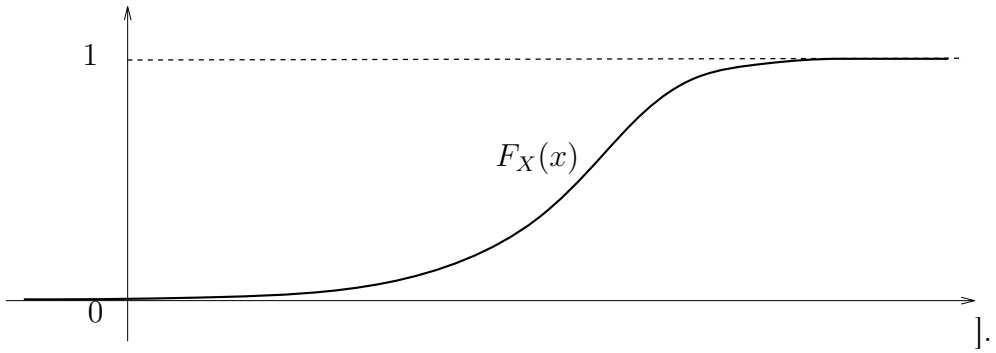


FIGURE 3.6 – Fonction de répartition d'une variable continue.

Exemple 3.2.7. Considérons le cas d'une variable X de loi uniforme sur l'intervalle $[0, 1]$. Sa densité f_X est constante sur l'intervalle $[0, 1]$ et vaut zéro ailleurs. On en déduit que F_X vaut zéro sur $]-\infty, 0]$, vaut 1 sur $[1, +\infty[$ et :

$$F_X(x) = \int_{-\infty}^{x} f(u)du = \int_{0}^{x} 1 \times du = x, \quad \text{pour} \quad x \in [0, 1].$$

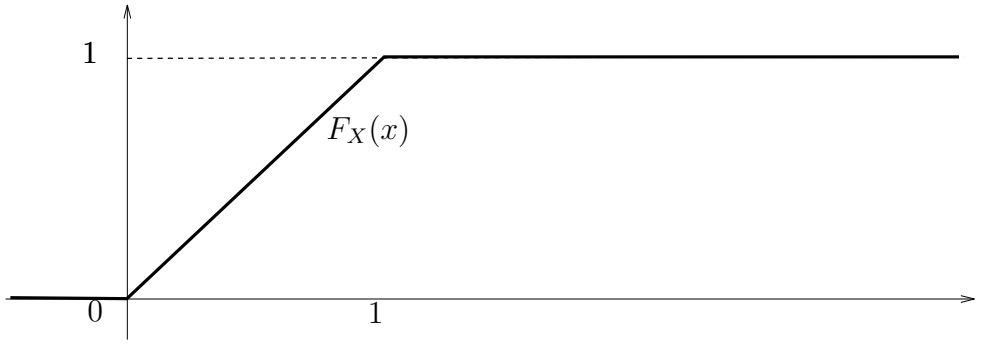


FIGURE 3.7 – Fonction de répartition d'une variable uniforme.

Exemple 3.2.8. Considérons le cas d'une variable X exponentielle de paramètre λ. Sa densité f_X est nulle sur $]-\infty, 0]$ et est donnée par $f_X(x) = \lambda \exp(-\lambda x)$ sur $[0, +\infty[$. On en déduit que F_X vaut zéro sur $]-\infty, 0]$, et vaut, pour $x > 0$:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f(u)du = \int_0^x \lambda \exp(-\lambda u)\, du \\ &= [-\exp(-\lambda u)]_0^x = 1 - \exp(-\lambda x). \end{aligned}$$

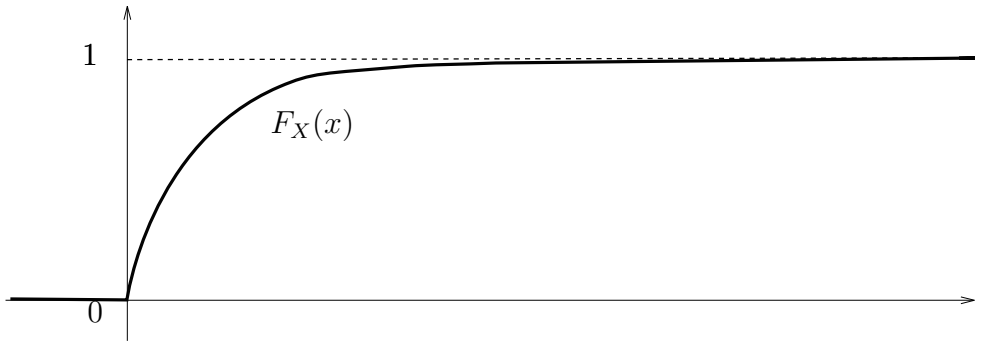


FIGURE 3.8 – Fonction de répartition d'une variable exponentielle.

3.3 Moments d'une variable aléatoire

Dans cette section, nous nous intéressons aux notions de moyenne et de variance d'une variable aléatoire. Ces deux notions seront fondamentales dans la partie "statistique" du cours. Nous définissons tout d'abord la notion d'espérance mathématique.

3.3.1 Espérance d'une variable aléatoire

La notion d'espérance généralise la notion bien connue de moyenne. Il s'agit précisément d'une moyenne pondérée. Dans les deux prochaines sections, nous donnons la définition de l'espérance mathématique d'une variable aléatoire discrète puis d'une variable continue.

Espérance d'une variable discrète

La notion de moyenne pondérée nous est tous familère, il suffit de penser au calcul de la moyenne au baccalauréat ou les différentes matières ont des coefficients distincts : pour un bac S option SVT, un 18 en bio est "plus intéressant" qu'un 18 en sport... L'espérance d'une variable aléatoire discrète est précisément un moyenne pondérée :

Définition 3.3.1. Soit X une variable aléatoire discrète à valeurs dans un ensemble au plus dénombrable $\{x_1, \ldots, x_n, \ldots\}$. On note $p_i := \mathbb{P}(X = x_i)$. Alors l'espérance de X, que l'on note $\mathbb{E}[X]$, est donnée par la formule :

$$\mathbb{E}[X] := \sum_{i=1}^{\infty} x_i p_i = \sum_{i=1}^{\infty} x_i \mathbb{P}_X(\{x_i\}) = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i).$$

Plus généralement, si h est une fonction de $\mathbb{R}$ dans $\mathbb{R}$, alors l'espérance de la variable $h(X)$ est donnée par la formule

$$\mathbb{E}[h(X)] := \sum_{i=1}^{\infty} h(x_i) p_i = \sum_{i=1}^{\infty} h(x_i) \mathbb{P}_X(\{x_i\}) = \sum_{i=1}^{\infty} h(x_i) \mathbb{P}(X = x_i).$$

Exemple 3.3.2. Par exemple, si X suit une loi de Bernoulli $\mathcal{B}(p)$ sur $\{0, 1\}$, alors l'espérance de X vaut

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

Exemple 3.3.3. Par exemple, si X suit une loi uniforme sur $\{1, 2, \ldots, n\}$, alors l'espérance de X vaut

$$\begin{aligned} \mathbb{E}[X] &= 1 \times \mathbb{P}(X = 1) + 2 \times \mathbb{P}(X = 2) + \ldots + n \times \mathbb{P}(X = n) \\ &= \frac{1 + 2 + \ldots + n}{n} = \frac{1}{n} \times \frac{n(n+1)}{2} = \frac{n+1}{2}. \end{aligned}$$

Espérance d'une variable continue

On peut généraliser la définition précédente au cadre continu, en remplaçant la somme discrète par une intégrale.

Définition 3.3.4. Soit X une variable aléatoire continue à valeur dans un intervalle $X(\Omega) \subset \mathbb{R}$ et admettant de densité $f(x)$. Alors l'espérance de X, que l'on note $\mathbb{E}[X]$, est donnée par la formule :

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

Plus généralement, si h est une fonction de $\mathbb{R}$ dans $\mathbb{R}$, alors l'espérance de la variable $h(X)$ est donnée par la formule

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

Exemple 3.3.5. Par exemple, si X suit une loi uniforme sur l'intervalle $[0,1]$, *i.e.* X admet la densité $f \equiv 1$ sur l'intervalle $[0,1]$, alors l'espérance de X vaut

$$\mathbb{E}[X] = \int_0^1 x f(x) dx = \int_0^1 x dx = \left[\frac{x^2}{2}\right]_0^1 = 1/2.$$

De même, l'espérance de la variable X^2 vaut :

$$\mathbb{E}[X^2] = \int_0^1 x^2 f(x) dx = \int_0^1 x^2 dx = \left[\frac{x^3}{3}\right]_0^1 = 1/3.$$

Exemple 3.3.6. Si X suit une loi exponentielle de paramètre λ, *i.e.* si X admet la densité $f(x) = \lambda e^{-\lambda x}$ sur l'intervalle $[0, +\infty[$, alors en intégrant par partie, on obtient que l'espérance de X vaut

$$\begin{aligned}\mathbb{E}[X] &= \int_0^{+\infty} x f(x) dx = \int_0^{+\infty} x\lambda e^{-\lambda x} dx = \left[-xe^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\ &= \left[\frac{-e^{-\lambda x}}{\lambda}\right]_0^{+\infty} = \frac{1}{\lambda}.\end{aligned}$$

Remarque 3.3.7. L'espérance mathématique n'est pas toujours définie. C'est en particulier le cas de la loi de Cauchy dont la densité sur $\mathbb{R}$ est donnée par $f(x) = \frac{1}{\pi(1+x^2)}$. Alors on a

$$\mathbb{E}[|X|] = \int_{-\infty}^{+\infty} \frac{|x|}{\pi(1+x^2)} dx = 2\int_0^{+\infty} \frac{x}{\pi(1+x^2)} dx = +\infty.$$

Propriétés de l'espérance

Nous donnons maintenant quelques propriétés de l'espérance, qui sont vérifiées que l'on se place dans le cas discret ou continu.

Proposition 3.3.8 (Linéarité de l'espérance). *Soient X et Y deux variables aléatoires et $c \in \mathbb{R}$ une constante. Alors on a :*

1. $\mathbb{E}[c] = c$;
2. $\mathbb{E}[cX + Y] = c \times \mathbb{E}[X] + \mathbb{E}[Y]$.

Proposition 3.3.9 (Positivité de l'espérance). *Soient X et Y deux variables aléatoires telles que $X \leqslant Y$ avec probabilité un, alors $\mathbb{E}[X] \leqslant \mathbb{E}[Y]$.*

3.3.2 Variance et autres moments

Définition 3.3.10. Soient X une variable aléatoire et m un entier strictement positif. On dit que X admet un moment d'ordre m si $\mathbb{E}[|X|^m] < +\infty$. Si c'est le cas, on

appelle moment d'ordre m la quantité $\mathbb{E}[X^m]$, c'est-à-dire selon que l'on est dans le cas discret ou continu :

$$\mathbb{E}[X^m] := \int_{X(\Omega)} x^m f(x) dx,$$

$$\mathbb{E}[X^m] := \sum_i x_i^m p_i = \sum_{i=1}^{\infty} x_i^m \mathbb{P}_X(\{x_i\}) = \sum_{i=1}^{\infty} x_i^m \mathbb{P}(X = x_i).$$

Exemple 3.3.11. Dans les exemples ci-dessus, on a vu que la loi uniforme sur $[0, 1]$ admet un moment d'ordre deux puisque $\mathbb{E}[X^2] = 1/3 < +\infty$. En revanche, la loi de Cauchy de densité $f(x) = \frac{1}{\pi(1+x^2)}$ n'admet pas de moment d'ordre un puisque $\mathbb{E}[|X|] = +\infty$.

Définition 3.3.12. Soient X une variable aléatoire qui admet des moments d'ordre un et deux, *i.e.* $\mathbb{E}[|X|] < +\infty$, $\mathbb{E}[|X|^2] < +\infty$. On appelle variance de X et on note $\mathrm{var}(X)$ la quantité

$$\mathrm{var}(X) := \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

La variance traduit la dispersion de la distribution de la variance autour de sa valeur moyenne. Étant un carré, la dimension de la variance n'est pas celle de la moyenne. C'est pourquoi on utilise plus souvent l'écart type, noté souvent σ, qui est la racine de la variance. On dit aussi que la variance traduit la notion d'incertitude. Plus la variance est faible, moins le résultat de l'expérience aléatoire est incertain. Le cas extrême est celui d'une variable aléatoire de variance nulle, qui est en fait déterministe.

Définition 3.3.13. On dit qu'une variable aléatoire Y est centrée réduite si sa moyenne $\mathbb{E}[Y]$ est nulle et si sa variance $\mathrm{var}(Y)$ est égale à un. Si X est une variable aléatoire qui admet des moments d'ordre un et deux, alors la variable $Y := \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{var}(X)}}$ est centrée réduite.

Exemple 3.3.14. Par exemple, si X suit une loi de Bernoulli $\mathcal{B}(p)$ sur $\{0, 1\}$, on a vu que l'espérance de X vaut

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

Le moment d'ordre deux vaut lui aussi p :

$$\mathbb{E}[X^2] = 0^2 \times \mathbb{P}(X = 0) + 1^2 \times \mathbb{P}(X = 1) = 0 \times (1 - p) + 1 \times p = p,$$

de sorte que la variance de X vaut $\mathrm{var}(X) = p - p^2 = p(1 - p)$.

Exemple 3.3.15. Dans l'exemple de la loi uniforme sur l'intervalle $[0, 1]$, on a vu que $\mathbb{E}[X] = 1/2$ et $\mathbb{E}[X^2] = 1/3$, on a donc $\mathrm{var}(X) = 1/3 - 1/4 = 1/12$.

Proposition 3.3.16. *Soient X et Y deux variables aléatoires, et a et b deux constantes réelles. Alors on a $var(aX+b) = a^2 var(X)$ et*

$$var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$$

où $cov(X,Y)$ est la covariance de X et Y définie par :

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

3.3.3 Moments des variables usuelles

On explicite ici les premiers moments des variables usuelles. On traite dans un premier temps le cas des variables discrètes, puis celui des variables continues admettant un densité.

Moments des variables discrètes usuelles

Le cas de la variable de Bernoulli a déja été traité dans l'exemple 3.3.14.

Loi de Bernoulli : on a vu que si X suit une loi de Bernoulli de paramètre $p \in [0,1]$, alors $\mathbb{E}[X] = p$ et $\text{var}(X) = p(1-p)$.

Loi uniforme : si X suit une loi uniforme sur un ensemble fini $E = \{1, \ldots, n\}$ alors on a vu dans l'exemple 3.3.3 que $\mathbb{E}[X] = n(n+1)/2$. Le calcul du moment d'ordre deux montre que $\text{var}(X) = (n^2-1)/12$, en effet, on a

$$\begin{aligned}\mathbb{E}[X^2] &= 1^2 \times \mathbb{P}(X=1) + 2^2 \times \mathbb{P}(X=2) + \ldots + n^2 \times \mathbb{P}(X=n) \\ &= \frac{1^2 + 2^2 + \ldots + n^2}{n} = \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.\end{aligned}$$

Loi binomiale : si $X_1, \ldots, X_n$ sont des variables "indépendantes" de loi de Bernoulli de paramètre p, alors la somme $S_n = X_1 + \ldots + X_n$ suit une loi binomiale $\mathcal{B}(n,p)$. Par linéarité de l'espérance, on en déduit que $\mathbb{E}[S_n] = np$ et $\text{var}(S_n) = np(1-p)$.

Loi géométrique : si X suit une loi géométrique de paramètre p sur $\{1, 2, \ldots\}$ alors $\mathbb{E}[X] = 1/p$ et $\text{var}(X) = 1-p$. En effet, on a

$$\begin{aligned}\mathbb{E}[X] &= \textstyle\sum_{k=1}^{+\infty} k \times p(1-p)^{k-1} = p \times \left(\sum_{k=0}^{+\infty} k \times (1-p)^{k-1}\right) \\ &= p \times \left(\sum_{k=0}^{+\infty} \times (1-p)^k\right)' = p \times \left(\frac{-1}{p}\right)' = 1/p.\end{aligned}$$

Par ailleurs, on montre que $\mathbb{E}[X^2] = (1-p) + 1/p^2$, d'où le résultat.

Loi de Poisson : si X suit une loi de Poisson de paramètres λ, alors $\mathbb{E}[X] = \lambda$ et $\mathrm{var}(X) = \lambda$. En effet, on a

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{+\infty} k \times e^{-\lambda}\frac{\lambda^k}{k!} = \lambda \times \left(e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}\right)\\ &= \lambda \times \left(e^{-\lambda}\sum_{\ell=0}^{+\infty}\frac{\lambda^{\ell}}{\ell!}\right) = \lambda.\\ \mathbb{E}[X^2] &= \sum_{k=0}^{+\infty} k^2 \times e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=0}^{+\infty}(k(k-1)+k) \times e^{-\lambda}\frac{\lambda^k}{k!}\\ &= \lambda^2\left(\sum_{k=2}^{+\infty} e^{-\lambda}\frac{\lambda^{k-2}}{(k-2)!}\right) + \lambda\left(\sum_{k=1}^{+\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!}\right) = \lambda^2 + \lambda.\end{aligned}$$

Moments des variables usuelles continues

On donne maintenant les moyennes et variances des variables continues les plus usuelles. Le cas d'une variable uniforme sur l'intervalle $[0, 1]$ a déjà été traité dans l'exemple 3.3.5.

Loi uniforme : si X suit une loi uniforme sur un intervalle $[a, b]$ alors l'espérance de X est $\mathbb{E}[X] = (b-a)/2$ et sa variance $\mathrm{var}(X) = (b-a)^2/12$.

Loi normale ou gaussienne : si X suit une loi normale de paramètres (μ, σ^2) alors la moyenne de X est $\mathbb{E}[X] = \mu$, et sa variance $\mathrm{var}(X) = \sigma^2$.

Loi exponentielle : si X suit une loi exponentielle de paramètre λ, on a vu dans l'exemple 5.1.10 que $\mathbb{E}[X] = 1/\lambda$. De même, on montre que $\mathbb{E}[X^2] = 1/\lambda^2 + 1/\lambda$. On a donc $\mathrm{var}(X) = 1/\lambda^2$.

Chapitre 4

Théorèmes limite fondamentaux

L'objet de ce chapitre est d'énoncer les deux théorèmes limite qui sont à la base de la théorie des probabilités et des statistiques à savoir, la loi des grands nombres et le théorème limite central. Pour se faire, nous généralisons tout d'abord la notion d'indépendance des évènements aux variables aléatoires, puis nous définissons différents modes de convergence qui vont nous permettre de traduire le fait qu'une suite de variables aléatoires converge vers une variable aléatoire limite.

4.1 Indépendance de variables aléatoires

Au chapitre précédent, nous avons introduit la notion d'indépendance de deux (ou plus d') évènements. Cette notion se généralise aux variables aléatoires.

4.1.1 Définitions équivalentes

Il existe plusieurs définitions équivalentes pour l'indépendance de variables aléatoires. La plus simple consiste à repasser par la notion d'indépendance pour les évènements.

Définition 4.1.1. Soient deux variables aléatoires X et Y définies sur un même espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que X et Y sont indépendantes, et on note parfois $X \perp Y$, si pour tous ensembles "mesurables" A et B dans $\mathbb{R}$, on a

$$\mathbb{P}(X \in A \text{ et } Y \in B) = \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B).$$

Plus généralement, on dit que des variables aléatoires $(X_i)_{i \in I}$ sont indépendantes, si pour toute famille d'évènements $(A_i)_{i \in I}$:

$$\mathbb{P}\left(\bigcap_{i \in I} \{X_i \in A_i\}\right) = \prod_{i \in I} \mathbb{P}(X_i \in A_i).$$

On peut se limiter à une famille d'évènements bien choisis et ainsi utiliser les fonctions de répartition. En outre, on peut envisager une définition utilisant l'espérance mathématique définie au chapitre précédent.

Définition 4.1.2 (équivalentes). Les variables X et Y sont indépendantes si pour tous $s, t \in \mathbb{R}$:

$$\mathbb{P}(X \leqslant s \text{ et } Y \leqslant t) = \mathbb{P}(X \leqslant s) \times \mathbb{P}(Y \leqslant t) = F_X(s) \times F_Y(t).$$

Les variables X et Y sont indépendantes si pour toutes fonctions continues bornées g et h :

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

Exemple 4.1.3. On considère le jet de deux dés modélisé par $\Omega = \{1, \ldots, 6\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$, et $\mathbb{P}$ uniforme. On note X le résultat du premier dé et Y le résultat du second. Alors X et Y sont des variables aléatoires indépendantes, pour tout $(k, \ell) \in \Omega$:

$$\mathbb{P}(X = k \text{ et } Y = \ell) = \mathbb{P}(X = k)\mathbb{P}(Y = \ell).$$

Exemple 4.1.4. Soient $p \in]0, 1[$ et X, Y deux variables aléatoires à valeurs dans l'ensemble $\{0, 1\}$ et telles que $\mathbb{P}(Y = 1) = p$, $\mathbb{P}(X = 0 \text{ et } Y = 1) = (1 - p)^2$, $\mathbb{P}(X = 0 \text{ et } Y = 0) = p(1 - p)$. Alors X et Y sont indépendantes. En effet, on a $\mathbb{P}(X = 0) = 1 - p$ et $\mathbb{P}(X = 1) = p$ puisque

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

.

On a donc bien

$$\mathbb{P}(X = 0 \text{ et } Y = 1) = \mathbb{P}(X = 0)\mathbb{P}(Y = 1) = (1 - p)^2,$$

$$\mathbb{P}(X = 0 \text{ et } Y = 0) = \mathbb{P}(X = 0)\mathbb{P}(Y = 0) = p(1 - p),$$

$$\mathbb{P}(X = 1 \text{ et } Y = 1) = \mathbb{P}(X = 1)\mathbb{P}(Y = 1) = p(1 - p),$$

$$\mathbb{P}(X = 1 \text{ et } Y = 0) = \mathbb{P}(X = 1)\mathbb{P}(Y = 0) = p^2.$$

Dans le cas de variables à densité, on peut encore donner la définition suivante.

Définition 4.1.5 (indépendance et densité). Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires continues admettant des densités $f_{X_1}, f_{X_2}, \ldots, f_{X_n}$. Les variables $X_1, X_2, \ldots, X_n$ sont indépendantes si et seulement si le vecteur $(X_1, X_2, \ldots, X_n)$ admet la densité $f_{X_1} \times f_{X_2} \times \ldots \times f_{X_n}$, c'est-à-dire, pour tout $[a_i, b_i[\subset \mathbb{R}$:

$$\mathbb{P}\left(\bigcap_{i \in I} \{X_i \in [a_i, b_i[\}\right) = \int_{\prod_{i=1}^n [a_i, b_i[} f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n) dx_1 \ldots dx_n.$$

Exemple 4.1.6. Soient X et Y deux variables aléatoires de loi exponentielle de paramètres λ et μ respectivement. On suppose que X et Y sont indépendantes. Alors le couple (X, Y) admet la densité $f_{(X,Y)}$ suivante :

$$f_{(X,Y)}(x, y) = f_X(x) \times f_Y(y) = \lambda e^{-\lambda x} \times \mu e^{-\mu y} = \lambda\mu e^{-\lambda x - \mu y}.$$

Coefficient de corrélation

La dépendance / relation entre deux variables aléatoires peut être quantifiée par la covariance comme vue précédemment. Cependant, à l'image de la moyenne et de la variance, la covariance est un moment donc possède une dimension ce qui la rend plus difficile à interpréter. C'est pourquoi on utilise plus généralement le coefficient de corrélation, indicateur sans dimension, défini par

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}.$$

Le coefficient de corrélation mesure la qualité de la relation linéaire entre deux variables aléatoires X et Y (i.e. de la forme $Y = aX + b$).

Proposition 4.1.7. *Pour toutes variables X et Y possédant un moment d'ordre deux, on a les propriétés suivantes :*

1. *$\rho(X,Y) \in [-1,1]$;*
2. *si $X \perp Y$, alors $\rho(X,Y) = 0$. La réciproque n'est pas vraie en général ;*
3. *si il existe une relation linéaire entre X et Y alors $\rho(X,Y) = \pm 1$.*

Exemple 4.1.8. On place au hasard deux billes dans deux boîtes A et B. On note X la variable aléatoire "nombre de billes dans la boîte A" et Y la variable aléatoire "nombre de boîtes vides". Les lois, espérances et variances de X, Y et XY sont :

$$\mathbb{P}(X=0) = 1/4, \quad \mathbb{P}(X=1) = 1/2, \quad \mathbb{P}(X=2) = 1/4, \quad \mathbb{E}[X] = 1, \text{var}(X) = 1/2,$$

$$\mathbb{P}(Y=0) = 1/2, \quad \mathbb{P}(Y=1) = 1/2, \quad \mathbb{E}[Y] = 1/2, \text{var}(Y) = 1/4,$$

$$\mathbb{P}(XY=0) = 3/4, \quad \mathbb{P}(XY=1) = 0, \quad \mathbb{P}(XY=2) = 1/4, \quad \mathbb{E}[XY] = 1/2.$$

Le coefficient de corrélation $\rho(X,Y)$ est nul car

$$\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 1/2 - 1 \times 1/2 = 0.$$

Cependant les variables X et Y ne sont pas indépendantes. En effet, X et Y ne peuvent s'annuler simultanément car il est impossible d'avoir à la fois aucune bille dans la boite A et aucune boite vide. On a donc

$$0 = \mathbb{P}(X = 0 \text{ et } Y = 0) \neq \mathbb{P}(X=0)\mathbb{P}(Y=0) = 1/4 \times 1/2 = 1/8.$$

4.2 Convergence de variables aléatoires

Les théorèmes limite qui font l'objet de ce chapitre concernent le comportement asymptotique de suites de variables aléatoires. Ils traduisent la convergence de ces suites vers des limites, qui peuvent être déterministes mais aussi aléatoires. Nous précisons ici en quel sens une suite de variables aléatoires peut admettre une limite.

4.2.1 Les différents types de convergence

Nous commençons par donner la définition de la converge en probabilité d'une suite de variables aléatoires.

Convergence en probabilités

Définition 4.2.1 (convergence en probabilités). Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définie sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que la suite (X_n) converge en probabilité vers une variable aléatoire X, et on note $X_n \xrightarrow{\mathbb{P}} X$ si pour tout $\varepsilon > 0$:

$$\lim_{n\to+\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] = 0,$$

ou de manière équivalente :

$$\lim_{n\to+\infty} \mathbb{P}\left[|X_n - X| \leqslant \varepsilon\right] = 1.$$

Exemple 4.2.2. Soit (X_n) une suite de variables indépendantes à valeurs dans l'ensemble $\{0, 1\}$ et telles que $\mathbb{P}(X_n = 0) = 1/n$, et donc $\mathbb{P}(X_n = 1) = 1 - 1/n$. Alors la suite (X_n) converge en probabilité vers la variable "aléatoire" constante égale à un. En effet, fixons $0 < \varepsilon < 1$. Lorsque n tend vers l'infini, on a

$$\mathbb{P}(|X_n - 1| > \varepsilon) = \mathbb{P}(X_n = 0) = 1/n \longrightarrow 0.$$

Exemple 4.2.3. Soit (X_n) une suite de variables aléatoires indépendantes telles que $\mathbb{P}(X_n = 2 - 1/n) = 1/3$ et $\mathbb{P}(X_n = 2 + 1/n) = 2/3$. Alors la suite (X_n) converge en probabilité vers la variable "aléatoire" constante égale à 2. En effet, fixons $0 < \varepsilon < 1$. On a toujours $|X_n - 2| = 1/n$ de sorte que pour $n > 1/\varepsilon$:

$$\mathbb{P}(|X_n - 2| > \varepsilon) = 0.$$

Exemple 4.2.4. Considérons une variable aléatoire X à valeurs dans $\{0, 1\}$ et telle que $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/2$. Pour tout entier $n \geqslant 1$, on définit la variable X_n de la façon suivante : si X vaut 1, alors X vaut 1 ; si X vaut zéro, alors X_n vaut $1/n$. Alors, n tend vers l'infini, la suite X_n converge en probabilité vers X. En effet, on a toujours $|X_n - X| = 1/n$. Pour $\varepsilon > 0$ fixé, dès que $n > 1/\varepsilon$, on a alors :

$$\mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Exemple 4.2.5. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes de loi uniforme sur l'intervalle $[0, 1]$. Alors, losque n tend vers l'infini, la suite de variables aléatoires $Y_n := \max(X_1, \ldots, X_n)$ converge en probabilité vers la constante 1. En effet, soit $0 < \varepsilon < 1$, on a

$$\begin{aligned}\mathbb{P}(|Y_n - 1| > \varepsilon) &= \mathbb{P}(\max(X_1, \ldots, X_n) < 1 - \varepsilon) \\ &= \mathbb{P}(X_1 < 1 - \varepsilon) \ldots \mathbb{P}(X_n < 1 - \varepsilon) = (1 - \varepsilon)^n \longrightarrow 0.\end{aligned}$$

Il existe une notion de convergence qui est plus forte que la convergence en probabilité, c'est la convergence presque sûre : toute suite qui converge presque sûrement converge en probabilité.

Convergence presque sûre

Définition 4.2.6 (convergence presque sûre)**.** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définie sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que la suite (X_n) converge presque sûrement vers une variable aléatoire X, et on note $X_n \xrightarrow{p.s.} X$, si il existe un sous ensemble A de Ω avec $\mathbb{P}(A) = 1$ et pour tout $\omega \in A$:

$$X_n(\omega) \to X(\omega).$$

Exemple 4.2.7. Considérons l'espace de probabilté $(\Omega, \mathcal{F}, \mathbb{P})$ où $\Omega = [0, 1]$, $\mathcal{F}$ est la tribu borélienne, et $\mathbb{P}$ est la probabilité uniforme sur $[0, 1]$. Pour tout $n \geqslant 1$, on considère la variable aléatoire X_n définit comme suit :

$$X_n : [0, 1] \to \mathbb{R}, \qquad X_n(\omega) = \min(n \times \omega, 1).$$

Soit $A :=]0, 1]$, on a $\mathbb{P}(A) = 1$. Par ailleurs, pour tout $\omega \in A$, on a $n \times \omega \to +\infty$, donc pour n assez grand $X_n(\omega) = 1$. La suite de variables aléatoires X_n converge donc presque sûrement vers 1 lorsque n tend vers l'infini.

Exemple 4.2.8. On reprend l'exemple 4.1.8. À la variable XY qui est à valeurs dans $\Omega = \{0, 1, 2\}$, on associe la suite Z_n définie par la formule :

$$Z_n = \left(1 - \frac{(XY - 1)^2}{2}\right)^n.$$

Lorsque XY vaut 0 ou 2, on a $Z_n = 1/2^n$ qui converge vers zéro lorsque n tend vers l'infini. Lorsque XY vaut 1, ce qui arrive avec probabilité $\mathbb{P}(XY = 1) = 0$, Z_n vaut 1. On peut donc affirmer que presque sûrement la suite de variables aléatoires converge presque sûrement vers zéro.

Nous définissons enfin un dernier mode de convergence, plus faible que les deux précédents, dont l'importance sera mise en évidence dans l'énoncé du théorème limite central.

Convergence en loi

Définition 4.2.9 (convergence en loi)**.** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définie sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que la suite (X_n) converge en loi vers une variable aléatoire X, et on note $X_n \xrightarrow{\mathcal{L}} X$, si la suite des fonctions de répartition F_{X_n} converge vers F_X en tout point de continuité de F_X, *i.e.* lorsque n tend vers l'infini, pour tout x où F_X ne "saute" pas :

$$F_{X_n}(x) = \mathbb{P}(X_n \leqslant x) \longrightarrow F_X(x) = \mathbb{P}(X \leqslant x).$$

Exemple 4.2.10. Soit (X_n) une suite de variables indépendantes et uniformes à valeurs dans l'ensemble $\{0, 1/2^n, 2/2^n, \dots, 1\}$, *i.e.* telles que

$$\mathbb{P}(X_n = k/n) = 1/(2^n + 1), \quad k \in \{0, 1, \dots, 2^n\}.$$

Alors la suite (X_n) converge en loi vers la variable aléatoire X de loi uniforme sur l'intervalle $[0, 1]$. En effet, soit $x \in [0, 1]$ alors pour tout n il existe k_n tel que $k_n/2^n \leqslant x < (k_n + 1)/2^n$ et donc $k_n/2^n \to x$. Dès lors,

$$\mathbb{P}(X_n \leqslant x) = \frac{k_n}{2^n} \to x = F_X(x).$$

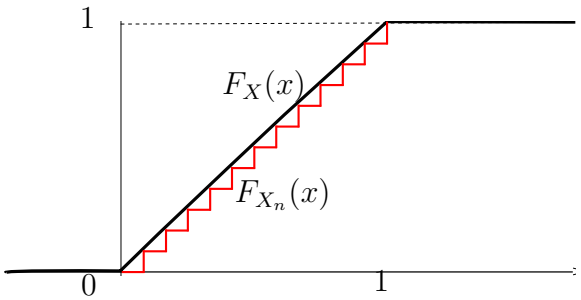


FIGURE 4.1 – Fonction de répartition des lois uniformes continue et discrètes.

Exemple 4.2.11. Les variables exponentielles sont caractérisées par leur fonction de répartititon : si $X \sim \mathcal{E}(\lambda)$ alors $\mathbb{P}(X > x) = e^{-\lambda x}$. Soient (a_n) une suite de nombres réels positifs tels que $\sum_1^{+\infty} a_n = 2$, et (X_n) des variables indépendantes de loi $\mathcal{E}(a_n)$. Alors la suite $Y_n = \min(X_1, \dots, X_n)$ converge en loi vers une variable de loi $\mathcal{E}(2)$. En effet, on a

$$\begin{aligned} \mathbb{P}(Y_n > x) &= \mathbb{P}(X_1 > x, X_2 > x, \dots, X_n > x) = \mathbb{P}(X_1 > x) \dots \mathbb{P}(X_n > x) \\ &= e^{-a_1 x} \times \dots \times e^{-a_n x} = \exp\left[-\left(\textstyle\sum_1^n a_k\right) x\right] \longrightarrow e^{-2x}. \end{aligned}$$

4.3 Les théorèmes limites

Nous pouvons à présent énoncer les deux théorèmes limite fondamentaux qui seront nos principaux outils dans la suite du cours, en particulier dans la partie du cours consacrée aux statistiques. Nous énonçons ainsi tout d'abord la loi des grands nombres puis le théorème limite central, en donnant à chaque fois des exemples d'applications de ces résultats.

4.3.1 Loi des grands nombres

La loi des grands nombres est le premier résultat fondamental de la théorie des probabilités. Elle concerne la moyenne arithmétique de variables aléatoires indépendantes et identiquement distribuées. Voici l'énoncé précis du théorème :

Théorème 4.3.1. *Soit (X_n) une suite de variables aléatoires indépendantes et de même loi, telle que $\mathbb{E}[|X_1|] < +\infty$. Alors lorsque n tend vers l'infini, on a*

$$\frac{S_n}{n} := \frac{X_1 + \ldots + X_n}{n} \xrightarrow{p.s. \text{ et } \mathbb{P}} \mathbb{E}[X_1],$$

autrement dit, pour tout $\varepsilon > 0$:

$$\lim_{n \to +\infty} \mathbb{P}\left[\left|\frac{S_n}{n} - \mathbb{E}[X_1]\right| > \varepsilon\right] = 0,$$

et même, il existe $A \subset \Omega$ tel que $\mathbb{P}(A) = 1$ et pour tout $\omega \in A$:

$$\frac{S_n(\omega)}{n} = \frac{X_1(\omega) + \ldots + X_n(\omega)}{n} \longrightarrow \mathbb{E}[X_1].$$

Remarque 4.3.2. La loi des grands nombres justifie la démarche intuitive suivante : pour connaître le résultat moyen d'une expérience aléatoire, on refait un grand nombre de fois l'expérience et on considère la moyenne arithmétique des résultats obtenus. En y réfléchissant bien, il n'est pas du tout clair a priori que la moyenne arithmétique des résultats soit une bonne approximation du résultat moyen. La loi des grands nombres justifie rigoureusement ce résultat intuitif.

Exemple : mutation d'un gène

Parmi de nombreuses causes, une certaine maladie est déclenchée par la mutation d'un gène sur un chromosome. Pour avoir une idée du nombre de personnes dans la population susceptibles d'être atteintes par cette maladie, on souhaite connaître la proportion de la population chez qui il y a eu mutation. On demande ainsi à n personnes de se soumettre à un test, on note $\{X_i = 0\}$ (resp. $\{X_i = 1\}$ les évènements il n'y a pas (resp. il y a) mutation chez la $i-$ème personne testée.

On fait l'hypothèse que les résultats des tests sont des réalisations de variables aléatoires indépendantes, autrement dit que les variables X_i sont indépendantes et

de loi de Bernoulli de paramètre p, la proportion théorique de mutation au sein de la population. D'après la loi des grands nombres, lorsque n tend vers l'infini, on a :

$$\frac{S_n}{n} := \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1] = p.$$

Autrement dit, avec une grande probabilité, si n est assez grand, la moyenne arithmétique S_n/n est proche de p. Une bonne valeur approchée de la proportion de personne chez qui la mutation est apparue est donc S_n/n. Dans la pratique, une valeur de n de l'ordre de 1000 ou 10000 fournit déjà une bonne approximation de p.

Exemple : nombre d'accidents

Afin de fixer ses primes pour l'année à venir, une compagnie d'assurance souhaite connaître le nombre moyen de sinistres auquels seront confrontés ses clients dans l'année. Les sinistres sont des évènements rares et l'expérience montrent que pour chaque client, leur nombre peut être modélisé par une variable de Poisson de paramètre λ. On suppose aussi que les nombres de sinistres pour deux clients distincts sont indépendants. La difficulté ici est choisir le paramètre $\lambda > 0$.

Pour se faire, la compagnie ouvre ses archives et observe le nombre de sinistres pour 100 de ses clients sur les 20 dernières années. On note ainsi X_i les nombres de sinistres individuels annuels, où $i = 1 \ldots 2000$. D'après la loi des grands nombres, si X_i est une suite de variables indépendantes de loi de Poisson $\mathcal{P}(\lambda)$, lorsque n tend vers l'infini, on a :

$$\frac{S_n}{n} := \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1] = \lambda.$$

Autrement dit, avec une grande probabilité, si n est assez grand, la moyenne arithmétique S_n/n est proche de λ. Une bonne valeur approchée du nombre moyen de sinistres par client chaque année est $S_{2000}/2000$.

Exemple : pile ou face

Vous jouez un grand nombre de fois de suite à pile ou face. Vous gagnez un euro à chaque pile et perdez un euro à chaque face. On note S_n le gain après n lancers. Ce gain peut sécrire sous la forme $S_n = \sum_1^n X_i$ où les X_i sont des variables aléatoires indépendantes de loi de Bernoulli $\mathcal{B}(\pm 1, p)$, $p \in [0, 1]$. Le cas d'une pièce équilibrée correspond bien sûr à $p = 1/2$. On s'intéresse aux questions du type : après n lancers, êtes-vous bénéficiaire ? quel est votre gain moyen ? etc.

La réponse à ces questions est donnée par la loi des grands nombres. En effet, d'après le théorème, lorsque n tend vers l'infini, on a :

$$\frac{S_n}{n} := \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1] = 2p - 1.$$

Ainsi, si $p > 1/2$ et n est assez grand, avec une grande probabilité on a $S_n \sim n(2p-1) > 0$ et vous êtes bénéficiaire. En revanche, lorsque $p < 1/2$, il vaut mieux arrêter de jouer rapidement sous peine d'être ruiné ! Le cas où $p = 1/2$ est plus

difficile à trancher. Pour ce faire, on a besoin d'un résultat plus fin que la loi des grands nombres.

4.3.2 Théorème limite central

La loi des grands nombres exprime le fait que la moyenne arithmétique d'une suite de variables indépendantes (X_n) de même loi converge vers la moyenne stochastique de la loi en question, c'est-à-dire $\mathbb{E}[X_1]$. Le théorème limite central est un raffinement de la loi des grands nombres : il précise à quelle vitesse a lieu cette convergence, et comment la moyenne arithmétique fluctue autour de sa limite.

Théorème 4.3.3. *Soit X_n une suite de variables aléatoires indépendantes et de même loi, telle que $\mathbb{E}[|X_1|] < +\infty$ et $\mathbb{E}[|X_1|^2 < +\infty$. On note $m = \mathbb{E}[X_1]$ et $\sigma^2 = var(X_1)$. Alors, lorsque n tend vers l'infini, on a*

$$\sqrt{n} \times \left(\frac{S_n}{n} - m \right) := \frac{(X_1 - m) + \ldots + (X_n - m)}{\sqrt{n}} \xrightarrow{loi} \mathcal{N}(0, \sigma^2),$$

ou de manière équivalente :

$$\frac{\sqrt{n}}{\sigma} \times \left(\frac{S_n}{n} - m \right) := \frac{(X_1 - m) + \ldots + (X_n - m)}{\sigma \times \sqrt{n}} \xrightarrow{loi} \mathcal{N}(0, 1).$$

Autrement dit, pour tout x dans $\mathbb{R}$, lorsque n tend vers l'infini, on a

$$\mathbb{P}\left[\frac{\sqrt{n}}{\sigma} \times \left(\frac{S_n}{n} - m \right) \leqslant x \right] \longrightarrow \mathbb{P}(\mathcal{N}(0, 1) \leqslant x),$$

ou encore pour tout intervalle $[a, b] \subset \mathbb{R}$:

$$\mathbb{P}\left[\frac{\sqrt{n}}{\sigma} \times \left(\frac{S_n}{n} - m \right) \in [a, b] \right] \longrightarrow \mathbb{P}(\mathcal{N}(0, 1) \in [a, b]).$$

Remarque 4.3.4. Le théorème limite central exprime le fait que dans la loi des grands nombres, les fluctuations autour de la moyenne limite sont de l'ordre de $1/\sqrt{n}$ et que la loi de ces fluctuations est universelle : elle est gaussienne et ne dépend pas la loi initiale des variables X_i :

$$S_n = n \times m + \sigma \times \sqrt{n}\, \mathcal{N}(0, 1) + o(\sqrt{n}),$$

$$i.e.\ \ \frac{S_n}{n} = m + \frac{\sigma}{\sqrt{n}}\, \mathcal{N}(0, 1).$$

L'universalité des fluctuations explique pourquoi la loi normale est omniprésente dans la modélisation de phénomènes aléatoires.

Exemple : mutation d'un gène

On reprend l'exemple de l'estimation de la proportion de mutation dans la population. On souhaite préciser l'erreur commise lorsque l'on approche la valeur théorique $p = \mathbb{E}[X_1]$ par la moyenne empirique S_n/n. On rappelle que la variance d'une variable de Bernoulli de paramètre p est $\sigma^2 = p(1-p)$. D'après le théorème limite central, lorsque n tend vers l'infini, on a

$$\frac{\sqrt{n}}{\sqrt{p(1-p)}} \times \left(\frac{S_n}{n} - p \right) := \frac{(X_1 - p) + \ldots + (X_n - p)}{\sqrt{np(1-p)}} \xrightarrow{loi} \mathcal{N}(0,1).$$

En prenant les valeurs absolues, on obtient :

$$\frac{\sqrt{n}}{\sqrt{p(1-p)}} \times \left| \frac{S_n}{n} - p \right| \xrightarrow{loi} |\mathcal{N}(0,1)|.$$

Soit $x_0 = 1.961$ de sorte que $\mathbb{P}(|\mathcal{N}(0,1)| > x_0) \leqslant 5\%$. Lorsque n tend vers l'infini, on a alors,

$$\mathbb{P}\left[\frac{\sqrt{n}}{\sqrt{p(1-p)}} \times \left| \frac{S_n}{n} - p \right| > x_0 \right] \longrightarrow \mathbb{P}(|\mathcal{N}(0,1)| > x_0) \leqslant 5\%.$$

Autrement dit,

$$\mathbb{P}\left(p \notin \left[\frac{S_n}{n} - \sqrt{\frac{p(1-p)}{n}} x_0, \frac{S_n}{n} + \sqrt{\frac{p(1-p)}{n}} x_0 \right] \right) \longrightarrow \mathbb{P}(|\mathcal{N}(0,1)| > x_0) \leqslant 5/\%.$$

Comme on a toujours $p(1-p) < 1/4$, on conclut que lorsque n tend vers l'infini :

$$\mathbb{P}\left(p \notin \left[\frac{S_n}{n} - \frac{x_0}{2\sqrt{n}}, \frac{S_n}{n} + \frac{x_0}{2\sqrt{n}} \right] \right) \leqslant 5\%.$$

Pour n assez grand, on peut donc affirmer qu'avec une probabilité supérieure à 95%, le taux de mutation moyen p appartient à l'intervalle

$$I_n := \left[\frac{S_n}{n} - \frac{x_0}{2\sqrt{n}}, \frac{S_n}{n} + \frac{x_0}{2\sqrt{n}} \right].$$

Exemple : nombre d'accidents

On reprend l'exemple précédent du nombre de sinistres. Là encore, on souhaite contrôler l'erreur commise en disant que le nombre moyen de sinistre λ est proche de S_n/n. On rappelle que la variance d'une loi de Poisson de paramètre λ est $\sigma^2 = \lambda$. D'après le théorème limite central, lorsque n tend vers l'infini, on a

$$\sqrt{\frac{n}{\lambda}} \times \left(\frac{S_n}{n} - \lambda \right) := \frac{(X_1 - \lambda) + \ldots + (X_n - \lambda)}{\sqrt{n\lambda}} \xrightarrow{loi} \mathcal{N}(0,1).$$

En prenant les valeurs absolues, on obtient alors :

$$\sqrt{\frac{n}{\lambda}} \times \left| \frac{S_n}{n} - \lambda \right| \xrightarrow{loi} |\mathcal{N}(0,1)|.$$

Soit $x_0 = 2.5759$ de sorte que $\mathbb{P}(|\mathcal{N}(0,1)| > x_0) < 1\%$. Lorsque n tend vers l'infini, on a alors,

$$\mathbb{P}\left[\sqrt{\frac{n}{\lambda}} \times \left| \frac{S_n}{n} - \lambda \right| > x_0 \right] \longrightarrow \mathbb{P}(|\mathcal{N}(0,1)| > x_0) < 1\%.$$

Autrement dit,

$$\mathbb{P}\left(\lambda \notin \left[\frac{S_n}{n} - \sqrt{\frac{\lambda}{n}} \times x_0, \frac{S_n}{n} + \sqrt{\frac{\lambda}{n}} \times x_0 \right] \right) \longrightarrow \mathbb{P}(|\mathcal{N}(0,1)| > x_0) < 1/\%.$$

On peut montrer que la convergence a encore lieu lorsque l'on remplace la variance λ par S_n/n, *i.e.*

$$\mathbb{P}\left(\lambda \notin \left[\frac{S_n}{n} - \sqrt{\frac{S_n}{n^2}} \times x_0, \ \frac{S_n}{n} + \sqrt{\frac{S_n}{n^2}} \times x_0 \right] \right) \leqslant 1\%.$$

Pour n assez grand, on peut donc affirmer qu'avec une probabilité supérieure à 99%, le nombre moyen d'accidents λ appartient à l'intervalle

$$I_n = \left[\frac{S_n}{n} - \sqrt{\frac{S_n}{n^2}} \times x_0, \ \frac{S_n}{n} + \sqrt{\frac{S_n}{n^2}} \times x_0 \right].$$

Exemple : pile ou face

On précise maintenant l'évolution du gain dans un jeu de pile ou face symétrique, *i.e.* lorsque $p = 1/2$. La loi des grands nombres donne :

$$\frac{S_n}{n} := \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1] = 2p - 1 = 0.$$

Le théorème limite central précise :

$$2\sqrt{n} \times \frac{S_n}{n} \xrightarrow{loi} \mathcal{N}(0,1).$$

Pour tout intervalle $[a,b] \in \mathbb{R}$, on a donc

$$\mathbb{P}\left(\frac{2S_n}{\sqrt{n}} \in [a,b] \right) \longrightarrow \mathbb{P}(\mathcal{N}(0,1) \in [a,b]) > 0.$$

Autrement dit, la gain normalisé $S_n/\sqrt{n}$ visite n'importe quel intervalle avec une probabilité strictement positive.

Exemple : prix d'une action

Le prix S_n d'une action au jour n est modélisé ainsi : $S_0 = s > 0$ est fixé, et $S_{n+1} = (1 + r + \sigma\varepsilon_{n+1})S_n$, où $r > 0$ est un taux fixe, $\sigma \in]0, 1 + r[$ est une volatilité fixe, et $(\varepsilon_n,\, n \in \mathbb{N})$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi de Bernoulli $B(\pm 1, 1/2)$. On souhaite répondre aux questions suivantes :

1. Étudier le comportement des suites $(\log S_n)/n$ et S_n.
2. Étudier le comportement de la suite $(\log S_n)/\sqrt{n}$ lorsque $(1 + r)^2 = 1 + \sigma^2$.
3. Étudier le comportement de la suite suivante :

$$\left[(1 + r)^2 - \sigma^2\right]^{\left(-1/(2\sqrt{n})\right)} \times S_n^{1/\sqrt{n}}.$$

On montre tout d'abord aisément par récurrence que pour $n > 0$, on a

$$S_n = \prod_{i=1}^{n} (1 + r + \sigma\varepsilon_i)s.$$

En prenant le logarithme, on obtient :

$$\log(S_n) = \log s + \sum_{i=1}^{n} Y_i, \quad \text{où l'on a posé } \; Y_i := \log(1 + r + \sigma\varepsilon_i).$$

Comme les variables ε_i, les variables Y_i sont indépendantes et identiquement distribuées. Par ailleurs, comme $0 < \sigma < 1 + r$, on a

$$\mathbb{E}[|Y_1|] = \frac{1}{2}|\log(1 + r + \sigma)| + \frac{1}{2}|\log(1 + r - \sigma)| < +\infty,$$

et

$$\mathbb{E}[Y_1^2] = \frac{1}{2}|\log(1 + r + \sigma)|^2 + \frac{1}{2}|\log(1 + r - \sigma)|^2 < +\infty.$$

Le calcul de l'espérance $m = \mathbb{E}[Y_1]$ et de la variance $\sigma^2 = \mathbb{E}[Y_1^2] - \mathbb{E}[Y_1]^2$ donne :

$$\begin{aligned}
m &= \tfrac{1}{2}\log(1 + r + \sigma) + \tfrac{1}{2}\log(1 + r - \sigma) \\
&= \tfrac{1}{2}\log\left((1 + r - \sigma) \times (1 + r - \sigma)\right) \\
&= \tfrac{1}{2}\log\left((1 + r)^2 - \sigma^2\right) = \log\left(\sqrt{(1 + r)^2 - \sigma^2}\right), \\
\sigma^2 &= \tfrac{1}{4}\log(1 + r + \sigma^2) + \tfrac{1}{4}\log(1 + r - \sigma)^2 - \tfrac{1}{2}\log(1 + r + \sigma)\log(1 + r - \sigma) \\
&= \tfrac{1}{4}\left(\log(1 + r + \sigma) - \log(1 + r - \sigma)\right)^2 = \tfrac{1}{4}\left(\log\left(\tfrac{1+r+\sigma}{1+r-\sigma}\right)\right)^2.
\end{aligned}$$

Il s'agit ici naturellement d'utiliser la loi des grands nombres et le théorème limite central.

1. Les variables Y_i satisfont aux hypothèses de la loi des grands nombres, d'après le théorème 4.3.1 on peut affirmer que lorsque n tend vers l'infini :

$$\frac{\log(S_n)}{n} = \frac{\log s}{n} + \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{\mathbb{P}} m = \mathbb{E}[Y_1].$$

On a donc, lorsque n tend vers l'infini :

$$S_n = \exp\left(n \times m + o_{\mathbb{P}}(n)\right).$$

Il y a donc une dichotomie selon que $m > 0$ ou $m < 0$. Si $m > 0$, c'est-à-dire si $(1+r)^2 > 1+\sigma^2$ le prix de l'action croît exponentiellement vite vers l'infini. $m < 0$, c'est-à-dire si $(1+r)^2 < 1+\sigma^2$ la prix de l'action tend exponentiellement vite vers zéro. Le cas $m = 0$ est plus subtil.

2. Le cas où $(1+r)^2 = 1+\sigma^2$ correspond au cas $m = 0$. Comme Y_i admet un moment d'ordre deux, on peut appliquer le théorème limite central, lorsque n tend vers l'infini :

$$\frac{\log(S_n)}{\sqrt{n}} \xrightarrow{loi} \mathcal{N}(0, \sigma^2).$$

Pour tout intervalle $[a, b] \in \mathbb{R}$, on a donc

$$\mathbb{P}\left(\frac{\log(S_n)}{\sqrt{n}} \in [a, b]\right) \longrightarrow \mathbb{P}(\mathcal{N}(0, \sigma^2) \in [a, b]) > 0.$$

Autrement dit, le prix normalisé $\log(S_n)/\sqrt{n}$ visite n'importe quel intervalle avec une probabilité strictement positive.

3. Plus généralement, le théorème 4.3.3 donne :

$$\sqrt{n}\left(\frac{\log(S_n)}{n} - m\right) \xrightarrow{loi} \mathcal{N}(0, \sigma^2),$$

c'est-à-dire :

$$\left(\frac{\log(S_n)}{\sqrt{n}} - m \times \sqrt{n}\right) \xrightarrow{loi} \mathcal{N}(0, \sigma^2),$$

et en prenant l'exponentielle :

$$\exp\left[\frac{\log(S_n)}{\sqrt{n}}\right] \exp\left[-m \times \sqrt{n}\right] \xrightarrow{loi} \exp\left(\mathcal{N}(0, \sigma^2)\right),$$

ou encore

$$S_n^{1/\sqrt{n}} e^{-m\sqrt{n}} = S_n^{1/\sqrt{n}} \times \left[(1+r)^2 - \sigma^2\right]^{\left(-1/(2\sqrt{n})\right)} \xrightarrow{loi} \exp\left(\mathcal{N}(0, \sigma^2)\right).$$

Deuxième partie

Éléments de statistiques

Chapitre 5

Estimation et intervalle de confiance

Nous abordons à présent la partie statistique du cours. L'objectif général de la statistique est de décrire / expliquer un phénomène aléatoire à partir d'un certain nombre d'observations de celui-ci. Le langage utilisé pour modéliser le phénomène aléatoire est naturellement celui de la théorie des probabilités. Le cas typique est celui-ci : on observe n fois un phénomène aléatoire de loi inconnue et on recueille ainsi des données $(x_1, \dots, x_n)$. On fait alors l'hypothèse que les données x_i sont les réalisations de variables aléatoires indépendantes X_i de même loi que la loi inconnue $\mathbb{P}_X$, c'est-à-dire $x_i = X_i(\omega)$ où $\mathbb{P}_{X_i} = \mathbb{P}_X$. L'objet de la statistique (inférentielle) est précisément d'estimer la loi $\mathbb{P}_X$, ou plus modestement d'estimer certaines de ses caractéristiques (moyenne, variance etc.).

5.1 Estimation paramétrique

Le plus souvent, on fait l'hypothèse que la loi inconnue $\mathbb{P}_X$ appartient à une famille de lois connue, famille indexée par un ou plusieurs paramètres. Par exemple, la loi inconnue peut être une loi de Bernoulli $\mathcal{B}(p)$, pour un certain réel $p \in [0, 1]$, elle peut être un loi de Poisson $\mathcal{P}(\lambda)$ ou exponentielle $\mathcal{E}(\lambda)$ de paramètre $\lambda > 0$, ou encore une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ avec $\mu \in \mathbb{R}$ et $\sigma^2 > 0$. Dans ce cas, estimer la loi inconnue $\mathbb{P}_X$ revient alors à estimer (*i.e.* deviner) la valeur du/des paramètre(s).

Exemple 5.1.1. Dans l'exemple de la mutation d'un gène du chapitre précédent, on sait a priori que les variables X_i sont à valeurs dans l'ensemble $\{0, 1\}$ de sorte que X_i suit une loi de Bernoulli de paramètre p inconnu. Déterminer la loi des variables X_i revient donc à déterminer la valeur du paramètre $p \in [0, 1]$.

On introduit alors la notion d'estimateur du/des paramètre(s) inconnu(s), il s'agit d'une fonction des données $(x_1, \dots, x_n)$ dont on espère qu'elle est un bonne approximation, en un sens à préciser, du/des paramètres inconnu(s).

Définition 5.1.2 (estimateur). On appelle *estimateur* de θ toute quantité $\widehat{\theta}_n$ qui est une fonction des données $(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega))$.

Remarque 5.1.3. Attention, un estimateur est une fonction des seules données connues $(x_1, \ldots, x_n)$, mais il ne doit pas, par définition, dépendre du paramètre inconnu que l'on souhaite estimer.

Il faut maintenant préciser ce que l'on entend par "être une bonne approximation du paramètre inconnu θ". La notion de biais prend en compte le fait qu'en moyenne, l'estimateur $\widehat{\theta}_n$ est proche de la valeur théorique inconnue :

Définition 5.1.4 (estimateur sans biais)**.** Le *biais* est d'un estimateur $\widehat{\theta}_n$ de θ est la différence : $\theta - \mathbb{E}[\widehat{\theta}_n]$. Si $\mathbb{E}[\widehat{\theta}_n] = \theta$, on dira que l'estimateur $\widehat{\theta}_n$ est sans biais. Si $\lim_{n\to\infty} \mathbb{E}[\widehat{\theta}_n] = \theta$, on dira que l'estimateur $\widehat{\theta}_n$ est *asymptotiquement sans biais.*

Par ailleurs, on veut que lorsque la taille de l'échantillon de données $(x_1, \ldots, x_n)$ devient grande, l'estimateur $\widehat{\theta}_n$ soit arbitrairement proche de la valeur théorique θ.

Définition 5.1.5 (estimateur consistant)**.** On dit que l'estimateur $\widehat{\theta}_n$ de la quantité θ est *consistant* si lorsque n tend vers l'infini, $\widehat{\theta}_n$ converge en probabilité vers θ.

Exemple 5.1.6. On reprend l'exemple du taux de mutation du chapitre précédent. On souhaite estimer le paramètre p de la loi de Bernoulli $\mathcal{B}(p)$. La quantité ci-dessous est un estimateur de p :

$$\widehat{p}_n := \frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n}.$$

En effet, $\widehat{p}_n(\omega) = (x_1 + \ldots + x_n)/n$ est bien une fonction des seules variables $(x_1, \ldots, x_n)$. Par ailleurs, c'est un estimateur sans biais puisque :

$$\mathbb{E}[\widehat{p}_n] := \frac{\mathbb{E}[S_n]}{n} = \frac{\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]}{n} = \frac{p + \ldots + p}{n} = p.$$

Enfin, c'est un estimateur consistant puisque d'après la loi des grands nombres :

$$\widehat{p}_n := \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1] = p.$$

On peut considérer de nombreux autres estimateurs de la quantité p, l'important est de garder à l'esprit que ce que l'on souhaite est approcher au mieux le paramètre p. Par exemple, $\widetilde{p}_n = X_1$ est bien une fonction des seules données. C'est donc un estimateur de p, et on peut ajouter qu'il est sans biais puisque si $X_1 \sim \mathcal{B}(p)$, on a $\mathbb{E}[X_1] = p$ et donc $\mathbb{E}[\widetilde{p}_n] = p$. En revanche, $\widetilde{p}_n$ n'est pas consistant puisqu'il ne dépend pas du nombre n de données. Au contraire, l'estimateur

$$\dot{p}_n := \frac{X_1 + \ldots + X_n}{n+1}$$

possède un biais de $p - \mathbb{E}[\dot{p}_n] = p/(n+1)$. Il est donc asymptotiquement sans biais, et d'après la loi des grands nombres, il est consistant.

5.1.1 Estimateurs empiriques

Nous introduisons maintenant une classe importante et naturelle d'estimateurs : les estimateurs empiriques. Ce sont les estimateurs construits à partir de somme de variables aléatoires et dont le comportement asymptotique peut être facilement décrit grâce à la loi des grands nombres et au théorème limite central.

Définition 5.1.7 (estimateurs empiriques)**.** On appelle moyenne empirique de l'échantillon $(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$ la moyenne arithmétique

$$\widehat{m}_n = \frac{x_1 + \ldots + x_n}{n} = \frac{X_1(\omega) + \ldots + X_n(\omega)}{n}.$$

On appelle variance empirique de l'échantillon $(x_1, \ldots, x_n)$ la quantité :

$$\widehat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{m}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \widehat{m}_n^2.$$

Si les variables X_i de loi inconnue admettent des moments d'ordre un et deux, alors la loi des grands nombres assure que la moyenne et la variance empirique sont des estimateurs consistants de la moyenne m et de la variance σ^2 théoriques. En effet, d'après la loi des grands nombres, on a

$$\widehat{m}_n = \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1] = m,$$

et

$$\frac{X_1^2 + \ldots + X_n^2}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1^2],$$

d'où

$$\widehat{\sigma}_n^2 \xrightarrow{\mathbb{P}} \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \sigma^2.$$

Exemple 5.1.8. On reprend l'exemple du nombre d'accidents envisagé au chapitre précédent. On fait l'hypothèse que les données sont des réalisations indépendantes de variables X_i de loi de Poisson $\mathcal{P}(\lambda)$ où λ est à déterminer. On a vu en cours et en TD que l'espérance et la variance d'une loi de Poisson sont $m = \mathbb{E}[X_i] = \lambda$ et $\sigma^2 = \mathrm{var}(X_i) = \lambda$. Pour estimer le paramètre λ, on peut donc naturellement choisir les estimateurs empiriques $\widehat{m}_n$ et $\widehat{\sigma}_n^2$.

Si le paramètre θ à déterminer s'écrit comme une fonction de la moyenne des variables de l'échantillon, c'est-à-dire si $\theta = g(\mathbb{E}[X])$ pour une certaine fonction continue g, alors un estimateur naturel de θ est donnée par :

$$\widehat{\theta}_n = g(\widehat{m}_n).$$

En effet, d'après la loi des grands nombres, lorsque n tend vers l'infini, on a alors

$$\widehat{\theta}_n = g(\widehat{m}_n) \xrightarrow{\mathbb{P}} g(\mathbb{E}[X]) = g(\theta),$$

autrement dit, l'estimateur $\widehat{\theta}_n$ est consistant.

Exemple 5.1.9. On recueille des données $(x_1, \ldots, x_n)$ dont on fait l'hypothèse qu'elles sont des réalisations indépendantes de variables X_i de loi uniforme sur un intervalle $[0, \theta]$ où θ est à déterminer. On a vu en cours et en TD que l'espérance d'une telle loi est $m = \mathbb{E}[X_i] = \theta/2$, autrement dit $\theta = 2\mathbb{E}[X]$. Alors un estimateur naturel de θ est donné par

$$\widehat{\theta}_n = \frac{2(x_1 + \ldots + x_n)}{n} = 2\widehat{m}_n.$$

En effet, d'après la loi des grands nombres, lorsque n tend vers l'infini, on a alors

$$\widehat{\theta}_n \xrightarrow{\mathbb{P}} 2\mathbb{E}[X] = \theta.$$

Exemple 5.1.10. On recueille des données $(x_1, \ldots, x_n)$ dont on fait l'hypothèse qu'elles sont des réalisations indépendantes de variables X_i de loi de exponentielle $\mathcal{E}(\lambda)$ où λ est à déterminer. On a vu en cours et en TD que l'espérance d'une loi exponentielle est $m = \mathbb{E}[X_i] = 1/\lambda$, autrement dit $\lambda = 1/\mathbb{E}[X]$. Alors un estimateur naturel de λ est donné par

$$\widehat{\lambda}_n = \frac{n}{x_1 + \ldots + x_n} = \frac{1}{\widehat{m}_n}.$$

En effet, d'après la loi des grands nombres, lorsque n tend vers l'infini, on a alors

$$\widehat{\lambda}_n \xrightarrow{\mathbb{P}} \frac{1}{\mathbb{E}[X]} = \lambda.$$

5.1.2 Maximum de vraissemblance

Nous venons de voir que la loi des grands nombres permet souvent de mettre en évidence des estimateurs naturels. Une autre façon de trouver de tels estimateurs est d'utiliser la méthode du maximum de vraissemblance décrite ci-dessous.

Considérons ainsi des réalisations x_i de variables aléatoires X_i indépendantes et de même loi, admettant une densité commune f_θ qui dépend du paramètre à estimer θ. Par exemple, les variables en question peuvent être des variables exponentielles de paramètre θ, de sorte que $f_\theta(x) = \theta \exp(-\theta x)$ pour $x \geqslant 0$.

Définition 5.1.11 (estimateur du maximum de vraissemblance)**.** Étant données des variables aléatoires de loi à densité f_θ, on appelle estimateur du maximum de vraissemblance la quantité

$$\arg\max_\theta \prod_{i=1}^{n} f_\theta(x_i),$$

ou de manière équivalente

$$\arg\max_\theta \sum_{i=1}^{n} \log\left(f_\theta(x_i)\right).$$

Dans le cas des variables exponentielles, on a

$$\prod_{i=1}^{n} f_\theta(x_i) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) = \theta^n \exp\left(-n\theta \widehat{m}_n\right)$$

où $\widehat{m}_n$ est la moyenne empirique. En prenant le logarithme, on obtient :

$$\sum_{i=1}^{n} \log\left(f_\theta(x_i)\right) = n\log(\theta) - n\theta\widehat{m}_n.$$

On cherche à trouver le maximum de cette fonction. Pour cela, on regarde quand sa dérivée par rapport à θ s'annulle. Le calcul donne :

$$\frac{\partial}{\partial\theta} n\log(\theta) - n\theta\widehat{m}_n = \frac{n}{\theta} - n\widehat{m}_n.$$

Cette expression s'annule si et seulement si $\theta = 1/\widehat{m}_n$, autrement dit :

$$\arg\max_\theta \sum_{i=1}^{n} \log\left(f_\theta(x_i)\right) = 1/\widehat{m}_n.$$

L'estimateur du maximum de vraissemblance n'est autre que $1/\widehat{m}_n$, *i.e.* on retrouve l'estimateur de l'exemple 5.1.10.

Exemple 5.1.12. On reprend l'exemple des variables uniforme sur l'intervalle $[0,\theta]$ où θ est à déterminer. La densité d'une telle variable est la fonction $f_\theta(x) = 1/\theta$ si $x \in [0,\theta]$ et zéro ailleurs. Dès lors,

$$\begin{aligned} V(\theta) := \textstyle\prod_{i=1}^{n} f_\theta(x_i) \quad &= \theta^{-n} \text{ si pour tout } i \;\; 0 \leqslant x_i \leqslant \theta, \text{ et zéro ailleurs} \\ &= \theta^{-n} \text{ si } \; 0 \leqslant \max x_i \leqslant \theta, \text{ et zéro ailleurs.} \end{aligned}$$

Le maximum de la fonction V est atteint en $\theta = \max x_i$, autrement dit, l'estimateur du maximum de vraissemblance du paramètre θ est ici donné par $\widehat{\theta}_n = \max_{i=1\ldots n} x_i$. Si l'on fixe un $\varepsilon > 0$, on a alors

$$\mathbb{P}(|\theta - \widehat{\theta}_n| > \varepsilon) = \mathbb{P}(\max X_i < \theta - \varepsilon) = \left(\frac{\theta-\varepsilon}{\theta}\right)^n \xrightarrow{n\to+\infty} 0.$$

Autrement dit, $\widehat{\theta}_n$ est un estimateur consistant de θ.

Remarque 5.1.13. Dans certains cas simples comme celui de l'estimation du paramètre d'une loi exponentielle envisagé ci-dessus, l'estimateur obtenu via la méthode du maximum de vraisemblance coïncide avec l'estimateur empirique. Ce n'est pas le cas en général comme en atteste le dernier exemple concernant la loi uniforme. Dans les cas où la maximisation de la vraisemblance est explicitement possible, et lorsqu'il diffère de l'estimateur empirique, on préfèrera l'estimateur du maximum de vraisemblance dont on peut montrer qu'il possède en général de meilleures propriétés asymptotiques.

5.2 Intervalles de confiance

Dans la section précédente, nous avons vu différentes méthodes pour estimer les paramètres d'une loi de probabilité inconnue. Nous avons par ailleurs introduit les notions de biais et de consistance qui permettent d'évaluer qualitativement un estimateur. Il arrive souvent dans la pratique que l'on veuille de plus évaluer quantitativement la qualité d'un estimateur : on peut par exemple chercher à savoir à quel vitesse (en fonction de la taille de l'échantillon) il converge vers la quantité à estimer, ou encore quelle est la probabilité de se tromper en disant que l'estimateur est proche de sa cible etc. La notion d'*intervalle de confiance*, on parle aussi de *zone de confiance*, permet précisément de quantifier la qualité d'un estimateur.

Définition 5.2.1. Soit $\alpha \in]0,1[$. On dit qu'un intervalle $I = I(X_1, \ldots, X_n)$ qui s'exprime en fonction de l'echantillon est un *intervalle de confiance* pour θ de niveau $1-\alpha$ si

$$\mathbb{P}(\theta \in I(X_1, \ldots, X_n)) = 1 - \alpha.$$

Lorsque $\mathbb{P}(\theta \in I(X_1, \ldots, X_n)) \geqslant 1-\alpha$, on parle d'intervalle de confiance de niveau $1-\alpha$ par excès.

Remarque 5.2.2. Les niveaux usuels sont 90%, 95% et 99% et correspondent respectivement à $\alpha = 10\%$, $\alpha = 5\%$ et $\alpha = 1\%$. Pour obtenir le maximum d'information, il faut s'efforcer de construire l'intervalle de confiance le moins large possible qui satisfait la condition de minoration donnée dans la définition.

Exemple 5.2.3. Considérons un $n-$échantillon $(X_1, \ldots, X_n)$ de variables aléatoires gaussiennes $\mathcal{N}(\mu, 1)$ où la moyenne μ est inconnue. Si $\widehat{X}_n$ désigne la moyenne empirique de l'échantillon, il facile de voir que la variable $Z = \sqrt{n} \times (\widehat{X}_n - \mu)$ a la même loi qu'une gaussienne $\mathcal{N}(0,1)$. Soit alors $\alpha = 5\%$, et $\beta = 1,960$ de sorte que $\mathbb{P}(|\mathcal{N}(0,1)| > \beta) = \alpha$. Alors, on a

$$\mathbb{P}(|Z| \leqslant \beta) = 1 - \alpha,$$

c'est-à-dire

$$\mathbb{P}\left(\mu \in \left[\widehat{X}_n - \frac{\beta}{\sqrt{n}}, \widehat{X}_n + \frac{\beta}{\sqrt{n}}\right]\right) = 1 - \alpha,$$

autrement dit, $I = [\widehat{X}_n - \beta/\sqrt{n}, \widehat{X}_n + \beta/\sqrt{n}]$ est un intervalle de confiance de niveau α pour le paramètre θ.

Définition 5.2.4. Soit $\alpha \in]0,1[$. On appelle intervalle de confiance asymptotique pour θ de niveau $1-\alpha$ une suite $I_n = I(X_1, \ldots, X_n)$ d'intervalles de confiance tels que

$$\lim_{n \to +\infty} \mathbb{P}(\theta \in I_n(X_1, \ldots, X_n)) = 1 - \alpha.$$

Exemple 5.2.5. On reprend les exemples du chapitre précédent sur les théorèmes limites fondamentaux. Dans le cas du taux de mutation d'un gène, d'après la loi des grands nombres, la moyenne empirique $\widehat{p}_n = S_n/n$ est un estimateur consistant du paramètre inconnu p. Soit $x_0 = 1.96$ de sorte que $\mathbb{P}(|\mathcal{N}(0,1)| > x_0) = 5\%$. D'après le théorème limite central, lorsque n tend vers l'infini, un intervalle de confiance asymptotique pour p de niveau 95% est donné par :

$$I_n := \left[\widehat{p}_n - \frac{x_0}{2\sqrt{n}},\ \widehat{p}_n + \frac{x_0}{2\sqrt{n}}\right].$$

De la même façon, dans le cas de l'estimation de la moyenne λ d'une loi de Poisson (nombre de sinistres), on a vu que si $x_0 = 2.5758$ l'intervalle suivant est un intervalle de confiance asymptotique pour λ de niveau 99% :

$$I_n = \left[\frac{S_n}{n} - \sqrt{\frac{S_n}{n^2}} \times x_0,\ \frac{S_n}{n} + \sqrt{\frac{S_n}{n^2}} \times x_0\right].$$

Exemple 5.2.6. Un sondage auprès d'un échantillon de n personnes sur leur intention de vote au second tour d'une élection indique que 46% des sondés veulent voter pour A et 54% pour B. On veut donner un intervalle de confiance asymptotique de niveau 95% de la proportion p des français qui souhaitent voter pour A. On peut modéliser les réponses des sondés (pris au hasard dans la population) par des variables aléatoires X_i de loi de Bernoulli $\mathcal{B}(p)$: $X_i = 1$ si la $i-$ème personne interrogée vote pour A, $X_i = 0$ si la $i-$ème personne interrogée vote pour B. D'après l'énoncé, la proportion de personne ayant l'intention de voter pour A, c'est-à-dire la moyenne empirique $\widehat{X}_n$ vaut 46%. Comme dans le cas du taux de mutation, si $x_0 = 1.96$ de sorte que $\mathbb{P}(|\mathcal{N}(0,1)| > x_0) = 5\%$, on montre qu'un intervalle de confiance asymptotique pour la proportion p est donné par :

$$I_n := \left[\widehat{X}_n - \frac{x_0}{2\sqrt{n}},\ \widehat{X}_n + \frac{x_0}{2\sqrt{n}}\right].$$

Si $n = 100$ on obtient ainsi l'intervalle

$$I_n = \left[0.46 - \frac{1.96}{2 \times 10}, 0.46 + \frac{1.96}{2 \times 10}\right] \approx [0.36, 0.55],$$

et l'issue de l'élection est très incertaine. Lorsque $n = 1000$, on trouve

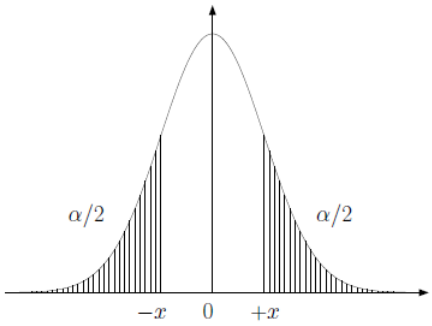
$$I_n = \left[0.46 - \frac{1.96}{2 \times \sqrt{1000}}, 0.46 + \frac{1.96}{2 \times \sqrt{1000}}\right] \approx [0.43, 0.49],$$

et avec 95 chances sur 100, on peut affirmer que le candidat A perdra l'élection.

Tables de la loi normale centrée réduite

$$2\int_x^\infty \mathrm{e}^{-y^2/2}\ \frac{dy}{\sqrt{2\pi}} = \mathbb{P}(|X| \geq x) = \alpha.$$

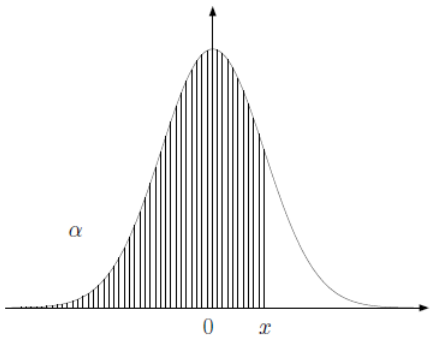
La table donne les valeurs de x en fonction de α. Par exemple $\mathbb{P}(|X| \geq 0.6280) \simeq 0.53$.



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	∞	2.5758	2.3263	2.1701	2.0537	1.9600	1.8808	1.8119	1.7507	1.6954
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.4395	1.4051	1.3722	1.3408	1.3106
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.1503	1.1264	1.1031	1.0803	1.0581
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.9346	0.9154	0.8965	0.8779	0.8596
0.4	0.8416	0.8239	0.8064	0.7892	0.7722	0.7554	0.7388	0.7225	0.7063	0.6903
0.5	0.6745	0.6588	0.6433	0.6280	0.6128	0.5978	0.5828	0.5681	0.5534	0.5388
0.6	0.5244	0.5101	0.4959	0.4817	0.4677	0.4538	0.4399	0.4261	0.4125	0.3989
0.7	0.3853	0.3719	0.3585	0.3451	0.3319	0.3186	0.3055	0.2924	0.2793	0.2663
0.8	0.2533	0.2404	0.2275	0.2147	0.2019	0.1891	0.1764	0.1637	0.1510	0.1383
0.9	0.1257	0.1130	0.1004	0.0878	0.0753	0.0627	0.0502	0.0376	0.0251	0.0125

$$\int_{-\infty}^x \mathrm{e}^{-y^2/2}\ \frac{dy}{\sqrt{2\pi}} = \mathbb{P}(X \leq x) = \alpha.$$

La table suivante donne les valeurs de $1-\alpha$ pour les grandes valeurs de x.



x	2	3	4	5	6	7	8	9	10
$1-\alpha$	2.28e-02	1.35e-03	3.17e-05	2.87e-07	9.87e-10	1.28e-12	6.22e-16	1.13e-19	7.62e-24

FIGURE 5.1 – Quantiles de la loi normale centrée réduite.

Chapitre 6

Tests statistiques

L'objectif d'un test d'hypothèses est de répondre à une question que l'on formalise de la manière suivante : au vu de l'observation d'un échantillon $(X_1, \dots, X_n)$, le paramètre θ du modèle est-il ou non dans un sous-ensemble de Θ appelé hypothèse nulle et noté H_0 ? Par exemple, si on s'intéresse au changement climatique, on peut par exemple travailler sur les données de température moyenne au mois d'août à Paris. Sur l'ensemble du vingtième siècle, ces températures moyennes en degrés Celsius sont bien décrites par une loi gaussienne $\mathcal{N}(20, 1)$. Sur les dix dernières années, on a observé les températures moyennes suivantes : $x = (x_1, \dots, x_{10}) = (22, 19, 21, 23, 20, 22, 24, 18, 20, 25)$, de sorte que $\widehat{x}_{10} = 21.4$ et $\widehat{\sigma}_{10} = 2.22$.

À partir de ces éléments, souhaite construire un test d'hypothèses pour savoir si la température moyenne a augmenté ces dix dernières années par rapport à l'ensemble du vingtième siècle. Bien sûr le fait que la moyenne empirique sur les dix dernières années dépasse 20 va plutôt dans le sens d'un réchauffement mais il faut procéder de manière plus fine pour pouvoir contrôler la probabilité d'affirmer à tort qu'il y a eu réchauffement.

6.1 Tests d'hypothèses

Comme précédemment dans le cas de l'estimation paramétrique, on supposera ici que les observations recueillies $x = (x_1, \dots, x_n)$ sont les réalisations d'un échantillon $X = (X_1, \dots, X_n)$ de variables aléatoires *i.i.d* de loi inconnue $\mathbb{P}_X$. On supposera de plus que la loi inconnue $\mathbb{P}_X$ appartient à une famille paramétrée de loi que l'on on notera $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$. Par exemple, la loi $\mathbb{P}_X$ pourra être une loi de Bernoulli *i.e.* $\mathcal{P} = \{\mathcal{B}(\theta), \theta \in \Theta = [0, 1]\}$, une loi exponentielle *i.e.* $\mathcal{P} = \{\mathcal{E}(\theta), \theta \in \Theta = \mathbb{R}_+^*\}$, une loi gaussienne *i.e.* $\mathcal{P} = \{\mathbb{P}_\theta = \mathcal{N}_\theta, \text{ avec } \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+^*\}$ etc.

Dans la suite, (H_0, H_1) désignera une partition de l'ensemble Θ des paramètres, *i.e.* on aura toujours : $\Theta = H_0 \cup H_1$ et $\emptyset = H_0 \cap H_1$. Par exemple dans le cas de variables de Bernoulli, on a $\Theta = [0, 1]$ et $H_0 := [0, 1/2[$, $H_1 := [1/2, 1]$ constituent une partition de Θ ; de même pour les ensembles $H_0 := \{1/4\}$ et $H_1 := [0, 1] \backslash \{1/4\}$.

6.1.1 Définitions

Ayant introduit la partition (H_0, H_1) de l'ensemble Θ des paramètres, nous pouvons à présent introduire la notion de test d'hypothèses qui consiste à construire à partir des données une règle de décision pour savoir si le paramètre inconnu θ de la loi est dans H_0 ou dans H_1.

Définition 6.1.1. On appelle test d'hypothèses une règle de décision qui au vu de l'observation X permet de décider si θ est dans l'ensemble H_0 appelé hypothèse nulle ou si θ est dans l'ensemble H_1 appelé hypotèse alternative. Un test est déterminé par sa région critique W qui constitue un sous-ensemble des valeurs possibles de $X = (X_1, \ldots, X_n)$. Lorsque l'on observe $x = (x_1, \ldots, x_n)$,

— si $x \in W$, alors on rejette H_0 et on accepte H_1 *i.e.* on décide que $\theta \in H_1$,

— si $x \notin W$, alors on accepte H_0 et on rejette H_1 *i.e.* on décide que $\theta \in H_0$.

Définition 6.1.2. On appelle erreur de première espèce le rejet de H_0 à tort. Cette erreur est mesurée par le risque de première espèce : $\theta \in H_0 \mapsto \mathbb{P}_\theta(X \in W)$. On appelle erreur de seconde espèce le rejet de H_1 à tort. Cette erreur est mesurée par le risque de seconde espèce : $\theta \in H_1 \mapsto P_\theta(X \in W^c)$. Par convention, on minimise en priorité le risque de première espèce.

Définition 6.1.3. On appelle niveau du test le nombre $\alpha(W) = \sup_{\theta \in H_0} \mathbb{P}_\theta(W)$. Parmi les tests de niveau inférieur à un seuil α fixé, on souhaite minimiser le risque de seconde espèce. En général, on choisit $\alpha = 10\%$, $\alpha = 5\%$ ou $\alpha = 1\%$.

Remarque 6.1.4. Lors d'un test, on minimise en priorité le risque de première espèce, aussi les rôles de l'hypothèse nulle H_0 et de l'hypothèse alternative H_1 ne sont pas symétriques. Le choix de H_0 parmi deux ensembles constituant une partition de Θ dépend donc du problème considéré : on choisit comme hypothèse nulle l'ensemble que l'on ne souhaite surtout pas voir rejeté à tort : hypothèse à laquelle on tient, hypothèse de prudence, hypothèse solidement établie etc. Par exemple, dans le test de dépistage d'une maladie, on souhaite surtout éviter de dire à une personne qu'elle est en bonne santé alors qu'elle est en fait malade. On choisit comme hypothèse nulle le fait d'être malade. Dans le cas du réchauffement climatique, un homme politique qui veut éviter de prendre des mesures si le réchauffement n'est pas avéré choisira comme hypothèse nulle "il n'y a pas réchauffement". Un écologiste choisira plutôt "il y a réchauffement".

Exemple 6.1.5. Commençons par un exemple très simple. On suppose que l'on observe une seule donnée $x_1 = 2,165$, réalisation d'une variable aléatoire X_1 de loi $\mathcal{N}(\mu, 1)$ où la moyenne μ appartient à l'ensemble à deux éléments $\{0, 5\}$. Au vu de cette observation, on souhaite construire un test pour décider, avec un niveau de sécurité de α, si $\mu = 0$ ou si $\mu = 5$. On privilégie la première hypothèse, et on pose donc $H_0 = \{\mu = 0\}$ et $H_1 = \{\mu = 5\}$.

— Considérons tout d'abord le cas où $\alpha = 5\%$. Soit $\beta_{5\%} = 1.64$ de sorte que

$\mathbb{P}(\mathcal{N}(0,1) > \beta_{5\%}) = 0.05$. On définit la région de rejet $W_{5\%} = \{X_1 > \beta_{5\%}\}$ de sorte que $\mathbb{P}_0(W_{5\%}) = 0.05$. Cette zone de rejet $W_{5\%}$ fournit un test de niveau 5% de H_0 contre H_1. Dans notre exemple, on a $x_1 = 2,165 > 1.64$, on rejette donc l'hypothèse $H_0 = \{\mu = 0\}$ au niveau 5%.

— Considérons maintenant le cas où $\alpha = 1\%$. Soit $\beta_{1\%} = 2.33$ de sorte que $\mathbb{P}(\mathcal{N}(0,1) > \beta_{1\%}) = 0.01$. On définit la région de rejet $W_{1\%} = \{X_1 > \beta_{1\%}\}$ de sorte que $\mathbb{P}_0(W_{1\%}) = 0.01$. Cette zone de rejet $W_{1\%}$ fournit un test de niveau 1% de H_0 contre H_1. Dans notre exemple, on a $x_1 = 2,165 < 2.33$, on accepte donc l'hypothèse $H_0 = \{\mu = 0\}$ au niveau 1%.

Exemple 6.1.6. On suppose maintenant que l'on dispose de n données $x_1, \ldots, x_n$ et qu'elles sont des réalisations de variables aléatoires gaussiennes, c'est-à-dire de loi du type $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \{\mu_0, \mu_1\}\}$ avec $\sigma^2 > 0$ connu et $\mu_0 > \mu_1$. On souhaite tester l'hypothèse $H_0 = \{\mu = \mu_0\}$ contre $H_1 = \{\mu = \mu_1\}$. On va bien sûr accepter H_0 (resp. H_1) si la moyenne empirique $\widehat{X}_n$ est grande (resp. petite), c'est-à-dire choisir la région critique de la forme $W = \{\widehat{X}_n < a\}$. Le choix $a = (\mu_0 + \mu_1)/2$, qui peut sembler naturel, ne permet pas de contrôler le risque de première espèce. Pour obtenir ce contrôle de la probabilité de rejeter H_0 à tort, on utilise le fait que sous H_0, la statistique de test $\widehat{X}_n$ suit la loi $\mathcal{N}(\mu_0, \sigma^2/n)$. Autrement dit, si $Z \sim \mathcal{N}(0,1)$,

$$\mathbb{P}_{(\mu_0,\sigma^2)}(\widehat{X}_n < a) = \mathbb{P}_{(\mu_0,\sigma^2)}\left(\mu_0 + \frac{\sigma Z}{\sqrt{n}} < a\right) = \mathbb{P}_{(\mu_0,\sigma^2)}\left(Z < \frac{\sqrt{n}(a-\mu_0)}{\sigma}\right).$$

En choisissant a de sorte que la dernière probabilité soit inférieure à α, on obtient donc un test de niveau α.

Exemple 6.1.7. On construit maintenant un test pour l'augmentation des températures moyennes à Paris au mois d'août. On suppose que les températures recueillies sur la dernière décennie

$$x = (x_1, \ldots, x_{10}) = (22, 19, 21, 23, 20, 22, 24, 18, 20, 25)$$

sont des réalisations de variables gaussiennes $\mathcal{N}(\mu, 1)$ où la moyenne μ est inconnue et on souhaite tester l'hypothèse nulle $H_0 := \{\mu \leqslant \mu_0 = 20\}$ contre l'hypothèse alternative $H_1 := \{\mu > \mu_0 = 20\}$. Pour cela, on introduit la statistique $\xi_n = \sqrt{n} \times (\widehat{X}_n - \mu_0)$. On remarque que $\xi_n \sim \mathcal{N}(\sqrt{n} \times (\mu - \mu_0), 1)$, autrement dit ξ_n a tendance à croître avec μ ce qui incite à choisir une région de rejet de la forme $W = \{\xi_n > a\}$ pour un seuil a à déterminer. Si $Z \sim \mathcal{N}(0,1)$, on a

$$\sup_{\mu \leqslant \mu_0} \mathbb{P}_{(\mu,1)}(\xi_n > a) = \sup_{\mu \leqslant \mu_0} \mathbb{P}_{(\mu,1)}(Z > a + \sqrt{n} \times (\mu - \mu_0)) = \mathbb{P}(Z > a).$$

Soit $a = 2.33$ de sorte que $\mathbb{P}(Z > a) \leqslant 0.01$. La zone de rejet $W = \{\xi_n > 2.33\}$ fournit alors un test de H_0 contre H_1 de niveau 99%.

Si on applique ce test aux données recueillies, on trouve $\xi_{10} = \sqrt{10} \times (\widehat{x}_{10} - \mu_0) = \sqrt{10} \times (21,4 - 20) \approx 4.33$. On a donc $\xi_{10} > a = 2.33$ et on rejette l'hypothèse H_0.

En fait on a $\mathbb{P}(Z > 4.33) \approx 4.7 \times 10^{-6}$, et on rejette l'hypothèse H_0 pour tous les niveaux $\alpha > 4.7 \times 10^{-6}$, c'est-à-dire à tous les niveaux usuels. Ainsi on peut conclure à l'augmentation des températures sur les dix dernières années.

6.2 Test du χ^2

Nous introduisons maintenant une classe de tests très utilisés en pratique : les tests du χ^2 (on lit $khi - deux$). Ces tests sont basés sur la loi du χ^2 qui comme la loi gaussienne centrée réduite est tabulée. On donne la table des quantiles de la loi χ^2 pour différents degrés de liberté en fin de chapitre.

Définition 6.2.1. On dit qu'une variable aléatoire X suit une loi du χ^2 à n degrés de liberté et on note $X \sim \chi^2(n)$ si X est à valeurs dans $\mathbb{R}^+$ et X admet la densité f_X suivante :

$$f_X(x) = \frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)}.$$

Si $X_1, X_2, \dots X_n$ sont des variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$, alors la variable $Z = X_1^2 + X_2^2 + \dots + X_n^2$ suit une loi $\chi^2(n)$. En particulier, si $Z \sim \chi^2(n)$, on a :

$$\mathbb{E}[Z] = n, \text{ et } \mathrm{var}(Z) = n.$$

Ci-dessous, on donne l'allure de la densité f_X pour différents degrés de liberté, c'est-à-dire pour différentes valeurs du paramètre n.

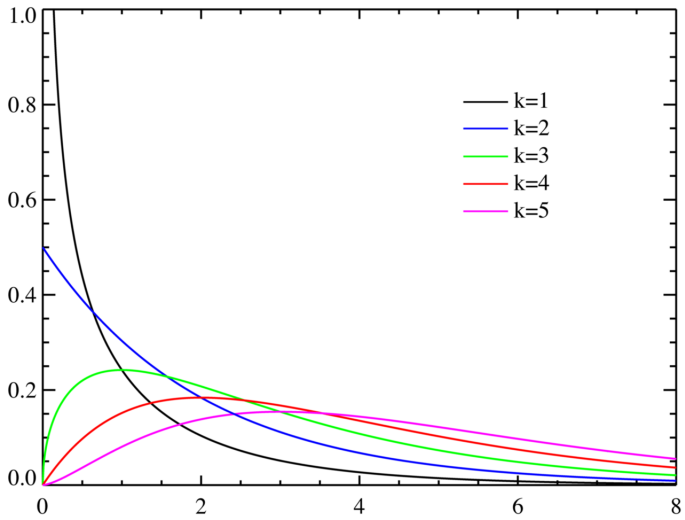


FIGURE 6.1 – Allure des densités des lois $\chi^2(k)$ pour différentes valeurs de k.

Les tests du χ^2 permettent de répondre à de nombreuses questions comme :

— un dé à six faces est-il pipé ? Pour cela on observe les fréquences d'apparition des faces lors de n lancers de ce dé et on les compare au vecteur $(1/6, \dots, 1/6)$.

— la répartition entre accouchements avec et sans césarienne dépend-elle du jour de la semaine ? Pour cela, on introduit $\widehat{q}_j$ les proportions d'accouchements qui ont lieu le $j-$ème jour de la semaine, $\widehat{r}_A$ et $\widehat{r}_S$ les proportions globales d'accouchements avec et sans césarienne et enfin $\widehat{p}_{j,l}$, $(j,l) \in \{1,\dots,7\} \times \{A,S\}$ la proportions d'accouchements qui ont lieu le $j-$ème jour de la semaine avec césarienne si $l = A$ et sans si $l = S$. Bien sûr, on va comparer la matrice $(\widehat{p}_{j,l})$ à la matrice $(\widehat{q}_j \widehat{r}_l)$.

Dans les deux prochaines sections, nous expliquons et illustrons la procédure de décision du χ^2 dans le cas de tests d'indépendance et d'adéquation à une loi donnée.

6.2.1 Test d'adéquation à une loi

Nous décrivons ici le test du χ^2 d'adéquation à une loi, qui comme son nom l'indique permet de décider si des observations sont des réalisations d'une loi donnée. On observe ainsi un échantillon $(x_1,\dots,x_n)$ de variables aléatoires $(X_1,\dots,X_n)$ *i.i.d.* à valeurs dans un ensemble fini $A = \{a_1,\dots,a_k\}$ et de loi inconnue $p = (p_1,\dots,p_k)$ où $\mathbb{P}_p(X_1 = a_j) = p_j$ pour $j \in \{1,\dots,k\}$. La loi p appartient à l'ensemble de paramètres $\Theta = \{(p_1,\dots,p_k) \in \mathbb{R}_+^k,\ p_1 + \dots + p_k = 1\}$.

On suppose par ailleurs donnée une loi a priori $p^0 = (p_1^0,\dots,p_k^0)$. On souhaite tester l'hypothèse nulle $H_0 = \{p = p^0\}$ contre l'hypothèse alternative $H_1 = \{p \neq p^0\}$, autrement dit on souhaite tester si les observations $(x_1,\dots,x_n)$ sont des réalisations de variables aléatoires $(X_1,\dots,X_n)$ de loi p^0.

Exemple 6.2.2. Dans le cas du dé à six faces évoqué plus haut, $A = \{1,\dots,6\}$ et $p^0 = (1/6,\dots,1/6)$. Tester $H_0 = \{p = p^0\}$ contre $H_1 = \{p \neq p^0\}$ revient à tester si le dé est pipé ou non.

Voici comment on met en oeuvre le test d'adéquation. Pour $j \in \{1,\dots,k\}$, on note $\widehat{p}_j := \frac{1}{n}\sum_{i=1}^n 1_{X_i = a_j}$ la fréquence empirique de a_j. Le vecteur des fréquences empiriques est alors $\widehat{p} = (\widehat{p}_1,\dots,\widehat{p}_k)$. L'idée qui est à la base du test est bien sûr que le vecteur $\widehat{p}$ est plus proche de p^0 sous l'hypothèse nulle H_0 que sous l'hypothèse alternative H_1. Afin de quantifier la "proximite", on utilise la pseudo-distance du χ^2 :

$$\xi_n := n \times \sum_{j=1}^k \frac{(\widehat{p}_j - p_j^0)^2}{p_j^0}.$$

On peut montrer le comportement asymptotique suivant :

Proposition 6.2.3. *Sous H_0, ξ_n converge en loi vers $Z \sim \chi^2(k-1)$. Sous H_1, ξ_n tend presque sûrement vers plus l'infini.*

Étant donné un niveau α (par exemple $\alpha = 5\%$) et un réel a tel que $\mathbb{P}(Z > a) = \alpha$, la zone de rejet $W_n = \{\xi_n > a\}$ fournit alors un test de niveau asymptotique α de $H_0 = \{p = p^0\}$ contre $H_1 = \{p \neq p^0\}$.

Remarque 6.2.4. En pratique, on considère que l'approximation en loi par $\chi^2(k-1)$ est valide sous H_0 si $n \times \min_{1\leqslant j\leqslant k} p_j^0 \geqslant 5$. Si cette condition n'est pas satisfaite, on peut regrouper les valeurs de a_j pour lesquelles p_j^0 est trop faible et augmenter ainsi le minimum.

Exemple 6.2.5. Lors de cent lancers d'un dé à six faces, on observe les résultats suivants :

x	1	2	3	4	5	6
$N(x)$	20	13	17	12	23	15

On veut tester au niveau de confiance 95% l'hypothèse $H_0 := \{\text{le dé n'est pas pipé}\}$ contre l'hypothèse $H_1 := \{\text{le dé est pipé}\}$. D'après les tables, si $Z \sim \chi^2(5)$, on a $\mathbb{P}(Z \geqslant 11,07) = 5\%$, autrement dit la zone de rejet est ici de la forme $\{\xi_n > 11.07\}$. Dans notre exemple, les fréquences observées sont :

x	1	2	3	4	5	6
$\widehat{p}_x$	0.2	0.13	0.17	0.12	0.23	0.15

On applique le test d'adéquation à la loi uniforme $p^0 = (1/6, \ldots, 1/6)$. La statistique de test vaut

$$\xi_{100} = 100 \times \sum_{j=1}^{6} \frac{(\widehat{p}_j - 1/6)^2}{1/6} = 600 \times \sum_{j=1}^{6} (\widehat{p}_j - 1/6)^2 \approx 5.36.$$

Comme $5.36 < 11.07$, on ne rejette pas au niveau de confiance 95% l'hypothèse H_0.

6.2.2 Test d'indépendance

Nous décrivons à présent la procédure du test d'indépendance du χ^2. La problématique est la suivante : on dispose d'un échantillon d'une loi à deux composantes $Z = (X, Y)$ et l'on souhaite déterminer si les variables X et Y sont indépendantes. Soient donc n données $(z_1, \ldots, z_n) = ((x_1, y_1), \ldots, (x_n, y_n))$ dont on suppose qu'elles sont les réalisations indépendantes de variables aléatoires $(Z_1, \ldots, Z_n) = ((X_1, Y_1), \ldots, (X_n, Y_n))$ à valeurs dans des ensembles finis :

$$X_i \in \{A_1, \ldots, A_k\}, \quad Y_i \in \{B_1, \ldots, B_\ell\}.$$

On note $p = (p_{jl}, 1 \leqslant j \leqslant k, 1 \leqslant l \leqslant \ell)$ la loi du couple $Z = (X, Y)$, c'est-à-dire :

$$p_{jl} = \mathbb{P}(Z = (A_j, B_l)) = \mathbb{P}(X = A_j, Y = B_l).$$

On introduit les fréquences empiriques

$$\widehat{p}_{jl} = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i = A_j, Y_i = B_l}, \quad \widehat{q}_j = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i = A_j}, \quad \widehat{r}_l = \frac{1}{n} \sum_{i=1}^{n} 1_{Y_i = B_l}.$$

La statistique de test

$$\xi_n = n \sum_{j,l} \frac{(\widehat{p}_{jl} - \widehat{q}_j \widehat{r}_l)^2}{\widehat{q}_j \widehat{r}_l}$$

mesure la distance entre la matrice $\widehat{p}$ des fréquences des couples (A_j, B_l) et la matrice $\widehat{qr}^*$ produit des fréquences marginales.

Proposition 6.2.6. *Sous H_0 et sous des hypothèses de régularité, ξ_n converge en loi vers $Z \sim \chi^2((k-1)(\ell-1))$. Sous H_1, ξ_n tend presque sûrement vers plus l'infini.*

Étant donné un niveau α (par exemple $\alpha = 5\%$) et un réel a tel que $\mathbb{P}(Z \geqslant a) = \alpha$, la zone de rejet $W_n = \{\xi_n > a\}$ fournit alors un test de niveau asymptotique α de $H_0 = \{X \text{ et } Y \text{ indépendantes}\}$ contre $H_1 = \{X \text{ et } Y \text{ non indépendantes}\}$.

Exemple 6.2.7. On désire étudier la répartition des naissances suivant le type du jour dans la semaine (jours ouvrables ou week-end) et suivant le mode d'accouchement (naturel ou par césarienne). Les données proviennent du "National Vital Statistics Report" et concernent les naissances aux USA en 1997.

Naissances	Naturelles	César.	Total
J.O.	2331536	663540	2995076
W.E.	715085	135493	850578
Total	3046621	799033	3845654

Naissances	Naturelles	César.	Total
J.O.	60.6%	17.3%	77.9%
W.E.	18.6%	3.5%	22.1%
Total	79.2%	20.8%	100.0%

On souhaite tester au niveau $0.1\% = 0.001$ l'hypothèse d'indépendance entre le type du jour de naissance (jour ouvrable ou week-end) et le mode d'accouchement (naturel ou césarienne).

Les fréquences observées sont $\widehat{p}_J = 0.779$, $\widehat{p}_W = 0.221$, $\widehat{p}_N = 0.792$, $\widehat{p}_C = 0.208$, $\widehat{p}_{JN} = 0.606$, $\widehat{p}_{JC} = 0.173$, $\widehat{p}_{WN} = 0.186$ et $\widehat{p}_{WC} = 0.035$ où les indices J, W, N, C signifient respectivement jour ouvrable, week-end, naissance naturelle, naissance par césarienne. On en déduit que $\widehat{p}_J\widehat{p}_N = 0.617$, $\widehat{p}_J\widehat{p}_C = 0.162$, $\widehat{p}_W\widehat{p}_N = 0.175$ et $\widehat{p}_W\widehat{p}_C = 0.046$. Sur ces observations, la statistique de test ξ_n pour le test d'indépendance est

$$\xi_{3845654}^{obs} = 16401.3.$$

Dans cet exemple, on a $k = 2$ et $\ell = 2$ et l'on a donc $\xi_{3845654} \approx \chi^2(1)$. D'après la table de la fin du chapitre, si $Z \sim \chi^2(1)$, on a $\mathbb{P}(Z > 10.83) = 0.001$. Autrement dit, la zone de rejet du test d'indépendance est $W = \{\xi_{3845654} > 10.83\}$. Comme $\xi_{3845654}^{obs} = 16401.3 > 10.83$, on rejette donc, au niveau 0.001, l'hypothèse d'indépendance entre le type du jour de naissance et le mode d'accouchement. Il y a plus de naissance par césarienne les jours ouvrables que les week-end.

Exemple 6.2.8. Un traitement est administré à trois doses différentes 1, 2 et 3, à un groupe de sujets atteints d'une même maladie. L'expérimentation est faite en double aveugle. On compte le nombre de guérisons pour chaque dose. Les résultats sont les suivants :

	Sujets guéris	Sujets non guéris	Total
Dose 1	30	30	60
Dose 2	42	35	77
Dose 3	58	31	89
Total	130	96	226

On souhaite déterminer avec un niveau de sécurité de 95% si l'efficacité du traitement est liée à la dose utilisée ? Cela revient à réaliser un test au niveau $\alpha = 5\%$ de l'hypothèse H_0 contre H_1 avec

$$H_0 := \{\text{dose et guérison sont indépendantes}\}$$

$$H_1 := \{\text{dose et guérison ne sont pas indépendantes}\}.$$

Si on note G pour guéri et M pour non guéri, les fréquences marginales observées sont ici : $\widehat{p}_G = 130/226$, $\widehat{p}_M = 96/226$, et $\widehat{p}_1 = 60/226$, $\widehat{p}_2 = 77/226$, $\widehat{p}_3 = 89/226$. D'autre part, on a $\widehat{p}_{G1} = 30/226$, $\widehat{p}_{G2} = 42/226$ et $\widehat{p}_{G3} = 58/226$, $\widehat{p}_{M1} = 30/226$, $\widehat{p}_{M2} = 35/226$ et $\widehat{p}_{M3} = 31/226$. La statistique du χ^2 est donnée par :

$$\xi_{226} = 226 \left(\sum_{j=1}^{3} \frac{(\widehat{p}_{Gj} - \widehat{p}_G \widehat{p}_j)^2}{\widehat{p}_G \widehat{p}_j} + \sum_{j=1}^{3} \frac{(\widehat{p}_{Mj} - \widehat{p}_M \widehat{p}_j)^2}{\widehat{p}_M \widehat{p}_j} \right).$$

Avec nos données, on trouve $\xi_{226}^{obs} = 3.80$. Le nombre de degrés de liberté est ici $k = (3-1) \times (2-1) = 2$ et l'on a $\mathbb{P}(\chi^2(2) > 5.99) = 5\%$. Autrement dit, l'ensemble $W := \{\xi_{226} > 5.99\}$ est une zone de rejet pour le test de H_0 contre H_1 au niveau 5%. On a ici $\xi_{226}^{obs} \notin W$ de sorte que l'on accepte l'hypothèse d'indépendance de la dose et de la guérison au niveau 5%.

Quantiles de la loi du χ^2

Soit $X_n \sim \chi^2(n)$. On pose :

$$\alpha = \mathbb{P}(X_n \geqslant x) = \int_x^{+\infty} \frac{y^{n-1}e^{-y/2}}{2^n\Gamma(n)}dy.$$

La table ci-dessous donne les valeurs de x en fonction de n et de α. Par exemple $\mathbb{P}(X_8 > 20.09) \approx 0.01$.

$n\backslash\alpha$	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.001
1	0.0002	0.0010	0.0039	0.0158	2.71	3.84	5.02	6.63	10.83
2	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	13.82
3	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	16.27
4	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	18.47
5	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	20.52
6	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	22.46
7	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	24.32
8	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	26.12
9	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	27.88
10	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	29.59
11	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	31.26
12	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	32.91
13	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	34.53
14	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	36.12
15	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	37.70
16	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	39.25
17	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	40.79
18	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	42.31
19	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	43.82
20	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	45.31
21	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	46.80
22	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	48.27
23	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	49.73
24	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	51.18
25	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	52.62
26	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	54.05
27	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	55.48
28	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	56.89
29	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	58.30
30	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	59.70

FIGURE 6.2 – Quantiles d'ordre α de la loi du χ^2 à n degrés de liberté.

Chapitre 7

Régression linéaire

Dans ce dernier chapitre, nous nous intéressons à la régression linéaire simple. Étant donnée une statistique double ou statistique bivariée $Z = (X, Y)$ où $X = (X_1, \ldots, X_n)$ et $Y = (Y_1, \ldots, Y_n)$, on cherche une relation du type $Y = f(X)$ où f est une fonction à déterminer. Le cas de la régression linéaire correspond au cas où la fonction f est linéaire, c'est-à-dire au cas où $f(x) = ax + b$ où a et b sont des constante. Graphiquement, étant donné le nuage de points (X_i, Y_i), cela revient à déterminer la droite qui "colle" le mieux au données.

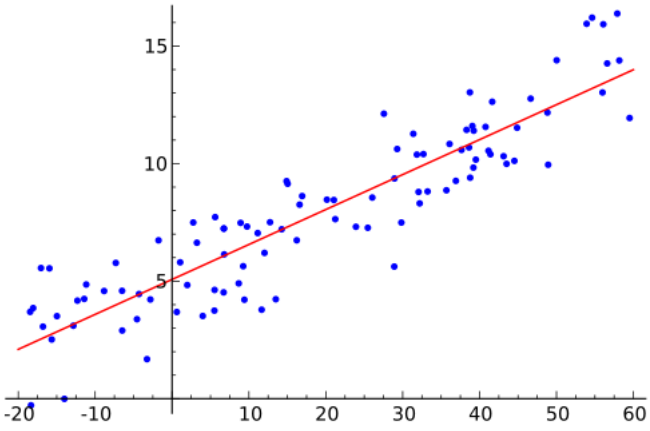


FIGURE 7.1 – Nuage de points et droite de régression.

L'objectif est double ici : il s'agit dans un premier temps d'expliquer les données Y_i en fonction des données X_i, et d'autre part d'essayer de prédire la valeur d'une nouvelle réalisation de la variable Y à partir d'une nouvelle réalisation de la variable X. Bien entendu, sauf cas exceptionnel, la relation $Y = f(X)$ ne peut être exacte. Aussi cherche-t-on la fonction f de sorte que la relation $Y = f(X)$ soit le plus près possible d'être vérifiée.

7.1 Régression linéaire simple

On se concentre ici sur le cas où $f(x) = ax + b$ avec des constantes a et b à déterminer. D'autres cas seront envisagés dans la section 7.3.

7.1.1 La problématique

La notion de proximité qu'on retient est celle qui conduit au calcul le plus simple : on cherche à minimiser la somme des carrés des distances à la droite, autrement dit on cherche le couple $(\widehat{a}_n, \widehat{b}_n)$, fonction des données (X_i, Y_i), qui minimise la fonction $\sigma^2 = \sigma^2(a, b)$:

$$\sigma^2 = \sigma^2(a, b) := \sum_{i=1}^{n} (Y_i - aX_i - b)^2.$$

La droite $D_{Y/X}$ d'équation $Y = \widehat{a}_n X + \widehat{b}_n$ correspondante est appelée la droite de régression au sens des moindres carrés (ou droite de régression) de $Y = (Y_1, \ldots, Y_n)$ par rapport à $X = (X_1, \ldots, X_n)$.

7.1.2 La solution

Pour résoudre le problème de minimisation, on rappelle quelques notations introduites dans les chapitres précédents. On considère ainsi les moyennes et variances empiriques suivantes :

$$\widehat{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad \widehat{Y}_n := \frac{1}{n} \sum_{i=1}^{n} Y_i,$$

$$\widehat{\sigma}_n^X := \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \widehat{X}_n^2, \qquad \widehat{\sigma}_n^Y := \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \widehat{Y}_n.$$

La covariance empirique κ_n^{XY} et la coefficient de corrélation linéaire ρ_n^{XY} sont alors donnés par les formules :

$$\kappa_n^{XY} := \frac{1}{n} \sum_{i=1}^{n} X_i Y_i - \widehat{X}_n \widehat{Y}_n, \qquad \rho_n^{XY} := \frac{\kappa_n^{XY}}{\sqrt{\widehat{\sigma}_n^X \widehat{\sigma}_n^Y}}.$$

On réécrit tout d'abord l'erreur quadratique σ^2 en faisant intervenir les moyennes et variances empiriques. On a ainsi :

$$\begin{aligned} \sum_{i=1}^{n} (Y_i - aX_i - b)^2 &= \sum_{i=1}^{n} \Big((Y_i - \widehat{Y}_n) - a(X_i - \widehat{X}_n) + (\widehat{Y}_n - a\widehat{X}_n - b) \Big)^2 \\ &= (\widehat{Y}_n - a\widehat{X}_n - b)^2 + a^2 \widehat{\sigma}_n^X - 2a\kappa_n^{XY} + \widehat{\sigma}_n^X, \end{aligned}$$

de sorte que, dès lors que la statistique X n'est pas constante *i.e.* $\sigma_n^X \neq 0$, il existe un unique couple $(\widehat{a}_n, \widehat{b}_n)$ qui minimise la la fonction $(a, b) \mapsto \sigma^2(a, b)$. Ce couple est

donné par :

$$\widehat{a}_n := \frac{\kappa_n^{XY}}{\widehat{\sigma}_n^X}, \qquad \widehat{b}_n := \widehat{Y}_n - \widehat{a}_n \widehat{X}_n.$$

En fonction de $(\widehat{a}_n, \widehat{b}_n)$, l'erreur quadratique s'écrit encore :

$$\sigma^2 = \sum_{i=1}^n (Y_i - aX_i - b)^2 = \widehat{\sigma}_n^Y - \kappa_n^{XY}/\widehat{\sigma}_n^X = (1 - |\rho_n^{XY}|^2)\widehat{\sigma}_n^Y.$$

Elle est nulle lorsque qu'existe une relation linéaire entre les statistiques X et Y, et faible en cas de relation quasi-linéaire. Notons que la droite de régression passe par le centre de gravité du nuage formé par les n points (X_i, Y_i).

7.2 Statisitique de la régression

On suppose dans cette section que les variables aléatoires $(Y_i)_{i=1\dots n}$ sont reliées à des données déterministes $(X_i)_{i=1\dots n}$ par une relation du type :

$$Y_i = aX_i + b + \varepsilon_i,$$

où les $(\varepsilon_i)_{i=1\dots n}$ sont indépendantes avec $\mathbb{E}[X_i] = 0$, $\mathrm{var}(X_i) = 1$. Les coefficients de régression $(\widehat{a}_n, \widehat{b}_n)$ fournissent alors des estimateurs des quantités a et b.

7.2.1 Propriétés des estimateurs de la régression

Les coefficients de régression $(\widehat{a}_n, \widehat{b}_n)$ sont de "bons" estimateurs des quantités a et b au sens suivant :

Proposition 7.2.1. *Les coefficients de régression $\widehat{a}_n$ et $\widehat{b}_n$ sont des estimateurs sans biais et consistants de a et b.*

Démonstration. Le coefficient $\widehat{a}_n$ s'écrit encore

$$\widehat{a}_n = a + \frac{1}{n\widehat{\sigma}_n^X} \sum_{i=1}^n (X_i - \widehat{X}_n)\varepsilon_i.$$

Comme les ε_i sont centrées, on a bien $\mathbb{E}[\widehat{a}_n] = a$. De la même façon, on vérifie que $\mathbb{E}[\widehat{b}_n] = \mathbb{E}[\widehat{Y}_n] - \mathbb{E}[\widehat{a}_n]\widehat{X}_n = \mathbb{E}[Y] - a\widehat{X}_n = b$, autrement dit, $\widehat{a}_n$ et $\widehat{b}_n$ sont sans biais. Par ailleurs, d'après la loi des grands nombres, lorsque n tend vers l'infini, on a les convergence en probabilité $\widehat{Y}_n \to \mathbb{E}[Y]$, $\widehat{X}_n \to \mathbb{E}[X]$, $\kappa_n^{XY} \to \mathrm{cov}(X, Y)$ et $\widehat{\sigma}_n^X \to \mathrm{var}(X)$. On en déduit les dernières convergences

$$\widehat{a}_n = \frac{\kappa_n^{XY}}{\widehat{\sigma}_n^X} \to \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)} = a, \quad \widehat{b}_n = \widehat{Y}_n - \widehat{a}_n\widehat{X}_n \approx \widehat{Y}_n - a\widehat{X}_n = b + \frac{1}{n}\sum_{i=1}^n \varepsilon_i \to b,$$

i.e. les estimateurs $\widehat{a}_n$ et $\widehat{b}_n$ sont consistants. □

7.2.2 Intervalle de confiance pour la prédiction

Il est naturel d'utiliser la droite $D_{Y/X}$ pour prédire une valeur supplémentaire Y_{n+1} de la statistique Y, connaissant une valeur supplémentaire X_{n+1} de la statistique X : $Y_{n+1} = aX_{n+1} + b$. La qualité d'une telle prédiction (qui a priori n'a vraiment de sens que pour X_{n+1} proche de l'intervalle $[\min X, \max X]$) dépend de la valeur de l'erreur quadratique σ^2 qui est la variance empirique de la statistique $Y - aX - b$. On peut déterminer un intervalle de confiance par sa largeur autour de la droite de régression, de la façon suivante : pour $p \in]0,1[$ fixé, soit r un réel strictement positif tel que

$$p \approx \frac{1}{n} \sum_{i=1}^{n} 1_{\{|Y_i - aX_i - b| \leqslant r\}}.$$

Pensant au théorème central limite, il est naturel d'approcher la loi empirique des $Y_i - aX_i - b$ par une gaussienne, nécessairement centrée et de variance σ^2. De sorte qu'on doit avoir

$$p \approx \mathbb{P}(|\mathcal{N}(0,1)| \leqslant r/\sigma).$$

On peut ainsi estimer que pour X_{n+1} proche de l'intervalle $[\min X, \max X]$, on doit avoir par exemple 95% de chance de trouver Y_{n+1} dans l'intervalle :

$$I_n := [\widehat{a}_n X_{n+1} + \widehat{b}_n - 1.96\sigma, \widehat{a}_n X_{n+1} + \widehat{a}_n + 1.96\sigma]$$

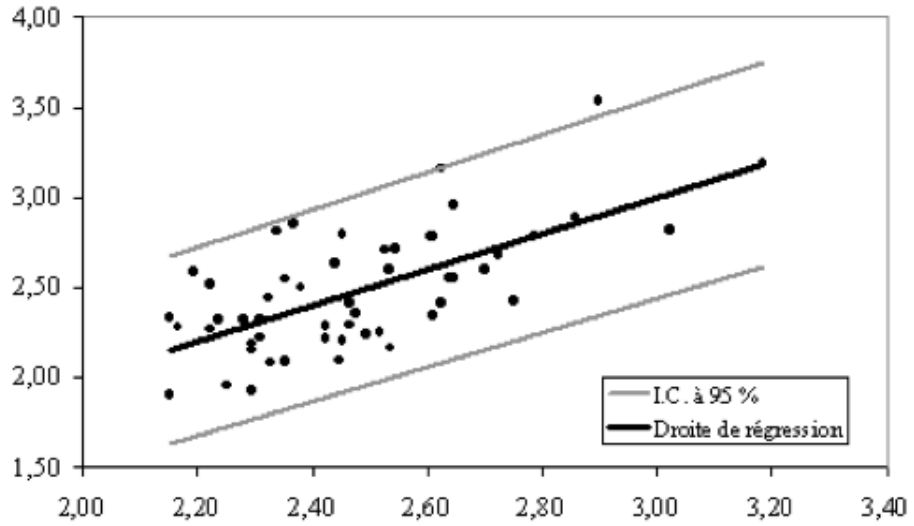


FIGURE 7.2 – Exemple d'intervalle de confiance pour la prédiction.

7.3 Au dela du cas linéaire

Bien entendu, on ne cherche pas toujours une relation linéaire entre les statistiques X et Y. Mais on peut bien souvent s'y ramener, par un changement de variable

élémentaire. Par exemple dans la cas suivant, on peut intuiter que la relation est de la forme $Y = be^{aX}$. On se ramène au cas linéaire en considérant les données $(X_i, \log Y_i)$.

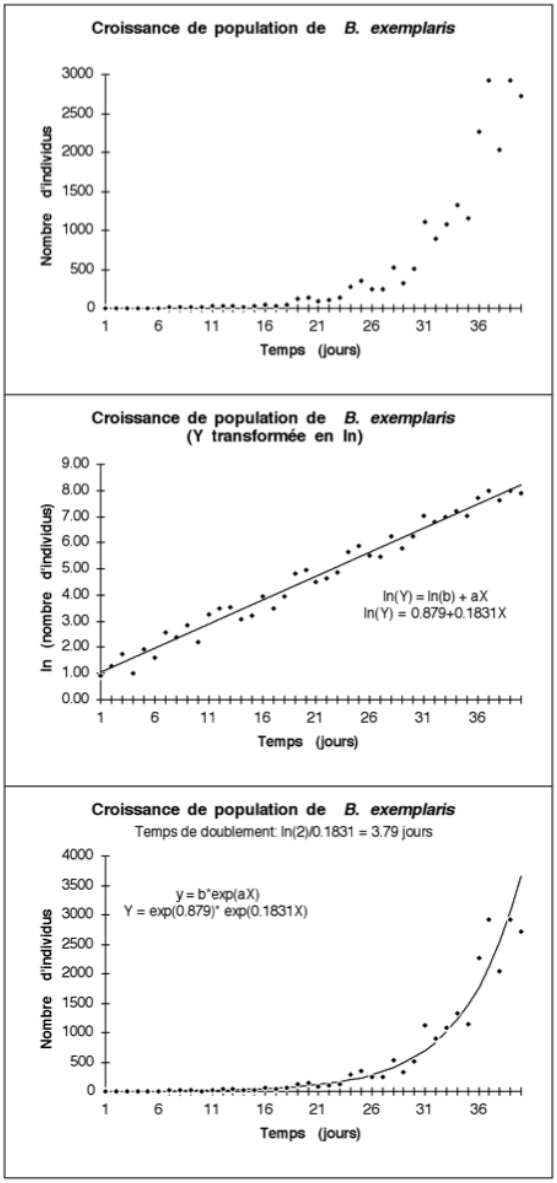


FIGURE 7.3 – Modèle exponentiel $Y = be^{aX}$ et regression linéaire.

De façcon analogue, si l'on espère que $Y = aX^b$, il suffit de considérer le nuage formé par les points $(\log X_j, \log Y_j)$; si l'on espère que $Y = \log(ae^X + b)$, il suffit de considérer le nuage formé par les points (e^{X_j}, e^{Y_j}) ; si l'on soupçonne que $Y = aX/(X^2 + b)$, il suffit de considérer le nuage formé par les points $(X_j^2, X_j/Y_j)$ etc.

On peut aussi envisager des relations plus complexes du types $Y = aX + b\sqrt{X} + c$ qui comme le montre l'exemple ci-dessous peuvent être mieux adaptées à certains type de données.

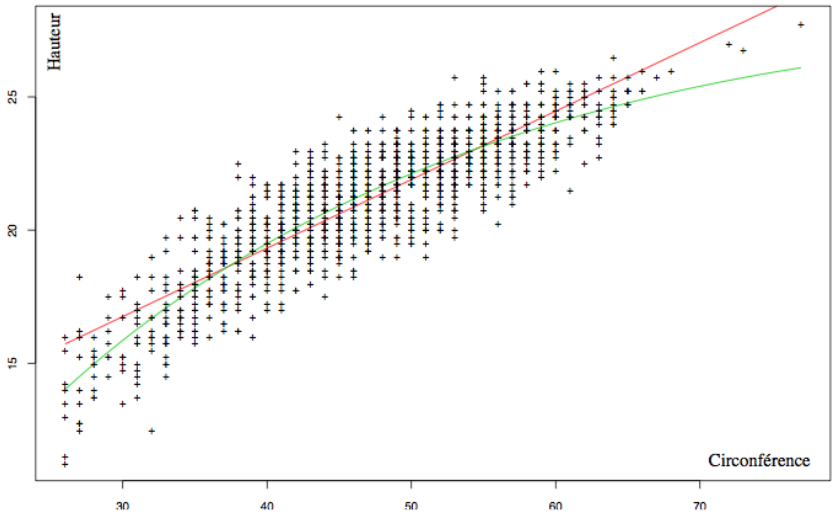


FIGURE 7.4 – En rouge la régression linéaire du type $Y = aX + b$ du nuage de points. En vert, la régression de type $Y = aX + b\sqrt{X} + c$ qui semble mieux adaptée aux données.