

Introduction aux probabilités

Table des matières

Introduction	1
1 Langage des probabilités et premières propriétés	3
1.1 Expériences aléatoires	3
1.1.1 Expérience aléatoire et univers	3
1.1.1.1 Notion d'expérience aléatoire	3
1.1.1.2 Notion d'événement aléatoire	4
1.1.2 Rappels de théorie des ensembles	5
1.1.3 Langage ensembliste VS langage probabiliste	6
1.1.4 Ensembles finis ou dénombrables	7
1.2 Notion d'espace probabilisé	7
1.2.1 Rappel sur les séries	7
1.2.2 Le cas des séries à terme général positif	9
1.2.3 Notion de probabilité	9
1.3 Premiers exemples d'espaces probabilisés	11
1.3.1 La mesure de Dirac	11
1.3.2 Probabilités sur un espace fini	12
1.3.3 Expériences d'urnes et dénombrement	15
1.3.3.1 Rappels de dénombrement	15
1.3.3.2 Nombre de parties à p éléments parmi un ensemble à n éléments	16
1.3.3.3 Expériences d'urnes	17
1.3.3.4 Tirage avec remise	17
1.3.3.5 Tirage sans remise	18

2	Probabilités conditionnelles et indépendance	19
	Introduction	19
2.1	Probabilités conditionnelles	20
2.2	Indépendance d'événements	23
2.2.1	Indépendance de deux événements	23
2.2.2	Indépendance mutuelle	24
2.2.3	Indépendance : loi de Bernoulli et loi binomiale	25
2.3	Double conditionnement, indépendance conditionnelle	26
2.4	Complément : construction d'expériences indépendantes	27
	Introduction	27
2.4.1	Cas de deux expériences	27
2.4.2	Généralisation à n expériences successives	28
3	Espaces de probabilité discrets et variables aléatoires	29
	Introduction	29
3.1	Construction d'un espace de probabilité discret	29
3.2	Variables aléatoires discrètes et leurs lois	30
3.2.1	Définition	30
3.2.2	Loi d'une variable aléatoire	31
3.3	Variables aléatoires discrètes réelles	33
3.3.1	Notion de fonction de répartition	33
3.3.2	Notions d'espérance et moments d'une variable aléatoire réelle	35
3.4	Lois usuelles	40
3.4.1	Cas fini	40
3.4.1.1	Loi de Dirac	40
3.4.1.2	Loi de Bernoulli	40
3.4.1.3	Loi binomiale	40

3.4.1.4	Loi hypergéométrique	42
3.4.2	Cas dénombrable	42
3.4.2.1	Loi géométrique	42
3.4.2.2	Loi de Poisson	44
3.5	Couples de variables aléatoires discrètes, lois conditionnelles, indépendance .	45
3.6	Vecteurs aléatoires, suites de variables aléatoires	49
3.6.1	Définitions et propriétés	49
3.6.2	Matrice de variance/covariance	50
3.6.3	Exemples	50
4	Probabilités sur \mathbb{R} et variables aléatoires à densité	53
	Introduction	53
4.1	Densités de probabilité sur \mathbb{R} et fonctions de répartition	54
4.1.1	Notion de densité de probabilité et exemples	54
4.1.2	Exercice	55
4.1.3	Notion de fonction de répartition	56
4.2	Probabilités à densité sur \mathbb{R}	58
4.3	Variables aléatoires à densité	59
4.4	Notion d'espérance	62
4.5	Étude des lois usuelles	64
4.5.1	Loi uniforme sur $[a, b]$	64
4.5.2	Loi exponentielle	64
4.5.3	Loi de Cauchy	65
4.5.4	Loi gaussienne	65
5	Théorèmes limites et estimation	67
	Introduction	67
5.1	Notions de convergence	67

Introduction	67
5.1.1 Convergence presque sûre	67
5.1.2 Convergence en probabilité	68
5.1.3 Convergence en moyenne quadratique	68
5.1.4 Convergence en loi	68
5.1.5 Liens entre ces convergences	70
5.2 Loi des Grands Nombres	70
5.3 Théorème Central Limite	71
5.4 Application du TCL : notion d'intervalle de confiance	71

Introduction

Le but de ce cours est de présenter les bases de la théorie des probabilités, qui trouve des applications dans de nombreux domaines des sciences (biologie, physique, finance, etc.). Cette théorie est une branche de l'analyse dont l'objet est l'étude mathématique de phénomènes où intervient le hasard. Le probabiliste s'intéresse donc à des expériences dont on connaît l'ensemble des résultats possibles mais dont un résultat particulier n'est pas connu à l'avance.

Le cadre rigoureux de la théorie des probabilités a été développé au XXⁱème siècle, principalement par Andreï Kolmogorov et Paul Lévy. Son axiomatique s'appuie sur la théorie de la mesure développée par Borel et Lebesgue.

Cependant, en dehors de cette axiomatique rigoureuse, les probabilités et les jeux de hasards ont été l'objet de travaux plus anciens, qui remontent au XVIIⁱème siècle (en particulier à Fermat et Pascal, puis à de Moivre, Bernoulli, Gauss, Laplace, entre autres).

L'intuition sous-jacente à la notion de « probabilité » est liée à la vision « fréquentielle » d'un événement : si on voit la probabilité d'un événement comme un nombre qui quantifie la chance qu'il a de se réaliser (ex : lancer d'une pièce équilibrée où $\mathbb{P}(Face) = \mathbb{P}(Pile) = \frac{1}{2}$), il est raisonnable de penser que cette probabilité doit pouvoir être approchée par la fréquence d'apparition de cet événement quand on répète un grand nombre de fois l'expérience de façon indépendante et identique. C'est cette intuition qui a guidé les travaux des mathématiciens cités plus haut et qui peut être (comme on le verra dans le dernier chapitre) rendue rigoureuse.



Langage des probabilités et premières propriétés

1.1 Expériences aléatoires

1.1.1 Expérience aléatoire et univers

1.1.1.1 Notion d'expérience aléatoire

On appelle expérience aléatoire une expérience qui, reproduite à l'identique, peut conduire à plusieurs résultats possibles et dont on ne connaît pas le résultat à l'avance. On note alors Ω l'ensemble (non vide) de tous les résultats possibles. Ω est appelé « univers » (ou encore « espace d'états »), associé à l'expérience. Un résultat possible de l'expérience est classiquement noté ω et appelé « réalisation » ou « éventualité ». Ainsi $\omega \in \Omega$.

Dans toute la suite, il sera important d'identifier correctement l'univers Ω associé à une expérience aléatoire donnée. Des exemples de tels univers sont :

1. Lancer d'une pièce de monnaie (les notations "Pile" et "Face" sont souvent remplacées par "0" et "1")

$$\Omega = \{0, 1\} \text{ et } \text{Card}(\Omega) = 2.$$

2. n lancers successifs d'une pièce de monnaie (ici l'ordre dans lequel on note les résultats a un sens) :

$$\Omega = \{0, 1\}^n = \{(a_1, \dots, a_n), a_i \in \{0, 1\}, i = 1, \dots, n\} \text{ et } \text{Card}(\Omega) = 2^n,$$

3. Un lancer simultané de n pièces de monnaie.

$$\Omega = \{(a_1, \dots, a_n), a_i \in \{0, 1\}, a_i \leq a_{i+1}, i = 1, \dots, n\} \text{ et } \text{Card}(\Omega) = n + 1$$

La différence ici avec le cas précédent est qu'il n'y a plus de notion d'ordre dans la notation des résultats : avoir "Pile" puis "Face" devient la même chose que "Face" puis "Pile" : tout ce qu'on observe, c'est un certain nombre de pièces tombées sur "Pile" (et le nombre complémentaire de pièces sur "Face"). Par conséquent, pour coder un tirage avec k "Pile", il suffit de considérer le vecteur où les k "zéro" sont rangés au début et les $n - k$ "un" à la fin, c'est-à-dire où les coordonnées sont rangées dans l'ordre croissant). Il y a autant de vecteurs qu'il y a de possibilités d'écrire de tels vecteurs avec des zéros, c'est-à-dire $n + 1$.

4. Une succession infinie de lancers d'une pièce de monnaie
 $\Omega = \{0, 1\}^{\mathbb{N}} = \{(a_i)_{i \geq 1}, a_i \in \{0, 1\}, i = 1, \dots, n\}$ et $\text{Card}(\Omega) = +\infty$.
 Ici l'univers est l'ensemble des suites à valeurs dans $\{0, 1\}$.
5. Lancer d'un dé à 6 faces
 $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\text{Card}(\Omega) = 6$.
6. Lancer simultané de deux dés à 6 faces :
 $\Omega = \{(a_1, a_2), a_i \in \{1, 2, 3, 4, 5, 6\}, a_1 \leq a_2\}$ et $\text{Card}(\Omega) = \sum_{i=1}^6 i = 21$.
 Ici la situation est analogue à l'item 3. L'ordre dans lequel on note les résultats ne compte plus : comme les deux lancers sont simultanés, avoir (3, 5) représente le même résultat que (5, 3). On garde donc le premier et on oublie le second.
7. Tirage aléatoire d'une carte dans un jeu de tarot
 $\Omega = \{1, \dots, 78\}$ et $\text{Card}(\Omega) = 78$
8. Génotype d'un enfant dont les parents ont pour gène soit A soit a
 $\Omega = \{(a, a), (a, A), (A, a), (A, A)\}$ et $\text{Card}(\Omega) = 4$
9. Nombre de naissances en France en 2014
 $\Omega = \mathbb{N}$ et $\text{Card}(\Omega) = +\infty$
 Un premier réflexe serait de dire qu'étant donné que chaque année, le nombre de naissance est fini (un nombre fini de mères donnent naissance à un nombre fini d'enfants), l'univers Ω devrait être fini. Le problème est ici que l'on a aucune connaissance a priori sur une borne supérieure sur ce nombre d'enfants d'une année sur l'autre. Donc l'univers naturel est \mathbb{N} .
10. Temps d'attente à une file d'attente (guichet de la poste, emails dans un serveur, etc.)
 $\Omega = [0, +\infty[$ et $\text{Card}(\Omega) = +\infty$
11. Mesure de la température au sommet de la Tour Eiffel le 1er janvier 2014 à 12h :
 $\Omega = \mathbb{R}$ et $\text{Card}(\Omega) = +\infty$
12. Mesure de la température au sommet de la Tour Eiffel le 1er janvier 2014 de 0h à 12h :
 $\Omega = \mathcal{C}([0, 12], \mathbb{R})$ et $\text{Card}(\Omega) = +\infty$
 Attention, ici, nous observons la température de façon continue au cours du temps : une observation est donc une fonction du temps $t \mapsto T(t)$ (Rq : nous n'étudierons pas des univers si compliqués dans ce cours).
13. Point d'impact d'une fléchette dans une cible de rayon R
 $\Omega = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq R^2\}$ et $\text{Card}(\Omega) = +\infty$.

On notera aussi à travers les exemples cités qu'on ne peut pas se contenter du cas où l'univers Ω est fini : il existe des cas intéressants où Ω est l'ensemble des réels, voire même un ensemble de fonctions.

On peut distinguer parmi les exemples cités plus hauts trois cas possibles :

- Le cas où Ω est fini (items 1, 2, 3, 5, 6, 7, 8),
- Le cas où Ω est infini mais dénombrable, c'est-à-dire qu'il est possible d'en énumérer ses éléments, (item 9).
- Le cas où Ω est infini non dénombrable (items 4, 10, 11, 12, 13).

1.1.1.2 Notion d'événement aléatoire

On appelle événement aléatoire un ensemble d'éventualités dont on peut dire au vu de l'expérience s'il est réalisé ou non. Un événement est donc représenté par une partie de Ω , ie un sous-ensemble de Ω .

Des événements tirés des expériences aléatoires précédentes sont par exemple :

- $A = \{\text{pile}\} = \{0\}$ (cf. item 1, dans la convention où "0" symbolise pile, bien sûr).
- $B = \{\text{le résultat du lancer de dé est pair}\} = \{2, 4, 6\}$ (cf. item 5),
- $C = \{\text{le génotype de l'enfant ne contient pas l'allèle } a\} = \{AA\}$ (item 8)
- $D = \{\text{j'ai attendu au moins 1h à la file d'attente}\} =]1, +\infty[$ (cf. item 10)
- On lance deux fois une pièce : l'événement $A = \text{"on obtient pile puis face"}$ est représenté par $\{(0, 1)\}$.

Ainsi un événement aléatoire est associé à un ensemble. Peut-on dire à l'inverse que tout sous-ensemble de Ω représente un événement ?

- Si Ω est fini ou dénombrable, la réponse est oui : tout ensemble représente un événement.
- Si Ω est non dénombrable, la réponse est non : il faut se restreindre à un sous ensemble strict des parties de Ω .

Dans ce cours, nous nous restreindrons à des cas où ce genre de difficulté n'apparaît pas. Ainsi, sauf mention du contraire, toute partie de Ω représentera un événement aléatoire.

1.1.2 Rappels de théorie des ensembles

Intuitivement, un ensemble est une collection d'éléments. Etant donné un élément ω et un ensemble Ω , on écrit $\omega \in \Omega$ si ω appartient à Ω et $\omega \notin \Omega$ sinon.

Il existe un ensemble ne possédant aucun élément : c'est l'ensemble vide, noté \emptyset .

Si A et B sont deux ensembles, on dit que A est inclus dans B (et on note $A \subset B$) si tout élément de A est élément de B . On a alors

$$A = B \Leftrightarrow (A \subset B \text{ et } B \subset A).$$

Si Ω est un ensemble, on note $\mathcal{P}(\Omega)$ l'ensemble de ses parties. Alors $A \subset \Omega$ équivaut à $A \in \mathcal{P}(\Omega)$. Si A et B sont deux sous-ensembles de Ω , on note A^c ou \bar{A} l'ensemble $\{\omega \in \Omega; \omega \notin A\}$, appelé le complémentaire de A . Note aussi $A \setminus B$ l'ensemble des éléments de Ω qui sont dans A et pas dans B . Ainsi $A \setminus B = A \cap B^c$.

Pour A et B deux sous-ensembles de Ω , si $A \cap B = \emptyset$, on dit que A et B sont disjoints.

Si $(A_n)_{n \geq 1}$ est une suite de sous-ensembles de Ω , on définit

$$\begin{aligned} \bigcup_{n \geq 1} A_n &:= \{\omega \in \Omega; \exists n \geq 1, \omega \in A_n\} \\ \bigcap_{n \geq 1} A_n &:= \{\omega \in \Omega; \forall n \geq 1, \omega \in A_n\} \end{aligned}$$

On a les égalités ensemblistes suivantes

$$\begin{aligned} \left(\bigcup_{n \geq 1} A_n\right)^c &= \bigcap_{n \geq 1} A_n^c \\ \left(\bigcap_{n \geq 1} A_n\right)^c &= \bigcup_{n \geq 1} A_n^c \end{aligned}$$

Démontrons ce fait : nous ne démontrons que le premier, le second se traite de la même manière : soit $\omega \in \Omega$. Alors $\omega \in \left(\bigcup_{n \geq 1} A_n\right)^c$ si et seulement si $\omega \notin \bigcup_{n \geq 1} A_n$ si et seulement si pour tout $n \geq 1$, $\omega \notin A_n$ ce qui équivaut à pour tout $n \geq 1$, $\omega \in A_n^c$, ce qui équivaut à $\omega \in \bigcap_{n \geq 1} A_n^c$. Il faut noter que nous avons procédé par équivalence, car il s'agit de montrer l'égalité de deux ensembles. Une autre façon aurait été de procéder par double inclusion.

On dit qu'une suite $(A_n)_{n \geq 1}$ de sous-ensembles de Ω constitue une partition de Ω si les deux conditions suivantes sont vérifiées :

- Les A_n sont deux-à-deux disjoints : pour tout $i \neq j$, $A_i \cap A_j = \emptyset$,
- $\Omega = \bigcup_{n \geq 1} A_n$.

Dans ce cas, il est usuel d'utiliser la notation suivante

$$\Omega = \bigsqcup_{n \geq 1} A_n$$

Définition 1.1 : Fonction indicatrice d'un ensemble

Soit Ω un ensemble et A un sous-ensemble de Ω . On appelle « fonction indicatrice de A », la fonction, notée $\mathbf{1}_A$, telle que pour tout ω dans Ω , $\mathbf{1}_A(\omega)$ vaut 1 si ω appartient à A et vaut 0 sinon.

Proposition 1.2

Soient A et B deux sous-ensembles de Ω . Alors,

$$\mathbf{1}_{A^c} = 1 - \mathbf{1}_A, \mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B, \mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \mathbf{1}_B.$$

Il faut noter que ce sont des égalités de fonctions : par exemple, la première égalité signifie que $\mathbf{1}_{A^c}(\omega) = 1 - \mathbf{1}_A(\omega)$, pour tout $\omega \in \Omega$.

Démonstration

Démontrons la dernière égalité (la plus compliquée), les autres sont laissées en exercice : comme il s'agit de prouver l'égalité de deux fonctions, il s'agit de montrer que pour tout $\omega \in \Omega$, les deux termes de l'égalité sont toujours égaux. Distinguons deux cas :

- Premier cas : $\omega \in A \cup B$. Dans ce cas $\mathbf{1}_{A \cup B}(\omega) = 1$. De plus, dans le cas où ω appartient à A mais pas à B , le second membre $\mathbf{1}_A(\omega) + \mathbf{1}_B(\omega) - \mathbf{1}_A(\omega)\mathbf{1}_B(\omega)$ vaut $1 + 0 - 1 \times 0 = 1$. Si au contraire ω appartient à B mais pas à A , le second membre devient $0 + 1 - 0 \times 1 = 1$. Reste le sous-cas où ω appartient à A et B auquel cas le second membre vaut $1 + 1 - 1 \times 1 = 1$. Dans tous les cas, les deux membres valent 1 et sont donc égaux.
- Deuxième cas : $\omega \notin A \cup B$, c'est-à-dire que ω n'appartient ni à A ni à B . Mais alors, $\mathbf{1}_{A \cup B}(\omega) = 0$ et le second membre vaut $0 + 0 - 0 \times 0 = 0$. Les deux membres valent 0 et sont donc égaux.

D'où l'égalité souhaitée.

1.1.3 Langage ensembliste VS langage probabiliste

Comme nous l'avons vu, un événement aléatoire est un ensemble. Nous établissons ici la traduction en terme ensembliste des opérations naturelles sur les événements :

- NON : la réalisation de l'événement contraire à A est représenté par son complémentaire A^c . Exemple : l'événement contraire de “le résultat d'un lancer de dé est pair” est $\{1, 3, 5\}$,
- ET : l'événement “ A et B sont réalisés” est représenté par l'intersection $A \cap B$ (le résultat de l'expérience se trouve dans A et dans B).
- OU : L'événement “ A ou B sont réalisés” est représenté par l'union $A \cup B$ (le résultat de l'expérience se trouve dans A ou dans B).
- IMPLICATION : le fait que la réalisation de l'événement A implique celle de B est représenté par $A \subset B$. Exemple : “le résultat du lancer de dé est 2” implique que “le résultat du lancer est pair”, i.e. $\{2\} \subset \{2, 4, 6\}$.
- INCOMPATIBLE : dire que A et B sont incompatibles (i.e. un résultat ne peut être à la fois dans A et dans B) se traduit par $A \cap B = \emptyset$.
- TOUJOURS VRAI : l'événement Ω est l'événement certain.
- IMPOSSIBLE : l'ensemble vide \emptyset est l'événement impossible.

1.1.4 Ensembles finis ou dénombrables

On dit qu'un ensemble Ω est fini si son nombre d'éléments est fini. Il est noté $|\Omega|$ ou $\text{Card}(\Omega)$ et est appelé cardinal de Ω .

Une autre notion utile dans le cadre de ce cours est la notion de dénombrabilité :

Définition 1.3

Soit E un ensemble. On dit que E est dénombrable s'il existe une bijection de E sur l'ensemble des entiers naturels \mathbf{N} . Autrement dit, si E est dénombrable, tout élément de E est numéroté et deux éléments différents sont numérotés différemment.

Propriété 1.4

Les ensembles suivants sont dénombrables :

$$\mathbf{N}, \mathbf{N} \times \mathbf{N}, \mathbf{Z}, \mathbf{Q}.$$

Propriété 1.5

Les ensembles suivants **ne sont pas** dénombrables :

$$\{0, 1\}^{\mathbf{N}}, [0, 1], \mathbf{R}.$$

Propriété 1.6

Une union finie ou dénombrable d'ensembles finis ou dénombrables est finie ou dénombrable : si I est fini ou dénombrable et $(A_i)_{i \in I}$ est telle que A_i est fini ou dénombrable pour tout i , alors $\bigcup_{i \in I} A_i$ est fini ou dénombrable.

1.2 Notion d'espace probabilisé

1.2.1 Rappel sur les séries

Nous ferons un usage constant dans la suite de ce cours des résultats sur les séries.

Auparavant, signalons que nous serons amenés très souvent à faire usage des opérations faisant intervenir $+\infty$ et $-\infty$. Nous utiliserons les conventions suivantes : $0 \times \infty = 0$, $a > 0 \Rightarrow a \times \infty = \infty$, $a < 0 \Rightarrow a \times \infty = -\infty$.

Dans tout ce paragraphe, $(u_n)_{n \geq 1}$ désigne une suite de nombres complexes.

Définition 1.7

- Pour tout $n \geq 1$, on désigne par $S_n = u_1 + \dots + u_n = \sum_{k=1}^n u_k$, la somme partielle de la suite (u_n) à l'ordre n . On dit que u_n est le terme général de la série $\sum_n u_n$.
- Dans le cas où la suite des sommes partielles $(S_n)_{n \geq 1}$ converge vers une limite finie S , on dit que la série $\sum_n u_n$ est convergente, et la limite S est notée $\sum_{n=1}^{\infty} u_n$.
- Dans le cas contraire, on dit que la série $\sum_n u_n$ est divergente.
- On dit que la série $\sum_n u_n$ est absolument convergente si la série $\sum_n |u_n|$ est convergente.

Remarque

Une remarque importante est la suivante : il s'agit de bien faire la différence entre les objets suivants :

- la « série » $\sum_n u_n$ (encore notée $\sum u_n$) qui est formellement la suite $(u_n, S_n)_{n \geq 1}$ où u_n est le terme général et S_n la somme partielle (en particulier $\sum_n u_n$ n'est pas nombre complexe, c'est une suite)
- la « somme partielle » $S_n = \sum_{k=1}^n u_k$ qui est un nombre complexe, dépendant de n ,
- la « somme » $\sum_{n=0}^{+\infty} u_n$ qui est la limite des sommes partielles S_n (qui n'a uniquement un sens que si la série $\sum_n u_n$ est convergente).

Propriété 1.8

Si la série de terme général u_n est absolument convergente, alors elle est convergente.

Exemple

- La série de terme général $u_n = r^n$ est convergente si et seulement si $|r| < 1$, et dans ce cas, $\sum_{n=1}^{+\infty} r^n = \frac{r}{1-r}$.
- Séries de Riemann : la série de terme général $u_n = \frac{1}{n^\alpha}$ est convergente si et seulement si $\alpha > 1$.
- La réciproque de la Proposition 1.8 est fausse : la série de terme général $u_n = \frac{(-1)^n}{n}$ est convergente mais pas absolument convergente.

Propriété 1.9 : Linéarité de la somme

Si $\sum_n u_n$ et $\sum_n v_n$ sont deux séries convergentes et $\lambda \in \mathbf{R}$, alors la série $\sum_n (\lambda u_n + v_n)$ est aussi convergente et on l'égalité

$$\sum_{n=1}^{+\infty} (\lambda u_n + v_n) = \lambda \sum_{n=1}^{+\infty} u_n + \sum_{n=1}^{+\infty} v_n.$$

Une propriété importante des séries absolument convergentes est qu'il est possible de sommer les termes en les regroupant par paquets :

Propriété 1.10 : Sommation par paquets

Soit $(I_k)_{k \in K}$ une partition de l'ensemble d'indices \mathbf{N} (où K est fini ou dénombrable). Soit $\sum_n u_n$ une série absolument convergente. Alors :

- pour tout $k \in K$, la série $\sum_{l \in I_k} u_l$ est convergente de somme notée v_k ,
- la série $\sum_{k \in K} v_k$ est absolument convergente,
- On a la relation d'associativité suivante

$$\sum_{n=1}^{+\infty} u_n = \sum_{k \in K} \sum_{l \in I_k} u_l.$$

Il faut voir la relation précédente comme suit : sommer les nombres complexes $(u_n)_{n \geq 1}$ revient à découper l'ensemble d'indices \mathbf{N} en paquets $I_k, k \in K$, à sommer à l'intérieur de chaque paquet I_k pour k fixé, puis ensuite à sommer les sommes intermédiaires correspondant à chaque paquet.

Remarque

Il faut aussi noter que ce résultat n'est plus vrai dans le cas convergent mais pas absolument convergent.

1.2.2 Le cas des séries à terme général positif

Dans ce cas (et dans ce cas seulement), les notions de convergence et d'absolue convergence bien entendu coïncident.

Propriété 1.11

Si $u_n \geq 0$ pour tout n , alors la suite des sommes partielles $(S_n)_{n \geq 1}$ est croissante (puisque $S_{n+1} - S_n = u_{n+1} \geq 0$). Deux cas sont alors possibles :

- Soit la suite $(S_n)_n$ est bornée, auquel cas la série est convergente,
- Soit $(S_n)_n$ n'est pas bornée, auquel cas la suite (S_n) tend vers $+\infty$.

Remarque

Autrement dit, pour une série à termes positifs (et seulement dans ce cas) on peut toujours donner un sens à la somme infinie $\sum_{n=1}^{+\infty} u_n$ dans $[0, +\infty[\cup \{+\infty\}$. Cette remarque va être particulièrement utile lorsque nous manipulerons des sommes de probabilités d'événements (qui sont dans $[0, 1]$ donc en particulier des réels positifs).

Propriété 1.12 : Sommation par paquets

Le résultat de la Proposition 1.10 reste vrai dans le cas de séries à termes positifs, notamment dans le cas où la somme $\sum_{n=1}^{+\infty} u_n$ est infinie.

1.2.3 Notion de probabilité

Nous en venons au dernier objet important de ce chapitre : la notion de probabilité.

Intuitivement, il s'agit d'attribuer à chaque événement A un réel $\mathbb{P}(A)$ compris entre 0 et 1 quantifiant la chance que cet événement soit réalisé, c'est-à-dire que le résultat de l'expérience se trouve dans A .

Toute définition raisonnable d'une probabilité se doit d'obéir à des règles naturelles en lien avec les opérations ensemblistes vues plus haut.

Définition 1.13

On appelle « probabilité » (ou bien « mesure de probabilité » ou encore « loi de probabilité ») sur Ω toute application \mathbb{P} de $\mathcal{P}(\Omega)$ dans $[0, 1]$ qui satisfait les propriétés suivantes :

1. $\mathbb{P}(\Omega) = 1$
2. Pour tout suite d'événements $(A_n)_{n \geq 1}$ **deux-à-deux-disjoints**, (i.e. pour tout $n \neq m$, $A_n \cap A_m = \emptyset$)

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

Remarque

L'égalité précédente généralise à une famille dénombrable l'égalité vue en terminale selon laquelle $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, **si A et B sont des événements disjoints**.

Remarque

Il faut bien sûr remarquer que cette égalité est **grossièrement fausse si les événements considérés ne sont pas deux-à-deux disjoints** : considérons l'expérience d'un jet de dé équilibré et les événements suivants : $A =$ “le résultat est un multiple de 3 ” et $B =$ “le résultat est pair”. Alors $\mathbb{P}(A) + \mathbb{P}(B) = 2/6 + 3/6 = 5/6$ alors que $\mathbb{P}(A \cup B) = \mathbb{P}(\{2, 3, 4, 6\}) = 4/6$.

Définition 1.14

On appelle « espace probabilisé » la donnée du triplet $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ où

- Ω est un univers,
- $\mathcal{P}(\Omega)$ est l'ensemble des parties de Ω (dont on rappelle qu'on suppose dans ce cours qu'il coïncide avec l'ensemble des événements),
- \mathbb{P} est une probabilité sur Ω .

Remarque

Si Ω n'est pas fini ou dénombrable, il faut remplacer $\mathcal{P}(\Omega)$ par un espace plus petit, appelé tribu, ce que nous n'aborderons pas dans ce cours.

Propriété 1.15

Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé. On a les propriétés suivantes :

1. $\mathbb{P}(\emptyset) = 0$.
2. Propriété d'additivité : pour tout A et B **disjoints** (ie $A \cap B = \emptyset$), $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
3. Pour tout événement A , $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
4. Pour tout événements A, B , $A \subset B \Rightarrow \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$,
5. Pour tout événements A, B **quelconques**, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
6. Pour tout suite croissante d'événements $(A_n)_{n \geq 1}$ (ie $A_n \subset A_{n+1}$ pour tout n), la suite de réels $(\mathbb{P}(A_n))_{n \geq 1}$ est croissante, convergente et sa limite pour $n \rightarrow \infty$ est $\mathbb{P}(\cup_{n \geq 1} A_n)$.
7. Pour tout suite décroissante d'événements $(A_n)_{n \geq 1}$ (ie $A_{n+1} \subset A_n$ pour tout n), la suite de réels $(\mathbb{P}(A_n))_{n \geq 1}$ est décroissante, convergente et sa limite pour $n \rightarrow \infty$ est $\mathbb{P}(\cap_{n \geq 1} A_n)$.
8. Pour tout suite **quelconque** d'événements $(A_n)_{n \geq 1}$ (non nécessairement deux-à-deux disjoints), on a $\mathbb{P}(\cup_{n \geq 1} A_n) \leq \sum_{n=1}^{+\infty} \mathbb{P}(A_n)$.

Preuve

1. Soit $a = \mathbb{P}(\emptyset) \geq 0$. Appliquons la définition d'une probabilité à $A_n = \emptyset$ pour tout n , deux-à-deux disjoints. Alors $a = \mathbb{P}(\emptyset) = \sum_{n \geq 1} \mathbb{P}(\emptyset) = \sum_{n \geq 1} a$. En particulier, cela implique que la série de terme général $a \geq 0$ est convergente. Nécessairement $a = 0$.
2. Il suffit d'appliquer la définition à la famille $(A_n)_{n \geq 1}$ définie par $A_1 = A$, $A_2 = B$ et $A_n = \emptyset$ pour tout $n \geq 3$ et utiliser le fait que $\mathbb{P}(\emptyset) = 0$.
3. C'est un cas particulier de l'item précédent avec $B = A^c$: $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$, car A et A^c sont disjoints.
4. Soient $A \subset B$. On peut alors écrire $B = A \cup (B \setminus A)$ avec A et $B \setminus A$ disjoints. Appliquant l'item 2 de cette proposition, on obtient que $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$, car $\mathbb{P}(B \setminus A) \geq 0$.
5. Soient A et B dans $\mathcal{P}(\Omega)$. Par application de l'item 2 précédent, les égalités suivantes sont vraies (faire un dessin) : $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$, $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)$ et $\mathbb{P}(A \cup B) = \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(B \cap A^c)$, car les ensembles considérés sont disjoints. Il vient donc que $\mathbb{P}(A \cup B) = (\mathbb{P}(A) - \mathbb{P}(A \cap B)) + \mathbb{P}(A \cap B) + (\mathbb{P}(B) - \mathbb{P}(B \cap A))$, ce qui donne après simplifications le résultat souhaité.
6. La suite A_n est croissante pour l'inclusion. En particulier, par l'item 4 précédent, la suite $(\mathbb{P}(A_n))_{n \geq 1}$ est croissante et bornée par 1 (ce sont des probabilités) : la limite de $(\mathbb{P}(A_n))_{n \geq 1}$ existe donc. Posons $B_1 = A_1$ et pour tout $n \geq 2$, $B_n = A_n \setminus A_{n-1}$. On a

donc $\mathbb{P}(B_1) = \mathbb{P}(A_1)$ et (cf. item 4.) $\mathbb{P}(B_n) = \mathbb{P}(A_n) - \mathbb{P}(A_{n-1})$ pour $n \geq 2$. En sommant ces égalités, pour n de 1 à $N \geq 1$, on obtient

$\sum_{n=1}^N \mathbb{P}(B_n) = \mathbb{P}(A_1) + (\mathbb{P}(A_2) - \mathbb{P}(A_1)) + \dots + (\mathbb{P}(A_N) - \mathbb{P}(A_{N-1}))$, ce qui constitue une somme télescopique : il reste le dernier terme $\sum_{n=1}^N \mathbb{P}(B_n) = \mathbb{P}(A_N)$.

Par ailleurs, les ensembles B_n sont deux-à-deux disjoints (par construction). En utilisant la définition d'une probabilité, on a donc $\mathbb{P}(\cup_{n \geq 1} B_n) = \sum_{n=1}^{+\infty} \mathbb{P}(B_n)$. On peut donc passer à la limite pour $N \rightarrow \infty$ dans $\sum_{n=1}^N \mathbb{P}(B_n) = \mathbb{P}(A_N)$ et obtenir que $\mathbb{P}(A_N)$ converge vers $\mathbb{P}(\cup_{n \geq 1} B_n)$. Or, par construction $\cup_{n \geq 1} B_n = \cup_{n \geq 1} A_n$, ce qui prouve l'item 6.

7. Nous pourrions refaire une démonstration semblable à l'item précédent mais il y a plus simple. Pour tout $n \geq 1$, considérons $B_n = A_n^c$. Alors la suite $(B_n)_{n \geq 1}$ est croissante pour l'inclusion. Appliquons l'item 6. à la suite $(B_n)_{n \geq 1}$: $\exists \lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \mathbb{P}(\cup_{n \geq 1} B_n)$. L'item 3. appliqué à chaque B_n donne $\mathbb{P}(A_n) = 1 - \mathbb{P}(B_n)$. En particulier, la limite de $\mathbb{P}(A_n)$ existe et vaut $1 - \mathbb{P}(\cup_{n \geq 1} B_n)$. Appliquant de nouveau l'item 3, cette dernière quantité vaut $\mathbb{P}(\cap_{n \geq 1} A_n)$, ce qui donne le résultat.
8. Soit $(A_n)_{n \geq 1}$ une suite quelconque d'événements. Par une récurrence immédiate à partir de l'item 5, il vient que, pour tout $n \geq 1$, $\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) \leq \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n)$. Le second terme de l'inégalité tend vers $\sum_{j=1}^{+\infty} \mathbb{P}(A_j)$ quand $n \rightarrow \infty$ (limite qui a toujours un sens car, encore une fois, c'est la somme partielle d'une série de terme général positif). D'autre part, si on pose $B_n = A_1 \cup A_2 \cup \dots \cup A_n$, c'est une suite d'événements croissants pour l'inclusion. Donc $\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n)$ converge (par l'item 6) vers $\mathbb{P}(\cup_{n \geq 1} B_n) = \mathbb{P}(\cup_{n \geq 1} A_n)$.

Remarque

Notons ici que pour l'item 8., on ne préjuge en rien de la valeur de la somme $\sum_{n=1}^{+\infty} \mathbb{P}(A_n)$: elle peut très bien valoir $+\infty$ (se rappeler la remarque 1.2.2) !

1.3 Premiers exemples d'espaces probabilisés

1.3.1 La mesure de Dirac

Un premier exemple de mesure de probabilité sur Ω est donné par

Propriété 1.16 : Mesure de Dirac

Fixons $\omega_0 \in \Omega$. L'application δ_{ω_0} de $\mathcal{P}(\Omega)$ dans $[0, 1]$ qui à A associe 1 si $\omega_0 \in A$ et 0 si $\omega_0 \notin A$ est une mesure de probabilité sur Ω . On l'appelle mesure de Dirac en ω_0 .

Preuve

Vérifions les deux axiomes de la définition d'une probabilité. On a clairement $\delta_{\omega_0}(\Omega) = 1$. Soit $(A_n)_{n \geq 1}$ une suite d'événements deux-à-deux disjoints. Deux cas sont possibles :

- Soit ω_0 n'appartient pas à $\cup_{n \geq 1} A_n$ auquel cas il n'appartient à aucun des A_n et la définition d'une probabilité est vérifiée ($0 = 0$).
- soit ω_0 appartient à $\cup_{n \geq 1} A_n$ auquel cas il appartient à un unique A_n (car les A_n sont deux-à-deux disjoints) et la définition d'une probabilité est vérifiée ($1 = 1$).

Cet exemple de mesure de probabilité est un peu particulier : cela correspond à une expérience aléatoire qui donne toujours le même résultat (ex : une pièce truquée qui donne toujours pile).

1.3.2 Probabilités sur un espace fini

Dans tout ce paragraphe, on suppose que l'univers Ω est fini de cardinal $|\Omega|$. On peut donc écrire $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ avec $n = |\Omega|$. L'intuition est que dans le cas d'un univers fini, pour définir une probabilité, il faut et il suffit de se donner n réels positifs p_1, \dots, p_n dont la somme vaut 1. Chaque p_i représente la probabilité d'obtenir le singleton $\{\omega_i\}$.

La probabilité d'un événement A se calcule alors comme la somme des probabilités des singletons qui constituent A .

Commençons par considérer un exemple. On considère le lancer de dé $\Omega = \{1, 2, 3, 4, 5, 6\}$. Dans ce cas, se donner une probabilité sur Ω , c'est se donner (p_1, \dots, p_6) , 6 réels positifs dont la somme vaut 1. Ici, p_i représente la probabilité de tomber sur la face numéro i (pour $i \in \{1, \dots, 6\}$). Le cas souvent considéré est le cas d'un dé équilibré où $p_i = \frac{1}{6}$ pour tout $i = 1, \dots, 6$ (mais ce n'est qu'un cas particulier).

Considérons l'événement $A = \{\text{le résultat du lancer est un nombre premier}\}$. Naturellement, $A = \{2, 3, 5\}$ (attention 1 n'est pas premier). La probabilité de A se calcule alors comme

$$\mathbb{P}(A) = p_2 + p_3 + p_5$$

Dans le cas où le dé est équilibré (c'est-à-dire $p_i = 1/6$ pour tout i), cette probabilité est bien sûr égale à $3/6 = 1/2$.

Il y a ici un jeu d'écriture qu'il s'agit de bien comprendre pour la suite. Dans cet exemple, l'égalité $\mathbb{P}(A) = p_2 + p_3 + p_5$ peut se réécrire sous la forme compacte

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

ce qui signifie que l'on somme les p_i en considérant uniquement les probabilités relatives aux éléments contenus dans A (c'est-à-dire $i = 2, 3$ et 5). Une autre manière utile de réécrire cette égalité est la suivante

$$\mathbb{P}(A) = 0 \times p_1 + 1 \times p_2 + 1 \times p_3 + 0 \times p_4 + 1 \times p_5 + 0 \times p_6.$$

Autrement dit nous avons fait apparaître cette fois-ci **tous les éléments** de $\Omega = \{1, 2, 3, 4, 5, 6\}$ mais en multipliant par 0 les probabilités dont nous n'avons pas besoin pour le calcul de $\mathbb{P}(A)$. Ce qu'il faut remarquer dans l'écriture précédente, c'est que le "0" ou le "1" devant chaque p correspond précisément à l'indicatrice de A appliqué à chaque élément $i \in \Omega = \{1, \dots, 6\}$: en effet, $\mathbf{1}_A(1) = 0$, $\mathbf{1}_A(2) = 1$, $\mathbf{1}_A(3) = 1$, $\mathbf{1}_A(4) = 0$, $\mathbf{1}_A(5) = 1$, $\mathbf{1}_A(6) = 0$.

Cette observation nous conduit donc à écrire

$$\mathbb{P}(A) = \mathbf{1}_A(1)p_1 + \mathbf{1}_A(2)p_2 + \mathbf{1}_A(3)p_3 + \mathbf{1}_A(4)p_4 + \mathbf{1}_A(5)p_5 + \mathbf{1}_A(6)p_6.$$

ce qu'on écrit de façon compacte en

$$\mathbb{P}(A) = \sum_{i=1}^6 \mathbf{1}_A(i)p_i = \sum_{i \in \Omega} \mathbf{1}_A(i)p_i.$$

Cette fois-ci, la somme porte sur l'intégralité des éléments de Ω , mais nous avons simplement multiplié par 0 (via l'indicatrice) les probabilités des singletons dont nous n'avons pas besoin pour le calcul de $\mathbb{P}(A)$. Cette écriture va nous être utile dans la suite.

La proposition suivante dit essentiellement qu'il n'y a qu'une seule façon de définir une probabilité sur un univers fini, et c'est la façon décrite en début de paragraphe.

Propriété 1.17

1) Soit \mathbb{P} une mesure de probabilité sur Ω fini. Notons $p_\omega = \mathbb{P}(\{\omega\})$ pour tout $\omega \in \Omega$. Alors pour tout $A \in \mathcal{P}(\Omega)$, on a l'égalité

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega = \sum_{\omega \in \Omega} p_\omega \mathbf{1}_A(\omega).$$

Autrement dit, \mathbb{P} est entièrement déterminée par la donnée des $(p_\omega)_{\omega \in \Omega}$.

2) Réciproquement, si $(p_\omega)_{\omega \in \Omega}$ est une famille indexée par Ω de réels positifs dont la somme vaut 1, alors il existe une unique probabilité \mathbb{P} sur Ω telle que $\mathbb{P}(\{\omega\}) = p_\omega$.

Preuve

1. Soit $A \in \mathcal{P}(\Omega)$. Alors A est l'union finie de ses singletons (car Ω est fini). Par additivité, on a donc $\mathbb{P}(A) = \mathbb{P}(\cup_{\omega \in A} \{\omega\}) = \sum_{\omega \in A} p_\omega = \sum_{\omega \in \Omega} p_\omega \mathbf{1}_A(\omega)$. La dernière égalité vient de ce que nous venons de dire dans l'introduction de ce paragraphe : l'indicatrice de A est simplement là pour éliminer les probabilités des singletons dont nous n'avons pas besoin pour le calcul de $\mathbb{P}(A)$, puisque si $\omega \notin A$, $\mathbf{1}_A(\omega) = 0$.
2. Réciproquement, si on se donne $(p_\omega)_{\omega \in \Omega}$ comme définis dans l'énoncé, définissons l'application \mathbb{P} de $\mathcal{P}(\Omega)$ dans $[0, 1]$ par $\mathbb{P}(A) = \sum_{\omega \in A} p_\omega$. Prouvons que \mathbb{P} ainsi construite est bien une probabilité. Tout d'abord $0 \leq \mathbb{P}(A) \leq \sum_{\omega \in \Omega} p_\omega = 1$, car les p_ω sont des réels positifs de somme 1. Donc \mathbb{P} est bien à valeurs dans $[0, 1]$. De plus $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p_\omega = 1$ par construction des p_ω .

Considérons maintenant une suite $(A_n)_{n \geq 1}$ d'événements deux-à-deux disjoints. Alors $\mathbb{P}(\cup_n A_n) = \sum_{\omega \in \cup_n A_n} p_\omega = \sum_{\omega \in \Omega} p_\omega \mathbf{1}_{\cup_n A_n}(\omega)$. Mais comme les A_n sont deux à deux disjoints, il est clair que $\mathbf{1}_{\cup_n A_n}(\omega) = \sum_{n \geq 1} \mathbf{1}_{A_n}(\omega)$ (dans cette série, seul un terme vaut 1, les autres 0). Par conséquent, $\mathbb{P}(\cup_n A_n) = \sum_{\omega \in \Omega} p_\omega \sum_{n \geq 1} \mathbf{1}_{A_n}(\omega) = \sum_{n \geq 1} \sum_{\omega \in \Omega} p_\omega \mathbf{1}_{A_n}(\omega) = \sum_{n \geq 1} \mathbb{P}(A_n)$.

Définition 1.18 : Loi de Bernoulli

Si $\Omega = \{0, 1\}$ (de cardinal 2) et si $p \in [0, 1]$, on appelle loi de Bernoulli, la mesure de probabilité sur Ω caractérisée par $\mathbb{P}(\{1\}) = p_1 = p$ et $\mathbb{P}(\{0\}) = p_0 = 1 - p$.

Remarque

La loi de Bernoulli correspond exactement à l'expérience de lancer d'une pièce de monnaie dont la probabilité d'avoir pile (= un succès) est p et d'avoir face (= un échec) est $1 - p$. Le cas d'une pièce non truquée correspond à $p = 1/2$.

On le voit donc ici, à un univers donné, il existe une multiplicité de probabilités possibles. Une parmi toutes les autres correspond à attribuer la même probabilité à chaque singleton :

Propriété 1.19 : Probabilité uniforme

Si $|\Omega|$ est fini, on appelle probabilité uniforme sur Ω la probabilité \mathbb{P} telle que $\mathbb{P}(\{\omega\})$ est indépendant du choix de $\omega \in \Omega$. Dans ce cas, on a nécessairement, pour tout $\omega \in \Omega$

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|},$$

et pour tout $A \in \mathcal{P}(\Omega)$

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

Preuve

C'est une conséquence immédiate de la proposition 1.17.

La probabilité uniforme est l'interprétation usuelle des expressions souvent rencontrées dans les exercices de probabilité de lycée comme “tirage au hasard”, “tirage sans distinction”, etc. Le lecteur doit se convaincre que ce genre d'expressions, beaucoup trop vagues, portent à confusion : il existe de multiples probabilités sur un même univers. Par exemple, tirer une pièce au hasard ne signifie rien en soi, sachant qu'il y a autant de façons différentes de le faire qu'il y a de choix du paramètre p dans la loi de Bernoulli.

Lorsqu'on voudra préciser que la probabilité sous-jacente est uniforme, on privilégiera les expressions du type “uniformément au hasard”.

Remarque

L'expression $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ dans le cas de la mesure uniforme justifie l'usage des techniques de dénombrement : la probabilité d'un événement est alors le rapport du nombre de cas favorables sur le nombre total de cas possibles.

Exemple

- Le lancer d'une pièce équilibrée correspond à la loi uniforme sur $\{P, F\}$: $\mathbb{P}(\{P\}) = \mathbb{P}(\{F\}) = \frac{1}{2}$,
- Le lancer d'un dé équilibré correspond à la probabilité uniforme sur $\{1, \dots, 6\}$: $\mathbb{P}(\{i\}) = \frac{1}{6}$.

Exemple

Evidemment, l'expression $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ n'a plus de sens dans le cas d'une probabilité qui n'est pas uniforme : si on considère un dé pipé tel que $\mathbb{P}(\{1\}) = 1/3$, $\mathbb{P}(\{2\}) = 1/6$ et $\mathbb{P}(\{i\}) = 1/8$ pour tout $3 \leq i \leq 6$ alors la probabilité d'obtenir un nombre pair est $\mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 1/6 + 1/8 + 1/8 = 5/12$ qui n'est pas égal à $3/6 = 1/2$ qu'on aurait obtenu dans le cas uniforme.

Exemple

Considérons l'expérience consistant à lancer deux dés équilibrés. Il est possible de modéliser cette expérience de deux manières :

- Soit on distingue l'ordre des résultats (lancers successifs) : l'univers est alors $\Omega = \{(i, j), 1 \leq i, j \leq 6\}$ de cardinal 36. Dans cette expérience, on distingue le lancer (1, 5) du lancer (5, 1).
- Dans ce cas, la probabilité est uniforme : la probabilité de l'événement $\{(i, j)\}$ vaut $1/36$ pour tout (i, j) .
- Soit on ne distingue pas l'ordre des lancers (lancers simultanés) (autrement dit, on identifie les lancers (1, 5) et (5, 1)). Dans ce cas, l'univers est alors $\Omega = \{(i, j), 1 \leq i \leq j \leq 6\}$ de cardinal $\sum_{i=1}^6 \sum_{j=i}^6 1 = \sum_{i=1}^6 (6 - i + 1) = 6(6 + 1)/2 = 21$. La probabilité sur Ω **n'est pas la probabilité uniforme** : pour tout i , $\mathbb{P}(\{(i, i)\}) = 1/36$ alors que si $i < j$, $\mathbb{P}(\{(i, j)\}) = 1/18$!

Calculons dans les deux cas la probabilité que la somme des deux chiffres valent 7.

Réponse : dans le premier cas, cet événement correspond à $\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$ de cardinal 6. La probabilité de cet événement est donc $6/36 = 1/6$ (car la probabilité est uniforme).

Dans le deuxième cas, cet événement correspond à $\{(1, 6), (2, 5), (3, 4)\}$. La probabilité de chacun des éléments de cet ensemble étant $1/18$, la probabilité de cet événement est donc $3/18 = 1/6$. Nous retrouvons le calcul précédent, et c'est bien normal : la valeur de la somme de deux nombres ne dépend pas de l'ordre de ces nombres.

La conclusion de cet exemple est qu'il faut faire très attention à la définition de l'univers sous-jacent à une expérience et que la probabilité naturelle d'une expérience n'est pas forcément la probabilité uniforme.

1.3.3 Expériences d'urnes et dénombrement

1.3.3.1 Rappels de dénombrement

Soient A_1, \dots, A_n une collection finie d'ensembles finis, supposés deux-à-deux disjoints. Alors

$$|A_1 \cup A_2 \cup \dots \cup A_n| = |A_1| + |A_2| + \dots + |A_n|.$$

Si $\Omega_1, \dots, \Omega_n$ est une collection finie d'ensembles finis, alors le produit cartésien $\Omega_1 \times \Omega_2 \times \dots \times \Omega_n$ est lui-même fini de cardinal $|\Omega_1| \cdot |\Omega_2| \dots |\Omega_n|$.

Propriété 1.20

Si A et B sont deux ensembles finis alors l'ensemble $\mathcal{F}(A, B)$ des applications de A dans B est lui aussi fini de cardinal $|B|^{|A|}$.

Preuve

Posons $n = |A|$ et $m = |B|$. Il y a autant de façons de choisir une fonction de A vers B qu'il y a de façons de choisir les images des n éléments de A par cette fonction. Pour chaque élément de A il y a m choix possibles d'une image, ce qui donne donc m^n choix possibles en tout.

Corollaire 1.21

Si A est fini, $\mathcal{P}(A)$, l'ensemble des parties de A , est lui aussi fini de cardinal $2^{|A|}$.

Preuve

L'application de $\mathcal{P}(A)$ dans $\mathcal{F}(A, \{0, 1\})$ (l'ensemble des fonctions de A dans l'ensemble $\{0, 1\}$) qui à chaque élément B de $\mathcal{P}(A)$ associe sa fonction indicatrice $\mathbf{1}_B$ est une bijection. En effet, cette application est surjective : si $f \in \mathcal{F}(A, \{0, 1\})$, posons $B = f^{-1}(\{1\})$, c'est-à-dire l'ensemble des éléments ω de A tels que $f(\omega) = 1$. Montrons que $f = \mathbf{1}_B$. En effet, pour tout $\omega \in B$, $f(\omega) = 1$, par définition de B . Si par contre $\omega \notin B$, alors $f(\omega) \neq 1$ (car sinon $\omega \in B$) et donc $f(\omega) = 0$. Donc $f = \mathbf{1}_B$.

Montrons maintenant que cette application est injective. Soient B et B' deux ensembles qui ont la même fonction indicatrice. Mais alors pour tout $\omega \in A$, $\omega \in B \Leftrightarrow \mathbf{1}_B(\omega) = 1 \Leftrightarrow \mathbf{1}_{B'}(\omega) = 1 \Leftrightarrow \omega \in B'$. Donc $B = B'$.

En particulier, le cardinal de $\mathcal{P}(A)$ est égal à celui de $\mathcal{F}(A, \{0, 1\})$ qui est $2^{|A|}$, d'après la propriété précédente.

Propriété 1.22

Si A et B sont deux ensembles finis de cardinaux respectifs $|A| = p$ et $|B| = n$, l'ensemble des injections de A dans B est lui aussi fini, de cardinal 0 si $p > n$ et de cardinal $\frac{n!}{(n-p)!}$ sinon.

Cas particulier où $p = n$: l'ensemble des bijections (on dit aussi dans ce cas "permutations") d'un ensemble à n éléments est de cardinal $n!$.

Preuve

Rappelons qu'une application de A dans B est injective si deux éléments distincts de A ont des images distinctes dans B . Il y a donc autant d'injections de A dans B qu'il y a de p -uplets formés d'éléments de B deux-à-deux distincts. Il y en a $n(n-1)(n-2) \dots (n-p+1) = \frac{n!}{(n-p)!}$.

1.3.3.2 Nombre de parties à p éléments parmi un ensemble à n éléments

Définition 1.23

Le nombre de sous-ensembles à p éléments pris parmi un ensemble à n éléments est noté $\binom{n}{p}$ et appelé «combinaison de p parmi n » ou encore «coefficient binomial de paramètre n et p ».

Remarque

On rappelle ici qu'un ensemble est par définition **non ordonné** : $\{2, 3\} = \{3, 2\}$.

Propriété 1.24

On a l'égalité suivante : pour tout $n, p \geq 0$,

$$\binom{n}{p} = \begin{cases} 0 & \text{if } p > n \\ \frac{n!}{p!(n-p)!} & \text{sinon.} \end{cases}$$

Preuve

Le cas $p > n$ étant évident, nous supposons donc que $p \leq n$. Soit E un ensemble à n éléments. Choisir une partie à p éléments parmi E c'est d'abord choisir les éléments qui le compose. On choisit le premier élément (il y a n choix possibles) puis le deuxième (il y a $n - 1$ choix possibles), etc. jusqu'au p ème (il y a $n - p + 1$ choix possibles). Un tel choix correspond à la donnée de p éléments **ordonnés** (a_1, \dots, a_p) . Il y a donc $n(n-1) \dots (n-p+1) = \frac{n!}{(n-p)!}$ choix possibles de tels vecteurs de taille p . Mais permuter les composantes du vecteur (a_1, \dots, a_p) va donner le même ensemble $\{a_1, \dots, a_p\}$ (ex : le choix de 2 en premier puis de 3 en deuxième conduit au même ensemble que le choix de 3 en premier et de 2 en deuxième, c'est-à-dire $\{2, 3\}$). Il faut donc diviser la quantité précédente par le nombre de permutations d'un ensemble à p éléments, ie $p!$, ce qui donne $\frac{n!}{p!(n-p)!}$.

Nous rappelons les propriétés essentielles des coefficients binomiaux :

Propriété 1.25

Soient $p \geq 1$, $n \geq 1$. On a

1. $\binom{n}{p} = \binom{n}{n-p}$,
2. Formule du triangle de Pascal : $\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}$,
3. Formule du binôme de Newton : pour tous réels a et b

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

Preuve

1. C'est évident d'après la formule. On peut aussi le voir en constatant que prendre une sous-partie de cardinal p revient au même que de prendre son complémentaire, de cardinal $n - p$.
2. On peut le vérifier par le calcul (exercice), mais on peut aussi raisonner de la manière suivante. L'ensemble des sous-parties à p éléments de l'ensemble $\{1, \dots, n\}$ se décompose en deux parties disjointes :
 - les ensembles qui contiennent l'élément 1. Pour déterminer un tel ensemble, il faut et il suffit alors de choisir les $p - 1$ éléments restants parmi les $n - 1$ éléments à notre disposition. Il y a donc $\binom{n-1}{p-1}$ tels ensembles.

- les ensembles qui ne contiennent pas l'élément 1. Pour déterminer un tel ensemble, il faut et il suffit de choisir les p éléments qui le composent parmi les $n - 1$ éléments restants. Il y a donc $\binom{n-1}{p}$ tels sous-ensembles.
- 3. On raisonne par récurrence sur $n \geq 0$. Le cas $n = 0$ est évident : $1 = 1$. Supposons la propriété vérifiée pour un certain $n \geq 0$ et prouvons-la au rang $n + 1$. On a par hypothèse de récurrence, $(a + b)^{n+1} = (a + b) \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = \sum_{k=0}^n \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=0}^n \binom{n}{k} a^k b^{n+1-k} = a^{n+1} + b^{n+1} + \sum_{k=0}^{n-1} \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=1}^n \binom{n}{k} a^k b^{n+1-k}$. Par un changement d'indice dans la première somme, il vient $(a + b)^{n+1} = a^{n+1} + b^{n+1} + \sum_{k=1}^n \binom{n}{k-1} a^k b^{n+1-k} + \sum_{k=1}^n \binom{n}{k} a^k b^{n+1-k} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n+1-k}$, par la formule du triangle de Pascal. D'où la propriété par récurrence.

1.3.3.3 Expériences d'urnes

Considérons le dispositif suivant : dans une urne se trouvent N boules, réparties entre N_1 boules rouges et $N - N_1$ boules noires. On suppose que les boules sont numérotées. On effectue un tirage de n boules dans cette urne et on s'intéresse à la probabilité que l'échantillon obtenu contienne k boules rouges (et donc $n - k$ boules noires).

Nous allons voir que l'on peut donner plusieurs sens à cette expérience, selon le type de tirage que l'on considère, chacun d'eux correspondant à un univers différent et menant à des calculs de probabilités différents. Ceci est en particulier l'illustration de l'importance de définir correctement l'univers associé à une expérience aléatoire donnée.

1.3.3.4 Tirage avec remise

On effectue le tirage avec remise (on s'autorise à tirer une boule plusieurs fois de suite). L'univers dans ce cas est l'ensemble des vecteurs de taille n dont les composantes sont parmi $\{1, \dots, N\}$, de cardinal N^n .

Le cardinal de l'événement $A_k = \text{"on tire } k \text{ boules rouges"}$ se calcule de la façon suivante : la première des étapes est de choisir l'ordre d'apparition des boules dans le tirage (indépendamment de leur numéro). Il y a autant de façons de répartir k boules rouges parmi n qu'il y a de choix de parties à k éléments parmi n c'est à dire $\frac{n!}{k!(n-k)!}$. Etant donnée cette répartition de couleur, il y a N_1 numéros possibles pour chaque boule rouge et $N - N_1$ numéros possibles pour chaque boule noire. Le cardinal de A_k est donc $\frac{n!}{k!(n-k)!} N_1^k (N - N_1)^{n-k}$. La probabilité de A_k est donc donnée par

$$\mathbb{P}(A_k) = \binom{n}{k} \left(\frac{N_1}{N}\right)^k \left(1 - \frac{N_1}{N}\right)^{n-k} = \binom{n}{k} p^k (1 - p)^{n-k},$$

où $p = \frac{N_1}{N}$.

Définition 1.26

On appelle loi binomiale de paramètres $n \geq 1$ et $p \in [0, 1]$ la probabilité sur l'ensemble $\{0, \dots, n\}$ donnée par, pour tout $0 \leq k \leq n$,

$$p_k = \binom{n}{k} p^k (1 - p)^{n-k}$$

La loi binomiale joue un rôle très important en probabilités, nous aurons l'occasion d'y revenir.

1.3.3.5 Tirage sans remise

On effectue ici le tirage de n boules sans remise (on ne peut pas tirer la même boule deux fois). L'univers Ω est alors l'ensemble des vecteurs de taille n dont les composantes sont dans $\{1, \dots, N\}$ et sont toutes distinctes (on ne peut pas tirer le même numéro deux fois). Son cardinal est $N(N-1) \dots (N-n+1) = \frac{N!}{(N-n)!}$.

Le cardinal de l'événement $A_k = \text{"tirage de } k \text{ boules rouges"}$ se calcule de la façon suivante : de même que pour le tirage avec remise, il y a autant de façons de répartir k boules rouges parmi n qu'il y a de choix de parties à k éléments parmi n c'est à dire $\frac{n!}{k!(n-k)!}$. Etant donnée cette répartition de couleur, il y a N_1 numéros possibles pour la première boule rouge, puis $N_1 - 1$ pour la deuxième etc. jusqu'à $N_1 - k + 1$ pour la k ème. De même, il y a $(N - N_1)(N - N_1 - 1) \dots (N - N_1 - (n - k) + 1)$ répartitions possibles de numéros pour les boules noires. Le cardinal de A est donc $\frac{n!}{k!(n-k)!} N_1(N_1 - 1) \dots (N_1 - k + 1)(N - N_1)(N - N_1 - 1) \dots (N - N_1 - (n - k) + 1) = \frac{n!}{k!(n-k)!} \frac{N_1!}{(N_1-k)!} \frac{(N-N_1)!}{(N-N_1-(n-k))!}$. La probabilité d'obtenir k boules rouges est donc

$$\mathbb{P}(A_k) = \frac{\binom{N_1}{k} \binom{N-N_1}{n-k}}{\binom{N}{n}}.$$

Définition 1.27

On appelle loi hypergéométrique de paramètres N , N_1 et n la probabilité sur $\{0, \dots, n\}$ donnée par, pour tout $0 \leq k \leq n$,

$$p_k = \frac{\binom{N_1}{k} \binom{N-N_1}{n-k}}{\binom{N}{n}}$$

Autre interprétation : une autre façon de voir P_k est de dire qu'obtenir sans remise k boules rouges, c'est tirer une famille de k boules rouges parmi les N_1 à disposition (il y en a $\binom{N_1}{k}$), puis tirer $n - k$ boules noires parmi les $N - N_1$ à disposition (il y en a $\binom{N-N_1}{n-k}$), parmi toutes les possibilités qui sont autant de parties à n éléments parmi N , au nombre de $\binom{N}{n}$.

Exemple

Pièces défectueuses dans une usine : si on sait que dans une chaîne de montage avec N pièces parmi lesquels M sont défectueuses, la probabilité d'obtenir k pièces défectueuses parmi un échantillon de taille n est $\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$.



Probabilités conditionnelles et indépendance

Introduction

Dans tout ce chapitre, $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est un espace probabilisé.

L'idée de conditionnement est simple : calculer une probabilité d'un événement A conditionnellement à un autre événement B consiste à calculer la probabilité de A en supposant que l'événement B est réalisé. Autrement dit, il s'agit de calculer une probabilité dans un cas où une partie de l'aléa de l'expérience a été (artificiellement) enlevé.

Considérons un exemple : soit l'expérience aléatoire consistant à tirer uniformément au hasard une carte parmi un paquet de 52 cartes. Soit l'événement

A = "la carte tirée est un roi".

L'univers associé à cette expérience est bien sûr $\Omega = \{1, \dots, 52\}$, muni de sa probabilité uniforme. Dans ce cadre, la probabilité de A est

$$\mathbb{P}(A) = \frac{4}{52} = \frac{1}{13}.$$

Supposons maintenant qu'un observateur extérieur regarde la carte tirée (sans dévoiler celle-ci) et donne l'information suivante

B = "la carte tirée est une figure".

Dans ce cas, cette information supplémentaire réduit drastiquement le champ des possibilités. Puisque nous savons maintenant que la carte est une figure, l'intuition nous dit que la probabilité de tirer un roi **sachant que (ou conditionnellement au fait que)** la carte est une figure est $\frac{4}{12} = \frac{1}{3}$.

Remarquons que l'exemple précédent semble indiquer que la probabilité de A sachant B est plus grande que la probabilité de A . Cette intuition est bien sûr fautive ! Par exemple, si maintenant on conditionne par l'événement C = "la carte tirée est un as", la probabilité de A sachant C sera nulle (donc plus petite que la probabilité de A).

2.1 Probabilités conditionnelles

Définition 2.1 : Probabilité conditionnelle

Soit B un événement tel que $\mathbb{P}(B) > 0$. Si A est un événement, on appelle « probabilité de A sachant B » (ou encore « probabilité de A conditionnellement à B ») la quantité suivante

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

Propriété 2.2

L'application $A \mapsto \mathbb{P}(A|B)$ ainsi définie est une probabilité sur Ω . On l'appelle « probabilité conditionnelle sachant B ». On la note souvent $\mathbb{P}(\cdot|B)$ ou bien \mathbb{P}_B .

Preuve

Vérifions les deux axiomes d'une probabilité :

1. $\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.
2. Soit $(A_n)_n$ une suite d'événements deux-à-deux disjoints. Alors $\mathbb{P}_B(\cup_{n \geq 1} A_n) = \frac{\mathbb{P}(B \cap (\cup_{n \geq 1} A_n))}{\mathbb{P}(B)} = \frac{\mathbb{P}(\cup_{n \geq 1} (B \cap A_n))}{\mathbb{P}(B)} = \sum_{n \geq 1} \frac{\mathbb{P}(B \cap A_n)}{\mathbb{P}(B)} = \sum_{n \geq 1} \mathbb{P}_B(A_n)$, car les $B \cap A_n$ sont deux-à-deux disjoints et \mathbb{P} est une probabilité.

Exemple

Revenons à notre exemple introductif : si A = “on tire un roi” et B = “on tire une figure” alors $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{4/52}{(12/52)} = \frac{1}{3}$: on retrouve bien le résultat donné par l'intuition initiale.

Exemple

On lance successivement deux pièces équilibrées et identiques. L'univers est

$$\Omega = \{(P, F), (P, P), (F, P), (F, F)\}$$

muni de la probabilité uniforme.

1. Quelle est la probabilité d'obtenir deux fois Face sachant que le premier jet donne Face ?

Il s'agit de calculer $\mathbb{P}(\{(F, F)\} | \{(F, F), (F, P)\}) = \frac{\mathbb{P}(\{(F, F)\})}{\mathbb{P}(\{(F, F), (F, P)\})} = \frac{(1/4)}{(1/2)} = \frac{1}{2}$.

2. Quelle est la probabilité d'obtenir deux fois Face, sachant que l'un des deux jets au moins donne Face ?

Ici on calcule $\mathbb{P}(\{(F, F)\} | \{(F, F), (F, P), (P, F)\}) = \frac{\mathbb{P}(\{(F, F)\})}{\mathbb{P}(\{(F, F), (F, P), (P, F)\})} = \frac{(1/4)}{(3/4)} = \frac{1}{3}$.

3. Mon voisin a deux enfants dont une fille, quel est la probabilité que l'autre soit un garçon ? C'est la reformulation de la question précédente.

La définition précédente permet de calculer la probabilité d'une intersection de deux événements : $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. La proposition suivante généralise ce fait à un nombre arbitraire d'événements :

Propriété 2.3 : Formule des probabilités composées

Si $n \geq 1$ et A_1, \dots, A_n des événements tels que $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$, alors

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Remarque

Une bonne façon de comprendre cette formule est de considérer (même si ce n'est pas forcément le cas) les événements A_1, \dots, A_n comme des **événements successifs** : on a l'événement A_1 , puis l'événement A_2 etc. Ainsi la formule précédente dit que la probabilité de l'intersection $\mathbb{P}(A_1 \cap \dots \cap A_n)$ est la probabilité du dernier événement sachant tout ce qui s'est passé avant $\mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1})$ multipliée par la probabilité de l'avant-dernier événement sachant tout ce qui s'est passé avant $\mathbb{P}(A_{n-1} | A_1 \cap \dots \cap A_{n-2})$, etc. jusqu'à la probabilité du deuxième sachant le premier $\mathbb{P}(A_2 | A_1)$ multipliée par la probabilité du premier $\mathbb{P}(A_1)$.

Attention

Bien sûr, il est possible de permuter les événements A_1, \dots, A_n dans la formule précédente. Il y a donc autant de variantes de cette formule que de permutations correspondantes. Par exemple, dans le cas où $n = 3$, on peut écrire

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

mais aussi (par exemple)

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

ou bien

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

Suivant les exercices, telle ou telle variante sera plus adaptée qu'une autre. L'idée est encore de respecter la chronologie de l'énoncé (cf exemple suivant).

Exemple

On tire sans remise trois cartes uniformément parmi un paquet de 52 cartes. Quelle est la probabilité de tirer trois piques ? Posons A l'événement "on tire un pique au premier tirage", B "on tire un pique au second tirage" et C "on tire un pique au troisième tirage". On cherche à calculer $\mathbb{P}(A \cap B \cap C)$. On applique la formule des probabilités **en respectant la chronologie de l'énoncé** (A se produit avant B qui se produit avant C) :

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|A \cap B) = \frac{13}{52} \frac{12}{51} \frac{11}{50}.$$

Toute autre variante de la formule des probabilités composées aurait conduit à des calculs plus compliqués, car ne respectant pas la chronologie imposée par l'énoncé.

Preuve : Preuve de la formule des probabilités composées.

Raisonnons par récurrence : le cas $n = 1$ est évident. Les cas $n = 2$ et $n = 3$ ne sont pas nécessaires pour la preuve mais traitons-les pour bien comprendre le mécanisme sous-jacent : pour $n = 2$ c'est la définition de la probabilité conditionnelle et pour $n = 3$,

$$\begin{aligned} \mathbb{P}(A_1 \cap A_2 \cap A_3) &= \mathbb{P}(A_3 | A_1 \cap A_2) \mathbb{P}(A_1 \cap A_2) \\ &= \mathbb{P}(A_3 | A_1 \cap A_2) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_1), \end{aligned}$$

par applications successives de la définition d'une probabilité conditionnelle. Supposons maintenant le résultat vrai pour $n \geq 1$. Prouvons le résultat pour $n + 1$:

$$\begin{aligned}
\mathbb{P}(A_1 \cap \dots \cap A_{n+1}) &= \mathbb{P}(A_{n+1} | A_1 \cap \dots \cap A_n) \mathbb{P}(A_1 \cap \dots \cap A_n), \text{ (déf. de la prob. conditionnelle)} \\
&= \mathbb{P}(A_{n+1} | A_1 \cap \dots \cap A_n) \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}) \dots \\
&\quad \mathbb{P}(A_3 | A_2 \cap A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_1),
\end{aligned}$$

par hypothèse de récurrence. D'où la propriété est vraie au rang $n + 1$ donc vraie pour tout n par récurrence.

Définition 2.4 : Système complet d'événements

On dit que les événements $E_n \subset \Omega$, $n \geq 1$ forment un système complet d'événements s'ils sont deux-à-deux disjoints et si $\mathbb{P}(\cup_{n \geq 1} E_n) = 1$.

Remarque

Cette définition généralise la notion de partition : ici, on ne suppose pas nécessairement que $\cup_{n \geq 1} E_n = \Omega$.

Propriété 2.5 : Formule des probabilités totales

Si $(E_n)_{n \geq 1}$ est un système complet d'événements, pour tout événement A ,

$$\mathbb{P}(A) = \sum_{n=1}^{+\infty} \mathbb{P}(A | E_n) \mathbb{P}(E_n).$$

Preuve

Comme $(E_n)_{n \geq 1}$ est un système complet d'événements,

$$\mathbb{P}(A) = \mathbb{P}(A \cap \cup_{n \geq 1} E_n) + \mathbb{P}(A \cap (\cup_{n \geq 1} E_n)^c).$$

Or, $A \cap (\cup_{n \geq 1} E_n)^c \subset (\cup_{n \geq 1} E_n)^c$ donc $\mathbb{P}(A \cap (\cup_{n \geq 1} E_n)^c) = 0$ et l'union $A \cap \cup_{n \geq 1} E_n = \cup_{n \geq 1} (A \cap E_n)$ est une union deux-à-deux disjointe. Par conséquent, $\mathbb{P}(A) = \sum_{n=1}^{+\infty} \mathbb{P}(A \cap E_n) = \sum_{n=1}^{+\infty} \mathbb{P}(A | E_n) \mathbb{P}(E_n)$.

Exemple

On dispose de deux urnes, l'une contenant 1 boule blanche et 3 noires, l'autre 2 blanches et 2 noires. On choisit uniformément au hasard parmi les deux urnes, puis on tire uniformément au hasard une boule dans l'urne choisie. On considère la probabilité de l'événement $A =$ "on tire une boule noire". L'univers Ω est $\{(i, j); i = 1, 2, j = 1, 2, 3, 4\}$, où i est le numéro de l'urne et j le numéro de la boule dans l'urne. On considère la partition $\Omega = E_1 \sqcup E_2$, où $E_i = \{(i, j), j = 1, 2, 3, 4\}$ qui est l'événement consistant à choisir l'urne i . Si on numérote en premier les boules noires, l'événement A s'écrit $\{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2)\}$. Pour tout i , la probabilité conditionnelle \mathbb{P}_{E_i} est la probabilité uniforme sur E_i . Donc $\mathbb{P}(A | E_1) = 3/4$ et $\mathbb{P}(A | E_2) = 1/2$. D'après la formule des probabilités totales, $\mathbb{P}(A) = \mathbb{P}(A | E_1) \mathbb{P}(E_1) + \mathbb{P}(A | E_2) \mathbb{P}(E_2) = \frac{3}{4} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{5}{8}$.

Le fait d'avoir à l'esprit une chronologie quand on manipule des probabilités conditionnelles fait que dans le calcul de $\mathbb{P}(A | B)$, A est vu comme la "conséquence" et B est vu comme la "cause". Cependant, on souhaite parfois calculer $\mathbb{P}(B | A)$, c'est-à-dire la probabilité de la "conséquence" A conditionnellement à la "cause" B . C'est l'objet de la proposition suivante :

Propriété 2.6 : Théorème de Bayes ou de "probabilité des causes"

Pour tout événement A tel que $\mathbb{P}(A) > 0$,

Si B est tel que $\mathbb{P}(B) > 0$,

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

De plus, soit $(E_n)_{n \geq 1}$ un système complet d'événements de Ω . On a, pour tout $n \geq 1$,

$$\mathbb{P}(E_n|A) = \frac{\mathbb{P}(A|E_n)\mathbb{P}(E_n)}{\sum_{n \geq 1} \mathbb{P}(A|E_n)\mathbb{P}(E_n)}.$$

Preuve

Par définition,

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

La deuxième formule vient de la précédente avec $B = E_n$ et de la formule des probabilités totales pour le dénominateur.

Exemple

Chaque don de sang est soumis à un test du VIH. On suppose que la probabilité qu'un individu soit atteint du virus est $r = 10^{-4}$. On suppose que la probabilité que le test soit positif pour une personne malade (=efficacité du test) est $p = 0.99$ et que la probabilité que le test soit positif pour une personne saine (=probabilité de fausse alarme) est $q = 0.05$. Quelle est la probabilité qu'une personne ayant un test positif soit malade ?

Réponse : un univers possible est Ω est $\{(I, T), I \in \{M, S\}, T \in \{P, N\}\}$ (I code l'état de l'individu (M : malade, S : sain) et T code le résultat du test (N : négatif, P : positif)). On cherche $\mathbb{P}(M|P)$. La formule de Bayes donne

$$\mathbb{P}(M|P) = \frac{\mathbb{P}(P|M)\mathbb{P}(M)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P|M)\mathbb{P}(M)}{\mathbb{P}(P|M)\mathbb{P}(M) + \mathbb{P}(P|S)\mathbb{P}(S)} = \frac{pr}{pr + q(1-r)} \approx 0.0019.$$

On peut être surpris de la petitesse du résultat trouvé. Cela est dû au fait qu'un test médical n'est pas infaillible et qu'une grande proportion de la population est saine, ce qui fait qu'on détecte souvent par erreur des individus sains : dépister une maladie coûte cher !

2.2 Indépendance d'événements

2.2.1 Indépendance de deux événements

Revenons à notre exemple introductif de tirage uniforme d'une carte. Supposons maintenant que l'information donnée par l'observateur extérieur est

$C = \text{«la carte tirée est un trèfle.»}$

Calculons $\mathbb{P}(A|C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \frac{(1/52)}{(13/52)} = \frac{1}{13}$. Cette dernière quantité est exactement égale à $\mathbb{P}(A)$. Autrement dit, le fait de conditionner par C n'a rien changé à la valeur de la probabilité initialement calculée. Intuitivement, cela signifie que conditionner par C n'a apporté aucune information quant à la détermination de A (ce n'est pas parce que je connais la couleur d'une carte que j'ai plus d'informations sur sa valeur). On dira alors que A et C sont « indépendants ».

Définition 2.7

Soient A et B deux événements. On dit que A et B sont « indépendants » si

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

Propriété 2.8 : Indépendance et probabilité conditionnelle

Soient A et B deux événements tels que $\mathbb{P}(B) > 0$. Alors A et B sont indépendants si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Preuve

A et B sont indépendants si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ ssi $\mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ car $\mathbb{P}(B) > 0$. Ce qui prouve le résultat.

On confond parfois la notion d'indépendance d'événements avec la notion de disjonction d'événements : ce sont deux notions très différentes !

Propriété 2.9

Soient A et B deux événements disjoints. Alors A et B sont indépendants si et seulement si $\mathbb{P}(A) = 0$ ou $\mathbb{P}(B) = 0$.

Preuve

A et B sont indépendants ssi $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Or $A \cap B = \emptyset$. Donc A et B sont indépendants ssi $\mathbb{P}(A)\mathbb{P}(B) = 0$, ce qui prouve le résultat.

Remarque

Autrement dit, deux événements disjoints de probabilité strictement positive **ne sont jamais indépendants**. C'est en fait très intuitif : le fait d'imposer que les deux événements soient disjoints implique justement que ces deux événements sont au contraire très dépendants (on pensera à l'analogie d'un métro bondé à l'heure de pointe : pour sortir de la rame, vous dépendez fortement des autres...).

Propriété 2.10

Si A et B sont indépendants, il en est de même pour A et B^c , A^c et B , ou A^c et B^c .

Preuve

Prouvons le dans le premier cas, les autres se prouvent de la même manière (exercice) : $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c)$.

2.2.2 Indépendance mutuelle

La notion d'indépendance définie dans le paragraphe précédent concerne l'indépendance de deux événements. On peut définir l'indépendance d'une famille quelconque d'événements mais la définition est plus subtile :

Définition 2.11

Soit une famille (finie ou infinie) d'événements $(A_i)_{i \in I}$.

- On dit que les événements A_i sont **deux-à-deux indépendants** si

$$\forall i \neq j, A_i \text{ et } A_j \text{ sont indépendants, i.e. } \mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

- On dit que les événements A_i sont **mutuellement indépendants** si pour toute partie finie $J \subset I$, on a $\mathbb{P}(\bigcap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i)$.

Remarque

En particulier, une famille finie d'événements (A_1, \dots, A_n) est indépendante si par définition pour tous indices (i_1, \dots, i_L) ($1 \leq L \leq n$) pris dans $\{1, \dots, n\}$ tels que $i_1 < i_2 < \dots < i_L$,

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_L}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_L}).$$

Exemple

Trois événements (A, B, C) sont mutuellement indépendants si

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B), \\ \mathbb{P}(A \cap C) &= \mathbb{P}(A)\mathbb{P}(C), \\ \mathbb{P}(B \cap C) &= \mathbb{P}(B)\mathbb{P}(C), \\ \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \end{aligned}$$

Remarque

La bonne définition de l'indépendance de plus de deux événements est la notion d'indépendance mutuelle. La notion d'indépendance deux-à-deux est plus faible et possède de mauvaises propriétés.

Propriété 2.12

L'indépendance mutuelle implique l'indépendance deux-à-deux mais la réciproque est fausse.

Preuve

Pour l'implication, il suffit de se restreindre aux parties $J \subset I$ qui ne contiennent que deux éléments.

Donnons un contre-exemple pour la réciproque : $\Omega = \{1, 2, 3, 4\}$ muni de sa probabilité uniforme. Alors $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{2, 3\}$ sont deux-à-deux indépendants : en effet $\mathbb{P}(A \cap B) = \mathbb{P}(\{1\}) = 1/4 = \mathbb{P}(A)\mathbb{P}(B)$ (idem pour les autres). Par contre, A , B et C ne sont pas mutuellement indépendants : $\mathbb{P}(A \cap B \cap C) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

2.2.3 Indépendance : loi de Bernoulli et loi binomiale

Le cadre de ce paragraphe est le suivant : on répète successivement et indépendamment n essais de Bernoulli de probabilité de succès égal à p . L'univers est donné par $\Omega = \{S, E\}^n$ (S : succès, E : échec) et comme les tirages sont indépendants, la probabilité d'un événement élémentaire $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ est donnée par

$$\mathbb{P}(\{\omega\}) = \prod_{i=1}^n p_{\omega_i}$$

où chacun de ces p_{ω_i} vaut soit p si $\omega_i = S$ ou soit $1 - p$ si $\omega_i = E$. Plus précisément, si $k(\omega)$ est le nombre de fois où un succès S apparaît dans l'épreuve ω on a

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

Exemple

Soit l'événement “le premier tirage est un succès, et tous les autres sont des échecs”. Il est représenté par (S, E, E, \dots, E) . Sa probabilité est $p(1-p)^{n-1}$.

Propriété 2.13

Pour tout $0 \leq k \leq n$, soit l'événement A_k défini par

A_k = “on a obtenu exactement k succès S au cours de l'expérience”. Alors, la probabilité de cet événement est donné par

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

Remarque

On retrouve la loi binomiale définie en fin de chapitre 1 : si μ est la loi binomiale de paramètres n et p , $\mu(\{k\})$ décrit la probabilité d'obtenir exactement k succès parmi n expériences de Bernoulli successives et indépendantes.

Preuve

A_k est constitué de l'ensemble des épreuves ω pour lesquelles $k(\omega) = k$. Toutes ces épreuves ont la même probabilité qui vaut $p^k(1-p)^{n-k}$. Il y a autant de telles épreuves qu'il y a de possibilités de choisir les succès parmi les n tirages, c'est-à-dire $\binom{n}{k}$.

2.3 Double conditionnement, indépendance conditionnelle

L'idée de ce paragraphe repose sur la Proposition 2.2 précédente : si C est un événement de probabilité positive, la probabilité conditionnelle sachant C , \mathbb{P}_C est elle-même une probabilité et donc obéit aux mêmes règles. Autrement dit, nous pouvons réécrire tout le début de ce cours avec \mathbb{P}_C à la place de \mathbb{P} .

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

- Probabilité conditionnelle d'une union : soient A, B, C trois événements, avec $\mathbb{P}(C) > 0$. Alors

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

- Double conditionnement : soient A, B, C trois événements, avec $\mathbb{P}(B \cap C) > 0$. Alors

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

- Formule des probabilités totales conditionnelle : pour tout événement C tel que $\mathbb{P}(C) > 0$, pour tout système complet d'événements $(E_n)_n$ et pour tout événement A

$$\mathbb{P}(A|C) = \sum_{n \geq 1} \mathbb{P}(A|E_n \cap C) \mathbb{P}(E_n|C)$$

- Indépendance conditionnelle : soient A, B, C trois événements, avec $\mathbb{P}(C) > 0$. On dit que A et B sont indépendants conditionnellement à C si

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

2.4 Complément : construction d'expériences indépendantes

Introduction

Ce paragraphe ne figure pas au programme de l'examen.

2.4.1 Cas de deux expériences

Supposons que nous réalisons deux expériences de façon successive, chacune représentée par (Ω_1, \mathbb{P}_1) et (Ω_2, \mathbb{P}_2) . Le nouvel univers est alors

$$\Omega = \Omega_1 \times \Omega_2.$$

La question est alors la suivante : de quelle manière doit-on définir une probabilité sur Ω de telle sorte que les deux expériences soient indépendantes (au sens où pour tous événements $A_1 \in \mathcal{P}(\Omega_1)$ et $A_2 \in \mathcal{P}(\Omega_2)$, A_1 et A_2 soient indépendants) ?

L'événement A qui consiste à observer A_1 puis A_2 s'écrit

$$A = A_1 \times A_2 = \{(\omega_1, \omega_2); \omega_1 \in A_1, \omega_2 \in A_2\}.$$

La seule façon de définir $\mathbb{P}(A)$ de sorte que A_1 et A_2 soient indépendants est donc

$$\mathbb{P}(A) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2).$$

En particulier la probabilité de tout singleton $\omega = (\omega_1, \omega_2) \in \Omega$ est donnée par

$$\mathbb{P}(\{\omega\}) = \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}).$$

Reste à prouver que ce choix est le bon.

Propriété 2.14

Définissons sur la probabilité \mathbb{P} sur Ω comme ci-dessus. Alors si A_1 est un événement lié à l'expérience (Ω_1, \mathbb{P}_1) et A_2 est un événement lié à (Ω_2, \mathbb{P}_2) , alors A_1 et A_2 sont indépendants dans l'expérience (Ω, \mathbb{P}) .

Preuve

Nous avons représenté A_1 comme un sous-ensemble de Ω_1 . Dans l'expérience Ω , cette représentation n'est pas correcte : le même événement sera représenté par

$$\tilde{A}_1 = A_1 \times \Omega_2,$$

(obtenir l'événement A_1 dans l'expérience Ω , c'est d'abord obtenir A_1 dans l'expérience Ω_1 , puis obtenir n'importe quoi dans l'expérience Ω_2). De même $\tilde{A}_2 = \Omega_1 \times A_2$ représente l'événement A_2 dans l'expérience globale. Mais alors, par définition de \mathbb{P} ,

$$\mathbb{P}(\tilde{A}_1) = \mathbb{P}_1(A_1) \times 1; \quad \mathbb{P}(\tilde{A}_2) = 1 \times \mathbb{P}_2(A_2).$$

Or $\tilde{A}_1 \cap \tilde{A}_2 = A_1 \times A_2$. Mais alors, par définition de \mathbb{P} , $\mathbb{P}(\tilde{A}_1 \cap \tilde{A}_2) = \mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$, ce qui est égal d'après les calculs précédents à $\mathbb{P}(\tilde{A}_1)\mathbb{P}(\tilde{A}_2)$. Conclusion :

$$\mathbb{P}(\tilde{A}_1 \cap \tilde{A}_2) = \mathbb{P}(\tilde{A}_1)\mathbb{P}(\tilde{A}_2),$$

ce qui donne le résultat.

2.4.2 Généralisation à n expériences successives

Soient $(\Omega_i, \mathbb{P}_i)_{i=1, \dots, n}$, n expériences. L'expérience globale consistant en la réalisation successive et indépendante de ces n expériences a pour univers

$$\Omega = \Omega_1 \times \dots \times \Omega_n.$$

Si pour $i = 1, \dots, n$, A_i est un événement associé à l'expérience (Ω_i, \mathbb{P}_i) , elle est représentée dans l'expérience globale par

$$\tilde{A}_i = \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n.$$

L'événement qui consiste en la réalisation de A_1 puis de A_2 , etc. puis de A_n est représenté par

$$A = A_1 \times A_2 \times \dots \times A_n = \tilde{A}_1 \cap \dots \cap \tilde{A}_n.$$

Si on définit la probabilité \mathbb{P} sur Ω en posant pour tout singleton

$$\mathbb{P}(\{\omega_1, \dots, \omega_n\}) = \mathbb{P}_1(\{\omega_1\}) \dots \mathbb{P}_n(\{\omega_n\})$$

alors les événements (\tilde{A}_i) sont mutuellement indépendants.



Espaces de probabilité discrets et variables aléatoires

Introduction

Dans tout ce chapitre, nous nous donnons un univers Ω dont on suppose qu'il est dénombrable. On le notera $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$. Dans les applications, la plupart du temps $\Omega = \mathbb{N}$, $\Omega = \mathbb{Z}$ ou encore $\Omega = \mathbb{Q}$.

3.1 Construction d'un espace de probabilité discret

Dans le chapitre précédent, nous avons vu que pour définir une probabilité sur un univers **fini**, il faut et il suffit de définir la probabilité de chacun des singletons de l'univers. Autrement dit, se donner une probabilité revient à se donner une suite finie de nombres positifs dont la somme vaut 1.

Dans le cas où Ω est dénombrable, la situation est très analogue : pour définir une probabilité sur Ω , il faut et il suffit de se donner une suite de réels positifs $(p_i)_{i \geq 1}$, chacun d'entre eux représentant la probabilité d'obtenir le singleton $\{\omega_i\}$. Ici, la condition que la somme vaut 1 est remplacée par le fait d'exiger que la série des p_i est convergente, de somme égale à 1.

Propriété 3.1

Si Ω est dénombrable,

1. Si \mathbb{P} est une probabilité sur Ω et si pour tout $\omega_i \in \Omega$ ($i \geq 1$), on note $p_i = \mathbb{P}(\{\omega_i\})$, alors les propriétés suivantes sont vraies :
 - pour tout $i \geq 1$, $p_i \in [0, 1]$,
 - la série de terme général p_i est convergente, de somme 1,
 - pour toute partie A de Ω , $\mathbb{P}(A) = \sum_{i \geq 1, \omega_i \in A} p_i$.
2. Réciproquement, si $(p_i)_{i \geq 1}$ est une suite de termes positifs, telle que la série de terme général p_i est convergente de somme 1, alors l'application \mathbb{P} définie par $\mathbb{P}(A) = \sum_{i \geq 1, \omega_i \in A} p_i$ (pour toute partie A de Ω) est une probabilité sur Ω .

Preuve

Cette démonstration est très analogue au cas fini, vu au Chapitre 1.

1. Le fait que les p_i sont dans $[0, 1]$ et que la série de terme général p_i est convergente de somme 1 vient du fait que \mathbb{P} est une probabilité et que Ω est l'union dénombrable disjointe de ses singletons. D'où $\mathbb{P}(\Omega) = 1 = \sum_{i=1}^{+\infty} p_i$.

Soit $A \in \mathcal{P}(\Omega)$. Alors A est l'union dénombrable et disjointe de ses singletons (car Ω est dénombrable). On a donc $\mathbb{P}(A) = \mathbb{P}(\cup_{\omega \in A} \{\omega\}) = \sum_i \mathbb{P}(\{\omega_i\} \cap A) = \sum_i \mathbb{P}(\{\omega_i\}) \mathbf{1}_A(\omega_i) = \sum_{i \geq 1} p_i \mathbf{1}_A(\omega_i)$.

2. Réciproquement, si on se donne $(p_i)_{i \geq 1}$ comme définis dans l'énoncé, définissons l'application \mathbb{P} de $\mathcal{P}(\Omega)$ dans $[0, 1]$ par $\mathbb{P}(A) = \sum_{i \geq 1, \omega_i \in A} p_i$. Prouvons que \mathbb{P} ainsi construite est bien une probabilité. Tout d'abord $0 \leq \mathbb{P}(A) \leq \sum_{i=1}^{+\infty} p_i = 1$, car les p_i sont des réels positifs dont la série converge de somme 1. Donc \mathbb{P} est bien à valeurs dans $[0, 1]$. De plus $\mathbb{P}(\Omega) = \sum_{i=1}^{+\infty} p_i = 1$ par construction des p_i .

Considérons maintenant une suite $(A_n)_{n \geq 1}$ d'événements deux-à-deux disjoints. Alors $\mathbb{P}(\cup_n A_n) = \sum_{i \geq 1} \mathbb{P}(\{\omega_i\} \cap \cup_n A_n) = \sum_{i \geq 1} \mathbb{P}(\{\omega_i\}) \mathbf{1}_{\cup_n A_n}(\omega_i)$. Mais comme les A_n sont deux à deux disjoints, il est clair que $\mathbf{1}_{\cup_n A_n}(\omega_i) = \sum_{n \geq 1} \mathbf{1}_{A_n}(\omega_i)$ (dans cette série, seul un terme vaut 1, les autres 0). Par conséquent, $\mathbb{P}(\cup_n A_n) = \sum_{i=1}^{+\infty} p_i \left(\sum_{n \geq 1} \mathbf{1}_{A_n}(\omega_i) \right) = \sum_{n \geq 1} \sum_{i=1}^{+\infty} p_i \mathbf{1}_{A_n}(\omega_i) = \sum_{n \geq 1} \sum_i \mathbb{P}(\{\omega_i\} \cap A_n) = \sum_{n \geq 1} \mathbb{P}(A_n)$.

Nous verrons dans la suite des exemples de telles probabilités.

Remarque

Rien n'empêche que pour certains $i \geq 1$, p_i soit nul. En particulier, le cas d'une probabilité sur un espace fini est un cas particulier du cas dénombrable (il suffit de considérer des p_i nuls à partir d'un certain rang).

Définition 3.2

Si \mathbb{P} est une loi discrète, on appelle **support** de \mathbb{P} l'ensemble des ω_i tels que $p_i = \mathbb{P}(\{\omega_i\}) \neq 0$. Le cas d'une loi de probabilité sur un ensemble fini correspond aux lois dont le support est fini.

Remarque

Il est important de noter qu'il **n'existe pas de probabilité uniforme dont le support est infini dénombrable** : la condition que la série $\sum p_i$ est convergente impose que la suite des p_i converge vers 0. A partir du moment où un nombre infini de p_i sont strictement positifs, la suite des p_i ne peut donc certainement pas être constante !

3.2 Variables aléatoires discrètes et leurs lois

3.2.1 Définition

Lors d'une expérience aléatoire, on est souvent moins intéressé par le résultat d'une expérience (par ex : le résultat d'un lancer de dé) que par les conséquences de ce tirage (par ex : le fait de gagner 3 euros si le résultat du dé est 6, et de perdre 2 euro sinon). C'est l'objet de la notion de variable aléatoire :

Définition 3.3

Une variable aléatoire X sur Ω est une application

$$X : \omega \in \Omega \mapsto X(\omega) \in E,$$

où E est un ensemble quelconque. Le plus souvent E sera $\mathbf{N}, \mathbf{Z}, \mathbf{Q}, \mathbf{R}, \mathbf{R}^d$, etc.

Exemple

Dans l'exemple introductif, la variable aléatoire en question est :

$$\begin{aligned} X : \Omega = \{1, \dots, 6\} &\longrightarrow \{-2, 3\} \\ 6 &\longmapsto 3 \\ \omega &\longmapsto -2 \text{ si } \omega \neq 6 \end{aligned}$$

Etant donnée une variable aléatoire X , nous pouvons nous intéresser maintenant aux événements du type

$$A = \{\omega \in \Omega, X(\omega) \in F\},$$

où F est un sous-ensemble de E . Cet événement, qui est donc (j'insiste) un sous-ensemble de Ω , est l'ensemble des éventualités ω dont l'image par X se trouve dans F . Autrement dit A est « l'image réciproque de F par X » :

$$A = X^{-1}(F).$$

Par commodité, on omet souvent la dépendance en ω et on note simplement

$$A = \{X \in F\}.$$

Attention

Des énormités sont souvent relevées dans les copies à ce propos : par exemple l'expression (malheureusement souvent rencontrée dans les copies) $\mathbb{P}(X)$ n'a absolument aucun sens ! X est une variable aléatoire, pas un événement ! On ne prend la probabilité que d'un événement. Par contre, l'expression $\mathbb{P}(X \in A)$ est tout-à-fait légitime : c'est la probabilité de l'événement $\{\omega \in \Omega, X(\omega) \in A\}$.

Je vous laisse donc imaginer l'humeur et la perplexité du correcteur quand il voit sur la copie des expressions du type $\mathbb{P}(X \cap Y) = \mathbb{P}(X)\mathbb{P}(Y)$... (que veut donc dire l'intersection de variables aléatoires !).

Exemple

Dans notre exemple introductif, l'événement consistant à perdre 2 euro est $X^{-1}(\{-2\}) = \{X = -2\} = \{\omega \in \Omega, X(\omega) = -2\} = \{1, 2, 3, 4, 5\}$.

3.2.2 Loi d'une variable aléatoire**Définition 3.4**

Si $X : \Omega \rightarrow E$ est une variable aléatoire, on définit l'application suivante

$$\begin{aligned} \mathbb{P}_X : \mathcal{P}(E) &\longrightarrow [0, 1] \\ F &\longmapsto \mathbb{P}_X(F) := \mathbb{P}(\{\omega \in \Omega, X(\omega) \in F\}) = \mathbb{P}(\{X \in F\}) \end{aligned}$$

\mathbb{P}_X est appelée la loi de la variable aléatoire X (ou encore distribution de X).

Propriété 3.5

Ainsi définie, \mathbb{P}_X est une probabilité sur E .

Remarque

Il faut bien faire la distinction entre \mathbb{P} qui est une probabilité sur Ω et \mathbb{P}_X qui est une probabilité sur E .

Preuve

Vérifions les deux axiomes d'une probabilité : tout d'abord $\mathbb{P}_X(E) = \mathbb{P}(\{\omega \in \Omega, X(\omega) \in E\}) = \mathbb{P}(\Omega) = 1$. Soit ensuite une suite (E_n) d'événements deux à deux disjoints de E . Alors

$$\begin{aligned}\mathbb{P}_X(\cup_{n \geq 1} E_n) &= \mathbb{P}(\{\omega \in \Omega, X(\omega) \in \cup_{n \geq 1} E_n\}), \\ &= \mathbb{P}(\cup_{n \geq 1} \{\omega \in \Omega, X(\omega) \in E_n\}), \\ &= \sum_{n=1}^{+\infty} \mathbb{P}(\{\omega \in \Omega, X(\omega) \in E_n\}), \text{ car les événements sont disjoints,} \\ &= \sum_{n=1}^{+\infty} \mathbb{P}_X(E_n).\end{aligned}$$

Exemple

Dans notre exemple, la probabilité \mathbb{P}_X (probabilité sur $E = \{-2, 3\}$) est entièrement déterminée par la donnée de $\mathbb{P}_X(\{-2\}) = \mathbb{P}(\{1, 2, 3, 4, 5\}) = \frac{5}{6}$ (et donc $\mathbb{P}_X(\{3\}) = \mathbb{P}(\{6\}) = \frac{1}{6}$).

Exemple

On considère l'expérience consistant à lancer une pièce équilibrée n fois. Ici $\Omega = \{P, F\}^n$. Soit la variable aléatoire X de Ω dans \mathbf{N} qui associe le nombre de fois où la pièce est tombée sur Pile. Alors la loi de X est donnée par

$$\mathbb{P}(X = k) = \frac{\text{nombre de tirages avec } k \text{ Piles}}{\text{nombre de tirage total}} = \frac{\binom{n}{k}}{2^n}, \quad k = 0, \dots, n.$$

Remarque

La loi d'une variable aléatoire dépend bien sûr du choix initial de l'univers Ω et de la probabilité \mathbb{P} : changer (Ω, \mathbb{P}) , c'est changer \mathbb{P}_X . Pour bien le voir, revenons à un exemple abordé en fin de Chapitre 2 : on tire n boules uniformément au hasard parmi N boules dont N_1 sont rouges et les autres sont noires.

Dans ce cas, la notion de variable aléatoire permet de reformuler les résultats trouvés au Chapitre 2. Notons X la variable aléatoire qui à un tirage ω associe le nombre de boules rouges présentes dans le tirage ω . Alors calculer la probabilité de l'événement "avoir k boules rouges", c'est précisément calculer $\mathbb{P}_X(\{k\}) = \mathbb{P}(X = k)$. Rappelons que \mathbb{P}_X est alors une probabilité sur l'ensemble $E = \{0, \dots, n\}$. En particulier, \mathbb{P}_X est très différente de la probabilité \mathbb{P} sous-jacente : \mathbb{P} est la loi uniforme sur Ω ensemble de vecteurs à valeurs dans $\{1, \dots, N\}$ alors que \mathbb{P}_X est une probabilité sur $\{0, \dots, n\}$ (qui en plus n'est pas uniforme).

Si le tirage est initialement avec remise, la loi \mathbb{P}_X est donnée par la loi binomiale de paramètres n et $p = \frac{N_1}{N}$:

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

Si maintenant on change l'univers Ω et on considère un tirage sans remise, nous avons vu que la loi de X est la loi hypergéométrique

$$\mathbb{P}_X(\{k\}) = \mathbb{P}(X = k) = \frac{\binom{N_1}{k} \binom{N-N_1}{n-k}}{\binom{N}{n}}.$$

Conclusion : changer (Ω, \mathbb{P}) , c'est changer \mathbb{P}_X .

Remarque

Nous n'avons rigoureusement défini de probabilité que sur un ensemble dénombrable. Or nous avons dit que E peut être égal à \mathbf{R} . Pour rester dans le cadre de notre définition, il suffit de constater qu'il suffit de définir une probabilité sur l'ensemble $Im(X) = \{X(\omega), \omega \in \Omega\}$ qui est un sous-ensemble dénombrable de E . Nous parlerons cependant de probabilité sur E .

3.3 Variables aléatoires discrètes réelles

3.3.1 Notion de fonction de répartition

On suppose ici que $E = \mathbf{R}$. Soit X une variable aléatoire de Ω dans \mathbf{R} .

Définition 3.6

On appelle fonction de répartition de X , la fonction $F_X : \mathbf{R} \mapsto [0, 1]$ définie par

$$\forall x \in \mathbf{R}, F_X(x) = \mathbb{P}_X([-\infty, x]) = \mathbb{P}(X \leq x).$$

Propriété 3.7

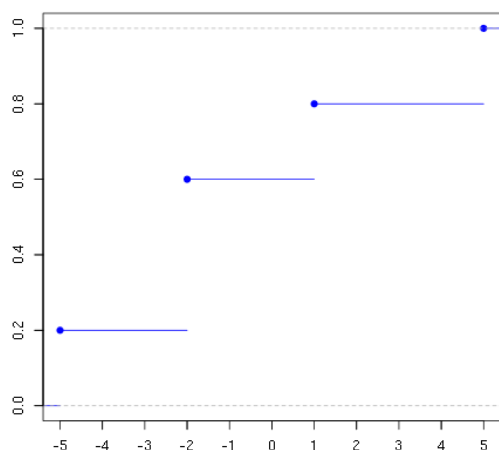
1. pour tout $a < b$, $F_X(b) - F_X(a) = \mathbb{P}_X([a, b]) = \mathbb{P}(a < X \leq b)$ (attention aux inégalités stricte et large!)
2. la fonction F_X est croissante sur \mathbf{R} ,
3. $\exists \lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\exists \lim_{x \rightarrow +\infty} F_X(x) = 1$,
4. la fonction F_X est continue à droite en tout point, et admet une limite à gauche en tout point x , notée $F_X(x^-)$,
5. F_X est constante par morceaux, et pour tout $x \in \mathbb{R}$, $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$.

Preuve

1. Soient $a < b$. Alors $F_X(b) - F_X(a) = \mathbb{P}(X \in]-\infty, b]) - \mathbb{P}(X \in]-\infty, a]) = \mathbb{P}(X \in]a, b])$.
2. Si $x \leq y$ alors $]-\infty, x] \subset]-\infty, y]$. Mais alors $\{\omega \in \Omega, X(\omega) \in]-\infty, x]\} \subset \{\omega \in \Omega, X(\omega) \in]-\infty, y]\}$. En prenant la probabilité \mathbb{P} de ces deux ensembles, on obtient donc $F_X(x) \leq F_X(y)$. Donc F_X est croissante.
3. Pour $x \in \mathbf{R}$, posons $p = \lfloor x \rfloor + 1 \in \mathbf{N}$. Alors par croissance de F_X , $F_X(x) \leq F_X(p) = \mathbb{P}_X([-\infty, p])$. Or, la suite d'ensembles $(]-\infty, p])_p$ est décroissante pour l'inclusion, quand $p \rightarrow -\infty$. Par la propriété d'une probabilité, $F_X(p)$ converge quand $p \rightarrow -\infty$ (et donc quand $x \rightarrow -\infty$) vers $\mathbb{P}_X(\cap_{p \geq 1}]-\infty, p]) = \mathbb{P}_X(\emptyset) = 0$.
De même $F_X(x) \geq F_X(q) = \mathbb{P}_X([-\infty, q])$ pour $q = \lfloor x \rfloor$. Or, la suite d'ensembles $(]-\infty, q])_q$ est croissante pour l'inclusion, quand $q \rightarrow \infty$. Donc par propriété d'une probabilité, $F_X(q) \rightarrow \mathbb{P}_X(\cup_{q \geq 1}]-\infty, q]) = \mathbb{P}_X(\mathbf{R}) = 1$.
4. Soit $x \in \mathbb{R}$ et $(x_n)_{n \geq 1}$ une suite qui tend vers x , telle que $x_n > x$ pour tout $n \geq 1$. Sans perte de généralité, on peut supposer (x_n) décroissante. Mais alors la suite $(]-\infty, x_n])_n$ est décroissante pour l'inclusion, donc $F_X(x_n)$ converge pour $n \rightarrow \infty$ vers $\mathbb{P}_X(\cap_{n \geq 1}]-\infty, x_n]) = \mathbb{P}_X([-\infty, x]) = F_X(x)$. Donc F_X est continue à droite en tout point. Prouvons qu'elle admet une limite à gauche en x . En effet, soit $(y_n)_{n \geq 1}$ une suite qui tend vers x par valeurs inférieures en croissant. Par le même argument, $F_X(y_n)$ converge quand $n \rightarrow \infty$ vers $\mathbb{P}_X(\cup_{n \geq 1}]-\infty, y_n]) = \mathbb{P}_X([-\infty, x[)$ qui est donc la limite, notée $F(x^-)$. On fera attention au fait que l'intervalle précédent est ouvert en x .

5. Notons $(x_i)_{i \geq 1}$ l'ensemble ordonné des valeurs prises par X . On peut supposer que $x_i < x_{i+1}$. Mais alors pour tout $x_i \leq s < t < x_{i+1}$, on a $F_X(t) - F_X(s) = \mathbb{P}_X([s, t]) = \mathbb{P}(X \in]s, t]) = 0$ car X ne prend pas ses valeurs dans $]s, t]$. Donc F_X est constante sur l'intervalle $[x_i, x_{i+1}[$, ce qui prouve le résultat.

L'égalité $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$ vient du calcul du paragraphe précédent.



Fonction de répartition d'une variable X telle que $\mathbb{P}(X = -5) = 0,2$, $\mathbb{P}(X = -2) = 0,4$, $\mathbb{P}(X = 1) = 0,2$, $\mathbb{P}(X = 5) = 0,2$.

FIGURE 3.1 – Fonction de répartition

Exemple : Exercice

- Si la loi de X est la loi de Dirac en $a \in \mathbf{R}$, alors $F_X(x) = \mathbf{1}_{[a, +\infty[}(x)$.
- Si la loi de X est telle que $\mathbb{P}(X = a) = p$ et $\mathbb{P}(X = b) = 1 - p$ (loi de Bernoulli de paramètre p en les valeurs $a < b$) alors $F_X(x) = p \cdot \mathbf{1}_{[a, b[}(x) + \mathbf{1}_{[b, +\infty[}(x)$.
- Plus généralement, si X ne prend qu'un nombre fini de valeurs $x_1 < x_2 < \dots < x_n$, (de probabilités respectives p_i) alors $F_X(x)$ vaut 0 si $x < x_1$, $p_1 + p_2 + \dots + p_i$ si $x_i \leq x < x_{i+1}$ ($i = 1, \dots, n$) et 1 si $x_{i+1} \leq x$.

Propriété 3.8

La fonction de répartition caractérise la loi de X . Autrement dit, deux variables aléatoires ayant même fonction de répartition ont même loi.

Preuve

En effet, si on note $(x_i)_i$ l'ensemble des valeurs possibles de X , alors $\mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) = F_X(x_i) - F_X(x_i^-)$.

On a de plus un résultat d'existence :

Propriété 3.9

Soit une fonction F de \mathbf{R} dans $[0, 1]$, vérifiant les propriétés de la proposition précédente. Alors il existe une variable aléatoire X telle que $F_X = F$.

Preuve

La fonction F étant constante par morceaux, ses points de discontinuités sont dénombrables $\Omega = \{x_i \mid i \in J\}$ pour un certain J dénombrable (ou fini), avec $x_i < x_{i+1}$. Posons alors $\mu(\{x_i\}) = F(x_i) - F(x_{i-1})$. Alors μ définit une probabilité sur Ω et la fonction $X : x \in \Omega \mapsto x$ a pour loi μ et fonction de répartition F .

3.3.2 Notions d'espérance et moments d'une variable aléatoire réelle

Soit X une variable aléatoire de (Ω, \mathbb{P}) dans \mathbf{R} .

Définition 3.10

Si la série de terme général $X(\omega_i)\mathbb{P}(\{\omega_i\})$ est absolument convergente, c'est-à-dire si

$$\sum_{i=1}^{+\infty} |X(\omega_i)| \mathbb{P}(\{\omega_i\}) < +\infty$$

on dit que X admet un moment d'ordre 1 (on dit encore que X est intégrable). Dans ce cas, la somme suivante existe

$$\sum_{i=1}^{+\infty} X(\omega_i) \mathbb{P}(\{\omega_i\})$$

et est notée $\mathbb{E}(X)$. $\mathbb{E}(X)$ est appelée espérance de X (ou encore moyenne de X).

Plus généralement, on dit que X admet un moment d'ordre k ($k \geq 1$) si X^k admet un moment d'ordre 1.

Remarque

Le fait que $\mathbb{E}(X)$ définie par la somme précédente existe vient du fait qu'une série absolument convergente est convergente.

Attention

- L'expression $\mathbb{E}(|X|) = \sum_{i=1}^{+\infty} |X(\omega_i)| \mathbb{P}(\{\omega_i\})$ (**attention aux valeurs absolues**) a toujours un sens dans $[0, +\infty[\cup \{+\infty\}$ car c'est une série à termes positifs. Dans ce cadre, écrire que X admet un moment d'ordre 1, c'est par définition écrire que $\mathbb{E}(|X|) < +\infty$.
- Par contre, il existe des variables X pour lesquelles l'expression $\mathbb{E}(X)$ n'a pas de sens ! Il est donc crucial qu'à chaque fois qu'on vous demande de calculer l'espérance d'une variable, vous ayez préalablement justifié que $\mathbb{E}(|X|) < +\infty$, et donc que $\mathbb{E}(X)$ a bien un sens.

Propriété 3.11

Si X est une constante égale à a alors $\mathbb{E}(X) = a$. En particulier, toute variable admet un moment d'ordre 0 et $\mathbb{E}(X^0) = 1$.

Preuve

Vérifions tout d'abord que la série de terme général $(|X(\omega_i)|\mathbb{P}(\{\omega_i\}))$ est convergente. En effet, ici $X(\omega_i) = a$ pour tout $i \geq 1$ donc $|X(\omega_i)|\mathbb{P}(\{\omega_i\}) = |a|\mathbb{P}(\{\omega_i\})$. Or ceci est le terme général d'une série convergente (par propriété d'une probabilité, $\sum_{i=1}^{+\infty} \mathbb{P}(\{\omega_i\}) = \mathbb{P}(\Omega) = 1 < +\infty$). Donc X admet un moment d'ordre 1 et $\mathbb{E}(X) = \sum_{i=1}^{+\infty} X(\omega_i)\mathbb{P}(\{\omega_i\})$ a bien un sens. Mais cette expression est ici égale à $\mathbb{E}(X) = \sum_{i=1}^{+\infty} a\mathbb{P}(\{\omega_i\}) = a \sum_{i=1}^{+\infty} \mathbb{P}(\{\omega_i\}) = a\mathbb{P}(\Omega) = a$.

Propriété 3.12 : Formule de transfert (1ère version)

Si X est une variable aléatoire admettant un moment d'ordre 1, on note $Im(X) = \{x_k, k \in K\}$ l'ensemble des valeurs prises par X . $Im(X)$ est fini ou dénombrable. Alors

$$\mathbb{E}(X) = \sum_{k \in K} x_k \mathbb{P}(X = x_k) = \sum_{k \in K} x_k \mathbb{P}_X(\{x_k\}).$$

Propriété 3.13 : Formule de transfert (2ième version)

Soit X une variable aléatoire. On note $Im(X) = \{x_k, k \in K\}$ l'ensemble des valeurs prises par X . $Im(X)$ est fini ou dénombrable. Alors pour tout fonction f réelle telle que $f(X)$ admette un moment d'ordre 1,

$$\mathbb{E}(f(X)) = \sum_{k \in K} f(x_k) \mathbb{P}(X = x_k) = \sum_{k \in K} f(x_k) \mathbb{P}_X(\{x_k\}).$$

Remarque

Cette proposition est appelée « formule de transfert » car elle transforme une expression faisant intervenir l'espace probabilisé de départ (Ω, \mathbb{P}) ($\sum_{i=1}^{+\infty} f(X(\omega_i)) \mathbb{P}(\{\omega_i\})$) en une expression faisant intervenir l'espace d'arrivée $(\mathbf{R}, \mathbb{P}_X)$: $\sum_{k \in K} f(x_k) \mathbb{P}_X(\{x_k\})$.

Cette formule fait écho au principe évoqué au début du paragraphe 3.2 : nous sommes finalement plus intéressés par les valeurs prises par X (et la probabilité \mathbb{P}_X d'obtenir ces valeurs) plutôt que par l'expérience aléatoire (Ω, \mathbb{P}) qui a permis de les obtenir.

Preuve

Pour tout $k \in K$, posons $\Omega_k = \{\omega_i, X(\omega_i) = x_k\} = X^{-1}(\{x_k\})$, c'est-à-dire l'ensemble des occurrences dont l'image par X est x_k . Alors, en sommant par paquets, et en utilisant que $(\Omega_k)_{k \in K}$ est une partition de Ω , et donc que $1 = \sum_{k \in K} \mathbf{1}_{\Omega_k}(\omega)$ pour tout ω ,

$$\begin{aligned} \mathbb{E}(f(X)) &= \sum_{i=1}^{+\infty} f(X(\omega_i)) \mathbb{P}(\{\omega_i\}), \\ &= \sum_{i=1}^{+\infty} \left(\sum_{k \in K} \mathbf{1}_{\Omega_k}(\omega_i) \right) f(X(\omega_i)) \mathbb{P}(\{\omega_i\}), \\ &= \sum_{k \in K} \left(\sum_{i=1}^{+\infty} \mathbf{1}_{\Omega_k}(\omega_i) f(X(\omega_i)) \mathbb{P}(\{\omega_i\}) \right), \\ &= \sum_{k \in K} f(x_k) \left(\sum_{i, \omega_i \in \Omega_k} \mathbb{P}(\{\omega_i\}) \right) = \sum_{k \in K} f(x_k) \mathbb{P}(\Omega_k), \\ &= \sum_{k \in K} f(x_k) \mathbb{P}(X = x_k). \end{aligned}$$

On remarque que X admet un moment d'ordre 1 si la série de terme général $(x_k \mathbb{P}(X = x_k))_{k \in K}$ est absolument convergente. C'est le critère que nous utiliserons usuellement.

Propriété 3.14

Pour tout $A \subset \mathbf{R}$, on a

$$\mathbb{P}(X \in A) = \mathbb{E}(\mathbf{1}_A(X)).$$

Preuve

$Y = \mathbf{1}_A(X)$ est une variable aléatoire à valeurs dans $\{0, 1\}$. Donc d'après précédemment,

$$\mathbb{E}(\mathbf{1}_A(X)) = \mathbb{E}(Y) = 0 \cdot \mathbb{P}(Y = 0) + 1 \cdot \mathbb{P}(Y = 1) = \mathbb{P}(Y = 1) = \mathbb{P}(X \in A).$$

Propriété 3.15 : Linéarité de l'espérance

Si X et Y ont toutes deux un moment d'ordre 1, alors pour tout $\alpha \in \mathbf{R}$, il en est de même de $\alpha X + Y$ et

$$\mathbb{E}(\alpha X + Y) = \alpha \mathbb{E}(X) + \mathbb{E}(Y).$$

Preuve

C'est une conséquence immédiate de la linéarité de la somme pour des séries convergentes.

Propriété 3.16 : Domination

- Si Y admet un moment d'ordre 1 et si $|X(\omega)| \leq |Y(\omega)|$ pour tout $\omega \in \Omega$, alors X admet un moment d'ordre 1 et $\mathbb{E}(|X|) \leq \mathbb{E}(|Y|)$.
- Si X et Y admettent un moment d'ordre 1 et si $X(\omega) \leq Y(\omega)$ pour tout $\omega \in \Omega$, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Preuve

Le fait que $|X(\omega)|\mathbb{P}(\{\omega\}) \leq |Y(\omega)|\mathbb{P}(\{\omega\})$ et que la série de terme général $|Y(\omega_i)|\mathbb{P}(\{\omega_i\})$ est convergente fait que la série de terme général $|X(\omega_i)|\mathbb{P}(\{\omega_i\})$ est aussi convergente et le fait que le passage à la limite conserve les inégalités donne le résultat.

En particulier, nous avons la remarque importante

Propriété 3.17

Toute variable bornée admet des moments de tout ordre.

Propriété 3.18

Si X est une variable positive ($X(\omega) \geq 0$ pour tout ω), alors

1. $\mathbb{E}(X) \geq 0$,
2. $\mathbb{E}(X) = 0$ implique que $\mathbb{P}(X = 0) = 1$. Dans ce cas, nous dirons que X est nulle, « presque sûrement ».

Preuve

Ici $x_k \geq 0$ pour tout $k \in K$. Donc $\mathbb{E}(X) = \sum_{k \in K} x_k \mathbb{P}(X = x_k) \geq 0$ et vaut 0 si et seulement si K est réduit au singleton $\{0\}$ ce qui veut dire que $\mathbb{P}(X = 0) = 1$.

Remarque

Évidemment, ce résultat est faux si on ne suppose pas la variable positive. La proposition précédente sera souvent appliquée pour $Y = |X|$ ou $Y = X^2$, où X est une variable quelconque : $\mathbb{E}(|X|) = 0$ (resp. $\mathbb{E}(X^2) = 0$) implique que $X = 0$ presque sûrement.

Propriété 3.19

Si X admet un moment d'ordre k pour $k \geq 1$ alors X admet un moment d'ordre $k - 1$.

Preuve

Cela vient de l'inégalité $|X|^{k-1} \leq |X|^k + 1$ et du résultat précédent (pour se convaincre de cette inégalité, on peut distinguer les cas $|X| \geq 1$ et $|X| < 1$).

Propriété 3.20 : Inégalité de Cauchy-Schwarz

Si X et Y admettent chacune un moment d'ordre 2 alors XY admet un moment d'ordre 1 et

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

Preuve

Le premier point vient de l'inégalité $|XY| \leq \frac{1}{2}(X^2 + Y^2)$. Pour le deuxième point, considérons la fonction $\varphi : t \mapsto \mathbb{E}((tX + Y)^2)$. φ est une fonction polynomiale de degré 2, de signe constant (positif) : $\varphi(t) = t^2\mathbb{E}(X^2) + 2t\mathbb{E}(XY) + \mathbb{E}(Y^2) \geq 0$. En particulier, le discriminant de ce polynôme est nécessairement négatif (car si ce dernier était positif, le polynôme changerait de signe, ce qui n'est pas le cas). Ce qui s'écrit : $0 \geq \Delta = 4\mathbb{E}(XY)^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2)$, ce qui donne $\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$, ce qui donne le résultat.

On en déduit le résultat suivant :

Propriété 3.21

Si X et Y admettent des moments d'ordre 2, alors il en est de même de $\alpha X + Y$.

Preuve

En effet, $|\alpha X + Y|^2 = \alpha^2 X^2 + 2\alpha X|Y| + Y^2$ qui est la somme de trois variables admettant un moment d'ordre 1.

Définition 3.22

Soit X admettant un moment d'ordre 2. Alors X admet un moment d'ordre 1 et on pose

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

$\mathbb{V}(X)$ est appelée variance de X . $\sigma = \sqrt{\mathbb{V}(X)}$ est appelé écart-type de X .

Remarque

La variance d'une variable aléatoire renseigne sur l'écart typique qui existe entre une variable X et sa moyenne. En particulier, dans le cas extrême où $\mathbb{V}(X) = 0$, X est nécessairement constante, égale à sa moyenne $\mathbb{E}(X)$. Plus la variance d'une variable aléatoire est grande, plus la variable aléatoire aura tendance à prendre des valeurs typiques éloignées de $\mathbb{E}(X)$.

Définition 3.23

Soient X et Y admettant des moments d'ordre 2. On définit la covariance de X et de Y par

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

Propriété 3.24

Soit X admettant un moment d'ordre 2,

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

Preuve

Développons :

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2), \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \end{aligned}$$

Propriété 3.25

Soient X et Y admettant des moments d'ordre 2. Alors,

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

Preuve

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)), \\ &= \mathbb{E}(XY) - 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

Propriété 3.26

Soient X et Y admettant un moment d'ordre 2 et α et β des réels. Alors les identités suivantes sont vraies :

1. $\mathbb{V}(\alpha X + \beta) = \alpha^2 \mathbb{V}(X)$,
2. $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{Cov}(X, Y)$,
3. $\text{Cov}(\alpha X + \beta, Y) = \alpha \text{Cov}(X, Y)$.

Preuve

1. $\mathbb{V}(\alpha X + \beta) = \mathbb{E}((\alpha X + \beta - \mathbb{E}(\alpha X + \beta))^2) = \mathbb{E}(\alpha^2(X - \mathbb{E}(X))^2) = \alpha^2 \mathbb{V}(X)$.
2. $\mathbb{V}(X + Y) = \mathbb{E}((X + Y - \mathbb{E}(X) - \mathbb{E}(Y))^2) = \mathbb{E}((X - \mathbb{E}(X))^2 + 2(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) + (Y - \mathbb{E}(Y))^2) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{Cov}(X, Y)$.
3. $\text{Cov}(\alpha X + \beta, Y) = \mathbb{E}((\alpha X + \beta - \mathbb{E}(\alpha X + \beta))(Y - \mathbb{E}(Y))) = \alpha \text{Cov}(X, Y)$.

Propriété 3.27

Soient X et Y ayant des moments d'ordre 2. Alors,

$$|\text{Cov}(X, Y)| \leq \sqrt{\mathbb{V}(X)\mathbb{V}(Y)}.$$

Preuve

C'est l'inégalité de Cauchy-Schwarz appliquée à $X - \mathbb{E}(X)$ et $Y - \mathbb{E}(Y)$.

Propriété 3.28 : Inégalité de Markov

Soit X une variable réelle admettant un moment d'ordre 1. Alors pour tout $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

Preuve

L'inégalité $a\mathbf{1}_{|X| \geq a} \leq |X|$ est vraie. Prenant l'espérance dans cette inégalité, on obtient $a\mathbb{E}(\mathbf{1}_{|X| \geq a}) \leq \mathbb{E}(|X|)$. Or, $\mathbb{E}(\mathbf{1}_{|X| \geq a}) = \mathbb{P}(|X| \geq a)$, d'où le résultat.

Propriété 3.29

Soit X une variable réelle admettant un moment d'ordre $k \geq 1$. Alors pour tout $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|^k)}{a^k}.$$

Preuve

Cela vient de l'inégalité $a^k \mathbf{1}_{|X| \geq a} \leq |X|^k$.

Propriété 3.30 : Inégalité de Bienaymé-Tchebychev

Soit X une variable réelle admettant un moment d'ordre 2. Alors pour tout $a > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| > a) \leq \frac{\mathbb{V}(X)}{a^2}$$

Preuve

C'est l'inégalité précédente appliquée à $X - \mathbb{E}(X)$ pour $k = 2$.

3.4 Lois usuelles

3.4.1 Cas fini

3.4.1.1 Loi de Dirac

Définition 3.31

On dit que X suit la loi de Dirac en $a \in \mathbf{R}$ si $\mathbb{P}(X = a) = 1$. Autrement dit, X est alors déterministe (non aléatoire), et $\mathbb{E}(X) = a$, $\mathbb{V}(X) = 0$.

3.4.1.2 Loi de Bernoulli

Définition 3.32

On dit que X suit la loi de Bernoulli de paramètre p si

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

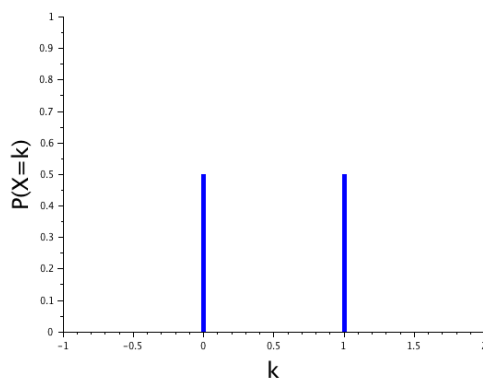
Exemple : on lance une pièce de monnaie non équilibrée (univers $\Omega = \{Pile, Face\}$ et $\mathbb{P}(Pile) = p = 1 - \mathbb{P}(Face)$). Alors $X(Pile) = 1$ et $X(Face) = 0$ suit une loi de Bernoulli de paramètre p .

Propriété 3.33

Si X suit une loi de Bernoulli alors $\mathbb{E}(X) = p$ et $\mathbb{V}(X) = p(1 - p)$.

Preuve

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



Loi de Bernoulli de paramètre $p = \frac{1}{2}$ (pièce équilibrée) : on attribue la même probabilité à l'échec 0 qu'au succès 1.

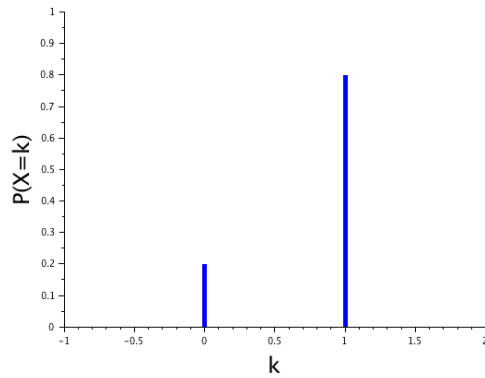
FIGURE 3.2 – Loi de Bernoulli

3.4.1.3 Loi binomiale

Définition 3.34 : Loi binomiale

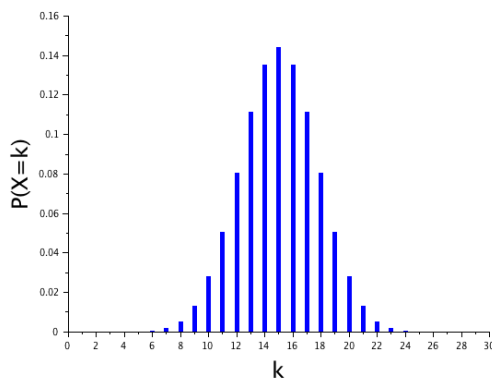
On dit que X suit une loi binomiale de paramètres $n \in \mathbf{N}^*$ et $p \in [0, 1]$ si pour tout $k = 0, \dots, n$

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



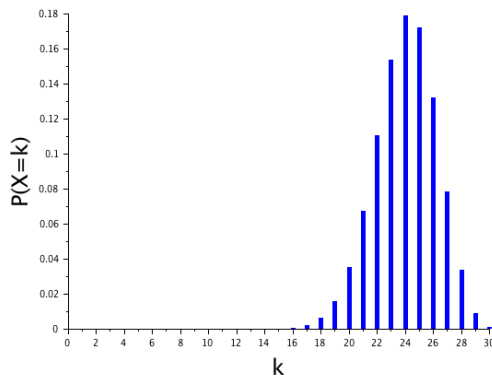
Loi de Bernoulli de paramètre $p = 0,8$ (pièce truquée) : on attribue une probabilité plus faible ($1 - p = 0,2$) à l'échec 0 qu'au succès 1 ($p = 0,8$).

FIGURE 3.3 – Loi de Bernoulli



Loi binomiale de paramètres $p = \frac{1}{2}$ et $n = 30$.

FIGURE 3.4 – Loi binomiale (30, 0.5)



Loi binomiale de paramètres $p = 0.8$ et $n = 30$.

FIGURE 3.5 – Loi binomiale (30, 0.8)

Exemple

Si X est le nombre de Pile dans un lancer de n pièces indépendantes de probabilité de succès p , alors X suit une loi binomiale de paramètres (n, p) (voir fin du chapitre précédent).

Propriété 3.35

Si X suit une loi binomiale de paramètres n et p alors

$$\mathbb{E}(X) = np \text{ et } \mathbb{V}(X) = np(1 - p).$$

Preuve

X étant bornée, X admet des moments à tout ordre et

$$\mathbb{E}(X) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}.$$

Considérons la fonction $x \mapsto f(x) = (x+1-p)^n = \sum_{k=0}^n \binom{n}{k} x^k (1-p)^{n-k}$. Alors f est (infiniment) dérivable sur $[0, 1]$ et d'une part, $f'(x) = n(x+1-p)^{n-1}$ et d'autre part, $f'(x) = \sum_{k=0}^n \binom{n}{k} k x^{k-1} (1-p)^{n-k}$. En particulier, $\mathbb{E}(X) = pf'(p) = pn$. De même

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k}, \\ &= \sum_{k=1}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}, \\ &= p^2 f''(p) + \mathbb{E}(X), \\ &= p^2 n(n-1) + np, \end{aligned}$$

ce qui donne $\mathbb{V}(X) = p^2 n(n-1) + np - n^2 p^2 = np(1-p)$.

3.4.1.4 Loi hypergéométrique**Définition 3.36**

On dit que X suit une loi hypergéométrique de paramètres N , M et n si

$$\mathbb{P}(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \text{ si } k \in \{0, \dots, n\},$$

et 0 sinon.

Exemple

On tire **sans remise** n boules parmi N dont M sont rouges et $N - M$ noires. On considère X = nombre de boule rouges tirées.

Proposition 3.37

Si X suit une loi hypergéométrique de paramètres (N, M, n) alors si $p = \frac{M}{N}$ est la proportion de boules rouges, $\mathbb{E}(X) = np$ et $\mathbb{V}(X) = np(1-p) \left(\frac{N-n}{N-1} \right)$

Preuve

Admis.

Remarque

On remarque que l'espérance est la même que dans le cas d'un tirage avec remise (loi binomiale) mais que la variance est plus petite : le fait de ne pas remettre les boules réduit l'aléa de l'expérience.

3.4.2 Cas dénombrable**3.4.2.1 Loi géométrique****Définition 3.38**

On dit que X suit une loi géométrique de paramètre $p \in]0, 1[$ si

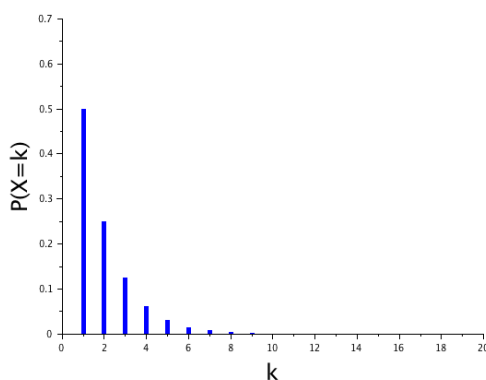
$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

Proposition 3.39

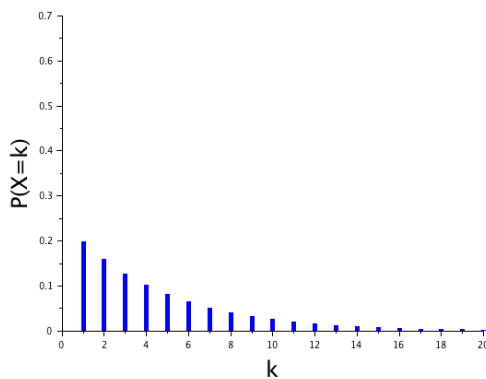
La loi définie plus haut est bien une probabilité sur \mathbb{N}^* : en effet, posons $p_k = p(1-p)^{k-1}$ pour tout $k \geq 1$. C'est une suite de réels positifs. Il suffit donc de vérifier que c'est le terme général d'une série convergente et que la somme vaut 1. En effet, $(1-p)^{k-1}$ est le terme général d'une série géométrique qui est bien convergente car $0 < 1-p < 1$ et nous avons $\sum_{k=1}^{+\infty} p_k = p \sum_{k=1}^{+\infty} (1-p)^{k-1} = p \frac{1}{1-(1-p)} = 1$. Nous avons bien défini une probabilité sur \mathbb{N}^* .

Remarque

Nous verrons à la fin de ce chapitre que la loi géométrique modélise le nombre de lancers successifs et indépendants d'une pièce de probabilité de Pile égale à p qu'il est nécessaire d'effectuer pour obtenir le premier Pile.

FIGURE 3.6 – Loi géométrique $p=0.5$

Loi géométrique de paramètre $p = \frac{1}{2}$. Attention : les valeurs de $\mathbb{P}(X = k)$ pour $k \geq 10$ sont trop faibles pour être visibles sur le graphe. Cependant, elles sont toutes non nulles : le support de la loi de X est \mathbb{N}^* tout entier.

FIGURE 3.7 – Loi géométrique $p=0.2$

Loi géométrique de paramètre $p = 0.2$. Attention : les valeurs de $\mathbb{P}(X = k)$ pour $k \geq 15$ sont trop faibles pour être visibles sur le graphe. Cependant, elles sont toutes non nulles : le support de la loi de X est \mathbb{N}^* tout entier.

Propriété 3.40

Si X suit une loi géométrique alors $\mathbb{E}(X) = \frac{1}{p}$ et $\mathbb{V}(X) = \frac{1-p}{p^2}$.

Preuve

X est une variable positive donc $\mathbb{E}(X) = \sum_{k=1}^{+\infty} kp(1-p)^{k-1}$ a toujours un sens dans $[0, +\infty]$. Pour tout $x \in [0, 1[$, posons $f(x) = \sum_{k=1}^{+\infty} x^k = \frac{x}{1-x}$ (série entière de rayon de convergence 1). Sur son rayon de convergence, f est dérivable et on peut dériver terme à terme : $f'(x) = \frac{1}{1-x} + \frac{x}{(1-x)^2} = \sum_{k=1}^{+\infty} kx^{k-1}$. En faisant $x = 1-p$, on obtient $\mathbb{E}(X) = pf'(1-p) = p \left(\frac{1}{p} + \frac{1-p}{p^2} \right) = \frac{1}{p}$. De même,

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=1}^{+\infty} k^2 p(1-p)^{k-1}, \\
&= \sum_{k=1}^{+\infty} k(k-1)p(1-p)^{k-1} + \sum_{k=1}^{+\infty} kp(1-p)^{k-1}, \\
&= p(1-p)f''(1-p) + pf'(1-p) = p(1-p)\frac{2}{p^3} + \frac{1}{p},
\end{aligned}$$

ce qui donne $\mathbb{V}(X) = p(1-p)\frac{2}{p^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

3.4.2.2 Loi de Poisson

Définition 3.41

On dit que X suit une loi de Poisson de paramètre $\lambda \in]0, +\infty[$ si pour tout $k \geq 0$

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

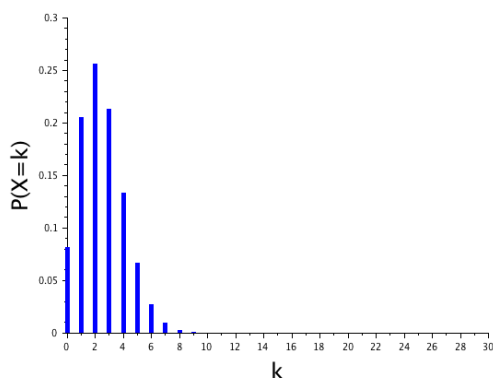


FIGURE 3.8 – Loi de Poisson

Loi de Poisson de paramètre $\lambda = 2.5$. Attention : les valeurs de $\mathbb{P}(X = k)$ pour $k \geq 10$ sont trop faibles pour être visibles sur le graphe. Cependant, elles sont toutes non nulles : le support de la loi de X est \mathbb{N} tout entier.

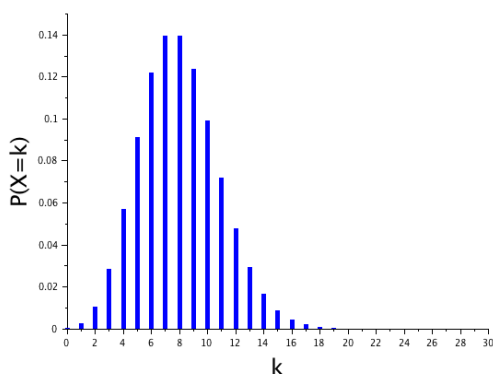


FIGURE 3.9 – Loi de Poisson

Loi de Poisson de paramètre $\lambda = 8$. Attention : les valeurs de $\mathbb{P}(X = k)$ pour $k \geq 20$ sont trop faibles pour être visibles sur le graphe. Cependant, elles sont toutes non nulles : le support de la loi de X est \mathbb{N} tout entier.

Proposition 3.42

La loi définie plus haut est bien une probabilité. En effet, posant pour tout $k \geq 0$, $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, on définit une suite de termes positifs qui constituent le terme général d'une série convergente (par critère de d'Alembert par exemple). De plus, reconnaissant le développement en série entière de la fonction exponentielle, il vient $\sum_{k=0}^{+\infty} p_k = e^{-\lambda} e^{\lambda} = 1$.

Propriété 3.43

Si X suit une loi de Poisson, alors $\mathbb{E}(X) = \lambda$ et $\mathbb{V}(X) = \lambda$.

Preuve

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{+\infty} e^{-\lambda} k \frac{\lambda^k}{k!}, \\ &= \lambda \sum_{k=1}^{+\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} = \lambda.\end{aligned}$$

De même,

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{k=0}^{+\infty} e^{-\lambda} k^2 \frac{\lambda^k}{k!}, \\ &= \sum_{k=0}^{+\infty} e^{-\lambda} k(k-1) \frac{\lambda^k}{k!} + \sum_{k=0}^{+\infty} e^{-\lambda} k \frac{\lambda^k}{k!} = \lambda^2 + \lambda,\end{aligned}$$

ce qui donne le résultat : $\mathbb{V}(X) = \lambda$.

3.5 Couples de variables aléatoires discrètes, lois conditionnelles, indépendance

Dans de nombreuses situations, nous aurons à étudier non pas une seule variable aléatoire, mais plusieurs variables aléatoires définies sur un même espace probabilisé (par exemple X et Y).

Définition 3.44

Soient X, Y deux variables aléatoires discrètes sur Ω à valeurs respectives dans E et F . L'application $\Omega \rightarrow E \times F$ définie par $\omega \mapsto (X(\omega), Y(\omega))$ est une variable aléatoire discrète appelée « couple aléatoire ».

La subtilité est que dans ce cas, la loi du vecteur (X, Y) n'est pas entièrement décrite par la connaissance des lois de ses composantes X et de Y . Considérons un exemple :

Exemple

On considère deux urnes A et B . Chacune d'entre elles a une chance sur deux de contenir soit 0 boule, soit 1 boule. Quel le nombre total de boules présentes ?

L'univers est ici $\Omega = \{0, 1\}^2$. Soient X_A et X_B les nombres respectifs de boules dans les urnes A et B . Ce qu'on cherche à connaître est (la loi de) $X = X_A + X_B$.

La réponse est qu'on ne peut pas répondre à la question en l'état, sans des précisions supplémentaires sur la façon dont est menée l'expérience (autrement dit, la probabilité \mathbb{P} sur l'univers Ω n'est pas entièrement définie par l'énoncé).

Notons tout d'abord que les lois de X_A et de X_B sont parfaitement connues : $\mathbb{P}(X_A = 0) = \mathbb{P}(X_A = 1) = \frac{1}{2}$ (idem pour X_B). Mais il existe plusieurs façons de construire l'expérience précédente en respectant les contraintes de l'énoncé :

1. Une première façon de procéder est de procéder de façon indépendante dans l'urne A et dans l'urne B . Alors X peut prendre les valeurs 0, 1, 2 et $\mathbb{P}(X = 0) = \mathbb{P}(X_A = 0 \cap X_B = 0) = \mathbb{P}(X_A = 0)\mathbb{P}(X_B = 0) = \frac{1}{4}$. De même $\mathbb{P}(X = 2) = \frac{1}{4}$ et $\mathbb{P}(X = 1) = \frac{1}{2}$.
2. Une autre façon de procéder est la suivante : on tire à pile ou face pour décider de mettre une boule dans l'urne A et **quoi qu'il arrive, on met un nombre de boule identique dans l'urne B**. Les lois respectives de X_A et de X_B n'ont pas changé, mais maintenant X ne prend plus ses valeurs que dans $\{0, 2\}$ et $\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = \frac{1}{2}$.

3. Une autre possibilité est de tirer à pile ou face pour décider de mettre une boule dans l'urne A et **quoi qu'il arrive, on met 0 boule dans l'urne B si on met 1 boule dans A et 1 boule dans B si on n'en a pas mise dans A**. Dans ce cas, les lois de X_A et X_B n'ont pas changé mais maintenant X vaut 1 tout le temps!

En fait, il y a une infinité d'expériences possibles vérifiant ces contraintes : ce qu'il manque dans l'énoncé, ce sont les relations de dépendance entre les urnes A et B . Pour définir entièrement l'expérience, il faut et il suffit de donner en plus les probabilités suivantes (qui définissent ce que nous appellerons « la loi jointe » de X_A et X_B) :

$$\begin{aligned}\mathbb{P}(X_A = 0 \cap X_B = 0), \\ \mathbb{P}(X_A = 0 \cap X_B = 1), \\ \mathbb{P}(X_A = 1 \cap X_B = 0), \\ \mathbb{P}(X_A = 1 \cap X_B = 1).\end{aligned}$$

A titre d'exercice, donner les valeurs des quatre probabilités précédentes dans les trois cas évoqués.

Définition 3.45

La loi du couple $Z = (X, Y)$ à valeurs dans $E \times F$, notée $\mathbb{P}_{X,Y}$, est appelée « loi jointe de X et Y ». La loi jointe de X et Y est la loi sur $E \times F$ définie pour $E' \in \mathcal{P}(E)$, $F' \in \mathcal{P}(F)$ par

$$\mathbb{P}_{X,Y}(E', F') = \mathbb{P}(\{\omega \in \Omega, X(\omega) \in E', Y(\omega) \in F'\}) = \mathbb{P}(X \in E', Y \in F').$$

Les lois de X et de Y sont appelées « lois marginales » du vecteur Z . Elles s'obtiennent facilement à partir de la loi jointe de X et de Y . En effet,

Propriété 3.46

Si on note $\{x_i, i \in I\}$ et $\{y_j, j \in J\}$ les valeurs prises respectivement par X et Y , alors pour tout $x \in E$, $y \in F$,

$$\begin{aligned}\mathbb{P}(X = x) &= \sum_{j \in J} \mathbb{P}_{X,Y}(x, y_j), \\ \mathbb{P}(Y = y) &= \sum_{i \in I} \mathbb{P}_{X,Y}(x_i, y).\end{aligned}$$

Preuve

Prouvons la première égalité, la seconde, semblable, est laissée à titre d'exercice. En utilisant la formule des probabilités totales (le système $(\{Y = y_j\})_{j \in J}$ étant un système complet d'événements),

$$\mathbb{P}(X = x) = \sum_{j \in J} \mathbb{P}(X = x, Y = y_j) = \sum_{j \in J} \mathbb{P}_{X,Y}(x, y_j).$$

Remarque

Si la loi jointe de X et Y détermine les lois de X et Y , la réciproque est fautive. Relire l'exemple de début de paragraphe concernant les deux urnes A et B : nous avons défini plusieurs lois jointes de X_A et de X_B distinctes sans changer les lois marginales.

Propriété 3.47

Pour toute fonction de deux variables $(x, y) \in E \times F \mapsto f(x, y) \in \mathbf{R}$ telle que $f(X, Y)$ est intégrable

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

Preuve

Evident, par la formule de transfert appliquée à la variable $Z = (X, Y)$.

Définition 3.48

Soit $y \in F$ tel que $\mathbb{P}(Y = y) \neq 0$. On appelle « loi conditionnelle de X sachant $Y = y$ » la probabilité $\mathbb{P}_{X|Y=y}$ sur E définie par

$$\mathbb{P}_{X|Y=y}(\{x\}) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

Ainsi, la loi $\mathbb{P}_{X|Y=y}$ est la loi \mathbb{P}_X conditionnée à l'événement $\{Y = y\}$.

Propriété 3.49

Si $\{y_j, j \in J\}$ est l'ensemble des valeurs prises respectivement par Y , alors pour tout $x \in E$,

$$\mathbb{P}(X = x) = \sum_{j \in J} \mathbb{P}(X = x | Y = y_j) \mathbb{P}(Y = y_j).$$

Preuve

Il s'agit de la formule des probabilités totales appliquée au système complet d'événements $(\{Y = y_j\})_{j \in J}$.

Propriété 3.50

Soit x tel que $\mathbb{P}(X = x) \neq 0$. Alors

$$\mathbb{P}(Y = y | X = x) = \frac{\mathbb{P}(X = x | Y = y) \mathbb{P}(Y = y)}{\sum_{j \in J} \mathbb{P}(X = x | Y = y_j) \mathbb{P}(Y = y_j)}.$$

Preuve

Il s'agit de la formule de Bayes.

Définition 3.51

Soient X et Y définies sur Ω . Si $\{x_i, i \in I\}$ et $\{y_j, j \in J\}$ sont les valeurs prises respectivement par X et Y , alors on dit que les variables X et Y sont indépendantes si, pour tout $i \in I$ et $j \in J$,

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

où de façon équivalente

$$\mathbb{P}(X = x_i | Y = y_j) = \mathbb{P}(X = x_i).$$

Exemple

1. Montrer que si X suit une loi de Poisson de paramètre λ et si Y suit une loi de Poisson de paramètre μ avec X et Y indépendantes, alors $Z = X + Y$ suit une loi de Poisson de paramètres $\lambda + \mu$.
2. Soit X une variable aléatoire de loi binomiale de paramètre $n \in \mathbf{N}^*$, $p \in]0, 1[$ et Y une variable aléatoire de loi binomiale de paramètre $m \in \mathbf{N}^*$, $p \in]0, 1[$. On suppose que $X + Y$ sont indépendantes. Quelle est la loi de $X + Y$?

Propriété 3.52

Si X et Y sont deux variables aléatoires à valeurs respectivement dans E et F . Les variables X et Y sont indépendantes si et seulement si pour toutes fonctions $f : E \rightarrow \mathbf{R}$, et $g : F \rightarrow \mathbf{R}$ bornées, on a

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).$$

Preuve

Cette condition est suffisante : si on l'applique aux fonctions $f = \mathbf{1}_{\{x_i\}}$ et $g = \mathbf{1}_{\{y_j\}}$ (qui sont des fonctions bornées), pour tout $i \in I$ et $j \in J$, on obtient

$$\begin{aligned} \mathbb{P}(X = x_i, Y = y_j) &= \mathbb{E}(\mathbf{1}_{\{x_i\}}(X)\mathbf{1}_{\{y_j\}}(Y)), \\ &= \mathbb{E}(\mathbf{1}_{\{x_i\}}(X))\mathbb{E}(\mathbf{1}_{\{y_j\}}(Y)), \\ &= \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j). \end{aligned}$$

Réciproquement, supposons que X et Y sont indépendantes et considérons deux fonctions f et g bornées. Alors, $f(X)g(Y)$ est intégrable car f et g sont bornées et

$$\begin{aligned} \mathbb{E}(f(X)g(Y)) &= \sum_{i \in I} \sum_{j \in J} f(x_i)g(y_j)\mathbb{P}(X = x_i, Y = y_j), \\ &= \sum_{i \in I} \sum_{j \in J} f(x_i)g(y_j)\mathbb{P}(X = x_i)\mathbb{P}(Y = y_j), \\ &= \left(\sum_{i \in I} f(x_i)\mathbb{P}(X = x_i)\right) \left(\sum_{j \in J} g(y_j)\mathbb{P}(Y = y_j)\right), \\ &= \mathbb{E}(f(X))\mathbb{E}(g(Y)). \end{aligned}$$

Ce qui donne le résultat.

Cette proposition se généralise aux fonctions telles que $f(X)$ et $g(Y)$ ont un moment d'ordre 2. Alors

Corollaire 3.53

Si X et Y sont deux variables aléatoires à valeurs dans \mathbf{R} admettant des moments d'ordre 2. Si X et Y sont indépendantes alors

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

On dit alors que X et Y sont décorrélées. La réciproque est fautive : deux variables peuvent être décorrélées sans pourtant être indépendantes.

En particulier, si X et Y sont indépendantes alors

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y).$$

Preuve

Il suffit de prendre $f(x) = x$ et $g(y) = y$ et d'appliquer les résultats vus dans les paragraphes précédents.

Propriété 3.54 : Contre-exemple

Voici un contre-exemple au fait que deux variables décorrélées ne sont pas forcément indépendantes. Posons X dont la loi est donnée par $\mathbb{P}(X = 1) = \mathbb{P}(X = 0) = \mathbb{P}(X = -1) = \frac{1}{3}$ et posons $Y = X^2$. On vérifie immédiatement que $\mathbb{E}(X) = 0$ et que $\mathbb{E}(XY) = \mathbb{E}(X^3) = 0$. Donc $\text{Cov}(X, Y) = 0$, les deux variables sont décorrélées.

Cependant, elles ne sont pas indépendantes : en effet on vérifie immédiatement (exercice) que $\mathbb{P}(X = 1)\mathbb{P}(Y = 1) = \frac{1}{3} \cdot \frac{2}{3} \neq \frac{1}{3}\mathbb{P}(X = 1, Y = 1) : X$ et Y ne sont pas indépendantes.

3.6 Vecteurs aléatoires, suites de variables aléatoires

3.6.1 Définitions et propriétés

On appelle suite de variables aléatoires une suite (finie ou infinie) de variables aléatoires $(X_n)_{n \geq 1}$ telle que $X_i : \Omega \rightarrow E_i$. Si cette suite est finie (de cardinal p), on appelle « vecteur aléatoire » le vecteur (X_1, \dots, X_p) . Le cas considéré dans le paragraphe précédent est le cas $p = 2$.

Définition 3.55

On dit que les variables aléatoires (X_1, \dots, X_p) ($p \geq 1$ fixé) sont indépendantes si pour tout $x_1, \dots, x_p \in E_1 \times \dots \times E_p$,

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_p = x_p) = \prod_{i=1}^p \mathbb{P}(X_i = x_i).$$

Définition 3.56

Soit une suite $(X_n)_{n \geq 1}$ de variables aléatoires. On dit que les variables X_n , $n \in \mathbf{N}$ sont indépendantes si les variables (X_1, \dots, X_p) sont indépendantes pour tout $p \geq 1$. Dans le cas où les variables X_n , $n \geq 1$ sont indépendantes et de même loi, on dit que ces variables sont indépendantes et identiquement distribuées, ce que nous noterons de manière abrégée en « i.i.d. ».

Propriété 3.57

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires. Alors les variables X_n , $n \geq 1$ sont indépendantes si et seulement si pour tout $p \geq 1$ et toutes fonctions f_1, \dots, f_p telles que $f_i : E_i \rightarrow \mathbf{R}$ est bornée, on a

$$\mathbb{E}(f_1(X_1)f_2(X_2) \dots f_p(X_p)) = \prod_{i=1}^p \mathbb{E}(f_i(X_i)).$$

Preuve

Même démonstration que dans le cas précédent.

Propriété 3.58

Soient X_1, \dots, X_p , p variables aléatoires à valeurs dans \mathbf{R} . Alors,

$$\begin{aligned} \mathbb{V}(\sum_{i=1}^p X_i) &= \sum_{1 \leq i, j \leq p} \text{Cov}(X_i, X_j), \\ &= \sum_{i=1}^p \mathbb{V}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j), \\ &= \sum_{i=1}^p \mathbb{V}(X_i) + 2 \sum_{1 \leq i < j \leq p} \text{Cov}(X_i, X_j). \end{aligned}$$

Preuve

$$\begin{aligned} \mathbb{V}(\sum_{i=1}^p X_i) &= \mathbb{E}((\sum_{i=1}^p X_i - \mathbb{E}(\sum_{i=1}^p X_i))^2), \\ &= \mathbb{E}((\sum_{i=1}^p X_i - \sum_{i=1}^p \mathbb{E}(X_i))^2), \\ &= \mathbb{E}((\sum_{i=1}^p (X_i - \mathbb{E}(X_i)))^2) \\ &= \mathbb{E}(\sum_{i,j=1}^p (X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))) \\ &= \sum_{i,j=1}^p \mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))). \end{aligned}$$

Les expressions trouvées dans la proposition sont des reformulations de l'expression précédente selon qu'on distingue les cas où $i = j$, $i < j$ et $i > j$.

Corollaire 3.59

Si X_1, \dots, X_p sont des variables discrètes réelles indépendantes, alors

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

Preuve

Dans ce cas, $\text{Cov}(X_i, X_j) = 0$ pour tout $i \neq j$.

3.6.2 Matrice de variance/covariance**Définition 3.60**

Pour $\mathbf{X} = (X_1, \dots, X_p)$ un vecteur aléatoire de taille $p \geq 1$ à valeurs dans \mathbf{R}^p , on définit son espérance $\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_p))$. De plus, on définit sa matrice de variance/covariance $\Gamma(\mathbf{X})$ par la matrice carrée de taille $p \geq 1$ suivante

$$\Gamma(\mathbf{X}) = \begin{pmatrix} \mathbb{V}(X_1) & \text{Cov}(X_1, X_2) & \text{Cov}(X_1, X_3) & \cdots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & \mathbb{V}(X_2) & \text{Cov}(X_2, X_3) & \cdots & \text{Cov}(X_2, X_p) \\ \text{Cov}(X_3, X_1) & \text{Cov}(X_3, X_2) & \mathbb{V}(X_3) & \cdots & \text{Cov}(X_3, X_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_p, X_1) & \text{Cov}(X_p, X_2) & \text{Cov}(X_p, X_3) & \cdots & \mathbb{V}(X_p) \end{pmatrix}$$

Propriété 3.61

La matrice $\Gamma(\mathbf{X})$ est symétrique donc diagonalisable et à valeurs propres positives. De plus, on a l'égalité

$$\Gamma(\mathbf{X}) = \mathbb{E}(\mathbf{X}\mathbf{X}^t) - \mathbb{E}(\mathbf{X})\mathbb{E}(\mathbf{X})^t.$$

Preuve

Exercice.

3.6.3 Exemples

Ces exemples sont fondamentaux : à retenir par cœur !

Propriété 3.62 : Loi de Bernoulli/ loi binomiale

Soient X_1, \dots, X_n , n variables i.i.d. de loi de Bernoulli de paramètre $p \in [0, 1]$. Alors la variable aléatoire $S_n = X_1 + \dots + X_n$ suit une loi binomiale de paramètres (n, p) .

Preuve

Déjà vu dans les chapitres précédents, mais réécrivons là dans le contexte de variables aléatoires. Par définition, S_n prend ses valeurs dans $\{0, \dots, n\}$ et pour tout k dans cet ensemble, $\mathbb{P}(S_n = k)$ est la probabilité qu'exactly k variables X_i parmi les n prennent la valeur 1 (et les autres prennent la valeur 0). Il y a autant de possibilité de choix de ces variables qu'il y a de choix de sous-ensembles I à k éléments. Ainsi,

$$\begin{aligned}
\mathbb{P}(S_n = k) &= \mathbb{P} \left(\bigcup_{I \subset \{0, \dots, n\}, \text{card}(I)=k} \bigcap_{i \in I} \{X_i = 1\} \cap_{i \in I^c} \{X_i = 0\} \right) \\
&= \sum_{I \subset \{0, \dots, n\}, \text{card}(I)=k} \mathbb{P}(\bigcap_{i \in I} \{X_i = 1\} \cap_{i \in I^c} \{X_i = 0\}), \text{ (union disjointe),} \\
&= \sum_{I \subset \{0, \dots, n\}, \text{card}(I)=k} \prod_{i \in I} \mathbb{P}(X_i = 1) \prod_{i \in I^c} \mathbb{P}(X_i = 0), \text{ (par indépendance des variables),} \\
&= \sum_{I \subset \{0, \dots, n\}, \text{card}(I)=k} p^k (1-p)^{n-k}, \text{ (car ce sont des variables i.i.d. de Bernoulli),} \\
&= \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

D'où le résultat.

Cela donne une façon plus simple de calculer les moyenne et variance d'une loi binomiale : en utilisant ce qu'on sait sur la loi de Bernoulli,

$$\begin{aligned}
\mathbb{E}(S_n) &= \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n) = np \\
\mathbb{V}(S_n) &= \mathbb{V}(X_1) + \dots + \mathbb{V}(X_n) = np(1-p).
\end{aligned}$$

La dernière égalité est bien sûr vraie car les variables sont indépendantes (rappel : la variance d'une somme de variables **indépendantes** est la somme des variances).

Propriété 3.63 : Loi de Bernoulli/ loi géométrique

Soit $(X_n)_{n \geq 1}$ une suite i.i.d. de loi de Bernoulli de paramètre $p \in]0, 1[$. On pose

$$T = \min \{i \geq 1, X_i = 1\}.$$

Alors T suit une loi géométrique de paramètre p .

Preuve

Pour tout $k \geq 1$,

$$\begin{aligned}
\mathbb{P}(T = k) &= \mathbb{P}(X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1), \\
&= \mathbb{P}(X_1 = 0) \mathbb{P}(X_2 = 0) \dots \mathbb{P}(X_{k-1} = 0) \mathbb{P}(X_k = 1), \text{ (par indépendance)} \\
&= (1-p)^{k-1} p.
\end{aligned}$$



Probabilités sur \mathbb{R} et variables aléatoires à densité

Introduction

Dans le chapitre précédent, nous avons traité le cas de variables aléatoires discrètes. Par exemple définir la loi d'une variable aléatoire X à valeurs dans \mathbf{N} revient à définir $p_k = \mathbb{P}(X = k)$ pour tout entier k . En particulier, pour tout couple d'entiers $k < l$,

$$\mathbb{P}(k < X \leq l) = \sum_{i=k+1}^l p_i. \quad (4.1)$$

Dans ce chapitre, nous nous intéressons à des variables à valeurs dans \mathbf{R} . Nous verrons que dans ce cas, toute notion raisonnable de la loi d'une telle variable X fait que pour tout réel x , $\mathbb{P}(X = x) = 0$. Autrement dit, il est impossible de construire une probabilité sur \mathbf{R} à partir de la probabilité des singletons. Ceci vient du fait que \mathbf{R} est un ensemble trop gros (au sens non dénombrable) pour qu'il soit possible de définir \mathbb{P} sur tout sous-ensemble de \mathbf{R} .

Plutôt que de définir des variables aléatoires à valeurs dans \mathbb{R} en toute généralité, nous allons considérer dans ce chapitre le cas particulier (qui suffit amplement dans les applications usuelles) des lois et des variables à densité. L'idée est d'écrire une égalité analogue 4.1 dans le cas discret : si X est une variable réelle à densité, nous écrirons : pour tous réels $a < b$

$$\mathbb{P}(a \leq X < b) = \int_a^b p(x) dx. \quad (4.2)$$

La sommation discrète $\sum_{i=k}^{l-1}$ dans 4.1 est devenue dans 4.2 une intégrale continue \int_a^b et la suite p_k dans 4.1 est devenue une fonction $x \mapsto p(x)$ dans 4.2. Nous appellerons cette fonction p une « densité de probabilité ». La différence majeure est que pour $x \in \mathbf{R}$, $p(x)$ **n'est pas** égale à $\mathbb{P}(X = x)$ (cette dernière quantité est nulle, comme nous l'avons dit plus haut). Intuitivement, il faut plutôt voir $p(x) dx$ comme une probabilité infinitésimale : $p(x) dx$ représente la probabilité que X se trouve dans le petit intervalle $[x, x + dx]$.

Attention : même si je viens de faire une analogie avec le chapitre précédent, il faudra bien faire la distinction entre une variable discrète et une variable continue ! La construction des ces deux objets n'est pas la même. En particulier, il ne faudra pas confondre les exemples vus dans ces deux chapitres bien distincts. Je pense notamment aux nombreux étudiants qui dans leur copie confondent une variable exponentielle (voir plus bas) avec une variable de Poisson, sous prétexte qu'il y a une exponentielle dans la définition d'une variable de Poisson...

4.1 Densités de probabilité sur \mathbb{R} et fonctions de répartition

4.1.1 Notion de densité de probabilité et exemples

Poursuivons l'analogie avec le cas discret entamée plus haut. Si $p_k \geq 0$ pour tout k , alors $p(x) \geq 0$ pour tout x . De même $\sum_{k=0}^{+\infty} p_k = 1$ devient $\int_{\mathbb{R}} p(x) dx = 1$. La définition suivante devient alors naturelle :

Définition 4.1

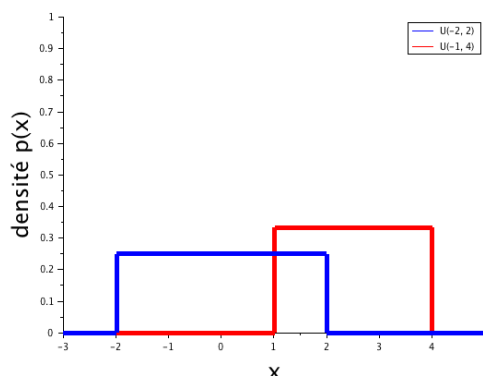
On appelle densité de probabilité sur \mathbb{R} une fonction $p : \mathbb{R} \mapsto \mathbb{R}$ telle que

1. $p(x) \geq 0$ pour tout $x \in \mathbb{R}$,
2. p est continue par morceaux et intégrable,
3. $\int_{\mathbb{R}} p(x) dx = 1$.

Nous rappelons qu'une fonction est dite continue par morceaux si sur tout segment, elle admet un nombre fini de discontinuités et admet une limite à droite et une limite à gauche en chacun de ces points. En fait, il est possible d'alléger cette hypothèse de continuité mais nous nous restreindrons à ce cas par simplicité.

Exemple : Exemples fondamentaux à connaître par coeur.

- Densité uniforme sur l'intervalle $[a, b]$, $(-\infty < a < b < +\infty)$: la densité uniforme sur l'intervalle $[a, b]$ est la fonction

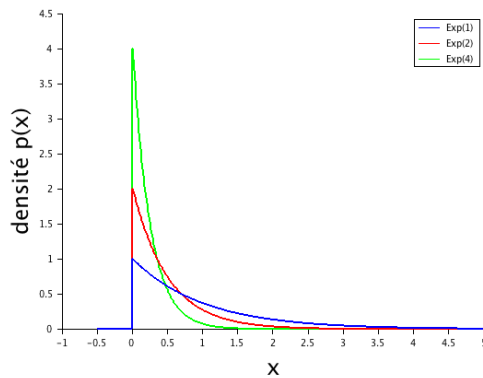


Densités des lois uniformes sur $[-1, 4]$ et $[-2, 2]$.

FIGURE 4.1 – Densités de lois uniformes sur un segment

$$x \mapsto p(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x), \text{ (i.e. } p(x) = \frac{1}{b-a} \text{ si } x \in [a, b] \text{ et } p(x) = 0 \text{ sinon).}$$

L'ensemble des densités uniformes forme donc une famille à deux paramètres (a, b) .



Densités de lois exponentielles pour différents paramètres $\lambda = 1, 2, 4$.

FIGURE 4.2 – Densités de lois exponentielles

- Densité exponentielle de paramètre $\lambda > 0$: il s'agit de la fonction p_λ définie par

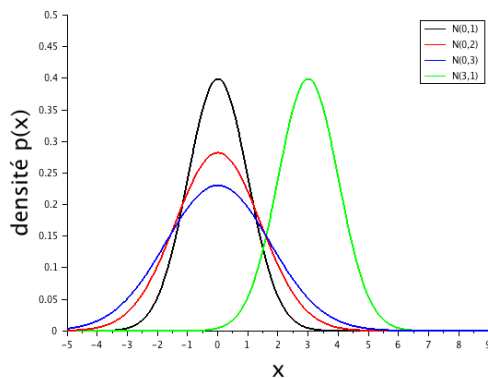
$$x \mapsto p_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}, \text{ (i.e. } p_\lambda(x) = \lambda e^{-\lambda x} \text{ si } x \geq 0 \text{ et } p_\lambda(x) = 0 \text{ sinon).}$$

L'ensemble des densités exponentielles forme une famille à un paramètre λ .

- Densité gaussienne (ou densité normale) $\mathcal{N}(m, \sigma^2)$, de paramètres $m \in \mathbf{R}$ et $\sigma > 0$: il s'agit de la densité p_{m, σ^2} définie par :

$$x \mapsto p_{m, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Il s'agit d'une famille à deux paramètres (m, σ^2) . Dans le cas où $m = 0$, $\sigma = 1$, on dit que la



Densités de lois gaussiennes pour différentes valeurs de m et σ^2 .

FIGURE 4.3 – Densités de lois gaussiennes

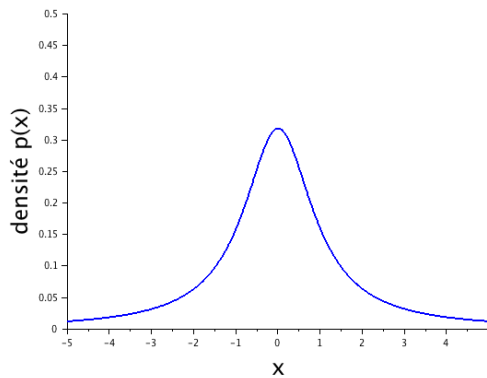
densité est **normale centrée réduite**. Le choix d'écrire σ^2 au lieu de σ deviendra clair plus bas dans ce chapitre.

- Densité de Cauchy : il s'agit de la fonction définie par

$$x \mapsto p(x) = \frac{1}{\pi(1+x^2)}.$$

4.1.2 Exercice

Montrer que les quatre exemples ci-dessus définissent bel et bien des densités de probabilité.



Densité de la loi de Cauchy : cette densité est paire (symétrique par rapport à l'axe des ordonnées).

FIGURE 4.4 – Loi de Cauchy

Pour la densité gaussienne, il est impossible de calculer une primitive de p_{m,σ^2} . Par contre, par passage aux intégrales doubles et changement de coordonnées polaires, il est possible de montrer que

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

Pour l'exemple de la densité de Cauchy, on pourra songer au changement de variables $y = \arctan(x)$.

4.1.3 Notion de fonction de répartition

On associe à une densité de probabilité sa fonction de répartition :

Définition 4.2

Soit p une densité de probabilité. Sa fonction de répartition F est la fonction de \mathbf{R} dans $[0, 1]$ définie par

$$\forall x \in \mathbf{R}, F(x) = \int_{-\infty}^x p(y) dy.$$

C'est la primitive de p qui tend vers 0 en $-\infty$.

Une fonction de répartition possède les propriétés suivantes :

Propriété 4.3

Soit p une densité de probabilité sur \mathbf{R} , continue par morceaux. Alors sa fonction de répartition F a les propriétés suivantes :

1. F est croissante,
2. $F(x) \rightarrow_{x \rightarrow -\infty} 0$, $F(x) \rightarrow_{x \rightarrow +\infty} 1$.
3. F est continue en tout point et dérivable sauf peut-être en un nombre dénombrable de points. En tout point où p est continue, F est dérivable et $F'(x) = p(x)$.

Réciproquement, toute fonction F vérifiant ces propriétés est la fonction de répartition d'une densité.

Preuve

1. Si $x \leq y$ alors $F(y) - F(x) = \int_x^y p(t) dt \geq 0$.
2. $\int_{-\infty}^x p(t) dt \rightarrow 0$ pour $x \rightarrow -\infty$ et $\int_{-\infty}^x p(t) dt \rightarrow \int_{-\infty}^{+\infty} p(t) dt = 1$ pour $x \rightarrow +\infty$, par propriété des intégrales convergentes.
3. Par propriété des primitives.

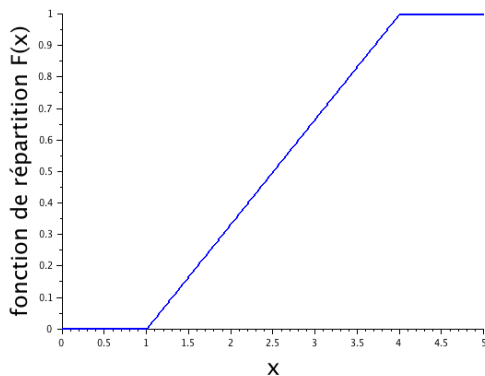
Exemple

Calculons les fonctions de répartition des densités définies précédemment :

- Densité uniforme sur $[a, b]$:

$$\forall x \in \mathbb{R}, F(x) = \begin{cases} 0, & \text{si } x \leq a, \\ \frac{x-a}{b-a}, & \text{si } a < x < b, \\ 1, & \text{si } x \geq b. \end{cases}$$

Elle est continue, mais non dérivable en a et b (points de discontinuité de $p_{a,b}$).



Fonction de répartition de la loi uniforme sur $[1, 4]$

FIGURE 4.5 – Fonction de répartition d'une loi uniforme sur $[1, 4]$

- Densité exponentielle de paramètres $\lambda > 0$:

$$\forall x \in \mathbb{R}, F(x) = \begin{cases} 0, & \text{si } x \leq 0, \\ 1 - e^{-\lambda x}, & \text{si } x > 0. \end{cases}$$

Elle est continue sur \mathbf{R} mais non dérivable en 0 (point de discontinuité de la densité).

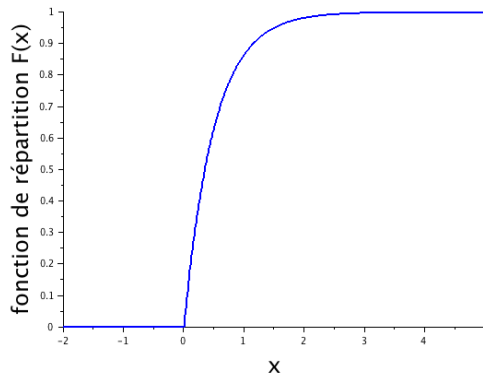
- Densité gaussienne : posant $G(x) := \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ la fonction de répartition de la densité normale centrée réduite, on a

$$\forall x \in \mathbb{R}, F(x) = G\left(\frac{x - m}{\sigma}\right).$$

Les valeurs de G ne sont pas calculables explicitement, mais sont tabulées.

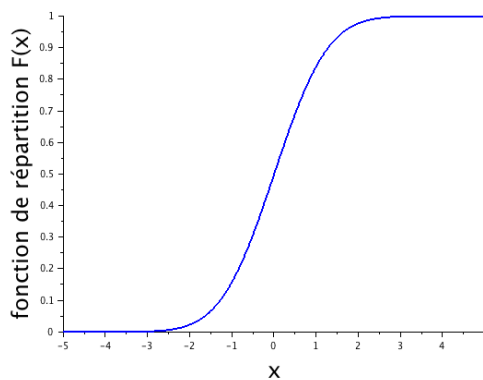
- Densité de Cauchy : sa fonction de répartition est donnée par

$$\forall x \in \mathbb{R}, F(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}.$$



Fonction de répartition de la loi exponentielle de paramètre $\lambda = 2$. Attention : les valeurs prises par F quand x est très grand sont proches de 1 sur le graphe, mais ne sont en fait jamais égales à 1.

FIGURE 4.6 – Fonction de répartition d’une loi exponentielle (2)



Fonction de répartition de la loi gaussienne centrée réduite. Attention : les valeurs prises par F quand x est très grand (resp. très petit) sont proches de 1 (resp. 0) sur le graphe, mais ne sont en fait jamais égales à 1 (resp. 0).

FIGURE 4.7 – Fonction de répartition d’une loi gaussienne $N(0, 1)$

4.2 Probabilités à densité sur \mathbb{R}

Pour des raisons techniques brièvement évoquées au début de ce chapitre, tout sous-ensemble de \mathbf{R} ne peut pas être un événement. On est obligé de se restreindre à une sous-partie stricte de l’ensemble des parties de \mathbf{R} , appelée tribu des boréliens.

Définition 4.4

On appelle tribu des boréliens de \mathbf{R} , notée $\mathcal{B}(\mathbf{R})$, la classe des sous-ensembles de \mathbf{R} comme étant la plus petite classe vérifiant les propriétés suivantes :

1. Elle contient tous les intervalles de la forme $]a, b]$, $-\infty < a < b < +\infty$,
2. Elle est stable par passage au complémentaire, ie $A \in \mathcal{B}(\mathbf{R}) \Rightarrow A^c \in \mathcal{B}(\mathbf{R})$,
3. Elle est stable par union dénombrable : si $(A_n)_{n \geq 1}$ est telle que $A_n \in \mathcal{B}(\mathbf{R})$ pour tout $n \geq 1$ alors $\bigcup_{n \geq 1} A_n \in \mathcal{B}(\mathbf{R})$.

La définition d’une probabilité sur \mathbf{R} est identique à celle introduite dans les premiers chapitres, à ceci près qu’on se restreint aux éléments de $\mathcal{B}(\mathbf{R})$:

Définition 4.5

On appelle « probabilité » sur \mathbf{R} toute application $\mathbb{P} : \mathcal{B}(\mathbf{R}) \mapsto [0, 1]$ telle que

1. $\mathbb{P}(\mathbf{R}) = 1$,

2. Pour toute suite $(A_n)_{n \geq 1}$ d'éléments de $\mathcal{B}(\mathbf{R})$ deux-à-deux disjoints,

$$\mathbb{P}(\cup_{n \geq 1} A_n) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

Remarque

Une remarque importante est que, les axiomes définissant une probabilité sur \mathbf{R} étant identiques aux chapitres précédents, tout ce qui a été défini ou prouvé concernant une probabilité restent valables ici. En particulier, les propriétés usuelles des probabilités, les notions d'indépendance et de probabilité conditionnelle sont inchangées.

Jusqu'ici nous n'avons pas encore démontré l'existence de probabilités sur \mathbf{R} . La notion de densité fournit précisément un moyen de construire des probabilités sur \mathbf{R} :

Propriété 4.6

Pour toute densité de probabilité p sur \mathbf{R} , il existe une unique probabilité \mathbb{P} sur \mathbf{R} vérifiant pour tout $-\infty < a < b < +\infty$,

$$\mathbb{P}([a, b]) = \int_a^b p(x) dx = F(b) - F(a).$$

On dit que \mathbb{P} a pour densité p .

Preuve

Admis.

4.3 Variables aléatoires à densité

Définition 4.7

On appelle variable aléatoire réelle définie sur un espace de probabilité (Ω, \mathbb{P}) toute application $X : \Omega \rightarrow \mathbf{R}$ telle que pour tout $A \in \mathcal{B}(\mathbf{R})$, $X^{-1}(A)$ est un événement de Ω .

Remarque

C'est précisément l'intérêt de se restreindre à la tribu des boréliens pour l'ensemble d'arrivée \mathbf{R} : si on avait considéré toutes les parties de \mathbf{R} , on n'aurait pas pu garantir que l'image réciproque de tout ensemble de \mathbf{R} par une variable aléatoire est un événement de Ω .

Définition 4.8

Soit un espace probabilisé (Ω, \mathbb{P}) et X une variable aléatoire réelle. On appelle loi de X la probabilité \mathbb{P}_X sur $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ définie par

$$\forall A \in \mathcal{B}(\mathbf{R}) \quad \mathbb{P}_X(A) := \mathbb{P}(\{\omega \in \Omega, X(\omega) \in A\})$$

On dit que X a pour loi de probabilité de densité p (on dit que X est une variable continue) si sa loi \mathbb{P}_X est une probabilité sur \mathbf{R} de densité p , c'est-à-dire si

1. p est une densité de probabilité sur \mathbf{R} ,
2. Pour tout $-\infty < a < b < +\infty$, $\mathbb{P}_X([a, b]) = \mathbb{P}(a < X \leq b) = \int_a^b p(x) dx$.

Remarque

En particulier, toutes les propriétés issues de la définition d'une densité de probabilité s'appliquent aux variables aléatoires à densité. En particulier, pour une variable X de densité p , on appellera fonction de répartition de X (notée F_X) la fonction de répartition de sa densité, ie pour tout $x \in \mathbb{R}$

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

Remarque

Si p est une densité de probabilité sur \mathbf{R} , il existe toujours une variable aléatoire X de densité p . En effet, soit $X : \Omega = \mathbf{R} \rightarrow \mathbf{R}$ définie par $X(\omega) = \omega$ et \mathbb{P} probabilité sur \mathbf{R} de densité p . Alors,

$$\begin{aligned} \{X \leq x\} &=]-\infty, x], \\ \{a < X \leq b\} &=]a, b]. \end{aligned}$$

En particulier, $\mathbb{P}(X \leq x) = \mathbb{P}(]-\infty, x]) = \int_{-\infty}^x p(t) dt$ et $\mathbb{P}(a < X \leq b) = \mathbb{P}(]a, b]) = \int_a^b p(x) dx$.

Remarque

On dit qu'une densité de probabilité p est concentrée sur $[a, b]$ si p est nulle sur le complémentaire. Alors X de densité p prend ses valeurs sur $[a, b]$ car $\mathbb{P}(X \in [a, b]) = \int_a^b p(x) dx = 1$.

Réciproquement,

Propriété 4.9

Soit X une variable aléatoire de densité p . Pour tout $-\infty \leq a < b \leq +\infty$, si $\mathbb{P}(X \in]a, b]) = 0$, alors sa densité p est nulle sur $[a, b]$, sauf peut-être en un nombre dénombrable de point de $]a, b[$.

Preuve

$\mathbb{P}(X \in]a, b]) = \int_a^b p(x) dx = 0$. Comme p est une fonction continue par morceaux, positive d'intégrale nulle sur $]a, b]$, elle est identiquement nulle, sauf peut-être en l'ensemble dénombrable des points de discontinuité.

La différence avec les variables discrètes est qu'une variable à densité a une probabilité nulle d'atteindre un point donné a priori :

Propriété 4.10

Si X est une variable aléatoire à densité, alors pour tout $x \in \mathbf{R}$

$$\mathbb{P}(X = x) = 0.$$

Preuve

Soit $n \geq 1$, $\mathbb{P}(x - \frac{1}{n} < X \leq x) = F(x) - F(x - \frac{1}{n})$. Les ensembles $\{x - \frac{1}{n} < X \leq x\}$ forment une suite décroissante et ont pour intersection $\{X = x\}$. Par propriété de continuité monotone décroissante d'une probabilité, on peut passer à la limite dans l'égalité précédente ce qui donne $\mathbb{P}(X = x) = F(x) - \lim_{n \rightarrow \infty} F(x - \frac{1}{n}) = 0$, car F est continue.

Corollaire 4.11

Pour tout $-\infty < a < b < +\infty$, $\mathbb{P}(X \in]a, b[) = \mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in [a, b[) = \mathbb{P}(X \in]a, b])$.

Corollaire 4.12

Soit X variable à densité et A ensemble dénombrable de \mathbf{R} (\mathbf{Q} par exemple) alors $\mathbb{P}(X \in A) = 0$.

Remarque

Deux densités de probabilité p et q qui sont égales, sauf peut-être en un point (ou en un nombre dénombrable de point), définissent la même probabilité \mathbb{P} . Par exemple, les deux fonctions $\lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}$ et $\lambda e^{-\lambda x} \mathbf{1}_{x > 0}$ définissent toutes deux la loi de densité exponentielle de paramètre λ .

Propriété 4.13

Respectivement, si F est une fonction croissante à valeurs dans $[0, 1]$, continue et dérivable sauf peut-être en un nombre fini de points, alors F est la fonction de répartition d'une variable aléatoire X qui admet une densité, donnée par $p(x) = F'(x)$ en tout point où F est dérivable. En dehors de ces points, on peut donner à p une valeur quelconque (0 par exemple).

Preuve

Si F est continue et dérivable en tout point $x \in \mathbf{R}$, alors $p(x) = F'(x)$ vérifie : p positive, continue par morceaux et, $\int_{-\infty}^{+\infty} p(t) dt = \lim_{+\infty} F - \lim_{-\infty} F = 1$ et pour tout $-\infty < a < b < +\infty$, $\int_a^b p(t) dt = F(b) - F(a)$. C'est donc une densité de probabilité.

Si F est non dérivable en un nombre fini de points, on obtient le même résultat en découpant \mathbf{R} en un nombre fini d'intervalles.

Remarque

Si F admet des discontinuités, X n'admet pas de densité.

Exemple

Soit $F(x) = \frac{x^2}{R^2} \mathbf{1}_{[0, R]}(x) + \mathbf{1}_{]R, +\infty[}(x)$. On vérifie que F vérifie les hypothèses du théorème précédent. En particulier, F est dérivable sauf en 0 et R de dérivée

$$p(x) = \begin{cases} 0 & \text{si } x < 0 \text{ et } x > R, \\ \frac{2x}{R^2} & \text{sinon.} \end{cases}$$

En $x = 0$ et $x = R$ on peut prendre ce qu'on veut (0 par exemple). F est la fonction de répartition de la densité p .

Un résultat important est le fait que la fonction de répartition caractérise la loi :

Propriété 4.14

Soient X et Y deux variables aléatoires réelles à densité de même fonction de répartition : $F_X(t) = F_Y(t)$ pour tout $t \in \mathbf{R}$. Alors X et Y ont même loi.

Preuve

C'est la conséquence de la proposition 4.6 : si X et Y ont la même fonction de répartition, $\mathbb{P}_X([a, b]) = \mathbb{P}_Y([a, b])$, pour tout $a < b$. Par unicité dans la proposition 4.6, $\mathbb{P}_X = \mathbb{P}_Y$.

4.4 Notion d'espérance

De la même manière que dans le cas discret, il est possible de définir une notion d'espérance.

Définition 4.15

Soit X une variable réelle de densité p . Soit f une fonction de \mathbf{R} dans \mathbf{R} telle que $x \mapsto f(x)p(x)$ est intégrable sur \mathbf{R} : $\int_{\mathbf{R}} |f(x)|p(x) dx < \infty$. Alors, l'intégrale $\int_{\mathbf{R}} f(x)p(x) dx \in \mathbf{R}$ existe et on note

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

$\mathbb{E}(f(X))$ est appelée espérance de $f(X)$.

Remarque

La définition précédente se généralise à des fonctions à valeurs complexes : si $f : \mathbf{R} \mapsto \mathbf{C}$ est telle que $x \mapsto |f(x)|p(x)$ est intégrable (ce qui équivaut au fait que la fonction partie réelle $x \mapsto \Re(f(x))p(x)$ et la fonction partie imaginaire $x \mapsto \Im(f(x))p(x)$ sont intégrables) alors l'espérance de $f(X)$ existe et est donnée par

$$\mathbb{E}(f(X)) = \int_{\mathbf{R}} \Re(f(x))p(x) dx + i \int_{\mathbf{R}} \Im(f(x))p(x) dx$$

Remarque

Cette définition est l'analogue de $\mathbb{E}(f(X)) = \sum_i f(x_i)\mathbb{P}(X = x_i)$ dans le cas discret. En fait, ces deux constructions sont des cas particuliers d'une même construction plus générale de la notion de probabilité et d'espérance, issue de la théorie de la mesure. Le lecteur ne sera donc pas étonné de retrouver des propositions similaires au chapitre précédent.

Propriété 4.16

Pour tout $a < b$,

$$\mathbb{P}(X \in]a, b]) = \mathbb{E}(\mathbf{1}_{]a, b]}(X)).$$

Preuve

La fonction $g : x \mapsto \mathbf{1}_{]a, b]}(x)p(x)$ est majorée par la densité p qui est intégrable sur \mathbf{R} . Par domination, g est intégrable et

$$\mathbb{E}(\mathbf{1}_{]a, b]}(X)) = \int_{\mathbf{R}} \mathbf{1}_{]a, b]}(x)p(x) dx = \mathbb{P}(X \in]a, b]).$$

Définition 4.17

Soit X une variable aléatoire réelle de densité p . On dit que X admet un moment d'ordre $k \geq 1$, si $x \mapsto |x|^k p(x)$ est intégrable, auquel cas l'intégrale $\int_{\mathbf{R}} x^k p(x) dx$ existe et on définit le k ième moment de X par

$$\mathbb{E}(X^k) = \int_{\mathbf{R}} x^k p(x) dx.$$

Dans le cas où $k = 1$, $\mathbb{E}(X)$ est appelé moyenne de X .

Propriété 4.18

Soit X une variable aléatoire réelle de densité p et f une fonction de \mathbf{R} dans \mathbf{R} .

1. S'il existe une fonction g positive telle que $|f(x)| \leq g(x)$ pour tout $x \in \mathbf{R}$, telle que $\int_{\mathbf{R}} g(x) dx < +\infty$ alors $f(X)$ est intégrable.
2. En particulier, s'il existe une constante $M > 0$ telle que pour tout $x \in \mathbf{R}$, $|f(x)| \leq M$, alors $f(X)$ est intégrable.

Preuve

1. Il s'agit du théorème de domination.
2. C'est l'item précédent appliqué à la fonction constante $g(x) = M$ pour tout x : en effet la fonction positive $x \mapsto Mp(x)$ est intégrable car $\int_{\mathbf{R}} Mp(x) dx = M \cdot 1 = M$.

Propriété 4.19

Si X est une variable aléatoire de densité admettant un moment d'ordre $k \geq 1$. Alors X admet un moment d'ordre $k - 1$.

Preuve

Cela vient de l'inégalité $|X|^{k-1} \leq 1 + |X|^k$ et de la proposition précédente.

Remarque

En particulier, si X admet un moment d'ordre 2, alors X admet un moment d'ordre 1.

Définition 4.20

Si X variable aléatoire à densité admettant un moment d'ordre 2, on définit la variance de X par

$$\begin{aligned} \mathbb{V}(X) &= \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2, \\ &= \int_{\mathbf{R}} (x - (\int_{\mathbf{R}} yp(y) dy))^2 p(x) dx, \\ &= \int_{\mathbf{R}} x^2 p(x) dx - (\int_{\mathbf{R}} xp(x) dx)^2. \end{aligned}$$

Comme dans le chapitre 3, il est facile de vérifier les égalités suivantes :

Propriété 4.21 : Inégalité de Markov

Si X est une variable à densité qui admet un moment d'ordre 1, alors pour tout $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$$

Preuve

Pour tout $a > 0$ et $x \in \mathbf{R}$, l'égalité suivante est vraie : $|x| \geq a \mathbf{1}_{|x| \geq a}$. Par positivité de la densité p et croissance de l'intégration, on en déduit $\int_{\mathbf{R}} |x| p(x) dx \geq a \int_{\mathbf{R}} \mathbf{1}_{|x| \geq a} p(x) dx$, ce qui est exactement : $\mathbb{E}(|X|) \geq a \mathbb{P}(|X| \geq a)$.

Corollaire 4.22

Si X est une variable à densité admettant un moment d'ordre k alors

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|^k)}{a^k}.$$

Corollaire 4.23 : Inégalité de Bienaymé-Tchebychev

Si X admet un moment d'ordre 2, alors

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathbb{V}(X)}{a^2}.$$

4.5 Étude des lois usuelles

4.5.1 Loi uniforme sur $[a, b]$

Propriété 4.24

Si X suit une loi uniforme sur $[a, b]$ alors X admet des moments de tout ordre et

$$\mathbb{E}(X) = \frac{a+b}{2}, \quad \mathbb{V}(X) = \frac{(b-a)^2}{12}.$$

Preuve

La densité de X étant nulle en dehors de $[a, b]$, $|X| \leq \max(|a|, |b|)$ avec probabilité 1 et est donc bornée. X admet donc des moments de tout ordre et $\mathbb{E}(X) = \frac{1}{b-a} \int_a^b x \, dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$. De même, $\mathbb{E}(X^2) = \frac{1}{b-a} \int_a^b x^2 \, dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$. Donc $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{(b-a)^2}{12}$.

4.5.2 Loi exponentielle

Propriété 4.25

Si X suit une loi exponentielle de paramètres $\lambda > 0$ alors X admet des moments de tout ordre et

$$\mathbb{E}(X) = \frac{1}{\lambda}, \quad \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

Preuve

La fonction $x \mapsto |x|^k e^{-\lambda x}$ est continue sur $[0, +\infty[$ donc intégrable sur tout segment $[0, A]$ (où $A > 0$). Noter qu'il n'y a pas de problème d'intégrabilité en 0 car la fonction $x \mapsto |x|^k e^{-\lambda x}$ est continue en ce point. Au voisinage de $+\infty$, cette fonction est dominée par $x \mapsto \frac{1}{x^2}$ par croissance comparée de l'exponentielle et d'un polynôme. Par critère de Riemann, cette fonction est donc intégrable sur $[1, +\infty[$. Donc X admet un moment d'ordre k pour tout $k \geq 1$. Par ailleurs, $\mathbb{E}(X) = \lambda \lim_{A \rightarrow \infty} \left(\int_0^A x e^{-\lambda x} \, dx \right)$. Ainsi pour tout $A > 0$, $\lambda \int_0^A x e^{-\lambda x} \, dx = \lambda \left(\left[x \frac{e^{-\lambda x}}{-\lambda} \right]_0^A + \int_0^A \frac{e^{-\lambda x}}{\lambda} \, dx \right) = \lambda \left(A \frac{e^{-\lambda A}}{-\lambda} + \frac{1}{\lambda^2} (1 - e^{-\lambda A}) \right)$. A la limite pour $A \rightarrow \infty$, on obtient que $\mathbb{E}(X) = \frac{1}{\lambda}$. De même, $\mathbb{E}(X^2) = \lim_{A \rightarrow \infty} \left(\lambda \int_0^A x^2 e^{-\lambda x} \, dx \right)$. Par intégrations par parties, $\lambda \int_0^A x^2 e^{-\lambda x} \, dx = \left[-x^2 e^{-\lambda x} \right]_0^A + 2 \int_0^A x e^{-\lambda x} \, dx = -A^2 e^{-\lambda A} + 2 \int_0^A x e^{-\lambda x} \, dx$. Passant à la limite pour $A \rightarrow \infty$ (toutes quantités convergentes), il vient $\mathbb{E}(X^2) = 2 \int_0^{+\infty} x e^{-\lambda x} \, dx = \frac{2}{\lambda} \mathbb{E}(X) = \frac{2}{\lambda^2}$. Ainsi $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$.

La loi exponentielle satisfait une propriété essentielle :

Propriété 4.26 : Absence de mémoire de la loi exponentielle

Si X suit une loi exponentielle de paramètre $\lambda > 0$, pour tout $s, t > 0$,

$$\mathbb{P}(X > s+t | X > s) = \mathbb{P}(X > t). \quad (4.3)$$

Remarque

La loi exponentielle modélise par exemple la durée de vie d'une ampoule, le temps d'attente à une file d'attente, etc.

Voyant X comme un temps d'attente, l'égalité 4.3 dit exactement la chose suivante : le fait que j'ai déjà attendu un temps s dans la file d'attente n'a aucune influence sur la connaissance du temps supplémentaire que je vais passer ensuite dans la file ; tout se passe comme si j'ai oublié avoir déjà passé un temps s dans la file d'attente. C'est ce qu'on appelle « l'absence de mémoire de la loi exponentielle ».

Preuve

Pour tout $s, t > 0$, on a

$$\begin{aligned}\mathbb{P}(X > s + t | X > s) &= \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > s)}, \\ &= \frac{\int_{s+t}^{+\infty} \lambda e^{-\lambda u} du}{\int_s^{+\infty} \lambda e^{-\lambda u} du}, \\ &= \frac{[-e^{-\lambda u}]_{s+t}^{+\infty}}{[-e^{-\lambda u}]_s^{+\infty}} = e^{-\lambda t} = \mathbb{P}(X > t).\end{aligned}$$

4.5.3 Loi de Cauchy

Loi de Cauchy :

Propriété 4.27

Si X suit une loi de Cauchy, alors X n'admet aucun moment :

$$\mathbb{E}(|X|) = +\infty.$$

Preuve

En effet, $\mathbb{E}(|X|) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} dx = \frac{2}{\pi} \int_0^{+\infty} \frac{x dx}{1+x^2} = \frac{1}{\pi} [\ln(1+x^2)]_0^{+\infty} = +\infty.$

4.5.4 Loi gaussienne**Propriété 4.28**

Si X suit une loi gaussienne de paramètres m, σ^2 alors X admet des moments de tout ordre et

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

Preuve

X admet des moments de tout ordre car $x \mapsto |x|^k e^{-\frac{(x-m)^2}{2\sigma^2}}$ est continue donc localement intégrable et dominée par la fonction $x \mapsto \frac{1}{x^2}$, intégrable au voisinage de $\pm\infty$. Le critère de Riemann permet de conclure.

Par conséquent, $\mathbb{E}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (m+u) \exp\left(-\frac{u^2}{2\sigma^2}\right) du$, par changement de variables $u = x - m$. Ainsi, (noter que les deux termes de la somme sont

convergentes) $\mathbb{E}(X) = m \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{u^2}{2\sigma^2}\right) du + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} u \exp\left(-\frac{u^2}{2\sigma^2}\right) du = m + 0$, par imparité de la fonction intégrée dans le deuxième terme.

Par ailleurs, $\mathbb{V}(X) = \mathbb{E}((X - m)^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x - m)^2 \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx$. Par changement de variables, $u = \frac{x-m}{\sigma}$, cette quantité est égale à $\sigma^2 \int_{-\infty}^{+\infty} u^2 e^{-\frac{u^2}{2}} du = \sigma^2$, par intégration par parties.



Théorèmes limites et estimation

Introduction

La motivation principale de ce dernier chapitre est d'aborder deux théorèmes fondamentaux de la théorie des probabilités qui ont aussi une grande utilité en statistiques :

- La Loi des Grands Nombres,
- Le Théorème Central Limite (ou Théorème de la Limite Centrale).

Ces deux théorèmes traitent de convergence de suites de variables aléatoires ; il faut tout d'abord donner un sens à cette notion de convergence.

5.1 Notions de convergence

Introduction

Il existe en fait plusieurs notions de convergence.

5.1.1 Convergence presque sûre

C'est la notion la plus proche de la notion de convergence simple pour une suite de fonctions (après tout une suite de variables aléatoires n'est jamais qu'une suite de fonctions).

Définition 5.1

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles définies sur (Ω, \mathbb{P}) . On dit que X_n converge presque sûrement vers une variable aléatoire X si

$$\mathbb{P} \left(\left\{ \omega \in \Omega, \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \right\} \right) = 1.$$

Autrement dit, une suite converge presque sûrement si $X_n(\omega)$ converge vers $X(\omega)$ sur un ensemble de ω de probabilité 1. Le « presque sûrement » vient du fait qu'on ne peut pas exclure a priori un ensemble de probabilité nulle où la convergence ci-dessus n'est pas vraie.

Attention

Il y a en fait une difficulté à définir rigoureusement cette notion de convergence presque sûre. Rien ne dit en effet qu'il est possible de prendre la probabilité de l'ensemble " X_n converge vers X ". Il est peu clair que cette partie soit un événement dont on peut calculer la probabilité. Pour définir rigoureusement cette notion, il faudrait introduire la notion de tribu, qui dépasse le cadre de ce cours. Contentons-nous d'une définition heuristique sur ce point.

Exemple

Soit X de loi exponentielle de paramètre $\lambda > 0$. Soit pour tout $n \geq 1$, $X_n = e^{-nX}$. On a $X_n(\omega) \rightarrow 0$ quand $n \rightarrow \infty$ si et seulement si $X(\omega) > 0$. En particulier $\mathbb{P}(X_n \rightarrow 0) = \mathbb{P}(X > 0) = 1$.

Cette notion de convergence est très forte et nécessite de connaître précisément les trajectoires d'une suite de variables aléatoires, ce qui n'est souvent pas le cas.

5.1.2 Convergence en probabilité

Définition 5.2

Soit $(X_n)_{n \geq 1}$ une suite de variables réelles. On dit que X_n converge en probabilité vers X si

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

5.1.3 Convergence en moyenne quadratique

Soit $(X_n)_{n \geq 1}$ une suite de variables réelles telle que X_n est de carré intégrable pour tout $n \geq 1$. Alors on dit que X_n converge en moyenne quadratique vers X si

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

Exemple

Soit X_n qui suit une loi normale centrée de variance $\frac{1}{n}$. Alors X_n converge en moyenne quadratique vers 0 car $\mathbb{E}((X_n - 0)^2) = \mathbb{V}(X_n) = \frac{1}{n} \rightarrow 0$, pour $n \rightarrow +\infty$.

5.1.4 Convergence en loi

Définition 5.3

Soit $(X_n)_{n \geq 1}$ variables aléatoires réelles (discrètes ou à densité). On dit que (X_n) converge en loi vers X (discrète ou à densité) si en tout point $x \in \mathbf{R}$ où la fonction de répartition F_X de X est continue, on a

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

Remarque

Cette notion de point de continuité de F_X ne se pose pas si X admet une densité puisque nous avons vu que dans ce cas, F_X est continue sur \mathbf{R} .

Propriété 5.4

Soit X_n et X des variables discrètes à valeurs dans \mathbb{N} . Alors X_n converge vers X en loi si et seulement si, pour tout $k \geq 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k).$$

Preuve

Nous savons que pour une variable discrète la fonction de répartition F_X est constante par morceaux, qu'elle est continue à droite en tout point, et que ses points de discontinuités sont contenus parmi les valeurs $k \in \mathbb{N}$. Or on a l'égalité suivante, vraie pour tout $k \geq 0$:

$$\mathbb{P}(X_n = k) = \mathbb{P}(X_n \in]k - \frac{1}{2}, k + \frac{1}{2}]) = F_{X_n}(k + \frac{1}{2}) - F_{X_n}(k - \frac{1}{2})$$

et la même égalité est vraie pour X . Or les points $k - \frac{1}{2}$ et $k + \frac{1}{2}$ sont des points de continuité de F_X (elle est constante au voisinage de ces points), donc cette quantité converge vers $F_X(k + \frac{1}{2}) - F_X(k - \frac{1}{2}) = \mathbb{P}(X = k)$. Réciproquement, soit x un point de continuité de F_X , c'est-à-dire un point de $\mathbb{R} \setminus \mathbb{N}$. Si $x < 0$, la convergence $F_{X_n}(x) = 0 \rightarrow_{n \rightarrow \infty} 0 = F_X(x)$ est évidente. Pour $x > 0$, on a l'égalité $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} \mathbb{P}(X = k)$ et une égalité semblable est vraie pour X_n . Par hypothèse, chaque suite $\mathbb{P}(X_n = k)$ converge vers $\mathbb{P}(X = k)$ et il y en a un nombre fini d'entre elles (pour $k = 0, \dots, \lfloor x \rfloor$). Donc $F_{X_n}(x)$ converge vers $F_X(x)$.

Une illustration importante de cette propriété est contenue dans la proposition suivante

Propriété 5.5 : Loi binomiale/ loi de Poisson

Si pour tout $n \geq 1$, X_n suit une loi binomiale de paramètres $(n, \lambda/n)$, pour $\lambda > 0$, alors X_n converge en loi vers X qui suit une loi de Poisson de paramètre λ .

Remarque

Ce théorème donne une interprétation de la loi de Poisson : nous savons en effet que la loi binomiale de paramètres n et p modélise le nombre de succès parmi n expériences de Bernoulli de paramètre p indépendantes. Ici le nombre d'expériences tend vers $+\infty$ alors que la probabilité de succès de chaque expérience tend vers 0.

On dit alors que la loi de Poisson est « la loi des événements rares » : elle modélise le nombre asymptotique de succès dans une grande population, chacun de ces succès ayant une probabilité faible de survenir.

Preuve

Il s'agit de montrer que pour tout $k \in \mathbb{N}$, $\mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k)$, pour $n \rightarrow \infty$. En effet, on a successivement

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

Or, le premier terme $\frac{n(n-1)\dots(n-k+1)}{n^k} = (1 - \frac{1}{n}) \dots (1 - k/n + 1/n)$ converge vers 1 pour $n \rightarrow \infty$ (noter ici qu'il y'a un nombre fini (par rapport à n) de termes). Reste à montrer que $(1 - \frac{\lambda}{n})^{n-k}$ converge vers $e^{-\lambda}$, pour $n \rightarrow \infty$: $(n-k) \ln(1 - \lambda/n) \sim -(n-k) \frac{\lambda}{n} \sim -\lambda$. En particulier, $(n-k) \ln(1 - \lambda/n) \rightarrow -\lambda$, pour $n \rightarrow \infty$. Par continuité de l'exponentielle, $(1 - \frac{\lambda}{n})^{n-k} \rightarrow e^{-\lambda}$ pour $n \rightarrow \infty$. Ce qui veut dire que pour tout $k \in \mathbf{N}$, $\mathbb{P}(X_n = k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$, ce qui prouve le résultat.

5.1.5 Liens entre ces convergences

La convergence presque-sûre (resp. en moyenne quadratique) implique la convergence en probabilité. La convergence en probabilité implique la convergence en loi.

Nous ne prouvons ici que l'implication suivante :

Propriété 5.6

Soit (X_n) une suite de variables de carré intégrable, telle que X_n converge vers X en moyenne quadratique. Alors X_n converge vers X en probabilité.

Preuve

Soit $\varepsilon > 0$. Prouvons que $\mathbb{P}(|X_n - X| \geq \varepsilon)$ converge vers 0 pour $n \rightarrow \infty$. En effet, $\mathbb{P}(|X_n - X| \geq \varepsilon) = \mathbb{P}(|X_n - X|^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}(|X_n - X|^2)}{\varepsilon^2}$, par inégalité de Markov. Or par hypothèse, $\mathbb{E}(|X_n - X|^2) \rightarrow 0$, pour $n \rightarrow \infty$, ce qui prouve le résultat.

5.2 Loi des Grands Nombres

C'est un des résultats les plus importants de la théorie des probabilités. Il légitime l'intuition selon laquelle, lorsqu'on lance un grand nombre de dés équilibrés de façon indépendante, la fréquence d'apparition de résultat 5 (par exemple) doit être proche de $\frac{1}{6}$. Le résultat que nous prouvons ici est une version faible de la loi des grands nombres, vous en verrez une version forte plus tard.

Théorème 5.7

Soit (X_n) suite de variables indépendantes et identiquement distribuées de carré intégrable. Soit $Z_n = \frac{1}{n} (X_1 + \dots + X_n)$, $n \geq 1$, la suite des moyennes empiriques de (X_n) . Alors pour $n \rightarrow \infty$, la suite (Z_n) converge en moyenne quadratique (et donc en probabilité) vers la moyenne commune des (X_n) , $\mathbb{E}(X_1)$. Autrement dit,

$$\mathbb{E} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right|^2 \right) \rightarrow 0, \text{ pour } n \rightarrow \infty.$$

Remarque

Une version forte de ce théorème existe : elle énonce que sous des hypothèses de moment d'ordre 1, la convergence ci-dessus est presque-sûre.

Preuve

Posons $Y_i = X_i - \mathbb{E}(X_1)$. Alors les variables Y_i sont indépendantes et identiquement distribuées et centrées $\mathbb{E}(Y_i) = 0$. De plus

$$\begin{aligned}\mathbb{E} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right|^2 \right) &= \mathbb{E} \left(\left| \frac{1}{n} \sum_{i=1}^n Y_i \right|^2 \right) = \mathbb{E} \left(\left| \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbb{E}(Y_i)) \right|^2 \right), \\ &= \mathbb{V} \left(\frac{1}{n} \sum_{i=1}^n Y_i \right) = \frac{1}{n^2} \mathbb{V} \left(\sum_{i=1}^n Y_i \right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(Y_i),\end{aligned}$$

car la variance d'une somme de variables indépendantes est la somme des variances. Par ailleurs, comme les variables sont identiquement distribuées, cette dernière quantité est égale à $\frac{1}{n} \mathbb{V}(Y_1) = \frac{1}{n} \mathbb{V}(X_1) \rightarrow 0$, pour $n \rightarrow \infty$.

5.3 Théorème Central Limite

La loi des grands nombres énonce que la moyenne empirique des X_i converge vers la moyenne $\mathbb{E}(X)$, c'est-à-dire que $U_n = \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X)$ tend vers 0. Comme vous l'avez vu en cours d'analyse, à partir du moment où une suite tend vers 0, une question naturelle est alors de se demander **à quelle vitesse** cette suite tend-elle vers 0. Autrement dit, il s'agit de trouver le terme suivant dans le développement asymptotique de U_n . L'idée de théorème central limite est de dire que si on multiplie U_n par \sqrt{n} , alors $\sqrt{n}U_n$ converge vers quelque chose de non-trivial. Cela signifie donc que la vitesse de convergence de la moyenne empirique vers la moyenne est en $\frac{1}{\sqrt{n}}$.

Théorème 5.8

Si (X_n) est une suite de variables aléatoires **indépendantes et identiquement distribuées de carré intégrable** (c'est-à-dire que $\mathbb{E}(X_1^2) < +\infty$) et si $\sigma = \sqrt{\mathbb{V}(X)}$, alors la suite

$$\frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right)$$

converge **en loi** quand $n \rightarrow \infty$ vers Z de loi gaussienne, centrée réduite : $Z \sim \mathcal{N}(0, 1)$.

Dans le cas de variables de Bernoulli de paramètre p , le théorème central limite prend le nom du théorème de de-Moivre-Laplace :

Propriété 5.9

Si (X_n) est une suite de variables i.i.d. de Bernoulli de paramètre p alors

$$\frac{X_1 + \dots + X_n - np}{\sqrt{n(p(1-p))}}$$

converge en loi vers Z de loi gaussienne centrée réduite, $\mathcal{N}(0, 1)$.

5.4 Application du TCL : notion d'intervalle de confiance

La motivation de ce paragraphe vient d'une problématique de statistique. Supposons qu'on effectue n mesures différentes d'un même phénomène (température d'une pièce à n différents instants, sondage d'opinion parmi une population de taille n , mesure des tailles de n étudiants d'un groupe de Td, etc.). On fait l'hypothèse que toutes ces observations correspondent à la réalisation de n variables aléatoires X_1, \dots, X_n dont on suppose qu'elles sont indépendantes

et identiquement distribuées, de moyenne commune $m \in \mathbb{R}$ et de variance σ^2 (on suppose l'existence d'un moment d'ordre 2 pour ces variables aléatoires).

Le problème statistique est d'estimer à partir de ces observations le paramètre inconnu m , c'est-à-dire de construire à partir de X_1, \dots, X_n , une approximation (on dit aussi **un estimateur**) de la moyenne inconnue m . La question est de savoir si l'approximation proposée est correcte (en un sens à préciser) et d'estimer l'erreur d'approximation.

La Loi des Grands Nombres et le Théorème Central Limite permettent de répondre à ces questions.

Un estimateur raisonnable de m est fourni par la moyenne empirique $\frac{1}{n}(X_1 + \dots + X_n)$: la Loi des Grands Nombres dit précisément que la moyenne empirique converge presque-sûrement vers le paramètre inconnu m , quand la taille de l'échantillon tend vers $+\infty$.

Pour une taille d'échantillon finie, nous commettons une erreur en approchant m par la moyenne empirique. Le TCL fournit une estimation de l'erreur commise.

Rappelons que la notion de convergence en loi revient à la convergence des fonctions de répartitions aux points de continuité de la fonction de répartition limite. Alors dans le cadre du TCL, la limite est gaussienne, de fonction de répartition toujours continue. Autrement dit, en utilisant le fait que $F_X(b) - F_X(a) = \mathbb{P}(X \in]a, b])$, une autre version du TCL est la suivante :

Propriété 5.10

Sous les hypothèses du TCL, pour tout $a < b$,

$$\mathbb{P}\left(a \leq \frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right) \leq b\right) \rightarrow \mathbb{P}(a \leq Z \leq b).$$

Choisissons maintenant $b = -a = 1.96$. Alors $\mathbb{P}(a \leq Z \leq b) \approx 0.95$. Le théorème précédent dit alors que pour n grand, la probabilité que m se trouve dans l'intervalle $I_n := \left[\frac{1}{n} \sum_{i=1}^n X_i - \frac{1.96\sigma}{\sqrt{n}}, \frac{1}{n} \sum_{i=1}^n X_i + \frac{1.96\sigma}{\sqrt{n}} \right]$ est (environ) 0.95. On dit que l'intervalle I_n est un « intervalle de confiance » pour le paramètre m à 95%.

Attention

L'intervalle de confiance précédent n'a un intérêt que si la variance σ^2 est connue. Si ce n'est pas le cas, une possibilité est de remplacer σ^2 par un estimateur de σ^2 lui-même calculé à partir de l'échantillon (X_1, \dots, X_n) , par exemple la variance empirique :

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n \left(X_i - \frac{1}{n} \sum_{k=1}^n X_k \right)^2.$$