

Notes de cours de
MAT 1700 – PROBABILITÉS 1

Hélène Guérin
Université du Québec à Montréal (UQAM)
heleneguerin.github.io

Ces notes de cours ont été construites notamment à partir du livre de Sheldon Ross *Initiation aux probabilités* et des notes de cours de J.F Renaud, Professeur à l'UQAM.

Table des matières

1	Éléments de combinatoire	5
1.1	Principe fondamental de dénombrement	5
1.2	Permutations	6
a)	Objets discernables (différents)	6
b)	Objets partiellement discernables	6
1.3	Combinaisons et arrangements	7
2	Axiomes de probabilités	11
2.1	Expérience aléatoire	11
a)	Ensemble fondamental d'une expérience aléatoire	11
b)	Évènement d'une expérience aléatoire	12
c)	Opérations sur les événements	12
2.2	Bases axiomatiques des probabilités	14
2.3	Ensembles fondamentaux à événements élémentaires équiprobables	16
2.4	Passage à la limite	19
3	Probabilités conditionnelles et indépendance	21
3.1	Probabilités conditionnelles	21
3.2	Formule de Bayes	24
3.3	Indépendance	26
a)	Indépendance entre plusieurs événements	27
b)	Expériences indépendantes	28
4	Variables aléatoires	31
4.1	Variables aléatoires et fonctions de répartition	31
a)	Fonction de répartition	33
b)	Calcul de probabilités et fonction de répartition	34
4.2	Variables aléatoires discrètes	35
a)	Fonction de masse	35
b)	Lien entre fonction de masse et fonction de répartition	36
c)	Lois discrètes usuelles	37
4.3	Variables aléatoires continues	41
a)	Propriétés	41
b)	Loi d'une v.a. transformée	43
c)	Lois à densité usuelles	44

5	Espérance d'une variable aléatoire	51
5.1	Définition de l'espérance	51
5.2	Espérance d'une transformation, Variance	54
a)	Cas général	54
b)	Variance	55
c)	Propriétés de l'espérance et de la variance	55
5.3	Fonctions génératrices	57
a)	Fonction génératrice des moments	57
b)	Fonction génératrice des probabilités	59
5.4	Formulaire	60
6	Couples de variables aléatoires et indépendance	61
6.1	Loi jointe	61
a)	Cas des variables discrètes	62
b)	Cas des variables à densité	64
6.2	Espérance d'une fonction réelle du couple (X, Y)	66
6.3	Variables aléatoires indépendantes	69
6.4	Loi d'une modification d'un couple de variables	71
a)	Loi de la somme de deux v.a. indépendantes	71
b)	Transformation générales des couples de variables à densité	73
6.5	Loi conditionnelle pour des variables discrètes	75
6.6	Loi conditionnelle pour des variables à densité	77
6.7	Covariance et corrélation	77
7	Théorèmes limites	81
7.1	Inégalités importantes	81
7.2	Différentes notions de convergence	82
7.3	Théorèmes limites importants	84
a)	La loi faible des grands nombres	84
b)	Quelques simulations	85
c)	Le théorème limite centrale	86
7.4	Rappels sur les variables gaussiennes	87

Chapitre 1

Probabilités : Éléments de combinatoire

Sommaire

1.1	Principe fondamental de dénombrement	5
1.2	Permutations	6
a)	Objets discernables (différents)	6
b)	Objets partiellement discernables	6
1.3	Combinaisons et arrangements	7

Dans ce premier chapitre, on va voir différentes façons de dénombrer des objets. Ces techniques vont vous être utiles dans la suite pour calculer la probabilité d'événements dans certains cas.

1.1 Principe fondamental de dénombrement

Supposons que vous souhaitez vous offrir une crème glacée. Le marchand propose 6 parfums de glace que l'on peut soit prendre dans un pot, soit dans un cône, soit dans un cône recouvert de chocolat. Supposons que vous êtes gourmands et que vous voulez essayer toutes les combinaisons possibles, mais en ne choisissant qu'un parfum à la fois. Combien de fois devrez-vous vous rendre chez le marchand de glace pour toutes les tester ?

Théorème 1.1. *Considérons deux expériences dont la première a m résultats possibles et la seconde n résultats, alors si on considère les deux expériences ensemble il y a en tout $m \times n$ résultats possibles.*

Dans le cas de notre marchand de glace, il y a donc $6 \times 3 = 18$ résultats possibles. Ce résultat se généralise à un plus grand nombre d'expériences. Notre marchand de glace peut aussi proposer d'ajouter des bonbons sur votre boule de glace que vous devez choisir parmi 4 possibilités.

Théorème 1.2. *Si on considère K expériences, la première ayant n_1 résultats possibles, la seconde n_2 , ainsi de suite jusqu'à la dernière qui a n_K résultats possibles, alors il y a un total de $n_1 \times n_2 \times \dots \times n_K$ résultats possibles pour les K expériences prises ensembles.*

Exemple 1.1. Combien de plaques minéralogiques portant un matricule de 7 caractères peut-on former si les trois premiers sont des lettres et les 4 derniers sont des chiffres ?

Réponse : $26^3 \times 10^4 = 175\,760\,000$. Si jamais on exclut que les lettres et les chiffres se répètent, on a alors $(26 \times 25 \times 24) \times (10 \times 9 \times 8 \times 7) = 78\,624\,000$ plaques possibles.

1.2 Permutations

a) Objets discernables (différents)

On souhaite ordonner de différentes façons des objets discernables (différents). Par exemple, combien de *mots* de trois lettres peut-on faire en utilisant chacune des lettres A , B et C . Il y a exactement 6 mots possibles : ABC , ACB , BAC , BCA , CAB , CBA . En effet, il y a trois choix possibles pour la première lettre, puis 2 choix pour la seconde et 1 seul choix pour la dernière, soit $3 \times 2 \times 1 = 6$ possibilités. Chaque arrangement est appelé **permutation**.

Théorème 1.3. *Le nombre de permutations de n objets discernables est $n! = n \times (n - 1) \times (n - 2) \dots \times 3 \times 2 \times 1$.*

Par convention $0! = 1$.

Remarque 1.4. *On a*

$$1! = 1, \quad 2! = 2, \quad 3! = 6, \quad 4! = 24, \quad 5! = 120, \quad 6! = 720, \quad 7! = 5040, \quad 8! = 40\,320 \dots$$

On remarque que la valeur des factoriels croît très rapidement.

Exemple 1.2. Vous souhaitez ranger votre bibliothèque. Vous possédez 4 livres de mathématiques, 2 romans en anglais et 5 romans en français. Il y a par conséquent $(4 + 2 + 5)! = 11! = 39\,916\,800$ façons de ranger ces livres.

Maintenant, vous souhaitez ranger les livres de façon à ce que les livres traitant d'un même sujet soit regroupés. Il y a 3 catégories et donc $3! = 6$ façons de choisir l'ordre dans lequel les catégories vont apparaître dans votre bibliothèque. Il y a ensuite $4!$ façon d'ordonner les livres de mathématiques, $2!$ façons de ranger les romans anglophones et $5!$ façons de ranger les romans francophones. Au total, vous avez $6 \times 4! \times 2! \times 5! = 34\,560$ possibilités.

b) Objets partiellement discernables

Nous avons calculé le nombre de permutations possibles de n éléments différents, maintenant on pourrait considérer des ensembles d'objets dont certains sont discernables et d'autres non. Par exemple, on considère un sac de billes contenant 3 billes Rouges, 2 billes Bleues. On tire une bille du sac et on la pose sur la table, puis on tire une autre bille qu'on pose à son tour sur la table. Combien de tirage de billes différents existe-t-il ?

Si toutes les billes étaient distinctes, il y aurait $(3 + 2)! = 5! = 120$ tirage possibles. Les billes rouges étant identiques, tirer trois billes rouges de suite ne représente qu'un seul tirage : $R_1 R_2 R_3 = R_2 R_1 R_3 = R_1 R_3 R_2 = \dots$. En fait, permuter les billes rouges ne change pas le résultat et il y a $3!$ permutations possibles des billes rouges. On fait le même raisonnement avec les billes bleues. Il y a par conséquent $\frac{5!}{3!2!} = 10$ résultats possibles.

Théorème 1.5. *Considérons n objets classés en K groupes, les objets de chaque groupe étant identiques. Il y a n_1 objets dans le premier groupe, n_2 dans le second, ainsi de suite jusqu'à n_K dans le $K^{\text{ème}}$ groupe, avec $n_1 + n_2 + \dots + n_K = n$.*

Il y a alors $\frac{n!}{n_1! n_2! \dots n_K!}$ permutations différentes de ces objets.

Exemple 1.3. Combien d'anagrammes peut-on faire avec le mot TORONTO ?

Réponse : Il y a $n = 7$ lettres à ordonner mais avec seulement $K = 4$ lettres qui sont différentes : le T est répété 2 fois et le O est répété 3 fois. Par conséquent, il y a $\frac{7!}{2!3!} = 420$ anagrammes possibles du mot TORONTO.

1.3 Combinaisons et arrangements

Considérons maintenant 5 lettres différentes : A, B, C, D, E et on souhaite former des mots de longueur 3 à partir de ces lettres sans répétition de lettres.

Si on tient compte de l'ordre (le mot ABC n'est pas le même que ACB), on a par conséquent 5 façons de choisir la première lettre, 4 de choisir la seconde et enfin 3 façons de choisir la dernière : soit $5 \times 4 \times 3 = 60$. Si on ne tient pas compte de l'ordre d'apparition des lettres (le mot ABC est le même que le mot ACB). Pour chaque choix de 3 lettres il y a $6 = 3!$ permutations des lettres qui représentent le même mot ($ABC = BAC = ACB = \dots$). Par conséquent, si on ne tient pas compte de l'ordre, il y a $\frac{5 \times 4 \times 3}{3!} = 10$ résultats possibles.

Théorème 1.6. *Le nombre de façons de choisir k objets parmi n objets distincts, sans répétition, en tenant compte de l'ordre est*

$$n(n-1) \dots (n-k+1) = \frac{n!}{(n-k)!}.$$

Chaque choix de ces k objets parmi n est appelé **arrangement** de k parmi n .

Le nombre de façon de choisir k objets parmi n sans répétition **sans tenir compte de l'ordre** est

$$\binom{n}{k} = \frac{n(n-1) \dots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.$$

Chaque choix de ces k objets parmi n est appelé **combinaison** de k parmi n . Le coefficient $\binom{n}{k}$ est appelé **coefficient binomial**.

Remarque 1.7. Soit $n \geq 1$ et $k \in \{0, \dots, n\}$, on a

$$\binom{n}{0} = 1, \quad \binom{n}{n} = 1, \quad \binom{n}{1} = n, \quad \binom{n}{n-1} = n, \quad \binom{n}{k} = \binom{n}{n-k}.$$

Exemple 1.4. Sur une classe de 20 étudiants, on souhaite créer un comité de 4 représentants. Il y a par conséquent

$$\binom{20}{4} = \frac{20 \times 19 \times 18 \times 17}{4 \times 3 \times 2 \times 1} = 4845$$

comités possibles. Chaque comité est une combinaison de 4 personnes parmi 20.

Si dans ce comité, on souhaite désigner un président, un vice-président, un secrétaire et un trésorier, alors l'ordre dans le lequel on choisit les individus est important et donc le nombre de comité est $20 \times 19 \times 18 \times 17 = 116\,280$. Chaque comité est un arrangement de 4 personnes parmi 20.

Exemple 1.5. On joue 10 fois de façon successive à pile ou face. Combien de façon peut-on avoir 3 fois « Pile » et 7 fois « Face » ?

Réponse : Il faut alors choisir les 3 lancers qui donneront « Pile » parmi les 10 lancers (les autres donneront automatiquement « Face »), Il y a $\binom{10}{3} = 120$ possibilités de position les lancers qui donneront « Pile ». On aurait pu choisir 7 lancers qui donneront « Face ». Le calcul donne le même résultat car $\binom{10}{3} = \binom{10}{7}$.

Proposition 1.8 (Relation de Pascal). *Pour tout $1 \leq k \leq n-1$, on a*

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Démonstration. Identifions l'objet 1. Quand on choisit k objets parmi n objets, il y a exactement deux cas :

- soit on a pris l'objet 1 dans notre choix, et donc il reste à choisir $k - 1$ objets parmi $n - 1$, il y a $\binom{n-1}{k-1}$ possibilités ;
- soit on n'a pas pris l'objet 1 dans notre choix, et donc il faut choisir k objets parmi $n - 1$, il y a $\binom{n-1}{k}$ possibilités.

Donc $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$. □

On peut ainsi calculer successivement la valeur du coefficient du binôme

$n \backslash k$	0	1	2	3	4	5	...
1	1	1					
2	1	2	1				
3	1	3	3	1			
4	1	4	6	4	1		
5	1	5	10	10	5	1	
⋮							

Théorème 1.9. (Théorème du binôme de Newton) *Pour tout $x, y \in \mathbb{R}$, $n \geq 1$, on a*

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

Démonstration. On peut le montrer par induction : pour $n = 1$, on a bien

$$\sum_{k=0}^1 \binom{1}{k} x^k y^{1-k} = y + x.$$

Supposons que l'égalité soit vraie pour n , regardons ce qui se passe pour $n + 1$:

$$\begin{aligned}
 (x + y)^{n+1} &= (x + y) \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} = \sum_{k=0}^n \binom{n}{k} x^{k+1} y^{n-k} + \sum_{k=0}^n \binom{n}{k} x^k y^{n-k+1} \\
 &= \sum_{j=1}^{n+1} \binom{n}{j-1} x^j y^{n-j+1} + \sum_{k=0}^n \binom{n}{k} x^k y^{n-k+1} \\
 &\quad \text{(décalage d'indice } j = k + 1) \\
 &= \sum_{i=1}^{n+1} \binom{n}{i-1} x^i y^{n-i+1} + \sum_{i=0}^n \binom{n}{i} x^i y^{n-i+1} \\
 &\quad \text{(changement d'indice } j \rightsquigarrow i \text{ et } k \rightsquigarrow i) \\
 &= x^{n+1} + \sum_{i=1}^n \left(\binom{n}{i-1} + \binom{n}{i} \right) x^i y^{n+1-i} + y^{n+1} \\
 &= x^{n+1} + \sum_{i=1}^n \binom{n+1}{i} x^i y^{n+1-i} + y^{n+1} \quad \text{(d'après la proposition 1.8)} \\
 &= \sum_{i=0}^{n+1} \binom{n+1}{i} x^i y^{n+1-i}.
 \end{aligned}$$

La propriété est donc vrai au rang $n + 1$.

Par induction, on vient de montrer que pour tout $n \geq 1$, $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$. □

Ce théorème permet de développer l'expression $(x + y)^n$:

$$\begin{aligned}(x + y)^2 &= x^2 + 2xy + y^2 \\(x + y)^3 &= x^3 + 3x^2y + 3xy^2 + y^3 \\(x + y)^4 &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4 \\&\vdots\end{aligned}$$

Application 1.1. Quel est le nombre de sous-ensembles d'un ensemble à n éléments ?

Le nombre de sous-ensemble de taille k est $\binom{n}{k}$, donc la réponse est

$$\sum_{k=0}^n \binom{n}{k} = 2^n.$$

On peut généraliser le résultat du Théorème du binôme de Newton. Au lieu de diviser en 2 groupes, on peut diviser la population en un plus grand nombre de groupes. Considérons un ensemble à n éléments que l'on divise en ℓ sous-groupes de taille respective $n_1, \dots, n_\ell$, avec $n_1 + \dots + n_\ell = n$. Le nombre de décomposition est

$$\binom{n}{n_1} \binom{n - n_1}{n_2} \dots \binom{n - n_1 - \dots - n_{\ell-1}}{n_\ell} = \frac{n!}{n_1! \dots n_\ell!}.$$

Cette quantité représente le nombre de répartitions de n objets en ℓ groupes distincts de tailles respectives $n_1, n_2, \dots, n_\ell$. La formule du binôme de Newton se généralise

$$(x_1 + x_2 + \dots + x_\ell)^n = \sum_{\substack{0 \leq n_1, n_2, \dots, n_\ell \leq n \\ n_1 + n_2 + \dots + n_\ell = n}} \frac{n!}{n_1! \dots n_\ell!} x_1^{n_1} x_2^{n_2} \dots x_\ell^{n_\ell}.$$

Exemple 1.6. Le nombre de bénévoles à la fête des neiges est de 20 et il faut 3 personnes à la patinoire, 6 clowns itinérants et 5 personnes à la sécurité, 2 à l'accueil et enfin 4 pour la musique. Il y a par conséquent

$$\frac{20!}{3!6!5!2!4!} = 97\,772\,875\,200$$

possibilités de répartir les volontaires dans les différentes tâches.

Exemple 1.7. Au cours d'éducation physique les 12 élèves doivent se répartir en deux équipes de 6 personnes. Le nombre de façon de répartir les élèves en deux groupes distincts est

$$\frac{12!}{6!6!} = 924$$

Les équipes n'étant pas distinctes, mais seulement 2 groupes d'enfants, la solution est par conséquent :

$$924/2! = 462$$

Chapitre 2

Axiomes de probabilités

Sommaire

2.1	Expérience aléatoire	11
a)	Ensemble fondamental d'une expérience aléatoire	11
b)	Évènement d'une expérience aléatoire	12
c)	Opérations sur les événements	12
2.2	Bases axiomatiques des probabilités	14
2.3	Ensembles fondamentaux à événements élémentaires équiprobables	16
2.4	Passage à la limite	19

2.1 Expérience aléatoire

On appelle expérience aléatoire une expérience dont le résultat n'est pas prévisible. Bien que le résultat n'est pas connu d'avance, on suppose qu'on connaît l'ensemble des résultats possibles. Le but de ce chapitre est d'évaluer les chances d'apparition de chaque résultat, on introduit ainsi la notion de probabilité.

a) Ensemble fondamental d'une expérience aléatoire

L'ensemble fondamental d'une expérience aléatoire est l'ensemble de tous les résultats possibles de l'expérience. Il n'est pas défini de façon unique. Cet ensemble est noté Ω . Il peut être fini, dénombrable, ou infini non dénombrable.

Un résultat de l'expérience aléatoire est souvent noté $\omega : \omega \in \Omega$.

Exemple 2.1. 1. On lance un dé à 6 faces, on s'intéresse au chiffre obtenu. L'ensemble fondamentale est alors fini, plus précisément

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

2. On jette une pièce autant de fois que nécessaire pour obtenir une fois « Face ». L'ensemble fondamentale est alors infini dénombrable. Si on note F pour « Face » et P pour « Pile », on a

$$\Omega = \{F, PF, PPF, PPPF, \dots, PPP \dots PPF, \dots\}.$$

3. On s'intéresse à la durée de vie d'une ampoule. L'ensemble fondamental est alors infini, non dénombrable : $\Omega = [0, +\infty[$.

4. On considère un assuré et on s'intéresse à son risque (i.e. à l'ensemble de ses sinistres) sur une année. Si on s'intéresse au nombre de sinistres au cours d'une année, alors $\Omega = \mathbb{N}$ (avec $\mathbb{N} = \{0, 1, 2, 3, \dots\}$). Si on s'intéresse au coût du premier sinistre, alors $\Omega =]0, +\infty]$. Si on s'intéresse aux instants des sinistres, alors $\Omega = \{(t_1, t_2, \dots, t_n) : t_i > 0, n \in \mathbb{N}^*\} \cup \{\oplus\}$, avec t_i la date du $i^{\text{ème}}$ sinistre et avec la convention que $\omega = \oplus$ s'il n'y a pas eu de sinistre.

b) Évènement d'une expérience aléatoire

Au lieu de s'intéresser à toute l'expérience, on ne s'intéresse qu'à certains résultats. Un **évènement** E est une combinaison de résultats possibles. C'est un sous-ensemble de l'ensemble fondamental, $E \subset \Omega$. Si le résultat de l'expérience est contenu dans E , on dit que l'évènement E est **réalisé**.

Exemple 2.2. 1. On lance deux dés distincts à 6 faces. L'ensemble fondamental est

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (2, 1), (2, 2), \dots, (6, 5), (6, 6)\}.$$

On remarque que $\text{card } \Omega = 6 \times 6 = 36$. On considère les événements suivants :

- la somme des points est 6 : $E = \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\}$,
- la somme des points est un multiple de 3 :
 $F = \{(1, 2), (2, 1), (1, 5), (5, 1), (2, 4), (4, 2), (3, 3), (3, 6), (6, 3), (4, 5), (5, 4), (6, 6)\}.$

2. On lance deux dés identiques à 6 faces. L'ensemble fondamental est

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (1, 6), (2, 2), (2, 3), \dots, (2, 6), \dots, (5, 6), (6, 6)\}.$$

Dans ce cas $\text{card } \Omega = 21$ et

- la somme des points est 6 : $E = \{(1, 5), (2, 4), (3, 3)\}$,
- la somme des points est un multiple de 3 :
 $F = \{(1, 2), (1, 5), (2, 4), (3, 3), (3, 6), (4, 5), (6, 6)\}.$

Toute expérience aléatoire comprend un **évènement certain** et un **évènement impossible**. L'évènement impossible, noté $\emptyset$, est le sous-ensemble de Ω qui ne contient pas d'éléments. L'évènement certain est le sous-ensemble qui contient tous les éléments, c'est l'ensemble fondamental Ω lui-même.

On appelle **évènement élémentaire** un évènement qui ne contient qu'un seul élément.

Exemple 2.3. On lance deux dés à 6 faces.

Si E_1 est l'évènement « la somme des points est supérieure à 13 », alors $E_1 = \emptyset$.

Si E_2 est l'évènement « la somme des points est inférieure à 15 », alors $E_2 = \{(1, 1), (1, 2), \dots, (5, 6), (6, 6)\} = \Omega$.

Si E_3 est l'évènement « la somme des points vaut 2 », alors $E_3 = \{(1, 1)\}$ est un évènement élémentaire.

c) Opérations sur les évènements

Soient E et F deux évènements.

$\leadsto$ faire des dessins.

La relation « l'évènement E implique l'évènement F » signifie que si E se réalise alors F se réalise aussi. On écrit alors $E \subset F$ (inclusion au sens large).

Si $E \subset F$ et $F \subset E$ alors $E = F$.

- **Négation** est l'évènement contraire de E , ou complémentaire de E par rapport à Ω . Il est noté : E^c (ou $\overline{E}$). E^c contient tous les éléments de Ω qui ne sont pas dans E , $E^c = \Omega \setminus E = \{\omega \in \Omega : \omega \notin E\}$.

E^c est l'évènement qui se réalise lorsque E ne se réalise pas.

- **Intersection** : L'évènement $G = E \cap F$ est réalisé lorsque E et F sont réalisés simultanément.

On a $\omega \in E \cap F$ si et seulement si $\omega \in E$ et $\omega \in F$.

- **Union** : L'évènement $H = E \cup F$ est réalisé lorsque l'un des événements E ou F est réalisé.

On a $\omega \in E \cup F$ si et seulement si $\omega \in E$ ou $\omega \in F$.

- **Différence** : Si $E \subset F$, on définit $F \setminus E = \{\omega \in F : \omega \notin E\} = F \cap E^c$.

On a $\omega \in F \setminus E$ si et seulement si $\omega \in F$ et $\omega \notin E$.

Deux événements E et F sont dits **incompatibles** ou **disjoints** si $E \cap F = \emptyset$.

Exemple 2.4. On jette deux dés distincts à 6 faces. On considère les événements suivants : E = « la somme des points vaut 6 », F = « la somme des points est un multiple de 3 », G = « avoir au moins un 6 ». On a $E \subset F$, $E \cap G = \emptyset$, $F \cap G = \{(3, 6), (6, 3), (6, 6)\}$ et G^c = « n'avoir aucun 6 ». On remarque que

$$\begin{aligned} E &= \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\} \\ F &= \{(1, 2), (2, 1), (1, 5), (5, 1), (2, 4), (4, 2), (3, 3), (3, 6), (6, 3), (4, 5), (5, 4), (6, 6)\} \\ G &= \{(1, 6), (6, 1), (2, 6), (6, 2), (3, 6), (6, 3), (4, 6), (6, 4), (5, 6), (6, 5), (6, 6)\} \end{aligned}$$

Propriété 2.1. :

— *Commutativité* : $E \cup F = F \cup E$ et $E \cap F = F \cap E$.

— *Associativité* : $(E \cup F) \cup G = E \cup (F \cup G)$ et $(E \cap F) \cap G = E \cap (F \cap G)$.

— *Distributivité* : $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ et $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$

— *Complétion* : $(E \cup F)^c = E^c \cap F^c$ et $(E \cap F)^c = E^c \cup F^c$.

De manière plus générale, $(\cup_{i=1}^n E_i)^c = \cap_{i=1}^n (E_i^c)$ et $(\cap_{i=1}^n E_i)^c = \cup_{i=1}^n (E_i^c)$

Extension des définitions :

On considère une suite infinie d'évènements $E_1, E_2, \dots$

— L'union $\cup_{i=1}^{\infty} E_i$ est l'évènement qui regroupe tous les points des événements $E_1, E_2, \dots$: on a « $\omega \in \cup_{i=1}^{\infty} E_i$ » si et seulement si « il existe un entier $i \geq 1$ tel que $\omega \in E_i$ ».

— L'intersection $\cap_{i=1}^{\infty} E_i$ est l'évènement composé des points communs à tous les événements $E_1, E_2, \dots$: on a « $\omega \in \cap_{i=1}^{\infty} E_i$ » si et seulement si « pour tous les entiers $i \geq 1$ on a $\omega \in E_i$ ».

Exemple 2.5. On lance un dé une infinité de fois. Alors l'évènement E = « avoir au moins un 6 », signifie qu'il existe un lancer qui donnera un 6. Par conséquent $\omega \in E$ est équivalent à « il existe $i \geq 1$ tel que le lancer i donne un 6 ».

Si on note E_i = « avoir un 6 au $i^{\text{ème}}$ lancer ». On a alors $E = \cup_{i=1}^{\infty} E_i$.

2.2 Bases axiomatiques des probabilités

Considérons une expérience dont l'ensemble fondamental est Ω .

Intuitivement, les chances d'apparition d'un événement est la fréquence d'apparition de cet événement lorsqu'on réalise une infinité de fois dans les mêmes conditions une expérience. On va définir maintenant de façon formelle ce qu'est une probabilité.

Définition 2.2. On admet que pour chaque événement E de Ω il existe une valeur $\mathbb{P}(E)$ appelée **probabilité de l'événement** E qui vérifie les axiomes suivants :

(A1) $0 \leq \mathbb{P}(E) \leq 1$;

(A2) $\mathbb{P}(\Omega) = 1$;

(A3) pour toute suite infinie $E_1, E_2, \dots$ d'événements disjoints (i.e., pour tout $i \neq j$, $E_i \cap E_j = \emptyset$), on a

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$$

$\mathbb{P}$ est appelé **probabilité** (ou mesure de probabilité).

Propriété 2.3 (Conséquences immédiates). On a

1. $\mathbb{P}(\emptyset) = 0$.
2. Si $E_1, E_2, \dots, E_n$ est une suite finie d'événements disjoints, on a

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n \mathbb{P}(E_i).$$

3. Pour tout événement E , $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
4. Si $E \subset F$, alors $\mathbb{P}(E) \leq \mathbb{P}(F)$.
5. Si $E \subset F$, on note $F \setminus E = F \cap E^c$. On a $\mathbb{P}(F \setminus E) = \mathbb{P}(F) - \mathbb{P}(E)$.

Preuve.

1. On considère la suite d'événements disjoints $E_1 = \Omega$ et $E_i = \emptyset$ pour $i \geq 2$. On a $\bigcup_{i=1}^{\infty} E_i = \Omega$. Par conséquent, d'après (A2) et (A3), on a

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i) = 1 + \sum_{i=2}^{\infty} \mathbb{P}(\emptyset).$$

Par conséquent, pour avoir $\sum_{i=2}^{\infty} \mathbb{P}(\emptyset) = 0$, on a forcément $\mathbb{P}(\emptyset) = 0$.

2. Il suffit de définir $E_i = \emptyset$ pour $i > n$. On a alors une suite (infinie) $(E_i)_{i \geq 1}$ d'événements disjoints, avec $\mathbb{P}(E_i) = 0$ pour $i > n$. De plus, $\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^n E_i$. Il suffit d'utiliser (A3) et le résultat précédent pour conclure.
3. Il suffit de prendre $E_1 = E$ et $E_2 = E^c$ et les deux résultats précédents pour conclure.
4. Si $E \subset F$, alors $F = E \cup (F \setminus E)$, avec E et $(F \setminus E)$ ensembles disjoints. D'après ce qui précède, on a $\mathbb{P}(F) = \mathbb{P}(E) + \mathbb{P}(F \setminus E)$ avec $\mathbb{P}(F \setminus E) \geq 0$ d'après (A1).

□

Exemple 2.6. Jetons une pièce bien équilibrée (même chance de tomber sur pile et sur face). On a $\Omega = \{P, F\}$. Comme $1 = \mathbb{P}(\Omega) = \mathbb{P}(\{F\}) + \mathbb{P}(\{P\})$, et comme la pièce est bien équilibrée ($\mathbb{P}(\{F\}) = \mathbb{P}(\{P\})$), on déduit que

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

Par contre, si la pièce est biaisée et qu'on a deux fois plus de chance d'avoir pile que face, on a alors

$$\mathbb{P}(\{P\}) = 2/3 \quad \mathbb{P}(\{F\}) = 1/3.$$

Théorème 2.4. Soit E et F deux événements quelconques, alors

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

Conséquence, pour tous événements E et F ,

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

Preuve. On remarque que $E \cup F = (E \setminus E \cap F) \cup (E \cap F) \cup (F \setminus E \cap F)$ (union d'ensembles disjoints). De plus, $E = (E \setminus E \cap F) \cup (E \cap F)$ et $F = (F \setminus E \cap F) \cup (E \cap F)$. Donc

$$\begin{aligned} \mathbb{P}(E) + \mathbb{P}(F) &= \mathbb{P}(E \setminus E \cap F) + \mathbb{P}(E \cap F) + \mathbb{P}(F \setminus E \cap F) + \mathbb{P}(E \cap F) \\ &= \mathbb{P}(E \cup F) + \mathbb{P}(E \cap F). \end{aligned}$$

□

Le théorème précédent se généralise à plus que 2 événements (Théorème 2.7, chapitre 2 du livre de Ross).

Proposition 2.5. On considère des événements $E_1, \dots, E_n$. On a alors

$$\begin{aligned} \mathbb{P}(E_1 \cup E_2 \cup \dots \cup E_n) &= \sum_{i=1}^n \mathbb{P}(E_i) - \sum_{i_1 < i_2} \mathbb{P}(E_{i_1} \cap E_{i_2}) + \sum_{i_1 < i_2 < i_3} \mathbb{P}(E_{i_1} \cap E_{i_2} \cap E_{i_3}) - \dots \\ &= \sum_{k=1}^n (-1)^{k+1} \left(\sum_{j_1 < \dots < j_k} \mathbb{P}(E_{j_1} \cap \dots \cap E_{j_k}) \right). \end{aligned}$$

Pour tous événements $E_1, \dots, E_n$, on a

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) \leq \sum_{i=1}^n \mathbb{P}(E_i).$$

Cas particulier d'un ensemble fondamental discret (fini ou dénombrable)

Considérons une expérience aléatoire dont l'ensemble fondamental Ω contient N éléments : $\Omega = \{\omega_1, \dots, \omega_N\}$.

On associe à chaque élément ω_i une probabilité $p_i := \mathbb{P}(\{\omega_i\})$.

La probabilité $\mathbb{P}$ est une fonction à valeurs dans $[0, 1]$ qui vérifie alors

- $\forall i \in 1, \dots, N, p_i \in [0, 1],$
- $\mathbb{P}(\Omega) = \sum_{i=1}^N \mathbb{P}(\{\omega_i\}) = p_1 + p_2 + \dots + p_N = 1.$

On peut étendre facilement ceci au cas où $\Omega = \{\omega_i\}_{i \geq 1}$ est dénombrable.

Exemple 2.7. On jette un dé à 6 faces : $\Omega = \{1, 2, \dots, 6\}$. On définit la probabilité de tomber sur une face du dé. On suppose que le dé est bien équilibré (non truqué), c'est à dire que chaque face a la même chance d'apparaître. On a donc $p_i = \mathbb{P}(\{i\}) = 1/6$ pour $i \in \{1, 2, \dots, 6\}$.

Soit E l'événement "obtenir un chiffre pair", on a alors

$$\mathbb{P}(E) = \mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 1/2.$$

Soit F l'événement "obtenir une valeur supérieure ou égale à deux". On remarque que F^c = "tomber sur 1", d'où

$$\mathbb{P}(F) = 1 - \mathbb{P}(F^c) = 1 - 1/6 = 5/6$$

Exemple 2.8. On jette un dé à 6 faces biaisé tel que on a deux fois plus de chances de tomber sur un nombre pair que sur un nombre impair. Tous les chiffres paires ont la même probabilité d'apparaître, et tous les chiffres impaires ont la même probabilité d'apparaître. Alors

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{5\}) = \frac{1}{9} \quad \text{et} \quad \mathbb{P}(\{2\}) = \mathbb{P}(\{4\}) = \mathbb{P}(\{6\}) = \frac{2}{9}.$$

2.3 Ensembles fondamentaux à événements élémentaires équiprobables

Considérons une expérience d'espace fondamental **fini** $\Omega = \{\omega_1, \dots, \omega_N\}$, de cardinal N , dont chaque événement élémentaire a la même chance d'apparaître. On dit que l'expérience est **équiprobable**.

Par conséquent, on a

$$\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\}) = \dots = \mathbb{P}(\{\omega_N\}) = \frac{1}{N}$$

Dans le cas d'une **expérience équiprobable**, pour tout événement E on a

$$\mathbb{P}(E) = \frac{\text{card}(E)}{N} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Attention ! Cette formule n'est vrai que dans le cas équiprobable.

Exemple 2.9. On lance un dé et on souhaite connaître la probabilité de tomber sur un nombre pair. On a $\Omega = \{1, 2, 3, 4, 5, 6\}$. On note E l'événement avoir un chiffre pair. Du coup, $E = \{2, 4, 6\}$.

- Si le dé est bien équilibré (cas d'équiprobabilité, car même chance de tomber sur chacune des faces). Alors

$$\mathbb{P}(E) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{3}{6} = \frac{1}{2}.$$

- Si le dé est biaisé de sorte qu'on a 2 fois plus de chance de tomber sur un chiffre pair que impair.
Alors

$$\mathbb{P}(E) = \mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{2}{9} \times 3 = \frac{2}{3} \neq \frac{1}{2}.$$

Exemple 2.10. On jette deux dés distincts (l'un rouge, l'autre vert) bien équilibrés et on se demande qu'elle est la probabilité que la somme des chiffres soit égale à 7.

Comme les dés sont de couleurs différentes, l'ensemble fondamental est

$$\Omega = \{(1, 1), (1, 2), (2, 1), (1, 3), (3, 1), \dots, (6, 5), (6, 6)\}$$

avec $\text{card}(\Omega) = 36$: les résultats $(1, 2)$ et $(2, 1)$ sont différents ($(1, 2)$ signifie que le dé rouge vaut 1 et le dé vert vaut 2).

L'expérience est équiprobable car chaque événement élémentaire a la même probabilité de survenir, $\mathbb{P}(\{(i, j)\}) = 1/36$ pour tout couple $(i, j) \in \Omega$.

L'événement « la somme des chiffres vaut 7 » est alors $E = \{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\}$ et donc $\mathbb{P}(E) = \frac{6}{36} = \frac{1}{6} = \frac{\text{card}(E)}{\text{card}(\Omega)}$.

On suppose que la personne est daltonienne, elle ne distingue pas la couleur rouge du vert. Ça ne change pas l'expérience aléatoire et donc la probabilité de l'événement E ne devrait pas changer.

Par contre, pour cette personne avoir "le dé rouge qui vaut 1 et le dé vert qui vaut 2" et avoir "le dé rouge qui vaut 2 et le dé vert qui vaut 1" sont les mêmes événements (elle ne peut pas les distinguer).

Par conséquent, l'ensemble fondamental est $\Omega' = \{(1, 1), (1, 2), (1, 3), \dots, (5, 6), (6, 6)\}$ et $\text{card}(\Omega') = 21$. On n'est plus dans une situation d'équiprobabilité car $\mathbb{P}(\{(1, 1)\}) \neq \mathbb{P}(\{(1, 2)\})$. En effet, $\mathbb{P}(\{(i, i)\}) = 1/36$ et pour $i < j$, $\mathbb{P}(\{(i, j)\}) = \frac{2}{36}$.

L'événement « la somme des chiffres vaut 7 » est alors $E' = \{(1, 6), (2, 5), (3, 4)\}$ et donc $\mathbb{P}(E') = \frac{2}{36} + \frac{2}{36} + \frac{2}{36} = \frac{1}{6} \left(\neq \frac{\text{card}(E')}{\text{card}(\Omega')} \right)$.

Conseil : Si on peut se ramener à une modélisation de l'expérience aléatoire telle que les événements élémentaires soient équiprobables, il est conseillé de le faire car les calculs sont plus simples.

Exemple 2.11. Trois boules sont tirées au hasard de façon successives dans une urne contenant 6 boules blanches et 5 noires (sans remise). Quelle est la probabilité de tirer 1 boule blanche et 2 boules noires ?

On se ramène à une modélisation équiprobable, en imaginant que les boules sont distinguables (par exemple, numérotées). Ceci ne change pas l'expérience, mais rend les calculs plus simples.

Si on ne tient pas compte de l'ordre dans lequel les boules apparaissent :

Le nombre d'éléments de Ω correspond au nombre de tirage de 3 boules parmi 11, soit $\text{card}(\Omega) = \binom{11}{3}$, tandis que $E = \{\text{tirer 1 boule blanche et 2 boules noires}\}$ a $\binom{6}{1} \binom{5}{2}$ éléments. Par conséquent,

$$\mathbb{P}(E) = \frac{\text{card}(E)}{\text{card}(\Omega)} = \frac{\binom{6}{1} \binom{5}{2}}{\binom{11}{3}} = \frac{4}{11}$$

Si on tient compte de l'ordre dans lequel les boules apparaissent :

On a alors $\text{card}(\Omega) = 11 \cdot 10 \cdot 9 = 990$ et $\text{card}(E) = 6 \times (5 \times 4) \times 3 = 120 \times 3 = 360$ (où le 3 correspond au nombre de positions de la boule blanche dans les trois tirages) et donc $\mathbb{P}(E) = \frac{4}{11}$.

On remarque dans l'exemple précédent, que tenir compte de l'ordre ou pas, ne change pas la probabilité. Dans les calculs, il faut cependant faire attention : si on tient compte de l'ordre pour le nombre de cas favorables, il faut aussi tenir compte de l'ordre pour le nombre de cas possibles. Il est souvent plus simple dans les calculs d'utiliser la modélisation qui ne tient pas compte de l'ordre.

Exemple 2.12. Si n personnes sont présentes dans une pièce, quelle est la probabilité que leurs anniversaires tombent sur des jours tous différents ?

$$\mathbb{P}(E) = \frac{365 \cdot 364 \cdots (365-n+1)}{365^n} = \frac{365!}{(365-n)!(365)^n}.$$

Exemple 2.13. Quelle est la probabilité d'avoir un jeu de 5 cartes dont 3 sont de la même valeur et une paire (différente du triplet) sur un jeu de 52 cartes ?

On fait le calcul sans tenir compte de l'ordre des cartes. Dans un jeu de 52 cartes, il y a 13 cartes pour chaque « couleur ». On suppose que les $\binom{52}{5}$ mains possibles sont équiprobables.

Il y a $\binom{4}{3}\binom{4}{2}$ choix possibles d'avoir trois valets et 2 dames. Ensuite, il y a 13 choix possible pour la valeur du triplets et 12 choix possibles pour la valeur de la paire. On a

$$\mathbb{P}(E) = \frac{13 \cdot 12 \cdot \binom{4}{3} \binom{4}{2}}{\binom{52}{5}}.$$

Exemple 2.14. Parmi les membres d'un club de sport, 36 jouent au tennis, 28 au squash et 18 au badminton, les autres jouant d'autres sports. Par ailleurs, 22 jouent au tennis et au squash, 12 tennis+bad, 9 squash+bad et enfin 4 pratiquent les trois sports. Quelle est la probabilité de pratiquer au moins un des trois sports ?

Soit N le nombre de membres total :

$$\begin{aligned} \mathbb{P}(T \cup S \cup B) &= \mathbb{P}(T) + \mathbb{P}(S) + \mathbb{P}(B) - \mathbb{P}(T \cap S) - \mathbb{P}(T \cap B) - \mathbb{P}(S \cap B) + \mathbb{P}(T \cap S \cap B) \\ &= \frac{36}{N} + \frac{28}{N} + \frac{18}{N} - \frac{22}{N} - \frac{12}{N} - \frac{9}{N} + \frac{4}{N} = \frac{43}{N} \end{aligned}$$

Exemple 2.15. Une soirée rassemble N personnes, qui laissent leur manteau dans l'entrée.

1. Quelle est la proba que tout le monde reparte avec son manteau ?
2. Quelle est la proba qu'aucun ne reparte avec son manteau ?
3. Quelle est la proba qu'exactement k personnes repartent avec leur manteau ?

Solution :

1. Soit $F = \{\text{chacun repart avec son propre manteau}\}$. Comme il y a $N!$ façon de répartir les manteaux et une seule qui permet à chacun de récupérer son manteau, on a alors $\mathbb{P}(F) = 1/N!$.
2. Attention cet événement n'est pas le complémentaire du précédent.

Soit $E = \{\text{aucune personne ne repart avec son propre manteau}\}$. Cet événement est difficile à décrire, cependant

$$E^c = \{\text{au moins une personne repart avec son propre manteau}\} = \bigcup_{i=1}^N E_i$$

où $E_i = \{\text{la personne } i \text{ repart avec son propre manteau}\}$. On a $\mathbb{P}(E_i) = \frac{1}{N}$ ($= \frac{(N-1)!}{N!}$), $\mathbb{P}(E_i \cap E_j) = \frac{1}{N(N-1)}$, $\mathbb{P}(E_{i_1} \cap \dots \cap E_{i_k}) = \frac{(N-k)!}{N!}$. Alors

$$\sum_{i_1 < \dots < i_k} \mathbb{P}(E_{i_1} \cap \dots \cap E_{i_k}) = \binom{N}{k} \frac{(N-k)!}{N!} = \frac{1}{k!}$$

Donc

$$\mathbb{P}(E^c) = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{N+1} \frac{1}{N!}.$$

Par conséquent, pour N grand, $\mathbb{P}(E) = 1 - \mathbb{P}(E^c) \simeq e^{-1} = 0.3679$.

3. On doit choisir k personnes parmi N qui prendront le bon manteau $\binom{N}{k}$. On sépare le groupe en deux sous-groupes : ceux qui prennent leur manteaux et ceux qui prennent un autre manteau. Par conséquent,

$$\begin{aligned} & \mathbb{P}(k \text{ personnes exactement ne repartent avec leur propre manteau}) \\ &= \binom{N}{k} \frac{(N-k)!}{N!} \left(1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + (-1)^{N-k+1} \frac{1}{(N-k)!} \right) \\ &= \frac{1}{k!} \left(1 - 1 + \frac{1}{2!} - \frac{1}{2!} + \dots + (-1)^{N-k+1} \frac{1}{(N-k)!} \right) \end{aligned}$$

2.4 Passage à la limite

Une suite d'événements est dite croissante si

$$E_1 \subset E_2 \subset \dots \subset E_n \subset E_{n+1} \subset \dots$$

et on note alors

$$\lim_{n \rightarrow \infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

Une suite d'événements est dite décroissante si

$$E_1 \supset E_2 \supset \dots \supset E_n \supset E_{n+1} \supset \dots$$

$$\lim_{n \rightarrow \infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

Théorème 2.6. Si $(E_n)_{n \geq 0}$ est une suite monotone d'événement (croissante ou décroissante), alors

$$\lim_{n \rightarrow \infty} \mathbb{P}(E_n) = \mathbb{P}\left(\lim_{n \rightarrow \infty} E_n\right)$$

Preuve. Si la suite est croissante, notons

$$F_1 = E_1, F_2 = E_2 \setminus E_1, \dots, F_n = E_n \setminus E_{n-1}, \dots$$

Les ensembles F_i sont disjoints et pour tout n

$$\bigcup_{i=1}^n F_i = \bigcup_{i=1}^n E_i = E_n \text{ et } \bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i$$

Par conséquent

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} F_i\right) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(F_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(F_i) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{i=1}^n F_i\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \lim_{n \rightarrow \infty} \mathbb{P}(E_n) \end{aligned}$$

Si la suite (E_n) est décroissante, alors la suite (E_n^c) est croissante et on obtient alors les résultat facilement $((\bigcup_{i=1}^{\infty} E_i)^c = \bigcap_{i=1}^{\infty} E_i^c)$ $\square$

Exemple 2.16. On lance un dé une infinité de fois et on cherche la probabilité qu'aucun 6 n'apparaisse. Une réalisation de l'expérience s'écrit sous la forme d'une suite infinie $\omega = (\omega_n)_{n \geq 1}$ avec ω_i le résultat du i ème lancer. On a donc $\omega_i \in \{1, \dots, 6\}$. Soit E l'événement n'avoir aucun 6. On a alors $\omega \in E$ est équivalent à "pour tout i , $\omega_i \neq 6$ ".

On introduit E_n l'événement n'avoir aucun 6 sur les n premiers lancers. On voit que la suite $(E_n)_{n \geq 1}$ est décroissante et $E = \bigcap_{n=1}^{\infty} E_n$.

Par conséquent,

$$\mathbb{P}(E) = \lim_{n \rightarrow +\infty} \mathbb{P}(E_n).$$

De plus,

$$\begin{aligned} \mathbb{P}(E_n) &= \mathbb{P}(\text{pas de 6 sur les } n \text{ premiers lancers et ensuite on s'en fiche}) \\ &= \mathbb{P}(\text{pas de 6 sur } n \text{ lancers}) = \frac{5^n}{6^n}. \end{aligned}$$

On déduit que $\mathbb{P}(E) = 0$, l'événement $E^c = \text{avoir un 6}$ est un événement certain.

Chapitre 3

Probabilités conditionnelles et indépendance

Sommaire

3.1	Probabilités conditionnelles	21
3.2	Formule de Bayes	24
3.3	Indépendance	26
a)	Indépendance entre plusieurs événements	27
b)	Expériences indépendantes	28

3.1 Probabilités conditionnelles

En 2021, au Québec, le nombre de cas déclarés de porteurs du virus de l'hépatite C (VHC_+) est 805 personnes¹ Un médecin avant de recevoir un patient inconnu en consultation peut donc estimer que ce patient a une probabilité de 0.5% d'être VHC_+ en supposant que sa clientèle ressemble globalement à l'ensemble de la population française (qui comporte environ 60 millions de personnes). Si en consultant le dossier du patient, il s'aperçoit que son patient a 10 ans, alors la probabilité que ce patient soit VHC_+ chute pour devenir inférieur à 10^{-4} (vu le mode de transmission du virus). Par contre si en consultant le dossier, il s'aperçoit que le patient est toxicomane par voie intraveineuse depuis 10 ans, les données épidémiologiques montrent que la probabilité que ce patient soit VHC_+ est alors supérieure à 50%. Si dans le dossier, il est seulement indiqué que le patient est asthmatique, alors le médecin n'aura aucune information supplémentaire vu qu'aucun lien n'a été établi entre l'asthme et le fait d'être porteur de VHC. Le médecin estimera donc que son patient a toujours une probabilité de 0.5% d'être VHC_+ . Le fait d'avoir l'asthme influe pas sur la probabilité d'être VHC_+ , on dit que ces événements sont indépendants. On voit bien dans cet exemple qu'une information complémentaire peut influencer ou non la probabilité d'être VHC_+ . La notion de probabilité conditionnelle permet de prendre en compte des informations complémentaires dans le calcul des probabilités.

Dans l'exemple suivant, on quantifie exactement le rôle d'une information complémentaire.

Exemple 3.1. On jette deux dés distincts. On a $\Omega = \{(1, 1), (1, 2), \dots, (5, 6), (6, 6)\}$. On a donc 36 événements élémentaires qui ont chacun la même probabilité d'apparaître, soit $1/36$.

1. Voir le site du [gouvernement du Canada](#).

On veut connaître la probabilité de l'événement E : "la somme des dés vaut 8". On remarque que $E = \{(2, 6), (6, 2), (3, 5), (5, 3), (4, 4)\}$, d'où $\mathbb{P}(E) = 5/36$.

Maintenant on cherche toujours la même probabilité, mais on sait déjà que le premier dé donne un 3. On note F l'événement : "le premier dé vaut 3".

Que vaut alors la probabilité que la somme vaut 8 sachant que le premier dé vaut 3? La probabilité recherchée est appelée probabilité conditionnelle de E sachant F et est notée $\mathbb{P}(E|F)$.

Sachant que le premier dé vaut 3, on ne regarde plus la même expérience, l'ensemble des possibilités ne sont plus les mêmes, c'est un sous-ensemble de Ω . L'aléa ne porte plus que sur le deuxième dé (c'est comme si on ne lançait plus qu'un dé). Il y a alors 6 résultats dans cette expérience : $\Omega' = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$ et chacune a la même probabilité d'apparaître, soit $\mathbb{P}(\{(3, i)\}) = 1/6$ pour $i \in \{1, \dots, 6\}$. Par conséquent la probabilité que la somme des deux dés soit égale à 8 sachant que le premier dé vaut 3 est $\mathbb{P}(E|F) = 1/6$ (il n'y a qu'une possibilité, le deuxième dé doit valoir 5).

On remarque que l'événement $E \cap F$ est l'événement "le premier dé donne 3 et la somme des chiffres vaut 8 lorsqu'on lance deux dés", on a alors $E \cap F = \{(3, 5)\}$ dans l'ensemble $\Omega = \{(1, 1), (1, 2), \dots, (5, 6), (6, 6)\}$ de cardinal 36.

D'autre part, on a $F = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$.

D'où $\mathbb{P}(E \cap F) = 1/36$ et $\mathbb{P}(F) = 6/36 = 1/6$. On remarque que

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

Définition 3.1. Soient E et F deux événements avec $\mathbb{P}(F) \neq 0$. La **probabilité conditionnelle** $\mathbb{P}(E|F)$ de E sachant F est définie par :

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

Proposition 3.2. Soit F un événement avec $\mathbb{P}(F) \neq 0$. Alors l'application

$$E \mapsto \mathbb{P}(E|F)$$

est une probabilité sur Ω .

Démonstration. Il faut vérifier que l'application vérifie les 3 axiomes :

1. Soit E un événement, on a $E \cap F \subset F$, donc $\mathbb{P}(E \cap F) \leq \mathbb{P}(F)$. Comme $\mathbb{P}$ est à valeurs positives, on déduit que $\mathbb{P}(E|F) \in [0, 1]$.
2. $\mathbb{P}(\Omega|F) = \frac{\mathbb{P}(\Omega \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F)}{\mathbb{P}(F)} = 1$.
3. Soit $(E_i)_{i \geq 1}$ une suite d'ensembles deux à deux disjoints, alors les ensembles $(E_i \cap F)_{i \geq 1}$ sont aussi deux à deux disjoints. Comme $(\cup_{i \geq 1} E_i) \cap F = \cup_{i \geq 1} (E_i \cap F)$ et $\mathbb{P}$ est une probabilité, on déduit que

$$\mathbb{P}(\cup_{i \geq 1} E_i|F) = \frac{\mathbb{P}(\cup_{i \geq 1} (E_i \cap F))}{\mathbb{P}(F)} = \sum_{i \geq 1} \mathbb{P}(E_i|F).$$

□

Conséquences : La probabilité conditionnelle vérifie les mêmes propriétés qu'une probabilité :

- Si $E \cap G = \emptyset$, alors $\mathbb{P}(E \cup G|F) = \mathbb{P}(E|F) + \mathbb{P}(G|F)$.
- $\mathbb{P}(E^c|F) = 1 - \mathbb{P}(E|F)$.

Attention ! $\mathbb{P}(E|F^c) \neq 1 - \mathbb{P}(E|F)$.

Exemple 3.2. Une pièce de monnaie équilibrée est lancée 2 fois. Quelle est la probabilité de A : « les deux jets donnent face » sachant l'événement B : « le premier donne face » ?

On peut se placer sur le nouvel espace fondamental $\Omega' = \{(F, F), (F, P)\}$ et donc $\mathbb{P}(A|B) = 1/2$, ou utiliser la formule vue plus tôt $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ sur l'espace $\Omega = \{(F, F), (F, P), (P, F), (P, P)\}$, on a donc $A \cap B = \{(F, F)\}$, soit $\mathbb{P}(A \cap B) = 1/4$ et $B = \{(F, F), (F, P)\}$, soit $\mathbb{P}(B) = 1/2$.

Quelle est la probabilité que les deux jets donnent face sachant l'événement C : « au moins un des jets donne face » ? On peut se placer sur le nouvel espace fondamental $\Omega' = \{(F, F), (F, P), (P, F)\}$ et donc $\mathbb{P}(A|C) = 1/3$ ou calculer $\mathbb{P}(A \cap C) = \mathbb{P}(A) = 1/4$ et $\mathbb{P}(C) = \mathbb{P}(\{(F, F), (F, P), (P, F)\}) = 3/4$.

Exemple 3.3. Quatre personnes, prénommées Nord, Sud, Est, Ouest, jouent au bridge et reçoivent chacun 13 cartes d'un jeu de 52, sachant que Nord et Sud cumulent 8 piques, quelle est la proba que Est ait exactement 3 des 5 piques restant ?

On note E l'événement « Est a exactement 3 piques » et F l'événement « Nord et Sud cumulent 8 piques ». Dans le cas présent la première méthode est la plus simple à utiliser, car l'événement F décrit mal comment les cartes sont distribuées entre Nord et Sud.

Si on considère le nouvel espace fondamental Ω' de l'expérience qui consiste à distribuer 13 cartes à deux joueurs sur un paquet de 26 cartes contenant $13 - 8 = 5$ piques, on a alors $\mathbb{P}(E|F) = \frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} = 0.339$

Propriété 3.3 (Règle de multiplication). Soient E, F, G trois événements. On a

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

Par conséquent

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

De façon générale, si $E_1, \dots, E_n$ sont des événements, on a

$$\mathbb{P}(E_1 \cap \dots \cap E_n) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)\mathbb{P}(E_3|E_1 \cap E_2) \dots \mathbb{P}(E_n|E_1 \cap E_2 \cap \dots \cap E_{n-1}).$$

Exemple 3.4. On coupe aléatoirement un jeu de 52 cartes en 4 piles de 13 cartes. Quelle est la probabilité que chaque pile contienne un as ? Notons E_1 l'as de pique est dans une des piles, E_2 l'as de pique et l'as de coeur sont dans des piles différentes, E_3 l'as de pique, de coeur et de carreau sont dans des piles différentes, E_4 les quatres as sont dans des piles différentes. On cherche $\mathbb{P}(E_4) = \mathbb{P}(E_4 \cap E_3 \cap E_2 \cap E_1)$ et

$$\mathbb{P}(E_4 \cap E_3 \cap E_2 \cap E_1) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)\mathbb{P}(E_3|E_2 \cap E_1)\mathbb{P}(E_4|E_3 \cap E_2 \cap E_1)$$

On a $\mathbb{P}(E_1) = 1$, $\mathbb{P}(E_2|E_1) = \frac{\binom{50}{12}}{\binom{51}{12}} = 39/51$ (50-12=38) car la pile de l'as de pique recevra 12 cartes parmi les 51 restantes, $\mathbb{P}(E_3|E_2) = \frac{\binom{49}{24}}{\binom{50}{24}} = 26/50$ puisque les piles contenant pique et coeur recevront 24 cartes des 50 restantes, $\mathbb{P}(E_4|E_3) = \frac{\binom{48}{36}}{\binom{49}{36}} = 13/49$ et donc

$$\mathbb{P}(E_4) = \frac{39 \cdot 26 \cdot 13}{51 \cdot 50 \cdot 49} = 0.105$$

3.2 Formule de Bayes

On considère deux événements E et F quelconques.

On peut E écrire comme l'union de deux événements disjoints : $E = (E \cap F) \cup (E \cap F^c)$, d'où

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c) \\ &= \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|F^c)\mathbb{P}(F^c) \\ \mathbb{P}(E) &= \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|F^c)(1 - \mathbb{P}(F))\end{aligned}\tag{3.1}$$

L'équation (3.1) s'appelle **formule des probabilités totales**.

Exemple 3.5. Une compagnie d'assurance répartit ses clients en deux classes : ceux qui sont enclins aux risques cardio-vasculaires (à haut risque) et ceux qui ne le sont pas (à bas risque). Elle constate que ceux à haut risque ont une probabilité de 0.4 d'avoir un accident dans l'année, tandis que ceux à bas risque ont une probabilité de 0.2. On suppose que 30% de la population est à haut risque. Quelle est alors la probabilité qu'un assuré ait un accident cardio-vasculaire durant l'année qui suit la signature de son contrat ?

On note A l'événement "l'assuré a un accident pendant l'année" et H l'événement "l'individu est à haut risque". Alors d'après la formule des probabilités totales (3.1), on a

$$\mathbb{P}(A) = 0.4 \times 0.3 + 0.2 \times 0.7 = 0.26.$$

La formule de Bayes permet de calculer les probabilités a posteriori d'un événement en fonction des probabilités a priori de cet événement, i.e. connaître $\mathbb{P}(F|E)$ quand on connaît $\mathbb{P}(E|F)$ et $\mathbb{P}(E|F^c)$.

Proposition 3.4. Soient E et F deux événements. La formule de Bayes s'écrit :

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

Démonstration. Il suffit d'écrire $\mathbb{P}(F|E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E|F)\mathbb{P}(F)}{\mathbb{P}(E)}$ et d'utiliser la formule des probabilités totales. □

Exemple 3.6. On reprend l'exemple 3.5 sur la compagnie d'assurance. On veut connaître la probabilité qu'un assuré ayant eu un accident cardio-vasculaire soit à bas risque, i.e. la probabilité $\mathbb{P}(B|A)$, où B est l'événement "l'assuré est à bas risque". On remarque que $B = H^c$. D'où

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|H^c)\mathbb{P}(H^c)}{\mathbb{P}(A)} = \frac{0.2 \times 0.7}{0.26} = 0.54.$$

Exemple 3.7. Un étudiant répond à une question à choix multiples, m réponses sont proposées. Il connaît la réponse avec proba p et avec proba $1 - p$ il choisit au hasard la réponse. Quelle est la probabilité que l'étudiant connaisse la réponse s'il y a répondu correctement ?

On note B l'événement avoir la bonne réponse et C l'événement connaître la réponse.

$$\mathbb{P}(C|B) = \frac{\mathbb{P}(B|C)\mathbb{P}(C)}{\mathbb{P}(B|C)\mathbb{P}(C) + \mathbb{P}(B|C^c)(1 - \mathbb{P}(C))} = \frac{1 \cdot p}{p + (1 - p)/m}$$

Si $p = 1/2$ et $m = 4$, alors $\mathbb{P}(C|B) = 4/5$.

Exemple 3.8. Une maladie affecte 0.5% de la population. Un test T permet de dépister cette maladie avec la fiabilité suivante :

T est positif pour 95% des personnes affectées par la maladie

T est négatif pour 99% des personnes non affectées par la maladie.

Quelle la probabilité qu'un individu ayant un test positif soit affecté par la maladie ?

On note M l'événement être malade et P l'événement le test est positif.

$$\mathbb{P}(M|P) = \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} = 95/294 \simeq 0.323.$$

En effet, sur 200 personnes, en moyenne 1 personne sera porteuse de la maladie et le test décèlera ce cas avec un proba de 0.95. Donc en moyenne sur 200 personnes, on détectera 0,95 des cas. Par contre, sur les 200 personnes (dont 199 saines), le test détectera en moyenne 199.01 = 1.99 faux positifs. La proportion de résultats correct est donc : $0.95/(0,9 + 1.99) = 0.323$.

Si la maladie n'était pas rare et touchait par exemple 10% de la population ($\mathbb{P}(M) = 0.1$), alors avec les mêmes taux de succès du test on aurait

$$\mathbb{P}(M|P) = \frac{0.95 \times 0.1}{0.95 \times 0.1 + 0.01 \times 0.9} \simeq 0.913.$$

Exemple 3.9. On choisit consécutivement et sans remise n boules d'une urne composée de r boules rouges et b boules bleues, avec $n \leq r + b$. Étant donné que k des n boules sont bleues, quelle est la proba que la première tirée soit bleue ?

On note E l'événement "la première boule tirée est bleue" et F l'événement " k des n boules tirées sont bleues".

Première méthode : En utilisant la formule $\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$

On va séparer ce qui se passe pour la première boule de ce qui se passe pour les $(n - 1)$ autres.

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(F|E)\mathbb{P}(E)}{\mathbb{P}(F)} \quad \text{avec}$$

- $\mathbb{P}(F)$ est la probabilité de tirer k boules bleues sur un choix au hasard de n boules dans une urne de $b + r$ boules

$$\mathbb{P}(F) = \frac{\binom{b}{k} \binom{r}{n-k}}{\binom{r+b}{n}},$$

avec $\binom{r+b}{n}$ nombre de façons de tirer n boules parmi $r + b$ sans tenir compte de l'ordre, $\binom{b}{k}$ nombre de façons de tirer k boules parmi b boules bleues, et $\binom{r}{n-k}$ nombre de façons de tirer $n - k$ boules parmi r boules rouges,

- $\mathbb{P}(E)$ est la probabilité que la première boule soit bleue (l'ordre compte, car l'événement prend en compte l'ordre des boules), on a

$$\mathbb{P}(E) = \frac{b}{r + b}$$

(on aurait aussi pu introduire le nombre $A_n^{r+b} = \frac{(r+b)!}{(r+b-n)!}$ d'arrangements de n boules parmi $r + b$ et

comme on est dans une expérience équiprobable, on a donc $\mathbb{P}(E) = \frac{b A_{n-1}^{r+b-1}}{A_n^{r+b}} = \frac{b}{r+b}$).

- $\mathbb{P}(F|E)$ est la probabilité que dans un choix au hasard de $n - 1$ boules dans une urne contenant $b - 1$ bleues et r rouges, on obtienne $k - 1$ bleues. Donc,

$$\mathbb{P}(F|E) = \frac{\binom{b-1}{k-1} \binom{r}{n-k}}{\binom{r+b-1}{n-1}}$$

Par conséquent, $\mathbb{P}(E|F) = \frac{k}{n}$.

Seconde méthode : En utilisant le nouvel espace fondamental... ne jamais délaisser cette méthode, elle peut être très efficace!

En se plaçant sur le nouvel ensemble fondamental, on regarde l'expérience suivante : on a n boules dont k sont bleues et on les ordonne au hasard. L'événement $E|F$ est « les boules sont rangées de telle façon qu'une boule bleue soit en première position ».

$$\mathbb{P}(E|F) = \frac{k(n-1)!}{n!} = \frac{k}{n}$$

En fait, si on suppose les boules discernables (par exemple numérotées de 1 à n), le nombre de façon de ranger n boules dont k sont bleues est $n!$ et le nombre de façon de les ranger avec une bleue en première position est $k(n-1)!$ où k est le nombre de choix possibles pour la boule bleue en première position et les $n-1$ autres boules peuvent être rangées n'importe comment.

La probabilité est donc $\mathbb{P}(E|F) = \frac{k}{n}$.

Généralisation

Les formules des probabilités totales et de Bayes peuvent être généralisées de la façon suivante : soit E , $F_1, \dots, F_n$ des événements avec F_i ensembles disjoints et $\bigcup_{i=1}^n F_i = \Omega$ (on dit que $(F_i)_{1 \leq i \leq n}$ est une **partition** de l'espace fondamental), alors

$$\begin{aligned} \mathbb{P}(E) &= \sum_{i=1}^n \mathbb{P}(E|F_i)\mathbb{P}(F_i) \\ \mathbb{P}(F_j|E) &= \frac{\mathbb{P}(E|F_j)\mathbb{P}(F_j)}{\sum_{i=1}^n \mathbb{P}(E|F_i)\mathbb{P}(F_i)}. \end{aligned}$$

Exemple 3.10. Une boîte contient 3 types d'ampoule différents. Selon l'emballage, c'est ampoules fonctionnent plus de 100h avec proba 0.7 pour le type 1 (T_1), 0.2 pour le type 2 (T_2) et 0.3 pour le type 3 (T_3). On suppose que la répartition des ampoules à l'uqam est de 20% de type 1, 30% de type 2 et 50% de type 3. Quelle est la probabilité qu'une ampoule choisie au hasard fonctionne plus de 100h? $A = \{\text{l'ampoule fonctionne plus de 100h}\}$

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A|T_1)\mathbb{P}(T_1) + \mathbb{P}(A|T_2)\mathbb{P}(T_2) + \mathbb{P}(A|T_3)\mathbb{P}(T_3) \\ &= 0.7 \times 0.2 + 0.2 \times 0.3 + 0.3 \times 0.5 = 0.35 \end{aligned}$$

Si l'ampoule fonctionne plus de 100h, avec quelle proba est-elle de type 1, 2 et enfin 3? On a

$$\mathbb{P}(T_1|A) = \frac{0.7 \cdot 0.2}{0.35} = 14/35$$

puis, 6/35 et enfin 15/35.

3.3 Indépendance

Dans le langage courant, on dit de deux événements qui ne sont pas liés entre eux qu'ils sont indépendants. Par conséquent, on a tendance à dire que si E et F sont indépendants, la connaissance de F ne donne aucune information utile pour la connaissance de l'événement E et donc de manière naturelle $\mathbb{P}(E|F)$ doit être égale à $\mathbb{P}(E)$. On va donner à partir de cette constatation la définition mathématique de l'indépendance :

Définition 3.5. Deux événements E et F sont indépendants si $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$.

Conséquence, si E et F indépendants, avec $\mathbb{P}(E) \neq 0$ et $\mathbb{P}(F) \neq 0$, on a $\mathbb{P}(E|F) = \mathbb{P}(E)$ et $\mathbb{P}(F|E) = \mathbb{P}(F)$.

Attention ! Les notions d'événements indépendants et d'événements disjoints n'ont aucun rapport. Par exemple, si on jette deux dés et on considère les événements E = "le premier dé vaut 3" et F = "le deuxième dé vaut 1". Les événements sont indépendants car on jette les dés de façon indépendante, mais les événements ne sont pas disjoints car l'événement $E \cap F = \{\text{le premier dé vaut 3 et le deuxième vaut 1}\}$ est possible.

Remarque 3.6. Soient E et F deux événements.

- Si E et F sont disjoints, alors $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$.
- Si E et F sont indépendants, alors $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E)\mathbb{P}(F)$.

Proposition 3.7. Si E et F sont deux événements indépendants, alors E et F^c le sont aussi.

Démonstration. On écrit E comme l'union de deux événements disjoints : $E = (E \cap F) \cup (E \cap F^c)$. D'où $\mathbb{P}(E) = \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c) = \mathbb{P}(E)\mathbb{P}(F) + \mathbb{P}(E \cap F^c)$. Ce qui implique que $\mathbb{P}(E \cap F^c) = \mathbb{P}(E) - \mathbb{P}(E)\mathbb{P}(F) = \mathbb{P}(E)\mathbb{P}(F^c)$ et donc E et F^c sont indépendants. $\square$

a) Indépendance entre plusieurs événements

Définition 3.8. On considère n événements $E_1, \dots, E_n$.

- Les événements sont **deux à deux indépendants** si pour tout $i \neq j$ distincts on a $\mathbb{P}(E_i \cap E_j) = \mathbb{P}(E_i)\mathbb{P}(E_j)$.
- Les événements sont **(mutuellement) indépendants** (ou totalement indépendants) si pour tout $k \leq n$, pour tout $i_1, \dots, i_k \in \{1, \dots, n\}$ distincts on a $\mathbb{P}(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_k}) = \mathbb{P}(E_{i_1})\mathbb{P}(E_{i_2}) \dots \mathbb{P}(E_{i_k})$.

Remarque 3.9. Si des événements sont mutuellement indépendants, alors ils sont indépendants deux à deux, mais la réciproque est fausse.

Par exemple, on jette deux dés, l'un rouge l'autre vert et on considère les événements suivants :

$$\begin{aligned} E &= \{\text{le dé rouge donne un chiffre impair}\} \\ F &= \{\text{le dé vert donne un chiffre impair}\} \\ G &= \{\text{la somme des deux dés est un chiffre impair}\}. \end{aligned}$$

E , F et G sont deux à deux indépendants mais pas mutuellement indépendants.

En effet, On a $\mathbb{P}(E) = \mathbb{P}(F) = \mathbb{P}(G) = 1/2$. En effet, comme les dés sont bien équilibrés l'expérience est équiprobable, et donc

$$\mathbb{P}(E) = \frac{3}{6} = \frac{1}{2} = \mathbb{P}(F)$$

$$\mathbb{P}(G) = \mathbb{P}(\text{dé rouge impair et dé vert pair}) + \mathbb{P}(\text{dé rouge pair et dé vert impair}) = \frac{3 \times 3}{36} + \frac{3 \times 3}{36} = \frac{1}{2}.$$

De plus, on a

$$\begin{aligned} \mathbb{P}(E \cap F) &= \mathbb{P}(\text{dé rouge impair et le dé vert impair}) = \frac{3 \times 3}{36} = \frac{1}{4} \Rightarrow E \text{ et } F \text{ sont donc indépendants.} \\ \mathbb{P}(E \cap G) &= \mathbb{P}(\text{dé rouge impair et dé vert pair}) = \mathbb{P}(E \cap F^c) \\ &= \mathbb{P}(E)\mathbb{P}(F^c) = 1/4 \quad \text{par indépendance entre } E \text{ et } F \\ &= \mathbb{P}(E)\mathbb{P}(G) \Rightarrow E \text{ et } G \text{ sont donc indépendants.} \end{aligned}$$

On montre de même que F et G sont indépendants. Par contre, $\mathbb{P}(E \cap F \cap G) = 0 \neq \mathbb{P}(E)\mathbb{P}(F)\mathbb{P}(G)$.

b) Expériences indépendantes

Un exemple basique consiste à considérer des jets successifs d'un dé. On peut considérer les lancers comme indépendants, car l'un des résultats n'aura pas d'influence sur les autres. Le résultat de chaque lancer est dit partiel. En fait pour chaque lancer l'ensemble fondamentale est $\Omega^{\text{partiel}} = \{1, 2, 3, 4, 5, 6\}$. Si on lance trois fois le dé, le résultat de toute l'expérience est

$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} = \Omega_1^{\text{partiel}} \times \Omega_2^{\text{partiel}} \times \Omega_3^{\text{partiel}},$$

avec $\Omega_i^{\text{partiel}}$ le résultat du i ème lancer.

On pourrait aussi considérer le cas d'un lancer de dé et d'un lancer de pièce. On aurait alors

$$\Omega = \Omega_1^{\text{partiel}} \times \Omega_2^{\text{partiel}},$$

avec $\Omega_1^{\text{partiel}} = \{1, 2, 3, 4, 5, 6\}$ et $\Omega_2^{\text{partiel}} = \{0, 1\}$.

On parle d'**épreuves**, si les expériences partielles d'une expérience aléatoire sont toutes identiques.

On parle **épreuves indépendantes** si de plus, comme dans le cas de lancers successifs d'un dé, les expériences partielles n'ont pas d'influence entre elles.

Si de plus les épreuves partielles n'ont que deux résultats possibles *succès/échec*, on parle d'**épreuves de Bernoulli**.

Quand on effectue n épreuves de Bernoulli, on peut s'intéresser au nombre de succès obtenus. L'ensemble fondamental est alors $\Omega = \{0, 1\}^n = \{(\omega_k)_{1 \leq k \leq n} : \omega_k \in \{0, 1\}\}$ ($\omega_k = 1$ si la k ème épreuve est un succès et 0 sinon). On a $\Omega_k^{\text{partiel}} = \{0, 1\}$ l'ensemble fondamental de la k ème épreuve.

On introduit les événements

$$F_k = \text{« avoir } k \text{ succès sur } n \text{ épreuves de Bernoulli »}.$$

et

$$E_k = \text{« la } k \text{ème épreuve est un succès »}.$$

Comme toutes les épreuves sont supposées identiques, on a

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

On note p cette probabilité.

On remarque que $F_0 = E_1^c \cap \dots \cap E_n^c$ et $F_n = E_1 \cap \dots \cap E_n$. Comme les expériences sont indépendantes, on a

$$\mathbb{P}(F_0) = (1 - p)^n \quad \text{et} \quad \mathbb{P}(F_n) = p^n.$$

De plus,

$$F_1 = (E_1 \cap E_2^c \cap \dots \cap E_n^c) \cup (E_1^c \cap E_2 \cap E_2^c \cap \dots \cap E_n^c) \cup \dots \cup (E_1^c \cap \dots \cap E_{n-1}^c \cap E_n)$$

Les événements de l'union sont disjoints et ont tous la même probabilité $p(1 - p)^{n-1}$. Il y a n événements, donc

$$\mathbb{P}(F_1) = np(1 - p)^{n-1} = \binom{n}{1} p^1 (1 - p)^{n-1}.$$

On peut généraliser le raisonnement pour F_k où on doit choisir k expériences parmi n qui donneront un succès et la probabilité de chaque possibilité est $p^k(1 - p)^{n-k}$. Par conséquent, la probabilité de l'événement F_k est

$$\mathbb{P}(\text{« avoir } k \text{ succès sur } n \text{ épreuves de Bernoulli »}) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Exemple 3.11. On jette une pièce $2n$ fois et on cherche la probabilité d'avoir autant de Pile que de Face, i.e avoir n fois Pile sur $2n$ lancers.

On a

$$\mathbb{P}(\text{avoir } n \text{ Piles sur } 2n \text{ lancers}) = \binom{2n}{n} \left(\frac{1}{2}\right)^n \left(\frac{1}{2}\right)^n.$$

Dans certaines situations, on devra considérer une infinité d'épreuves de Bernoulli. On peut par exemple considérer les événements $G_n = \text{"avoir que des succès lors des } n \text{ premières épreuves"}$.

Alors $G = \bigcap_{n \geq 1} G_n$ représente le fait de n'obtenir que des succès. Les G_n étant une suite décroissantes d'événements, on a

$$\mathbb{P}(G) = \lim_{n \rightarrow +\infty} \mathbb{P}(G_n) = \lim_{n \rightarrow +\infty} p^n = \begin{cases} 0 & \text{si } 0 \leq p < 1 \\ 1 & \text{si } p = 1. \end{cases}$$

Exemple 3.12. Deux joueurs A et B jouent à Pile ou Face à tour de rôle. Le joueur A commence. Le premier joueur qui obtient pile gagne. Quelle est la probabilité que le joueur A gagne?

Le joueur A gagne si un jet impair donne Pile. La suite de lancer est une suite d'épreuves de Bernoulli indépendantes. On note E_n le jeu s'arrête au n ième lancer.

On a $E_n = \text{"avoir Face sur les } n-1 \text{ premiers lancers et Pile au } n \text{ième lancer"}$. Les lancers étant indépendants, on a $\mathbb{P}(E_n) = \left(\frac{1}{2}\right)^{n-1} \frac{1}{2} = \frac{1}{2^n}$.

Comme $A = \bigcup_{n \geq 0} E_{2n+1}$ et comme les événements E_n sont disjoints, on déduit

$$\mathbb{P}(A \text{ gagne}) = \sum_{n=0}^{\infty} \mathbb{P}(E_{2n+1}) = \sum_{n=0}^{\infty} \frac{1}{2^{2n+1}} = \frac{1}{2} \frac{1}{1 - \frac{1}{4}} = \frac{2}{3}.$$

Chapitre 4

Variables aléatoires

Sommaire

4.1	Variables aléatoires et fonctions de répartition	31
a)	Fonction de répartition	33
b)	Calcul de probabilités et fonction de répartition	34
4.2	Variables aléatoires discrètes	35
a)	Fonction de masse	35
b)	Lien entre fonction de masse et fonction de répartition	36
c)	Lois discrètes usuelles	37
4.3	Variables aléatoires continues	41
a)	Propriétés	41
b)	Loi d'une v.a. transformée	43
c)	Lois à densité usuelles	44

4.1 Variables aléatoires et fonctions de répartition

On considère une expérience aléatoire. L'ensemble fondamental Ω est l'ensemble des résultats. $\omega \in \Omega$ est le résultat d'une réalisation de l'expérience.

Souvent on ne s'intéresse pas à tous les résultats d'une expérience aléatoire, mais juste à certaines quantités associées.

Définition 4.1. Une variable aléatoire (v.a.) X est une fonction définie sur Ω et à valeurs dans $\mathbb{R}$

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega) \end{aligned}$$

qui à chaque résultat ω associe un nombre réel $X(\omega)$.

$X(\omega)$ est appelé « une réalisation de X ».

Rappel La fonction indicatrice de l'ensemble A est définie par

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A. \end{cases}$$

- Exemple 4.1.** 1. On lance un dé et on s'intéresse au résultat X du lancer. On a alors $\Omega = \{1, \dots, 6\}$ et $X(\omega) = \omega$. Si on s'intéresse au carré Y du résultat, on a $Y(\omega) = \omega^2$.
2. On lance deux dés et on s'intéresse à la somme des chiffres obtenus. On a alors $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\} = \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \dots, 6\}\}$, où ω_i est le résultat du $i^{\text{ème}}$ dé. On pose alors

$$X(\omega) = X(\omega_1, \omega_2) = \omega_1 + \omega_2.$$

3. On lance une pièce et on regarde le résultat. $\Omega = \{Pile, Face\}$. On pose

$$X(\omega) = \begin{cases} 1 & \text{si } \omega = Face \\ 0 & \text{si } \omega = Pile \end{cases}$$

On peut écrire $X(\omega) = \mathbb{1}_{\{Face\}}(\omega)$.

4. On lance 4 pièces identiques et on regarde le nombre de Face. On a

$$\Omega = \{Pile, Face\}^4 = \{\omega = (\omega_1, \dots, \omega_4) : \omega_i \in \{Pile, Face\} \forall i \in \{1, \dots, 4\}\}.$$

Alors $X(\omega_1, \dots, \omega_4) = \sum_{i=1}^4 \mathbb{1}_{\{Face\}}(\omega_i)$.

Souvent, on triche et on simplifie directement le problème en notant l'espace fondamental $\Omega = \{0, 1\}^4 = \{\omega = (\omega_1, \dots, \omega_4) : \omega_i \in \{0, 1\} \forall i \in \{1, \dots, 4\}\}$, la valeur 0 correspondant à Pile et la valeur 1 correspondant à Face. Alors $X(\omega_1, \dots, \omega_n) = \sum_{i=1}^4 \omega_i$.

5. On regarde la durée de vie d'une ampoule. On a alors $\Omega = [0, +\infty[$ et $X(\omega) = \omega$.
6. Il y a 3 ampoules dans la classe et on s'intéresse au premier instant où il faudra changer une ampoule. On a $\Omega = [0, +\infty[\times [0, +\infty[\times [0, +\infty[$. $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega$ où ω_i représente la durée de vie de l'ampoule i . Alors $X(\omega) = \min(\omega_1, \omega_2, \omega_3)$.

Proposition 4.2. Si X est une variable aléatoire et une fonction $h : \mathbb{R} \rightarrow \mathbb{R}$, alors $Y = h(X)$ est une variable aléatoire :

$$\begin{aligned} Y : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto Y(\omega) = h(X(\omega)) = h \circ X(\omega). \end{aligned}$$

Définition 4.3. Soit X une v.a. Soit B un sous-ensemble de $\mathbb{R} : B \subset \mathbb{R}$. L'ensemble $\{X \in B\}$ est le sous-ensemble de Ω défini par

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}.$$

$\{X \in B\}$ est un événement. On peut donc calculer sa probabilité, qui est notée $\mathbb{P}(X \in B)$.

Notation

- Soit $b \in \mathbb{R}$ un réel. Si $B = \{b\}$, on note $\{X = b\}$ l'ensemble $\{X \in \{b\}\}$.
- Soit $a, b \in \mathbb{R}$ tels que $a \leq b$. Si $B = [a, b]$, on note $\{a \leq X \leq b\} = \{X \in [a, b]\}$. De même avec $B =]a, b]$, $\{a < X \leq b\} = \{X \in]a, b]\}$. Si $B =]-\infty, b]$, $\{X \leq b\} = \{X \in]-\infty, b]\}$.
- Si on a deux variables X et Y , on note $\{X = Y\} = \{\omega \in \Omega : X(\omega) = Y(\omega)\}$ et $\{X \leq Y\} = \{\omega \in \Omega : X(\omega) \leq Y(\omega)\}$.

Exemple 4.2. On lance deux dés équilibrés. On note X la somme des deux chiffres obtenus. On a alors $\Omega = \{1, \dots, 6\}^2$ et

$$\begin{aligned} \{X = 6\} &= \{X \in \{6\}\} = \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\} \\ \{X \leq 3\} &= \{X \in]-\infty, 3]\} = \{X = 1\} \cup \{X = 2\} \cup \{X = 3\} = \{(1, 1), (1, 2), (2, 1)\} \end{aligned}$$

Par conséquent, $\mathbb{P}(X = 6) = 5/36$ et $\mathbb{P}(X \leq 3) = 3/36$.

a) Fonction de répartition

Définition 4.4. Soit X une variable aléatoire. On appelle fonction de répartition de X la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par, pour $x \in \mathbb{R}$,

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}).$$

La fonction F_X définit la loi de probabilité de la variable aléatoire X .

Définition 4.5. On dira que deux variables aléatoires X et Y ont même loi si elles ont la même fonction de répartition :

$$\forall x \in \mathbb{R} \quad F_X(x) = F_Y(x).$$

On note alors $X \stackrel{\mathcal{L}}{=} Y$.

Proposition 4.6. La fonction F_X est croissante, continue à droite et $\lim_{x \rightarrow +\infty} F_X(x) = 1$, $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

Une fonction f est continue à droite en a si $\lim_{\substack{x \rightarrow a \\ x > a}} f(x) = f(a)$.

Démonstration. • si $x \leq y$, on a $\{X \leq x\} \subset \{X \leq y\}$. Par conséquent, $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$. La fonction est bien croissante.

- si $(b_n)_{n \geq 1}$ est une suite croissante qui tend vers $+\infty$, alors la suite d'événements $\{X \leq b_n\}$ est croissante et donc

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X \leq b_n) = \mathbb{P}(\cup_n \{X \leq b_n\}) = \mathbb{P}(X < +\infty) = 1$$

si $(b_n)_{n \geq 1}$ est une suite décroissante qui tend vers $-\infty$, alors la suite d'événements $\{X \leq b_n\}$ est décroissante et donc

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X \leq b_n) = \mathbb{P}(\cap_n \{X \leq b_n\}) = \mathbb{P}(X \leq -\infty) = 0$$

- La continuité à droite est une conséquence de la propriété de continuité monotone de $\mathbb{P}$.

□

Proposition 4.7. Si F est une fonction telle que $F : \mathbb{R} \rightarrow [0, 1]$, croissante, continue à droite et telle que $\lim_{x \rightarrow +\infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$, alors il existe une variable aléatoire X telle que F soit la fonction de répartition de cette variable $F = F_X$.

Cette proposition dit qu'il existe une variable aléatoire telle que $F = F_X$, mais en général on ne sait pas la construire explicitement.

Exemple 4.3. 1. On joue à Pile ou Face avec une pièce bien équilibrée et X est le résultat du jeu. Alors

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{2} & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

Refaire le calcul si la pièce est biaisée de sorte que Pile apparaît 2 fois plus que Face : $\mathbb{P}(\text{Pile}) = 2/3$.

2. On lance un dé équilibré et X est le résultat du jeu, alors $F_X(x) = [x]/6$ pour $x \in [0, 6]$, où $[.]$ est la fonction partie entière.

3. On lance deux dés et on regarde le maximum des chiffres obtenu X . La fonction de répartition de X est

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1; \\ \frac{1}{36} & \text{si } x \in [1, 2[; \\ \frac{4}{36} = \frac{1}{9} & \text{si } x \in [1, 2[; \\ \frac{9}{36} = \frac{1}{4} & \text{si } x \in [2, 3[; \\ \frac{16}{36} = \frac{4}{9} & \text{si } x \in [3, 4[; \\ \frac{25}{36} & \text{si } x \in [4, 5[; \\ 1 & \text{si } x \geq 6; \end{cases}$$

b) Calcul de probabilités et fonction de répartition

Propriété 4.8. Soit $a, b \in \mathbb{R}$ avec $a \leq b$ et X une variable aléatoire. On note

$$F_X(a^-) = \lim_{\substack{x \rightarrow a \\ x < a}} F_X(x).$$

On a

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

Par conséquent, la fonction F_X est continue en a si et seulement si $\mathbb{P}(X = a) = 0$ (car alors $F_X(a^-) = F_X(a)$).

Démonstration. • On sait que si $E \subset F$, on a $\mathbb{P}(F \setminus E) = \mathbb{P}(F) - \mathbb{P}(E)$. Comme

$$\{a < X \leq b\} = \{X \leq b\} \setminus \{X \leq a\}.$$

on déduit

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

- On remarque que

$$\{X < a\} = \bigcup_{n \geq 1} \left\{ X \leq a - \frac{1}{n} \right\}$$

Comme la suite d'événements $(\{X \leq a - \frac{1}{n}\})_{n \geq 1}$ est croissante, on a

$$\mathbb{P}(X < a) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq a - \frac{1}{n}) = \lim_{n \rightarrow \infty} F_X(a - \frac{1}{n}) := F_X(a^-)$$

□

Exemple 4.4.

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ x/2 & \text{si } 0 \leq x < 1 \\ 2/3 & \text{si } 1 \leq x < 2 \\ 11/12 & \text{si } 2 \leq x < 3 \\ 1 - \frac{1}{4x} & \text{si } x \geq 3. \end{cases}$$

Alors $\mathbb{P}(X < 3) = F_X(3^-) = \frac{11}{12}$, $\mathbb{P}(X > \frac{1}{2}) = 1 - F_X(\frac{1}{2}) = 1 - \frac{1}{4} = \frac{3}{4}$. $P(X = 2) = 1/4$, $P(X = 1/2) = 0$.

4.2 Variables aléatoires discrètes

Une variable aléatoire X est dite discrète si l'ensemble de ses valeurs $\{X(\omega) : \omega \in \Omega\}$ est au plus dénombrable (i.e., fini ou infini dénombrable). Il est de la forme $\{x_1, \dots, x_n\}$ ou $\{x_1, x_2, \dots\}$.

Exemple 4.5. 1. X le résultat d'un lancer de dé à 6 faces, on a $X \in \{1, \dots, 6\}$.
 2. Y le nombre de Face obtenu sur 4 lancers d'une pièce, on a $Y \in \{0, 1, 2, 3, 4\}$.
 3. S la somme des chiffres obtenu lorsqu'on lance deux dés, on a $S \in \{2, \dots, 12\}$.
 4. On lance deux dés et on considère Z le nombre de lancers nécessaires pour obtenir un double « 6 », on a $Z \in \{1, 2, 3, \dots\}$.

a) Fonction de masse

Définition 4.9. Une fonction p est une fonction de masse sur les valeurs $\{x_1, x_2, \dots\}$ si

1. $p(x_i) > 0$, pour tout $i \geq 1$,
2. $\sum_{i \geq 1} p(x_i) = 1$.

On dit que l'ensemble $\{x_1, x_2, \dots\}$ est le support de la fonction de masse p .

Exemple 4.6. La fonction p définie par $p(-2) = 1/2$, $p(0) = c$, $p(3) = 0.4$ est une fonction de masse sur $\{-2, 0, 3\}$ si $c = 0.1$, car on doit avoir $p(-2) + p(0) + p(3) = 1$.

Exemple 4.7. Soit $\lambda > 0$. On considère la fonction p définie par $p(k) = c \frac{\lambda^k}{k!}$ pour $k \in \mathbb{N} = \{0, 1, 2, 3, \dots\}$. p est une fonction de masse sur $\mathbb{N}$ si $c > 0$ et

$$1 = \sum_{k=0}^{\infty} p(k) = c \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = ce^{\lambda}.$$

Par conséquent, $c = e^{-\lambda}$.

Définition 4.10. Soit X une variable discrète, avec X est à valeurs dans $\{x_1, x_2, \dots\}$ (les x_k étant distincts). La fonction de masse de la variable X est la fonction p_X définie par $p_X(x_k) := \mathbb{P}(X = x_k)$.

Remarque 4.11. On a bien $p_X(x_k) > 0$ et

$$\sum_{k \geq 1} p_X(x_k) = \sum_{k \geq 1} \mathbb{P}(X = x_k) = \mathbb{P}\left(X \in \bigcup_{k \geq 1} \{x_k\}\right) = \mathbb{P}(X \in \{x_1, x_2, \dots\}) = 1,$$

car $\{x_1, x_2, \dots\}$ est l'ensemble des valeurs de X .

Exemple 4.8. 1. On lance un dé et on regarde X le chiffre obtenu. Alors la fonction de masse de X est $p(k) = \frac{1}{6}$ pour $k \in \{1, \dots, 6\}$.

2. On lance 4 fois une pièce dont la probabilité de tomber sur Face est p et on regarde le nombre Y de Face obtenu. On a $\Omega = \{\omega = (\omega_1, \omega_2, \omega_3, \omega_4) : \omega_i \in \{0, 1\}\}$ (0 correspond à Pile et 1 correspond à Face). Alors $Y(\omega) = \sum_{i=1}^4 \omega_i$. Y est à valeurs dans $\{0, \dots, 4\}$, c'est une variable discrète.

Chaque lancer est une expérience de Bernoulli, et les lancers sont indépendants entre eux.. On a donc

$$p_Y(0) = (1-p)^4, \quad p_Y(1) = 4p(1-p)^3, \quad p_Y(2) = \binom{4}{2} p^2(1-p)^2, \\ p_Y(3) = \binom{4}{3} p^3(1-p), \quad p_Y(4) = p^4.$$

Si la pièce est bien équilibrée, $p = \frac{1}{2}$ et donc on a $p_Y(k) = \binom{4}{k} \left(\frac{1}{2}\right)^4$ pour $k \in \{0, \dots, 4\}$.

Ce raisonnement se généralise quand on considère n lancers. La fonction de masse du nombre de Face sur n lancers est $p_Y(k) = \binom{n}{k} p^k (1-p)^{n-k}$, où $\binom{n}{k}$ représente le nombre de façon de placer les k Face sur n lancers.

3. On lance deux dés et on regarde la fonction de masse de la somme S des chiffres obtenus. On doit calculer

k	2	3	4	5	6	7	8	9	10	11	12
$p_X(k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

4. Le nombre Z de lancers nécessaires pour obtenir un double « 6 », on a, pour $k \geq 1$ par indépendance entre les lancers,

$$p_Z(k) = \mathbb{P}(\underbrace{\text{pas de double « 6 », } \dots, \text{ pas de double « 6 »}}_{k-1 \text{ fois}}, \text{ avoir un double « 6 »})$$

$$= \left(1 - \frac{1}{36}\right)^{k-1} \frac{1}{36} = \left(\frac{35}{36}\right)^{k-1} \frac{1}{36}.$$

b) Lien entre fonction de masse et fonction de répartition

Soit X une variable discrète à valeurs dans $\{x_1, x_2, x_3, \dots\}$.

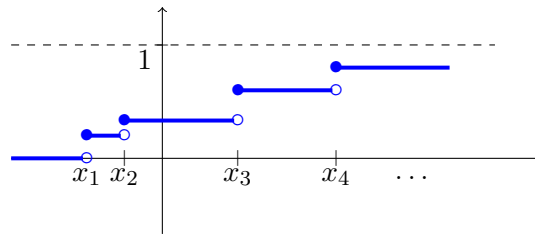
Soit $B \subset \mathbb{R}$. On veut calculer $\mathbb{P}(X \in B)$. On a

$$\begin{aligned} \mathbb{P}(X \in B) &= \mathbb{P}(X \in B \text{ et } X \in \{x_1, x_2, x_3, \dots\}) \\ &= \mathbb{P}(X \in (B \cap \{x_1, x_2, x_3, \dots\})) \\ &= \sum_{k: x_k \in B} p_X(x_k). \end{aligned}$$

En prenant $B =]-\infty, x]$, on a $x_k \in B \Leftrightarrow x_k \leq x$ et par conséquent, la fonction de répartition vérifie,

$$F_X(x) = \sum_{k: x_k \leq x} p_X(x_k).$$

La fonction de répartition d'une variable discrète X est constante par morceaux. On a



Réciproquement,

$$p_X(x_k) = F_X(x_k) - F_X(x_k^-).$$

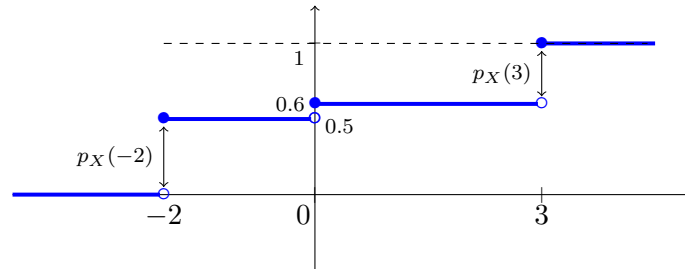
Conséquence, connaître la fonction masse permet de connaître la fonction de répartition et inversement. La fonction de masse caractérise donc la loi de probabilité de la variable X (i.e, connaître la fonction de masse suffit à connaître la loi de X).

Propriété 4.12. Deux variables aléatoires X et Y ont la même loi si leur ensemble de valeurs est identique et elles ont la même fonction de masse : $p_X = p_Y$.

Quand on cherche la loi d'une variable,

- d'abord, on cherche l'ensemble des valeurs de cette variable ;
- puis, on calcule soit la fonction de masse, soit la fonction de répartition.

Exemple 4.9. Soit X une variable de fonction de masse $p_X(-2) = 1/2$, $p_X(0) = 0.1$, $p_X(3) = 0.4$. Sa fonction de répartition F_X est



Exemple 4.10. On lance deux dés et on s'intéresse au plus grand chiffre obtenu Y . La variable Y est à valeurs dans $\{1, 2, \dots, 6\}$, donc $\mathbb{P}(Y = k) = 0$ pour $k \notin \{1, 2, \dots, 6\}$.

k	1	2	3	4	5	6
$p_Y(k)$	1/36	3/36	5/36	7/36	9/36	11/36
$F_Y(k)$	1/36	4/36	9/36	16/36	25/36	1

Exemple 4.11. Soit X une v.a. sur $\{-1, 0, 1, 2, \dots\}$ avec $p_X(k) = \left(\frac{1}{2}\right)^{k+2}$. p_X est bien une fonction de masse. On a

$$\begin{aligned}\mathbb{P}(X \leq 3) &= \mathbb{P}(X \in \{-1, 0, 1, 2, 3\}) = p_X(-1) + p_X(0) + p_X(1) + p_X(2) + p_X(3) \\ \mathbb{P}(|X| \leq 2) &= \mathbb{P}(X \in \{-1, 0, 1, 2\}) = p_X(-1) + p_X(0) + p_X(1) + p_X(2) \\ \mathbb{P}(X > 27) &= \mathbb{P}(X \in \{28, 29, \dots\}) = \sum_{k=28}^{+\infty} p_X(k) = \sum_{k=28}^{+\infty} \left(\frac{1}{2}\right)^{k+2} = \sum_{k=30}^{+\infty} \left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^{29}.\end{aligned}$$

La fonction de répartition de X est, pour $x \in [n, n+1[$ avec $n \geq -1$,

$$F_X(x) = \sum_{k=-1}^n \left(\frac{1}{2}\right)^{k+2} = \sum_{k=0}^{n+1} \left(\frac{1}{2}\right)^{k+1} = \frac{1}{2} \frac{1 - \left(\frac{1}{2}\right)^{n+2}}{1 - \frac{1}{2}} = 1 - \left(\frac{1}{2}\right)^{n+2}.$$

c) Loïs discrètes usuelles

Loi uniforme (discrète)

On dit qu'une v.a. X suit une loi uniforme discrète sur $\{x_1, \dots, x_n\}$, avec $x_1 < \dots < x_n$, si les événements $\{X = x_k\}$ sont équiprobables. Sa fonction de masse est donnée par

$$p_X(x_k) = \frac{1}{n}, \quad \forall k \in \{1, \dots, n\} \quad (n \text{ est le cardinal de l'ensemble des valeurs prises par } X).$$

On note alors $X \sim \text{Unif}\{x_1, \dots, x_n\}$.

Sa fonction de répartition s'écrit

$$F_X(x) = \begin{cases} 0 & \text{si } x < x_1 \\ \frac{k}{n} & \text{si } x \in [x_k, x_{k+1}[\\ 1 & \text{si } x \geq x_n. \end{cases}$$

Exemple 4.12. Le résultat d'un dé à 6 faces suit la loi uniforme sur $\{1, \dots, 6\}$.

Loi de Bernoulli, $\mathcal{B}er(p)$, avec $p \in]0, 1[$

Une variable aléatoire X de Bernoulli est une variable qui ne prend que deux valeurs : échec (auquel on associe la valeur 0) et succès (auquel on associe la valeur 1) d'une expérience. Cette expérience est appelée **épreuve de Bernoulli**.

X est à valeurs dans $\{0, 1\}$ et sa fonction de masse est donnée par :

$$p_X(1) = p \quad p_X(0) = 1 - p.$$

La fonction de répartition est

$$F_X(x) = (1 - p)\mathbb{1}_{[0, +\infty[}(x) + p\mathbb{1}_{[1, +\infty[}(x).$$

Exemple 4.13. Si E un événement alors $X = \mathbb{1}_E$ est une variable de Bernoulli.

Par exemple, E peut être l'événement "l'assuré a eu au moins un accident au cours de la dernière année".

Loi Binomiale, $\mathcal{B}(n, p)$, avec $p \in]0, 1[$, $n \in \mathbb{N}$, $n \geq 1$

La loi Binomiale est utilisée pour modéliser un « sondage avec remise ». C'est la loi du nombre de succès lorsqu'on renouvelle n fois de manière indépendante une épreuve de Bernoulli de paramètre p . On note X le nombre de succès obtenus à l'issue des n épreuves. Sa loi s'appelle loi Binomiale de paramètres n et p . La variable X est à valeurs dans $\{0, 1, \dots, n\}$ et on a

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ avec } k \in \{0, 1, \dots, n\}.$$

$\binom{n}{k}$ représente la nombre de façon de répartir les k succès sur n expériences, p^k est la probabilité d'avoir k succès sur k expériences et $(1 - p)^{n-k}$ est la probabilité d'avoir $n - k$ succès sur $n - k$ expériences.

Rappel de la formule du binôme de Newton :

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

On vérifie facilement avec la formule du binôme de Newton que

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

La fonction de répartition de la loi binomiale n'a pas d'expression simple

$$F_X(x) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \mathbb{1}_{k \leq x}.$$

Remarque 4.13 (Non fait en cours). Soit X le nombre de succès sur n épreuves de Bernoulli. On peut écrire X à l'aide des résultats de chaque épreuve de Bernoulli. On note X_i le résultat de la $i^{\text{ème}}$ expérience :

$$X_i = \begin{cases} 1 & \text{si la } i^{\text{ème}} \text{ expérience est réussie} \\ 0 & \text{si la } i^{\text{ème}} \text{ expérience est un échec} \end{cases}$$

On a alors $X = X_1 + \dots + X_n$, avec $X_i \sim \text{Ber}(p)$ indépendantes.

Exemple 4.14. Un portefeuille d'assurance contient 1000 clients. On s'intéresse aux clients qui ont eu au moins un accident dans l'année.

Loi Géométrique, $\mathcal{G}(p)$, avec $p \in]0, 1[$

La loi géométrique est la loi du premier succès.

Exemple 4.15. On lance une pièce truquée jusqu'à ce qu'on obtienne une fois « Face ». On note p la probabilité de tomber sur « Face ». On note X le nombre de lancers nécessaires pour avoir un succès.

La variable X est à valeurs dans $\{1, 2, \dots\} = \mathbb{N}^*$ et, pour $k \geq 1$,

$$p_X(k) = p(1-p)^{k-1}.$$

Rappel sur la série géométrique : Pour tout $q \neq 1$ on a,

$$1 + q^2 + \dots + q^n = \frac{1 - q^{n+1}}{1 - q}$$

et si $|q| < 1$,

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1 - q}.$$

On peut calculer explicitement la fonction de répartition : $F_X(x) = 0$ si $x < 1$ et

$$F(x) = 1 - (1-p)^k \quad \text{si } x \in [k, k+1[.$$

La loi géométrique possède la propriété d'absence de mémoire : $\forall k \geq 1, n \geq 0$

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

Loi Binomiale négative, $\text{BinNeg}(r, p)$, avec $p \in]0, 1[, r \geq 1$

C'est une généralisation de la loi géométrique. La variable X représente le nombre d'épreuve de Bernoulli à effectuer pour obtenir le $r^{\text{ème}}$ succès.

La variable X est à valeurs dans $\{r, r+1, \dots\}$ et, pour $k \geq r$,

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r},$$

où $\binom{k-1}{r-1}$ est le nombre de façon de répartir $r-1$ succès parmi $k-1$ expériences (la dernière expérience étant forcément un succès), p^r est les probabilité d'avoir que des succès sur r expériences et p^{k-r} est la probabilité d'avoir des échecs sur $k-r$ expériences.

Loi de Poisson, $\mathcal{Poi}(\lambda)$, avec $\lambda > 0$

La loi de Poisson est utilisée pour modéliser le comptage d'événements rares, c'est à dire des événements ayant une faible probabilité de réalisation : maladies rares, accidents, pannes, tremblements de terre,...

Exemple 4.16. Le nombre d'accident d'un portefeuille en assurance est souvent modélisé par une loi de Poisson.

La loi est donnée par $p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ avec $k \in \mathbb{N}$.

Rappel sur la série exponentielle :

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Sa fonction de répartition n'a pas de formule simple

$$F_X(x) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \mathbf{1}_{k \leq x} = \sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k}{k!} e^{-\lambda},$$

où $\lfloor x \rfloor$ est la partie entière de x .

Autre loi : loi hypergéométrique, avec $n, m, N \in \mathbb{N}$, avec $n, m \leq N$

(non fait en cours)

Cette loi permet de modéliser un sondage sans remise. On tire sans remise n boules dans une urne contenant N boules dont m sont Rouges et $N - m$ sont bleues.

On s'intéresse au nombre X de boule Rouge tirées. Cette variable est à valeurs dans $\{0, \dots, \max(n, m)\}$ et

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

4.3 Variables aléatoires continues

Considérons la durée de vie d'une ampoule. On conçoit facilement que la probabilité que cette durée de vie vaille exactement une certaine valeur est nulle. La fonction de répartition d'une telle variable est par conséquent continue.

On étudiera dans le cours *Probabilité 1* seulement des variables qui sont soit discrètes, soit continues. Cependant, il existe des variables plus complexes (ni discrètes, ni continues).

Définition 4.14. Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est appelée *densité de probabilité* si

1. pour tout $x \in \mathbb{R}$, $f(x) \geq 0$.

2. $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Exemple 4.17. La fonction $f(x) = \frac{x}{8} \mathbb{1}_{[0,4]}(x) = \begin{cases} \frac{x}{8} & \text{si } 0 \leq x \leq 4 \\ 0 & \text{sinon} \end{cases}$ est une densité.

Définition 4.15. Une variable aléatoire X est à **densité**, ou **continue**, s'il existe une fonction de densité f_X définie sur $\mathbb{R}$ telle que la fonction de répartition de X s'écrit

$$\forall x \in \mathbb{R} \quad F_X(x) = \int_{-\infty}^x f_X(t)dt.$$

La fonction f_X est appelée la (fonction de) densité de X .

Définition 4.16. On appelle le support de la loi d'une v.a. aléatoire X continue de densité f_X comme l'ensemble $S \subset \mathbb{R}$ des réels x tels que $f_X(x) > 0$.

Le support peut être vu comme l'ensemble des valeurs possibles de la v.a. X .

a) Propriétés

On a vu que $\mathbb{P}(X \in]a, b]) = F_X(b) - F_X(a)$. Par conséquent, pour une variable à densité X , on a

$$\mathbb{P}(X \in]a, b]) = \int_{-\infty}^b f_X(t)dx - \int_{-\infty}^a f_X(t)dx = \int_a^b f_X(x)dx$$

ce qui correspond à l'aire de la surface comprise entre la courbe de f_X et l'axe des abscisses sur l'intervalle $[a, b]$.

De manière générale, si $B \subset \mathbb{R}$ est un sous-ensemble de $\mathbb{R}$, on a

$$\mathbb{P}(X \in B) = \int_{-\infty}^{+\infty} f_X(t) \mathbb{1}_{x \in B} dt = \int_B f_X(x)dx.$$

Propriété 4.17. Soit X une variable aléatoire à densité.

Alors, pour tout $a \in \mathbb{R}$,

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

Par conséquent, la fonction de répartition d'une variable à densité est continue.

Démonstration. Soit $x \in \mathbb{R}$. On considère l'intervalle réduit à un point $B = \{a\}$. On a

$$\mathbb{P}(X = a) = \mathbb{P}(X \in B) = \int_a^a f(y)dy = 0.$$

□

Comme $\mathbb{P}(X = a) = 0 = \mathbb{P}(X = b)$. Si X v.a. à densité, on a

$$\mathbb{P}(X \in]a, b]) = \mathbb{P}(X \in]a, b[) = \mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in [a, b[).$$

Proposition 4.18. Soit X une variable aléatoire de fonction de répartition F_X . Si F_X est continue sur $\mathbb{R}$ et dérivable sur $\mathbb{R}$ (sauf peut-être en un nombre fini de points), alors X est une variable à densité f_X avec

$$f_X(x) = F'_X(x).$$

Proposition 4.19. Deux variables continues X et Y ont la même loi si et seulement si elles ont la même densité :

$$f_X(x) = f_Y(x) \quad \forall x \in \mathbb{R}.$$

Exemple 4.18. 1. Soit X une v.a. de densité $f(x) = \frac{x}{8} \mathbb{1}_{[0,4]}$.

On a

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2}{16} & \text{si } x \in [0, 4[\\ 1 & \text{si } x \geq 4. \end{cases}$$

Par conséquent,

$$\mathbb{P}(X < 3) = \frac{9}{16}, \quad \mathbb{P}(X \in [1, 2]) = \frac{4}{16} - \frac{1}{16} = \frac{3}{16}, \quad \mathbb{P}(X \geq 15) = 0.$$

2. Soit X une v.a. de densité $f_X(x) = ce^{-2(x-1)} \mathbb{1}_{x>1}$.

f_X est bien une densité si

$$1 = \int_{-\infty}^{+\infty} f_X(x) dx = c \int_1^{\infty} e^{-2(x-1)} dx = c \left[\frac{1}{2} e^{-2x} \right]_0^{+\infty} = \frac{c}{2},$$

donc $c = 2$. On a alors

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ 1 - e^{-2x} & \text{si } x \geq 1. \end{cases}$$

Par conséquent,

$$\mathbb{P}(X < 3) = 1 - e^{-6}, \quad \mathbb{P}(|X| \leq 2) = \mathbb{P}(X \in [1, 2]) = e^{-2} - e^{-4}, \quad \mathbb{P}(X \geq 15) = e^{-30}.$$

3. Soit X une v.a. de densité $f(x) = \frac{1}{8} \mathbb{1}_{[-4,0]} + e^{-2x} \mathbb{1}_{x>0}$.

On a

$$F_X(x) = \begin{cases} 0 & \text{si } x < -4 \\ \frac{x+4}{8} & \text{si } x \in [-4, 0[\\ 1 - \frac{1}{2} e^{-2x} & \text{si } x \geq 0. \end{cases}$$

Par conséquent,

$$\begin{aligned} \mathbb{P}(X < 3) &= 1 - \frac{1}{2} e^{-6}, \\ \mathbb{P}(|X| \leq 2) &= \mathbb{P}(-2 \leq X \leq 2) = F_X(2) - F_X(-2) = \frac{3}{4} - \frac{1}{2} e^{-4}, \\ \mathbb{P}(X \geq 15) &= \frac{1}{2} e^{-30}. \end{aligned}$$

b) Loi d'une v.a. transformée

Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ une fonction et X une v.a. à densité f_X connue. On s'intéresse à la loi de la v.a. $Y = h(X) : \forall \omega \in \Omega, Y(\omega) = h(X(\omega))$. On peut par exemple souhaiter connaître la loi de $Y = X^2$, $Y = e^X$, $Y = \sqrt{X^2 + X + 1}$, ...

Dans le cas discret, si X est une v.a. discrète alors Y est une v.a. discrète. Mais si X est à densité, la variable Y n'est pas forcément à densité. Y peut-être à densité, ou discrète, ou plus compliquée.

Méthode pour trouver la loi de Y

1. Identifier l'ensemble des valeurs de Y .
2. Chercher la fonction de répartition F_Y de Y (ou la fonction de masse si Y est discrète).
Si F_Y est continue et dérivable (sauf peut-être en un nombre fini de points), alors Y est à densité obtenue en dérivant la fonction de répartition de Y .

$$f_Y(y) = F'_Y(y) \quad \text{avec} \quad F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(h(X) \leq y).$$

Si X est une v.a. à densité et h une fonction telle que $Y = h(X)$ soit à densité, alors on peut donc exprimer la densité f_Y de Y à l'aide de la densité f_X de X et de la fonction h .

Exemple 4.19. Soit X une v.a. de densité $f(x) = \frac{x}{8} \mathbb{1}_{[0,4]}$. On a calculé sa fonction de répartition :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2}{16} & \text{si } x \in [0, 4[\\ 1 & \text{si } x \geq 4. \end{cases}$$

1. Si on prend la fonction $h(x) = \mathbb{1}_{[0,1]}(x)$, on a $Y = \mathbb{1}_{X \in [0,1]} = \begin{cases} 1 & \text{si } X \in [0, 1] \\ 0 & \text{sinon.} \end{cases}$

La variable Y est à valeurs dans $\{0, 1\}$. C'est une variable de Bernoulli. On a

$$\mathbb{P}(Y = 1) = \mathbb{P}(X \in [0, 1]) = \int_0^1 \frac{x}{8} dx = \frac{1}{16}.$$

Donc $Y \sim \text{Ber}(1/16)$.

2. Si on prend $h(x) = [x]$, la fonction partie entière.
 $Y = [X]$ est une variable discrète à valeurs dans $\{0, 1, 2, 3\}$. Sa fonction de masse est donnée par, $k \in \{0, 1, 2, 3\}$,

$$p_Y(k) = \mathbb{P}(X \in [k, k+1]) = F_X(k+1) - F_X(k) = \frac{2k+1}{16}.$$

Par conséquent, $p_Y(0) = \frac{1}{16}$, $p_Y(1) = \frac{3}{16}$, $p_Y(2) = \frac{5}{16}$ et $p_Y(3) = \frac{7}{16}$.

3. Si $h(x) = ax$.

Si $a > 0$, $Y = aX$ est à valeurs dans $[0, 4a]$. On a donc $F_Y(y) = 0$ si $y < 0$ et $F_Y(y) = 1$ si $y > 4a$. Soit $y \in [0, 4a]$, on a

$$F_Y(y) = \mathbb{P}(aX \leq y) = \mathbb{P}(X \leq y/a) = \frac{y^2}{16a^2}.$$

F_Y est continue sur $\mathbb{R}$ car $F_Y(0^-) = 0 = F_Y(0)$ et $\lim_{y \rightarrow 4a^-} F_Y(y) = 1 = F_Y(4a)$. Elle est dérivable sur $\mathbb{R} \setminus \{0, 4a\}$. Donc la densité de Y est

$$f_Y(y) = F'_Y(y) = \frac{y}{8a^2} \mathbb{1}_{[0, 4a]}(y).$$

Si $a < 0$, Y est à valeurs dans $[4a, 0]$. Soit $y \in [4a, 0]$, on a

$$F_Y(y) = \mathbb{P}(aX \leq y) = \mathbb{P}(X \geq y/a) = 1 - \frac{y^2}{16a^2}.$$

F_Y est continue sur $\mathbb{R}$ car $F_Y(4a^-) = 0 = F(4a)$ et $\lim_{y \rightarrow 0^-} F_Y(y) = 1 = F_Y(0)$. Elle est dérivable sur $\mathbb{R} \setminus \{0, 4a\}$. Donc la densité de Y est

$$f_Y(y) = -\frac{y}{8a^2} \mathbb{1}_{[4a, 0]}(y).$$

4. Si $h(x) = e^x$.

$Y = e^X$ est à valeurs dans $[1, e^4]$. On a donc $F_Y(y) = 0$ si $y < 1$ et $F_Y(y) = 1$ si $y > e^4$. Soit $y \in [1, e^4]$, comme $x \mapsto \ln(x)$ est croissante

$$F_Y(y) = \mathbb{P}(e^X \leq y) = \mathbb{P}(X \leq \ln(y)) = \frac{\ln^2(y)}{16}.$$

La fonction est continue sur $\mathbb{R}$ ($F_Y(1) = 0 = F_Y(1^-)$ et $F_Z(e^4) = 1 = F_Y(e^4)$) et dérivable sur $\mathbb{R} \setminus \{1, e^4\}$, donc la densité de Y est

$$f_Y(y) = \frac{\ln(y)}{8y} \mathbb{1}_{[1, e^4]}(y).$$

5. Si $h(x) = (x - 2)^2$.

La variable $Y = (X - 2)^2$ est à valeurs dans $[0, 4]$. On a donc $F_Y(y) = 0$ si $y < 0$ et $F_Y(y) = 1$ si $y > 4$. Soit $y \in [0, 4]$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}((X - 2)^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X - 2 \leq \sqrt{y}) = \mathbb{P}(2 - \sqrt{y} \leq X \leq 2 + \sqrt{y}) \\ &= F_X(2 + \sqrt{y}) - F_X(2 - \sqrt{y}) = \frac{(2 + \sqrt{y})^2}{16} - \frac{(2 - \sqrt{y})^2}{16} = \frac{\sqrt{y}}{2}. \end{aligned}$$

La fonction est continue sur $\mathbb{R}$ ($F_Y(0) = 0 = F_Y(0^-)$ et $F_Y(4) = 1 = F_Y(4^-)$) et dérivable sur $\mathbb{R} \setminus \{0, 4\}$, donc la densité de Y est

$$f_Y(y) = \frac{1}{4\sqrt{y}} \mathbb{1}_{]0, 4[}(y).$$

c) Lois à densité usuelles

Loi Uniforme, $\mathcal{U}([a, b])$, avec $a, b \in \mathbb{R}$, $a < b$.

Cette loi est l'analogie continu de l'équiprobabilité dans le cas discret. Elle permet de modéliser le tirage d'un nombre aléatoire dans l'intervalle $[a, b]$.

Exemple 4.20. Lors d'une étude du comportement animal, on a relâché des oiseaux dont l'orientation a été rendue très difficile. On s'attend alors à ce que les oiseaux choisissent au hasard leur direction. On peut modéliser la direction prise par un oiseau de la façon suivante. On considère X l'angle entre le nord et la direction prise par l'oiseau (selon le sens des aiguilles d'une montre). La variable X suit une loi uniforme entre 0 et 360 degrés.

On note $X \sim \mathcal{U}([a, b])$ et le support de X est $[a, b]$. Densité :

$$\begin{aligned} f(x) &= \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) \\ &= \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

Fonction de répartition :

$$\begin{aligned} F(x) &= \frac{x-a}{b-a} \mathbb{1}_{[a,b]}(x) + \mathbb{1}_{[b,+\infty)}(x) \\ &= \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b) \\ 1 & \text{si } x \geq b \end{cases} \end{aligned}$$

Loi Exponentielle, $\mathcal{E}(\lambda)$, avec $\lambda > 0$ un réel.

Cette densité de probabilité permet en général de modéliser des durées de vie d'êtres non soumis au vieillissement (par exemple, la durée de vie d'une bactérie) ou des temps d'attente (par exemple, le temps d'attente entre deux sinistres).

- Exemple 4.21.**
1. Dans une substance radioactive, la désintégration des noyaux se fait de façon spontanée. Le nombre de désintégration sur un intervalle de temps fixé suit une loi de Poisson. Par contre le temps d'attente entre deux désintégrations est modélisé par une loi exponentielle.
 2. Dans un portefeuille d'assurance, le nombre de sinistres sur une année est souvent modélisé par une loi de Poisson. Le temps d'attente entre deux sinistres est modélisé par une loi exponentielle.

On note $X \sim \mathcal{E}(\lambda)$ et le support de X est $]0, +\infty[$.

Densité :

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x} \mathbb{1}_{x>0} \\ &= \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0. \end{cases} \end{aligned}$$

Fonction de répartition :

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x} & \text{si } x \geq 0. \end{cases}$$

La loi exponentielle est la seule loi continue qui vérifie la propriété d'absence de mémoire : Si $X \sim \mathcal{E}(\lambda)$, alors pour tout $s, t > 0$ $\mathbb{P}(X > t + s | X > s) = \mathbb{P}(X > s) = e^{-\lambda t}$.

La loi exponentielle est celle de la mortalité des êtres qui ne seraient pas soumis au vieillissement : à chaque moment ils ont la même probabilité de mourir dans l'unité de temps qu'il leur reste, quelque soit leur âge.

Loi Gamma, $\text{Gamma}(\alpha, \lambda)$, avec $\alpha, \lambda > 0$ réels.

α est appelé paramètre de forme (shape) et λ paramètre d'échelle (scale).

Cette densité de probabilité est notamment utilisée pour modéliser le coût d'un sinistre en assurance.

On note $X \sim \text{Gamma}(\alpha, \lambda)$ et le support de X est $]0, +\infty[$.

Densité :

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{x>0}$$

où la fonction $\Gamma(\cdot)$ est définie par

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx.$$

On remarque que si $\alpha = 1$, $\Gamma(1) = 1$ et donc on retrouve la loi exponentielle.

Propriété 4.20. Pour tout $\alpha > 0$, $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$.

On déduit que si $n \in \mathbb{N}$, $n \geq 1$, $\Gamma(n) = (n-1)!$.

Démonstration. Par intégration par partie avec $u = x^\alpha$ et $v' = e^{-x}$, on a

$$\begin{aligned} \Gamma(\alpha + 1) &= \int_0^{+\infty} x^\alpha e^{-x} dx = [-x^\alpha e^{-x}]_0^\infty + \alpha \int_0^{+\infty} x^{\alpha-1} e^{-x} dx \\ &= \alpha \Gamma(\alpha). \end{aligned}$$

On déduit par induction, que pour tout $n \in \mathbb{N}$, $n \geq 1$, $\Gamma(n) = (n-1)!$. □

La fonction de répartition de X n'a pas d'expression simple en général :

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \int_0^x \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} dy & \text{sinon.} \end{cases}$$

Loi Normale (ou loi Gaussienne), $\mathcal{N}(\mu, \sigma^2)$, avec $\mu \in \mathbb{R}$, $\sigma > 0$ réel.

La loi Normale est une loi centrale dans la théorie des probabilités. Elle est notamment très utilisée en statistique. Une grandeur influencée par un grand nombre de paramètres indépendants est souvent modélisée par une loi normale (par exemple, les erreurs de mesures lors d'une expérience).

On note $X \sim \mathcal{N}(\mu, \sigma^2)$ et le support de X est $\mathbb{R}$.

Densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ avec } x \in \mathbb{R}.$$

La loi $\mathcal{N}(0, 1)$ est appelée loi normale centrée réduite (ou loi normale standard).

La fonction de répartition de X n'a pas d'expression simple. On note en général Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Propriété 4.21. soit Z une variable de loi $\mathcal{N}(0, 1)$. Alors sa fonction de répartition vérifie

1. $\Phi(0) = \frac{1}{2}$;
2. $\Phi(-x) = 1 - \Phi(x)$;

3. Si $Z \sim \mathcal{N}(0, 1)$, alors $\mathbb{P}(|Z| < x) = 2\Phi(x) - 1$.

Démonstration. 1. Par parité de la fonction $t \mapsto e^{-\frac{t^2}{2}}$, on a

$$1 = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 2 \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

D'où le résultat.

2. On utilise le changement de variable $z = -t$,

$$\Phi(-x) = \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1 - \Phi(x).$$

3. On a $\mathbb{P}(|Z| < x) = \mathbb{P}(-x < Z < x) = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$. □

Propriété 4.22.

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Réciproquement, si $Z \sim \mathcal{N}(0, 1)$, alors $X = \sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$.

Démonstration. Supposons que $X \sim \mathcal{N}(\mu, \sigma^2)$. Soit $x \in \mathbb{R}$, en utilisant le changement de variables $t = \frac{y-\mu}{\sigma}$, on a

$$F_Z(z) = \mathbb{P}(X \leq \sigma z + \mu) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\sigma z + \mu} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt = \Phi(z).$$

Par conséquent, si $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mathbb{P}\left(\frac{X-\mu}{\sigma} \leq z\right) = \Phi(z)$ et donc $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$. Et réciproquement. □

Remarque 4.23. On remarque sur la table de la loi gaussienne que si $Z \sim \mathcal{N}(0, 1)$, alors $\mathbb{P}(|Z| \leq 4) \simeq 1$. Donc même si Z est à valeurs dans tout $\mathbb{R}$, avec probabilité proche de 1 ses valeurs sont comprises entre -4 et 4 .

Autres lois à densités (il n'est pas demandé de les connaître.)

Loi Log-Normale, $LN(m, \sigma^2)$, avec $m \in \mathbb{R}$, $\sigma > 0$ réel.

Une variable X suit une loi Log-Normale $LN(m, \sigma^2)$ si $\ln(X)$ suit une loi normale $\mathcal{N}(m, \sigma^2)$.

Cette loi permet de modéliser par exemple le temps de survie des bactéries en présence de désinfectant, le dosage de certains médicaments, la sévérité d'un sinistre en assurance, etc.

Densité :

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln(x) - m)^2}{2\sigma^2}} \text{ avec } x > 0.$$

Loi de Weibull, $W(a, b)$, avec $a > 0$, $b > 0$ réels.

La loi de Weibull est utilisée en démographie pour modéliser le vieillissement et en épidémiologie pour modéliser la distribution de probabilité de la durée d'incubation d'une maladie infectieuse.

Densité :

$$f(x) = \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} \exp\left(-\left(\frac{x}{b}\right)^a\right) \text{ avec } x > 0.$$

Fonction de répartition :

$$F(x) = 1 - \exp\left(-\left(\frac{x}{b}\right)^a\right) \text{ avec } x > 0.$$

On constate que pour $a = 1$, on retrouve la loi exponentielle. La loi exponentielle est celle de la mortalité des être vivants qui ne seraient pas soumis au vieillissement : à chaque moment ils ont la même probabilité de mourir dans l'unité de temps qu'il leur reste, quelque soit leur âge. Plus a est grand plus le vieillissement se fait pesant (la mortalité augmente avec l'âge). Le cas $a < 1$ correspond à un monde dans lequel plus on vieillirait, moins forte serait la probabilité de mourir dans l'unité de temps qui vient.

Loi Beta, $\text{beta}(a, b)$, avec $a, b > 0$ réels.

On note $X \sim \text{beta}(a, b)$. Son support est $]0, 1[$ et sa densité est

$$f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{]0,1[}(x).$$

Loi de Pareto, avec $\alpha > 0$ et $\lambda > 0$

Densité

$$f_X(x) = f(x; \alpha, \lambda) = \frac{\alpha \lambda^\alpha}{(x + \lambda)^{\alpha+1}} \mathbb{1}_{x>0}$$

Fonction de répartition

$$F_X(x; \alpha, \lambda) = \begin{cases} 1 - \left(\frac{\lambda}{x+\lambda}\right)^\alpha & \text{si } x \geq 0 \\ 0 & \text{si } x < 0. \end{cases}$$

FONCTION DE RÉPARTITION DE LA LOI NORMALE STANDARD

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

z	0.841	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
$\Phi(z)$	0.8000	0.9000	0.9500	0.9750	0.9800	0.9900	0.9950	0.9975	0.9990	0.9995

Chapitre 5

Espérance d'une variable aléatoire

Sommaire

5.1	Définition de l'espérance	51
5.2	Espérance d'une transformation, Variance	54
a)	Cas général	54
b)	Variance	55
c)	Propriétés de l'espérance et de la variance	55
5.3	Fonctions génératrices	57
a)	Fonction génératrice des moments	57
b)	Fonction génératrice des probabilités	59
5.4	Formulaire	60

5.1 Définition de l'espérance

L'espérance mathématique représente la valeur moyenne pondérée de la variable aléatoire en tenant compte des probabilités d'apparition de chaque réalisation dans le cas discret, ou de la densité dans le cas continu.

Définition 5.1.

Variables discrètes

(a) Pour une variable prenant un nombre fini de valeurs :

Soit X une v.a. discrète à valeurs dans $\{x_1, x_2, \dots, x_n\}$ de fonction de masse p_X . On définit l'**espérance** de X comme le réel

$$\mathbb{E}[X] = \sum_{k=1}^n x_k p_X(x_k).$$

(b) Pour une variable prenant un nombre dénombrable de valeurs :

Soit X une v.a. discrète à valeurs dans $\{x_1, x_2, \dots\}$ de fonction de masse p_X . Si X est à valeurs positives ou si

$$\sum_{k=1}^{\infty} |x_k| p_X(x_k) < +\infty,$$

on définit l'espérance de X comme

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} x_k p_X(x_k).$$

Variables continues

Soit X une v.a. de densité f_X . Si X est à valeurs positives ou si $\int_{-\infty}^{+\infty} |x|f_X(x)dx < +\infty$, on définit l'**espérance** de X comme le réel

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

Si $\mathbb{E}[X] = 0$, on dit que la variable est centrée.

L'espérance est un réel (ce n'est pas une variable aléatoire!). Si $a \in \mathbb{R}$, alors $\mathbb{E}[a] = a$.

Exemple 5.1. 1. On lance un dé et on regarde X le chiffre obtenu. Alors $\mathbb{E}[X] = \sum_{k=1}^6 k p_X(k) = \frac{1}{6} \sum_{k=1}^6 k = 3.5$.

2. On lance 4 fois une pièce dont la probabilité de tomber sur Face est $1/2$. Alors

$$\begin{aligned} \mathbb{E}[X] &= 0 \times \left(\frac{1}{2}\right)^4 + 1 \times 4 \left(\frac{1}{2}\right)^4 + 2 \times 6 \left(\frac{1}{2}\right)^4 + 3 \times 4 \left(\frac{1}{2}\right)^4 + 4 \times \left(\frac{1}{2}\right)^4 \\ \mathbb{E}[X] &= 2 \end{aligned}$$

3. Si $X \sim \mathcal{B}(p)$, $\mathbb{E}[X] = p \times 1 + (1-p) \times 0 = p$.

4. Si $X \sim \text{Bin}(n, p)$, alors

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} = np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} = np. \end{aligned}$$

5. Si $X \sim \text{Pois}(\lambda)$.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda.$$

6. Si X est une v.a. de densité $f(x) = \frac{x}{8} \mathbb{1}_{[0,4]}$, on a

$$\mathbb{E}[X] = \int_0^4 \frac{x^2}{8} dx = \frac{8}{3}.$$

7. Si $X \sim \mathcal{U}[a, b]$, on a

$$\mathbb{E}[X] = \frac{1}{b-a} \int_a^b x dx = \frac{a+b}{2}.$$

8. Si $X \sim \mathcal{Exp}(\lambda)$, on a par intégration par parties

$$\mathbb{E}[X] = \int_0^\infty \lambda x e^{-\lambda x} dx = \left[-x e^{-\lambda x} \right]_0^\infty + \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Autre méthode : On rappelle que $\lambda^2 x e^{-\lambda x} \mathbb{1}_{x>0}$ est la densité de la loi Gamma(2, λ), donc son intégrale est égale à 1. Par conséquent,

$$\mathbb{E}[X] = \int_0^\infty \lambda x e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^\infty \lambda^2 x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

9. Si $X \sim \text{Gamma}(\alpha, \lambda)$, on a

$$\mathbb{E}[X] = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\lambda x} = \frac{\Gamma(\alpha+1)}{\lambda \Gamma(\alpha)} \int_0^\infty \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\lambda x}$$

On reconnaît à l'intérieur de l'intégrale la densité de la loi Gamma($\alpha+1$, λ). Comme $\Gamma(\alpha+1) = \alpha \Gamma(\alpha)$, on déduit que

$$\mathbb{E}[X] = \frac{\alpha}{\lambda}.$$

10. Si $X \sim \mathcal{N}(0, 1)$, La variable X est à valeurs dans $\mathbb{R}$. On remarque que $\int_{-\infty}^{+\infty} |x| e^{-x^2/2} dx < +\infty$ grâce à la fonction exponentielle. De plus, comme la fonction est impaire, on a

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{x^2/2} dx = 0.$$

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, comme dans le cas précédent on a $\int_{-\infty}^{+\infty} |x| e^{-(x-\mu)^2/2\sigma^2} dx < +\infty$. On a

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{+\infty} \frac{x-\mu}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{+\infty} \frac{y}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy + \mu \\ &= \mu, \end{aligned}$$

en utilisant le changement de variable $y = x - \mu$, le calcul précédent et le fait que l'intégrale d'une densité est égale à 1.

11. si X est une v.a. de densité $f(x) = \frac{1}{2} \mathbb{1}_{[-1,0]} + e^{-2x} \mathbb{1}_{x>0}$.

$$\mathbb{E}[X] = \frac{1}{2} \int_{-1}^0 x dx + \int_0^\infty x e^{-2x} dx = -\frac{1}{4} + \frac{1}{4} = 0.$$

Exemple 5.2. Soit $E \subset \Omega$ un événement. On considère la variable $X = \mathbb{1}_E$. On a

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \in E \\ 0 & \text{si } \omega \notin E. \end{cases}$$

La variable X est à valeurs dans $\{0, 1\}$, elle suit la loi de bernoulli $\mathcal{B}(\mathbb{P}(E))$ car $p_X(1) = \mathbb{P}(E)$. Par conséquent, $\mathbb{E}[X] = \mathbb{P}(E)$.

5.2 Espérance d'une transformation, Variance

On a vu précédemment que si X est une v.a. et $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction, alors $Y = g(X)$ est une v.a. On peut relier l'espérance de Y avec l'espérance de X sans avoir besoin de chercher la loi de Y .

a) Cas général

Proposition 5.2 (Espérance d'une transformation d'une v.a.).

1. **Cas discret** : si X une v.a. discrète à valeurs dans $\{x_1, x_2, \dots\}$. Alors la variable $Y = g(X)$ est à valeurs dans $\{g(x_1), g(x_2), \dots\}$ (attention, on peut avoir $g(x_i) = g(x_j)$ pour $i \neq j$).

(a) Si X est une variable à valeur dans $\{x_1, \dots, x_n\}$, l'espérance de $Y = g(X)$ est

$$\mathbb{E}[g(X)] = \sum_{k=1}^n g(x_k) p_X(x_k).$$

(b) Si X est une variable à valeur dans $\{x_1, x_2, \dots\}$ et si g est positive ou si $\sum_{k=1}^{\infty} |g(x_k)| p_X(x_k) < \infty$, alors l'espérance de $Y = g(X)$ est

$$\mathbb{E}[g(X)] = \sum_{k=1}^{\infty} g(x_k) p_X(x_k).$$

2. **Cas continu** : Si X une v.a. de densité f_X , alors

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx,$$

si g est à valeurs positives ou si $\int_{-\infty}^{+\infty} |g(x)| f_X(x) dx < +\infty$.

Conséquence 5.3.

$$\mathbb{E}[X^2] = \sum_{k=1}^{\infty} (x_k)^2 p_X(x_k), \quad \mathbb{E}[e^X] = \sum_{k=1}^{\infty} e^{x_k} p_X(x_k).$$

Exemple 5.3. 1. On considère une v.a. X de fonction de masse $p(-2) = 1/2$, $p(0) = 0.1$, $p(3) = 0.4$.
Alors

$$\mathbb{E}[X] = -2 \times \frac{1}{2} + 0 \times 0.1 + 3 \times 0.4 = 0.2 = 0.2$$

$$\mathbb{E}[X^2] = (-2)^2 \times \frac{1}{2} + 0^2 \times 0.1 + 3^2 \times 0.4 = 0.2 = 5.6$$

$$\mathbb{E}[e^X] = e^{-2} \times \frac{1}{2} + e^0 \times 0.1 + e^3 \times 0.4 = 0.2 = 8.202.$$

2. Soit X de densité $f_X(x) = \frac{x}{8} \mathbf{1}_{[0,4]}(x)$. On a

$$\mathbb{E}[X^2] = \int_0^4 x^2 \frac{x}{8} dx = \frac{1}{8 \times 4} [x^4]_0^4 = 8$$

$$\mathbb{E}[e^X] = \int_0^4 e^x \frac{x}{8} dx = \frac{1}{8} [(x-1)e^x]_0^4 = \frac{1}{8} (3e^4 + 1).$$

b) Variance

Définition 5.4. Soit X une variable aléatoire. Soit $n \geq 1$.

Le **moment d'ordre n** de X est défini comme $\mathbb{E}[X^n]$. Donc, dans le cas discret

$$\mathbb{E}[X^n] = \sum_{k \geq 1} (x_k)^n p_X(x_k)$$

et dans le cas continu

$$\mathbb{E}[X^n] = \int_{-\infty}^{+\infty} x^n f_X(x) dx.$$

Définition 5.5. La **variance** de X est définie par

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

Dans le cas discret, on a

$$\text{Var}(X) = \sum_k (x_k - \mathbb{E}[X])^2 p_X(x_k)$$

et dans le cas continu

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 f_X(x) dx.$$

L'**écart-type** $\sigma_X = \sqrt{\text{Var}(X)}$ est la distance moyenne de la variable à sa moyenne. C'est une mesure de dispersion de la variable aléatoire.

Si $\text{Var}(X) = 1$, on dit que la variable X est réduite.

On remarque que la variance est toujours positive (alors que $\mathbb{E}[X] \in \mathbb{R}$).

c) Propriétés de l'espérance et de la variance

Propriété 5.6. 1. Si $X \geq 0$, alors $\mathbb{E}[X] \geq 0$.

2. S'il existe $a, b \in \mathbb{R}$ tels que $a \leq X \leq b$, alors $a \leq \mathbb{E}[X] \leq b$.

3. $\mathbb{E}[X]$ est une constante réelle, donc $\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$.

4. L'espérance est linéaire : si $a, b \in \mathbb{R}$, alors $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

5. Si X est à valeurs positives et X à densité alors

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > y) dy.$$

6. $\text{Var}(X) \geq 0$ et $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. Par conséquent, si X est discrète, on a

$$\text{Var}(X) = \sum_{k \geq 1} (x_k)^2 p_X(k) - \left(\sum_{k \geq 1} x_k p_X(k) \right)^2$$

et si X est à densité,

$$\text{Var}(X) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \left(\int_{-\infty}^{+\infty} x f_X(x) dx \right)^2.$$

7. Si $a, b \in \mathbb{R}$, $\text{Var}(aX + b) = a^2 \text{Var}(X)$.
8. $\text{Var}(X) = 0$ si et seulement si la variable X est constante.

Démonstration. 1. Si $X \geq 0$, ça signifie que $x_k \geq 0, \forall k \geq 1$ dans le cas discret et $f_X(x)$ à support dans $\mathbb{R}^+$ dans le cas à densité, et donc on conclut en utilisant la définition de l'espérance.

2. Si $a \leq X \leq b$, ça signifie que $a \leq x_k \leq b \forall k \geq 1$ dans le cas discret et f_X à support dans $[a, b]$ (i.e., $f_X(x) = f_X(x) \mathbb{1}_{[a,b]}(x)$) dans le cas à densité, et donc on conclut en utilisant la définition de l'espérance.

3. $\mathbb{E}[X]$ est une constante réelle, donc sa valeur moyenne est elle-même.

4. Si X est discrète, on a par linéarité de la somme

$$\mathbb{E}[aX + b] = \sum_{k \geq 1} (ax_k + b)p_k = a \sum_{k \geq 1} x_k p_k + b \sum_{k \geq 1} p_k = a\mathbb{E}[X] + b,$$

car $(p_k)_{k \geq 0}$ est une fonction de masse. Si X est à densité, on a par linéarité de l'intégrale

$$\mathbb{E}[aX + b] = \int_{-\infty}^{+\infty} (ax + b)f_X(x)dx = a \int_{-\infty}^{+\infty} x f_X(x)dx + b \int_{-\infty}^{+\infty} f_X(x)dx = a\mathbb{E}[X] + b,$$

car f_X est une densité de probabilité.

5. Si X est positive on a

$$\int_0^\infty \mathbb{P}(X > y)dy = \int_0^\infty \left(\int_y^\infty f_X(x)dx \right) dy = \int_0^\infty \left(\int_0^x dy \right) f_X(x)dx = \int_0^\infty x f_X(x)dx = \mathbb{E}[X].$$

6. Par définition, on a toujours $\text{Var}(X) \geq 0$. De plus, en développant le carré et par linéarité de l'espérance, on obtient

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

7. Par définition de la variance et linéarité de l'espérance, on a

$$\text{Var}(aX + b) = \mathbb{E}[(aX + b - \mathbb{E}[aX + b])^2] = \mathbb{E}[a^2(X - \mathbb{E}[X])^2] = a^2 \text{Var}(X).$$

8. Si $X = c$ est constante, on a $\mathbb{E}[X] = c$ et donc $\text{Var}(X) = 0$. Si $\text{Var}(X) = 0$, on a alors $\mathbb{E}[(X - \mathbb{E}[X])^2] = 0$. Comme la variable $(X - \mathbb{E}[X])^2$ est positive ceci implique qu'elle est nulle et donc $X = \mathbb{E}[X]$, qui est une constante.

□

Exemple 5.4. 1. Si X v.a. de fonction de masse $p(-2) = 1/2, p(0) = 0.1, p(3) = 0.4$. Alors

$$\begin{aligned} \mathbb{E}[X] &= 0.2 & \mathbb{E}[X^2] &= 5.6 \\ \text{Var}(X) &= 5.6 - 0.04 = 5.56. \end{aligned}$$

2. Soit X v.a. de fonction de masse $p_X(-1) = 1/4, p_X(0) = 1/2$, et $p_X(1) = 1/4$. Et Y v.a. de fonction de masse $p_Y(-10) = 1/3, p_Y(0) = 1/3$, et $p_Y(10) = 1/3$. On a $\mathbb{E}[X] = 0 = \mathbb{E}[Y]$. Par contre, $\text{Var}(X) = 1/2$ et $\text{Var}(Y) = 200/3 > \text{Var}(X)$. La variable Y est plus dispersée que la variable X .

3. Si $X \sim \mathcal{B}(n, p)$. On a déjà calculé que $\mathbb{E}[X] = np$. On calcule maintenant

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

Par conséquent,

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

4. On sait que si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors on peut écrire $X = \sigma Z + \mu$ avec $Z \sim \mathcal{N}(0, 1)$. Donc $\text{Var}(X) = \sigma^2 \text{Var}(Z)$. Calculons la variance de Z . On a $\mathbb{E}[Z] = 0$, donc

$$\begin{aligned}\text{Var}(Z) = \mathbb{E}[Z^2] &= \int_{-\infty}^{+\infty} \frac{z^2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 2 \int_0^{+\infty} \frac{z^2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \left[-\frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]_0^{+\infty} + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= 1.\end{aligned}$$

Par conséquent, si $X \sim \mathcal{N}(\mu, \sigma^2)$, on a $\mathbb{E}[X] = \mu$ et $\text{Var}(X) = \sigma^2$.

5.3 Fonctions génératrices

Ces fonctions permettent de caractériser la loi d'une variable aléatoire.

a) Fonction génératrice des moments

Définition 5.7. Soit X une variable aléatoire, la fonction génératrice des moments est la fonction définie sur $\mathbb{R}$ à valeurs dans $[0, +\infty]$, pour $t \in \mathbb{R}$,

$$M_X(t) = \mathbb{E}[e^{tX}].$$

Attention, l'espérance peut être infinie pour certaines valeurs de t et donc cette fonction n'est pas toujours définie sur $\mathbb{R}$ tout entier.

Cependant, on a toujours $M_X(0) = 1$. Dans le cas où la fonction génératrice des moments existe, on a le résultat suivant.

Théorème 5.8. Si X et Y sont deux v.a. alors X et Y ont la même loi si et seulement si $M_X(t) = M_Y(t)$ pour tout $t \in \mathbb{R}$.

Proposition 5.9. Si M_X est la fonction génératrice des moments de X , alors $\forall k \geq 1$,

$$\mathbb{E}[X^k] = M_X^{(k)}(0).$$

Démonstration. On suppose que l'on peut dériver la fonction M_X . On a, pour $k \geq 1$,

$$M_X(t) = \mathbb{E}[e^{tX}], \quad M'_X(t) = \mathbb{E}[Xe^{tX}], \quad M''_X(t) = \mathbb{E}[X^2e^{tX}], \quad M_X^{(k)}(t) = \mathbb{E}[X^k e^{tX}].$$

D'où le résultat. $\square$

Exemple 5.5. 1. Si $X \sim \mathcal{Ber}(p)$, on a $M_X(t) = pe^t + (1-p)$.

2. Si $X \sim \mathcal{B}(n, p)$, on a $M_X(t) = \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} = (pe^t + (1-p))^n$, d'après la formule du binôme de Newton.

On peut en déduire les moments de la loi Binomiale. On a

$$\begin{aligned} M'_X(t) &= npe^t(pe^t + (1-p))^{n-1}, \\ M''_X(t) &= npe^t(pe^t + (1-p))^{n-1} + n(n-1)(pe^t)^2(pe^t + (1-p))^{n-2}. \end{aligned}$$

Par conséquent,

$$\mathbb{E}[X] = M'_X(0) = np, \quad \mathbb{E}[X^2] = M''_X(0) = np + n(n-1)p^2.$$

Comme $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, on obtient $\text{Var}(X) = np(1-p)$.

3. Si $X \sim \mathcal{U}[a, b]$,

$$M_X(t) = \frac{1}{b-a} \int_a^b e^{tx} dx = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

4. Si $X \sim \mathcal{Geom}(p)$, si $t < -\ln(1-p)$,

$$\begin{aligned} M_X(t) &= \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} = pe^t \sum_{k=0}^{\infty} (e^t(1-p))^k \\ &= \frac{pe^t}{1 - (1-p)e^t} \\ &= \frac{p}{e^{-t} - (1-p)} \end{aligned}$$

et si $t \geq -\ln(1-p)$, M_X n'est pas défini.

On remarque que

$$\begin{aligned} M'_X(t) &= pe^{-t}(e^{-t} - (1-p))^{-2}, \\ M''_X(t) &= -pe^{-t}(e^{-t} - (1-p))^{-2} + 2pe^{-2t}(e^{-t} - (1-p))^{-3}. \end{aligned}$$

Par conséquent,

$$\mathbb{E}[X] = M'_X(0) = \frac{1}{p}, \quad \mathbb{E}[X^2] = M''_X(0) = -\frac{1}{p} + \frac{2}{p^2}.$$

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

5. Si $X \sim \mathcal{Exp}(\lambda)$.

$$M_X(t) = \int_0^{\infty} e^{xt} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-x(\lambda-t)} dx.$$

Cette intégrale n'est finie que si $\lambda - t > 0$, pour $t < \lambda$, on a

$$M_X(t) = \lambda \left[-\frac{1}{\lambda-t} e^{-x(\lambda-t)} \right]_0^{\infty} = \frac{\lambda}{\lambda-t}.$$

b) Fonction génératrice des probabilités

[Non fait en cours]

Dans le cas des variables discrètes, on utilise aussi une autre fonction génératrice.

Définition 5.10. La fonction génératrice des probabilités est définie pour les variables discrètes à valeurs dans $\mathbb{N}$. Soit X une v.a. à valeurs entières, la fonction génératrice des probabilités est la fonction définie sur $\mathbb{R}$ à valeurs dans $[0, +\infty]$, pour $t \in \mathbb{R}$, par

$$G_X(t) = \mathbb{E}[t^X] = \sum_{k=0}^{\infty} t^k p_X(k).$$

Théorème 5.11. Si X et Y sont deux v.a. à valeurs dans $\mathbb{N}$, alors X et Y ont la même loi si et seulement si $G_X(t) = G_Y(t)$ pour tout $t \in \mathbb{R}$.

Proposition 5.12. Si G_x est la fonction génératrice des probabilités de X , alors $\forall k \geq 1$,

$$p_X(k) = G_X^{(k)}(0)/k!.$$

Démonstration. On suppose que l'on peut dériver la fonction G_X . on a alors,

$$G_X(t) = \sum_{k=0}^{\infty} t^k p_X(k), \quad G'_X(t) = \sum_{k=0}^{\infty} k t^{k-1} p_X(k), \quad G''_X(t) = \sum_{k=0}^{\infty} k(k-1) t^{k-2} p_X(k), \dots$$

En faisant un changement d'indice dans la somme et en prenant la valeur en 0, on obtient le résultat. $\square$

Remarque 5.13. Si $t > 0$, $G_X(t) = M_X(\ln t)$.

Exemple 5.6. 1. Si $X \sim \mathcal{Ber}(p)$, on a $G_X(t) = pt + (1 - p)$.

2. Si $X \sim \mathcal{B}(n, p)$, on a $G_X(t) = (pt + (1 - p))^n$.

3. Si $X \sim \mathcal{Geom}(p)$, si $t < 1/(1 - p)$,

$$G_X(t) = \frac{pt}{1 - t(1 - p)}.$$

5.4 Formulaire

• Lois discrètes usuelles

Loi	Fonction de masse	Moments	$M_X(t)$
<i>Bernoulli</i>	$p_X(1) = p, p_X(0) = 1 - p$	$\mathbb{E}[X] = p, \text{Var}(X) = p(1 - p)$	$pe^t + 1 - p$
<i>Binomiale</i>	$\binom{n}{k} p^k (1 - p)^{n-k}, 0 \leq k \leq n$	$\mathbb{E}[X] = np, \text{Var}(X) = np(1 - p)$	$(pe^t + 1 - p)^n$
<i>Géométrique</i>	$p(1 - p)^{k-1}, k \geq 1$	$\mathbb{E}[X] = \frac{1}{p}, \text{Var}(X) = \frac{1-p}{p^2}$	$\frac{pe^t}{1 - (1 - p)e^t}$
<i>Poisson</i>	$\frac{\lambda^k}{k!} e^{-\lambda}, k \geq 0$	$\mathbb{E}[X] = \lambda, \text{Var}(X) = \lambda$	$e^{\lambda(e^t - 1)}$
<i>Binomiale Négative</i>	$\binom{k-1}{r-1} p^r (1 - p)^{k-r}, k \geq r$	$\mathbb{E}[X] = \frac{r}{p}, \text{Var}(X) = \frac{r(1-p)}{p^2}$	$\left(\frac{pe^t}{1 - (1 - p)e^t} \right)^r$

• Lois à densité usuelles

Loi	Densité	Moments	$M_X(t)$
<i>Uniforme</i> $\mathcal{U}([a, b]), a < b$	$\frac{1}{b - a} \mathbf{1}_{x \in [a, b]}$	$\mathbb{E}[X] = \frac{a+b}{2}, \text{Var}(X) = \frac{(b-a)^2}{12}$	$\frac{e^{bt} - e^{at}}{t(b - a)}$
<i>Exponentielle</i> $\mathcal{E}(\lambda), \lambda > 0$	$\lambda e^{-\lambda x} \mathbf{1}_{x > 0}$	$\mathbb{E}[X] = \frac{1}{\lambda}, \text{Var}(X) = \frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$ si $t < \lambda$
<i>Normale</i> $\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mathbb{E}[X] = \mu, \text{Var}(X) = \sigma^2$	$e^{\mu t} e^{\frac{\sigma^2 t^2}{2}}$
<i>Gamma</i> $G(\alpha, \lambda), \alpha, \lambda > 0$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{x > 0}$	$\mathbb{E}[X] = \frac{\alpha}{\lambda}, \text{Var}(X) = \frac{\alpha}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - t} \right)^\alpha$ si $t < \lambda$

Chapitre 6

Couples de variables aléatoires et indépendance

Sommaire

6.1	Loi jointe	61
a)	Cas des variables discrètes	62
b)	Cas des variables à densité	64
6.2	Espérance d'une fonction réelle du couple (X, Y)	66
6.3	Variables aléatoires indépendantes	69
6.4	Loi d'une modification d'un couple de variables	71
a)	Loi de la somme de deux v.a. indépendantes	71
b)	Transformation générales des couples de variables à densité	73
6.5	Loi conditionnelle pour des variables discrètes	75
6.6	Loi conditionnelle pour des variables à densité	77
6.7	Covariance et corrélation	77

On considère deux variables aléatoires X et Y . On aimerait connaître s'il y a influence d'une variable sur l'autre et la quantifier.

On va la plupart du temps se limiter à l'étude de couples de variables aléatoires, on peut bien sûr étendre les notions introduites à des n -uplets $(X_1, X_2, \dots, X_n)$ de variables.

Exemple 6.1. On peut se poser la question de l'influence des catastrophes météorologiques (tempêtes, ouragans, tsunamis, ...) sur les cours de la bourse. La variable X modélisera alors les catastrophes météorologiques et la variable Y le cours de la bourse.

6.1 Loi jointe

Définition 6.1. Soient X et Y deux variables aléatoires. La **loi jointe** (aussi appelée loi conjointe, ou encore loi bivariable) du couple (X, Y) est définie par sa **fonction de répartition** $F_{(X,Y)}$:

$$\begin{aligned} F_{(X,Y)} : \mathbb{R} \times \mathbb{R} &\rightarrow [0, 1] \\ (x, y) &\mapsto \mathbb{P}(X \leq x, Y \leq y) \end{aligned}$$

Les **lois marginales** du couple (X, Y) sont la loi $\mathcal{L}(X)$ de X et la loi $\mathcal{L}(Y)$ de Y .

Remarque 6.2. À partir de la loi du couple, on retrouve facilement les lois marginales. En effet, soit $u \in \mathbb{R}$, $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y \in \mathbb{R}) = \lim_{y \rightarrow +\infty} F_{(X,Y)}(x, y)$.

Par contre, des lois marginales on ne peut pas déduire la loi du couple, car elles ne rendent pas compte des liens qui existent entre les variables.

Exemple 6.2. Soit U une v.a. de loi uniforme sur $[0, 1]$. On pose $X = U$ et $Y = 1 - U$. On souhaite la loi du couple (X, Y) .

On sait que les variables sont à valeurs dans $[0, 1]$, donc le couple (X, Y) est à valeurs dans $[0, 1] \times [0, 1]$. Pour $(x, y) \in [0, 1]^2$, on a

$$F_{(X,Y)}(x, y) = \mathbb{P}(U \leq x, 1 - U \leq y) = \mathbb{P}(1 - y \leq U \leq x) = \begin{cases} x + y - 1 & \text{si } 1 \leq x + y \\ 0 & \text{si } 1 > x + y. \end{cases}$$

De plus, si $x \leq 0$ ou $y \leq 0$, on a $F_{(X,Y)}(x, y) = 0$. Si $x \geq 1$ et $y \in [0, 1]$, $F_{(X,Y)}(x, y) = \mathbb{P}(Y \leq y) = \mathbb{P}(1 - U \leq y) = y$ et si $x \in [0, 1]$, $y \geq 1$, $F_{(X,Y)}(x, y) = \mathbb{P}(X \leq x) = \mathbb{P}(U \leq x) = x$.

a) Cas des variables discrètes

Propriété 6.3. Soient X et Y deux variables discrètes, X à valeurs dans $\mathbb{D}_X = \{x_i\}_{i \in I}$ et Y à valeurs dans $\mathbb{D}_Y = \{y_j\}_{j \in J}$. La loi du couple (X, Y) est définie par la fonction de masse du couple :

$$p_{(X,Y)}(x, y) := \mathbb{P}(X = x, Y = y) \quad \text{avec } x \in \mathbb{D}_X \text{ et } y \in \mathbb{D}_Y.$$

L'ensemble des couples (x, y) tels que $p_{(X,Y)}(x, y) > 0$ s'appelle le support du couple (X, Y) .

Remarque 6.4. On obtient les lois marginales de façon simple. Soit $x \in \mathbb{D}_X$, on a

$$\begin{aligned} \mathbb{P}(X = x) &= \mathbb{P}(X = x \text{ et } Y \in \mathbb{D}_Y) \\ &= \sum_{j \geq 1} \mathbb{P}(X = x, Y = y_j) \end{aligned}$$

De même, pour $y \in \mathbb{D}_Y$, on a $\mathbb{P}(Y = y) = \sum_{i \geq 1} \mathbb{P}(X = x_i, Y = y)$

Dans le cas où les variables sont discrètes et prennent un petit nombre de valeurs, on écrit en général la loi du couple sous la forme d'un tableau :

$Y \backslash X$	...	Somme des colonnes
$\vdots$	$\mathbb{P}(X = x, Y = y)$	$\mathbb{P}(Y = y)$
Somme des lignes	$\mathbb{P}(X = x)$	

Exemple 6.3. On lance une pièce truquée 3 fois. La probabilité de tomber sur "Pile" est $2/3$. Soit X le nombre de "Face" obtenu dans les deux premiers jets et Y le nombre de "Face" obtenu dans les deux derniers jets. La loi de (X, Y) est donnée par

$y \backslash x$	0	1	2	$\mathbb{P}(Y = y)$
0	$\left(\frac{2}{3}\right)^3 = \frac{8}{27}$	$\frac{1}{3} \left(\frac{2}{3}\right)^2 = \frac{4}{27}$	0	$4/9$
1	$\frac{1}{3} \left(\frac{2}{3}\right)^2 = \frac{4}{27}$	$\frac{1}{3} \left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2 \frac{2}{3} = \frac{6}{27}$	$\left(\frac{1}{3}\right)^2 \frac{2}{3} = \frac{2}{27}$	$4/9$
2	0	$\left(\frac{1}{3}\right)^2 \frac{2}{3} = \frac{2}{27}$	$\left(\frac{1}{3}\right)^3 = \frac{1}{27}$	$1/9$
$\mathbb{P}(X = x)$	$4/9$	$4/9$	$1/9$	

On retrouve la loi binomiale $\mathcal{Bin}(2, 1/3)$ pour les loi de X et de Y .

Exemple 6.4. On effectue une suite infinie de lancers indépendants d'un dé équilibré. On note les lancers à partir de 1. On définit les deux variables aléatoires :

X est égale au numéro du lancer qui donne le premier 6,

Y est égale au nombre de 5 obtenus avant le premier 6.

Le couple (X, Y) est à valeurs dans $\mathbb{N}^* \times \mathbb{N}$ avec $Y < X$. Par conséquent si $k \geq n$, $\mathbb{P}(X = n, Y = k) = 0$ et si $k < n$,

$$\begin{aligned} \mathbb{P}(X = n, Y = k) &= \mathbb{P}(\text{"k fois 5 et pas de 6 sur les } n-1 \text{ premiers lancers et un 6 au } n^{\text{ème}} \text{ lancer."}) \\ &= \binom{n-1}{k} \left(\frac{1}{6}\right)^k \left(\frac{4}{6}\right)^{n-1-k} \frac{1}{6} = \binom{n-1}{k} \frac{4^{n-1-k}}{6^n}. \end{aligned}$$

Propriété 6.5. Si $A \subset \mathbb{R}$ et $B \subset \mathbb{R}$, alors

$$\mathbb{P}((X, Y) \in A \times B) = \sum_{x \in A} \sum_{y \in B} p_{(X, Y)}(x, y).$$

Si $C \subset \mathbb{R}^2$, alors

$$\mathbb{P}((X, Y) \in C) = \sum_{(x, y) \in C} p_{(X, Y)}(x, y)$$

Exemple 6.5. Reprendre les exemples précédents et calculer $\mathbb{P}(X > Y + 1)$.

1. $\mathbb{P}(X > Y + 1) = \mathbb{P}(X > 1, Y = 0) + \mathbb{P}(X > 2, Y = 1) + \mathbb{P}(X > 3, Y = 2) = 0$.
- 2.

$$\begin{aligned} \mathbb{P}(X > Y + 1) &= \sum_{n > k+1} p_{(X, Y)}(n, k) = \sum_{n=2}^{\infty} \sum_{k=0}^{n-2} p_{(X, Y)}(n, k) = \sum_{k=0}^{\infty} \sum_{n=k+2}^{\infty} p_{(X, Y)}(n, k) \\ &= 1 - \mathbb{P}(X \leq Y + 1) = 1 - \mathbb{P}(X = Y + 1) \quad \text{car on a par définition } X \geq Y + 1 \\ &= 1 - \sum_{n=1}^{\infty} \mathbb{P}(X = n, Y = n - 1) = 1 - \sum_{n=1}^{\infty} \left(\frac{1}{6}\right)^n = 1 - \frac{1}{6} \frac{1}{1 - 1/6} = \frac{4}{5}. \end{aligned}$$

Exemple 6.6. Une urne contient 2 boules rouges, 5 blanches et 3 vertes. On pioche 3 boules sans remise. On note X le nombre de boules rouges et Y le nombre boules blanches. Fonction de masse conjointe de (X, Y) est

$$p_{(X, Y)}(x, y) = \frac{\binom{2}{x} \binom{5}{y} \binom{3}{3-x-y}}{\binom{10}{3}}$$

si $x = 0, 1, 2$, $y = 0, 1, 2, 3$ et $x + y \leq 3$. Sinon elle vaut 0.

$y \backslash x$	0	1	2	$\mathbb{P}(Y = y)$
0	$\frac{1}{120}$	$\frac{6}{120}$	$\frac{3}{120}$	$\frac{10}{120}$
1	$\frac{15}{120}$	$\frac{30}{120}$	$\frac{5}{120}$	$\frac{50}{120}$
2	$\frac{30}{120}$	$\frac{20}{120}$	0	$\frac{50}{120}$
3	$\frac{10}{120}$	0	0	$\frac{10}{120}$
$\mathbb{P}(X = x)$	$\frac{56}{120}$	$\frac{56}{120}$	$\frac{8}{120}$	1

On déduit que

$$\mathbb{P}(X \geq Y) = \frac{10}{120} + \frac{30}{120} + \frac{5}{120} = \frac{45}{120}$$

ainsi que

$$\mathbb{P}(X = 2 | X \geq Y) = \frac{\frac{8}{120}}{\frac{45}{120}} = \frac{8}{45}.$$

b) Cas des variables à densité

Définition 6.6. Une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est appelée densité sur $\mathbb{R}^2$ si

(i) $f(x, y) \geq 0$ pour tout $(x, y) \in \mathbb{R}^2$,

(ii) $\iint_{\mathbb{R}^2} f(x, y) dx dy = 1$.

Définition 6.7. La loi du couple de v.a. (X, Y) est dite à densité s'il existe une fonction de densité $f_{(X,Y)}$ de deux variables telle que la fonction de répartition du couple vérifie pour tout $(x, y) \in \mathbb{R}^2$

$$\mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^x \left(\int_{-\infty}^y f_{(X,Y)}(s, t) dt \right) ds = \int_{-\infty}^y \left(\int_{-\infty}^x f_{(X,Y)}(s, t) ds \right) dt$$

Le support de la loi jointe de (X, Y) est l'ensemble des couples (x, y) tels que $f_{(X,Y)}(x, y) > 0$.

La densité jointe caractérise la loi du couple (X, Y) .

Exemple 6.7. On considère la densité jointe

$$f_{(X,Y)}(x, y) = c(x^2 + xy/2) \mathbb{1}_{[0,1] \times [0,2]}(x, y).$$

Il s'agit d'une densité si $c = \frac{6}{7}$. Le support de la loi est $[0, 2] \times [0, 1]$.

On peut facilement retrouver la densité à partir de la fonction de répartition. En effet, si le couple (X, Y) a pour densité $f_{(X,Y)}$, sa fonction de répartition s'écrit :

$$F_{(X,Y)}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(u, v) dv du$$

En dérivant une fois par rapport à chacune des variables la fonction de répartition, on obtient

$$f_{(X,Y)}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x, y).$$

Remarque 6.8. De même que pour les variables discrètes, on peut retrouver facilement les lois marginales. Soit $x \in \mathbb{R}$, on a

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(X \in]-\infty, x] \text{ et } Y \in \mathbb{R}) = \mathbb{P}((X, Y) \in]-\infty, x] \times \mathbb{R}) \\ &= \int_{-\infty}^x \left(\int_{\mathbb{R}} f_{(X,Y)}(s, t) dt \right) ds. \end{aligned}$$

La variable X est une variable continue de densité $f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy$.

De même, la variable Y est une variable continue de densité $f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx$.

Pour $C \subset \mathbb{R}^2$, on a

$$\mathbb{P}((X, Y) \in C) = \iint_C f_{(X,Y)}(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbb{1}_{(x,y) \in C} f_{(X,Y)}(x, y) dx dy,$$

et si $C = A \times B$, avec $A \subset \mathbb{R}$ et $B \subset \mathbb{R}$, par le théorème de Fubini,

$$\mathbb{P}((X, Y) \in A \times B) = \int_A \int_B f_{(X,Y)}(x, y) dy dx = \int_A \left(\int_B f_{(X,Y)}(x, y) dy \right) dx = \int_B \left(\int_A f_{(X,Y)}(x, y) dx \right) dy.$$

Exemple 6.8. 1. Considérons le couple (X, Y) de densité $f_{(X,Y)}(x, y) = \frac{6}{7}(x^2 + xy/2)\mathbb{1}_{[0,1] \times [0,2]}(x, y)$. Cette fonction est bien une densité de probabilité car à valeurs positives et d'intégrale 1. On en déduit la densité de X :

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy = \int_0^2 \frac{6}{7}(x^2 + \frac{xy}{2}) dy \\ &= \frac{6}{7}(2x^2 + x) \quad \text{si } x \in [0, 1]. \end{aligned}$$

On peut calculer par exemple $\mathbb{P}(X > Y)$:

$$\begin{aligned} \mathbb{P}(X > Y) &= \int \int_{y < x} f_{(X,Y)}(x, y) dx dy = \int_0^1 \left(\int_0^x \frac{6}{7}(x^2 + \frac{xy}{2}) dy \right) dx \\ &= \frac{6}{7} \int_0^1 \frac{5}{4} x^3 dx = \frac{15}{14 \times 4}. \end{aligned}$$

2. Considérons le couple (X, Y) de densité $f_{(X,Y)}(x, y) = xe^{-y}\mathbb{1}_{0 < x < y}$.

Cette fonction est bien une densité de probabilité car à valeurs positives et

$$\iint f_{(X,Y)}(x, y) dx dy = \int_0^{\infty} x \left(\int_x^{\infty} e^{-y} dy \right) dx = \int_x^{\infty} xe^{-x} dx = 1.$$

On en déduit la densité de X . La variable X est à valeurs dans $]0, +\infty[$ et pour $x > 0$, on a

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy = x \int_x^{+\infty} e^{-y} dy \\ &= xe^{-x}. \end{aligned}$$

Donc X suit la loi Gamma(2, 1).

La variable Y est à valeurs dans $]0, +\infty[$ et pour $y > 0$, la densité de Y est donnée par

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx = e^{-y} \int_0^y x dx \\ &= \frac{y^2}{2} e^{-y}. \end{aligned}$$

Donc Y suit la loi Gamma(3, 1).

On peut calculer par exemple $\mathbb{P}(X < \frac{Y}{2})$:

$$\begin{aligned} \mathbb{P}\left(X < \frac{Y}{2}\right) &= \int \int \mathbb{1}_{x < \frac{y}{2}} f_{(X,Y)}(x, y) dx dy = \int_0^\infty x \left(\int_{2x}^\infty e^{-y} dy \right) dx \\ &= \int_0^\infty x e^{-2x} dx = \frac{1}{2} \int_0^\infty x 2e^{-2x} dx = \frac{1}{4}, \end{aligned}$$

car la dernière intégrale correspond à l'espérance d'une variable exponentielle de paramètre 2.

6.2 Espérance d'une fonction réelle du couple (X, Y)

Définition 6.9. Soit $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction.

1. Si (X, Y) est un couple de v.a. discrètes à valeurs dans $\mathbb{D}$ de fonction de masse $p_{(X,Y)}$. On définit $\mathbb{E}[h(X, Y)]$ comme

$$\mathbb{E}[h(X, Y)] = \sum_{(x,y) \in \mathbb{D}} h(x, y) p_{(X,Y)}(x, y).$$

2. Si (X, Y) est un couple de v.a. de densité $f_{(X,Y)}$. On définit $\mathbb{E}[h(X, Y)]$ comme

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

Remarque 6.10. Si h est positive ou h borné, on peut toujours calculer $\mathbb{E}[h(X, Y)]$. Sinon, il faut d'abord vérifier que $\mathbb{E}[h(X, Y)]$ est bien définie, i.e. $\mathbb{E}[|h(X, Y)|]$ est finie.

Exemple 6.9. 1. Un joueur lance un dé et une pièce (bien équilibrés) de façon indépendantes. Si la pièce donne Face, alors il gagne deux fois la valeur affichée par le dé, sinon il gagne la moitié de la valeur du dé. Calculer le gain espéré.

On note X le résultat du dé et Y le résultat de la pièce.

Le gain est alors $G = 2X\mathbb{1}_{Y=1} + \frac{X}{2}\mathbb{1}_{Y=0} = h(X, Y)$. La loi de (X, Y) est la loi uniforme sur $\{1, \dots, 6\} \times \{0, 1\}$. Donc

$$\begin{aligned} \mathbb{E}[G] &= \frac{1}{12} \sum_{x=1}^{12} \sum_{y=0}^1 h(x, y) = \frac{1}{12} \sum_{x=1}^6 h(x, 0) + h(x, 1) \\ &= \frac{1}{12} \left(\sum_{x=1}^6 2x + \sum_{x=1}^6 \frac{x}{2} \right) = \frac{1}{12} \left(2 \times 21 + \frac{1}{2} \times 21 \right) = \frac{35}{8}. \end{aligned}$$

2. On considère (X, Y) de densité $\mathbf{1}_{[0,1] \times [0,1]}$. Calculer $\mathbb{E}[(X - Y)^2]$ et $\mathbb{E}[|X - Y|]$.

$$\begin{aligned}\mathbb{E}[(X - Y)^2] &= \int_0^1 \int_0^1 (x - y)^2 dx dy = \int_0^1 \left[\frac{1}{3}(x - y)^3 \right]_0^1 dy \\ &= \frac{1}{3} \int_0^1 ((1 - y)^3 + y^3) dy = \frac{1}{12} [-(1 - y)^4 + y^4]_0^1 = \frac{1}{6}.\end{aligned}$$

$$\begin{aligned}\mathbb{E}[|X - Y|] &= \int_0^1 \int_0^1 |x - y| dx dy = \int_0^1 \left(\int_0^x (x - y) dy + \int_x^1 (y - x) dy \right) dx \\ &= \int_0^1 \left(\frac{x^2}{2} + \frac{(1 - x)^2}{2} \right) dx = \frac{1}{3}.\end{aligned}$$

Définition 6.11. La **covariance** de deux v.a. X et Y est le réel

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

L'espérance $\mathbb{E}[XY]$ est calculée à partir de la loi jointe de (X, Y) :

1. dans le cas discret, lorsque la somme a un sens,

$$\mathbb{E}[XY] = \sum_{x \in \mathbb{D}_X, y \in \mathbb{D}_Y} xy p_{(X,Y)}(x, y)$$

2. dans le cas continu, lorsque l'intégrale a un sens,

$$\mathbb{E}[XY] = \int \int_{\mathbb{R}^2} xy f_{(X,Y)}(x, y) dx dy.$$

Exemple 6.10. 1. Soit (X, Y) le couple défini dans l'exemple 6.3, où X est le nombre de Face sur les deux premiers lancers et Y le nombre de face sur les 2 derniers lancers lorsqu'on lance 3 fois une pièce. Calculons la covariance de (X, Y) . Comme X et Y suivent la loi Binomiale $\mathcal{Bin}(2, 1/3)$, on a $\mathbb{E}[X] = 2/3 = \mathbb{E}[Y]$. De plus

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{i=0}^2 \sum_{j=0}^2 ij p_{(X,Y)}(i, j) = p_{(X,Y)}(1, 1) + 2p_{(X,Y)}(1, 2) + 2p_{(X,Y)}(2, 1) + 4p_{(X,Y)}(2, 2) \\ &= \frac{6}{27} + 2\frac{2}{27} + 2\frac{2}{27} + 4\frac{1}{27} = \frac{18}{27} = \frac{2}{3}.\end{aligned}$$

$$\text{Donc } \text{Cov}(X, Y) = \frac{2}{3} - \frac{2}{3} \times \frac{2}{3} = \frac{2}{9}.$$

2. Soit (X, Y) un couple de densité $f_{(X,Y)}(x, y) = xe^{-y}\mathbf{1}_{0 < x < y}$. Calculons sa covariance. Comme $X \sim \text{Gamma}(2, 1)$ et $Y \sim \text{Gamma}(3, 1)$. On sait que $\mathbb{E}[X] = 2$ et $\mathbb{E}[Y] = 3$. De plus,

$$\begin{aligned}\mathbb{E}[XY] &= \int_0^\infty \int_0^\infty x^2 ye^{-y} \mathbf{1}_{0 < x < y} dx dy = \int_0^\infty ye^{-y} \left(\int_0^y x^2 dx \right) dy \\ &= \int_0^\infty ye^{-y} \left[\frac{x^3}{3} \right]_0^y dy = \int_0^\infty \frac{y^4}{3} e^{-y} dy \\ &= \frac{\Gamma(5)}{3} \int_0^\infty \frac{y^4}{\Gamma(5)} e^{-y} dy = \frac{\Gamma(5)}{3} = 8,\end{aligned}$$

car $\frac{y^4}{\Gamma(5)} e^{-y} \mathbf{1}_{y > 0}$ est la densité de la loi $\text{Gamma}(5, 1)$.

Par conséquent, $\text{Cov}(X, Y) = 8 - 2 \times 3 = 2$.

Propriété 6.12. L'espérance est linéaire : si $a, b \in \mathbb{R}$ et X, Y deux variables aléatoires. Alors

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

Si $(a_i)_{1 \leq i \leq n}$ des réels et $(X_i)_{1 \leq i \leq n}$ des variables aléatoires, alors

$$\mathbb{E}\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i \mathbb{E}[X_i].$$

Propriété 6.13. Soient X, Y deux variables aléatoires telles que $X \leq Y$ (au sens, $\forall \omega \in \Omega, X(\omega) \leq Y(\omega)$). Alors

$$\mathbb{E}[X] \leq \mathbb{E}[Y].$$

Démonstration. La variable $Z = Y - X \geq 0$ et donc $\mathbb{E}[Z] \geq 0$. Par linéarité de l'espérance, $\mathbb{E}[Z] = \mathbb{E}[Y] - \mathbb{E}[X]$, et on conclut facilement. $\square$

Proposition 6.14. Soient X et Y deux v.a. Alors

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

Démonstration. En développant le carré, on obtient le résultat :

$$\begin{aligned} \text{Var}(X + Y) &= E[(X + Y - E[X + Y])^2] = E[(X - E[X] + Y - E[Y])^2] \\ &= E[(X - E[X])^2] + E[(Y - E[Y])^2] + 2E[(X - E[X])(Y - E[Y])] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y). \end{aligned}$$

$\square$

Exemple 6.11. Les n étudiants de la classe font une fête de Noël. Chaque étudiant reprend un manteau au hasard en repartant. On note N le nombre de personne repartant avec son propre manteau. On voudrait connaître le nombre moyen de personnes repartant avec son manteau.

On a $N \in \{0, \dots, n\}$. La fonction de masse de N est difficile à obtenir (on a montré dans un chapitre précédent par la formule d'inclusion-exclusion que $\mathbb{P}(N = 0) = \sum_{k=0}^n \frac{(-1)^k}{k!}$). Donc chercher la loi de N n'est pas envisageable.

Si on note E_i : la personne i repart avec son manteau, on remarque que

$$N = \sum_{i=1}^n \mathbf{1}_{E_i}.$$

Donc $\mathbb{E}[N] = \sum_{i=1}^n \mathbb{P}(E_i) = \sum_{i=1}^n \frac{1}{n} = 1$.

6.3 Variables aléatoires indépendantes

Définition 6.15. Deux variables aléatoires X et Y sont indépendantes si pour tous $A \subset \mathbb{R}$ et $B \subset \mathbb{R}$ les événements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants, i.e.

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

On note souvent $X \perp\!\!\!\perp Y$.

Les variables aléatoires $X_1, \dots, X_d$ sont indépendantes si pour tous $A_1 \subset \mathbb{R}, \dots, A_d \subset \mathbb{R}$, on a

$$\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \prod_{i=1}^d \mathbb{P}(X_i \in A_i).$$

Une suite de variables $(X_n)_{n \geq 0}$ indépendantes est une suite telle que pour toute sous partie finie $I \subset \mathbb{N}$, les variables $(X_i)_{i \in I}$ sont indépendantes.

Exemple 6.12. 1. Si $X \sim \text{Unif}([-1, 1])$ et $Y = |X|$, alors elles ne sont pas indépendantes car

$$0 = \mathbb{P}(X \in]-1/2, 1/2[, Y > 3/4) \neq \mathbb{P}(-1/2 \leq X \leq 1/2)\mathbb{P}(Y > 3/4)$$

2. On lance deux dés, on note X le résultat du premier dé et S la somme de chiffres apparus. Les événements $\{X = 3\}$ et $\{S = 7\}$ sont indépendants, mais les variables ne sont pas indépendantes car $\{X = 5\}$ et $\{S = 12\}$ non indépendants.

En effet, $\mathbb{P}(X = 3) = \mathbb{P}(X = 5) = \frac{1}{6}$, $\mathbb{P}(X = 3, S = 7) = \mathbb{P}(X = 5, S = 12) = \frac{1}{36}$. Par contre, $\mathbb{P}(S = 7) = \frac{1}{6}$ et $\mathbb{P}(S = 12) = \frac{1}{36}$.

Proposition 6.16. Soit $X_1, \dots, X_d$ des variables indépendantes, alors pour tout $p \in \{1, \dots, d\}$, toute fonction de $X_1, \dots, X_p$ est indépendante de toute fonction de $X_{p+1}, \dots, X_d$.

En fait si $I \cap J = \emptyset$, on a $(X_i)_{i \in I}$ et $(X_j)_{j \in J}$ indépendantes.

Propriété 6.17. Soit f et g deux fonctions. Si X et Y sont indépendantes, alors $f(X)$ et $g(Y)$ sont indépendantes.

Théorème 6.18. (Admis)

Deux v.a. X et Y sont indépendantes

ssi la fonction de répartition du couple vérifie pour tout $(x, y) \in \mathbb{R}^2$

$$F_{(X,Y)}(x, y) = F_X(x)F_Y(y);$$

ssi (dans le cas discret) pour tout $x \in \mathbb{D}_X$ et $y \in \mathbb{D}_Y$,

$$p_{(X,Y)}(x, y) = p_X(x)p_Y(y);$$

ssi (dans le cas continu), la densité du couple vérifie pour tout $(x, y) \in \mathbb{R}^2$

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y).$$

Ce théorème est encore vrai quand on considère d variables $X_1, \dots, X_d$.

Remarque 6.19. Si X, Y sont discrets, pour montrer que X, Y sont indépendantes, il faut montrer que pour tout (x, y) , on a $p_{(X,Y)}(x, y) = p_X(x)p_Y(y)$. Par contre, pour montrer qu'ils ne sont pas indépendants, il suffit de trouver UN couple (x, y) tel que $p_{(X,Y)}(x, y) \neq p_X(x)p_Y(y)$.

Exemple 6.13. Soit $X \sim \text{Ber}(0.3)$ et $Y \sim \text{Bin}(2, 0.1)$ indépendantes. Trouver la loi de $Z = XY$.
 Z est à valeurs dans $\{0, 1, 2\}$. On a, par indépendance,

$$\mathbb{P}(Z = 1) = \mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = 1)\mathbb{P}(Y = 1) = 0.3 \times 2 \times 0.1 \times 0.9 = 0.054$$

$$\mathbb{P}(Z = 2) = \mathbb{P}(X = 1, Y = 2) = \mathbb{P}(X = 1)\mathbb{P}(Y = 2) = 0.3 \times (0.1)^2 = 0.003$$

$$\mathbb{P}(Z = 0) = 1 - \mathbb{P}(Z = 1) - \mathbb{P}(Z = 2).$$

Exemple 6.14. Deux personnes se donnent rendez-vous. L'heure d'arrivée de chacune de ces deux personnes sur les lieux est une variable uniforme entre midi et une heure. Ces deux variables sont indépendantes.

1. Quelle est la probabilité qu'ils arrivent au même instant ?

Notons X, Y les heures d'arrivées des personnes, elle ont pour loi à densité $f(x) = \mathbb{1}_{[12,13]}(x)$. Par indépendance, la densité du couple est alors $f_{X,Y}(x, y) = \mathbb{1}_{[12,13]}(x)\mathbb{1}_{[12,13]}(y)$.

La probabilité $\mathbb{P}(X = Y)$ qu'ils arrivent au même instant est nulle car le couple est à densité. En effet,

$$\mathbb{P}(X = Y) = \int_{12}^{13} \int_{12}^{13} \mathbb{1}_{x=y} dx dy = 0.$$

2. Quelle est la probabilité que le premier arrivé doive attendre plus de 10 minutes ?

On a

$$\begin{aligned} \mathbb{P}(|X - Y| > 1/6) &= \mathbb{P}(\max(X, Y) > \min(X, Y) + 1/6) \\ &= \mathbb{P}(X > Y + 1/6) + \mathbb{P}(Y > X + 1/6) = 2\mathbb{P}(X > Y + 1/6) \quad (\text{par symétrie}) \\ &= 2 \int_{12}^{13-1/6} \left(\int_{y+1/6}^{13} dx \right) dy = 25/36. \end{aligned}$$

3. Si les deux personnes se donnent un rendez-vous plus précis, à midi exactement par exemple. La loi uniforme est-elle adaptée au problème ?

Non, si elles se donnent rendez-vous à midi, il faut une loi qui charge plus du côté de midi.

Exemple 6.15. Soit X, Y deux v.a. indépendantes, alors les lois du maximum $U = \max(X, Y)$ et du minimum $V = \min(X, Y)$ sont faciles à calculer à partir de la fonction de répartition.

On a

$$\begin{aligned} F_U(x) &= \mathbb{P}(X \leq x, Y \leq x) = F_X(x)F_Y(x) \\ F_V(x) &= 1 - \mathbb{P}(X > x, Y > x) = 1 - (1 - F_X(x))(1 - F_Y(x)). \end{aligned}$$

Cas de deux lois exponentielles : on suppose $X \sim \text{Exp}(\lambda)$ et $Y \sim \text{Exp}(\mu)$, alors $V = \min(X, Y) \sim \text{Exp}(\lambda + \mu)$.

Théorème 6.20. (Admis)

Deux v.a. X et Y sont indépendantes

ssi pour toutes fonctions h, g telles que les espérances sont bien définies,

$$E[h(X)g(Y)] = E[h(X)]E[g(Y)],$$

ssi la fonction génératrice des moments du couple vérifie pour tout (s, t) tels que les fonctions sont bien définies,

$$M_{(X,Y)}(s, t) = M_X(s)M_Y(t) \quad \text{où } M_{(X,Y)}(s, t) = E[e^{sX+tY}].$$

Ce théorème se généralise à d variables aléatoires $X_1, \dots, X_d$.

Conséquence 6.21. Si X et Y deux v.a. indépendantes, alors $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0$ et donc

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) = \text{Var}(X) + \text{Var}(Y).$$

Exemple 6.16. En calculant la fonction génératrice des moments de $X + Y$, on obtient facilement les résultats suivants.

1. Soient X et Y deux variables indépendantes, de loi de Poisson de paramètre respectif λ et μ . Alors $X + Y$ suit la loi de Poisson de paramètre $\lambda + \mu$.

En effet, on rappelle que $M_X(t) = e^{\lambda(e^t-1)}$. Donc, par indépendance

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\lambda(e^t-1)}e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}.$$

On reconnaît la fonction génératrice des moments de la loi de Poisson de paramètre $(\lambda + \mu)$.

2. Loi de la somme de deux variables indépendantes de loi Gamma : $X \sim \text{Gamma}(\alpha, \lambda)$ et $Y \sim \text{Gamma}(\beta, \lambda)$.

Alors $M_X(t) = \left(\frac{\lambda}{\lambda-t}\right)^\alpha$ pour $t < \lambda$. On obtient $X + Y \sim \text{Gamma}(\alpha + \beta, \lambda)$.

Exemple 6.17. Une loi binomiale $\mathcal{B}(n, p)$ de paramètres n et p correspond au modèle suivant :

On renouvelle n fois de manière indépendante une épreuve de Bernoulli de paramètre p . On compte le nombre de succès obtenus à l'issue des n épreuves et on appelle S la variable aléatoire correspondant à ce nombre de succès. Donc si on note X_i le résultat du $i^{\text{ème}}$ lancer

$$X_i = \mathbb{1}_{\text{avoir un succès au } i^{\text{ème}} \text{ lancer}}$$

les variables $X_1, \dots, X_n$ sont des variables de Bernoulli de paramètre p , indépendantes. On a $S = X_1 + \dots + X_n$.

On en déduit facilement à partir de la fonction génératrice des moments que si $S_1 \sim \mathcal{B}(n_1, p)$ et $S_2 \sim \mathcal{B}(n_2, p)$ indépendantes, alors $S_1 + S_2 \sim \mathcal{B}(n_1 + n_2, p)$.

Exercice 1. Trouver la loi de la somme de n v.a. exponentielles de même paramètres, indépendantes.

6.4 Loi d'une modification d'un couple de variables

a) Loi de la somme de deux v.a. indépendantes

Soient X et Y deux variables indépendantes. On veut trouver la loi de la somme.

Dans l'exemple précédent, on a vu que si X et Y sont indépendantes alors la fonction génératrice des moments de $X + Y$ est

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

Si X et Y ont des lois connues dont on connaît la FGM, alors il peut être intéressant d'utiliser ce résultat pour trouver la loi de $X + Y$ mais il ne donnera rien si on ne reconnaît pas une loi connue pour la loi de Z .

On va voir des méthodes qui fonctionnent toujours.

- Si $X \in \mathbb{D}_X$ et $Y \in \mathbb{D}_Y$ discrètes, alors pour tout z on a

$$\begin{aligned} p_{X+Y}(z) &= \sum_{x \in \mathbb{D}_X} p_{(X,Y)}(x, z-x) = \sum_{x \in \mathbb{D}_X} p_X(x) p_Y(z-x) \\ &= \sum_{y \in \mathbb{D}_Y} p_X(z-y) p_Y(y) \end{aligned}$$

- Si X et Y à densité, alors pour tout z on a

$$\begin{aligned} F_{X+Y}(z) &= \iint_{\mathbb{R}^2} \mathbb{1}_{x+y \leq z} f_X(x) f_Y(y) dx dy \\ &= \int_{\mathbb{R}} f_X(x) \left(\int_{-\infty}^{z-x} f_Y(y) dy \right) dx \\ &= \int_{\mathbb{R}} f_X(x) F_Y(z-x) dx \end{aligned}$$

En dérivant par rapport à z , on obtient que la densité de $X+Y$ vérifie

$$\begin{aligned} f_{X+Y}(z) &= \int_{\mathbb{R}} f_X(x) f_Y(z-x) dx \\ &= \int_{\mathbb{R}} f_X(z-x) f_Y(x) dx \end{aligned}$$

On définit le produit de convolution $f \star g(z) = \int f(x) g(z-x) dx = \int f(z-x) g(x) dx$. Par conséquent, si X et Y sont indépendantes, on a

$$f_{X+Y}(z) = f_X \star f_Y(z).$$

Exemple 6.18. 1. Loi de la somme de deux v.a. indépendantes de loi de Poisson : supposons que $X \sim \mathcal{Pois}(\lambda)$ et $Y \sim \mathcal{Pois}(\mu)$. On a $X+Y$ à valeurs dans $\mathbb{N}$ et pour $n \geq 0$,

$$\begin{aligned} p_{X+Y}(n) &= \mathbb{P}(X+Y=n) = \sum_{k=0}^{\infty} \mathbb{P}(X=k, Y=n-k) \\ &= \sum_{k=0}^{\infty} \mathbb{P}(X=k) \mathbb{P}(Y=n-k) \quad \text{par indépendance} \\ &= \sum_{k=0}^n \frac{\lambda^k}{k!} e^{-\lambda} \frac{\mu^{n-k}}{(n-k)!} e^{-\mu} \quad \text{car } \mathbb{P}(Y=n-k) = 0 \text{ si } k > n \\ &= \frac{1}{n!} e^{-(\lambda+\mu)} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} = \frac{(\lambda+\mu)^n}{n!} e^{-(\lambda+\mu)}, \end{aligned}$$

d'après la formule du binôme de Newton. On retrouve que $X+Y \sim \mathcal{Pois}(\lambda+\mu)$.

Loi de la somme de deux v.a. indépendantes de loi Exponentielle : supposons que $X \sim \mathcal{Exp}(\lambda)$ et $Y \sim$

$\mathcal{Exp}(\mu)$. On a $X + Y$ à valeurs dans $\mathbb{R}^+$ et sa densité est donnée par

$$\begin{aligned} f_{X+Y}(z) &= \int_0^\infty f_X(x) f_Y(z-x) dx \\ &= \lambda \mu \int_0^z e^{-\lambda x} e^{-\mu(z-x)} dx \quad \text{car } Y \text{ est à valeurs dans } \mathbb{R}^+ \\ &= \lambda \mu e^{-\mu z} \int_0^z e^{-(\lambda-\mu)x} dx \\ &= \begin{cases} \lambda \mu \frac{e^{-\mu z} - e^{-(\lambda)\mu z}}{\lambda - \mu} & \text{si } \lambda \neq \mu \\ \lambda^2 x e^{-\mu z} & \text{si } \lambda = \mu. \end{cases} \end{aligned}$$

On retrouve que $X + Y \sim \text{Gamma}(2, \lambda)$ quand $\lambda = \mu$.

b) Transformation générales des couples de variables à densité

Soit (X, Y) un couple de v.a. de densité $f_{(X,Y)}$. Les variables ne sont pas forcément indépendantes. Soit $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ une transformation. A priori, comme dans le cas des v.a. réelles, le couple $(U, V) = g(X, Y)$ peut-être à densité, ou ne pas l'être.

Si la transformation g est bijective et de classe $\mathcal{C}^1$, alors $(U, V) = g(X, Y)$ est un couple à densité.

Cherchons la densité de (U, V) .

Soit $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction bornée. On sait que

$$\begin{aligned} \mathbb{E}[h(U, V)] &= \iint h(u, v) f_{(U,V)}(u, v) du dv \\ &= \mathbb{E}[h(g(X, Y))] = \iint h(g(x, y)) f_{(X,Y)}(x, y) dx dy \end{aligned}$$

On effectue le changement de variables $(u, v) = g(x, y)$. En inversant la fonction g , on a

$$(x, y) = g^{-1}(u, v) \quad \text{et} \quad dx dy = |\det(J_{g^{-1}}(u, v))| du dv,$$

où $J_{g^{-1}}$ est la matrice jacobienne de g^{-1} .

On a alors

$$\begin{aligned} \mathbb{E}[h(U, V)] &= \iint h(u, v) f_{(U,V)}(u, v) du dv \\ &= \iint h(u, v) f_{(X,Y)}(g^{-1}(u, v)) |\det(J_{g^{-1}}(u, v))| du dv \end{aligned}$$

ATTENTION : le domaine d'intégration est modifié lorsqu'on utilise le théorème de changement de variables.

Rappel : La matrice jacobienne J_f d'une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ avec $f(u, v) = (f_1(u, v), f_2(u, v))$ est

$$J_f = \begin{bmatrix} \frac{\partial f_1}{\partial u}(u, v) & \frac{\partial f_2}{\partial u}(u, v) \\ \frac{\partial f_1}{\partial v}(u, v) & \frac{\partial f_2}{\partial v}(u, v) \end{bmatrix}.$$

Le déterminant d'une matrice 2×2 est

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Exemple 6.19. (X, Y) de densité $f_{(X,Y)}(x, y) = \frac{6}{7}(x^2 + \frac{xy}{2})\mathbb{1}_{[0,1] \times [0,2]}(x, y)$ et on considère le couple (U, V) avec $U = X + Y$ et $V = X - Y$. Trouver la densité de (U, V) .

Soit h une fonction bornée, on a

$$\mathbb{E}[h(U, V)] = \iint_{[0,1] \times [0,2]} h(x+y, x-y) \frac{6}{7} \left(x^2 + \frac{xy}{2}\right) dx dy.$$

On pose $u = x + y$ et $v = x - y$. on a alors $x = \frac{u+v}{2}$ et $y = \frac{u-v}{2}$. La matrice jacobienne est donc

$$J = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Donc $|\det(J)| = 1/2$ et donc

$$\mathbb{E}[h(U, V)] = \iint_D h(u, v) \frac{3}{7} \left(\left(\frac{u+v}{2}\right)^2 + \frac{u^2 - v^2}{8} \right) du dv.$$

Il faut encore déterminer le domaine D . On remarque que $u \in [0, 3]$ et $v \in [-2, 1]$, avec

$$0 < u + v < 2 \quad \text{et} \quad 0 < u - v < 4.$$

Donc

$$f_{(U,V)}(u, v) = \frac{3}{28} \left((u+v)^2 + \frac{1}{2}(u^2 - v^2) \right) \mathbb{1}_D(u, v),$$

avec $D = \{(u, v) \in [0, 3] \times [-2, 1] : 0 < u + v < 2 \text{ et } 0 < u - v < 4\}$.

Remarque 6.22. Soit (X, Y) couple de densité $f_{(X,Y)}$ et $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction. Si on cherche la loi de $U = g(X, Y)$ en utilisant le calcul précédent, on pose en général $V = k(X, Y)$ de manière à ce que l'application $(x, y) \mapsto (g(x, y), k(x, y))$ soit inversible. Le choix de la fonction k dépend de l'exercice, mais souvent $V = X$, ou $V = Y$, ou $V = X + Y$ ou $V = X - Y$ sont des bonnes options.

6.5 Loi conditionnelle pour des variables discrètes

(PAS VU EN COURS, DONC PAS AU PROGRAMME DE L'EXAMEN.)

On se place sur $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

Définition 6.23. On considère deux variables aléatoires X et Y discrètes à valeurs respectivement dans $\mathbb{D}_X$ et $\mathbb{D}_Y$. On fixe $y \in \mathbb{D}_Y$.

La loi conditionnelle de X sachant que $Y = y$ est donnée, pour tout $x \in \mathbb{D}_X$ par

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

Remarque 6.24. • Si les variables X et Y sont indépendantes, alors $\forall y \in \mathbb{D}_Y$, la loi conditionnelle de X sachant que $Y = y$ est juste la loi de X :

$$p_{X|Y}(x|y) = p_X(x) \quad \forall x \in \mathbb{D}_X.$$

• La loi conditionnelle est une loi de probabilité et donc

$$\sum_{x \in \mathbb{D}_X} p_{X|Y}(x|y) = 1.$$

Exemple 6.20. On lance deux dés et on souhaite connaître la loi de la somme des chiffres sachant la valeur maximale. On note X la somme des chiffres et Y la valeur maximale. On a $\mathbb{D}_X = \{2, \dots, 12\}$ et $\mathbb{D}_Y = \{1, \dots, 6\}$.

Si $Y = 1$, on a forcément $X = 2$ et donc $p_{X|Y}(x|1) = 1$ si $x = 2$ et 0 sinon.

Si $Y = 2$, en utilisant

$$p_{X|Y}(x|y) = \frac{p_{(X,Y)}(x, y)}{p_Y(y)},$$

on a alors $p_Y(y) = \frac{3}{36}$ et donc

$$p_{X|Y}(x|2) = \begin{cases} \frac{2}{3} & \text{si } x = 3 \\ \frac{1}{3} & \text{si } x = 4 \\ 0 & \text{sinon.} \end{cases}$$

Si $Y = 5$, on a alors $p_Y(y) = \frac{5}{36}$ et donc

$$p_{X|Y}(x|5) = \begin{cases} \frac{2}{5} & \text{si } x = 4, 5 \\ \frac{1}{5} & \text{si } x = 6 \\ 0 & \text{sinon.} \end{cases}$$

Si $Y = y$, on a alors $p_Y(y) = \frac{2(y-1)+1}{36} = \frac{2y-1}{36}$ et donc

$$p_{X|Y}(x|y) = \begin{cases} \frac{2}{2y-1} & \text{si } x = y+1, \dots, 2y-1 \\ \frac{1}{2y-1} & \text{si } x = 2y \\ 0 & \text{sinon.} \end{cases}$$

Définition 6.25. On considère deux variables aléatoires X et Y discrètes à valeurs respectivement dans $\mathbb{D}_X$ et $\mathbb{D}_Y$. On fixe $y \in \mathbb{D}_Y$.

On définit l'espérance conditionnelle de X sachant que $Y = y$ comme l'espérance de la loi conditionnelle de X sachant que $Y = y$:

$$\mathbb{E}[X|Y = y] = \sum_{x \in \mathbb{D}_X} x p_{X|Y}(x|y).$$

On remarque que l'espérance conditionnelle $\mathbb{E}[X|Y = y]$ est une fonction de y : on peut définir la fonction $g(y) = \mathbb{E}[X|Y = y]$.

On va noter $\mathbb{E}[X|Y]$ la variable aléatoire $g(Y)$, c'est à dire la variable aléatoire qui vaut $\mathbb{E}[X|Y = y]$ sur l'événement $\{Y = y\}$.

Théorème 6.26 (Formule des espérances itérées). Soient X, Y deux variables d'espérance finie. La variable aléatoire $\mathbb{E}[X|Y]$ a la même espérance que X :

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

Démonstration. Comme $\mathbb{E}[X|Y = y]$ est l'espérance de la loi conditionnelle, on a

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X|Y]] &= \mathbb{E}[g(Y)] = \sum_{y \in \mathbb{D}_Y} \mathbb{E}[g(y)] p_Y(y) \\ &= \sum_{y \in \mathbb{D}_Y} \mathbb{E}[X|Y = y] p_Y(y) \\ &= \sum_{y \in \mathbb{D}_Y} \sum_{x \in \mathbb{D}_X} x p_{X|Y}(x|y) p_Y(y) = \sum_{y \in \mathbb{D}_Y} \sum_{x \in \mathbb{D}_X} x p_{(X,Y)}(x, y) \\ &= \sum_{x \in \mathbb{D}_X} x \left(\sum_{y \in \mathbb{D}_Y} p_{(X,Y)}(x, y) \right) \\ &= \sum_{x \in \mathbb{D}_X} x p_X(x) = \mathbb{E}[X]. \end{aligned}$$

□

Exemple 6.21. Soient X et Y deux variables indépendantes, de loi de Poisson de paramètre respectif λ et μ . Déterminer la loi conditionnelle de X sachant $\{X + Y = n\}$.

Comme Y est positive, on sait que si $X + Y = n$, alors on a forcément $X \leq n$.

D'après l'exercice 6.16, on sait déjà que $X + Y \sim \mathcal{P}(\lambda + \mu)$. Soit $k \in \mathbb{N}$, si $k > n$ $p_{X|X+Y}(k|n) = 0$ et si $k \leq n$

$$\begin{aligned} p_{X|X+Y}(k|n) &= \frac{\mathbb{P}(X = k, Y = n - k)}{\mathbb{P}(X + Y = n)} \stackrel{\text{Indép.}}{=} \frac{\mathbb{P}_X(k) \mathbb{P}_Y(n - k)}{\mathbb{P}_{X+Y}(n)} \\ &= \binom{n}{k} \left(\frac{\lambda}{\lambda + \mu} \right)^k \left(\frac{\mu}{\lambda + \mu} \right)^{n-k}. \end{aligned}$$

La loi conditionnelle de X sachant que $X + Y = n$ est la loi $\mathcal{B}(n, \frac{\lambda}{\lambda + \mu})$.

On déduit que l'espérance conditionnelle de X sachant que $X + Y = n$ est $n \frac{\lambda}{\lambda + \mu}$.

On a donc $\mathbb{E}[X|X + Y] = \frac{\lambda}{\lambda + \mu} (X + Y)$ avec $X + Y \sim \mathcal{Poi}(\lambda + \mu)$.

On retrouve bien $\mathbb{E}[\mathbb{E}[X|X + Y]] = \mathbb{E}[X]$.

6.6 Loi conditionnelle pour des variables à densité

Lorsqu'on avait défini la notion de probabilité conditionnelle pour les événements $\mathbb{P}(A|B)$ pour un événement B tel que $\mathbb{P}(B) \neq 0$. Par conséquent, conditionné par rapport à l'événement $\{Y = y\}$ pose problème quand Y est à densité.

Cependant, on peut quand même définir la loi conditionnelle de la manière suivante.

Définition 6.27. On considère une couple de v.a. (X, Y) de densité $f_{(X,Y)}$. Soit y tel que $f_Y(y) > 0$. On définit la densité conditionnelle de X sachant $Y = y$ par

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}.$$

Ainsi, Si $A \subset \mathbb{R}$, on a

$$\mathbb{P}(X \in A | Y = y) = \int \mathbb{1}_{x \in A} f_{X|Y}(x|y) dx.$$

Exemple 6.22. Soit (X, Y) de densité $f_{(X,Y)} = \frac{1}{y} e^{-x/y} e^{-y} \mathbb{1}_{(x,y) \in]0, +\infty[^2}$. Soit $y > 0$. On souhaite calculer $\mathbb{P}(X > 1 | Y = y)$.

On a pour $x > 0$

$$f_{X|Y}(x|y) = \frac{\frac{1}{y} e^{-x/y} e^{-y}}{\int_0^{+\infty} \frac{1}{y} e^{-x/y} e^{-y} dx} = \frac{\frac{1}{y} e^{-x/y} e^{-y}}{e^{-y}} = \frac{1}{y} e^{-x/y}.$$

Par conséquent, $X|Y = y$ suit la loi exponentielle $\mathcal{Exp}(1/y)$. Donc,

$$\mathbb{P}(X > 1 | Y = y) = \int_1^{+\infty} \frac{1}{y} e^{-x/y} dx = e^{-1/y},$$

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

Remarque 6.28. • Si X et Y sont indépendantes, on a $f_{X|Y}(x|y) = f_X(x)$.

• On définit aussi l'espérance conditionnelle comme

$$\mathbb{E}[X|Y = y] = \int x f_{X|Y}(x|y) dx,$$

et on remarque que $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$.

6.7 Covariance et corrélation

La covariance permet d'estimer la dépendance entre deux variables aléatoires. On rappelle la définition.

Définition 6.29. La covariance de deux v.a. X et Y est le réel

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

On définit la **corrélation** entre deux variables X, Y comme

$$\rho_{(X,Y)} := \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \mathbb{E} \left[\left(\frac{X - \mathbb{E}[X]}{\sigma_X} \right) \left(\frac{Y - \mathbb{E}[Y]}{\sigma_Y} \right) \right],$$

où $\sigma_X = \sqrt{\text{Var}(X)}$ et $\sigma_Y = \sqrt{\text{Var}(Y)}$.

L'espérance $\mathbb{E}[XY]$ est calculée à partir de la loi jointe de (X, Y) :

1. dans le cas discret, lorsque la somme a un sens,

$$\mathbb{E}[XY] = \sum_{x \in \mathbb{D}_X, y \in \mathbb{D}_Y} xyp_{(X,Y)}(x, y)$$

2. dans le cas continu, lorsque l'intégrale a un sens,

$$\mathbb{E}[XY] = \int \int_{\mathbb{R}^2} xyf_{(X,Y)}(x, y)dxdy.$$

Propriété 6.30. *La covariance vérifie les propriétés suivantes*

1. Si X et Y sont indépendantes, alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ et donc $\text{Cov}(X, Y) = 0$.

La réciproque est fautive ! Par contre,

Si $\text{Cov}(X, Y) \neq 0$, alors X et Y ne sont pas forcément indépendantes.

2. $\text{Cov}(X, X) = \text{Var}(X)$.
3. Symétrie : $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
4. bilinéarité : $\forall a, b \in \mathbb{R} \forall X, Y, Z$

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

et par symétrie $\text{Cov}(X, aY + bZ) = a\text{Cov}(X, Y) + b\text{Cov}(X, Z)$.

5. On a toujours $\rho_{(X,Y)} \in [-1, 1]$.

En effet, d'après l'inégalité de Cauchy-Schwartz, on a

$$|\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]| \leq \sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}.$$

On en déduit que $(a_i)_{1 \leq i \leq n}$, $(b_i)_{1 \leq i \leq n}$ des réels et $(X_i)_{1 \leq i \leq n}$, $(Y_i)_{1 \leq i \leq n}$ des variables aléatoires, alors

$$\text{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{i=1}^n b_i Y_i\right) = \sum_{i=1}^n \sum_{j=1}^n a_i b_j \text{Cov}(X_i, Y_j)$$

On dira que deux variables sont fortement corrélées (ou encore fortement dépendantes) si $|\rho(X, Y)|$ est proche de 1. D'après l'inégalité de Cauchy-Schwarz, $|\rho(X, Y)| = 1$ si et seulement si il existe des réels a, b tels que $Y = aX + b$.

Si $\rho(X, Y) = 0$ on dit que les variables ne sont pas corrélées, si $\rho(X, Y) > 0$, on dit qu'elles sont positivement corrélées et si $\rho(X, Y) < 0$, on dit qu'elles sont négativement corrélées.

Proposition 6.31. *Soient X et Y deux v.a. Alors*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

Par conséquent, si X et Y sont indépendantes, on a

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Démonstration. En développant le carré, on obtient le résultat :

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

□

Remarque 6.32. De manière générale, si $X_1, \dots, X_n$ indépendantes, on a

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

Exemple 6.23. Une étude médicale sur l'effet du tabac est menée dans un hôpital. Les 2278 patients sont divisés en deux groupes : ceux atteints d'un cancer pulmonaire ($X = 1$) et les autres ($X = 0$). Les membres de chaque groupe sont ensuite répartis selon le nombre Y de paquets de cigarettes fumés par jour.

Cancer pulmonaire	Nombre de paquets de cigarettes					Total
	0	1	2	3	4	
0	1247	492	319	58	9	2125
1	66	50	28	6	3	153
Total	1313	542	347	64	12	2278

On souhaite étudier l'association entre cancer pulmonaire et la consommation de cigarette en calculant la covariance.

La proportion de personnes atteintes d'un cancer pulmonaire est 6.72%, le nombre moyen de paquets de cigarettes consommés est 0.65, on obtient

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{1 \times 1 \times 50 + 1 \times 2 \times 28 + 1 \times 3 \times 6 + 1 \times 4 \times 3}{2278} - 0.0672 \times 0.65 \\ &= 0.02 \end{aligned}$$

La covariance est positive, le résultat indique qu'il y a un lien positif entre la déclaration du cancer et la consommation de cigarettes.

Exemple 6.24. On lance une pièce truquée 3 fois. La probabilité de tomber sur "Pile" est $2/3$. Soit X le nombre de "Face" obtenu dans les deux premiers jets et Y le nombre de "Face" obtenu dans les deux derniers jets. La loi de (X, Y) est donnée par

$y \backslash x$	0	1	2	$\mathbb{P}(Y = y)$
0	$\frac{8}{27}$	$\frac{4}{27}$	0	$4/9$
1	$\frac{4}{27}$	$\frac{6}{27}$	$\frac{2}{27}$	$4/9$
2	0	$\frac{2}{27}$	$\frac{1}{27}$	$1/9$
$\mathbb{P}(X = x)$	$4/9$	$4/9$	$1/9$	

On a

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[Y] = \frac{4}{9} + \frac{2}{9} = \frac{2}{3} \\ \mathbb{E}[X^2] &= \mathbb{E}[Y^2] = \frac{4}{9} + \frac{4}{9} = \frac{8}{9} \\ \mathbb{E}[XY] &= \sum xy p_{(X,Y)}(x, y) = \frac{6}{27} + \frac{4}{27} + \frac{4}{27} + \frac{4}{27} = \frac{2}{3} \\ \text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{2}{9} \\ \text{Var}(X) &= \text{Var}(Y) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{4}{9}. \end{aligned}$$

Par conséquent, $\rho_{(X,Y)} = \frac{1}{2}$. Comme $\rho_{(X,Y)} > 0$, donc X et Y sont corrélées positivement (plus X est grand, plus Y a tendance à être grand, et réciproquement).

Chapitre 7

Théorèmes limites

Sommaire

7.1	Inégalités importantes	81
7.2	Différentes notions de convergence	82
7.3	Théorèmes limites importants	84
a)	La loi faible des grands nombres	84
b)	Quelques simulations	85
c)	Le théorème limite centrale	86
7.4	Rappels sur les variables gaussiennes	87

On va définir ici la convergence de suite de variables aléatoires. On va voir qu'on peut définir plusieurs notions de convergence, mais on n'abordera que deux notions de convergence : la convergence en loi et la convergence en probabilité, mais il en existe d'autres. On donnera ensuite les deux principaux théorèmes de convergence, très utilisés en statistique.

7.1 Inégalités importantes

L'inégalité de Markov

Soit X une variable positive et un réel $a > 0$, alors

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[X]}{a}.$$

Démonstration. La variable X étant positive, on a

$$\mathbb{E}[X] = \mathbb{E}[X\mathbb{I}_{X>a}] + \mathbb{E}[X\mathbb{I}_{X\leq a}] \geq \mathbb{E}[X\mathbb{I}_{X>a}] \geq a\mathbb{P}(X > a).$$

□

L'inégalité de Bienaymé-Tchebichev

Soit X une variable dans L^2 et un réel $a > 0$, alors

$$\mathbb{P}(|X - \mathbb{E}[X]| > a) \leq \frac{\text{Var}(X)}{a^2}.$$

Démonstration. On applique l'inégalité de Markov à la variable $|X - \mathbb{E}[X]|^2$ et on obtient le résultat car $\mathbb{P}(|X - \mathbb{E}[X]| > a) = \mathbb{P}(|X - \mathbb{E}[X]|^2 > a^2)$. $\square$

Exercice 2. Estimer en utilisant l'inégalité de Bienaymé-Tchebichev le nombre de lancers nécessaires d'une pièce bien équilibrée pour que la moyenne du nombre de piles obtenu soit dans l'intervalle $[1/2 - 0.01, 1/2 + 0.01]$ avec une probabilité supérieure ou égale à 0.96 ?

7.2 Différentes notions de convergence

On considère une suite de variables $(X_n)_{n \geq 1}$ et la variable X .

• Convergence en probabilité

La suite $(X_n)_{n \geq 1}$ converge en probabilité vers X ssi

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

On note alors $X_n \xrightarrow[n \rightarrow +\infty]{\text{proba}} X$.

La convergence en probabilité signifie que la probabilité que X_n soit loin de X (à distance plus grande que ε pour $\varepsilon > 0$) converge vers 0 quand n tend vers $+\infty$.

Exemple 7.1. Le nombre moyen de piles sur n lancer d'une pièce bien équilibrée tend en probabilité vers $1/2$.

On note X_i la résultat du $i^{\text{ème}}$ lancer, avec $X_i = 1$ si on obtient pile. Les variables sont toutes de même loi, on a $X_i \sim \mathcal{Ber}(1/2)$. De plus, les lancers étant indépendants, les variables $(X_i)_{i \geq 1}$ sont indépendantes. Le nombre moyen de piles sur n lancers est

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Par linéarité de l'espérance, on remarque que

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \mathbb{E}[X_1 + \dots + X_n] = \frac{\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]}{n} = \frac{1}{2}.$$

Soit $\varepsilon > 0$, on a par l'inégalité de Bienaymé-Tchebychev,

$$\mathbb{P}\left(\left|\bar{X}_n - \frac{1}{2}\right| > \varepsilon\right) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2}$$

Par ailleurs,

$$\begin{aligned} \text{Var}(\bar{X}_n) &= \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) \\ &= \frac{1}{n^2} (\text{Var}(X_1) + \dots + \text{Var}(X_n)) \quad \text{car les variables sont indépendantes} \\ &= \frac{1}{n^2} \times n \times \frac{1}{4} = \frac{1}{4n} \quad \text{car les variables sont de même loi } \mathcal{Ber}(1/2). \end{aligned}$$

On déduit que pour tout $\varepsilon > 0$, $\lim_{n \rightarrow +\infty} \mathbb{P}\left(\left|\bar{X}_n - \frac{1}{2}\right| > \varepsilon\right) = 0$ et donc le nombre moyen de piles converge en probabilité vers $1/2$.

Propriété 7.1. Soient $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ deux suites de v.a. qui convergent respectivement vers X et Y en probabilité. Alors

1. si f est continue sur $\mathbb{R}$ alors $(f(X_n))_{n \geq 1}$ converge en probabilité vers $f(X)$,
2. $(X_n Y_n)_{n \geq 1}$ converge en probabilité vers XY , et $(X_n + Y_n)_{n \geq 1}$ converge en probabilité vers $X + Y$.

• Convergence en loi

La suite $(X_n)_{n \geq 1}$ converge en loi vers X ssi il y a convergence des fonctions de répartition en tout point x où F_X est **continue**

$$F_{X_n}(x) \xrightarrow{n \rightarrow +\infty} F_X(x)$$

On note alors $X_n \xrightarrow[n \rightarrow +\infty]{loi} X$.

La convergence en loi signifie que la loi de X_n converge vers la loi de X quand n tend vers $+\infty$. Par conséquent l'ensemble des valeurs de X_n converge vers l'ensemble des valeurs de X .

Proposition 7.2.

- Dans le cas général :

$X_n \xrightarrow[n \rightarrow +\infty]{loi} X$ ssi il y a convergence des fonctions génératrices des moments : pour tout $t \in \mathbb{R}$,

$$M_{X_n}(t) \xrightarrow{n \rightarrow +\infty} M_X(t).$$

- Dans le cas où (X_n) et X sont des v.a. discrètes :

$X_n \xrightarrow[n \rightarrow +\infty]{loi} X$ ssi $\forall x \in \mathbb{D}_X, \mathbb{P}(X_n = x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X = x)$.

Exemple 7.2. On montre à l'aide des fonctions génératrices des moments que si X_n suit une loi Binomiale $\mathcal{B}(n, p_n)$ avec $np_n \xrightarrow{n \rightarrow +\infty} \lambda > 0$, alors la loi de X_n converge vers la loi de Poisson $\mathcal{P}(\lambda)$.

Propriété 7.3. 1. Soit $(X_n)_{n \geq 1}$ une suite qui converge en loi vers X . Si f est continue sur $\mathbb{R}$ alors $(f(X_n))_{n \geq 1}$ converge en loi vers $f(X)$.

2. Soit f une fonction continue sur $\mathbb{R}^2$.

Si $(X_n, Y_n) \xrightarrow[n \rightarrow +\infty]{loi} (X, Y)$, alors $f(X_n, Y_n) \xrightarrow[n \rightarrow +\infty]{loi} f(X, Y)$.

Attention, si on a juste f continue, $X_n \xrightarrow[n \rightarrow +\infty]{loi} X$ et $Y_n \xrightarrow[n \rightarrow +\infty]{loi} Y$, on ne peut rien conclure sur $f(X_n, Y_n)$. Notamment on ne peut rien dire sur la convergence de $(X_n Y_n)_{n \geq 1}$ ou de $(X_n + Y_n)_{n \geq 1}$.

• Lien entre les deux notions de convergence

$$\text{"convergence en proba"} \Rightarrow \text{"convergence en loi"}$$

Attention la réciproque est fautive! Cependant, on a le résultat suivant :

Remarque 7.4. Si $X_n \xrightarrow[n \rightarrow +\infty]{loi} a$ et a est constante, alors $X_n \xrightarrow[n \rightarrow +\infty]{proba} a$

7.3 Théorèmes limites importants

a) La loi faible des grands nombres

On s'intéresse à la variable aléatoire $\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ appelée **moyenne empirique** et on souhaite connaître son comportement quand n grandit. On a le théorème suivant :

Théorème 7.5. Soit X une v.a. ayant un moment d'ordre 1 fini.

Soient $(X_n)_{n \geq 0}$ une suite de variables indépendantes et de même loi que X . Alors

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{} E[X] \quad \text{en proba.}$$

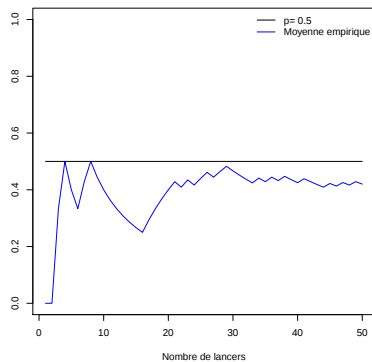
Démonstration. Preuve dans le cas où $X \in L^2$. On suppose $E[X] = 0$ (sinon il suffit de travailler avec la suite $(X_n - E[X])_{n \geq 0}$). Soit $\varepsilon > 0$. On a

$$P\left(\left|\frac{X_1 + \dots + X_n}{n}\right| \geq \varepsilon\right) \stackrel{\text{Bienaymé-Tchebichev}}{\leq} \frac{\text{Var}(X_1 + \dots + X_n)}{n^2 \varepsilon^2} \stackrel{\text{Indépendance}}{\leq} \frac{\text{Var}(X)}{n \varepsilon^2}.$$

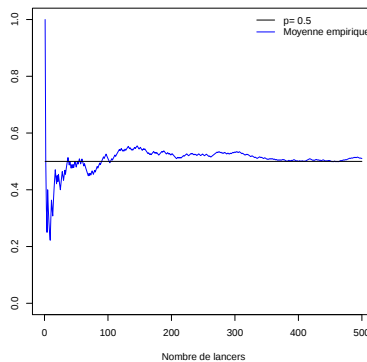
D'où le résultat. $\square$

b) Quelques simulations

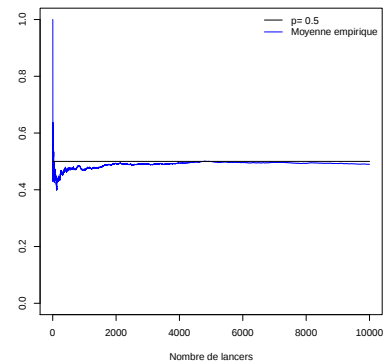
Considérons l'exemple du jeu "Pile ou Face" avec une pièce bien équilibrée. La probabilité de tomber sur "Pile" est alors $p = 1/2$. Le joueur mise sur "Pile". On répète plusieurs fois l'expérience et on observe l'évolution de la fréquence du nombre de succès $\bar{X}_n$ en fonction du nombre de lancers n :



pour $n = 50$

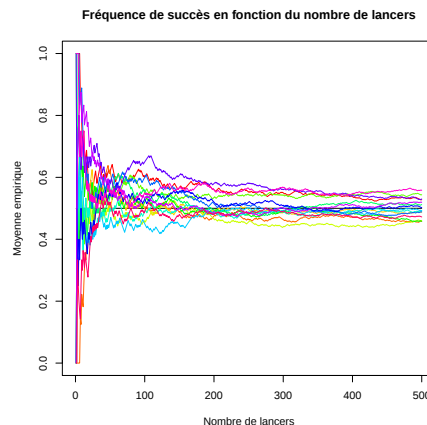


pour $n = 500$

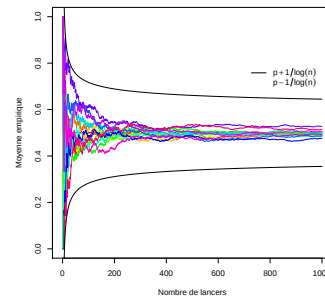
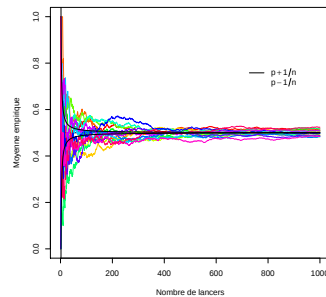


pour $n = 10\,000$.

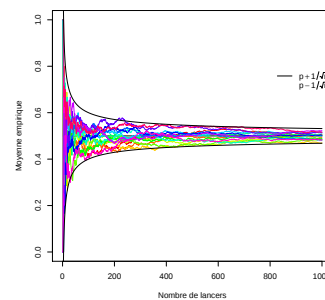
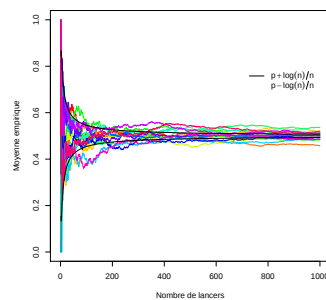
On trace plusieurs réalisations de la trajectoire aléatoire $n \mapsto \bar{X}_n$ (chaque réalisation ayant une couleur différente sur le graphique), afin d'essayer d'identifier la vitesse de convergence. On obtient le résultat suivant



On observe que la convergence est plutôt lente. Essayons différentes fonctionnelles pour évaluer la vitesse de convergence (courbe tracée en noir).



la vitesse semble moins rapide que n la vitesse semble plus rapide que $\ln(n)$



la vitesse semble moins rapide que $\frac{n}{\ln(n)}$ la vitesse semble de l'ordre de $\sqrt{n}$.

Le théorème central limite va nous fournir la vitesse de convergence de la loi des grands nombres.

c) Le théorème limite centrale

On se demande maintenant à quelle vitesse la moyenne empirique converge vers l'espérance.

Théorème 7.6. Soit X une v.a. ayant un moment d'ordre 2 fini. On pose $m = E[X]$ et $\sigma^2 = \text{Var}(X)$. Soient $X_1, X_2, \dots, X_n$ des variables indépendantes et de même loi que X . Alors

$$\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + X_2 + \dots + X_n}{n} - m \right) \xrightarrow[n \rightarrow +\infty]{} Z \quad \text{en loi, avec } \mathcal{L}(Z) = \mathcal{N}(0, 1).$$

Donc $\forall a, b \in \mathbb{R}, a < b$,

$$\mathbb{P}\left(a \leq \frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + X_2 + \dots + X_n}{n} - m \right) \leq b\right) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}(a \leq Z \leq b) = \int_a^b e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$$

La vitesse de convergence de la moyenne empirique est donc assez lente car elle est en $\frac{\sqrt{n}}{\sigma}$ (bien sûr, plus σ est petit plus la vitesse est rapide).

Exemple 7.3. 1. Si X_i sont des v.a. indépendantes de loi $\mathcal{B}(p)$, alors $S_n = \sum_{i=1}^n X_i$ suit une loi $\mathcal{B}(n, p)$. Par conséquent, pour n grand on approche

$$\frac{S_n - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0, 1)$$

2. Si X_i sont des v.a. indépendantes de loi $\mathcal{P}(\lambda)$, alors $S_n = \sum_{i=1}^n X_i$ suit une loi $\mathcal{P}(n\lambda)$. Par conséquent, pour n grand on approche

$$\frac{S_n - n\lambda}{\sqrt{n\lambda}} \sim \mathcal{N}(0, 1)$$

Règles d'usage

1. On approche la loi $\mathcal{P}(\lambda)$ par $\mathcal{N}(\lambda, \lambda)$ dès que $\lambda \geq 10$.
2. On approche la loi $\mathcal{B}(n, p)$ par $\mathcal{N}(np, np(1-p))$ dès que $np \geq 10$ et $n(1-p) \geq 10$.
3. Si $np < 10$ et $p \leq 0.1$, on approche la loi $\mathcal{B}(n, p)$ par $\mathcal{P}(np)$.

Exemple 7.4. Un restaurateur peut servir 75 repas, uniquement sur réservation. La pratique montre que 20% des clients ayant réservés ne viennent pas.

Combien le restaurateur doit-il accepter de réservations indépendantes pour avoir une probabilité supérieure ou égale à 0.9 de pouvoir servir tous les clients qui se présenteront ?

Notons n le nombre de réservation que le restaurateur accepte. Le nombre de personnes qui viennent effectivement au restaurant suit la loi binomiale $\mathcal{Bin}(n, 0.8)$. On note $X_i = \mathbf{1}_{\text{le } i^{\text{ème}} \text{ client vient au restaurant}}$.

On cherche le plus grand n tel que

$$\mathbb{P}\left(\sum_{i=1}^n X_i \leq 75\right) \geq 0.9,$$

ce qu'on peut réécrire

$$\mathbb{P}\left(\sum_{i=1}^n X_i \leq 75\right) = \mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{0.8 * 0.2}}(\bar{X}_n - 0.8) \leq \frac{\sqrt{n}}{\sqrt{0.8 * 0.2}}\left(\frac{75}{n} - 0.8\right)\right)$$

D'après le théorème central limite,

$$\mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{0.8 * 0.2}}(\bar{X}_n - 0.8) \leq \frac{\sqrt{n}}{\sqrt{0.8 * 0.2}}\left(\frac{75}{n} - 0.8\right)\right) \simeq \mathbb{P}\left(Z \leq \frac{\sqrt{n}}{\sqrt{0.8 * 0.2}}\left(\frac{75}{n} - 0.8\right)\right)$$

D'après la table gaussienne, cette probabilité vaut 0.9 si $\frac{\sqrt{n}}{\sqrt{0.8 * 0.2}}\left(\frac{75}{n} - 0.8\right) \simeq 1.285$. En résolvant cette équation du second de degré on obtient $n = 93$ (car on doit avoir $\frac{75}{n} - 0.8 > 0$).

Exercice 3. Une compagnie d'assurance se propose d'assurer 100000 clients contre le vol. Les sommes en euros (la plupart du temps nulles) $X_1, \dots, X_{100000}$ qu'aura à rembourser chaque année la compagnie aux clients sont des v.a. indépendantes d'espérance 75 et d'écart type 750.

Quelle somme cette compagnie d'assurance doit-elle faire payer à chaque client par an pour que ses frais évalués à 1,5 millions d'euros soient couverts avec une probabilité supérieure ou égale à 0.999 ?

7.4 Rappels sur les variables gaussiennes

Définition 7.7. Une variable aléatoire **gaussienne** (ou **normale**) X d'espérance m et de variance σ^2 a pour densité

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad \text{avec } x \in \mathbb{R}.$$

On dit que X suit la loi $\mathcal{N}(m, \sigma^2)$.

Proposition 7.8. 1. La fonction génératrice des moments d'une variable aléatoire X de loi $\mathcal{N}(m, \sigma^2)$ est :

$$L_X(t) = E[e^{tX}] = e^{mt} e^{\frac{\sigma^2 t^2}{2}} \quad \text{pour } t \in \mathbb{R}.$$

2. Si $a, b \in \mathbb{R}$, X une variable gaussienne $\mathcal{N}(m, \sigma^2)$, alors $Y = aX + b$ est une variable gaussienne $\mathcal{N}(am + b, a^2\sigma^2)$.
3. Soit $X_1, \dots, X_n$ des variables gaussiennes réelles indépendantes d'espérance m_i et de variance σ_i^2 respectives. Soient $a_1, \dots, a_n$ des réels, alors $Y = a_1X_1 + \dots + a_nX_n$ est une variable gaussienne $\mathcal{N}(\sum_{i=1}^n a_i m_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$.

Démonstration. 1. Calculons la fonction génératrice des moments d'une variable gaussienne :

$$\begin{aligned} L_X(t) &= E[e^{tX}] = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} e^{tx} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-m-\sigma^2 t)^2 + m^2 - (m+\sigma^2 t)^2}{2\sigma^2}} dx \\ &= e^{\frac{\sigma^4 t^2 + 2\sigma^2 mt}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-m-\sigma^2 t)^2}{2\sigma^2}} dx = e^{mt} e^{\frac{\sigma^2 t^2}{2}}. \end{aligned}$$

On reconnaît la densité de $\mathcal{N}(m + \sigma^2 t, \sigma^2)$ dans l'intégrale, on a donc $\frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(x-m-\sigma^2 t)^2}{2\sigma^2}} dx = 1$.

2. On calcule la fonction génératrice des moments de Y :

$$\begin{aligned} L_Y(t) &= E[e^{atX+bt}] = e^{bt} E[e^{(at)X}] = e^{bt} e^{mat} e^{\sigma^2 a^2 t^2 / 2} \\ &= e^{(am+b)t} e^{(a\sigma)^2 t^2 / 2} \end{aligned}$$

On reconnaît la fonction génératrice des moments de la loi gaussienne $\mathcal{N}(am + b, a^2\sigma^2)$, et comme la fonction génératrice des moments caractérise la loi, on a le résultat.

3. On utilise encore la fonction génératrice des moments. Comme les variables X_i sont indépendantes, la fonction génératrice des moments de la somme est le produit des transformées de Laplace :

$$g_Y(t) = E[e^{ta_1X_1 + \dots + ta_nX_n}] = \prod_{i=1}^n E[e^{ta_iX_i}] = \prod_{i=1}^n e^{\sum_{i=1}^n a_i m_i t} e^{\sum_{i=1}^n a_i^2 \sigma_i^2 t^2 / 2}$$

La variable Y suit par conséquent la loi gaussienne $\mathcal{N}(\sum_{i=1}^n a_i m_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$. □

Remarque 7.9. 1. La loi gaussienne est bien connue. N'importe quel ordinateur donne les valeurs de la fonction de répartition.

2. Soit X une variable de loi $\mathcal{N}(0, 1)$, alors les valeurs de X sont "concentrées" entre -4 et 4 . En effet, la probabilité $\mathbb{P}(|X| > 4) = 5.10^{-5}$ est très faible.
3. Du fait de la symétrie de la loi Normale $\mathcal{N}(0, 1)$ par rapport à 0 , on a

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(X \geq -x) = 1 - \mathbb{P}(X \leq x), \\ \text{et donc } \mathbb{P}(|X| \leq x) &= 2\mathbb{P}(X \leq x) - 1 = 2F_X(x) - 1. \end{aligned}$$

On donne maintenant la table de la loi gaussienne qui permet de connaître la valeur de la fonction de répartition F d'une loi $\mathcal{N}(0, 1)$ en fonction de la valeur de x .

FONCTION DE RÉPARTITION DE LA LOI NORMALE STANDARD

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

z	0.841	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
$\Phi(z)$	0.8000	0.9000	0.9500	0.9750	0.9800	0.9900	0.9950	0.9975	0.9990	0.9995