
0. TABLE DES MATIÈRES

1	Analyse Combinatoire	2
1.1	Principe fondamental de dénombrement	2
1.2	Permutations	2
1.3	Combinaisons	3
2	Axiome des probabilités	5
2.1	Ensemble fondamental et événements	6
2.2	Axiomes des probabilités	8
2.3	Ensembles fondamentaux à probabilités équiprobables	12
3	Probabilités conditionnelles	14
3.1	Formule des probabilités totales et formule de Bayes	16
3.2	Indépendance de deux événements	18
3.3	Indépendance de trois événements ou plus	19
3.4	Suite d'épreuves indépendantes	20
3.5	Fonction de probabilité conditionnelle	22
4	Variables Aléatoires	23
4.1	Variables aléatoires discrètes	25
4.2	Exemples de variables aléatoires discrètes	30
4.2.1	Variable aléatoire de Bernoulli et binomiale	30
4.2.2	Variable aléatoire de Poisson	32
4.2.3	Variable aléatoire géométrique	34
4.2.4	Variable aléatoire binomiale négative	37
4.2.5	Variable aléatoire hypergéométrique	38
4.3	Espérance de la somme de deux variables aléatoires discrètes	39
4.4	Variables aléatoires continues	40
4.5	Exemples de variables aléatoires continues	46
4.5.1	Variable aléatoire uniforme	46
4.5.2	Variable aléatoire gaussienne	47
4.5.3	Variable aléatoire exponentielle	50
4.5.4	Variable aléatoire Gamma	52
4.6	Fonction d'une variable continue	53
5	Variables aléatoires simultanées	55
5.1	Variables aléatoires indépendantes	58
5.2	Somme de variables aléatoires indépendantes	61
5.3	Distributions conditionnelles	65
5.4	Changement de variables	67
6	Espérance Mathématique	72
6.1	Espérance conditionnelle	78
6.2	Fonction génératrice des moments	85

7 Théorèmes limites	89
7.1 Inégalités de probabilité	89
7.2 Loi des grands nombres	94
7.3 Théorème central limite	96

1. ANALYSE COMBINATOIRE

Commençons par un exemple et considérons 4 antennes qui ont deux états possibles : chaque antenne est soit fonctionnelle ou soit défectueuse. On place alors les antennes sur une ligne et on dira que le réseau fonctionnera s'il n'y a pas deux antennes défectueuses consécutives. Si on sait que parmi les 4 antennes, 2 sont défectueuses, quelle est la probabilité que le réseau fonctionne ? Dans ce cas simple, on peut regarder simplement tous les états possibles. Si 0 représente une antenne défectueuse et 1 une antenne fonctionnelle alors tous les états sont donnés par la liste suivante

0011 → Défectueux 0101 → Fonctionnel 1010 → Fonctionnel
 0110 → Fonctionnel 1001 → Défectueux 1100 → Défectueux

Ainsi le réseau sera fonctionnel avec probabilité $\frac{3}{6} = \frac{1}{2}$ car il y a 3 réseaux fonctionnels sur les 6 réseaux possibles de 4 antennes avec 2 antennes défectueuses. Dans ce cas précis, nous pouvions faire le dénombrement explicite de tous les états possibles mais cela n'est plus possible si l'on veut considérer le cas général de n antennes avec m antennes défectueuses. Nous allons donc faire de l'analyse combinatoire.

1.1 Principe fondamental de dénombrement

Nous commençons par donner le principe le plus important de dénombrement.

THÉORÈME 1.1: Principe de dénombrement

Si on a r expériences avec la première ayant n_1 résultats possibles, la deuxième $n_2, \dots$, la r -ième n_r alors on a un total de

$$n_1 \times n_2 \times \dots \times n_r = \prod_{i=1}^r n_i$$

résultats possibles.

EXEMPLE 1. Un comité est constitué de 3 étudiants de première année, 4 de deuxième année, 5 de troisième année et 2 de quatrième année. Nous voulons créer un sous-comité de 4 étudiants avec un élève de chaque classe. Combien y a-t-il de tels sous-comités possibles ?

Nous avons 4 expériences pour le choix de chaque étudiant : 3 résultats possibles pour le choix de l'étudiant de première année, 4 pour celui de deuxième année, 5 pour celui de troisième année et 2 résultats pour celui de quatrième année. Ainsi le nombre total est de $3 \times 4 \times 5 \times 2 = 120$ sous-comités possibles.

EXEMPLE 2. Combien y a-t-il de fonctions $f : \{1, \dots, n\} \rightarrow \{0, 1\}$?

Nous avons n expériences pour chaque valeur de $f(i)$ pour i allant de 1 à n et chaque expérience a 2 résultats : 0 ou 1. Ainsi le nombre total de telles fonctions est de $2 \times \dots \times 2 = \prod_{i=1}^n 2 = 2^n$.

1.2 Permutations

Si on se donne les 3 lettres a, b, c . Combien y a-t-il de manières d'arranger les lettres ? On voit que l'on peut écrire

$abc \quad acb \quad bac \quad bca \quad cab \quad cba$

et on a ainsi 6 manières d'arranger ces lettres. Chaque arrangement de lettres est appelé une *permutation* des lettres a, b et c .

En utilisant le principe de dénombrement, on voit que l'on a 3 expériences : la première a 3 résultats possibles (a , b ou c). La deuxième expérience a 2 résultats possibles (si a a été choisie lors de la première expérience alors il y a soit b ou c). Enfin, pour la troisième expérience, on a qu'un résultat qui est la lettre restante. Ainsi par le principe de dénombrement on retrouve bien un nombre de permutations de $3 \times 2 \times 1 = 6$. On peut faire le même raisonnement pour n objets discernables et on obtient le résultat suivant.

THÉORÈME 1.2

Le nombre de permutations à n objets est de

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

EXEMPLE 3. Dans un tournoi d'échec, on a 6 hommes et 4 femmes avec un classement final. Combien y a-t-il de classements finaux possibles en comptant tous les participants (sans égalité)?

On voit que l'on a $6 + 4 = 10$ objets discernables et a donc $10! = 3\,628\,800$ permutations possibles. On sépare maintenant le tournoi en un tournoi homme et un tournoi femme. Combien y a-t-il de classements?

On a maintenant 2 expériences : le tournoi homme a $6! = 720$ résultats possibles et le tournoi femme a $4! = 24$ résultats possibles. Ainsi par le principe de dénombrement on a $6! \times 4! = 17\,280$ classements possibles.

Dans ces exemples, tous les objets sont discernables entre eux, que se passe-t-il lorsque certains objets sont indiscernables? Par exemple, combien y a-t-il de façons d'arranger les lettres du mot SERREE? Si on numérotait les lettres pour les discerner $S_1E_1R_1R_2E_2E_3$ alors on aurait 6 objets indiscernables et ainsi $6!$ permutations possibles. Mais si on fixe une permutation, par exemple $R_1R_2E_1S_1E_2E_3$ alors nous avons beaucoup de permutations qui donne le même motif,

$$\begin{array}{lll} R_1R_2E_1S_1E_2E_3 & R_2R_1E_1S_1E_2E_3 & R_1R_2E_2S_1E_1E_3 \\ R_2R_1E_2S_1E_1E_3 & R_1R_2E_1S_1E_3E_2 & R_2R_1E_1S_1E_3E_2 \\ R_1R_2E_3S_1E_1E_2 & R_2R_1E_3S_1E_1E_2 & R_1R_2E_3S_1E_2E_1 \\ R_2R_1E_3S_1E_2E_1 & R_1R_2E_2S_1E_3E_1 & R_2R_1E_2S_1E_3E_1 \end{array}$$

Il y a donc 12 permutations qui donne RRESEE. En fait, on voit que l'on a 3 expériences : l'ordre des R qui ont 2 résultats possibles R_1R_2 ou R_2R_1 , l'ordre des E qui ont $3! = 6$ résultats possibles et l'ordre des S qui n'a qu'un seul résultat. Ainsi on a $3! \times 2! \times 1 = 12$ résultats par le principe de dénombrement. Chaque permutation est donc comptée 12 fois et le nombre total est donc de $\frac{6!}{3!2!} = 60$ permutations possibles de RRESEE. Ce résultat se généralise facilement par le théorème suivant.

THÉORÈME 1.3

Il y a $\frac{n!}{n_1!n_2!\cdots n_r!}$ permutations différentes de n objets parmi lesquels n_1 sont indiscernables entre eux, n_2 autres aussi, $\dots$, n_r autres aussi.

1.3 Combinaisons

Considérons un autre problème et comptons le nombre de groupes de 3 objets que l'on peut former à partir de $\{1, 2, 3, 4, 5\}$. Ici on parle de groupes non ordonnés ou *combinaison*.

On pourrait tout d'abord vouloir utiliser le principe de dénombrement et dire que pour le premier objet nous avons 5 choix puis 4 choix pour le deuxième et enfin 3 choix pour le troisième qui nous donnerait par le principe de dénombrement $5 \times 4 \times 3 = 60$ possibilités. Mais attention, on compte plusieurs fois le même groupe! En effet, les groupes $\{1, 2, 3\} = \{3, 2, 1\} = \{1, 3, 2\}$ sont les mêmes par exemple. Pour un groupe de 3 objets, on voit ainsi que l'on a $3!$ permutations possibles du groupe et ainsi le nombre de combinaisons de 3 objets parmi 5 est de $\frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10$.

De manière générale, pour r objets parmi n on a

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

THÉORÈME 1.4

Le nombre de combinaisons de k objets parmi n est de

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

EXEMPLE 4. Reprenons l'exemple des antennes : on a n antennes et m sont défectueuses. Si deux antennes consécutives sont défectueuses alors le réseau ne fonctionne pas. Quelle est la probabilité que le réseau fonctionne ?

Tout d'abord, combien y a-t-il de possibilités totales ? On peut remarquer qu'il suffit de choisir où sont les antennes défectueuses et on a ainsi $\binom{n}{m}$ possibilités. Maintenant, nous cherchons le nombre de réseaux qui fonctionnent. Considérons les $n - m$ antennes qui fonctionnent, représentée par des 1

$$\uparrow 1 \uparrow 1 \uparrow 1 \uparrow 1 \uparrow 1 \uparrow 1 \uparrow 1 \uparrow.$$

Entre chaque antenne fonctionnelle, on peut placer des antennes défectueuses mais on ne peut avoir au plus qu'une antenne défectueuse entre chaque antenne fonctionnelle sinon le réseau est défectueux. On remarque que les emplacements sont données par les flèches et que s'il y a $n - m$ antennes fonctionnelles, il y a $n - m + 1$ emplacements. On le voit facilement ici avec $n - m = 8$ puisqu'il y a 9 flèches.

On doit ainsi choisir les emplacements des m antennes défectueuses. On remarque que si $m > n - m + 1$ ou $m > \frac{n+1}{2}$ alors par le principe des tiroirs, il y a forcément un emplacement qui contient deux antennes défectueuses et le réseau sera nécessairement défectueux et la probabilité recherchée est donc de 0. Sinon, on a $\binom{n-m+1}{m}$ choix possibles. Par exemple si $m = 4$, on peut faire le choix

$$\uparrow 1 \uparrow 1 \uparrow 1 \uparrow 1 \uparrow 1 \uparrow 1 \uparrow 1 \uparrow \rightarrow 011010110111.$$

Ainsi, la probabilité recherchée est de

$$\frac{\binom{n-m+1}{m}}{\binom{n}{m}} = \frac{\frac{(n-m+1)!}{m!(n-2m+1)!}}{\frac{n!}{m!(n-m)!}} = \frac{(n-m+1)!(n-m)!}{n!(n-2m+1)!}.$$

Dans le cas $n = 4$ et $m = 2$, on retrouve bien

$$\frac{\binom{3}{2}}{\binom{4}{2}} = \frac{3!2!}{4!1!} = \frac{1}{2}.$$

Dans le cas illustré ci-dessus avec $n = 12$ et $m = 4$, on obtient une probabilité de

$$\frac{9!8!}{12!5!} \approx 0.254545.$$

THÉORÈME 1.5

Les coefficients $\binom{n}{k}$ sont appelés *coefficients binomiaux* car pour tout $x, y \in \mathbb{R}$ nous avons

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

Démonstration. On pourrait faire une preuve par récurrence mais on peut aussi faire une preuve combina-

toire de ce résultat. En effet, on peut écrire

$$(x + y)^n = (x + y)(x + y)(x + y) \dots (x + y).$$

Lorsque l'on développe ce terme, pour avoir k fois x (et ainsi $n - k$ fois y), on doit choisir k fois x parmi les n termes et on a donc par définition des coefficients binomiaux $\binom{n}{k}$ possibilités. Comme on a toutes les possibilités de $x^k y^{n-k}$ pour $k \in \{0, \dots, n\}$, cela nous donne la formule

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

□

Une remarque importante et que pour $x = y = 1$, on obtient la formule

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

On peut généraliser ces coefficients en considérons le problème suivant : supposons que l'on a n objets distincts que nous divisons en r groupes de taille respective $n_1, n_2, \dots, n_r$ avec $\sum_{i=1}^r n_i = n$, combien y a-t-il de manières de faire ?

Pour le premier groupe, on voit que l'on a $\binom{n}{n_1}$ choix. Après ce choix, nous avons $\binom{n-n_1}{n_2}$ choix pour le deuxième groupe puis $\binom{n-n_1-n_2}{n_3}$ pour le troisième et ainsi de suite jusqu'à $\binom{n-n_1-n_2-\dots-n_{r-1}}{n_r}$ choix pour le dernier groupe. Par le principe de dénombrement le nombre de manières est donc de

$$\begin{aligned} & \binom{n}{n_1} \binom{n-n_1}{n_2} \dots \binom{n-n_1-n_2-\dots-n_{r-1}}{n_r} \\ &= \frac{n!}{n_1! (n-n_1)!} \times \frac{(n-n_1)!}{n_2! (n-n_1-n_2)!} \times \dots \times \frac{(n-n_1-n_2-\dots-n_{r-1})!}{n_r! (n-n_1-n_2-\dots-n_{r-1}-n_r)!} \\ &= \frac{n!}{n_1! n_2! \dots n_r! (n-n_1-n_2-\dots-n_r)!} \end{aligned}$$

On remarque maintenant que par définition $n - n_1 - n_2 - \dots - n_r = 0$ et que par convention $0! = 1$ et on obtient ainsi le résultat suivant.

THÉORÈME 1.6

Le nombre de façon de diviser n objets distincts en r groupes de tailles respectives $n_1, n_2, \dots, n_r$ avec $\sum_{i=1}^r n_i = n$ est de

$$\frac{n!}{n_1! n_2! \dots n_r!} =: \binom{n}{n_1, n_2, \dots, n_r}$$

Ces coefficients sont appelés *coefficients multinomiaux* car pour tout $x_1, \dots, x_r \in \mathbb{R}$ nous avons

$$(x_1 + \dots + x_r)^n = \sum_{\substack{n_1, \dots, n_r=0 \\ n_1 + \dots + n_r = n}}^n \binom{n}{n_1, n_2, \dots, n_r} x_1^{n_1} x_2^{n_2} \dots x_r^{n_r}.$$

2. AXIOME DES PROBABILITÉS

En probabilités, on réalise des expériences dont le résultat n'est pas connue exactement. Par exemple, un lancer d'une pièce peut donner Pile ou Face, un dé peut donner une valeur entre 1 et 6. L'ensemble des résultats possibles d'une expérience s'appelle l'ensemble fondamental ou univers.

2.1 Ensemble fondamental et événements

DÉFINITION 2.1

L'ensemble fondamental, ou univers, dénoté Ω est l'ensemble de tous les résultats possibles.

EXEMPLE 5. — Si on lance une pièce de monnaie alors $\Omega = \{\text{Pile}, \text{Face}\}$.

- Si on lance un dé à 6 faces alors $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Si on lance deux dés et on garde la somme alors $\Omega = \{2, 3, 4, \dots, 12\}$.
- Si on considère l'ordre d'arrivée d'une course entre 7 chevaux alors

$$\Omega = \{\text{toutes les permutations de } (1, 2, \dots, 7)\}.$$

En particulier $\#\Omega = 7!$.

- Si on jette deux pièces de monnaie alors $\Omega = \{(P, P), (P, F), (F, P), (F, F)\}$.
- Si on considère la durée de vie d'une machine à laver (en mois) alors $\Omega = \mathbb{R}_+ = \{x \in \mathbb{R}, x \geq 0\}$.

DÉFINITION 2.2

Un sous-ensemble E de Ω est appelé un événement.

EXEMPLE 6. Si $\Omega = \{(P, P), (P, F), (F, P), (F, F)\}$ alors $E = \{\text{la première pièce est Pile}\} = \{(P, P), (P, F)\} \subset \Omega$ est un événement.

Si $\Omega = \mathbb{R}_+$, alors $E = \{\text{la machine a une durée de vie d'au moins 3 mois}\} = \{x \in \mathbb{R}, x \geq 3\} \subset \Omega$ est aussi un événement.



Ce n'est pas la définition exacte d'un événement, ces sous-ensembles doivent être des ensembles *mesurables* mais nous ne considérerons pas ces subtilités dans ce cours.

Nous avons plusieurs opérations sur les événements que nous décrivons dans la définition suivante.

DÉFINITION 2.3

On définit les opérations suivantes sur les événements

- Union : Si $E, F \subset \Omega$ alors $E \cup F \subset \Omega$ est l'événement qui contient chaque élément de E ou de F .
- Intersection : Si $E, F \subset \Omega$ alors $E \cap F \subset \Omega$ est l'événement des réalisations qui sont dans E et dans F .
- Complémentaire : Si $E \subset \Omega$ est un événement alors $E^c \subset \Omega$ est l'événement des réalisations qui ne sont pas dans E .

Par exemple si on considère le lancer de deux pièces et $E = \{(P, P), (P, F)\}$ et $F = \{(F, P), (P, P)\}$ alors nous avons

$$E \cup F = \{(P, P), (P, F), (F, P)\} = \{\text{On obtient au moins un Pile}\}.$$

De même, nous avons

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

On remarque que si par exemple $E = \{(P, P)\}$ et $F = \{(F, F)\}$ alors il n'y a pas d'éléments dans l'intersection de ces deux ensembles, on écrit alors

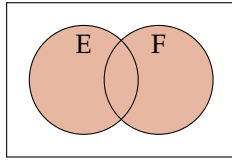
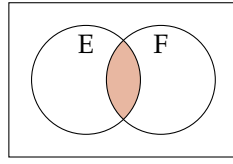
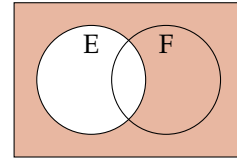
$$E \cap F = \emptyset.$$

On peut généraliser ces définitions à une suite d'ensembles plutôt que deux ensembles. Si $E_1, E_2, \dots \subset \Omega$ alors on peut définir les événements $\bigcup_{n=1}^{+\infty} E_n$ et $\bigcap_{n=1}^{+\infty} E_n$.

Enfin, en gardant toujours le lancer de deux pièces, on voit que si $E = \{(P, P), (P, F)\}$ alors $E^c = \{(F, F), (F, P)\}$. On remarque que par définition, on a toujours $E \cup E^c = \Omega$ et $E \cap E^c = \emptyset$. Au passage, cela nous donne que l'on a

$$S^c = \emptyset \quad \text{et} \quad \emptyset^c = S.$$

On peut représenter ces opérations en utilisant un diagramme de Venn.


 $E \cup F$

 $E \cap F$

 E^c

THÉORÈME 2.4

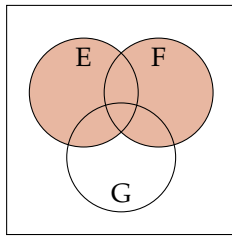
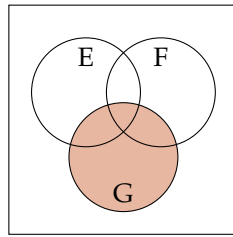
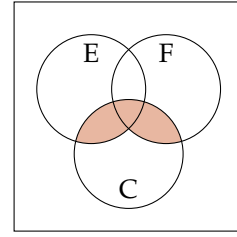
On a les égalités suivantes pour $E, F, G \subset \Omega$

- Commutativité : $E \cap F = F \cap E$ et $E \cup F = F \cup E$.
- Associativité : $(E \cap F) \cap G = E \cap (F \cap G)$ et $(E \cup F) \cup G = E \cup (F \cup G)$
- Distributivité : $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ et $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.

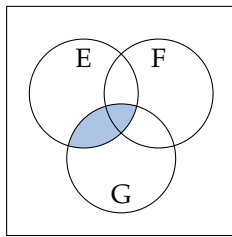
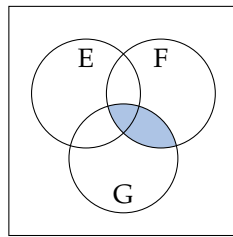
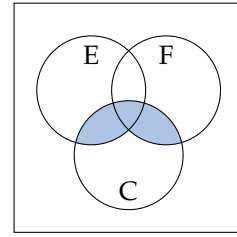
Démonstration. Pour prouver une égalité d'ensemble $A = B$ on prouve souvent $A \subset B$ puis $B \subset A$ et pour prouver que $A \subset B$, on prend un élément quelconque de A et on montre qu'il appartient aussi à B . Ici, on peut aussi prouver ces égalités d'ensemble en utilisant les diagrammes de Venn. Faisons un exemple pour l'égalité

$$(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$$

alors l'ensemble pour l'ensemble $(E \cup F) \cap G$ on peut colorier $E \cup F$ puis G et prendre ensuite la partie qui est coloriée dans les deux diagrammes. Ce sera l'intersection.


 $E \cup F$

 G

 $(E \cup F) \cap G$

Pour l'ensemble $(E \cap G) \cup (F \cap G)$ on peut regarder $E \cap G$ et $F \cap G$ séparément puis prendre l'union des deux ensembles.


 $E \cap G$

 $F \cap G$

 $(E \cap G) \cup (F \cap G)$

On voit bien ainsi que l'ensemble coloriée est le même dans les deux cas. □

Nous avons aussi des relations entre le complémentaire et l'union et l'intersection appelées les règles de De Morgan.

THÉORÈME 2.5

Soient $E_1, \dots, E_n \subset \Omega$ alors

$$\left(\bigcup_{k=1}^n E_k \right)^c = \bigcap_{k=1}^n E_k^c \quad \text{et} \quad \left(\bigcap_{k=1}^n E_k \right)^c = \bigcup_{k=1}^n E_k^c.$$

Démonstration. Soit $x \in (\bigcup_{k=1}^n E_k)^c$ alors par définition du complémentaire $x \notin \bigcup_{k=1}^n E_k$ et ainsi pour tout $k \in \{1, \dots, n\}$, $x \notin E_k$ ce que l'on peut réécrire par pour tout $k \in \{1, \dots, n\}$, $x \in E_k^c$ et ainsi nous avons $x \in \bigcap_{k=1}^n E_k^c$. Ainsi, nous avons prouvé que

$$\left(\bigcup_{k=1}^n E_k \right)^c \subset \bigcap_{k=1}^n E_k^c.$$

Pour l'autre inclusion, soit $x \in \bigcap_{k=1}^n E_k^c$ alors par définition de l'intersection, pour tout $k \in \{1, \dots, n\}$, $x \in E_k^c$ et par définition du complémentaire pour tout $k \in \{1, \dots, n\}$ nous avons $x \notin E_k$ et ainsi $x \notin \bigcup_{k=1}^n E_k$ et ainsi $x \in (\bigcup_{k=1}^n E_k)^c$ ce qui nous donne que

$$\bigcap_{k=1}^n E_k^c \subset \left(\bigcup_{k=1}^n E_k \right)^c \text{ et ainsi } \bigcap_{k=1}^n E_k^c = \left(\bigcup_{k=1}^n E_k \right)^c.$$

Pour la deuxième formule, on peut utiliser la première formule que l'on vient de prouver! En effet, si on applique la formule à la suite d'événements $E_1^c, \dots, E_n^c$ alors la formule devient

$$\left(\bigcup_{k=1}^n E_k^c \right)^c = \bigcap_{k=1}^n (E_k^c)^c \Leftrightarrow \left(\bigcup_{k=1}^n E_k^c \right)^c = \bigcap_{k=1}^n E_k \Leftrightarrow \bigcup_{k=1}^n E_k^c = \left(\bigcap_{k=1}^n E_k \right)^c$$

où on a passé au complémentaire dans la dernière équivalence. $\square$

2.2 Axiomes des probabilités

La théorie des probabilités de ce cours sera entièrement basée sur les trois axiomes de la définition suivante.

DÉFINITION 2.6

Soit Ω un ensemble fondamental, pour chaque événement E il existe une valeur $\mathbb{P}(E)$ ou probabilité de E telle que

- $0 \leq \mathbb{P}(E) \leq 1$.
- $\mathbb{P}(\Omega) = 1$.
- Si $E_1, E_2, \dots$ est une suite d'ensembles mutuellement exclusifs, ou 2 à 2 disjoints, i.e $E_i \cap E_j = \emptyset$ pour tout $i \neq j$, alors

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

EXEMPLE 7. On considère le résultat d'un dé équilibré $\Omega = \{1, 2, 3, 4, 5, 6\}$ avec

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \dots = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

Si on considère l'événement $\{\text{le résultat est pair}\} = \{2, 4, 6\}$ alors en utilisant le troisième axiome des probabilités, on obtient

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

On remarque que l'on pouvait utiliser l'axiome des probabilités car les événements $\{2\}$, $\{4\}$ et $\{6\}$ sont disjoints, le résultat ne peut pas à la fois être 2 et 4 par exemple.

Le théorème suivant nous donne des propriétés importantes sur les probabilités d'événements. On va le prouver en n'utilisant seulement les axiomes de probabilité ainsi que la théorie des ensembles que nous avons vu dans la sous-section précédente.

THÉORÈME 2.7

Soient E, F deux événements, on a

- $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
- Si $E \subset F$, $\mathbb{P}(E) \leq \mathbb{P}(F)$.
- On a $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$.

Démonstration. On remarque que par définition de complémentaire on a $E \cup E^c = \Omega$ et $E \cap E^c = \emptyset$. Ainsi, en utilisant le deuxième axiome, on obtient que

$$\mathbb{P}(E \cup E^c) = \mathbb{P}(\Omega) = 1.$$

De plus, en utilisant le troisième axiome, on obtient que

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

En combinant ces deux équations, on obtient

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

Pour la deuxième propriété, on note que comme $E \subset F \subset \Omega$, nous avons

$$F = E \cup F = \Omega \cap (E \cup F) = (E \cup E^c) \cap (E \cup F) = E \cup (E^c \cap F)$$

où on a utilisé la distributivité de l'union sur l'intersection. Maintenant, on a E et $E^c \cap F$ qui sont mutuellement exclusifs car

$$E \cap (E^c \cap F) = (E \cap E^c) \cap F = \emptyset \cap F = \emptyset$$

en utilisant l'associativité de l'intersection. Ainsi par le troisième axiome des probabilités, nous obtenons que

$$\mathbb{P}(F) = \mathbb{P}(E \cup (E^c \cap F)) = \mathbb{P}(E) + \mathbb{P}(E^c \cap F) \geq \mathbb{P}(E)$$

car $\mathbb{P}(E^c \cap F) \in [0, 1]$ par le premier axiome des probabilités.

Pour la dernière propriété, on commence par utiliser la même formule qu'on a prouvé précédemment

$$E \cup F = E \cup (E^c \cap F)$$

avec

$$E \cap (E^c \cap F) = \emptyset$$

et ainsi par le troisième axiome des probabilités,

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(E^c \cap F). \quad (1)$$

Maintenant, on peut écrire

$$F = \Omega \cap F = (E \cup E^c) \cap F = (E \cap F) \cup (E^c \cap F).$$

De plus, cette union est disjointe car

$$(E \cap F) \cap (E^c \cap F) = (E \cap E^c) \cap F = \emptyset.$$

Ainsi, on a

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

Lorsque l'on remplace cette formule dans (1), on obtient la formule voulue. $\square$

EXEMPLE 8. Jeanne a deux livres : elle aimera le premier avec probabilité $\frac{1}{2}$, le deuxième avec probabilité $\frac{2}{5}$ et les deux avec probabilité $\frac{3}{10}$. Quelle est la probabilité qu'elle n'aime aucun livre ?

On considère l'événement

$$\begin{aligned} A &= \{\text{elle n'aime aucun livre}\} \\ &= \{\text{elle aime au moins un livre}\}^c \\ &= (\{\text{elle aime le premier livre}\} \cup \{\text{elle aime le deuxième livre}\})^c. \end{aligned}$$

Ainsi par le théorème précédent, on obtient

$$\begin{aligned} \mathbb{P}(A) &= 1 - \mathbb{P}(\{\text{elle aime le premier livre}\} \cup \{\text{elle aime le deuxième livre}\}) \\ &= 1 - (\mathbb{P}(\text{elle aime le premier livre}) + \mathbb{P}(\text{elle aime le deuxième livre}) - \mathbb{P}(\text{elle aime les deux livres})) \\ &= 1 - \left(\frac{1}{2} + \frac{2}{5} - \frac{3}{10}\right) = \frac{2}{5}. \end{aligned}$$

Que se passe-t-il si on considère 3 événements E , F et G . On commence par appliquer la formule une première fois et on a

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

En réappliquant la formule, on obtient

$$\mathbb{P}(E \cup F \cup G) = \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) - \mathbb{P}(E \cap F) - \mathbb{P}((E \cap G) \cup (F \cap G)). \quad (2)$$

En regardant le dernier terme de plus près et en appliquant la formule une autre fois,

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

En insérant cette formule dans (2), on obtient finalement

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

On peut généraliser cette formule à une union de n événements qui est donnée dans le théorème suivant.

THÉORÈME 2.8

Soient $E_1, \dots, E_n$ n événements

$$\begin{aligned} \mathbb{P}(E_1 \cup \dots \cup E_n) &= \sum_{i=1}^n \mathbb{P}(E_i) - \sum_{i_1 < i_2} \mathbb{P}(E_{i_1} \cap E_{i_2}) + \sum_{i_1 < i_2 < i_3} \mathbb{P}(E_{i_1} \cap E_{i_2} \cap E_{i_3}) \\ &\quad + \dots + (-1)^{r+1} \sum_{i_1 < \dots < i_r} \mathbb{P}(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_r}) + \dots \\ &\quad + (-1)^{n+1} \mathbb{P}(E_1 \cap E_2 \cap \dots \cap E_n) \end{aligned}$$

que l'on peut réécrire

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{r=1}^n (-1)^{r+1} \sum_{i_1 < i_2 < \dots < i_r} \mathbb{P}\left(\bigcap_{j=1}^r E_{i_j}\right)$$

Pour certaines suites d'événements, on peut aussi faire un passage à la limite, ce sera important dans plusieurs résultats venant plus tard dans le cours.

DÉFINITION 2.9

Une suite d'événements $E_1, E_2, \dots$ est dite

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

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

DÉFINITION 2.10

Si $(E_n)_{n \geq 1}$ est une suite croissante, on définit

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

Si $(E_n)_{n \geq 1}$ est une suite décroissante, on définit

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

THÉORÈME 2.11

Si $(E_n)_{n \geq 1}$ est une suite croissante ou décroissante alors

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

Démonstration. On commence par supposer que la suite d'événements est croissante. L'idée est d'écrire $\bigcup_{n=1}^{\infty} E_n$ comme une union disjointe d'événements et d'utiliser le troisième axiome des probabilités. On définit la suite d'événements

$$F_1 = E_1 \quad F_n = E_n \cap E_{n-1}^c \text{ pour } n \geq 2.$$

En effet, par exemple

$$F_1 \cup F_2 = E_1 \cup (E_2 \cap E_1^c) = (E_1 \cup E_2) \cap (E_1 \cup E_1^c) = (E_1 \cup E_2) \cap \Omega = E_1 \cup E_2.$$

De même, on voit facilement que $\bigcup_{n=1}^{+\infty} F_n = \bigcup_{n=1}^{+\infty} E_n$ mais on remarque aussi que par construction $F_i \cap F_j = \emptyset$ pour $i \neq j$. Ainsi, on peut maintenant écrire

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

En utilisant dans la dernière étape que $\bigcup_{i=1}^n E_i = E_n$ car la suite d'événements est croissante et $E_1 \subset E_2 \subset \dots \subset E_n$. Si maintenant la suite est décroissante, alors la suite d'événements (E_i^c) est croissante et ainsi

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i^c\right) = \lim_{n \rightarrow \infty} \mathbb{P}(E_n^c)$$

mais par la règle de de Morgan, on obtient ainsi

$$\mathbb{P}\left(\left(\bigcap_{n=1}^{\infty} E_i\right)^c\right) = \lim_{n \rightarrow \infty} \mathbb{P}(E_n^c)$$

ce qui donne en utilisant la règle de calcul des probabilités de complémentaires

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} E_i\right) = \lim_{n \rightarrow \infty} \mathbb{P}(E_n).$$

□

2.3 Ensembles fondamentaux à probabilités équiprobables

Il arrive souvent (mais pas tout le temps) que sur un ensemble fondamental fini, chaque événement élémentaire ait la même probabilité. Ainsi si on écrit $\Omega = \{1, \dots, n\}$ alors cela veut dire que

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

Or, par le premier axiome des probabilités, nous savons que $\mathbb{P}(\Omega) = 1$ et comme

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

pour tout $j \in \{1, \dots, n\}$, on obtient que pour tout $j \in \{1, \dots, n\}$,

$$\mathbb{P}(\{j\}) = \frac{1}{n}.$$

Cela a pour conséquence directe le théorème suivant.

THÉORÈME 2.12

Soit Ω un ensemble fondamental fini. Si pour tout $x, y \in \Omega$, $\mathbb{P}(\{x\}) = \mathbb{P}(\{y\})$ alors pour tout événement $E \subset \Omega$,

$$\mathbb{P}(E) = \frac{\#E}{\#\Omega} := \frac{\text{nombre d'éléments de } E}{\text{nombre d'éléments de } \Omega}$$

EXEMPLE 9. On lance deux dés équilibrés, quelle est alors la probabilité que la somme des deux dés soit 7? On commence par définir l'ensemble fondamental donné par le résultat des deux dés

$$\Omega = \{1, \dots, 6\}^2 = \{(i, j), i \in \{1, \dots, 6\}, j \in \{1, \dots, 6\}\}.$$

On remarque que $\#\Omega = 6^2 = 36$ par le principe de dénombrement. Maintenant si considère l'événement donné par {la somme fait 7} alors

$$E = \{\text{la somme fait 7}\} = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

et ainsi la probabilité est donnée par

$$\mathbb{P}(E) = \frac{\#E}{\#\Omega} = \frac{6}{36} = \frac{1}{6}.$$

EXEMPLE 10. On considère un bol avec 5 boules noires et 6 boules blanches. On tire 3 boules, quelle est la probabilité de tirer une boule blanche et 2 boules noires? On commence toujours par définir notre ensemble fondamental, ici on va considérer l'ensemble des boules tirées (non-ordonnées)

$$\Omega = \{\text{ensemble des boules tirées non-ordonnées}\},$$

ainsi $\#\Omega$ est le nombre de combinaisons à 3 éléments parmi $6 + 5 = 11$ boules i.e $\#\Omega = \binom{11}{3}$. On considère l'événement $E = \{\text{tirer une boule blanche et deux boules noires}\}$ alors par le principe de dénombrement nous avons que

$$\#E = \binom{6}{1} \binom{5}{2}$$

puisque que l'on doit choisir une boule blanche parmi les 6 et deux boules noires parmi les 5. Ainsi la probabilité recherchée est de

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

EXEMPLE 11. On a n boules dont une est en or, les autres sont indistinguables. On tire k boules l'une après les autres de manière équiprobable sans remise. Quelle est la probabilité de tirer la boule spéciale? On

commence toujours par donner l'ensemble fondamental

$$\Omega = \{\text{tirage de } k \text{ boules ordonné}\}.$$

Alors on voit que $\#\Omega = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$ car nous avons n choix pour la première boule, puis $n-1$ pour la deuxième et ainsi de suite. Par le principe de dénombrement, on multiplie ces nombres de résultats. Définissons les événements

$$A_i = \{\text{La boule en or est tirée au } i\text{-ième tirage}\}.$$

On remarque que la probabilité recherchée peut alors s'écrire

$$\begin{aligned} \mathbb{P}(\text{la boule en or est tirée}) &= \mathbb{P}(\exists i \in \{1, \dots, k\}, \text{ la boule est tirée au } i\text{-ième tirage}) = \mathbb{P}\left(\bigcup_{i=1}^k A_i\right) \\ &= \sum_{i=1}^k \mathbb{P}(A_i) \end{aligned}$$

Nous devons donc calculer $\mathbb{P}(A_i) = \frac{\#A_i}{\#\Omega}$. Si on sait que la boule est tirée au i -ième tirage, alors pour le premier tirage, nous avons $n-1$ possibilités puisque l'on peut tirer n'importe quelle boule à part la boule en or. De même, pour le deuxième tirage, nous avons $n-2$ possibilités (toutes les boules à part la boule en or et la boule tirée au premier tirage). Ce raisonnement, fonctionne jusqu'au $(i-1)$ -ième tirage. Au i -ième tirage, il n'y a qu'une possibilité puisque nous devons tirer la boule en or. Pour le tirage suivant, nous avons maintenant $(n-i)$ possibilités car toutes les boules sont possibles à part les i boules qui ont déjà été tirées et ensuite $(n-i-1)$, $(n-i-2)$, $\dots$, $(n-k+1)$ pour le dernier tirage. Par le principe de dénombrement, nous avons

$$\#A_i = (n-1)(n-2)\dots(n-(i-1)) \times 1 \times (n-i)(n-i-1)\dots(n-k+1) = \frac{(n-1)!}{(n-k)!}.$$

Ainsi, on voit que

$$\mathbb{P}(A_i) = \frac{\#A_i}{\#\Omega} = \frac{\frac{(n-1)!}{(n-k)!}}{\frac{n!}{(n-k)!}} = \frac{1}{n}$$

et finalement

$$\mathbb{P}(\text{la boule en or est tirée}) = \sum_{i=1}^k \frac{1}{n} = \frac{k}{n}.$$

EXEMPLE 12. On répartit 52 cartes parmi 4 joueurs de manière équiprobable. Quelle est la probabilité que chaque joueur ait reçu un as? L'ensemble fondamental est donnée par la répartition des cartes et son cardinal est donnée par le coefficient multinomial $\binom{52}{13,13,13,13}$. On considère l'événement

$$E = \{\text{chaque joueur reçoit un as}\}$$

Pour calculer le nombre de possibilités donnant cela, on peut remarquer que l'on a deux expériences : la distribution des as qui nous donne 4! résultats et la distribution du reste des cartes qui nous donne le coefficient multinomial $\binom{48}{12,12,12,12}$ et ainsi la probabilité recherchée est de

$$\mathbb{P}(E) = \frac{4! \binom{48}{12,12,12,12}}{\binom{52}{13,13,13,13}} = \frac{2197}{20825} \approx 0.1055.$$

EXEMPLE 13. On considère n personnes, on suppose que leur anniversaire sont équiprobables (et on oublie le 29 Février). Quelle est la probabilité que leur anniversaire soit tous différents? Le nombre de possibilités total est alors de 365^n . La probabilité qu'ils soient tous différents est de $365 \times 364 \times \dots \times (365 - n + 1)$ et

ainsi

$$\mathbb{P}(n \text{ anniversaires soient tous différents}) = \frac{\frac{365!}{(365-n)!}}{365^n} = \frac{365!}{(365-n)!365^n}.$$

Il est intéressant de voir que si $n \geq 23$ alors on a $\mathbb{P}(E) = \frac{1}{2}$. Il suffit donc de 23 personnes pour avoir plus de la moitié de chance d'avoir deux personnes partageant un anniversaire !

EXEMPLE 14. On considère N invités portant une casquette. Chaque invité jette sa casquette au milieu de la pièce. On mélange alors les casquettes et les invités choisissent au hasard, chacun leur tour, une casquette. Quelle est la probabilité que personne n'ait choisi sa casquette ? L'ensemble fondamental est donnée par tous les vecteurs de N nombres différents correspondant au numéro de chaque casquette choisie. En particulier, cela correspond aux permutations de $\{1, \dots, N\}$ et on a $\#\Omega = N!$. On commence par réécrire l'événement recherché

$$\begin{aligned} \{\text{personne n'a choisi sa casquette}\} &= \{\text{au moins une personne a choisi sa casquette}\}^c \\ &= \left(\bigcup_{i=1}^N \{\text{la personne } i \text{ a choisi sa casquette}\} \right)^c \end{aligned}$$

On définit alors $E_i = \{\text{la personne } i \text{ a choisi sa casquette}\}$. Il faut faire attention car les E_i ne sont pas des événements disjoints, on peut avoir deux personnes qui choisissent leur casquette, on utilise alors la formule d'inclusion-exclusion

$$\mathbb{P}\left(\bigcup_{i=1}^N E_i\right) = \sum_{r=1}^N (-1)^{r+1} \sum_{i_1 < i_2 < \dots < i_r} \mathbb{P}\left(\bigcap_{j=1}^r E_{i_j}\right).$$

On peut calculer la probabilité des intersections, en effet, en comptant le nombre de possibilités de l'événement $E_{i_1} \cap \dots \cap E_{i_r}$, alors les personnes $i_1, \dots, i_r$ n'ont qu'une seule possibilité de casquette (la leur) et les autres ont $(N-r)(N-r-1) \dots 1 = (N-r)!$ possibilités car ils peuvent choisir n'importe quelle casquette à part les r venant de $i_1, \dots, i_r$. Ainsi, on obtient la formule

$$\mathbb{P}\left(\bigcup_{i=1}^N E_i\right) = \sum_{r=1}^N (-1)^{r+1} \sum_{i_1 < i_2 < \dots < i_r} \frac{(N-r)!}{N!} = \sum_{r=1}^N (-1)^{r+1} \frac{(N-r)!}{N!} \sum_{i_1 < i_2 < \dots < i_r} 1.$$

Il faut maintenant compter le nombre de termes dans cette somme, cela consiste à choisir les r indices $i_1, \dots, i_r$ parmi les N indices possibles, c'est-à-dire $\binom{N}{r}$ possibilités. Ainsi, la probabilité recherchée devient

$$\mathbb{P}\left(\bigcup_{i=1}^N E_i\right) = \sum_{r=1}^N (-1)^{r+1} \frac{(N-r)!}{N!} \binom{N}{r} = \sum_{r=1}^N (-1)^{r+1} \frac{(N-r)!}{N!} \times \frac{N!}{r!(N-r)!} = \sum_{r=1}^N \frac{(-1)^{r+1}}{r!}.$$

Finalement, cela nous donne

$$\mathbb{P}(\text{personne n'a choisi sa casquette}) = 1 - \mathbb{P}\left(\bigcup_{i=1}^N E_i\right) = 1 - \sum_{r=1}^N \frac{(-1)^{r+1}}{r!} = \sum_{r=0}^N \frac{(-1)^r}{r!}.$$

On remarque alors que si on fait tendre le nombre d'invités vers l'infini alors on obtient

$$\mathbb{P}(\text{personne n'a choisi sa casquette}) \xrightarrow{N \rightarrow \infty} \sum_{r=0}^{+\infty} \frac{(-1)^r}{r!} = \frac{1}{e}$$

3. PROBABILITÉS CONDITIONNELLES

On a vu que dans le cadre d'un ensemble fondamental à probabilités équiprobables, nous pouvions calculer $\mathbb{P}(E) = \frac{\#E}{\#\Omega}$. Il arrive cependant que notre connaissance de l'expérience nous donne un nouvel

ensemble fondamental. Par exemple, si Ω correspond aux lancers de deux pièces, nous avons

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

avec $\#\Omega = 4$. Ainsi, si on considère l'événement $E = \{\text{au moins une Face}\} = \{(P, F), (F, P), (F, F)\}$ alors nous avons

$$P(E) = \frac{\#E}{\#\Omega} = \frac{3}{4}.$$

Mais si je sais maintenant que la première pièce est Pile, que devient la probabilité de l'événement $\{\text{au moins une Face}\}$? Notre nouvel ensemble fondamental devient alors

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

et l'événement $E = \{\text{au moins une Face}\} = \{(P, F)\}$. Alors la probabilité recherchée est maintenant

$$\frac{\#\{(P, F)\}}{\#\{(P, F), (P, P)\}} = \frac{1}{2}.$$

On voit que la probabilité de l'événement E sachant que la première pièce est Pile n'est pas la même que la probabilité de E ! Pour généraliser cette observation à une définition, on peut écrire

$$\frac{\#\{(P, F)\}}{\#\{(P, F), (P, P)\}} = \frac{\#\{(P, F), (F, P), (F, F)\} \cap \{(P, F), (P, P)\}}{\#\{(P, F), (P, P)\}} = \frac{\#E \cap F}{\#F} = \frac{\#E \cap F}{\#\Omega} \times \frac{\#\Omega}{\#F} = \frac{P(E \cap F)}{P(F)}.$$

DÉFINITION 3.1

Soient $E, F \subset \Omega$ deux événements, la probabilité de E conditionnellement à F , ou sachant F , est si $P(F) \neq 0$

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

On peut l'utiliser aussi de l'autre façon

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

EXEMPLE 15. Jean est sûr à 80% d'avoir laissé ces clés dans sa veste. À 40%, elles sont dans sa poche gauche et à 40% dans sa poche droite. Il regarde dans la poche gauche et elles n'y sont pas, quelle est la probabilité de les trouver dans la poche de droite?

On définit les événements

$$G = \{\text{Elles sont à gauche}\} \quad \text{et} \quad D = \{\text{Elles sont à droites}\}.$$

On recherche alors la probabilité

$$P(D|G^c) = \frac{P(D \cap G^c)}{P(G^c)} = \frac{P(D)}{1 - P(G)} = \frac{2}{3}.$$

EXEMPLE 16. On considère une urne avec 8 boules rouges et 4 blanches. On tire sans remise deux boules de manière équiprobable. Quelle est la probabilité de tirer deux boules rouges?

On remarque que l'on sait calculer directement cette probabilité puisque

$$P(\text{Tirer deux boules rouges}) = \frac{\binom{8}{2}}{\binom{12}{2}}.$$

Retrouvons (et calculons) cet résultat d'une autre façon. Si on considère les événements

$$R_1 = \{\text{tirer une boule rouge au premier tirage}\} \quad \text{et} \quad R_2 = \{\text{tirer une boule rouge au deuxième tirage}\}$$

alors on a

$$\mathbb{P}(\text{Tirer deux boules rouges}) = \mathbb{P}(R_1 \cap R_2) = \mathbb{P}(R_1) \mathbb{P}(R_2|R_1) = \frac{8}{12} \times \frac{7}{11} = \frac{14}{33}.$$

THÉORÈME 3.2

Soient $E_1, \dots, E_n \subset \Omega$ des événements

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

que l'on peut réécrire

$$\mathbb{P}\left(\bigcap_{i=1}^n E_i\right) = \mathbb{P}(E_1) \prod_{i=1}^{n-1} \mathbb{P}\left(E_{i+1} \middle| \bigcap_{k=1}^i E_k\right)$$

3.1 Formule des probabilités totales et formule de Bayes

Soient $E, F \subset \Omega$ deux événements avec $\mathbb{P}(F) \neq 0$, alors on peut écrire

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

En utilisant la formule du conditionnement, on peut réécrire cette équation comme

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

C'est la formule des probabilités totales avec deux événements.

EXEMPLE 17. Pour une compagnie d'assurance, on peut répartir les clients en 2 groupes :

- Les clients enclin aux accidents pour qui $\mathbb{P}(\text{avoir un accident en 1 an}) = 0.4$.
- Ceux qui ne le sont pas pour qui $\mathbb{P}(\text{avoir un accident en 1 an}) = 0.2$.

On estime que 30% de la population sont enclins aux accidents, quelle est la probabilité qu'un nouveau client ait un accident dans sa première année ?

On définit les événements

$$E = \{\text{Le client aura un accident dans l'année}\} \quad \text{et} \quad F = \{\text{Le client est enclin aux accidents}\}.$$

Alors par la formule des probabilités totales, on peut écrire

$$\mathbb{P}(E) = \mathbb{P}(E|F) \mathbb{P}(F) + \mathbb{P}(E|F^c) \mathbb{P}(F^c) = 0.4 \times 0.3 + 0.2 \times 0.7 = 0.26.$$

Un nouveau client a un accident dans l'année, quelle est la probabilité qu'il soit enclin aux accidents ?

On cherche maintenant

$$\mathbb{P}(F|E) = \frac{\mathbb{P}(F \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(F) \mathbb{P}(E|F)}{\mathbb{P}(E)} = \frac{0.3 \times 0.4}{0.26} = \frac{6}{13}.$$

La formule des probabilités totales se généralise à plusieurs événements disjoints.

THÉORÈME 3.3: Loi des probabilités totales

Soient E un événement et $F_1, \dots, F_n \subset \Omega$ n événements disjoints i.e pour tout $i \neq j$, $F_i \cap F_j = \emptyset$ et tels que $\Omega = \bigcup_{i=1}^n F_i$ et $\mathbb{P}(F_i) \neq 0$ pour tout $i \in \{1, \dots, n\}$, alors

$$\mathbb{P}(E) = \sum_{i=1}^n \mathbb{P}(E|F_i) \mathbb{P}(F_i).$$

Démonstration. On peut écrire

$$E = E \cap \Omega = E \cap \left(\bigcup_{i=1}^n F_i \right) = \bigcup_{i=1}^n (E \cap F_i).$$

Mais on a pour $i \neq j$,

$$(E \cap F_i) \cap (E \cap F_j) = E \cap (F_i \cap F_j) = \emptyset.$$

Ainsi, nous avons une union disjointe et par le troisième axiome des probabilités, nous pouvons écrire

$$\mathbb{P}(E) = \sum_{i=1}^n \mathbb{P}(E \cap F_i) = \sum_{i=1}^n \mathbb{P}(E|F_i) \mathbb{P}(F_i).$$

□

Dans l'exemple de la compagnie d'assurance ci-dessus, nous avons écrit

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

Nous pouvons continuer en utilisant la formule des probabilités totales au dénominateur pour écrire

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

C'est la formule de Bayes avec deux événements.

EXEMPLE 18. On considère un questionnaire. L'étudiant connaît la réponse à la question avec probabilité p et met une réponse au hasard avec probabilité $1 - p$. Lorsqu'il répond au hasard, il a une probabilité $\frac{1}{m}$ de répondre correctement. Quelle est la probabilité qu'un étudiant connaisse la réponse sachant qu'il a répondu correctement ?

On commence par définir les événements

$$E = \{\text{L'étudiant connaît la réponse}\} \quad \text{et} \quad F = \{\text{L'étudiant a répondu correctement}\}.$$

On cherche alors

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(F|E)\mathbb{P}(E)}{\mathbb{P}(F|E)\mathbb{P}(E) + \mathbb{P}(F|E^c)\mathbb{P}(E^c)} = \frac{1 \times p}{1 \times p + \frac{1}{m} \times (1 - p)} = \frac{mp}{mp + 1 - p}.$$

On peut bien sûr généraliser cette formule à de multiples événements.

THÉORÈME 3.4: Formule de Bayes

Soient E un événement et $F_1, \dots, F_n \subset S$ n événements disjoints i.e pour tout $i \neq j$, $F_i \cap F_j = \emptyset$ et tels que $S = \bigcup_{i=1}^n F_i$ alors pour tout $j \in \{1, \dots, n\}$,

$$\mathbb{P}(F_j|E) = \frac{\mathbb{P}(E|F_j) \mathbb{P}(F_j)}{\sum_{i=1}^n \mathbb{P}(E|F_i) \mathbb{P}(F_i)}$$

EXEMPLE 19. On considère 3 cartes : une carte a deux faces bleues, une carte a une face verte et une face bleue et une carte a deux faces vertes. Après un mélange, on tire une carte au hasard et on montre une face. Sachant que la face montrée est verte, quelle est la probabilité que l'autre soit bleue ?

On définit les événements

$$VV = \{\text{la carte a deux faces vertes}\}, BB = \{\text{la carte a deux faces bleues}\} \text{ et } VB = \{\text{la carte est bicolore}\}.$$

On considère aussi l'événement $V = \{\text{la face montrée est verte}\}$. Alors, on cherche

$$\mathbb{P}(VB|V) = \frac{\mathbb{P}(V|VB)\mathbb{P}(VB)}{\mathbb{P}(V|VB)\mathbb{P}(VB) + \mathbb{P}(V|VV)\mathbb{P}(VV) + \mathbb{P}(V|BB)\mathbb{P}(BB)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{1}{3}.$$

3.2 Indépendance de deux événements

Dans beaucoup d'exemples, on remarque que $\mathbb{P}(E) \neq \mathbb{P}(E|F)$, en d'autres termes l'information donnée par la connaissance de F a changé la probabilité de E . Si ce n'est pas le cas, c'est-à-dire que $\mathbb{P}(E) = \mathbb{P}(E|F)$, on dit que E et F sont *indépendants*. Le fait de savoir que F s'est réalisé ne change pas la probabilité de E . Ainsi, cela peut s'écrire

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

DÉFINITION 3.5

Deux événements $E, F \subset S$ sont dits indépendants si

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

EXEMPLE 20. On tire deux pièces de monnaie équilibrées et on considère les événements suivants

$$E = \{\text{la première pièce fait P}\} \quad \text{et} \quad F = \{\text{la deuxième pièce fait F}\}.$$

Alors on considère l'ensemble fondamental

$$\Omega = \{(P, P), (P, F), (F, P), (F, F)\} \quad \text{avec} \quad \#\Omega = 4.$$

et on obtient les probabilités

$$\begin{aligned} \mathbb{P}(E) &= \mathbb{P}(\{(P, F), (P, P)\}) = \frac{2}{4} = \frac{1}{2}, & \mathbb{P}(F) &= \mathbb{P}(\{(P, F), (F, F)\}) = \frac{2}{4} = \frac{1}{2} \\ \mathbb{P}(E \cap F) &= \mathbb{P}(\{(P, F)\}) = \frac{1}{4}. \end{aligned}$$

On voit que l'on a bien $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$ et les événements E et F sont donc bien indépendants.

EXEMPLE 21. On lance maintenant deux dés et considère les événements

$$E_1 = \{\text{la somme des deux dés fait 6}\} \quad \text{et} \quad E_2 = \{\text{le premier dé fait 4}\}.$$

On considère l'ensemble fondamental

$$\Omega = \{1, \dots, 6\}^2 \quad \text{avec} \quad \#\Omega = 36.$$

Alors on a

$$\begin{aligned} \mathbb{P}(E_1) &= \mathbb{P}(\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}) = \frac{5}{36} \\ \mathbb{P}(E_2) &= \mathbb{P}(\{(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6)\}) = \frac{1}{6}. \end{aligned}$$

Et on voit que l'on a $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(\{(4, 2)\}) = \frac{1}{36}$ alors nous avons

$$\mathbb{P}(E_1 \cap E_2) = \frac{1}{36} \neq \frac{5}{36} = \mathbb{P}(E_1)\mathbb{P}(E_2)$$

et ainsi les événements E_1 et E_2 ne sont pas indépendants. Maintenant si on considère l'événement

$$E_3 = \{\text{la somme des deux dés fait 7}\}$$

alors

$$\mathbb{P}(E_3) = \mathbb{P}(\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}) = \frac{6}{36} = \frac{1}{6}.$$

Et nous avons

$$\mathbb{P}(E_3 \cap E_2) = \mathbb{P}(\{(4,3)\}) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = \mathbb{P}(E_2)\mathbb{P}(E_3)$$

et les événements E_3 et E_2 sont donc indépendants!

THÉORÈME 3.6

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

Démonstration. Il suffit de voir que

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

où l'on a utilisé dans la deuxième égalité que l'union est disjointe et le troisième des axiomes de probabilités et dans la troisième égalité que les deux événements sont indépendants. On peut maintenant réécrire cette égalité comme

$$\mathbb{P}(E \cap F^c) = \mathbb{P}(E)(1 - \mathbb{P}(F)) = \mathbb{P}(E)\mathbb{P}(F^c)$$

i.e les événements E et F^c sont indépendants. □

3.3 Indépendance de trois événements ou plus

Soient E , F et G trois événements, si E et F , E et G , et E et G sont indépendants, est-il vrai que E et $F \cap G$ sont indépendants? Considérons l'exemple suivant, on jette deux dés avec $\Omega = \{1, \dots, 6\}^2$ et $\#\Omega = 36$ et on considère

$$E = \{\text{la somme des deux dés fait 7}\}, \quad F = \{\text{le premier dé est un 4}\} \quad \text{et} \quad G = \{\text{le second est un 3}\}.$$

On a vu que E et F sont indépendants et le même raisonnement donne que E et G sont indépendants. Comme les deux dés sont lancés indépendamment F et G sont aussi indépendants. Maintenant, on voit que

$$\mathbb{P}(E|F \cap G) = 1 \neq \mathbb{P}(E) = \frac{1}{6}.$$

On remarque donc qu'avoir indépendance des événements deux à deux n'est pas suffisant pour avoir indépendance de trois événements. Cette forme d'indépendance est donnée par la définition suivante

DÉFINITION 3.7

Trois événements E , F et G sont dits (totalement) indépendants si

$$\begin{aligned} \mathbb{P}(E \cap F) &= \mathbb{P}(E)\mathbb{P}(F), \quad \mathbb{P}(E \cap G) = \mathbb{P}(E)\mathbb{P}(G), \quad \mathbb{P}(F \cap G) = \mathbb{P}(F)\mathbb{P}(G) \\ \text{et} \quad \mathbb{P}(E \cap F \cap G) &= \mathbb{P}(E)\mathbb{P}(F)\mathbb{P}(G). \end{aligned}$$

On remarque donc que les événements doivent être deux à deux indépendants avec la dernière condition en plus. Dans l'exemple ci-dessus, on avait

$$\mathbb{P}(E \cap F \cap G) = \mathbb{P}(F \cap G) = \frac{1}{36} \quad \text{et} \quad \mathbb{P}(E)\mathbb{P}(F)\mathbb{P}(G) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}.$$

On remarque que si E , F , et G sont totalement indépendants alors E est indépendant de tout ensemble formé

à partir de F et G . Par exemple, on peut calculer

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

On peut généraliser cette définition à plus de trois événements. On doit considérer tout sous-ensemble fini d'entre eux.

DÉFINITION 3.8

Si $E_1, \dots, E_n$ sont n événements, ils sont dits indépendants ou totalement indépendants si pour tout $k \in \{1, \dots, n\}$ et pour tout $i_1, i_2, \dots, i_k \in \{1, \dots, n\}$, on a

$$\mathbb{P}\left(\bigcap_{\ell=1}^k E_{i_\ell}\right) = \prod_{\ell=1}^k \mathbb{P}(E_{i_\ell})$$

Ceci fonctionne avec $n = \infty$ en choisissant tout sous-ensemble fini d'entre eux est indépendant.

3.4 Suite d'épreuves indépendantes

Souvent, on réalise des épreuves indépendantes i.e une suite d'expériences telles que leur résultat n'influe pas sur les autres. Par exemple un lancers de n pièces ou de dés.

EXEMPLE 22. On réalise une suite d'épreuves indépendantes. Chaque épreuve donne un succès avec probabilité p et un échec avec probabilité $1 - p$. Calculons les probabilités des événements suivants

- Il y a au moins un succès parmi les n premières épreuves.
- Il y a exactement k succès parmi les n premières épreuves.
- Toutes les épreuves sont des succès.

Pour le premier événement, on peut passer au complémentaire. En effet, le complémentaire devient alors que tous les événements sont des échecs. Alors si on écrit

$$E_i = \{\text{l'épreuve } i \text{ a été un échec}\}$$

alors on recherche

$$\begin{aligned}\mathbb{P}(\text{au moins un succès parmi les } n \text{ épreuves}) &= 1 - \mathbb{P}(\text{toutes les épreuves sont des échecs}) \\ &= 1 - \mathbb{P}\left(\bigcap_{i=1}^n E_i\right) \\ &= 1 - \prod_{i=1}^n \mathbb{P}(E_i) = 1 - (1 - p)^n.\end{aligned}$$

Pour le deuxième événement, si on sait où sont les k succès et les $n - k$ échecs, par exemples les k premiers sont les succès, alors la probabilité de cet événement est donnée par

$$\mathbb{P}\left(\left(\bigcap_{i=1}^k E_i^c\right) \cap \left(\bigcap_{i=k+1}^n E_i\right)\right) = \prod_{i=1}^k \mathbb{P}(E_i^c) \times \prod_{i=k+1}^n \mathbb{P}(E_i) = p^k (1 - p)^{n-k}.$$

Mais ceci n'est qu'un exemple, il reste à choisir où sont les k succès et les $n - k$ échecs et on sait qu'il y a $\binom{n}{k}$ choix! Ainsi on obtient

$$\mathbb{P}(\text{il y a exactement } k \text{ succès}) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Pour la dernière probabilité, on remarque que la probabilité de n'avoir que des succès parmi les n premières

épreuves est de

$$\mathbb{P} \left(\bigcap_{i=1}^n E_i^c \right) = \prod_{i=1}^n \mathbb{P}(E_i^c) = p^n.$$

Mais on remarque que $F_n = \bigcap_{i=1}^n E_i^c$ est une suite d'événements décroissante et ainsi par la continuité des probabilités, on peut écrire

$$\mathbb{P} \left(\bigcap_{i=1}^{+\infty} E_i^c \right) = \mathbb{P} \left(\lim_{n \rightarrow \infty} \bigcap_{i=1}^n E_i^c \right) = \lim_{n \rightarrow \infty} \mathbb{P} \left(\bigcap_{i=1}^n E_i^c \right) = \lim_{n \rightarrow \infty} p^n = \begin{cases} 0 & \text{si } p < 1, \\ 1 & \text{sinon.} \end{cases}$$

EXEMPLE 23. On considère une suite d'épreuves indépendantes qui consiste à jeter deux dés et à garder leur somme. Quelle est la probabilité de voir un 5 avant un 7? On considère les événements suivants

$$E_n = \{\text{il n'y a eu ni 5, ni 7 pendant les } n-1 \text{ premières épreuves puis un 5 à la } n\text{-ième}\}$$

alors on remarque que

$$\mathbb{P}(\text{voir un 5 avant un 7}) = \mathbb{P} \left(\bigcup_{n=1}^{+\infty} E_n \right) = \sum_{n=1}^{+\infty} \mathbb{P}(E_n)$$

car les événements E_n sont deux à deux disjoints. Maintenant on peut calculer

$$\begin{aligned} \mathbb{P}(\text{obtenir 5 lors d'une épreuve}) &= \mathbb{P}(\{(1,4), (2,3), (3,2), (4,1)\}) = \frac{4}{36} = \frac{1}{9}, \\ \mathbb{P}(\text{obtenir 7 lors d'une épreuve}) &= \frac{1}{6}. \end{aligned}$$

et ainsi on peut calculer

$$\mathbb{P}(E_n) = \frac{1}{9} \mathbb{P} \left(1 - \frac{1}{6} - \frac{1}{9} \right)^{n-1} = \frac{1}{9} \left(\frac{13}{18} \right)^{n-1}.$$

Finalement, on peut calculer

$$\mathbb{P}(\text{voir un 5 avant un 7}) = \sum_{n=1}^{+\infty} \frac{1}{9} \left(\frac{13}{18} \right)^{n-1} = \frac{1}{9} \sum_{n=0}^{+\infty} \left(\frac{13}{18} \right)^n = \frac{1}{9} \times \frac{1}{1 - \frac{13}{18}} = \frac{2}{5}.$$

On peut aussi utiliser un autre raisonnement, appelé le "conditionnement à un pas". Considérons les 4 événements suivants

$$\begin{aligned} E &= \{5 \text{ apparaît avant } 7\}, \quad F = \{\text{première épreuve est } 5\} \\ G &= \{\text{première épreuve est } 7\}, \quad H = \{\text{première épreuve est ni } 5 \text{ ni } 7\}. \end{aligned}$$

Alors on voit que F , G et H sont des événements disjoints et $F \cup G \cup H = \Omega$ ainsi par la formule des probabilités totales, on a

$$\begin{aligned} \mathbb{P}(E) &= \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|G)\mathbb{P}(G) + \mathbb{P}(E|H)\mathbb{P}(H) \\ &= 1 \times \frac{1}{9} + 0 \times \frac{1}{6} + \mathbb{P}(E) \times \frac{13}{18}. \end{aligned}$$

Pour calculer cela, on remarque que si la première épreuve est 5 alors 5 apparaît avant 7 et que si cette première épreuve donne 7 alors 5 n'apparaît pas avant 7. Enfin, si H se réalise, alors le résultat de la première épreuve n'est pas important pour savoir si 5 apparaît avant 7 et ainsi $\mathbb{P}(E \cap H) = \mathbb{P}(E)$. On peut réécrire cette équation comme

$$\frac{5}{18} \mathbb{P}(E) = \frac{1}{9} \Leftrightarrow \mathbb{P}(E) = \frac{2}{5}.$$

3.5 Fonction de probabilité conditionnelle

On peut définir la fonction $\mathbb{P}(\cdot|F)$ sur les événements de Ω . Cela définit une nouvelle fonction de probabilité.

THÉORÈME 3.9

Soit $F \subset S$ un événement. On peut définir la fonction $\mathbb{P}(\cdot|F)$ qui suit les propriétés d'une fonction de probabilité.

- Pour tout événement E , $0 \leq \mathbb{P}(E|F) \leq 1$.
- $\mathbb{P}(S|F) = 1$.
- Si $E_1, E_2, \dots$ est une suite d'événements deux à deux disjoints i.e. pour tout $i \neq j$, $E_i \cap E_j = \emptyset$

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} E_n \middle| F\right) = \sum_{i=1}^{+\infty} \mathbb{P}(E_i|F).$$

Démonstration. On remarque tout d'abord que

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} \in [0, 1] \quad \text{car} \quad E \cap F \subset F \Rightarrow \mathbb{P}(E \cap F) \leq \mathbb{P}(F).$$

De plus, nous avons

$$\mathbb{P}(\Omega|F) = \frac{\mathbb{P}(\Omega \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F)}{\mathbb{P}(F)} = 1.$$

Finalement, si $(E_i)_i$ est une suite d'événements deux à deux disjoints alors,

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty} E_i \middle| F\right) = \frac{\mathbb{P}\left(\left(\bigcup_{i=1}^{+\infty} E_i\right) \cap F\right)}{\mathbb{P}(F)} = \frac{\mathbb{P}\left(\bigcup_{i=1}^{+\infty} (E_i \cap F)\right)}{\mathbb{P}(F)} = \frac{\sum_{i=1}^{+\infty} \mathbb{P}(E_i \cap F)}{\mathbb{P}(F)} = \sum_{i=1}^{+\infty} \mathbb{P}(E_i|F).$$

□

Ainsi, on remarque que si on définit $Q(E) = \mathbb{P}(E|F)$ alors toutes les formules que l'on a vu pour l'instant sont aussi valides pour Q . Par exemple, la formule des probabilités totales donne

$$Q(E_1) = Q(E_1|E_2)Q(E_2) + Q(E_1|E_2^c)Q(E_2^c)$$

et on peut calculer

$$Q(E_1|E_2) = \frac{Q(E_1 \cap E_2)}{Q(E_2)} = \frac{\mathbb{P}(E_1 \cap E_2|F)}{\mathbb{P}(E_2|F)} = \frac{\frac{\mathbb{P}(E_1 \cap E_2 \cap F)}{\mathbb{P}(F)}}{\frac{\mathbb{P}(E_2 \cap F)}{\mathbb{P}(F)}} = \frac{\mathbb{P}(E_1 \cap (E_2 \cap F))}{\mathbb{P}(E_2 \cap F)} = \mathbb{P}(E_1|E_2 \cap F).$$

Ainsi, on peut réécrire cette formule comme

$$\mathbb{P}(E_1|F) = \mathbb{P}(E_1|E_2 \cap F)\mathbb{P}(E_2|F) + \mathbb{P}(E_1|E_2^c \cap F)\mathbb{P}(E_2^c|F).$$

On peut aussi définir une forme d'indépendance conditionnelle.

DÉFINITION 3.10

Deux événements sont conditionnellement indépendants selon $F \subset S$ si

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

4. VARIABLES ALÉATOIRES

Lors d'expériences, on s'intéresse à une fonction du résultat nous donnant un nombre. Par exemple, si on lance deux dés et considère leur somme ou la valeur du premier dé, si on lance n pièces et on compte le nombre de faces. On remarque alors que ces valeurs sont des fonctions de l'ensemble fondamental : pour chaque réalisation, on obtient une valeur. Ces fonctions sont appelées *variable aléatoire*.

DÉFINITION 4.1

Une variable aléatoire est une fonction $X : \Omega \rightarrow \mathbb{R}$.

Si on considère le lancer de deux dés alors $\Omega = \{1, \dots, 6\}^2$ et si considère X la somme des deux dés alors on a, par exemple,

$$X(\{(5, 2)\}) = 7 \quad \text{et} \quad X(\{(4, 1)\}) = 5.$$

On remarque que l'on a pris un élément de l'ensemble fondamental pour obtenir une valeur numérique.

EXEMPLE 24. On lance 3 pièces et on définit Y le nombre de Faces obtenues. On considère l'ensemble fondamental

$$\Omega = \{P, F\}^3 \quad \text{avec} \quad \#\Omega = 2^3 = 8$$

par le principe de dénombrement. On peut alors calculer la probabilité des différentes valeurs de Y . Tout d'abord, on remarque que $Y \in \{0, 1, 2, 3\}$ et on a alors

$$\begin{aligned} \mathbb{P}(Y = 0) &= \mathbb{P}(\{(P, P, P)\}) = \frac{1}{8}, \\ \mathbb{P}(Y = 1) &= \mathbb{P}(\{(P, P, F), (P, F, P), (F, P, P)\}) = \frac{3}{8}, \\ \mathbb{P}(Y = 2) &= \mathbb{P}(\{(P, F, F), (F, P, F), (F, F, P)\}) = \frac{3}{8}, \\ \mathbb{P}(Y = 3) &= \mathbb{P}(\{(F, F, F)\}) = \frac{1}{8}. \end{aligned}$$

On remarque que Y ne peut bien prendre que des valeurs entre 0 et 3 et on a bien

$$\mathbb{P}\left(\bigcup_{i=0}^3 \{Y = i\}\right) = \sum_{i=0}^3 \mathbb{P}(Y = i) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1.$$

EXEMPLE 25. On répète le jet d'une pièce jusqu'à obtenir Face ou après n jets. On suppose que les jets sont indépendants et que la probabilité d'obtenir Face est p . Tout d'abord, on peut définir l'ensemble fondamental comme étant

$$\Omega = \{P^k F, k \in \{0, 1, \dots, n-1\} \cup \{P^n\}\} \quad \text{avec} \quad P^n F = \underbrace{(P, \dots, P)}_{n \text{ fois}}, F).$$

En effet, soit on fait une suite de Piles puis un Face qui arrête l'expérience ou alors on a fait n Piles et on arrête l'expérience. On considère la variable aléatoire X correspondant au nombre de jets réalisés. On remarque que les valeurs possibles de X sont $\{1, \dots, n\}$. Calculons les premières probabilités de X . On a

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

On peut généraliser ce raisonnement pour $\mathbb{P}(X = k)$ jusqu'à $k = n-1$. En effet, on doit avoir une suite de $k-1$ Piles puis 1 Face à la fin et ainsi

$$\mathbb{P}(X = k) = \mathbb{P}(P^{k-1}F) = (1-p)^{k-1}p \quad \text{pour} \quad k \in \{1, \dots, n-1\}.$$

Pour $\mathbb{P}(X = n)$ ceci est légèrement différent, il y a deux façons d'obtenir $X = n$, soit on a fait $n-1$ Piles puis

une Face ou alors on a fait n Piles et ainsi

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

Vérifions que les probabilités calculées somment bien à 1,

$$\mathbb{P}\left(\bigcup_{k=1}^n \{X = k\}\right) = \sum_{k=1}^n \mathbb{P}(X = k) = \sum_{k=1}^{n-1} (1-p)^{k-1}p + (1-p)^{n-1} = p \frac{1 - (1-p)^{n-1}}{1 - (1-p)} + (1-p)^{n-1} = 1.$$

EXEMPLE 26. On collectionne des coupons, on suppose qu'il y a N types et on choisit un coupon de manière équiprobable et indépendant à chaque fois. Soit T le temps i.e le nombre de choix pour avoir tous les types de coupons. On va calculer $\mathbb{P}(T > n)$, on observe que

$$\begin{aligned} \mathbb{P}(T > n) &= \mathbb{P}(\text{il existe un type de coupon que l'on n'a pas reçu avant le } n\text{-ième tirage}) \\ &= \mathbb{P}\left(\bigcup_{j=1}^N \underbrace{\{\text{le type } j \text{ n'a pas été reçu avant le } n\text{-ième tirage}\}}_{=: A_j}\right) \\ &= \mathbb{P}\left(\bigcup_{j=1}^N A_j\right). \end{aligned}$$

On obtient la probabilité d'une union d'événements mais il faut faire attention car cette union n'est pas disjointe. En effet, on pourrait tout à fait avoir que deux types n'ont pas encore été reçus avant le n -ième tirage. On va donc utiliser la formule d'inclusion-exclusion

$$\mathbb{P}\left(\bigcup_{j=1}^N A_j\right) = \sum_{r=1}^N (-1)^{r+1} \sum_{i_1 < i_2 < \dots < i_r} \mathbb{P}\left(\bigcap_{j=1}^r A_{i_j}\right).$$

Calculons maintenant la probabilité de ces intersections, pour $r \geq 1$, on a

$$\mathbb{P}\left(\bigcap_{j=1}^r A_{i_j}\right) = \mathbb{P}(\text{les coupons ne sont pas de types } i_1, \dots, i_r \text{ dans les } n \text{ premiers lancers}) = \left(\frac{N-r}{N}\right)^n.$$

En effet, pour chaque choix de coupons, on tire de manière indépendante parmi les $N - r$ types qui ne sont pas $i_1, \dots, i_r$. On remarque que cette probabilité ne dépend pas du choix du type $i_1, \dots, i_r$ mais seulement du nombre de types r . Ainsi on obtient le résultat

$$\mathbb{P}(T > n) = \sum_{r=1}^N (-1)^{r+1} \sum_{i_1 < \dots < i_r} \left(\frac{N-r}{N}\right)^n = \sum_{r=1}^N (-1)^{r+1} \binom{N}{r} \left(\frac{N-r}{N}\right)^n$$

où on a utilisé que le nombre de termes dans la somme sur $i_1 < \dots < i_r$ correspond au choix de ces indices parmi N et ainsi donne un coefficient binomial $\binom{N}{r}$. On peut finir en écrivant

$$\begin{aligned} \mathbb{P}(T = n) &= \mathbb{P}(\{T > n-1\} \cap \{T \leq n\}) = 1 - \mathbb{P}(\{T \leq n-1\} \cup \{T > n\}) \\ &= 1 - \mathbb{P}(T \leq n-1) - \mathbb{P}(T > n) \\ &= \mathbb{P}(T > n-1) - \mathbb{P}(T > n). \end{aligned}$$

Cette formule est importante à connaître car utile s'il est plus facile de calculer la probabilité d'inégalité plutôt que d'égalité. Finalement, on peut calculer cette différence et on obtient

$$\mathbb{P}(T = n) = \sum_{r=1}^N (-1)^{r+1} \binom{N}{r} \frac{r}{N} \left(\frac{N-r}{N}\right)^{n-1}.$$

Sur l'exemple précédent, on remarque que la probabilité d'inégalité nous donne toute l'information sur la variable aléatoire. Ceci est en fait général et on peut définir la *fonction de répartition* d'une variable aléatoire par la définition suivante.

DÉFINITION 4.2

Soit X une variable aléatoire, on définit sa fonction de répartition par une fonction $F : \mathbb{R} \rightarrow [0, 1]$,

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

En particulier, F est toujours croissante et on a

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

4.1 Variables aléatoires discrètes

DÉFINITION 4.3

Une variable aléatoire X est dite discrète si elle ne prend qu'un nombre dénombrable (et possiblement fini) de valeurs. On définit sa loi de probabilité par

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

Ainsi, si X est une variable aléatoire discrète prenant les valeurs $\{x_1, x_2, \dots\}$ alors on a

$$p(x_i) \geq 0 \quad \forall i \in \mathbb{N} \quad \text{et} \quad p(a) = 0 \quad \text{sinon.}$$

En particulier, on obtient donc que $\sum_{i=1}^{+\infty} p(x_i) = 1$.

EXEMPLE 27. Si X est la somme de deux dés lancés indépendamment alors on a

$$\begin{aligned} \mathbb{P}(X = 2) &= \frac{1}{36} = \mathbb{P}(X = 12), & \mathbb{P}(X = 3) &= \frac{1}{18} = \mathbb{P}(X = 11), \\ \mathbb{P}(X = 4) &= \frac{1}{12} = \mathbb{P}(X = 10), & \mathbb{P}(X = 5) &= \frac{1}{9} = \mathbb{P}(X = 9), \\ \mathbb{P}(X = 6) &= \frac{5}{36} = \mathbb{P}(X = 8), & \mathbb{P}(X = 7) &= \frac{1}{6}. \end{aligned}$$

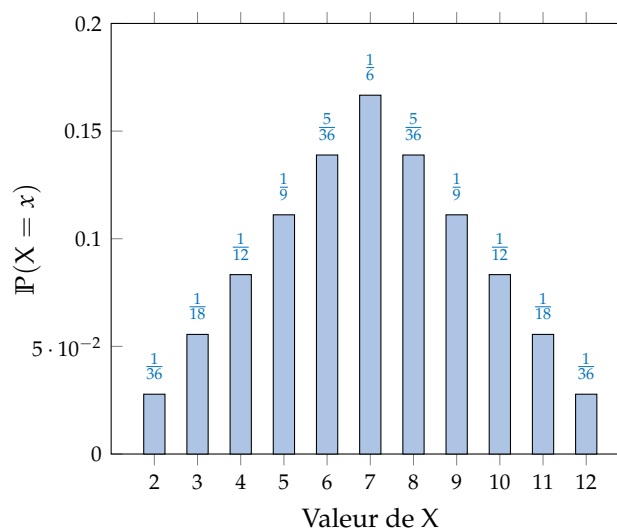


FIGURE 1 – On peut représenter la loi de probabilité sous cette forme.

Dans le cadre des variables aléatoires discrètes, la fonction de répartition devient pour X une variable aléatoire à valeurs dans $\{x_1, x_2, \dots\}$,

$$F(x) = \sum_{x \leq a} p(x) = \sum_{j, x_j \leq a} p(x_j).$$

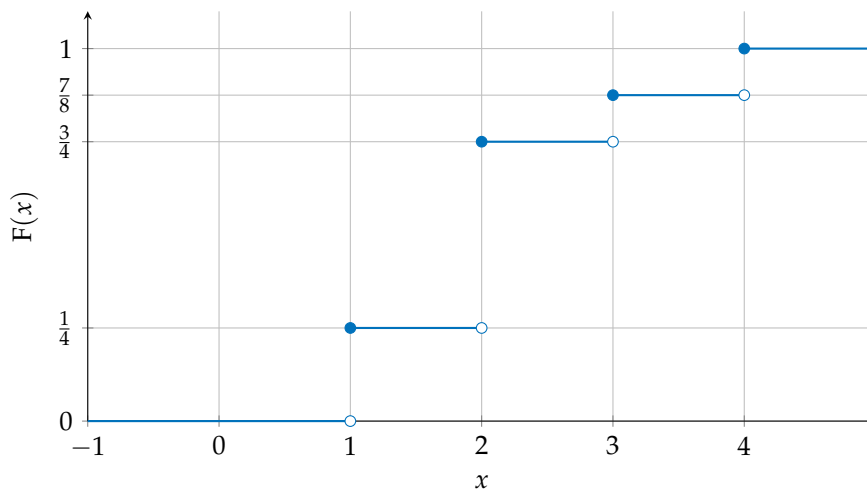
On remarque que si $x, y \in [x_{j-1}, x_j)$ alors on a $F(x) = F(y)$ mais, en revanche, $F(x_j) - F(x_{j-1}) = p(x_j)$. On obtient donc une fonction *en escalier* : elle est constante entre deux valeurs de X mais fait un saut de taille $p(x_j)$ à chaque x_j .

EXEMPLE 28. On considère la loi de probabilité suivante,

$$p(1) = \frac{1}{4}, \quad p(2) = \frac{1}{2}, \quad p(3) = \frac{1}{8}, \quad p(4) = \frac{1}{8}.$$

on peut alors donner le graphe de la fonction de répartition en utilisant le principe donné ci-dessus : entre chaque valeur, la fonction est constante et à chaque valeur de X , la fonction fait un saut de taille $p(x_j)$. On peut aussi écrire cette fonction sous cette forme

$$F(x) = \begin{cases} 0 & \text{si } x < 1, \\ \frac{1}{4} & \text{si } 1 \leq x < 2, \\ \frac{3}{4} & \text{si } 2 \leq x < 3, \\ \frac{7}{8} & \text{si } 3 \leq x < 4, \\ 1 & \text{si } x \geq 4. \end{cases}$$



On peut faire aussi le travail inverse, si on nous donne la fonction de répartition : les valeurs pour lesquelles la fonction a un saut sont les valeurs possibles de la variable aléatoire et la taille des sauts nous donne la loi de probabilité. On donne un récapitulatif par la proposition suivante.

PROPOSITION 4.4. Soit X une variable aléatoire discrète à valeurs dans $\{x_1, x_2, \dots\}$ alors sa fonction de répartition est donnée par

$$F(x) = \sum_{j, x_j \leq x} \mathbb{P}(X = x_j).$$

De plus, si $x, y \in [x_{j-1}, x_j)$ alors $F(x) = F(y)$ et on a $F(x_j) - F(x_{j-1}) = \mathbb{P}(X = x_j)$. C'est donc une fonction en escalier.

On définit maintenant l'espérance d'une variable aléatoire discrète qui correspond à la valeur moyenne de la variable aléatoire.

DÉFINITION 4.5

Soit X une variable aléatoire discrète, on définit son espérance par

$$\mathbb{E}[X] = \sum_{x, p(x)>0} xP(X=x).$$

Si X est à valeurs dans $\{x_1, x_2, \dots\}$, alors

$$\mathbb{E}[X] = \sum_{j=1}^{+\infty} x_j P(X=x_j).$$

C'est la moyenne pondérée des valeurs de X où les poids sont donnés par les probabilités.

EXEMPLE 29. Si X est le résultat d'un dé équilibré alors son espérance est donnée par

$$\mathbb{E}[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{21}{6} = \frac{7}{2}.$$

On voit bien ici que c'est le résultat *moyen* d'un dé mais ce n'est pas nécessairement un résultat possible de la variable aléatoire. Ici, un dé ne donnera jamais $\frac{7}{2}$.

EXEMPLE 30. On définit la fonction indicatrice pour un événement $A \subset \Omega$,

$$\mathbb{1}_A = \begin{cases} 1 & \text{si } A \text{ se réalise,} \\ 0 & \text{sinon.} \end{cases}$$

Alors on a, $P(\mathbb{1}_A = 1) = P(A)$ et $P(\mathbb{1}_A = 0) = 1 - P(A)$ et ainsi

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

EXEMPLE 31. On considère 120 écoliers divisés en 3 bus :

- un bus avec 36 élèves,
- un bus avec 40 élèves,
- un bus avec 44 élèves.

On choisit un élève au hasard, calculons l'espérance du nombre d'élèves dans le bus qu'il a pris que l'on note X . On calcule tout d'abord la loi de probabilité

$$P(X=36) = \frac{36}{120} = \frac{3}{10}, \quad P(X=40) = \frac{40}{120} = \frac{1}{3} \quad \text{et} \quad P(X=44) = \frac{44}{120} = \frac{11}{30}.$$

Ainsi, on peut calculer l'espérance,

$$\mathbb{E}[X] = 36 \times \frac{3}{10} + 40 \times \frac{1}{3} + 44 \times \frac{11}{30} = \frac{1208}{30} = \frac{604}{15} \approx 40,266.$$

On remarque que l'espérance est plus grande que 40 qui correspond à la moyenne du nombre d'élèves des trois bus! Ceci vient du fait qu'on a plus de chance de choisir un élève dans le bus de 44 élèves que celui avec 36 élèves.

Souvent, on ne veut pas calculer simplement l'espérance d'une variable aléatoire mais l'espérance d'une fonction d'une variable aléatoire $\mathbb{E}[f(X)]$.

EXEMPLE 32. Parfois on peut le faire directement. Par exemple si X est donnée par

$$P(X=-1) = 0.2 \quad P(X=0) = 0.5 \quad \text{et} \quad P(X=1) = 0.3.$$

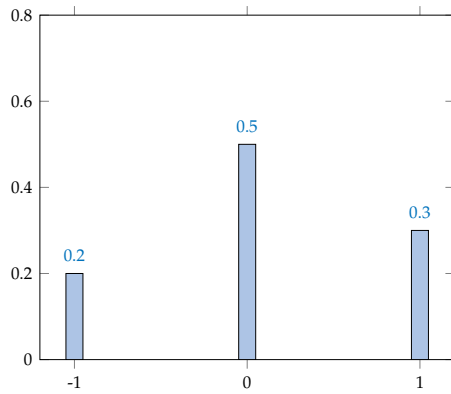


FIGURE 2 – Loi de probabilité

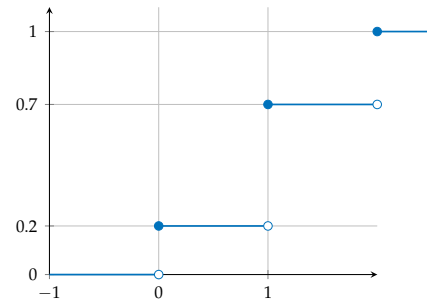


FIGURE 3 – Fonction de répartition

Calculons $\mathbb{E}[X^2]$, on peut calculer la loi de probabilité de la variable aléatoire $Y = X^2$ à valeurs dans $\{0, 1\}$. On obtient

$$\mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = \frac{1}{2}, \quad \mathbb{P}(Y = 1) = \mathbb{P}(X = 1) + \mathbb{P}(X = -1) = \frac{1}{2}$$

et ainsi

$$\mathbb{E}[X^2] = \mathbb{E}[Y] = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}.$$

⚠ On n'a pas $\mathbb{E}[X^2] = \mathbb{E}[X]^2$.

On voit qu'on a en fait

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

et cela se généralise !

THÉORÈME 4.6

Si X une variable aléatoire discrète à valeurs dans $\{x_1, x_2, \dots\}$ alors pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E}[g(X)] = \sum_{j=1}^{+\infty} g(x_j) \mathbb{P}(X = x_j).$$

Démonstration. On remarque que pour plusieurs valeurs x_i de X alors $g(x_i)$ peuvent être égaux. Cela était le cas ci-dessus puisque $1^2 = (-1)^2$. Soit $\{y_1, y_2, \dots\} = \{g(x_1), g(x_2), \dots\}$. Nous avons

$$\begin{aligned} \sum_{i \geq 1} g(x_i) \mathbb{P}(X = x_i) &= \sum_{j \geq 1} \sum_{i, g(x_i) = y_j} g(x_i) \mathbb{P}(X = x_i) = \sum_{j \geq 1} \sum_{i, g(x_i) = y_j} y_j \mathbb{P}(X = x_i) = \sum_{j \geq 1} y_j \sum_{i, g(x_i) = y_j} \mathbb{P}(X = x_i) \\ &= \sum_{j \geq 1} y_j \mathbb{P}(g(X) = y_j) \\ &= \mathbb{E}[g(X)]. \end{aligned}$$

□

THÉORÈME 4.7

L'espérance est linéaire : si X est une variable aléatoire discrète et $a, b \in \mathbb{R}$ alors nous avons

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b.$$

Démonstration. On a

$$\mathbb{E}[aX + b] = \sum_{x, p(x) > 0} (ax + b)p(x) = a \sum_{x, p(x) > 0} xp(x) + b \sum_{x, p(x) > 0} p(x) = a\mathbb{E}[X] + b.$$

□

Il y a un cas particulier de fonction de variable aléatoire qui correspond au monôme $f(x) = x^n$. Ceci s'appelle les *moments* de la variable aléatoire.

DÉFINITION 4.8

Le moment (non-centré) d'ordre n d'une variable aléatoire X est $\mathbb{E}[X^n]$.

On voit que l'espérance d'une variable aléatoire nous donne de l'information sur X , sa valeur moyenne. Mais par exemple, les trois variables aléatoires

$$X = 0, \quad Y = \begin{cases} -1 & \text{avec probabilité } \frac{1}{2}, \\ 1 & \text{avec probabilité } \frac{1}{2}, \end{cases} \quad Z = \begin{cases} -10^6 & \text{avec probabilité } \frac{1}{2}, \\ 10^6 & \text{avec probabilité } \frac{1}{2} \end{cases}$$

ont la même valeur moyenne, ici 0, mais sont pourtant très différentes ! On peut mesurer les variations de X par rapport à son espérance en définissant la *variance* d'une variable aléatoire.

DÉFINITION 4.9

Soit X une variable aléatoire, on définit sa variance par

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

Son écart-type est $\sigma = \sqrt{\text{Var}(X)}$.

LEMME 4.10. On a $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Démonstration. On a

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

□

EXEMPLE 33. Si X est le résultat d'un dé alors on a vu que $\mathbb{E}[X] = \frac{7}{2}$ et on a

$$\mathbb{E}[X^2] = 1^2\mathbb{P}(X=1) + 2^2\mathbb{P}(X=2) + 3^2\mathbb{P}(X=3) + 4^2\mathbb{P}(X=4) + 5^2\mathbb{P}(X=5) + 6^2\mathbb{P}(X=6) = \frac{91}{6}.$$

Ainsi, la variance d'un lancer de dé équilibré est de

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{91}{6} - \frac{7^2}{2^2} = \frac{35}{12}.$$

Contrairement à l'espérance, la variance n'est pas linéaire mais elle a tout de même une propriété similaire.

PROPOSITION 4.11. Soit $a, b \in \mathbb{R}$ et X une variable aléatoire alors

$$\text{Var}(aX + b) = a^2\text{Var}(X).$$

Démonstration. On a

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

□

4.2 Exemples de variables aléatoires discrètes

4.2.1 Variable aléatoire de Bernoulli et binomiale

L'exemple le plus simple de variable aléatoire discrète correspond à réaliser une expérience qui a 2 résultats : un succès avec probabilité p où un échec avec probabilité $1 - p$. On définit alors la variable aléatoire qui prend la valeur 1 si on a un succès et 0 sinon. C'est une variable dite de *Bernoulli*.

DÉFINITION 4.12

Une variable aléatoire X est dite de Bernoulli de paramètre $p \in (0, 1)$ ou $X \sim \text{Ber}(p)$ si

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

On peut généraliser cette définition au nombre de succès sur n épreuves indépendantes.

DÉFINITION 4.13

La variable aléatoire X comptant le nombre de succès de n épreuves indépendantes qui réussissent avec probabilité p est dite binomiale de paramètre (n, p) ou $X \sim \text{Bin}(n, p)$. On a

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad \text{pour } i \in \{0, \dots, n\}.$$

On remarque ainsi que si $X \sim \text{Ber}(p)$ alors $X \sim \text{Bin}(1, p)$. Comment obtenons-nous la loi de probabilité ? En fait, on l'a déjà fait plus haut : si $i \leq n$ alors on choisit quelles expériences seront des succès, on a donc $\binom{n}{i}$ choix et par l'indépendance des épreuves, la probabilité devient $(1 - p)^{n-i} p^i$. Si $i > n$ alors $\mathbb{P}(X = i) = 0$ et finalement on obtient

$$\mathbb{P}(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad \text{pour } i \in \{0, \dots, n\}.$$

On remarque que par la formule du binôme, on a

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

EXEMPLE 34. On jette cinq pièces équilibrées de manière indépendantes, on compte le nombre de Piles obtenus. Donnons la loi de probabilité de X .

Par définition, on sait que X suit une loi binomiale de paramètre $(5, \frac{1}{2})$ et ainsi,

$$\begin{aligned} \mathbb{P}(X = 0) &= \binom{5}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = \frac{1}{32}, & \mathbb{P}(X = 1) &= \binom{5}{1} \frac{1}{2} \times \left(\frac{1}{2}\right)^4 = \frac{5}{32}, \\ \mathbb{P}(X = 2) &= \binom{5}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = \frac{5}{16}, & \mathbb{P}(X = 3) &= \binom{5}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{5}{16}, \\ \mathbb{P}(X = 4) &= \binom{5}{4} \left(\frac{1}{2}\right)^4 \times \frac{1}{2} = \frac{5}{32}, & \mathbb{P}(X = 5) &= \binom{5}{5} \left(\frac{1}{2}\right)^5 = \frac{1}{32}. \end{aligned}$$

EXEMPLE 35. On réalise le jeu suivant : un joueur mise sur un numéro entre 1 et 6. On jette ensuite 3 dés et il gagne \$ i où i est le nombre de fois que le nombre choisi apparaît et il perd \$1 s'il n'y en a pas. Voudrions-nous jouer à ce jeu ?

Pour répondre à cela, on peut calculer le gain moyen du joueur. On sait que le nombre de fois que le nombre choisi apparaît est une loi binomiale de paramètre $(3, \frac{1}{6})$. Ainsi, si on désigne par X le gain du

joueur on obtient

$$\begin{aligned} \mathbb{P}(X = -1) &= \binom{3}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^3 = \frac{125}{216}, & \mathbb{P}(X = 1) &= \binom{3}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^2 = \frac{75}{216}, \\ \mathbb{P}(X = 2) &= \binom{3}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^1 = \frac{15}{216}, & \mathbb{P}(X = 3) &= \binom{3}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^0 = \frac{1}{216}. \end{aligned}$$

Ainsi le gain moyen du joueur est donc de

$$\mathbb{E}[X] = (-1) \times \frac{125}{216} + 1 \times \frac{75}{216} + 2 \times \frac{15}{216} + 3 \times \frac{1}{216} = -\frac{17}{216} < 0$$

En moyenne, le joueur est donc perdant! On peut modifier le jeu pour avoir un jeu équilibré, par exemple on peut modifier la perte du joueur si aucun des dés n'a son numéro. Si on dénote par a la perte du joueur et X_a son nouveau gain alors son espérance devient

$$\mathbb{E}[X_a] = (-a) \times \frac{125}{216} + 1 \times \frac{75}{216} + 2 \times \frac{15}{216} + 3 \times \frac{1}{216} = \frac{108 - 125a}{216}.$$

En choisissant $a = \frac{108}{125} = 0.864$, on obtient un gain moyen nul et le jeu est équilibré.

THÉORÈME 4.14

Soit $X \sim \text{Bin}(n, p)$ alors

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

De plus lorsque k croît, $\mathbb{P}(X = k)$ grandit jusqu'à $\lfloor (n + 1)p \rfloor = k$ puis décroît.

Démonstration. Calculons l'espérance d'une variable aléatoire $X \sim \text{Bin}(n, p)$. On a

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

Dans cette suite d'égalité, on a en fait prouvé la formule

$$i \binom{n}{i} = n \binom{n-1}{i-1}. \quad (3)$$

On continue en changeant d'indice $j = i - 1$,

$$\mathbb{E}[X] = n \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j+1} (1-p)^{n-(j+1)} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np(p + 1 - p)^{n-1} = np.$$

Pour la variance on va utiliser la formule de König-Huygens, $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. On commence ainsi à calculer

$$\mathbb{E}[X^2] = \sum_{i=0}^n i^2 \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^n i \cdot i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^n i n \binom{n-1}{i-1} p^i (1-p)^{n-i}$$

où on a utilisé la formule (3). On continue avec le même changement d'indice $j = i - 1$,

$$\mathbb{E}[X^2] = n \sum_{j=0}^{n-1} (j+1) \binom{n-1}{j} p^{j+1} (1-p)^{n-(j+1)} = np \sum_{j=0}^{n-1} (j+1) \binom{n-1}{j} p^j (1-p)^{n-1-j}$$

$$= np \sum_{j=0}^{n-1} (j+1) \mathbb{P}(Y=j) = np \mathbb{E}[Y+1]$$

avec $Y \sim \text{Bin}(n-1, p)$. Or, on vient de calculer l'espérance d'une variable binomiale et on obtient donc

$$\mathbb{E}[X^2] = np \mathbb{E}[Y+1] = np((n-1)p+1)$$

et finalement,

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

Enfin, pour voir la croissance de $\mathbb{P}(X=k)$, on peut calculer

$$\frac{\mathbb{P}(X=k+1)}{\mathbb{P}(X=k)} = \frac{\binom{n}{k+1} p^{k+1} (1-p)^{n-(k+1)}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{p}{1-p} \cdot \frac{n!}{(k+1)!(n-k-1)!} \cdot \frac{(n-k)!k!}{n!} = \frac{p(n-k)}{(1-p)(k+1)}$$

et ceci est plus petit que 1 pour $p(n-k) \leq (1-p)(k+1) \Leftrightarrow k \geq (n+1)p$. $\square$

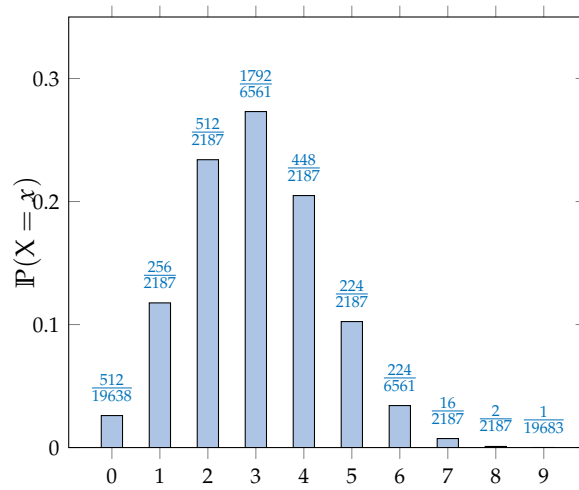


FIGURE 4 – Loi de probabilité de $X \sim \text{Bin}\left(9, \frac{1}{3}\right)$

4.2.2 Variable aléatoire de Poisson

DÉFINITION 4.15

Une variable aléatoire X prenant des valeurs dans $\mathbb{N}$ est dite de Poisson de paramètre $\lambda > 0$ si pour tout $i \in \mathbb{N}$,

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

On remarque que l'on a bien

$$\sum_{i=0}^{+\infty} \mathbb{P}(X=i) = \sum_{i=0}^{+\infty} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{i=0}^{+\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1.$$

Il y a un lien avec loi binomiale, si on prend $X \sim \text{Bin}(n, p)$ et on dénote $\lambda = np$, alors on peut écrire

$$\mathbb{P}(X=i) = \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} = \frac{\lambda^i}{i!} \left(1 - \frac{\lambda}{n}\right)^n \cdot \frac{n(n-1) \cdots (n-i+1)}{n^i} \cdot \frac{1}{\left(1 - \frac{\lambda}{n}\right)^i}.$$

Si on considère n grand et p petit de telle sorte que $\lambda = np$ reste constant alors avec i fixé, on remarque que

$$\left(1 - \frac{\lambda}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-\lambda}, \quad \frac{n(n-1) \cdots (n-i+1)}{n^i} \xrightarrow{n \rightarrow \infty} 1, \quad \frac{1}{\left(1 - \frac{\lambda}{n}\right)^i} \xrightarrow{n \rightarrow \infty} 1$$

et ainsi,

$$\mathbb{P}(X = i) \xrightarrow{n \rightarrow \infty} \frac{\lambda^i}{i!} e^{-\lambda} = \mathbb{P}(Y = i)$$

avec $Y \sim \text{Poi}(\lambda)$.

EXEMPLE 36. Un objet est défectueux avec probabilité 0.1. Calculons la probabilité que 10 objets comprennent au plus un objet défectueux.

Le nombre d'objets défectueux X est donc une variable aléatoire de loi binomiale de paramètre $\left(10, \frac{1}{10}\right)$. On cherche donc

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \binom{10}{0} \left(\frac{9}{10}\right)^{10} + \binom{10}{1} \frac{1}{10} \cdot \left(\frac{9}{10}\right)^9 \approx 0.7361.$$

On remarque qu'ici nous avons $np = 1$. Si on approche la variable aléatoire X par $Y \sim \text{Poi}(1)$, alors on aurait

$$\mathbb{P}(Y \leq 1) = \mathbb{P}(Y = 0) + \mathbb{P}(Y = 1) = e^{-1} \frac{1^0}{0!} + e^{-1} \frac{1}{1!} = e^{-1} + e^{-1} \approx 0.7358.$$

On voit ainsi que même avec $n = 10$, on a une approximation plutôt bonne!

THÉORÈME 4.16

Soit $X \sim \text{Poi}(\lambda)$ alors

$$\mathbb{E}[X] = \lambda \quad \text{et} \quad \text{Var}(X) = \lambda.$$

Démonstration. On commence par calculer l'espérance

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

On calcule maintenant le deuxième moment,

$$\mathbb{E}[X^2] = \sum_{i=0}^{+\infty} i^2 \mathbb{P}(X = i) = \sum_{i=1}^{+\infty} i^2 \frac{\lambda^i}{i!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{i=1}^{+\infty} i \frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{+\infty} (j+1) \frac{\lambda^j}{j!}$$

et on peut ainsi utiliser le calcul de l'espérance pour écrire

$$\mathbb{E}[X^2] = \lambda e^{-\lambda} \left(\sum_{j=0}^{+\infty} j \frac{\lambda^j}{j!} + \sum_{j=0}^{+\infty} \frac{\lambda^j}{j!} \right) = \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda}) = \lambda(\lambda + 1).$$

Finalement, on obtient

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

□

On remarque que l'on pouvait deviner ces valeurs en utilisant l'approximation de la loi binomiale par la loi de Poisson, en effet si $X \sim \text{Bin}(n, p)$ et $\lambda = np$ avec $n \rightarrow +\infty$ et $p \rightarrow 0$ alors

$$\mathbb{E}[X] = np = \lambda, \quad \text{Var}(X) = np(1-p) \rightarrow \lambda.$$

L'approximation de Poisson est en fait vraie pour approcher le nombre de succès d'épreuves "faiblement dépendantes". Nous n'allons pas formaliser ceci ici mais nous allons l'utiliser dans l'exemple suivant.

EXEMPLE 37. On considère n personnes qui ont chacune une probabilité de $\frac{1}{365}$ d'avoir un anniversaire donné (on oublie le 29 Février). On considère alors tous les couples de personnes, on sait qu'il y en a $\binom{n}{2} = \frac{n(n-1)}{2}$ et ainsi nous voulons introduire pour l'approximation de Poisson, le paramètre

$$\lambda = \frac{n(n-1)}{2} \cdot \frac{1}{365} = \frac{n(n-1)}{730}.$$

On remarque que les événements $E_{ij} = \{\text{le couple } (i, j) \text{ a le même anniversaire}\}$ ne sont pas indépendants entre eux mais nous allons tout de même admettre l'approximation de Poisson avec $Y \sim \text{Poi}(\lambda)$,

$$\begin{aligned} \mathbb{P}(\text{Aucun anniversaire en commun}) &= \mathbb{P}\left(0 \text{ succès parmi les } \binom{n}{2} \text{ couple}\right) \approx \mathbb{P}(Y = 0) \\ &= \exp\left(-\frac{n(n-1)}{730}\right). \end{aligned}$$

Ainsi, si on cherche le nombre minimal de personnes pour que cette probabilité soit plus petite que $\frac{1}{2}$, on cherche

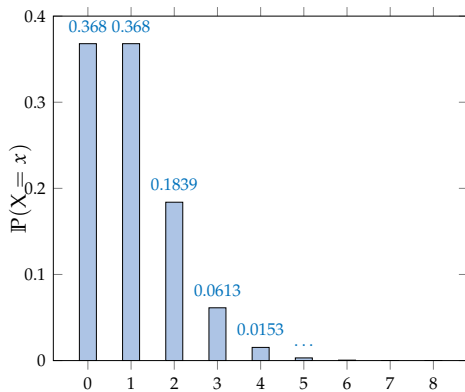
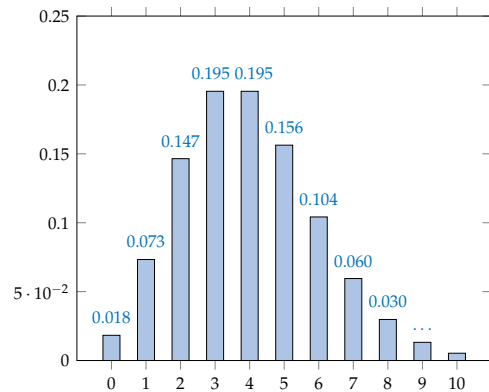
$$\exp\left(-\frac{n(n-1)}{730}\right) \leq \frac{1}{2} \iff n(n-1) \geq 730 \ln(2) \iff n \geq 23.$$

Ainsi, si on a plus de 23 personnes, on a une chance sur deux que deux personnes aient le même anniversaire. Si on veut faire la même chose pour que trois personnes aient le même anniversaire alors on peut généraliser le calcul avec

$$\binom{n}{3} = \frac{n(n-1)(n-2)}{6}, \quad \lambda = \frac{n(n-1)(n-2)}{6} \cdot \left(\frac{1}{365}\right)^2 = \frac{n(n-1)(n-2)}{799350}.$$

Et ainsi pour avoir une probabilité d'avoir un anniversaire en commun pour 3 personnes de plus de $\frac{1}{2}$, on doit avoir

$$n(n-1)(n-2) \geq 799350 \ln(2) \iff n = 84.$$

FIGURE 5 – Loi de probabilité de $X \sim \text{Poi}(1)$ FIGURE 6 – Loi de probabilité de $X \sim \text{Poi}(4)$

4.2.3 Variable aléatoire géométrique

DÉFINITION 4.17

Soit une série d'épreuves indépendantes avec une probabilité de succès p avec $p \in (0, 1)$. Soit X le nombre d'épreuves nécessaires pour obtenir un premier succès, X est dit une loi géométrique de paramètre p et pour tout $i \in \mathbb{N}^*$,

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

On vérifie que l'on a bien

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

EXEMPLE 38. On a B boules blanches et N boules noires. On tire avec remise jusqu'à obtenir une boule noire. Quelle est la probabilité de devoir faire exactement n tirages ? On voit que le nombre de tirages effectués est une variable aléatoire géométrique de paramètre $\frac{N}{B+N}$ qui est la probabilité de tirer une boule noire. Ainsi, on cherche si $X \sim \text{Geo}\left(\frac{N}{B+N}\right)$,

$$\mathbb{P}(X = n) = \frac{N}{B+N} \left(\frac{B}{B+N}\right)^{n-1} = \frac{NB^{n-1}}{(B+N)^n}.$$

Si on cherche maintenant la probabilité de devoir faire au moins n tirages, alors on peut calculer de deux façons cette probabilité, tout d'abord en écrivant

$$\begin{aligned} \mathbb{P}(X \geq n) &= \mathbb{P}\left(\bigcup_{k=n}^{+\infty} \{X = k\}\right) = \sum_{k=n}^{+\infty} \mathbb{P}(X = k) = \sum_{k=n}^{+\infty} \frac{N}{B+N} \left(\frac{B}{B+N}\right)^{k-1} = \frac{N}{B+N} \sum_{\ell=n-1}^{+\infty} \left(\frac{B}{B+N}\right)^{\ell} \\ &= \frac{N}{B+N} \cdot \frac{\left(\frac{B}{B+N}\right)^{n-1}}{1 - \frac{B}{B+N}} \\ &= \left(\frac{B}{B+N}\right)^{n-1}. \end{aligned}$$

On aurait pu trouver ce résultat directement en disant que devoir faire au moins n tirages est le même événement que faire $n-1$ échecs i.e tirer une boule blanche. Par indépendance des tirages, on obtient une probabilité de $\left(\frac{B}{B+N}\right)^{n-1}$.

THÉORÈME 4.18

Soit $X \sim \text{Geo}(p)$ alors

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

Démonstration. On calcule

$$\mathbb{E}[X] = \sum_{k=1}^{+\infty} k \mathbb{P}(X = k) = \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = p \sum_{k=1}^{+\infty} k(1-p)^{k-1}.$$

L'idée est de voir que si on définit pour $|x| < 1$, $f(x) = \sum_{k=0}^{+\infty} x^k$ alors on a $f'(x) = \sum_{k=0}^{+\infty} kx^{k-1}$ et ainsi

$$\mathbb{E}[X] = pf'(1-p).$$

Mais on a aussi $f(x) = \frac{1}{1-x}$ et ainsi $f'(x) = \frac{1}{(1-x)^2}$. Finalement, on obtient donc

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

Pour calculer la variance, on calcule le second moment et on obtient

$$\mathbb{E}[X^2] = \sum_{k=1}^{+\infty} k^2(1-p)^{k-1}p = p \sum_{k=1}^{+\infty} k^2(1-p)^{k-1}.$$

De la même manière, on peut définir pour $|x| < 1$, $g(x) = \sum_{k=1}^{+\infty} kx^k$ et on a $g'(x) = \sum_{k=1}^{+\infty} k^2x^{k-1}$ et ainsi

$$\mathbb{E}[X^2] = pg'(1-p).$$

Mais on peut calculer g et g' en réalisant que

$$g(x) = \sum_{k=1}^{+\infty} kx^k = x \sum_{k=1}^{+\infty} kx^{k-1} = xf'(x) = \frac{x}{(1-x)^2} \quad \text{et ainsi} \quad g'(x) = \frac{1+x}{(1-x)^3}.$$

Finalement, on obtient donc

$$\mathbb{E}[X^2] = p \frac{1+(1-p)}{(1-(1-p))^3} = \frac{2}{p^2} - \frac{1}{p}.$$

Et ainsi

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

□

La loi géométrique a une propriété très importante. Elle est dite *sans mémoire* que l'on définit par le théorème suivant.

THÉORÈME 4.19

Soit X une variable aléatoire $\text{Geo}(p)$ alors pour tout $k \in \mathbb{N}^*$ et $n \in \mathbb{N}^*$,

$$\mathbb{P}(X > k+n | X > n) = \mathbb{P}(X > k).$$

De plus la loi géométrique de paramètre p est la seule variable aléatoire dans $\mathbb{N}^*$ ayant cette propriété telle que $1-p = \mathbb{P}(X > 1)$.

Démonstration. On a

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

Ainsi, on peut calculer la probabilité conditionnelle

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

Pour montrer que c'est la seule, on va montrer par récurrence que si on a la propriété d'absence de mémoire alors on a $\mathbb{P}(X > n) = (1-p)^n$. Pour $n = 1$, on a par hypothèse que $1-p = \mathbb{P}(X > 1)$. Maintenant, si on suppose que ceci est vrai pour un $n \in \mathbb{N}^*$ alors on a

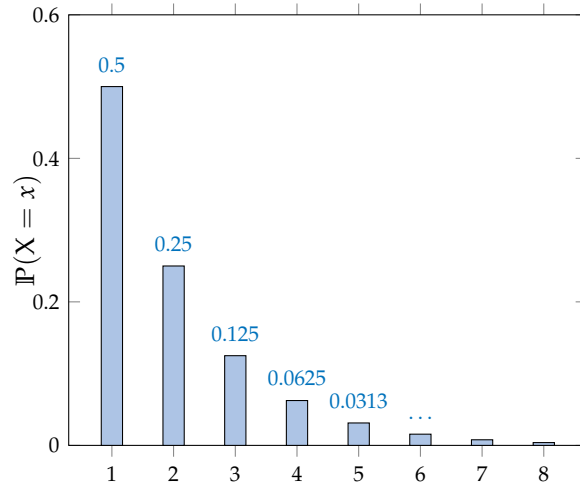
$$\begin{aligned} \mathbb{P}(X > n+1) &= \mathbb{P}(\{X > n+1\} \cap \{X > n\}) = \mathbb{P}(X > n+1 | X > n) \mathbb{P}(X > n) = \mathbb{P}(X > 1) \mathbb{P}(X > n) \\ &= (1-p) \cdot (1-p)^n \\ &= (1-p)^{n+1}. \end{aligned}$$

On peut ainsi retrouver

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

□

Cette propriété est bien cohérente avec la définition de la loi géométrique comme le premier succès d'une série d'épreuves. En effet, si on a eu n échecs, avoir un succès après $k+n$ échecs correspond à oublier ces n premiers échecs et attendre k échecs.


 FIGURE 7 – Loi de probabilité de $X \sim \text{Geo}\left(\frac{1}{2}\right)$

4.2.4 Variable aléatoire binomiale négative

DÉFINITION 4.20

On réalise une série d'épreuves indépendantes avec probabilité de succès $p \in (0, 1)$ jusqu'à obtenir r succès. On définit X le nombre d'épreuves nécessaires, on dit que X suit une loi binomiale négative de paramètre (r, p) et pour tout $n \in \{r, r+1, \dots\}$,

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

On remarque que cette définition est claire par rapport à sa définition, en effet on doit tout d'abord avoir $r-1$ succès parmi les $n-1$ premières épreuves ce qui donne $\binom{n-1}{r-1} p^{r-1} (1-p)^{n-r}$. Ensuite, on sait que la n -ième épreuve est un succès avec probabilité p . On obtient bien finalement $\binom{n-1}{r-1} p^r (1-p)^{n-r}$.

THÉORÈME 4.21

Soit $X \sim \text{BinNeg}(r, p)$ alors

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

Démonstration. Faisons la preuve du calcul de l'espérance. Nous laissons la preuve de la variance en exercice. On commence par rappeler la formule que nous avons vu plus haut

$$n \binom{n-1}{r-1} = r \binom{n}{r}.$$

Ainsi on a

$$\mathbb{E}[X] = \sum_{n=r}^{+\infty} n \binom{n-1}{r-1} p^r (1-p)^{n-r} = \sum_{n=r}^{+\infty} r \binom{n}{r} p^r (1-p)^{n-r} = \frac{r}{p} \sum_{n=r}^{+\infty} \binom{n}{r} p^{r+1} (1-p)^{n-r}.$$

Il reste à prouver que la somme est égale à 1. Mais on remarque que l'on peut écrire

$$\begin{aligned} \sum_{n=r}^{+\infty} \binom{n}{r} p^{r+1} (1-p)^{n-r} &= \sum_{m=r+1}^{+\infty} \binom{m-1}{r} p^{r+1} (1-p)^{m-1-r} = \sum_{m=r+1}^{+\infty} \binom{m-1}{(r+1)-1} p^{r+1} (1-p)^{m-(r+1)} \\ &= \sum_{m=r+1}^{+\infty} \mathbb{P}(Y = m) = 1 \end{aligned}$$

avec $Y \sim \text{BinNeg}(r+1, p)$. □

On remarque que $X \sim \text{BinNeg}(1, p) \sim \text{Geo}(p)$. En effet on a

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

EXEMPLE 39. Calculons la moyenne du nombre de jets de dé pour obtenir 4 fois la valeur 1. On remarque que ce nombre de jets est donné par $X \sim \text{BinNeg}\left(4, \frac{1}{6}\right)$ et ainsi

$$\mathbb{E}[X] = \frac{4}{\frac{1}{6}} = 24.$$

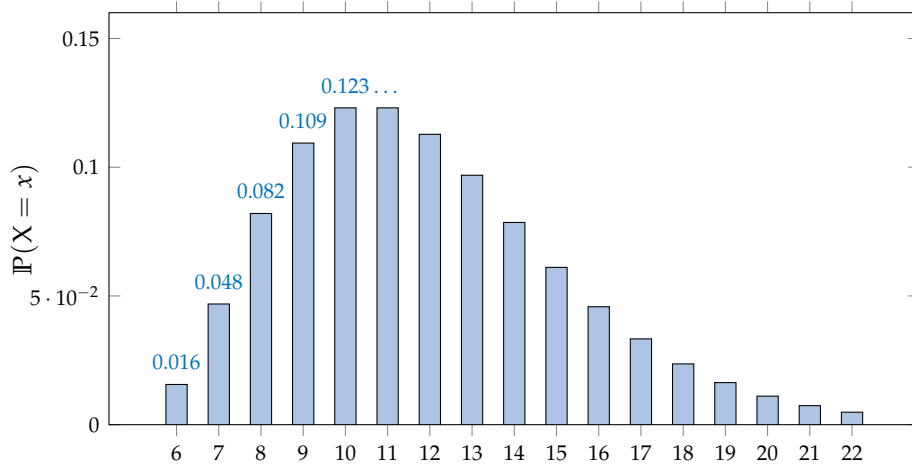


FIGURE 8 – Loi de probabilité de $X \sim \text{BinNeg}\left(6, \frac{1}{2}\right)$

4.2.5 Variable aléatoire hypergéométrique

DÉFINITION 4.22

On tire sans remise n boules d'une urne contenant N boules dont m sont blanches et $N - m$ sont noires. On définit X comme le nombre de boules blanches tirées, on dit que X suit une loi hypergéométrique de paramètre (n, m, N) et on a pour tout $i \in \{0, \dots, n\}$

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

avec la convention que $\binom{n}{k} = 0$ pour $k < 0$ ou $k > n$.

EXEMPLE 40. On achète des composants par paquets de 10. On en regarde 3 et on accepte le lot si les 3 composants fonctionnent. On suppose que 30% des lots ont 4 composants défectueux et que 70% des lots ont 1 composant défectueux. Calculons la proportion des lots que l'on rejettera. On définit

$$A = \{\text{On accepte le lot}\}$$

alors par la formule des probabilités totales on a

$$\mathbb{P}(A) = \mathbb{P}(A|4 \text{ défectueux}) \mathbb{P}(4 \text{ défectueux}) + \mathbb{P}(A|1 \text{ défectueux}) \mathbb{P}(1 \text{ défectueux}).$$

Si on sait que l'on a 4 composants défectueux, alors le lot est accepté si on choisit 3 composants parmi les 6

composants qui fonctionnent et ainsi on a si $X \sim \text{HyperGéo}(3, 6, 10)$

$$\mathbb{P}(A|4 \text{ défectueux}) = \mathbb{P}(X = 3) = \frac{\binom{6}{3}\binom{4}{0}}{\binom{10}{3}} = \frac{20 \times 1}{120} = \frac{1}{6}.$$

De même, si $Y \sim \text{HyperGéo}(3, 9, 10)$ alors on a

$$\mathbb{P}(A|1 \text{ défectueux}) = \mathbb{P}(Y = 3) = \frac{\binom{9}{3}\binom{1}{0}}{\binom{10}{3}} = \frac{84 \times 1}{120} = \frac{7}{10}.$$

Finalement, on obtient

$$\mathbb{P}(A) = \frac{1}{6} \times \frac{3}{10} + \frac{7}{10} \times \frac{7}{10} = \frac{27}{50}.$$

Ainsi, on rejettera en moyenne 46% des lots.

THÉORÈME 4.23

Soit $X \sim \text{HyperGeo}(n, m, N)$ alors

$$\mathbb{E}[X] = \frac{nm}{N} \quad \text{et} \quad \text{Var}(X) = \frac{nm}{N} \left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N} \right)$$

Démonstration. Faisons la preuve de l'espérance, on laisse la preuve de la variance en exercice. On commence par écrire

$$\mathbb{E}[X] = \sum_{i=0}^n i \mathbb{P}(X = i) = \sum_{i=1}^n i \mathbb{P}(X = i) = \sum_{i=1}^n i \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}} = \sum_{i=1}^n \frac{m \binom{m-1}{i-1} \binom{N-m}{n-i}}{\binom{N}{n}}$$

où on a utilisé la formule $i \binom{m}{i} = m \binom{m-1}{i-1}$. Si on utilise aussi la même formule $\binom{N}{n} = \frac{N}{n} \binom{N-1}{n-1}$ alors on obtient

$$\mathbb{E}[X] = m \sum_{i=1}^n \frac{\binom{m-1}{i-1} \binom{N-m}{n-i}}{\frac{N}{n} \binom{N-1}{n-1}} = \frac{nm}{N} \sum_{j=0}^n \frac{\binom{m-1}{j} \binom{N-m}{n-1-j}}{\binom{N-1}{n-1}} = \frac{nm}{N} \sum_{j=0}^n \mathbb{P}(Y = j) = \frac{nm}{N}$$

avec $Y \sim \text{HyperGéo}(n-1, m-1, N-1)$. □

4.3 Espérance de la somme de deux variables aléatoires discrètes

Si on considère deux variables aléatoires X et Y , comment pouvons-nous calculer $\mathbb{E}[X + Y]$? Revenons tout d'abord à la définition d'une variable aléatoire comme fonction de Ω vers $\mathbb{R}$, $X : \Omega \rightarrow \mathbb{R}$ et supposons que Ω soit dénombrable. Commençons par l'exemple suivant : on considère 5 lancers de pièces de telle sorte que

$$\Omega = \{(P, F)\}^5$$

et dénotons X le nombre de Faces dans les 2 premiers lancers et Y le nombre de Faces dans les deux derniers lancers. Alors si on considère l'événement $\{F, P, F, P, F\}$ alors

$$X(\{F, P, F, P, F\}) = 2 \quad \text{et} \quad Y(\{F, P, F, P, F\}) = 1.$$

Ainsi, si $Z = X + Y$, on a

$$Z(\{F, P, F, P, F\}) = (X + Y)(\{F, P, F, P, F\}) = X(\{F, P, F, P, F\}) + Y(\{F, P, F, P, F\}) = 3.$$

Revoir les variables aléatoires sous cette forme nous permet notamment de réécrire l'espérance d'une variable aléatoire ainsi que de calculer l'espérance d'une somme finie de variables aléatoires.

THÉORÈME 4.24

Pour Ω un ensemble fondamental dénombrable et X une variable aléatoire, on a

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}).$$

En particulier, si $X_1, \dots, X_n$ sont n variables aléatoires définies sur S alors

$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i].$$

Démonstration. On suppose que X est à valeurs dans $\{x_1, x_2, \dots\}$ alors si on définit l'événement

$$\Omega_i = \{X = x_i\} = \{\omega \in \Omega, X(\omega) = x_i\},$$

on a

$$\mathbb{E}[X] = \sum_{i \geq 1} x_i \mathbb{P}(X = x_i) = \sum_{i \geq 1} x_i \mathbb{P}(\Omega_i) = \sum_{i \geq 1} x_i \sum_{\omega \in \Omega_i} \mathbb{P}(\{\omega\}) = \sum_{i \geq 1} \sum_{\omega \in \Omega_i} X(\omega) \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}).$$

Pour la somme de variables aléatoires on voit que si on définit $Z = \sum_{i=1}^n X_i$ alors

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{\omega \in \Omega} Z(\omega) \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} (X_1 + \dots + X_n)(\omega) \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} X_1(\omega) \mathbb{P}(\{\omega\}) + \dots + \sum_{\omega \in \Omega} X_n(\omega) \mathbb{P}(\{\omega\}) \\ &= \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]. \end{aligned}$$

□

EXEMPLE 41. Calculons la valeur moyenne de la somme des valeurs de n dés. Si X_i est la valeur du dé i alors $X = \sum_{i=1}^n X_i$ et ainsi comme $\mathbb{E}[X] = \frac{7}{2}$ on obtient

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \frac{7n}{2}.$$

EXEMPLE 42. On suppose qu'on a n expériences qui réussissent avec probabilité p_i pour $i = 1, \dots, n$ et ces expériences ne sont pas nécessairement indépendantes. Quelle est le nombre moyen de succès? Si on considère

$$X_i = \begin{cases} 1 & \text{si succès,} \\ 0 & \text{si échec.} \end{cases}$$

Alors $X = \sum_{i=1}^n X_i$ et on a

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p_i.$$



On voit que l'espérance de la somme de deux variables aléatoires est toujours donnée par la somme des espérances même si les variables aléatoires ne sont pas indépendantes.

4.4 Variables aléatoires continues

On suppose maintenant que l'ensemble fondamental Ω est infini indénombrable, par exemple $\Omega = \mathbb{R}_+$ ou $\Omega = \mathbb{R}$. On définit une variable aléatoire continue par la définition suivante.

DÉFINITION 4.25

$X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire continue (ou absolument continue) s'il existe une fonction $f : \mathbb{R} \rightarrow \mathbb{R}_+$ positive telle que pour tout $a \leq b$,

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

f est appelée la densité de X .

On remarque tout d'abord que cela donne une condition sur la densité. En effet, en prenant a et b vers l'infini, on obtient donc

$$\mathbb{P}(X \in \mathbb{R}) = 1 = \int_{-\infty}^{+\infty} f(x) dx.$$

De plus, nous avons que

$$\mathbb{P}(X = a) = \mathbb{P}(a \leq X \leq a) = \int_a^a f(x) dx = 0.$$

En particulier, on remarque donc que pour une variable aléatoire continue, nous avons

$$\mathbb{P}(X < a) = \mathbb{P}(X \leq a) = \int_{-\infty}^a f(x) dx.$$

Même si $\mathbb{P}(X = a) = 0$ pour tout a , nous pouvons voir la densité de manière informelle comme une version infinitésimale de ce résultat. En effet, on a pour tout $\varepsilon > 0$,

$$\mathbb{P}(x \leq X \leq x + \varepsilon) = \int_x^{x+\varepsilon} f(y) dy \underset{\varepsilon \rightarrow 0}{\approx} f(x) \varepsilon$$

On peut ainsi réécrire cette approximation comme

$$f(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{P}(x \leq X \leq x + \varepsilon).$$

De manière informelle (et non rigoureuse), on peut réécrire cela de la façon suivante en remplaçant ε par dx ,

$$f(x) dx = \mathbb{P}(x \leq X \leq x + dx).$$

Même si nous ne pouvons pas utiliser directement ce résultat, car non rigoureux, il peut nous permettre de comprendre certains résultats sur les variables aléatoires continues.

EXEMPLE 43. La durée de fonctionnement (en heures) d'un ordinateur avant sa première panne X est une variable continue avec densité donnée par

$$f(x) = \begin{cases} \lambda \exp\left(-\frac{x}{100}\right) & \text{si } x \geq 0, \\ 0 & \text{sinon} \end{cases} =: \lambda \exp\left(-\frac{x}{100}\right) \mathbb{1}_{x \geq 0}.$$

Calculons la probabilité que cette durée soit entre 50 et 150 heures. Nous pouvons utiliser la définition d'une variable aléatoire continue pour écrire

$$\begin{aligned} \mathbb{P}(50 \leq X \leq 150) &= \int_{50}^{150} f(x) dx = \int_{50}^{150} \lambda \exp\left(-\frac{x}{100}\right) dx = \left[-100\lambda \exp\left(-\frac{x}{100}\right)\right]_{50}^{150} \\ &= 100\lambda \left(\frac{1}{\sqrt{e}} - \frac{1}{\sqrt{e^3}}\right). \end{aligned}$$

Cela nous donne une réponse en fonction du paramètre λ mais nous pouvons le calculer en utilisant la propriété des densités. En effet, nous savons que

$$1 = \int_{-\infty}^{+\infty} f(x) dx = \int_0^{+\infty} \lambda \exp\left(-\frac{x}{100}\right) dx = 100\lambda \Rightarrow \lambda = \frac{1}{100}.$$

Finalement, nous obtenons

$$\mathbb{P}(50 \leq X \leq 150) = \frac{1}{\sqrt{e}} - \frac{1}{\sqrt{e^3}} \approx 0.384.$$

EXEMPLE 44. La durée de vie d'un composant (en heures) est donnée par la densité

$$f(x) = \frac{100}{x^2} \mathbb{1}_{x>100}.$$

Une machine a 5 composants indépendants. Calculons la probabilité que 2 des 5 composants ne marchent plus lors des 150 premières heures. On peut commencer par vérifier que l'on a bien une densité, c'est bien une fonction positive et de plus

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} \frac{100}{x^2} \mathbb{1}_{x>100} dx = \int_{100}^{+\infty} \frac{100}{x^2} dx = \left[-\frac{100}{x} \right]_{100}^{+\infty} = 1.$$

Si on considère un des composants, en définissant l'événement

$$E_i = \{\text{le composant } i \text{ ne marche plus pendant les 150 premières heures}\}$$

alors

$$\mathbb{P}(E_i) = \int_0^{150} f(x) dx = \int_{100}^{150} \frac{100}{x^2} dx = \left[-\frac{100}{x} \right]_{100}^{150} = -\frac{100}{150} + 1 = \frac{1}{3}.$$

Ainsi par indépendance des composants, la probabilité recherchée est la probabilité qu'une variable aléatoire binomiale de paramètre $(5, \frac{1}{3})$ soit égale à 2 et on obtient donc

$$\binom{5}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = \frac{80}{243}.$$

THÉORÈME 4.26

Soit X une variable aléatoire continue de densité f et soit F sa fonction de répartition alors pour tout $x \in \mathbb{R}$,

$$F(x) = \int_{-\infty}^x f(y) dy.$$

En particulier, F est continue et si f est continue en x alors

$$F'(x) = f(x).$$

Démonstration. La preuve vient du fait que

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(y) dy.$$

Par le théorème fondamental de l'analyse, nous avons

$$F'(x) = \frac{d}{dx} \int_{-\infty}^x f(y) dy = f(x).$$

□

EXEMPLE 45. Soit X une variable continue de densité f_X , calculons la densité de la variable aléatoire $Y = 2X$ en fonction de f_X . On commence par écrire

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(2X \leq x) = \mathbb{P}\left(X \leq \frac{x}{2}\right) = F_X\left(\frac{x}{2}\right).$$

Ainsi, en dérivant, on obtient que

$$f_Y(x) = F'_Y(x) = \frac{1}{2} F'_X\left(\frac{x}{2}\right) = \frac{1}{2} f_X\left(\frac{x}{2}\right).$$

DÉFINITION 4.27

Soit X une variable aléatoire continue de densité f alors

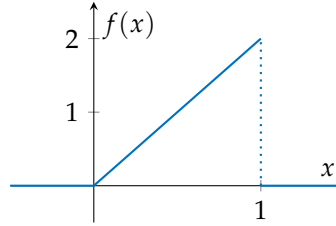
$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx.$$

On remarque que l'on peut faire le lien avec le cas discret où nous avons,

$$\mathbb{E}[X] = \sum_x x\mathbb{P}(X=x) \leftrightarrow \int_{\mathbb{R}} x\mathbb{P}(x \leq X \leq x+dx) \leftrightarrow \int_{\mathbb{R}} xf(x)dx.$$

EXEMPLE 46. Calculons l'espérance de X avec X une variable continue de densité

$$f(x) = 2x\mathbb{1}_{0 \leq x \leq 1}.$$



Commençons par vérifier que l'on a bien une densité

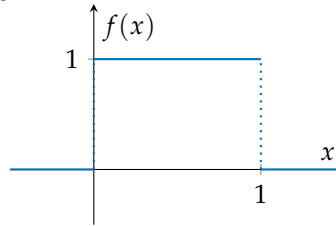
$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^1 2xdx = [x^2]_0^1 = 1.$$

Et ainsi

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 2x^2dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}.$$

EXEMPLE 47. On considère maintenant X de densité

$$f(x) = \mathbb{1}_{0 \leq x \leq 1}.$$

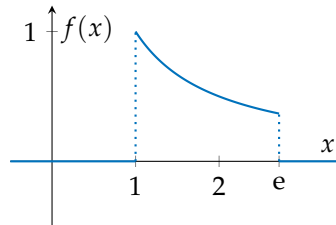


Calculons $\mathbb{E}[e^X]$. Si on définit $Y = e^X$ alors Y est à valeurs dans $[1, e]$ et de plus pour $x \in [1, e]$,

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(e^X \leq x) = \mathbb{P}(X \leq \ln(x)) = \int_0^{\ln(x)} dy = \ln(x).$$

Ainsi, en dérivant, on obtient

$$f_Y(x) = F'_Y(x) = \frac{1}{x}\mathbb{1}_{1 \leq x \leq e}.$$



Ainsi, on peut calculer

$$\mathbb{E}[e^X] = \int_{-\infty}^{+\infty} yf_Y(y)dy = \int_1^e \frac{y}{y}dy = e - 1.$$

Dans ce dernier exemple, on voit que pour calculer l'espérance nous avons du connaître la fonction inverse de e^X , ce n'est pas toujours le cas. On a cependant un théorème similaire au cas discret qui nous permet de calculer ce type d'espérance!

THÉORÈME 4.28

Si X est une variable aléatoire à densité f alors pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$, nous avons

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

On peut retrouver beaucoup plus facilement le résultat de l'exemple,

$$\mathbb{E}[e^X] = \int_{-\infty}^{+\infty} e^x f(x)dx = \int_0^1 e^x dx = e - 1.$$

Pour prouver ce théorème, nous commençons par prouver une autre formule de l'espérance.

THÉORÈME 4.29

Si X est une variable aléatoire continue positive alors

$$\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X > x) dx.$$

Démonstration. Si f est la densité de X alors nous avons

$$\begin{aligned} \int_0^{+\infty} \mathbb{P}(X > x) dx &= \int_0^{+\infty} \left(\int_x^{+\infty} f(y) dy \right) dx = \int_0^{+\infty} \left(\int_0^{+\infty} f(y) \mathbb{1}_{y \geq x} dy \right) dx \\ &= \int_0^{+\infty} \left(\int_0^{+\infty} f(y) \mathbb{1}_{y \geq x} dx \right) dy \\ &= \int_0^{+\infty} f(y) \left(\int_0^y dx \right) dy \\ &= \int_0^{+\infty} y f(y) dy = \mathbb{E}[X]. \end{aligned}$$

□

Nous sommes maintenant prêts à prouver le Théorème 4.28.

Démonstration du Théorème 4.28. Commençons par considérer le cas où g est positive. Dans ce cas,

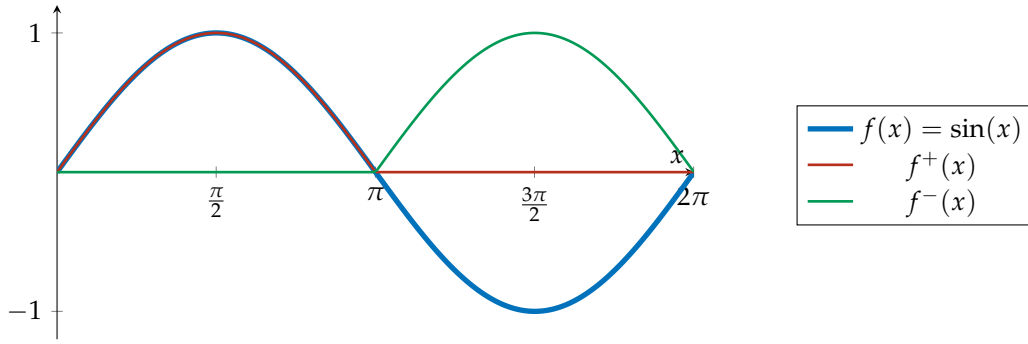
$$\begin{aligned} \mathbb{E}[g(X)] &= \int_0^{+\infty} \mathbb{P}(g(X) > x) dx = \int_0^{+\infty} \left(\int_{y, g(y) > x} f(y) dy \right) dx = \int_0^{+\infty} \left(\int_{-\infty}^{+\infty} f(y) \mathbb{1}_{g(y) > x} dy \right) dx \\ &= \int_{-\infty}^{+\infty} \left(\int_0^{+\infty} f(y) \mathbb{1}_{g(y) > x} dx \right) dy \\ &= \int_{-\infty}^{+\infty} f(y) \left(\int_0^{g(y)} dx \right) dy \\ &= \int_{-\infty}^{+\infty} g(y) f(y) dy. \end{aligned}$$

Dans le cas d'une fonction quelconque, on peut faire la décomposition illustrée en Figure 9.

$$g(x) = g^+(x) - g^-(x) \quad \text{avec} \quad g^+(x) = \max(g(x), 0) \quad \text{et} \quad g^-(x) = \max(-g(x), 0).$$

Le but de cette décomposition est que g^+ et g^- sont toutes deux des fonctions positives et on peut donc écrire

$$\begin{aligned} \mathbb{E}[g(x)] &= \mathbb{E}[g^+(x) - g^-(x)] = \mathbb{E}[g^+(x)] - \mathbb{E}[g^-(x)] = \int_{-\infty}^{+\infty} g^+(x) f(x) dx - \int_{-\infty}^{+\infty} g^-(x) f(x) dx \\ &= \int_{-\infty}^{+\infty} (g^+(x) - g^-(x)) f(x) dx \\ &= \int_{-\infty}^{+\infty} g(x) f(x) dx. \end{aligned}$$


 FIGURE 9 – Illustration de la décomposition en partie positive et négative pour la fonction $f(x) = \sin(x)$

□

EXEMPLE 48. Un transporteur doit livrer des marchandises à une usine à une heure précise.

- S’il arrive en avance de s minutes, le camion doit attendre sur place jusqu’à l’ouverture du quai de déchargement. Ce temps d’attente immobilise le camion et le chauffeur, ce qui engendre un coût proportionnel à l’avance : as euros.
- S’il arrive en retard de s minutes, la chaîne de production de l’usine est ralentie ou arrêtée, ce qui engendre un coût de rs euros.

Le temps de trajet T entre le dépôt et l’usine est aléatoire (dépendant par exemple du trafic) et suit une densité f . À quelle heure le transporteur devrait-il quitter le dépôt afin de minimiser le coût moyen (espéré) de cette livraison ?

Si on part t minutes avant le rendez-vous alors nous pouvons donner la fonction du coût en fonction du temps de trajet T ,

$$C_t(X) = \begin{cases} a(t - X) & \text{si } X \leq t, \\ r(X - t) & \text{si } X \geq t. \end{cases}$$

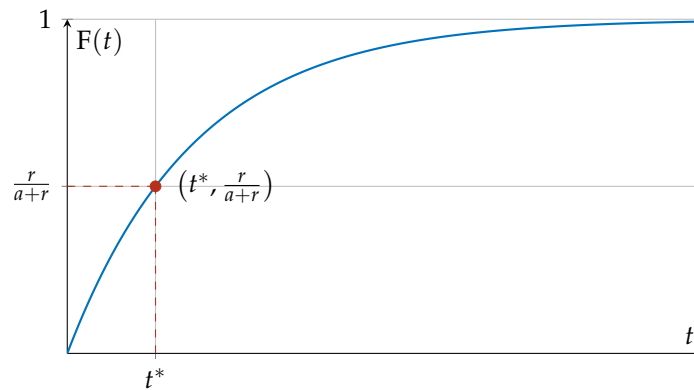
Ainsi, le coût moyen est donc de

$$\begin{aligned} \mathbb{E}[C_t(X)] &= \int_0^{+\infty} C_t(x)f(x)dx = \int_0^t a(t-x)f(x)dx + \int_t^{+\infty} r(x-t)f(x)dx \\ &= at \int_0^t f(x)dx - a \int_0^t xf(x)dx + r \int_t^{+\infty} xf(x)dx - rt \int_t^{+\infty} f(x)dx. \end{aligned}$$

Ainsi, en dénotant $F(t)$ la fonction de répartition de X , en dérivant cette fonction en t , nous obtenons

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[C_t(X)] &= a \int_0^t f(x)dx + atf(t) - atf(t) - rt f(t) - r \int_t^{+\infty} f(x)dx + rt f(t) = aF(t) - r(1 - F(t)) \\ &= (a + r)F(t) - r. \end{aligned}$$

Ainsi, si on cherche le minimum du coût moyen, on cherche t^* tel que $F(t^*) = \frac{r}{a+r}$.



Nous avons les mêmes propriétés de l'espérance et de la variance que dans le cas discret.

THÉORÈME 4.30

Pour $a, b \in \mathbb{R}$ et X une variable aléatoire continue on a toujours

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b.$$

De même si

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

alors

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

EXEMPLE 49. On reprend la variable aléatoire X de densité $f(x) = 2x\mathbb{1}_{0 \leq x \leq 1}$. On a vu que $\mathbb{E}[X] = \frac{2}{3}$ et nous avons

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^1 2x^3 dx = \left[\frac{x^4}{2} \right]_0^1 = \frac{1}{2}.$$

Finalement, cela nous donne

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

4.5 Exemples de variables aléatoires continues

4.5.1 Variable aléatoire uniforme

DÉFINITION 4.31

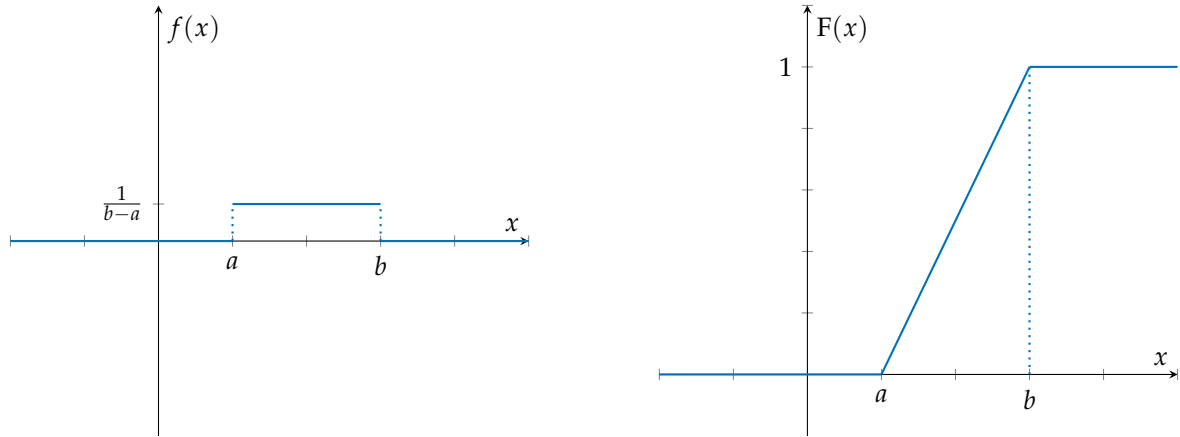
Une variable aléatoire est dite uniforme sur (a, b) si elle a pour densité

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

Elle a pour fonction de répartition

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

On appelle cette variable aléatoire uniforme car la densité est la même sur tout l'intervalle.


THÉORÈME 4.32

Soit $X \sim \text{Unif}((a, b))$ alors

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

Démonstration. On commence par calculer

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{2(b-a)} (b^2 - a^2) = \frac{a+b}{2}.$$

Pour la variance, on calcule

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{3(b-a)} (b^3 - a^3) = \frac{b^2 + ab + a^2}{3}.$$

Finalement, on obtient

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}.$$

□

EXEMPLE 50. Les bus passent toutes les 15 minutes à partir de 7 heures. Un usager arrive entre 7h et 7h30 selon une loi uniforme. Calculons la probabilité qu'il doive attendre moins de 5 minutes. Si $X \sim \text{Unif}((0, 30))$ correspondant au nombre de minutes entre 7h et l'heure d'arrivée de l'usager, alors pour attendre moins de 5 minutes nous devons avoir

$$\mathbb{P}(\{10 < X < 15\} \cup \{25 < X < 30\}) = \mathbb{P}(10 < X < 15) + \mathbb{P}(25 < X < 30) = \int_{10}^{15} \frac{dx}{30} + \int_{25}^{30} \frac{dx}{30} = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}.$$

Si on veut calculer la probabilité d'attendre plus de 10 minutes, alors on peut faire le même raisonnement mais cette fois-ci on calcule

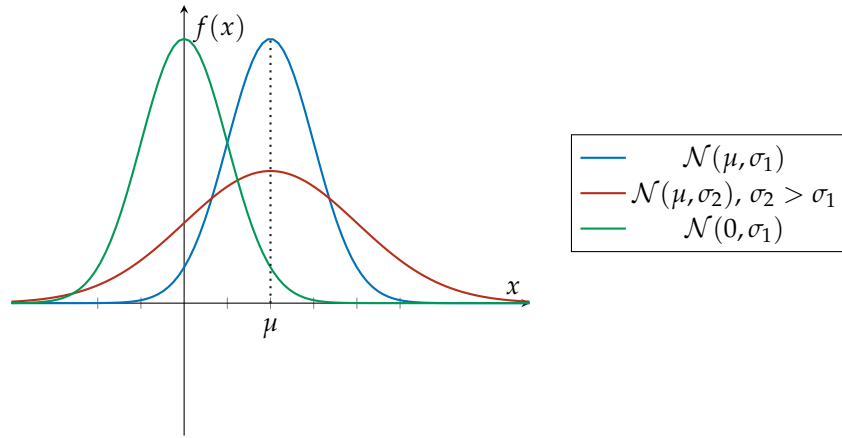
$$\mathbb{P}(\{0 < X < 5\} \cup \{15 < X < 20\}) = \frac{1}{3}.$$

4.5.2 Variable aléatoire gaussienne

DÉFINITION 4.33

Une variable aléatoire gaussienne, ou normale, de paramètre μ et σ^2 a pour densité

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$



Il n'est pas évident de voir que $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1$. On peut commencer par poser $u = \frac{x-\mu}{\sigma}$ alors $du = \frac{1}{\sigma} dx$ et ainsi nous voulons calculer

$$I = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du.$$

Ensuite, nous avons

$$I^2 = \frac{1}{2\pi} \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{u^2}{2}\right) du \right) \left(\int_{-\infty}^{+\infty} \exp\left(-\frac{v^2}{2}\right) dv \right) = \frac{1}{2\pi} \iint_{\mathbb{R}^2} \exp\left(-\frac{u^2+v^2}{2}\right) dudv.$$

On passe maintenant en coordonnées polaires avec $u = r \cos(\theta)$ et $v = r \sin(\theta)$ et $dudv = r dr d\theta$. Ainsi, l'intégrale devient

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{+\infty} \exp\left(-\frac{r^2}{2}\right) r dr d\theta = \frac{1}{2\pi} \int_0^{2\pi} \left[-\exp\left(-\frac{r^2}{2}\right) \right]_0^{+\infty} d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta = 1.$$

THÉORÈME 4.34

Si $X \sim \mathcal{N}(\mu, \sigma^2)$ et si $a, b \in \mathbb{R}$ alors

$$aX + b \sim \mathcal{N}(a\mu + b, \sigma^2 a^2).$$

Démonstration. Soit $Y = aX + b$ alors on a

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(aX + b \leq x) = \mathbb{P}\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right).$$

Ainsi, en dérivant on obtient

$$f_Y(x) = \frac{1}{a} f_X\left(\frac{x-b}{a}\right) = \frac{1}{\sqrt{2\pi}\sigma a} \exp\left(-\frac{\left(\frac{x-b}{a} - \mu\right)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma a} \exp\left(-\frac{(x - (a\mu + b))^2}{2\sigma^2 a^2}\right)$$

qui correspond bien à la densité d'une variable aléatoire gaussienne $\mathcal{N}(a\mu + b, \sigma^2 a^2)$. $\square$

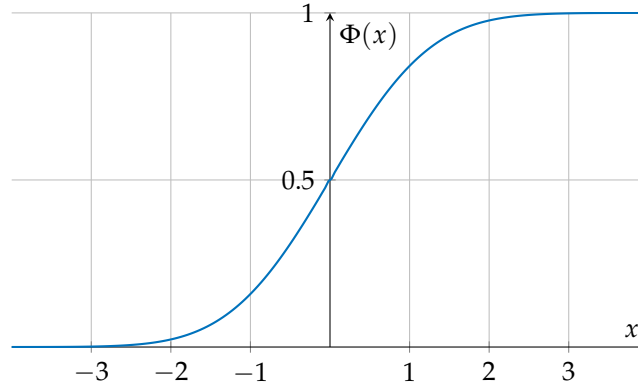
En particulier, cela montre que si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$ qui s'appelle aussi la loi normale centrée réduite. On remarque que l'on ne peut pas calculer explicitement la fonction de répartition d'une variable aléatoire gaussienne, on définit donc la fonction, si $X \sim \mathcal{N}(0, 1)$,

$$\Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \mathbb{P}(X \leq x).$$

Mais on peut obtenir toutes les fonctions de répartition de variables gaussiennes grâce à Φ puisque si $Y \sim$

$\mathcal{N}(\mu, \sigma^2)$,

$$\mathbb{P}(Y \leq x) = \mathbb{P}(\sigma X + \mu \leq x) = \mathbb{P}\left(X \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$



THÉORÈME 4.35

Soit $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $\mathbb{E}[X] = \mu$ et $\text{Var}(X) = \sigma^2$.

Démonstration. Considérons le cas où $Z \sim \mathcal{N}(0, 1)$. Alors on peut calculer

$$\mathbb{E}[Z] = \int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \left[-\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)\right]_{-\infty}^{+\infty} = 0.$$

Maintenant, nous avons donc par une intégration par parties,

$$\text{Var}(Z) = \mathbb{E}[Z^2] = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} dx = \left[-\frac{x e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0 + 1 = 1.$$

En utilisant la linéarité de l'espérance nous avons donc

$$\mathbb{E}[X] = \mathbb{E}[\sigma Z + \mu] = \sigma \mathbb{E}[Z] + \mu = \mu.$$

Et

$$\text{Var}(X) = \text{Var}(\sigma Z + \mu) = \sigma^2 \text{Var}(Z) = \sigma^2.$$

□

EXEMPLE 51. Exprimons en fonction de Φ , $\mathbb{P}(X > 0)$ et $\mathbb{P}(|X - 3| > 6)$ avec $X \sim \mathcal{N}(3, 9)$. On commence par

$$\mathbb{P}(X > 0) = \mathbb{P}\left(\frac{X - 3}{\sqrt{9}} > \frac{0 - 3}{\sqrt{9}}\right) = \mathbb{P}(Z > -1) = 1 - \mathbb{P}(Z \leq -1) = 1 - \Phi(-1) = \Phi(1).$$

On remarque que l'on a utilisé l'égalité $\Phi(-x) = 1 - \Phi(x)$ car avec $v = -u$,

$$\int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 1 - \int_{-x}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 1 - \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv.$$

Aussi nous pouvons écrire

$$\begin{aligned} \mathbb{P}(|X - 3| > 6) &= \mathbb{P}(X < -3) + \mathbb{P}(X > 9) = \mathbb{P}\left(\frac{X - 3}{3} < \frac{-3 - 3}{3}\right) + \mathbb{P}\left(\frac{X - 3}{3} > \frac{9 - 3}{3}\right) \\ &= \mathbb{P}(Z < -2) + \mathbb{P}(Z > 2) \\ &= \Phi(-2) + 1 - \Phi(2) = 2 - 2\Phi(2). \end{aligned}$$

EXEMPLE 52. La valeur au risque (VAR) est définie comme la valeur ν telle que la probabilité que la perte soit plus grande que ν est de $\frac{1}{100}$. Si on suppose que le gain de l'investissement est donnée par une variable aléatoire gaussienne $X \sim \mathcal{N}(\mu, \sigma^2)$ alors on cherche

$$\mathbb{P}(X < -\nu) = \frac{1}{100} \Leftrightarrow \mathbb{P}\left(\frac{X - \mu}{\sigma} < -\frac{\nu + \mu}{\sigma}\right) = \frac{1}{100} \Leftrightarrow \Phi\left(-\frac{\nu + \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\nu + \mu}{\sigma}\right) = \frac{1}{100}.$$

Ainsi, on obtient ν par l'équation

$$\Phi\left(\frac{\mu + \nu}{\sigma}\right) = \frac{99}{100} \Leftrightarrow \frac{\mu + \nu}{\sigma} \approx 2.33 \quad \text{i.e.} \quad \nu \approx 2.33\sigma - \mu.$$

L'origine de la loi gaussienne vient de cette observation d'Abraham de Moivre qui considère la limite d'une loi binomiale.

THÉORÈME 4.36

Soit $X \sim \text{Bin}(n, p)$ alors

$$\mathbb{P}\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) = \int_a^b \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \mathbb{P}(a \leq \mathcal{N}(0, 1) \leq b)$$

C'est un exemple du théorème central limite que l'on verra de manière plus générale à la fin du cours.

4.5.3 Variable aléatoire exponentielle

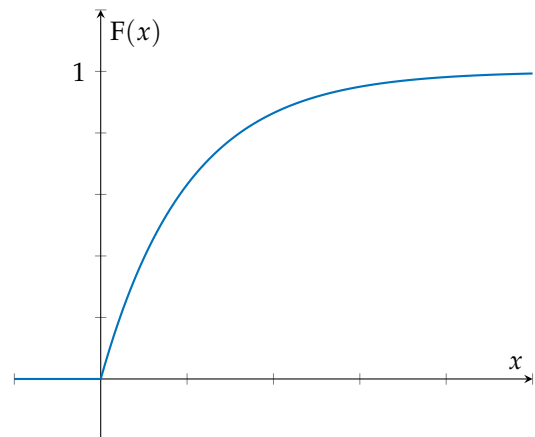
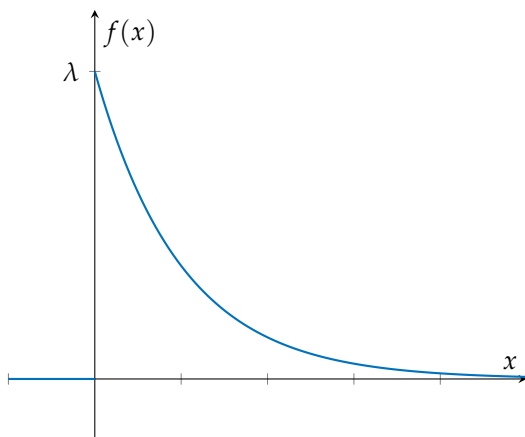
DÉFINITION 4.37

Une variable aléatoire est dite exponentielle de paramètre $\lambda > 0$ si elle est à densité

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

Elle a pour fonction de répartition

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



On remarque que l'on a bien

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^{+\infty} \lambda \exp^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_0^{+\infty} = 1.$$

THÉORÈME 4.38

 Soit $X \sim \text{Exp}(\lambda)$ alors

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

Démonstration. On calcule par une intégration par parties

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx = \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \left[-x e^{-\lambda x} \right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = 0 + \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^{+\infty} = \frac{1}{\lambda}.$$

Pour la variance, on calcule en faisant deux intégrations par parties,

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{\mathbb{R}} x^2 f(x) dx = \int_0^{+\infty} \lambda x^2 e^{-\lambda x} dx = \left[-x^2 e^{-\lambda x} \right]_0^{+\infty} + \int_0^{+\infty} 2x e^{-\lambda x} dx \\ &= 0 + \left[-\frac{2x}{\lambda} e^{-\lambda x} \right]_0^{+\infty} + \int_0^{+\infty} \frac{2}{\lambda} e^{-\lambda x} dx \\ &= \frac{2}{\lambda^2}. \end{aligned}$$

Ainsi

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

□

Les variables exponentielles modélisent souvent des temps entre des événements.

EXEMPLE 53. La durée (en minutes) d'une conversation au téléphone X est modélisée par une variable exponentielle de paramètre $\lambda = \frac{1}{10}$. Lorsque que l'on arrive à la cabine téléphonique, quelqu'un vient de rentrer! Calculons la probabilité d'attendre plus de 10 minutes. On cherche donc

$$\mathbb{P}(X > 10) = 1 - \mathbb{P}(X \leq 10) = 1 - \left(1 - e^{-\frac{10}{10}}\right) = \frac{1}{e} \approx 0.368.$$

Si on veut la probabilité d'attendre entre 10 et 20 minutes alors on cherche

$$\mathbb{P}(10 < X < 20) = \mathbb{P}(X \leq 20) - \mathbb{P}(X \leq 10) = 1 - e^{-\frac{20}{10}} - \left(1 - e^{-\frac{10}{10}}\right) = \frac{1}{e} - \frac{1}{e^2} \approx 0.233$$

On a la même propriété d'absence de mémoire que pour la variable géométrique,

THÉORÈME 4.39

 Si $X \sim \text{Exp}(\lambda)$ alors pour tout $s, t \geq 0$,

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

Démonstration. Par la définition des probabilités conditionnelles, nous avons que

$$\mathbb{P}(X > s + t | X > t) = \frac{\mathbb{P}(X > s + t, X > t)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(X > s).$$

□

EXEMPLE 54. Une batterie de voiture dure en moyenne 10 000 km. On modélise sa durée de vie X par une variable aléatoire exponentielle. Un étudiant souhaite entreprendre un long voyage en voiture de Montréal jusqu'à Vancouver, soit environ 5 000 km. Quelle est la probabilité que la batterie ne tombe pas en panne avant la fin du voyage?

L'avantage de la loi exponentielle c'est que l'on a pas besoin de savoir combien de km on a déjà parcourue

avec la voiture car si cette distance est d alors

$$\mathbb{P}(X > d + 5000 | X > d) = \mathbb{P}(X > 5000) = 1 - e^{-\frac{5000}{10000}} = 1 - \frac{1}{\sqrt{e}} \approx 0.393.$$

Si on avait utilisé une autre loi que la variable exponentielle alors on a

$$\mathbb{P}(X > 5000 + d | X > d) = \frac{\mathbb{P}(X > 5000 + d)}{\mathbb{P}(X > d)} = \frac{1 - F(d + 5000)}{1 - F(d)}$$

mais cette valeur dépend de d dans le cas général.

4.5.4 Variable aléatoire Gamma

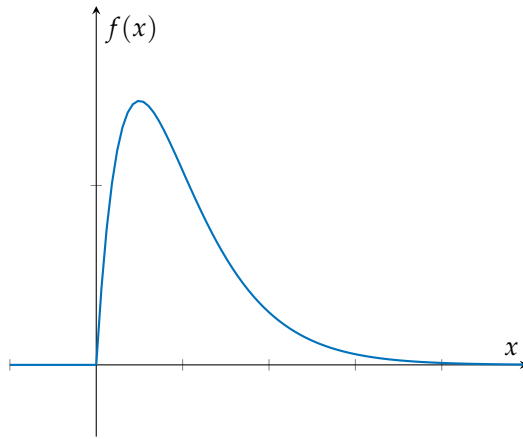
DÉFINITION 4.40

Une variable aléatoire suit une loi Gamma de paramètre (s, λ) avec $\lambda, s > 0$ si elle a pour densité

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

avec

$$\Gamma(s) = \int_0^{\infty} e^{-y} y^{s-1} dy.$$



La fonction Γ a été introduite par Leonhard Euler dans le but de généraliser la factorielle. En effet, on a

$$\Gamma(s) = \int_0^{+\infty} e^{-y} y^{s-1} dy = \left[-y^{s-1} e^{-y} \right]_0^{+\infty} + \int_0^{+\infty} e^{-y} (s-1) y^{s-2} dy = (s-1) \Gamma(s-2).$$

Ainsi, pour $n \in \mathbb{N}$, on a

$$\Gamma(n) = (n-1) \Gamma(n-1) = (n-1)(n-2) \Gamma(n-2) = \cdots = (n-1)(n-2) \cdots 2 \cdot 1 \cdot \Gamma(1).$$

Or $\Gamma(1) = \int_0^{+\infty} e^{-y} dy = 1$ et ainsi

$$\Gamma(n) = (n-1)!.$$

Pour $s \in \mathbb{N}$, la loi Gamma représente souvent la n -ième arrivée d'un événement.

EXEMPLE 55. Supposons qu'entre 0 et t , le nombre d'arrivée d'événements $N(t)$ est une variable aléatoire de Poisson de paramètre λt . Calculons la densité du temps d'arrivée T_n du n -ième événement.

$$\mathbb{P}(T_n \leq t) = \mathbb{P}(N(t) \geq n) = \sum_{j=n}^{+\infty} \mathbb{P}(N(t) = j) = \sum_{j=n}^{+\infty} \frac{e^{-\lambda t} (\lambda t)^j}{j!}.$$

On retrouve la densité en dérivant la fonction de répartition

$$\begin{aligned} \frac{d}{dt} \mathbb{P}(T_n \leq t) &= \sum_{j=n}^{+\infty} \frac{-\lambda e^{-\lambda t} (\lambda t)^j}{j!} + \sum_{j=n}^{+\infty} \frac{\lambda e^{-\lambda t} j (\lambda t)^{j-1}}{j!} = \sum_{j=n}^{+\infty} \frac{-\lambda e^{-\lambda t} (\lambda t)^j}{j!} + \sum_{j=n}^{+\infty} \frac{\lambda e^{-\lambda t} (\lambda t)^{j-1}}{(j-1)!} \\ &= - \sum_{j=n}^{+\infty} \frac{\lambda e^{-\lambda t} (\lambda t)^j}{j!} + \sum_{j=n-1}^{+\infty} \frac{\lambda e^{-\lambda t} (\lambda t)^j}{j!} \\ &= \frac{\lambda e^{-\lambda t} (\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

qui correspond bien à la densité d'une loi Gamma de paramètre (n, λ) .

THÉORÈME 4.41

Soit $X \sim \text{Gamma}(s, \lambda)$ alors

$$\mathbb{E}[X] = \frac{s}{\lambda}, \quad \text{Var}(X) = \frac{s}{\lambda^2}.$$

Démonstration. On a

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx = \int_0^{+\infty} \frac{x \lambda e^{-\lambda x} (\lambda x)^{s-1}}{\Gamma(s)} dx = \frac{1}{\Gamma(s)} \int_0^{+\infty} \frac{\lambda e^{-\lambda x} (\lambda x)^s}{\lambda} dx = \frac{1}{\lambda \Gamma(s)} \Gamma(s+1) = \frac{s}{\lambda}.$$

De même,

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f(x) dx = \frac{1}{\Gamma(s)} \int_0^{+\infty} \frac{\lambda e^{-\lambda x} (\lambda x)^{s+1}}{\lambda^2} dx = \frac{1}{\lambda^2 \Gamma(s)} \Gamma(s+2) = \frac{s(s+1)}{\lambda^2}.$$

Ainsi

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{s(s+1)}{\lambda^2} - \frac{s^2}{\lambda^2} = \frac{s}{\lambda^2}.$$

□

4.6 Fonction d'une variable continue

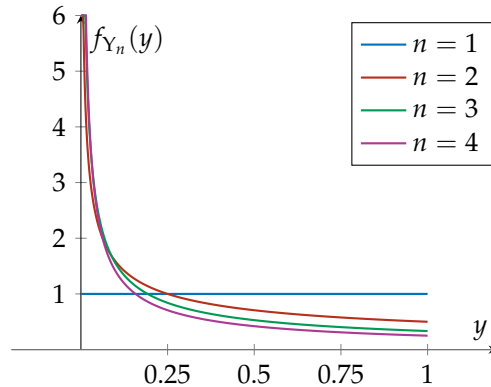
Souvent, on considère $g(X)$ avec X une variable aléatoire continue.

EXEMPLE 56. Soit $X \sim \text{Unif}((0, 1))$. On considère $Y_n = X^n$. Calculons la densité de Y_n . On remarque que les valeurs de Y_n sont aussi entre 0 et 1 et on a pour $0 < y < 1$, on a

$$F_{Y_n}(y) = \mathbb{P}(Y_n \leq y) = \mathbb{P}(X^n \leq y) = \mathbb{P}\left(X \leq y^{\frac{1}{n}}\right) = y^{\frac{1}{n}}.$$

On obtient donc la densité de Y_n en dérivant et on obtient

$$f_{Y_n}(y) = \frac{1}{n} y^{\frac{1}{n}-1} \mathbb{1}_{0 < y < 1}.$$



EXEMPLE 57. Si X est une variable aléatoire de densité f_X , donnons la densité de $Y = X^2$. On remarque que $Y \geq 0$ et ainsi si $y > 0$, on a

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

En dérivant, on obtient donc

$$f_Y(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) \mathbb{1}_{y>0}.$$

Même question pour $Z = |X|$ alors on a pour $z > 0$,

$$\mathbb{P}(Z \leq z) = \mathbb{P}(|X| \leq z) = \mathbb{P}(-z \leq X \leq z) = F_X(z) - F_X(-z).$$

Ainsi, on obtient la densité

$$f_Z(z) = (f_X(z) + f_X(-z)) \mathbb{1}_{z>0}.$$

THÉORÈME 4.42

Soit X une variable aléatoire continue de densité f_X . Soit g une fonction strictement monotone (strictement croissante ou décroissante) et dérivable. La densité de la variable aléatoire $Y = g(X)$ est donnée par

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & \text{si } y = g(x), \\ 0 & \text{sinon.} \end{cases}$$

Démonstration. Supposons que $y = g(x)$ pour un certain x et que g est croissante de telle sorte que g^{-1} est aussi croissante. Ainsi

$$\mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

En dérivant, on obtient la densité

$$f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dy} (g^{-1}(y)) = f_X(g^{-1}(y)) \left| \frac{d}{dy} (g^{-1}(y)) \right|$$

car g^{-1} est croissante et sa dérivée est donc positive. Dans le cas où g est décroissante alors le même calcul donne

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

Et ainsi,

$$f_Y(y) = -f_X(g^{-1}(y)) \frac{d}{dy} (g^{-1}(y)) = f_X(g^{-1}(y)) \left| \frac{d}{dy} (g^{-1}(y)) \right|$$

car g^{-1} est décroissante et sa dérivée est donc négative. S'il n'existe pas de x telle que $y = g(x)$ alors comme g est monotone soit $\mathbb{P}(Y \leq y) = 0$ ou 1 et en dérivant on obtient bien que la densité est nulle. $\square$

EXEMPLE 58. Soit X une variable aléatoire positive de densité f et $Y_n = X^n$. Alors on a $g(x) = x^n$ qui est strictement monotone et dérivable pour $x > 0$. On a de plus que $g^{-1}(y) = y^{\frac{1}{n}}$ et ainsi $\frac{d}{dy} g^{-1}(y) = \frac{1}{n} y^{\frac{1}{n}-1}$. Finalement, on obtient

$$f_{Y_n}(y) = \frac{1}{n} y^{\frac{1}{n}-1} f_X(y^{\frac{1}{n}}) \mathbb{1}_{y>0}.$$

EXEMPLE 59. Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $Y = e^X$ est appelée une loi log-normale. Calculons sa densité, nous avons $g(x) = e^x$ qui est strictement croissante et dérivable, pour $y > 0$ on a $g^{-1}(y) = \ln(y)$ et ainsi $\frac{d}{dy} g^{-1}(y) = \frac{1}{y}$ ce qui nous donne que

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\ln(y) - \mu)^2}{2\sigma^2}\right) \mathbb{1}_{y>0}.$$

5. VARIABLES ALÉATOIRES SIMULTANÉES

Pour l'instant, nous avons considéré des variables aléatoires isolées. Nous n'avons pas comparé ou additionner plusieurs variables aléatoires.

DÉFINITION 5.1

Pour toute paire de variables aléatoire (X, Y) , on définit leur fonction de répartition jointe, conjointe ou simultanée par

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y), \quad x, y \in \mathbb{R}.$$

On retrouve les fonctions de répartition, dites marginales, de X et de Y par

$$F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y) \quad \text{et} \quad F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y).$$

On peut écrire tous les événements en terme de la fonction de répartition jointe et des fonctions de répartition marginales. Par exemple

$$\begin{aligned} \mathbb{P}(X > x, Y > y) &= \mathbb{P}(\{X > x\} \cap \{Y > y\}) = 1 - \mathbb{P}((\{X > x\} \cap \{Y > y\})^c) \\ &= 1 - \mathbb{P}(\{X \leq x\} \cup \{Y \leq y\}) \\ &= 1 - \mathbb{P}(X \leq x) - \mathbb{P}(Y \leq y) + \mathbb{P}(X \leq x, Y \leq y) \\ &= 1 - F_X(x) - F_Y(y) + F_{X,Y}(x, y). \end{aligned}$$

Pour les lois discrètes, on considère leur loi de probabilité jointe.

DÉFINITION 5.2

Si X et Y ont deux variables aléatoires discrètes, leur loi de probabilité jointe est donnée par

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y).$$

On retrouve les lois de probabilité, dites marginales, de X et de Y par

$$p_X(x) = \sum_{y, p(x,y) > 0} p_{X,Y}(x, y) \quad \text{et} \quad p_Y(y) = \sum_{x, p(x,y) > 0} p_{X,Y}(x, y).$$

EXEMPLE 60. On tire 3 boules d'une urne contenant 3 boules rouges, 4 boules blanches et 5 boules bleues. On définit X le nombre de boules rouges tirées et Y le nombre de boules blanches tirées. On voit aussi que les valeurs possibles de X et de Y sont $\{0, 1, 2, 3\}$. On peut calculer la loi jointe directement. Par exemple,

$$\begin{aligned} \mathbb{P}(X = 0, Y = 0) &= \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{1}{22}, \quad \mathbb{P}(X = 1, Y = 0) = \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{3}{22} \\ \mathbb{P}(X = 2, Y = 0) &= \frac{\binom{3}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{3}{44}, \quad \mathbb{P}(X = 3, Y = 0) = \frac{\binom{3}{3}}{\binom{12}{3}} = \frac{1}{220}. \end{aligned}$$

On peut calculer de la même façon toute la loi de probabilité. On remarque aussi que comme on tire que 3 boules si $i + j > 3$ alors $\mathbb{P}(X = i, Y = j) = 0$. Pour retrouver la loi marginale de Y par exemple, on somme sur les valeurs de X . Par exemple dans ce cas, on obtient

$$\mathbb{P}(Y = 0) = \sum_{x, p(x,0) > 0} \mathbb{P}(X = x, Y = 0) = \frac{1}{22} + \frac{3}{22} + \frac{3}{44} + \frac{1}{220} = \frac{14}{55}.$$

On écrit cela sous forme d'un tableau de la façon suivante

$Y \backslash X$	0	1	2	3	$P(Y = y)$
0	$\frac{1}{22}$	$\frac{3}{22}$	$\frac{3}{44}$	$\frac{1}{220}$	$\frac{14}{55}$
1	$\frac{2}{11}$	$\frac{3}{11}$	$\frac{3}{55}$	0	$\frac{28}{55}$
2	$\frac{3}{22}$	$\frac{9}{110}$	0	0	$\frac{12}{55}$
3	$\frac{1}{55}$	0	0	0	$\frac{1}{55}$
$P(X = x)$	$\frac{21}{55}$	$\frac{27}{55}$	$\frac{27}{220}$	$\frac{1}{220}$	

TABLE 1 – Loi de probabilité jointe de (X, Y) avec les lois marginales.**DÉFINITION 5.3**

X et Y sont deux variables continues conjointement si il existe une fonction $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ positive, appelée densité jointe de X et de Y , telle que pour tout A événement de $\mathbb{R}^2$

$$P((X, Y) \in A) = \iint_A f_{X,Y}(x, y) dx dy.$$

On peut retrouver la densité par la fonction de répartition avec

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = f_{X,Y}(x, y)$$

lorsque la dérivée existe.

On peut aussi retrouver la densité marginale de X et de Y en fonction de la densité jointe en l'intégrant.

THÉORÈME 5.4

Si X et Y sont deux variables aléatoires continues conjointement alors X et Y sont des variables continues et leur densité marginale est donnée par

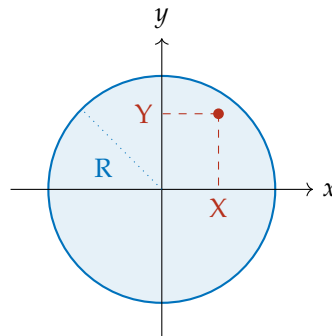
$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy \quad \text{et} \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx.$$

Démonstration. On a

$$P(X \in A) = P(X \in A, Y \in \mathbb{R}) = \int_{\mathbb{R}} \int_A f_{X,Y}(x, y) dx dy = \int_A \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy \right) dx$$

et on retrouve la définition d'une densité en posant $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy$. De même pour la densité marginale de Y . $\square$

EXEMPLE 61. On lance un point uniformément sur le disque de centre 0 et de rayon R . On dénote X son abscisse et Y son ordonnée.



Comme le point est lancé uniformément sur le disque, la densité jointe est constante sur le disque de telle sorte que l'on peut écrire

$$f_{X,Y}(x,y) = \alpha \mathbb{1}_{x^2+y^2 \leq R^2}.$$

Pour trouver α , on rappelle que l'on a une densité qui s'intègre à 1 et ainsi

$$1 = \iint_{\mathbb{R}^2} f_{X,Y}(x,y) dx dy = \iint_{x^2+y^2 \leq R^2} f_{X,Y}(x,y) dx dy = \alpha \text{Aire}(\text{Disque de rayon } R) = \alpha \pi R^2.$$

et ainsi $\alpha = \frac{1}{\pi R^2}$. On peut ainsi retrouver les lois marginales de X et de Y et on a pour $|x| \leq R$,

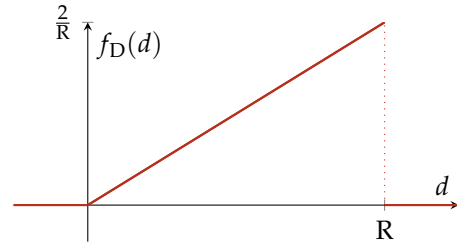
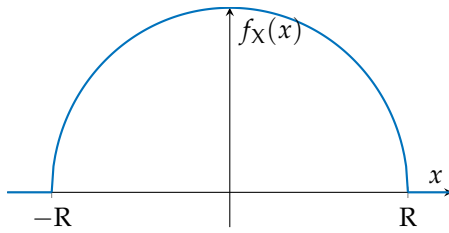
$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y) dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} f_{X,Y}(x,y) dy = \frac{2}{\pi R^2} \sqrt{R^2 - x^2} \mathbb{1}_{-R \leq x \leq R}.$$

De même pour $f_Y(y)$. Si on considère D la distance à l'origine du point alors on a pour $0 \leq d \leq R$,

$$\mathbb{P}(D \leq d) = \mathbb{P}(X^2 + Y^2 \leq d^2) = \mathbb{P}(X^2 + Y^2 \leq R^2) = \iint_{x^2+y^2 \leq d^2} f_{X,Y}(x,y) dx dy = \frac{d^2}{R^2}.$$

On obtient donc la densité en dérivant et ainsi

$$f_D(d) = \frac{2d}{R^2} \mathbb{1}_{0 \leq d \leq R}.$$



On peut alors calculer l'espérance de la distance à l'origine,

$$\mathbb{E}[D] = \int_0^R \frac{2d^2}{R^2} dd = \frac{2}{R^2} \left[\frac{d^3}{3} \right]_0^R = \frac{2R^3}{3R^2} = \frac{2R}{3}.$$

On peut généraliser ces définitions à plus de deux variables aléatoires.

DÉFINITION 5.5

Soient $X_1, \dots, X_n$ n variables aléatoires définies sur le même ensemble fondamental. On définit la fonction de répartition jointe de $X_1, \dots, X_n$ par $F : \mathbb{R}^n \rightarrow [0, 1]$

$$F(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n).$$

On dit qu'elles sont conjointement continues si il existe une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$ telle que pour tout événement A de $\mathbb{R}^n$,

$$\mathbb{P}((X_1, \dots, X_n) \in A) = \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Si elles sont discrètes, on définit la loi jointe de probabilité

$$p(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

5.1 Variables aléatoires indépendantes

DÉFINITION 5.6

Deux variables aléatoires X et Y sont dites indépendantes si pour tout $x, y \in \mathbb{R}$,

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \mathbb{P}(Y \leq y) = F_X(x)F_Y(y).$$

Si X et Y sont discrètes alors elles sont indépendantes si et seulement si pour tout $x, y \in \mathbb{R}$,

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

Si X et Y sont jointement continues alors elles sont indépendantes si et seulement si pour tout $x, y \in \mathbb{R}$,

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

EXEMPLE 62. On réalise $n + m$ épreuves indépendantes qui ont une probabilité p de succès. On définit X le nombre de succès lors des n premières épreuves et Y le nombre de succès dans les m dernières épreuves. On a alors

$$\mathbb{P}(X = x, Y = y) = \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y} = \mathbb{P}(X = x) \mathbb{P}(Y = y).$$

Ainsi X et Y sont indépendantes. Si maintenant, on considère Z le nombre de succès total dans les $n + m$ épreuves alors on a

$$\mathbb{P}(X = 1, Z = 0) = 0$$

car on ne peut pas avoir un succès dans les n premières épreuves et 0 succès en total. Mais on a

$$\mathbb{P}(X = 1) = np(1-p)^{n-1} \quad \text{et} \quad \mathbb{P}(Z = 0) = (1-p)^{n+m}.$$

Ainsi, on a

$$\mathbb{P}(X = 1, Z = 0) \neq \mathbb{P}(X = 1) \mathbb{P}(Z = 0)$$

et X et Z ne sont pas indépendantes.

EXEMPLE 63. Le nombre de clients d'un magasin en un jour est modélisé par une variable aléatoire de Poisson de paramètre λ . On note p la probabilité qu'un client soit un homme. Montrons que le nombre de clients masculins X est donnée par une loi de Poisson de paramètre λp et le nombre de clients féminins Y suit une loi de Poisson de paramètre $\lambda(1-p)$ et que X et Y sont indépendantes.

Nous allons calculer $\mathbb{P}(X = i, Y = j)$ mais nous ne connaissons que la loi du nombre total de clients $X + Y$. Nous allons donc conditionner sur la valeur de $X + Y$ et nous pouvons écrire

$$\begin{aligned} \mathbb{P}(X = i, Y = j) \\ = \mathbb{P}(X = i, Y = j | X + Y = i + j) \mathbb{P}(X + Y = i + j) + \mathbb{P}(X = i, Y = j | X + Y \neq i + j) \mathbb{P}(X + Y \neq i + j) \end{aligned}$$

par la loi des probabilités totales et bien sur $\mathbb{P}(X = i, Y = j | X + Y \neq i + j) = 0$ de telle sorte que

$$\mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i, Y = j | X + Y = i + j) \mathbb{P}(X + Y = i + j).$$

Nous pouvons calculer

$$\mathbb{P}(X = i, Y = j | X + Y = i + j) = \binom{i+j}{i} p^i (1-p)^j \quad \text{et} \quad \mathbb{P}(X + Y = i + j) = e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}.$$

Ainsi, en combinant ces deux calculs, on peut écrire

$$\mathbb{P}(X = i, Y = j) = \frac{(i+j)!}{i!j!} (\lambda p)^i (\lambda(1-p))^j \frac{e^{-\lambda(p+(1-p))}}{(i+j)!} = e^{-\lambda p} \frac{(\lambda p)^i}{i!} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}.$$

Ainsi, si on somme sur les valeurs de Y , comme on sait par la loi de Poisson que

$$\sum_{j=0}^{+\infty} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} = 1$$

on obtient

$$\mathbb{P}(X = i) = \sum_{j=0}^{+\infty} \mathbb{P}(X = i, Y = j) = e^{-\lambda p} \frac{(\lambda p)^i}{i!}$$

et de même

$$\mathbb{P}(Y = j) = \sum_{i=0}^{+\infty} \mathbb{P}(X = i, Y = j) = e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}.$$

Ainsi, on voit bien que

$$\mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i)\mathbb{P}(Y = j)$$

et donc $X \sim \text{Poi}(\lambda p)$, $Y \sim \text{Poi}(\lambda(1-p))$ et X et Y sont indépendantes.

EXEMPLE 64. Deux femmes se donnent rendez-vous. Elles arrivent selon une loi uniforme entre 12h et 13h de manière indépendantes. Quelle est la probabilité que la première arrivée attende plus de 10 minutes?

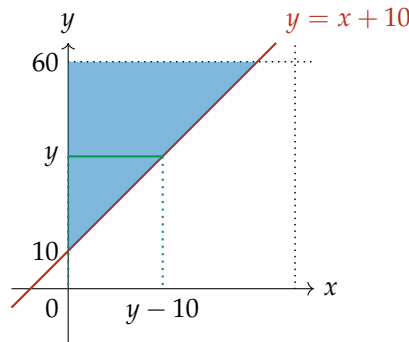
Soit X le temps d'arrivée d'une des deux femmes et Y le temps d'arrivée de l'autre. Alors on a que $X, Y \sim \text{Unif}(0, 60)$ et que X et Y sont indépendantes. On cherche donc

$$\mathbb{P}(X + 10 < Y) + \mathbb{P}(Y + 10 < X) = 2\mathbb{P}(X + 10 < Y)$$

car X et Y ont la même loi. On peut maintenant calculer

$$\begin{aligned} 2\mathbb{P}(X + 10 < Y) &= 2 \iint_{x+10 < y} f_{X,Y}(x, y) dx dy = 2 \iint_{x+10 < y} f_X(x) f_Y(y) dx dy \\ &= 2 \iint_{x+10 < y} \frac{1}{60^2} \mathbb{1}_{0 \leq x \leq 60} \mathbb{1}_{0 \leq y \leq 60} dx dy. \end{aligned}$$

Pour calculer cette double intégrale, on peut dessiner le domaine d'intégration donnée par les conditions $0 \leq x \leq 60$, $0 \leq y \leq 60$ et $x + 10 < y$,



Cela nous donne ainsi,

$$2\mathbb{P}(X + 10 < Y) = \frac{1}{1800} \int_{10}^{60} \left(\int_0^{y-10} dx \right) dy = \frac{1}{1800} \int_{10}^{60} (y - 10) dy = \frac{1}{1800} \left[\frac{(y - 10)^2}{2} \right]_{10}^{60} = \frac{25}{36}.$$

THÉORÈME 5.7

Les variables aléatoires continues X et Y sont indépendantes si et seulement si la densité conjointe de X et de Y peut s'écrire pour tout $x, y \in \mathbb{R}$,

$$f_{X,Y}(x, y) = h(x)g(y).$$

De même, les variables aléatoires discrètes X et Y sont indépendantes si et seulement si la loi jointe de probabilité de X et de Y peut s'écrire pour tout $x, y \in \mathbb{R}$,

$$p_{X,Y}(x, y) = h(x)g(y)$$

Démonstration. Nous faisons la preuve dans le cas continu. Si X, Y sont indépendantes alors on sait que $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. On suppose maintenant que l'on a $f_{X,Y}(x, y) = h(x)g(y)$ alors on a

$$1 = \iint_{\mathbb{R}^2} f_{X,Y}(x, y) dx dy = \iint_{\mathbb{R}^2} h(x)g(y) dx dy = \left(\int_{\mathbb{R}} g(y) dy \right) \left(\int_{\mathbb{R}} h(x) dx \right) = C_1 C_2$$

avec $C_1 = \int_{\mathbb{R}} g(y) dy$ et $C_2 = \int_{\mathbb{R}} h(x) dx$. Maintenant, on sait que

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy = \int_{\mathbb{R}} g(y)h(x) dy = C_1 h(x)$$

et de même $f_Y(y) = C_2 g(y)$. Ainsi, on obtient

$$f_{X,Y}(x, y) = h(x)g(y) = h(x)g(y)C_1 C_2 = f_X(x)f_Y(y)$$

et les variables aléatoires X et Y sont indépendantes. $\square$

EXEMPLE 65. On considère la densité jointe

$$f_{X,Y}(x, y) = 6e^{-2x}e^{-3y}\mathbb{1}_{x>0}\mathbb{1}_{y>0}.$$

Alors on voit que l'on peut factoriser la densité comme un produit d'une fonction de x et d'une fonction de y et ainsi X et Y sont indépendantes. De plus, on peut prendre $h(x) = 2e^{-2x}$ et $g(y) = 3e^{-3y}$ et on voit que $X \sim \text{Exp}(2)$ et $Y \sim \text{Exp}(3)$.

Si maintenant on considère la densité jointe

$$f_{X,Y}(x, y) = 24xy\mathbb{1}_{0<x<1}\mathbb{1}_{0<y<1}\mathbb{1}_{0<x+y<1},$$

on voit que même si $24xy$ peut se factoriser en une fonction de x et de y , la condition $\mathbb{1}_{0<x+y<1}$ empêche une factorisation en $h(x)g(y)$ et les variables aléatoires X et Y ne sont pas indépendantes. En particulier, on voit que le domaine de valeurs doit être un rectangle alors qu'ici, c'est un triangle.

DÉFINITION 5.8

Soient $X_1, \dots, X_n$ n variables aléatoires, elles sont dites indépendantes si pour tout $x_1, \dots, x_n$,

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i).$$

Si on en a une infinité, on définit l'indépendance en prenant tout sous-ensemble fini de ces variables aléatoires.

EXEMPLE 66. Soient X, Y, Z trois variables aléatoires indépendantes et de même loi $\text{Unif}(0, 1)$. Calculons $\mathbb{P}(X \geq YZ)$. On a

$$f_{X,Y,Z}(x, y, z) = f_X(x)f_Y(y)f_Z(z) = \mathbb{1}_{0<x<1}\mathbb{1}_{0<y<1}\mathbb{1}_{0<z<1}$$

et ainsi on peut calculer

$$\begin{aligned} \mathbb{P}(X \geq YZ) &= \iiint_{x \geq yz} f_{X,Y,Z}(x, y, z) dx dy dz = \int_0^1 \int_0^1 \int_{yz}^1 dx dy dz = \int_0^1 \int_0^1 (1 - yz) dy dz = \int_0^1 \left(1 - \frac{z}{2}\right) dz \\ &= \frac{3}{4}. \end{aligned}$$

5.2 Somme de variables aléatoires indépendantes

Supposons que l'on connaisse la loi de X et de Y qui sont par exemple de densité f_X et f_Y et que X et Y soient indépendantes. Alors si on cherche

$$\begin{aligned} F_{X+Y}(z) &= \mathbb{P}(X + Y \leq z) = \iint_{x+y \leq z} f_X(x) f_Y(y) dx dy = \int_{\mathbb{R}} \int_{-\infty}^{z-y} f_X(x) f_Y(y) dx dy \\ &= \int_{\mathbb{R}} f_Y(y) \left(\int_{-\infty}^{z-y} f_X(x) dx \right) dy \\ &= \int_{\mathbb{R}} F_X(z-y) f_Y(y) dy. \end{aligned}$$

On définit le *produit de convolution* de la façon suivante,

$$(F_X * f_Y)(z) = \int_{\mathbb{R}} F_X(z-y) f_Y(y) dy.$$

Si on veut obtenir la densité de $X + Y$, on dérive

$$f_{X+Y}(z) = \frac{d}{dz} F_{X+Y}(z) = \int_{\mathbb{R}} \frac{d}{dz} F_X(z-y) f_Y(y) dy = \int_{\mathbb{R}} f_X(z-y) f_Y(y) dy = (f_X * f_Y)(z).$$

DÉFINITION 5.9

Soient X et Y deux variables aléatoires continues indépendantes alors la fonction de répartition de $X + Y$

$$F_{X+Y}(z) = \mathbb{P}(X + Y \leq z) = (F_X * f_Y)(z) := \int_{-\infty}^{\infty} F_X(z-y) f_Y(y) dy$$

et sa densité est donnée par

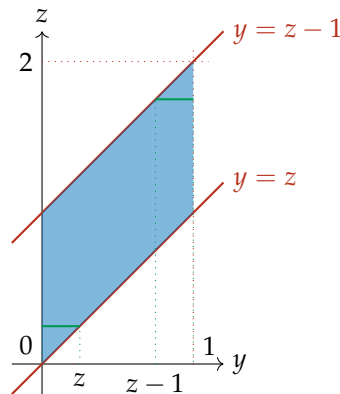
$$f_{X+Y}(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy.$$

$f * g$ est appelé le produit de convolution de f et g .

EXEMPLE 67. Soit $X, Y \sim \text{Unif}(0, 1)$ et X et Y sont indépendantes. Calculons la densité de $X + Y$. On remarque déjà que les valeurs possibles de $X + Y$ sont entre 0 et 2, ensuite pour $z \in [0, 2]$,

$$f_{X+Y}(z) = (f_X * f_Y)(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y) dy = \int_0^1 f_X(z-y) dy = \int_0^1 \mathbb{1}_{0 < z-y < 1} dy = \int_0^1 \mathbb{1}_{y < z} \mathbb{1}_{y > z-1} dy.$$

On peut dessiner le domaine délimité par les conditions, $0 < y < 1$, $0 < z < 2$, $y < z$ et $y > z - 1$,

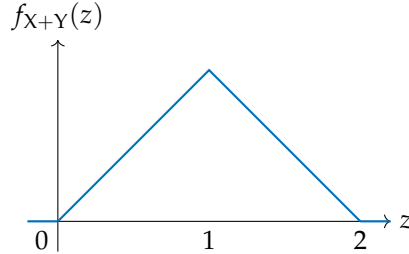


On voit donc que l'on doit diviser l'intégration en deux cas, si $z \in [0, 1]$ alors

$$f_{X+Y}(z) = \int_0^z dy = z$$

et si $z \in [1, 2]$ alors

$$f_{X+Y}(z) = \int_{z-1}^1 dy = 2 - z.$$



EXEMPLE 68. En moyenne, combien faut-il de variables aléatoires uniformes sur $[0, 1]$ indépendantes pour que leur somme dépasse 1. On définit donc si $X_1, X_2, \dots$ est une famille de variables aléatoires $\text{Unif}(0, 1)$ indépendantes,

$$N = \min\{n \in \mathbb{N}, X_1 + \dots + X_n > 1\}.$$

Calculons la loi de N , on remarque tout d'abord que

$$\mathbb{P}(N > n) = \mathbb{P}(X_1 + \dots + X_n \leq 1) = F_{X_1 + \dots + X_n}(1) =: F_n(1).$$

La loi d'une somme de variables aléatoires uniformes est compliquée mais nous sommes intéressés qu'à la valeur de la fonction de répartition en 1. En fait, on peut montrer par récurrence que pour $0 \leq x \leq 1$, on a $F_n(x) = \frac{x^n}{n!}$. En effet, on a déjà vu que $F_1(x) = x$ pour $x \in [0, 1]$ ci-dessus. Maintenant si $F_n(x) = \frac{x^n}{n!}$ pour un certain n donné alors on a

$$\begin{aligned} F_{n+1}(x) &= (F_n(x) * f_{X_{n+1}})(x) = \int_{\mathbb{R}} F_n(x-y) f_{X_{n+1}}(y) dy = \int_0^1 F_n(x-y) dy = \int_0^1 \frac{(x-y)^n}{n!} \mathbb{1}_{0 < x-y < 1} dy \\ &= \int_0^x \frac{(x-y)^n}{n!} dx \\ &= \frac{x^{n+1}}{(n+1)!}. \end{aligned}$$

Ainsi, on obtient en utilisant $x = 1$ que

$$\mathbb{P}(N > n) = \frac{1}{n!}.$$

On peut maintenant obtenir la loi de probabilité en utilisant

$$\mathbb{P}(N = n) = \mathbb{P}(N > n-1) - \mathbb{P}(N > n) = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{n-1}{n!}.$$

On peut ainsi calculer l'espérance

$$\mathbb{E}[N] = \sum_{n=2}^{+\infty} \frac{n(n-1)}{n!} = \sum_{n=2}^{+\infty} \frac{1}{(n-2)!} = e.$$

THÉORÈME 5.10

Si $X \sim \text{Gamma}(s, \lambda)$ et $Y \sim \text{Gamma}(t, \lambda)$ et X et Y sont indépendantes alors $X + Y \sim \text{Gamma}(s + t, \lambda)$.
En particulier, si $X_1, \dots, X_n$ sont indépendantes et identiquement distribuées $\text{Exp}(\lambda)$ alors $X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda)$.

Démonstration. On rappelle la densité d'une variable $\text{Gamma}(s, \lambda)$,

$$f_X(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{s-1}}{\Gamma(s)} \mathbb{1}_{x>0}.$$

En utilisant le produit de convolution, on obtient

$$\begin{aligned} f_{X+Y}(z) &= (f_X * f_Y)(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y) dy = \int_0^z \frac{\lambda e^{-\lambda y} (\lambda y)^{t-1} \lambda e^{-\lambda(z-y)} (\lambda(z-y))^{s-1}}{\Gamma(s)\Gamma(t)} dy \\ &= \frac{\lambda^2 \lambda^{t+s-2} e^{-\lambda z}}{\Gamma(s)\Gamma(t)} \int_0^z y^{t-1} (z-y)^{s-1} dy. \end{aligned}$$

On réalise dans l'intégrale le changement de variable $x = \frac{y}{z}$ alors cela devient

$$f_{X+Y}(z) = \frac{\lambda e^{-\lambda z} \lambda^{t+s-1}}{\Gamma(s)\Gamma(t)} \int_0^1 (zx)^{t-1} (z(1-x))^{s-1} z dx = \lambda e^{-\lambda z} (\lambda z)^{t+s-1} \frac{\int_0^1 x^{t-1} (1-x)^{s-1} dx}{\Gamma(s)\Gamma(t)}.$$

On reconnaît pour le terme $\lambda e^{-\lambda z} (\lambda z)^{t+s-1}$ la densité non renormalisée d'une loi Gamma de paramètre $(s+t, \lambda)$. L'autre terme ne dépend pas de z et comme on sait que $f_{X+Y}(z)$ est une densité, on obtient automatiquement la formule

$$\frac{\int_0^1 x^{t-1} (1-x)^{s-1} dx}{\Gamma(s)\Gamma(t)} = \Gamma(s+t) \quad (4)$$

et ainsi $X+Y \sim \text{Gamma}(s+t, \lambda)$. □

THÉORÈME 5.11

Si $X_1, \dots, X_n$ sont indépendantes et identiquement distribuées $\mathcal{N}(0, 1)$ alors $Z = X_1^2 + \dots + X_n^2$ est distribuée selon la loi du χ_n^2 et on a $\chi_n^2 \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$.

Démonstration. Pour le cas $n = 1$, on a $Z \sim \mathcal{N}(0, 1)^2$, on avait vu que pour $z > 0$,

$$f_{X^2}(z) = \frac{1}{2\sqrt{z}} (f_X(\sqrt{z}) + f_X(-\sqrt{z})) = \frac{1}{\sqrt{z}} f_X(\sqrt{z}) = \frac{1}{\sqrt{2\pi z}} e^{-\frac{z}{2}} = \frac{\frac{1}{2} e^{-\frac{z}{2}} (\frac{z}{2})^{\frac{1}{2}-1}}{\sqrt{\pi}}$$

et on reconnaît la densité d'une loi Gamma $(\frac{1}{2}, \frac{1}{2})$. par le théorème précédent sur les sommes de variables Gamma indépendantes, on obtient ainsi que $\chi_n^2 \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$ et a ainsi pour densité

$$f_{\chi_n^2}(z) = \frac{z^{\frac{n}{2}-1} e^{-\frac{z}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \mathbb{1}_{z>0}.$$

□

THÉORÈME 5.12

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes telles que $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ alors

$$\sum_{i=1}^n X_i \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

Démonstration. On commence par le cas facile où $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, \sigma^2)$ alors par le produit de convolution on obtient

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y) dy = \int_{\mathbb{R}} \frac{1}{2\pi\sigma} e^{-\frac{y^2}{2\sigma^2} - \frac{(z-y)^2}{2}} dy = \int_{\mathbb{R}} \frac{1}{2\pi\sigma} \exp\left(-\frac{y^2}{2\sigma^2} - \frac{y^2}{2} + \frac{2zy}{2} - \frac{z^2}{2}\right) dy.$$

L'idée est maintenant de réécrire les termes à l'intérieur de l'exponentielle, si on définit ν de telle sorte que

$$\frac{1}{\nu^2} = 1 + \frac{1}{\sigma^2} = \frac{\sigma^2 + 1}{\sigma^2} \quad \text{i.e.} \quad \nu = \frac{\sigma}{\sqrt{1 + \sigma^2}}$$

alors

$$\frac{y^2}{2\sigma^2} + \frac{y^2}{2} - zy = \frac{1}{2\nu^2} (y^2 - 2\nu^2 zy) = \frac{1}{2\nu^2} (y^2 - 2\nu^2 zy + \nu^4 z^2 - \nu^4 z^2) = \frac{1}{2\nu^2} (y - \nu^2 z)^2 - \frac{\nu^2 z^2}{2}.$$

Ainsi, en rentrant cela dans le calcul de densité, on obtient

$$f_{X+Y}(z) = \frac{1}{2\pi\sigma} e^{-\frac{z^2}{2} + \frac{\nu^2 z^2}{2}} \int_{\mathbb{R}} \exp\left(-\frac{(y - \nu^2 z)^2}{2\nu^2}\right) dy.$$

Maintenant, on sait que

$$\int_{\mathbb{R}} \exp\left(-\frac{(y - \nu^2 z)^2}{2\nu^2}\right) dy = \sqrt{2\pi}\nu$$

ce qui donne

$$f_{X+Y}(z) = \frac{1}{\sqrt{2\pi}} \frac{\nu}{\sigma} \exp\left(-\frac{z^2}{2} (1 - \nu^2)\right).$$

Avec la définition de ν , on obtient

$$f_{X+Y} = \frac{1}{\sqrt{2\pi(1 + \sigma^2)}} \exp\left(-\frac{z^2}{2(1 + \sigma^2)}\right)$$

et ainsi $X + Y \sim \mathcal{N}(0, 1 + \sigma^2)$.

Pour $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ alors on a

$$X + Y = \sigma_1 \left(\frac{X - \mu_1}{\sigma_1} + \frac{Y - \mu_2}{\sigma_1} \right) + \mu_1 + \mu_2.$$

Mais $\frac{X - \mu_1}{\sigma_1} \sim \mathcal{N}(0, 1)$ et $\frac{Y - \mu_2}{\sigma_1} \sim \mathcal{N}\left(0, \frac{\sigma_2^2}{\sigma_1^2}\right)$, et par ce qu'on vient de prouver

$$X + Y \sim \sigma_1 \mathcal{N}\left(0, 1 + \frac{\sigma_2^2}{\sigma_1^2}\right) + \mu_1 + \mu_2 \sim \mathcal{N}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right).$$

On obtient le résultat pour n variables gaussiennes par récurrence immédiate. □

EXEMPLE 69. Un objet financier a pour prix $S(n)$ lors de la n -ième semaine. Un modèle pour l'évolution de ce prix est la suivante

$$\frac{S(n)}{S(n-1)} \sim \log \mathcal{N}(\mu, \sigma^2) \text{ et sont indépendants entre eux.}$$

Calculons la probabilité que le prix croisse lors de la première et deuxième semaine. Pour la première semaine, on cherche donc

$$\mathbb{P}\left(\frac{S(1)}{S(0)} > 1\right) = \mathbb{P}\left(\ln\left(\frac{S(1)}{S(0)}\right) > 0\right) = \mathbb{P}\left(\mathcal{N}(\mu, \sigma^2) > 0\right) = \Phi\left(\frac{\mu}{\sigma}\right).$$

Pour que le prix croisse lors des deux premières semaines, par indépendance des ratios des prix entre chaque semaine, on obtient la probabilité $\Phi\left(\frac{\mu}{\sigma}\right)^2$.

Calculons maintenant la probabilité que le prix soit plus élevé après deux semaines. Ce n'est pas la même chose car le prix pourrait décroître lors de la première semaine puis croître fortement lors de la deuxième

semaine. On cherche donc

$$\mathbb{P}\left(\frac{S(2)}{S(0)} > 1\right) = \mathbb{P}\left(\frac{S(2)}{S(1)} \frac{S(1)}{S(0)} > 1\right) = \mathbb{P}\left(\ln\left(\frac{S(2)}{S(1)}\right) + \ln\left(\frac{S(1)}{S(0)}\right) > 0\right) = \mathbb{P}(Z_1 + Z_2 > 0)$$

avec Z_1 et Z_2 deux variables gaussiennes $\mathcal{N}(\mu, \sigma^2)$ indépendantes. Ainsi,

$$\mathbb{P}\left(\frac{S(2)}{S(0)} > 1\right) = \mathbb{P}\left(\mathcal{N}(2\mu, 2\sigma^2) > 0\right) = \Phi\left(\frac{\sqrt{2}\mu}{\sigma}\right).$$

On peut aussi faire une opération similaire au produit de convolution dans le cas discret.

THÉORÈME 5.13

Soient $X \sim \text{Poi}(\lambda_1)$ et $Y \sim \text{Poi}(\lambda_2)$ avec X et Y indépendantes alors $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$.

Démonstration. On commence par écrire

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^n \mathbb{P}(X = k, Y = n - k) = \sum_{k=0}^n \mathbb{P}(X = k) \mathbb{P}(Y = n - k) = \sum_{k=0}^n e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!}$$

On réarrange les termes pour écrire

$$\mathbb{P}(X + Y = n) = e^{-(\lambda_1 + \lambda_2)} \sum_{k=0}^n \frac{\lambda_1^k \lambda_2^{n-k}}{k! (n-k)!} = \frac{e^{-\lambda_1 - \lambda_2}}{n!} \sum_{k=0}^n \binom{n}{k} \lambda_1^k \lambda_2^{n-k} = e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^n}{n!}.$$

□

5.3 Distributions conditionnelles

On rappelle que pour deux événements E et F , nous avons défini $\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$. Dans le cas discret, on peut alors définir la distribution conditionnelle de la façon suivante.

DÉFINITION 5.14

Soient X et Y deux variables aléatoires discrètes alors la loi de probabilité conditionnelle de X sous $Y = y$ est

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

On peut aussi définir la fonction de répartition conditionnelle de X sous $Y = y$

$$f_{X|Y}(x|y) = \mathbb{P}(X \leq x | Y = y).$$

On remarque que si X et Y sont indépendantes alors

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

Ainsi la distribution conditionnelle de X sous $Y = y$ est simplement la distribution de X .

EXEMPLE 70. On donne la loi jointe de X et Y par

$$p_{X,Y}(0,0) = 0.4, \quad p_{X,Y}(0,1) = 0.2, \quad p_{X,Y}(1,0) = 0.1, \quad p_{X,Y}(1,1) = 0.3.$$

Quelle est la loi de X conditionnellement à $Y = 1$? On peut calculer

$$p_{X|Y}(x|1) = \frac{p_{X,Y}(x,1)}{p_Y(1)}.$$

On commence par calculer $p_Y(1)$ en sommant sur les valeurs possibles de X ,

$$p_Y(1) = p_{X,Y}(0,1) + p_{X,Y}(1,1) = 0.2 + 0.3 = 0.5.$$

Ainsi, on calcule

$$p_{X|Y}(0|1) = \frac{p_{X,Y}(0,1)}{p_Y(1)} = 0.4, \quad p_{X|Y}(1|1) = \frac{p_{X,Y}(1,1)}{p_Y(1)} = 0.6.$$

EXEMPLE 71. Soient X et Y deux variables aléatoires indépendantes telles que $X \sim \text{Poi}(\lambda_1)$ et $Y \sim \text{Poi}(\lambda_2)$. Calculons la loi de X conditionnellement à $X + Y = n$.

$$\mathbb{P}(X = k | X + Y = n) = \frac{\mathbb{P}(X = k, X + Y = n)}{\mathbb{P}(X + Y = n)} = \frac{\mathbb{P}(X = k, Y = n - k)}{\mathbb{P}(X + Y = n)} = \frac{\mathbb{P}(X = k)\mathbb{P}(Y = n - k)}{\mathbb{P}(X + Y = n)}.$$

Comme on sait que $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$, on obtient

$$\mathbb{P}(X = k | X + Y = n) = e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \frac{n!}{(\lambda_1 + \lambda_2)^n} e^{\lambda_1 + \lambda_2} = \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k}.$$

Et ainsi on obtient que la loi de X conditionnellement à $X + Y = n$ est une loi binomiale de paramètres $\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$.

On peut définir la même chose dans le cas continu en remplaçant la loi de probabilité par la densité.

DÉFINITION 5.15

Soient X et Y deux variables aléatoires de densité jointe $f_{X,Y}$ alors la densité conditionnelle de X sous $Y = y$, si $f_Y(y) \neq 0$ est définie par

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

On retrouve la fonction de répartition conditionnelle de X sous $Y = y$ par

$$F_{X|Y}(x|y) = \mathbb{P}(X \leq x | Y = y) = \int_{-\infty}^x f_{X|Y}(u|y) du.$$

On remarque que même si $\mathbb{P}(Y = y) = 0$ car Y est continue la définition a du sens (du moins que $f_Y(y) > 0$). On peut le voir formellement en se rappelant de la définition formelle de la densité en écrivant

$$\begin{aligned} f_{X|Y}(x|y) dx &= \frac{f_{X,Y}(x,y) dx dy}{f_Y(y) dy} = \frac{\mathbb{P}(x \leq X \leq x + dx, y \leq Y \leq y + dy)}{\mathbb{P}(y \leq Y \leq y + dy)} \\ &= \mathbb{P}(x \leq X \leq x + dx | y \leq Y \leq y + dy). \end{aligned}$$

EXEMPLE 72. On considère deux variables aléatoires X et Y de densité jointe

$$f_{X,Y}(x,y) = \frac{e^{-\frac{x}{y}} e^{-y}}{y} \mathbb{1}_{x>0} \mathbb{1}_{y>0}.$$

Calculons $\mathbb{P}(X > 1 | Y = y)$. On commence par calculer la densité marginale de Y , on a

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y) dx = \int_0^{+\infty} \frac{e^{-\frac{x}{y}} e^{-y}}{y} dx = e^{-y} \left[-e^{-\frac{x}{y}} \right]_0^{+\infty} = e^{-y} \mathbb{1}_{y>0}.$$

On remarque ainsi que Y suit une loi exponentielle de paramètre 1. Maintenant, on peut calculer la loi de X conditionnellement à $Y = y$ et on a pour $x, y > 0$,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{e^{-\frac{x}{y}} e^{-y}}{y}}{e^{-y}} = e^{-\frac{x}{y}} \mathbb{1}_{x>0}.$$

Ainsi conditionnellement à $Y = y$, X suit une loi $\text{Exp}(y)$. On peut maintenant calculer la probabilité voulue,

$$\mathbb{P}(X > 1 | Y = 1) = \int_1^{+\infty} f_{X|Y}(x|y) dx = \int_1^{+\infty} \frac{e^{-\frac{x}{y}}}{y} dx = \left[-e^{-\frac{x}{y}} \right]_1^{+\infty} = e^{-\frac{1}{y}}.$$

On peut aussi avoir des lois discrètes et continues ensemble.

DÉFINITION 5.16

Si X est une variable aléatoire de densité f_X et N une variable aléatoire discrète de loi p_N alors la densité conditionnelle de X sous $N = n$ si $\mathbb{P}(N = n) \neq 0$ est

$$f_{X|N}(x|n) = \mathbb{P}(N = n | X = x) \frac{f_X(x)}{p_N(n)}.$$

EXEMPLE 73. On a $n + m$ épreuves indépendantes avec une probabilité de succès X avec $X \sim \text{Unif}(0, 1)$. Calculons la distribution conditionnelle de X sachant qu'il y a eu n succès. Si $X = x$ alors le nombre de succès N suit une loi binomiale de paramètres $(n + m, x)$ alors la densité conditionnelle devient

$$f_{X|N}(x|n) = \frac{\mathbb{P}(N = n | X = x) f_X(x)}{\mathbb{P}(N = n)} = \frac{\binom{n+m}{n} x^n (1-x)^m}{\mathbb{P}(N = n)} \mathbb{1}_{0 < x < 1}.$$

On remarque que l'on a pas besoin de calculer $\mathbb{P}(N = n)$ car cela ne dépend pas de x et on sait que $\int_{\mathbb{R}} f_{X|N}(x|n) dx = 1$ ainsi on a nécessairement

$$\mathbb{P}(N = n) = \int_0^1 \binom{n+m}{n} x^n (1-x)^m dx.$$

Ainsi, cela donne

$$f_{X|N}(x|n) = \frac{x^n (1-x)^m}{\int_0^1 x^n (1-x)^m dx} = \frac{\Gamma(n+m+2)}{\Gamma(n+1)\Gamma(m+1)} x^n (1-x)^m = \frac{(n+m+1)!}{n!m!} x^n (1-x)^m$$

où on a utilisé la formule (4). Cette distribution est appelée une loi Beta de paramètre $(n+1, m+1)$.

5.4 Changement de variables

Similairement au cas d'intégrales multiples, nous avons un théorème de changement de variables pour les variables aléatoires continues qui nous permettent de calculer des densités.

THÉORÈME 5.17

Soient $X_1, X_2, \dots, X_n$ n variables aléatoires conjointement continues de densité $f_{X_1, \dots, X_n}$. Si on définit,

$$Y_1 = g_1(X_1, \dots, X_n), \dots, Y_n = g_n(X_1, \dots, X_n)$$

et si on suppose que le système d'équations

$$\begin{cases} y_1 = g_1(x_1, \dots, x_n), \\ \vdots \\ y_n = g_n(x_1, \dots, x_n). \end{cases} \iff \begin{cases} x_1 = h_1(y_1, \dots, y_n), \\ \vdots \\ x_n = h_n(y_1, \dots, y_n). \end{cases}$$

On suppose aussi que les fonctions $g_1, \dots, g_n$ sont dérivables et que leurs dérivées partielles de premier ordre sont continues et que

$$J(x_1, \dots, x_n) = \det \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(x_1, \dots, x_n) & \dots & \frac{\partial g_1}{\partial x_n}(x_1, \dots, x_n) \\ \vdots & & \vdots \\ \frac{\partial g_n}{\partial x_1}(x_1, \dots, x_n) & \dots & \frac{\partial g_n}{\partial x_n}(x_1, \dots, x_n) \end{bmatrix} \neq 0.$$

Alors on peut écrire la densité jointe de Y_1, Y_2 par

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = \frac{1}{|J(h_1(y_1, \dots, y_n), \dots, h_n(y_1, \dots, y_n))|} f_{X_1, \dots, X_n}(h_1(y_1, \dots, y_n), \dots, h_n(y_1, \dots, y_n))$$

avec $x_i = h_i(y_1, \dots, y_n)$.

EXEMPLE 74. Soit (X_1, X_2) de densité f_{X_1, X_2} et $Y_1 = X_1 + X_2$ et $Y_2 = X_1 - X_2$. Calculons la densité de (Y_1, Y_2) . Nous avons

$$Y_1 = g_1(X_1, X_2), \quad Y_2 = g_2(X_1, X_2) \quad \text{avec} \quad g_1(x_1, x_2) = x_1 + x_2 \quad \text{et} \quad g_2(x_1, x_2) = x_1 - x_2.$$

Nous avons le système d'équations

$$\begin{cases} y_1 = x_1 + x_2, \\ y_2 = x_1 - x_2 \end{cases} \Leftrightarrow \begin{cases} x_1 = \frac{y_1 + y_2}{2} =: h_1(y_1, y_2), \\ x_2 = \frac{y_1 - y_2}{2} =: h_2(y_1, y_2). \end{cases}$$

On calcule les dérivées partielles de g_1 et g_2 et on a

$$\partial_{x_1} g_1(x_1, x_2) = 1 = \partial_{x_2} g_1(x_1, x_2) = \partial_{x_1} g_2(x_1, x_2) = -\partial_{x_2} g_2(x_1, x_2).$$

Ainsi, le jacobien est donné par

$$J(x_1, x_2) = \det \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = -1 - 1 = -2 \neq 0 \text{ pour tout } x_1, x_2 \in \mathbb{R}.$$

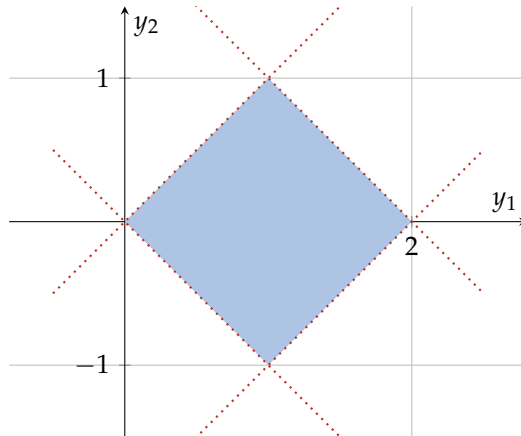
On obtient finalement la densité jointe de Y_1 et Y_2 ,

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{1}{|-2|} f_{X_1, X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right) = \frac{1}{2} f_{X_1, X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right).$$

Par exemple si X_1 et X_2 sont deux variables uniformes sur $(0, 1)$ et indépendantes alors on a

$$x_1 \in (0, 1), x_2 \in (0, 1) \Leftrightarrow y_1 + y_2 \in (0, 2), \quad y_1 - y_2 \in (0, 2).$$

$$f_{Y_1, Y_2} = \frac{1}{2} \mathbb{1}_{0 < y_1 + y_2 < 2} \mathbb{1}_{0 < y_1 - y_2 < 2}$$



Si X_1 et X_2 sont indépendantes et $X_1 \sim \text{Exp}(\lambda_1)$ et $X_2 \sim \text{Exp}(\lambda_2)$ alors on obtient la densité

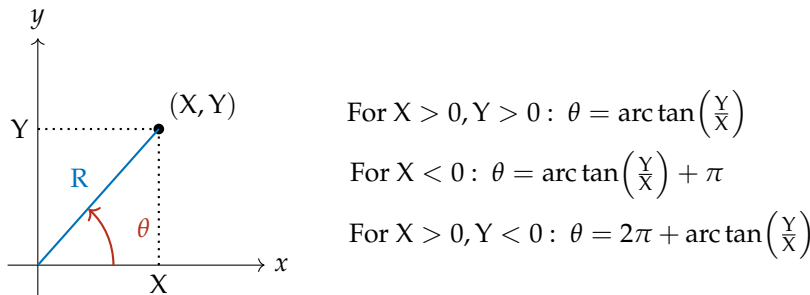
$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= \frac{\lambda_1 \lambda_2}{2} e^{-\lambda_1(y_1+y_2)} e^{-\lambda_2(y_1-y_2)} \mathbb{1}_{y_1+y_2>0} \mathbb{1}_{y_1-y_2>0} \\ &= \frac{\lambda_1 \lambda_2}{2} e^{-(\lambda_1+\lambda_2)y_1} e^{-(\lambda_1-\lambda_2)y_2} \mathbb{1}_{y_1+y_2>0} \mathbb{1}_{y_1-y_2>0} \end{aligned}$$

Si X_1 et X_2 sont deux variables normales centrées réduites indépendantes alors

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= \frac{1}{2\sqrt{2\pi}} e^{-\frac{(y_1+y_2)^2}{8}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_1-y_2)^2}{8}} = \frac{1}{4\pi} \exp\left(-\frac{y_1^2}{8} - \frac{y_2^2}{8} - \frac{2y_1y_2}{8} - \frac{y_1^2}{8} - \frac{y_2^2}{8} + \frac{2y_1y_2}{8}\right) \\ &= \frac{1}{\sqrt{4\pi}} e^{-\frac{y_1^2}{4}} \frac{1}{\sqrt{4\pi}} e^{-\frac{y_2^2}{4}}. \end{aligned}$$

Ainsi, Y_1 et Y_2 sont indépendantes et $Y_1, Y_2 \sim \mathcal{N}(0, 2)$.

EXEMPLE 75. On suppose que $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 1)$ et on regarde le point (X, Y) dans le plan en coordonnées polaires. On a bien sur $R = \sqrt{X^2 + Y^2}$, pour l'angle θ il faut faire un peu attention car la formule dépend des coordonnées du point (X, Y) ,



Regardons pour l'instant, le cas où $X > 0$ et $Y > 0$, c'est-à-dire $\theta \in (0, \frac{\pi}{2})$. On a

$$R = g_1(X, Y) = \sqrt{X^2 + Y^2}, \quad \theta = g_2(X, Y) = \arctan\left(\frac{Y}{X}\right).$$

On sait que

$$X = R \cos(\theta) = h_1(R, \theta), \quad Y = R \sin(\theta) = h_2(R, \theta).$$

On calcule les dérivées partielles de g_1 et de g_2

$$\begin{aligned} \partial_{x_1} g_1(x, y) &= \frac{x}{\sqrt{x^2 + y^2}}, & \partial_{x_2} g_1(x, y) &= \frac{y}{\sqrt{x^2 + y^2}}, \\ \partial_{x_1} g_2(x, y) &= \frac{-y}{x^2 + y^2}, & \partial_{x_2} g_2(x, y) &= \frac{x}{x^2 + y^2}. \end{aligned}$$

Si $X > 0$ et $Y > 0$ les dérivées existent et sont continues et ainsi on peut calculer le Jacobien

$$J(x, y) = \det \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix} = \frac{x^2 + y^2}{r^3} = \frac{1}{r} \neq 0.$$

Or on a la densité conditionnelle

$$f_{X,Y}(x, y | X > 0, Y > 0) = \frac{f_{X,Y}(x, y) \mathbb{1}_{x>0, y>0}}{\mathbb{P}(X > 0, Y > 0)} = 4f_{X,Y}(x, y) = \frac{2}{\pi} e^{-\frac{x^2+y^2}{2}}.$$

Le changement de variable nous donne donc

$$f_{R,\theta}(r, \theta | X > 0, Y > 0) = \frac{2}{\pi} r e^{-\frac{r^2}{2}} \mathbb{1}_{r>0} \mathbb{1}_{0<\theta<2\pi}.$$

On peut calculer de même

$$f_{R,\theta}(r, \theta | X > 0, Y < 0) = \frac{2}{\pi} r e^{-\frac{r^2}{2}} \mathbb{1}_{r>0} \mathbb{1}_{\frac{3\pi}{2} < \theta < 2\pi}, \quad f_{R,\theta}(r, \theta | X < 0) = \frac{2}{\pi} r e^{-\frac{r^2}{2}} \mathbb{1}_{r>0} \mathbb{1}_{\frac{\pi}{2} < \theta < \frac{3\pi}{2}}.$$

Finalement, avec la loi des probabilités totales, on obtient

$$f_{R,\theta}(r, \theta) = \frac{1}{2\pi} r e^{-\frac{r^2}{2}} \mathbb{1}_{r>0} \mathbb{1}_{0 < \theta < 2\pi}.$$

On remarque ainsi que R et θ sont indépendants et $\theta \sim \text{Unif}(0, 2\pi)$. On appelle la loi de R la loi de Rayleigh. Si on avait plutôt regardé la loi jointe de (R^2, θ) alors le jacobien serait devenu

$$J(x, y) = \det \begin{bmatrix} 2x & 2y \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{bmatrix} = \frac{2x^2 + 2y^2}{r^2} = 2 \neq 0.$$

Et la densité finale serait alors

$$f_{R^2, \theta}(d, \theta) = \frac{1}{4\pi} e^{-\frac{d}{2}} \mathbb{1}_{d>0} \mathbb{1}_{0 < \theta < 2\pi}.$$

R^2 et θ sont bien sur toujours indépendants mais cette fois $R^2 \sim \text{Exp}\left(\frac{1}{2}\right)$. Cette observation nous permet de simuler des variables aléatoires gaussiennes à partir du moment où l'on sait simuler des variables uniformes entre 0 et 1. La première observation est que $-2 \ln(U) \sim \text{Exp}\left(\frac{1}{2}\right)$ si $U \sim \text{Unif}(0, 1)$, en effet pour $x > 0$,

$$\mathbb{P}(-2 \ln(U) < x) = \mathbb{P}(U < e^{-\frac{x}{2}}) = 1 - e^{-\frac{x}{2}}$$

qui correspond à la fonction de répartition de la loi exponentielle de paramètre $\frac{1}{2}$. Ainsi si U_1 et U_2 sont deux variables uniformes entre 0 et 1 indépendantes alors on peut poser

$$R^2 = -2 \ln(U_1) \quad \text{et} \quad \theta = 2\pi U_2$$

de telle façon que $X_1 = R \cos(\theta)$ et $X_2 = R \sin(\theta)$ sont deux variables aléatoires gaussiennes indépendantes. On veut donc poser

$$X_1 = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2), \quad X_2 = \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$$

et on vient de voir que $X_1, X_2 \sim \mathcal{N}(0, 1)$ et X_1 et X_2 sont indépendantes.

EXEMPLE 76. Soient X_1, X_2, X_3 trois variables aléatoires indépendantes normales centrées réduites. Donnons la densité jointe de Y_1, Y_2, Y_3 avec

$$Y_1 = X_1 + X_2 + X_3, \quad Y_2 = X_1 - X_2, \quad Y_3 = X_1 - X_3.$$

On a les fonctions

$$g_1(x_1, x_2, x_3) = x_1 + x_2 + x_3, \quad g_2(x_1, x_2, x_3) = x_1 - x_2, \quad g_3(x_1, x_2, x_3) = x_1 - x_3.$$

On peut résoudre le système,

$$\begin{cases} y_1 = x_1 + x_2 + x_3, \\ y_2 = x_1 - x_2, \\ y_3 = x_1 - x_3 \end{cases} \iff \begin{cases} x_1 = \frac{y_1 + y_2 + y_3}{3} = h_1(y_1, y_2, y_3), \\ x_2 = \frac{y_1 - 2y_2 + y_3}{3} = h_2(y_1, y_2, y_3), \\ x_3 = \frac{y_1 + y_2 - 2y_3}{3} = h_3(y_1, y_2, y_3). \end{cases}$$

Toutes les dérivées de g_i sont constantes donc bien continues et on calcule le jacobien

$$J(x_1, x_2, x_3) = \det \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} = 1 \cdot \begin{vmatrix} 1 & -1 \\ 1 & 0 \end{vmatrix} - 1 \cdot \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = 3 \neq 0.$$

Finalement, on obtient la densité

$$f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) = \frac{1}{3} f_{X_1, X_2, X_3} \left(\frac{y_1 + y_2 + y_3}{3}, \frac{y_1 - 2y_2 + y_3}{3}, \frac{y_1 + y_2 - 2y_3}{3} \right)$$

avec la densité jointe de X_1, X_2, X_3 donnée par

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \frac{1}{(2\pi)^{\frac{3}{2}}} e^{-\frac{x_1^2 + x_2^2 + x_3^2}{2}}.$$

On veut donc développer

$$(y_1 + y_2 + y_3)^2 + (y_1 - 2y_2 + y_3)^2 + (y_1 + y_2 - 2y_3)^2 = 3y_1^2 + 6y_2^2 + 6y_3^2 - 6y_2y_3.$$

Finalement, on obtient

$$f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) = \frac{1}{3(2\pi)^{\frac{3}{2}}} e^{-\frac{y_1^2}{6}} e^{-\frac{y_2^2}{3}} e^{-\frac{y_3^2}{3}} e^{\frac{y_2 y_3}{3}}.$$

Les variables aléatoires ne sont donc pas indépendantes.

EXEMPLE 77. Soient $X_1, \dots, X_n$ des variables indépendantes $\text{Exp}(\lambda)$ et on considère

$$Y_i = X_1 + \dots + X_i.$$

On sait déjà que comme $X_i \sim \text{Gamma}(1, \lambda)$ alors $Y_i \sim \text{Gamma}(i, \lambda)$. Mais calculons maintenant la densité jointe de $Y_1, \dots, Y_n$. On a

$$Y_i = g_i(X_1, \dots, X_n) = X_1 + \dots + X_i.$$

De plus, on peut résoudre le système

$$\begin{cases} y_1 = x_1, \\ y_2 = x_1 + x_2, \\ \vdots \\ y_n = x_1 + \dots + x_n \end{cases} \iff \begin{cases} x_1 = y_1 = h_1(y_1, \dots, y_n), \\ x_2 = y_2 - y_1 = h_2(y_1, \dots, y_n), \\ \vdots \\ x_n = y_n - y_{n-1} = h_n(y_1, \dots, y_n) \end{cases}$$

On calcule le jacobien facilement,

$$J(x_1, \dots, x_n) = \det \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} = 1 \neq 0$$

où on a calculé le déterminant car la matrice est triangulaire inférieure et le déterminant est donc les produits des éléments diagonaux. Maintenant, on sait que la densité jointe des X_i est

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n \lambda e^{-\lambda x_i} \mathbb{1}_{x_i > 0}$$

ce qui donne par le changement de variables, avec la convention $y_0 = 0$,

$$\begin{aligned} f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) &= \prod_{i=1}^n \lambda e^{-\lambda(y_i - y_{i-1})} \mathbb{1}_{y_i - y_{i-1} > 0} = \lambda^n e^{-\lambda \sum_{i=1}^n (y_i - y_{i-1})} \mathbb{1}_{0 < y_1 < y_2 < \dots < y_n} \\ &= \lambda^n e^{-\lambda y_n} \mathbb{1}_{0 < y_1 < y_2 < \dots < y_n}. \end{aligned}$$

Si on veut retrouver la loi marginale de Y_n (que l'on connaît déjà), on peut le faire en intégrant une par une les variables.

$$\begin{aligned} f_{Y_2, \dots, Y_n}(y_2, \dots, y_n) &= \int_0^{y_2} \lambda^n e^{-\lambda y_n} \mathbb{1}_{0 < y_2 < \dots < y_n} dy_1 = \lambda^n y_2 e^{-\lambda y_n} \mathbb{1}_{0 < y_2 < \dots < y_n} \\ f_{Y_3, \dots, Y_n}(y_3, \dots, y_n) &= \int_0^{y_3} \lambda^n y_2 e^{-\lambda y_n} \mathbb{1}_{0 < y_3 < \dots < y_n} dy_2 = \lambda^n \frac{y_3^2}{2} e^{-\lambda y_n} \mathbb{1}_{0 < y_3 < \dots < y_n}, \\ &\vdots \\ f_{Y_n}(y_n) &= \int_0^{y_n} \lambda^n \frac{y_{n-1}^{n-2}}{(n-2)!} e^{-\lambda y_n} \mathbb{1}_{y_n > 0} dy_{n-1} = \frac{\lambda^n y_n^{n-1} e^{-\lambda y_n}}{(n-1)!} \mathbb{1}_{y_n > 0} \end{aligned}$$

qui correspond bien à une loi Gamma(n, λ).

6. ESPÉRANCE MATHÉMATIQUE

On rappelle la définition,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx \text{ dans le cas continu, } \mathbb{E}[X] = \sum_x x p_X(x) \text{ dans le cas discret.}$$

Nous allons voir plusieurs propriétés de l'espérance dans cette section.

THÉORÈME 6.1

Si $\mathbb{P}(a \leq X \leq b) = 1$ alors $a \leq \mathbb{E}[X] \leq b$.

Démonstration. Faisons la preuve dans le cas continu, on peut écrire

$$\mathbb{E}[X] = \mathbb{E}[X \mathbb{1}_{a \leq X \leq b}] = \int_{\mathbb{R}} x \mathbb{1}_{a \leq x \leq b} f_X(x) dx = \int_a^b x f_X(x) dx.$$

Maintenant, on peut borner

$$\int_a^b x f_X(x) dx \geq a \int_a^b f_X(x) dx = a \int_{\mathbb{R}} f_X(x) dx = a.$$

De même,

$$\int_a^b x f_X(x) dx \leq b \int_a^b f_X(x) dx = b \int_{\mathbb{R}} f_X(x) dx = b$$

□

THÉORÈME 6.2

Si X et Y sont deux variables aléatoires discrètes de loi jointe $p_{X,Y}$ alors

$$\mathbb{E}[g(X, Y)] = \sum_y \sum_x g(x, y) p_{X,Y}(x, y).$$

Si X et Y sont deux variables aléatoires continues de densité jointe $f_{X,Y}$ alors

$$\mathbb{E}[g(X, Y)] = \iint_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) dx dy.$$

Démonstration. Développons la preuve dans le cas continu et commençons par le cas où g est positive. On a déjà vu que

$$\mathbb{E}[g(X, Y)] = \int_0^\infty \mathbb{P}(g(X, Y) > t) dt = \int_0^\infty \iint_{g(x, y) > t} f_{X, Y}(x, y) dx dy dt = \int_0^\infty \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{g(x, y) > t} f_{X, Y}(x, y) dx dy dt.$$

On échange l'ordre d'intégration, pour écrire

$$\begin{aligned} \mathbb{E}[g(X, Y)] &= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_0^\infty \mathbb{1}_{t < g(x, y)} f_{X, Y}(x, y) dt dx dy = \int_{\mathbb{R}} \int_{\mathbb{R}} \int_0^{g(x, y)} f_{X, Y}(x, y) dt dx dy \\ &= \iint_{\mathbb{R}^2} g(x, y) f_{X, Y}(x, y) dx dy. \end{aligned}$$

Pour g une fonction quelconque, on réalise la décomposition $g = g^+ - g^-$. □

EXEMPLE 78. On considère une route de longueur L . On suppose qu'un accident a lieu sur la route avec une loi $\text{Unif}(0, L)$ et de même une ambulance se trouve sur la route selon une loi indépendante $\text{Unif}(0, L)$. Calculons la distance moyenne entre une ambulance et l'accident. En dénotant X la distance sur la route de l'accident et Y celle de l'ambulance, on cherche à calculer

$$\mathbb{E}[|X - Y|] \quad \text{avec} \quad f_{X, Y}(x, y) = \frac{1}{L^2} \mathbb{1}_{0 < x < L} \mathbb{1}_{0 < y < L}.$$

Ainsi, on calcule

$$\mathbb{E}[|X - Y|] = \frac{1}{L^2} \int_0^L \int_0^L |x - y| dx dy$$

et on commence par calculer l'intégrale intérieure

$$\begin{aligned} \frac{1}{L^2} \int_0^L |x - y| dx &= \frac{1}{L^2} \int_0^y (y - x) dx + \frac{1}{L^2} \int_y^L (x - y) dx = \frac{1}{L^2} \left[-\frac{(y - x)^2}{2} \right]_0^y + \frac{1}{L^2} \left[\frac{(x - y)^2}{2} \right]_y^L \\ &= \frac{y^2}{2L^2} + \frac{(L - y)^2}{2} \\ &= \frac{1}{L^2} \left(\frac{L^2}{2} - Ly + y^2 \right). \end{aligned}$$

Maintenant, on finit en intégrant ceci par rapport à y et on obtient

$$\mathbb{E}[|X - Y|] = \int_0^L \frac{1}{L^2} \left(\frac{L^2}{2} - Ly + y^2 \right) dy = \frac{1}{L^2} \left(\frac{L^3}{2} - \frac{L^3}{2} + \frac{L^3}{3} \right) = \frac{L}{3}.$$

Nous avons déjà vu la linéarité de l'espérance dans le cas discret mais ceci marche pour toute variable aléatoire.

THÉORÈME 6.3

Si X et Y sont deux variables aléatoires d'espérance finie alors

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Démonstration. Faisons le dans le cas continu, on écrit

$$\begin{aligned} \mathbb{E}[X + Y] &= \iint_{\mathbb{R}^2} (x + y) f_{X, Y}(x, y) dx dy = \iint_{\mathbb{R}^2} x f_{X, Y}(x, y) dy dx + \iint_{\mathbb{R}^2} y f_{X, Y}(x, y) dx dy \\ &= \int_{\mathbb{R}} x f_X(x) dx + \int_{\mathbb{R}} y f_Y(y) dy \\ &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

□

Cela nous donne que l'espérance d'une somme finie de variables aléatoires est la somme des espérances. En revanche, le résultat pour une somme infinie n'est pas vrai, on n'a pas toujours



$$\mathbb{E} \left[\sum_{i=1}^{\infty} X_i \right] = \sum_{i=1}^{\infty} \mathbb{E} [X_i].$$

Il y a cependant des cas où on peut le faire

- Si $\mathbb{P}(X_i \geq 0) = 1$ pour tout $i \in \mathbb{N}$.
- Si $\sum_{i=1}^{\infty} \mathbb{E} [|X_i|] < \infty$.

EXEMPLE 79. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées d'espérance μ . On définit la moyenne empirique de l'échantillon

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

On remarque alors que

$$\mathbb{E} [\bar{X}] = \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E} [X_i] = \frac{1}{n} n\mu = \mu.$$

On peut aussi utiliser la linéarité de l'espérance pour retrouver la moyenne d'une loi binomiale par exemple. On rappelle que $X \sim \text{Bin}(n, p)$ correspond au nombre de succès sur n lancers avec probabilité p de succès. Si on définit la variable aléatoire

$$X_i = \begin{cases} 1 & \text{si succès au } i\text{-ième lancer,} \\ 0 & \text{sinon,} \end{cases}$$

alors on peut écrire $X = \sum_{i=1}^n X_i$ et ainsi comme $\mathbb{E} [X_i] = p$ alors

$$\mathbb{E} [X] = \sum_{i=1}^n \mathbb{E} [X_i] = np.$$

EXEMPLE 80. On a N types de cartes à collectionner qui ont toutes la même rareté. Calculons le nombre moyen de cartes à rassembler pour toutes les obtenir.

Soit X ce nombre, l'idée est d'écrire $X = X_1 + \dots + X_N$ avec

X_1 : nombre de cartes à rassembler pour obtenir un type de cartes,

X_2 : nombre de cartes à rassembler après X_1 pour obtenir un deuxième type de cartes,

$\vdots$

X_i : nombre de cartes à rassembler après $X_1 + \dots + X_{i-1}$ pour obtenir un i -ième type de cartes.

En particulier, on remarque que $X_1 = 1$. Si on possède i types de cartes alors, on remarque que

$$\mathbb{P}(X_{i+1} = k) = \left(\frac{i}{N} \right)^{k-1} \cdot \frac{N-i}{N}$$

car cela veut dire que les $k-1$ premières cartes rassemblées ont été choisies parmi les i types déjà connus et la k -ième a été choisie parmi les $N-i$ types pas encore découverts. On reconnaît donc une loi géométrique de paramètre $\frac{N-i}{N}$ et on sait donc que $\mathbb{E}[X_i] = \frac{N}{N-i}$. Cela nous donne finalement par la linéarité de l'espérance

$$\mathbb{E} [X] = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \dots + N = N \left(\sum_{i=1}^N \frac{1}{i} \right).$$

En particulier, on remarque que lorsque N est grand, on a l'équivalence

$$\mathbb{E}[X] \sim N \ln(N).$$

On peut aussi utiliser les probabilités et les propriétés de l'espérance pour résoudre des problèmes déterministes comme dans l'exemple suivant.

EXEMPLE 81. On considère 52 arbres mis en cercle et on suppose que 15 écureuils vivent dans ce bosquet. Montrons qu'il existe 7 arbres adjacents où vivent 3 écureuils au moins.

On commence par tirer un voisinage de 7 arbres au hasard et on définit X le nombre d'écureuils qui vivent dans le voisinage et on définit

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ième écureuil vit dans le voisinage,} \\ 0 & \text{sinon.} \end{cases}$$

Alors on a $X = \sum_{i=1}^{15} X_i$ et $\mathbb{E}[X] = \sum_{i=1}^{15} \mathbb{E}[X_i]$. On remarque alors que $\mathbb{E}[X_i] = \frac{7}{52}$ et ainsi

$$\mathbb{E}[X] = \frac{15 \times 7}{52} = \frac{105}{52} > 2.$$

Si tous les voisinages avaient moins de deux écureuils alors $X \leq 2$ mais alors nous aurions $\mathbb{E}[X] \leq 2$ or nous avons $\mathbb{E}[X] > 2$ donc il existe au moins un voisinage avec au moins 3 écureuils.

THÉORÈME 6.4

Soient X et Y deux variables aléatoires indépendantes et h et g deux fonctions alors

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

Démonstration. On suppose que X et Y sont conjointement continues alors

$$\begin{aligned} \mathbb{E}[g(X)h(Y)] &= \iint_{\mathbb{R}^2} g(x)h(y)f_{X,Y}(x,y)dx dy = \iint_{\mathbb{R}^2} g(x)h(y)f_X(x)f_Y(y)dx dy \\ &= \left(\int_{\mathbb{R}} g(x)f_X(x)dx \right) \left(\int_{\mathbb{R}} h(y)f_Y(y)dy \right) \\ &= \mathbb{E}[g(X)]\mathbb{E}[h(Y)]. \end{aligned}$$

□

DÉFINITION 6.5

La covariance de X et Y est définie par

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

THÉORÈME 6.6

Si X et Y sont indépendantes alors $\text{Cov}(X, Y) = 0$. La réciproque est fausse.

Démonstration. On a déjà vu que si X et Y sont indépendants alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ et ainsi leur covariance est nulle. Pour voir que la réciproque est fausse, regardons l'exemple suivant. On définit

$$X : \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{3}$$

et

$$Y = \begin{cases} 0 & \text{si } X \neq 0, \\ 1 & \text{sinon.} \end{cases}$$

On voit alors que $XY = 0$ et ainsi $\mathbb{E}[XY] = 0$ et de plus $\mathbb{E}[X] = 0$ donc on a bien $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0$ et

$\text{Cov}(X, Y) = 0$. En revanche, X et Y ne sont pas indépendants pour le voir on peut calculer

$$\mathbb{P}(Y = 1, X = 0) = \mathbb{P}(X = 0) = \frac{1}{3} \quad \text{mais} \quad \mathbb{P}(Y = 1)\mathbb{P}(X = 0) = \frac{1}{3^2} = \frac{1}{9}.$$

□

THÉORÈME 6.7

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- $\text{Cov}(X, X) = \text{Var}(X)$.
- $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$.
- $\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$.

Démonstration. On voit facilement que

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

De plus

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

Aussi,

$$\text{Cov}(aX, Y) = \mathbb{E}[aXY] - \mathbb{E}[aX]\mathbb{E}[Y] = a\mathbb{E}[XY] - a\mathbb{E}[X]\mathbb{E}[Y] = a\text{Cov}(X, Y).$$

Finalement, en dénotant $\mu_i = \mathbb{E}[X_i]$ et $\nu_j = \mathbb{E}[Y_j]$

$$\begin{aligned} \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) &= \mathbb{E}\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i, \sum_{j=1}^m Y_j - \sum_{j=1}^m \nu_j\right)\right] = \mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^m (X_i - \mu_i)(Y_j - \nu_j)\right] \\ &= \sum_{i=1}^n \sum_{j=1}^m \mathbb{E}[(X_i - \mu_i)(Y_j - \nu_j)] \\ &= \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j). \end{aligned}$$

□

EXEMPLE 82. Montrons que la moyenne empirique $\bar{X}$ et $X_i - \bar{X}$ sont décorrélées pour $X_1, \dots, X_n$ un échantillon de variables aléatoires i.i.d d'espérance μ et de variance σ^2 i.e.

$$\text{Cov}(X_i - \bar{X}, \bar{X}) = 0.$$

On commence par calculer

$$\text{Cov}(X_i - \bar{X}, \bar{X}) = \text{Cov}(X_i, \bar{X}) - \text{Cov}(\bar{X}, \bar{X}) = \text{Cov}\left(X_i, \frac{1}{n} \sum_{j=1}^n X_j\right) - \text{Var}(\bar{X})$$

Or, on a

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n}.$$

Finalement,

$$\text{Cov}(X_i, X - i - \bar{X}) \frac{1}{n} \sum_{j=1}^n \text{Cov}(X_i, X_j) - \frac{\sigma^2}{n} = \frac{1}{n} \sigma^2 - \frac{\sigma^2}{n} = 0.$$

THÉORÈME 6.8

Soient $X_1, \dots, X_n$ n variables aléatoires alors

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j).$$

Si $X_1, \dots, X_n$ sont indépendantes deux à deux alors

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

Démonstration. On utilise le théorème précédent pour écrire

$$\begin{aligned} \text{Var} \left(\sum_{i=1}^n X_i \right) &= \text{Cov} \left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j \right) = \sum_{i,j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Cov}(X_i, X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j). \end{aligned}$$

Si X_i et X_j sont indépendants deux à deux alors $\text{Cov}(X_i, X_j) = 0$ pour $i \neq j$ et ainsi

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

□

EXEMPLE 83. On considère X une variable aléatoire binomiale de paramètre (n, p) alors on peut écrire $X = \sum_{i=1}^n X_i$ avec $X_i = \mathbb{1}_{\text{succès au } i\text{-ième lancer}}$. Comme les lancers sont indépendants on a alors

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = \sum_{i=1}^n \left(\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \right) = n(p - p^2) = np(1 - p).$$

EXEMPLE 84. On définit la variance empirique pour un échantillon $X_1, \dots, X_n$ des variables aléatoires i.i.d telle que $\mathbb{E}[X] = \mu$ et $\text{Var}(X) = \sigma^2$ par

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{avec} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Calculons $\mathbb{E}[S^2]$, on a

$$(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu).$$

Or, on a $\sum_{i=1}^n (X_i - \mu) = n\bar{X} - n\mu$ de telle sorte que l'on peut écrire

$$(n-1)S^2 = \sum_{i=1}^n (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2n(\bar{X} - \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2.$$

En prenant l'espérance et en utilisant les propriétés de la variance de la somme variables aléatoires indépendantes

$$\mathbb{E} \left[(n-1)S^2 \right] = n\sigma^2 - n \text{Var}(\bar{X}).$$

Finalement,

$$\mathbb{E} \left[S^2 \right] = \frac{1}{n-1} (n\sigma^2 - \sigma^2) = \sigma^2.$$

DÉFINITION 6.9

On définit le coefficient de corrélation par

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

THÉORÈME 6.10

On a toujours $-1 \leq \rho(X, Y) \leq 1$.

Démonstration. On dénote $\text{Var}(X) = \sigma_X^2$ et $\text{Var}(Y) = \sigma_Y^2$. On a alors

$$\text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \geq 0$$

et en développant

$$\text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} + 2\frac{\text{Cov}(X, Y)}{\sigma_X\sigma_Y} = 2(1 + \rho(X, Y)).$$

Ainsi,

$$\text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \geq 0 \iff \rho(X, Y) \geq -1.$$

De même,

$$\text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) \geq 0 \iff 2(1 - \rho(X, Y)) \geq 0 \iff \rho(X, Y) \leq 1.$$

□

On verra que $\text{Var}(X) = 0$ implique que X est constante avec probabilité 1. Ainsi, par ce que l'on a utilisé dans la preuve ci-dessus,

$$\rho(X, Y) = -1 \Rightarrow \text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = 0 \Leftrightarrow Y = -\frac{\sigma_Y}{\sigma_X}X + a.$$

De même,

$$\rho(X, Y) = 1 \Rightarrow Y = \frac{\sigma_Y}{\sigma_X}X + a.$$

Ceci correspond aux cas extrêmes mais on voit ainsi que si $\rho(X, Y) \geq 0$ alors si l'un augmente, l'autre a tendance à augmenter aussi. Si $\rho(X, Y) \leq 0$ alors si l'un augmente, l'autre a tendance à diminuer. Enfin si $\rho(X, Y) = 0$ les deux variables aléatoires sont décorréliées (mais attention, elles ne sont pas nécessairement indépendantes).

6.1 Espérance conditionnelle

Si X et Y sont deux variables aléatoires discrètes, on a défini si $p_Y(y) \neq 0$,

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

Dans le cas continu, on a défini si $f_Y(y) > 0$,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Cela nous pousse donc à la définition suivante de l'espérance conditionnelle.

DÉFINITION 6.11

Si X et Y deux variables aléatoires discrètes, on définit l'espérance conditionnelle de X sous $Y = y$, si $\mathbb{P}(Y = y) > 0$,

$$\mathbb{E}[X|Y = y] = \sum_x x \mathbb{P}(X = x|Y = y) = \sum_x x p_{X|Y}(x|y).$$

Si X et Y sont deux variables aléatoires continues, on définit l'espérance conditionnelle de X sous $Y = y$, si $f_Y(y) > 0$,

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

EXEMPLE 85. Soit $X \sim \text{Bin}(n, p)$ et $Y \sim \text{Bin}(n, p)$ deux variables aléatoires indépendantes. Calculons $\mathbb{E}[X|X + Y = m]$. On commence par écrire

$$\mathbb{P}(X = k|X + Y = m) = \frac{\mathbb{P}(X = k, X + Y = m)}{\mathbb{P}(X + Y = m)} = \frac{\mathbb{P}(X = k, Y = m - k)}{\mathbb{P}(X + Y = m)} = \frac{\mathbb{P}(X = k)\mathbb{P}(Y = m - k)}{\mathbb{P}(X + Y = m)}.$$

On peut calculer

$$\mathbb{P}(X + Y = m) = \sum_{k=0}^m \mathbb{P}(X + Y = m, X = k) = \sum_{k=0}^m \mathbb{P}(X = k, Y = m - k) = \sum_{k=0}^m \mathbb{P}(X = k)\mathbb{P}(Y = m - k).$$

Ainsi, on obtient

$$\begin{aligned} \mathbb{P}(X + Y = m) &= \sum_{k=0}^m \binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-m+k} = p^m (1-p)^{2n-m} \sum_{k=0}^m \binom{n}{k} \binom{n}{m-k} \\ &= \binom{2n}{m} p^m (1-p)^{2n-m}. \end{aligned}$$

Ainsi,

$$\mathbb{P}(X = k|X + Y = m) = \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-m+k}}{\binom{2n}{m} p^m (1-p)^{2n-m}} = \frac{\binom{n}{k} \binom{n}{m-k}}{\binom{2n}{m}}.$$

Ainsi, on voit que la loi de X sous $X + Y = m$ est une loi Hypergéométrique de paramètre $(m, n, 2n)$. On connaît ainsi son espérance et on a

$$\mathbb{E}[X|X + Y = m] = \frac{mn}{2n} = \frac{m}{2}.$$

EXEMPLE 86. On considère X et Y de densité jointe,

$$f_{X,Y}(x, y) = \frac{e^{-\frac{x}{y}} e^{-y}}{y} \mathbb{1}_{x>0} \mathbb{1}_{y>0}.$$

Calculons $\mathbb{E}[X|Y = y]$, on a déjà calculé

$$f_{X|Y}(x|y) = \frac{e^{-\frac{x}{y}}}{y} \mathbb{1}_{x>0}.$$

et

$$f_Y(y) = e^{-y} \mathbb{1}_{y>0}.$$

On peut donc calculer

$$\mathbb{E}[X|Y = y] = \int_0^{\infty} x \frac{e^{-\frac{x}{y}}}{y} dy = y.$$

Tout comme on a vu que la probabilité conditionnelle est une vraie probabilité, l'espérance conditionnelle est une vraie espérance et les théorèmes que l'on a vus sont aussi vrais. Par exemple, on a le théorème suivant,

THÉORÈME 6.12

On a

$$\mathbb{E} \left[\sum_{i=1}^n X_i \middle| Y = y \right] = \sum_{i=1}^n \mathbb{E} [X_i | Y = y]$$

et si g est une fonction alors

$$\mathbb{E} [g(X) | Y = y] = \begin{cases} \sum_x g(x) p_{X|Y}(x|y) & \text{dans le cas discret,} \\ \int_{-\infty}^{+\infty} g(x) f_{X|Y}(x|y) dx & \text{dans le cas continu.} \end{cases}$$

Ici on a défini une fonction $h(y) = \mathbb{E} [X | Y = y]$. On peut alors mettre Y en argument de cette fonction et il en dénote ceci $\mathbb{E} [X | Y] := h(Y)$. En particulier, cette fonction est aléatoire même si c'est une espérance ! On se donne connu la variable aléatoire Y et on calcule alors l'espérance de X sachant Y .

DÉFINITION 6.13

Si on pose $h(y) = \mathbb{E} [X | Y = y]$ alors on définit

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

THÉORÈME 6.14

Si X et Y sont deux variables aléatoires alors

$$\mathbb{E} [X] = \mathbb{E} [\mathbb{E} [X | Y]].$$

Cela s'écrit

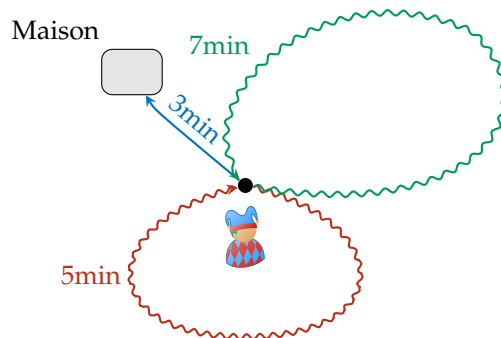
$$\mathbb{E} [X] = \begin{cases} \sum_y \mathbb{E} [X | Y = y] \mathbb{P}(Y = y) & \text{dans le cas discret,} \\ \int_{-\infty}^{+\infty} \mathbb{E} [X | Y = y] f_Y(y) dy & \text{dans le cas continu.} \end{cases}$$

Démonstration. Faisons la preuve dans le cas discret,

$$\begin{aligned} \sum_y \mathbb{E} [X | Y = y] \mathbb{P}(Y = y) &= \sum_y \sum_x x \mathbb{P}(X = x | Y = y) \mathbb{P}(Y = y) = \sum_y \sum_x x \mathbb{P}(X = x, Y = y) \\ &= \sum_x x \mathbb{P}(X = x) = \mathbb{E}[X]. \end{aligned}$$

□

EXEMPLE 87. Un homme saoul veut rentrer chez lui : il voit trois chemins. Le premier met 3 minutes et il rentrera chez lui, le deuxième met 5 minutes et revient au point de départ, le troisième met 7 minutes et revient au point de départ. S'il revient au point de départ, il a oublié ce qu'il a fait et reprend un chemin au hasard. En moyenne, combien de temps met-il à rentrer chez lui ?



On définit X le temps de trajet et Y le numéro du premier chemin choisi, par la loi des probabilités totales,

on peut écrire

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X|Y=1]\mathbb{P}(Y=1) + \mathbb{E}[X|Y=2]\mathbb{P}(Y=2) + \mathbb{E}[X|Y=3]\mathbb{P}(Y=3).$$

Maintenant, on a $\mathbb{E}[X|Y=1] = 3$ puisque qu'on sait qu'il va rentrer chez lui en 3 minutes. Ensuite on a $\mathbb{E}[X|Y=2] = 5 + \mathbb{E}[X]$ car il va marcher pendant 5 minutes puis revenir au point de départ et il devra encore faire le chemin pour rentrer chez lui. De la même façon, on a $\mathbb{E}[X|Y=3] = 7 + \mathbb{E}[X]$. Finalement, on a

$$\mathbb{E}[X] = \frac{1}{3}(15 + 2\mathbb{E}[X]) \iff \mathbb{E}[X] = 15.$$

EXEMPLE 88. Soit N une variable aléatoire discrète d'espérance $\mathbb{E}[N]$ et soient $X_1, \dots$ une famille de variables aléatoires i.i.d d'espérance μ et indépendantes de N . Calculons $\mathbb{E}\left[\sum_{i=1}^N X_i\right]$. On somme donc un nombre aléatoire de variables aléatoires et on prend ensuite l'espérance. On commence par écrire

$$\mathbb{E}\left[\sum_{i=1}^N X_i\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^N X_i \middle| N\right]\right] = \sum_{n=1}^{+\infty} \mathbb{E}\left[\sum_{i=1}^N X_i \middle| N=n\right] \mathbb{P}(N=n) = \sum_{n=1}^{+\infty} \sum_{i=1}^n \mathbb{E}[X_i|N=n] \mathbb{P}(N=n).$$

On utilise maintenant l'indépendance des X_i et de N pour écrire

$$\mathbb{E}\left[\sum_{i=1}^N X_i\right] = \sum_{n=1}^{+\infty} \sum_{i=1}^n \mathbb{E}[X_i] \mathbb{P}(N=n) = \sum_{n=1}^{+\infty} n\mu \mathbb{P}(N=n) = \mathbb{E}[N]\mu.$$

Dans ce calcul, on a calculé $h(n) = \mathbb{E}\left[\sum_{i=1}^N X_i|N=n\right] = n\mu$ et ainsi on obtient que

$$\mathbb{E}\left[\sum_{i=1}^N X_i \middle| N\right] = h(N) = N\mu.$$

EXEMPLE 89. On considère la densité jointe

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right).$$

Calculons $\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{\mathbb{E}[XY] - \mu_X\mu_Y}{\sigma_X\sigma_Y}$. On a vu dans une feuille d'exercice que la loi de X conditionnellement à $Y = y$ est donnée par $\mathcal{N}\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1-\rho^2)\right)$ et que la loi marginale de Y est une $\mathcal{N}(\mu_Y, \sigma_Y^2)$. Or on a

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY|Y]] = \int_{\mathbb{R}} \mathbb{E}[XY|Y=y] f_Y(y) dy$$

et

$$\mathbb{E}[XY|Y=y] = \mathbb{E}[Xy|Y=y] = y\mathbb{E}[X|Y=y] = y\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y)\right).$$

Ainsi,

$$\mathbb{E}[XY|Y] = Y\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(Y^2 - \mu_Y Y)$$

et finalement

$$\mathbb{E}[XY] = \mu_Y\mu_X + \rho\frac{\sigma_X}{\sigma_Y}\left(\mathbb{E}[Y^2] - \mu_Y^2\right) = \mu_Y\mu_X + \rho\frac{\sigma_X}{\sigma_Y}\text{Var}(Y) = \mu_Y\mu_X + \rho\sigma_X\sigma_Y.$$

Ainsi,

$$\text{Corr}(X,Y) = \frac{\mathbb{E}[XY] - \mu_X\mu_Y}{\sigma_X\sigma_Y} = \rho.$$

EXEMPLE 90. On peut aussi utiliser l'espérance conditionnelle pour calculer des variances. Par exemple si $N \sim \text{Geo}(p)$, recalculons la variance de N . On rappelle que N correspond au nombre d'épreuves pour

obtenir un succès ainsi si $Y = \mathbb{1}_{\text{premier lancer est un succès}}$, on peut calculer

$$\mathbb{E}[N^2] = \mathbb{E}[\mathbb{E}[N^2|Y]].$$

Mais on a $\mathbb{E}[N^2|Y=1] = 1$ et $\mathbb{E}[N^2|Y=0] = \mathbb{E}[(1+N)^2]$ car si la première épreuve est un échec, alors on est revenu au point de départ. Ainsi, on a

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

D'où,

$$\mathbb{E}[N^2] = \frac{1}{p} + \frac{2(1-p)}{p^2} = \frac{2-p}{p^2}.$$

Finalement,

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

EXEMPLE 91. Soit $U \sim \text{Unif}(0,1)$, on suppose que la loi de X conditionnellement à $U = p$ est $\text{Bin}(n, p)$. Quelle est la loi de X ?

On va conditionner sur les valeurs de U pour écrire

$$\begin{aligned} \mathbb{P}(X=i) &= \int_{\mathbb{R}} \mathbb{P}(X=i|U=p) f_U(p) dp = \int_0^1 \mathbb{P}(X=i|U=p) dp = \int_0^1 \binom{n}{i} p^i (1-p)^{n-i} dp \\ &= \frac{n!}{i!(n-i)!} \int_0^1 p^i (1-p)^{n-i} dp. \end{aligned}$$

On a vu dans (4) que

$$\int_0^1 p^i (1-p)^{n-i} dp = \frac{\Gamma(i+1)\Gamma(n-i+1)}{\Gamma(n+2)} = \frac{i!(n-i)!}{(n+1)!}.$$

Finalement, cela nous donne

$$\mathbb{P}(X=i) = \frac{n!}{i!(n-i)!} \cdot \frac{i!(n-i)!}{(n+1)!} = \frac{1}{n+1}.$$

Ainsi, X suit une loi uniforme sur $\{0, \dots, n\}$.

EXEMPLE 92. Soient X et Y deux variables aléatoires continues et indépendantes. Calculons $\mathbb{P}(X, Y)$. On a

$$\begin{aligned} \mathbb{P}(X < Y) &= \int_{\mathbb{R}} \mathbb{P}(X < Y|Y=y) f_Y(y) dy = \int_{\mathbb{R}} \mathbb{P}(X < y|Y=y) f_Y(y) dy = \int_{\mathbb{R}} \mathbb{P}(X < y) f_Y(y) dy \\ &= \int_{\mathbb{R}} F_X(y) f_Y(y) dy. \end{aligned}$$

On peut aussi utiliser le conditionnement pour retrouver le produit de convolution

$$\begin{aligned} \mathbb{P}(X+Y < z) &= \int_{\mathbb{R}} \mathbb{P}(X+Y < z|Y=y) f_Y(y) dy = \int_{\mathbb{R}} \mathbb{P}(X < z-y|Y=y) f_Y(y) dy \\ &= \int_{\mathbb{R}} \mathbb{P}(X < z-y) f_Y(y) dy \\ &= \int_{\mathbb{R}} F_X(z-y) f_Y(y) dy. \end{aligned}$$

En dérivant par rapport à z , on retrouve

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y) dy.$$

DÉFINITION 6.15

Soient X et Y deux variables aléatoires X et Y alors on définit la variance conditionnelle de X par rapport à Y ,

$$\text{Var}(X|Y) = \mathbb{E} \left[(X - \mathbb{E}[X|Y])^2 | Y \right]$$

THÉORÈME 6.16

On a

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y])$$

Démonstration. On peut aussi écrire $\text{Var}(X|Y) = \mathbb{E}[X^2|Y] - \mathbb{E}[X|Y]^2$ et ainsi

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

Mais on a

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

Finalement, on obtient

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

□

EXEMPLE 93. Le nombre de personnes qui se sont rendues à une gare ferroviaire au temps t est modélisé par une loi de Poisson de paramètre λt . Si le premier train arrive selon une loi $\text{Unif}(0, T)$ indépendante au nombre de passagers, calculons l'espérance et la variance du nombre de passagers dans le train.

On définit les variables aléatoires

N_t : Nombre de passagers dans la gare au temps t ,

Y : Heure d'arrivée du train.

On est donc intéressé par la variable aléatoire N_Y . On peut calculer

$$\mathbb{E}[N_Y|Y = t] = \mathbb{E}[N_t|Y = t] = \mathbb{E}[N_t] = \lambda t.$$

Ainsi, on a l'espérance conditionnelle $\mathbb{E}[N_Y|Y] = \lambda Y$ et on peut donc calculer l'espérance

$$\mathbb{E}[N_Y] = \mathbb{E}[\mathbb{E}[N_Y|Y]] = \mathbb{E}[\lambda Y] = \frac{\lambda T}{2}.$$

Pour calculer la variance de N_Y , on a

$$\text{Var}(N_Y|Y = t) = \text{Var}(N_t|Y = t) = \text{Var}(N_t) = \lambda t.$$

Et ainsi, $\text{Var}(N_Y|Y) = \lambda Y$. Par la formule que l'on vient de voir on a

$$\text{Var}(N_Y) = \mathbb{E}[\text{Var}(N_Y|Y)] + \text{Var}(\mathbb{E}[N_Y|Y]) = \mathbb{E}[\lambda Y] + \text{Var}(\lambda Y) = \frac{\lambda T}{2} + \lambda^2 \text{Var}(Y) = \frac{\lambda T}{2} + \frac{\lambda^2 T^2}{12}.$$

EXEMPLE 94. Soit N une variable aléatoire sur $\mathbb{N}$, $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires i.i.d telle que $\mathbb{E}[X_i] = \mu$ et $\text{Var}(X_i) = \sigma^2$. Calculons la variance de $\sum_{i=1}^N X_i$.

On sait que $\mathbb{E}[\sum_{i=1}^N X_i|N] = N\mathbb{E}[X] = N\mu$ et que $\text{Var}(\sum_{i=1}^N X_i|N) = N\sigma^2$. Ainsi, on peut utiliser la formule pour calculer

$$\text{Var}\left(\sum_{i=1}^N X_i\right) = \mathbb{E}\left[\text{Var}\left(\sum_{i=1}^N X_i\right)\middle|N\right] + \text{Var}\left(\mathbb{E}\left[\sum_{i=1}^N X_i\right]\middle|N\right) = \mathbb{E}[N]\sigma^2 + \mu^2 \text{Var}(N).$$

On se donne X et Y deux variables aléatoires et on veut prédire Y sur ce que l'on sait de X . si $g(X)$ est

le *prédicteur* de Y , on veut que cette quantité soit proche de Y . Une façon de quantifier ce “proche” est de vouloir minimiser

$$\mathcal{L}(g(X), Y) = \mathbb{E} \left[(Y - g(X))^2 \right].$$

Le théorème suivant nous donne que le meilleur prédicteur est en fait l’espérance conditionnelle.

THÉORÈME 6.17

Pour tout fonction g , on a

$$\mathbb{E} \left[(Y - g(X))^2 \right] \geq \mathbb{E} \left[(Y - \mathbb{E}[Y|X])^2 \right].$$

Démonstration. On calcule

$$\begin{aligned} \mathbb{E} \left[(Y - g(X))^2 | X \right] &= \mathbb{E} \left[(Y - \mathbb{E}[Y|X] + \mathbb{E}[Y|X] - g(X))^2 | X \right] \\ &= \mathbb{E} \left[(Y - \mathbb{E}[Y|X])^2 | X \right] + \mathbb{E} \left[(\mathbb{E}[Y|X] - g(X))^2 | X \right] + 2\mathbb{E} \left[(Y - \mathbb{E}[Y|X]) (\mathbb{E}[Y|X] - g(X)) | X \right]. \end{aligned}$$

Maintenant, on voit que $\mathbb{E}[Y|X] - g(X)$ est une fonction que de X est ainsi, si on conditionne sur les valeurs de X cela correspond à une constante et on peut donc le passer en dehors de l’espérance pour écrire

$$\begin{aligned} \mathbb{E} \left[(Y - \mathbb{E}[Y|X]) (\mathbb{E}[Y|X] - g(X))^2 | X \right] &= (\mathbb{E}[Y|X] - g(X)) \mathbb{E} \left[(Y - \mathbb{E}[Y|X]) | X \right] \\ &= (\mathbb{E}[Y|X] - g(X)) (\mathbb{E}[Y|X] - \mathbb{E}[Y|X]) = 0. \end{aligned}$$

Cela donne ainsi que

$$\mathbb{E} \left[(Y - g(X))^2 \right] = \mathbb{E} \left[(Y - \mathbb{E}[Y|X])^2 \right] + \mathbb{E} \left[(\mathbb{E}[Y|X] - g(X))^2 \right] \geq \mathbb{E} \left[(Y - \mathbb{E}[Y|X])^2 \right]$$

var le deuxième terme est positif. $\square$

EXEMPLE 95. Un signal est envoyé en A et est modélisé par $S \sim \mathcal{N}(\mu, \sigma^2)$. Il est perçu en B par R qui conditionnellement à $S = s$, $R \sim \mathcal{N}(s, 1)$. Si on sait que le signal reçu est $R = r$, quel est le meilleur prédicteur du signal envoyé?

Commençons par regarder la densité conditionnelle de S par rapport à $R = r$. On commence par la densité jointe

$$f_{S,R}(s, r) = f_S(s) f_{R|S}(r|s) = \frac{1}{2\pi\sigma} e^{-\frac{(s-\mu)^2}{2\sigma^2}} e^{-\frac{(r-s)^2}{2}} = \frac{1}{2\pi\sigma} \exp \left(-\frac{1}{2} \left(\frac{(s-\mu)^2}{\sigma^2} + (r-s)^2 \right) \right).$$

En développant, on va écrire

$$\begin{aligned} \frac{(s-\mu)^2}{\sigma^2} + (r-s)^2 &= s^2 \left(\frac{1}{\sigma^2} + 1 \right) - s \left(\frac{2\mu}{\sigma^2} + 2r \right) + \frac{\mu^2}{\sigma^2} + r^2 \\ &= \frac{1+\sigma^2}{\sigma^2} \left(s^2 - 2s \frac{\sigma^2}{1+\sigma^2} \left(\frac{\mu + r\sigma^2}{\sigma^2} \right) \right) + \frac{\mu^2}{\sigma^2} + r^2 \\ &= \frac{1+\sigma^2}{\sigma^2} \left(s^2 - 2s \frac{\mu + r\sigma^2}{1+\sigma^2} + \left(\frac{\mu + r\sigma^2}{1+\sigma^2} \right)^2 \right) - \frac{1+\sigma^2}{\sigma^2} \left(\frac{\mu + r\sigma^2}{1+\sigma^2} \right)^2 + \frac{\mu^2}{\sigma^2} + r^2 \\ &= \frac{1+\sigma^2}{\sigma^2} \left(s - \frac{\mu + r\sigma^2}{1+\sigma^2} \right)^2 - \frac{(\mu + r\sigma^2)^2}{\sigma^2(1+\sigma^2)} + \frac{\mu^2 + r^2\sigma^2}{\sigma^2} \end{aligned}$$

où on a développé et complété le carré. On obtient finalement la densité jointe

$$f_{S,R}(s, r) = \frac{1}{2\pi\sigma} \exp \left(-\frac{1}{2\sigma^2} \left(\mu^2 + r\sigma^2 - \frac{(\mu + r\sigma^2)^2}{1+\sigma^2} \right) \right) \exp \left(-\frac{\left(s - \frac{\mu + r\sigma^2}{1+\sigma^2} \right)^2}{\frac{\sigma^2}{1+\sigma^2}} \right).$$

On peut donc obtenir la loi marginale de R , on remarque qu'on a séparé la variable s de telle sorte que seule la seconde exponentielle sera intégrée et cela correspond à une intégrale gaussienne de variance $\frac{\sigma^2}{1+\sigma^2}$ et cela nous donne donc

$$\begin{aligned} f_R(r) &= \int_{\mathbb{R}} f_{S|R}(s, r) ds = \frac{1}{2\pi\sigma} \exp\left(-\frac{1}{2\sigma^2} \left(\mu^2 + r\sigma^2 - \frac{(\mu + r\sigma^2)^2}{1 + \sigma^2}\right)\right) \sqrt{2\pi \frac{\sigma^2}{1 + \sigma^2}} \\ &= \frac{1}{\sqrt{2\pi(1 + \sigma^2)}} \exp\left(-\frac{1}{2\sigma^2} \left(\mu^2 + r\sigma^2 - \frac{(\mu + r\sigma^2)^2}{1 + \sigma^2}\right)\right). \end{aligned}$$

Finalement, cela nous donne

$$f_{S|R}(s|r) = \frac{f_S(s)f_{R|S}(r|s)}{f_R(r)} = \frac{1}{\sqrt{2\pi \frac{\sigma^2}{1+\sigma^2}}} \exp\left(-\frac{\left(s - \frac{\mu + r\sigma^2}{1+\sigma^2}\right)^2}{2 \frac{\sigma^2}{1+\sigma^2}}\right).$$

On voit donc que

$$S|R = r \sim \mathcal{N}\left(\frac{\mu + r\sigma^2}{1 + \sigma^2}, \frac{\sigma^2}{1 + \sigma^2}\right)$$

et ainsi $\mathbb{E}[S|R = r] = \frac{\mu + r\sigma^2}{1 + \sigma^2}$. Le meilleur prédicteur est donc donné par

$$\mathbb{E}[S|R] = \frac{\mu}{1 + \sigma^2} + \frac{\sigma^2}{1 + \sigma^2} R.$$

6.2 Fonction génératrice des moments

DÉFINITION 6.18

Pour $t \in \mathbb{R}$ et X une variable aléatoire, on définit sa fonction génératrice des moments par

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

On dit qu'elle existe si il existe $\delta > 0$ tel que $M_X(t) < \infty$ pour tout $t \in (-\delta, \delta)$. On a

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

On remarque que pour certain t , on peut avoir $M(t) = +\infty$. On l'appelle fonction génératrice des moments pour la raison suivante, en dérivant par rapport à t , on a

$$M'_X(t) = \mathbb{E}[Xe^{tX}] \implies M'(0) = \mathbb{E}[X].$$

De même, on a

$$M''_X(t) = \mathbb{E}\left[\frac{d}{dt}(Xe^{tX})\right] = \mathbb{E}[X^2 e^{tX}] \implies M''_X(0) = \mathbb{E}[X^2].$$

On peut faire ceci ainsi de suite et bien avoir $M_X^{(n)}(0) = \mathbb{E}[X^n]$. On peut ainsi utiliser la fonction génératrice pour calculer des espérances ou des variances par exemple. Calculons la fonction génératrice des moments de certaines lois.

THÉORÈME 6.19

Si $W \sim \text{Bin}(n, p)$ alors

$$M_W(t) = (pe^t + 1 - p)^n \text{ pour tout } t \in \mathbb{R}$$

Si $X \sim \text{Poi}(\lambda)$ alors

$$M_X(t) = e^{\lambda(e^t - 1)} \text{ pour tout } t \in \mathbb{R}.$$

Si $Y \sim \text{Exp}(\lambda)$ alors

$$\mathbb{E} \left[e^{tX} \right] = \begin{cases} \frac{\lambda}{\lambda - t} & \text{si } t < \lambda, \\ +\infty & \text{sinon.} \end{cases}$$

Si $Z \sim \mathcal{N}(\mu, \sigma^2)$ alors

$$\mathbb{E} \left[e^{tX} \right] = e^{t\mu + \frac{t^2\sigma^2}{2}} \text{ pour tout } t \in \mathbb{R}.$$

Démonstration. Si $W \sim \text{Bin}(n, p)$ alors on a

$$\mathbb{E} \left[e^{tX} \right] = \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (e^t p)^k (1-p)^{n-k} = (e^t p + 1 - p)^n.$$

Si on veut retrouver l'espérance et la variance, on calcule

$$M'_W(t) = n p e^t (e^t p + 1 - p)^{n-1} \implies M'_W(0) = \mathbb{E}[X] = np.$$

Pour le second moment on dérive deux fois

$$M''_W(t) = n p e^t (e^t p + 1 - p)^{n-1} + n(n-1) p^2 e^{2t} (e^t p + 1 - p)^{n-2}$$

qui nous donne

$$M''_W(0) = [X^2] = np + n(n-1)p^2 = np(1 + (n-1)p).$$

En combinant ces deux calculs, on retrouve $\text{Var}(W) = np(1-p)$.

Pour $X \sim \text{Poi}(\lambda)$, on a

$$M_X(t) = \mathbb{E} \left[e^{tX} \right] = \sum_{k=0}^{+\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda + \lambda e^t}.$$

En dérivant, on retrouve

$$M'_X(t) = \lambda e^t e^{\lambda(e^t-1)} \implies M'_X(0) = \mathbb{E}[X] = \lambda.$$

Et encore

$$M''_X(t) = \lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t} e^{\lambda(e^t-1)} \implies M''_X(0) = \mathbb{E}[X^2] = \lambda + \lambda^2$$

et on retrouve $\text{Var}(X) = \lambda$.

Pour $Y \sim \text{Exp}(\lambda)$, on calcule

$$M_Y(t) = \mathbb{E} \left[e^{tY} \right] = \int_0^{+\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{(t-\lambda)x} dx = \left[\frac{\lambda}{t-\lambda} e^{(t-\lambda)x} \right]_0^{+\infty} = \begin{cases} \frac{\lambda}{\lambda-t} & \text{si } t < \lambda, \\ +\infty & \text{sinon.} \end{cases}$$

On remarque ainsi que la fonction génératrice n'est pas définie partout. On peut calculer facilement, pour $t < \lambda$,

$$M'_Y(t) = \frac{\lambda}{(\lambda-t)^2} \implies M'_Y(0) = \mathbb{E}[Y] = \frac{1}{\lambda}.$$

Et

$$M''_Y(t) = \frac{2\lambda}{(\lambda-t)^3} \implies M''_Y(0) = \mathbb{E}[Y^2] = \frac{2}{\lambda^2}$$

et on retrouve bien $\text{Var}(Y) = \frac{1}{\lambda^2}$.

Enfin, pour $Z_0 \sim \mathcal{N}(0, 1)$, on calcule

$$\begin{aligned} M_{Z_0}(t) = \mathbb{E} \left[e^{tZ_0} \right] &= \int_{\mathbb{R}} \frac{e^{tx - \frac{x^2}{2}}}{\sqrt{2\pi}} dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} (x^2 - 2tx + t^2 - t^2) \right) dx = e^{\frac{t^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx \\ &= e^{\frac{t^2}{2}}. \end{aligned}$$

Si $Z \sim \mathcal{N}(\mu, \sigma^2)$ alors on sait que Z a la même loi que $\sigma Z_0 + \mu$ et ainsi

$$M_Z(t) = \mathbb{E} \left[e^{t(\mu + \sigma Z_0)} \right] = e^{t\mu} \mathbb{E} \left[e^{t\sigma Z_0} \right] = e^{t\mu + \frac{t^2 \sigma^2}{2}}.$$

On peut dériver

$$M'_Z(t) = \left(\mu + \sigma^2 t \right) e^{t\mu + \frac{t^2 \sigma^2}{2}} \implies M'_Z(0) = \mathbb{E}[Z] = \mu.$$

Et

$$M''_Z(t) = \sigma^2 e^{t\mu + \frac{t^2 \sigma^2}{2}} + (\mu + \sigma^2 t)^2 e^{t\mu + \frac{t^2 \sigma^2}{2}} \implies M''_Z(0) = \mathbb{E}[Z^2] = \sigma^2 + \mu^2$$

et on retrouve $\text{Var}(Z) = \sigma^2$. □

Une propriété très importante de la fonction génératrice des moments est que, comme la fonction de répartition, elle caractérise la distribution. Nous énonçons ce théorème sans le prouver.

THÉORÈME 6.20

La fonction génératrice des moments caractérise la distribution de la variable aléatoire i.e. si X et Y sont deux variables aléatoires et si il existe $\delta > 0$ tel que pour tout $t \in (-\delta, \delta)$, $M_X(t), M_Y(t)$ existent et $M_X(t) = M_Y(t)$ alors pour tout $x \in \mathbb{R}$, $F_X(x) = F_Y(x)$

EXEMPLE 96. Si on sait que $M_X(t) = \frac{1}{2^{10}} (e^t + 1)^{10}$ alors on peut réécrire

$$M_X(t) = \left(\frac{1}{2} e^t + \frac{1}{2} \right)^{10} = \left(\frac{1}{2} e^t + \left(1 - \frac{1}{2} \right) \right)^{10}$$

et on remarque ainsi que $X \sim \text{Bin} \left(10, \frac{1}{2} \right)$.

EXEMPLE 97. Si on sait que $M_X(t) = e^{3(e^t - 1)}$ que vaut $\mathbb{P}(X = 0)$?

On sait que si $X \sim \text{Poi}(\lambda)$ alors $M_X(t) = e^{\lambda(e^t - 1)}$ et ainsi ici $X \sim \text{Poi}(3)$ et $\mathbb{P}(X = 0) = \frac{1}{e^3}$.

Nous avons aussi une propriété très importante sur la fonction génératrice des moments de la somme de variables aléatoires indépendantes.

THÉORÈME 6.21

Si X et Y sont deux variables aléatoires indépendantes alors

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Démonstration. La preuve vient simplement du fait que

$$\mathbb{E} \left[e^{t(X+Y)} \right] = \mathbb{E} \left[e^{tX} e^{tY} \right] = \mathbb{E} \left[e^{tX} \right] \mathbb{E} \left[e^{tY} \right] = M_X(t) M_Y(t).$$

□

EXEMPLE 98. On peut utiliser ce résultat pour par exemple retrouver la formule de la fonction génératrice des moments d'une loi binomiale de paramètre (n, p) . On sait que X est le nombre de succès sur n épreuves indépendantes et ainsi

$$X = \sum_{i=1}^n X_i \quad \text{avec} \quad X_i = \begin{cases} 1 & \text{si le } i\text{-ième lancer est un succès,} \\ 0 & \text{sinon.} \end{cases}$$

Or on calcule facilement

$$\mathbb{E} \left[e^{tX_i} \right] = 1 - p + e^t p$$

d'où on obtient

$$M_X(t) = \prod_{i=1}^n M_{X_i}(t) = (1 - p + e^t p)^n.$$

De plus, on remarque aussi que si $Y \sim \text{Bin}(m, p)$ et est indépendante de X alors

$$M_{X+Y}(t) = M_X(t)M_Y(t) = (pe^t + 1 - p)^n (pe^t + 1 - p)^m = (pe^t + 1 - p)^{n+m}.$$

Comme la fonction génératrice des moments caractérise la distribution, on sait alors que $X + Y \sim \text{Bin}(n + m, p)$.

EXEMPLE 99. Si $X \sim \text{Poi}(\lambda)$ et $Y \sim \text{Poi}(\mu)$ avec X et Y indépendantes alors

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\lambda(e^t-1)} e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$$

et on remarque ainsi que $X + Y \sim \text{Poi}(\lambda + \mu)$.

EXEMPLE 100. On peut faire la même avec la somme de deux variables gaussiennes. Si $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ et X et Y sont indépendantes alors

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\mu_1 t + \frac{t^2 \sigma_1^2}{2}} e^{\mu_2 t + \frac{t^2 \sigma_2^2}{2}} = e^{(\mu_1 + \mu_2)t + \frac{t^2(\sigma_1^2 + \sigma_2^2)}{2}}$$

et on voit donc facilement que $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

EXEMPLE 101. On considère $X \sim \chi_n^2$, on rappelle que cela veut dire que $X = \sum_{i=1}^n X_i^2$ avec (X_i) une famille de variables aléatoires i.i.d normales centrées réduites. Calculons la fonction génératrice de X . Par indépendance, on sait que

$$M_X(t) = \prod_{i=1}^n M_{X_i}(t) = (M_{X_1}(t))^n$$

car les X_i ont tous la même loi. Ensuite on calcule

$$M_{X_1}(t) = \mathbb{E} \left[e^{tX_1^2} \right] = \int_{\mathbb{R}} \frac{e^{tx^2 - \frac{x^2}{2}}}{\sqrt{2\pi}} dx = \int_{\mathbb{R}} \frac{e^{-\frac{1}{2}(1-2t)x^2}}{\sqrt{2\pi}} dx.$$

Si on pose $\sigma^2 = \frac{1}{1-2t} > 0$ pour $t < \frac{1}{2}$ alors on peut calculer

$$\int_{\mathbb{R}} \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}} dx = \sigma = \frac{1}{\sqrt{1-2t}}.$$

Si $t \geq \frac{1}{2}$, on voit que l'intégrale est infinie. Enfin, on obtient pour la loi χ_n^2 ,

$$M_X(t) = \begin{cases} \frac{1}{(1-2t)^{\frac{n}{2}}} & \text{si } t < \frac{1}{2} \\ +\infty & \text{sinon.} \end{cases}$$

EXEMPLE 102. On peut reprendre l'exemple où on a une loi $Y \sim \text{Unif}(0, 1)$ et on suppose que X conditionnellement à $Y = p$ suit une loi $\text{Bin}(n, p)$. On avait vu que $X \sim \text{Unif}(\{0, \dots, n\})$, retrouvons le avec la fonction génératrice des moments. On a

$$\mathbb{E} \left[e^{tX} \middle| Y = p \right] = (pe^t + 1 - p)^n.$$

Et ainsi,

$$\mathbb{E} \left[e^{tX} \right] = \mathbb{E} \left[\mathbb{E} \left[e^{tX} \middle| Y \right] \right] = \int_0^1 (pe^t + 1 - p)^n dp.$$

En faisant le changement de variable $y = pe^t + 1 - p$ avec $dy = (e^t - 1) dp$, on obtient

$$\mathbb{E} [e^{tX}] = \frac{1}{e^t - 1} \int_1^{e^t} y^n dy = \frac{1}{n+1} \cdot \frac{e^{(n+1)t} - 1}{e^t - 1} = \frac{1}{n+1} \sum_{k=0}^n e^{kt} = \sum_{k=0}^n e^{kt} \mathbb{P}(U = k) = \mathbb{E} [e^{tU}]$$

avec $U \sim \text{Unif}(\{0, \dots, n\})$.

On peut aussi définir une version simultanée de la fonction génératrice des moments par la définition suivante.

DÉFINITION 6.22

Pour $X_1, \dots, X_n$ n variables aléatoires on définit leur fonction génératrice des moments conjoints M_n par

$$M_n(\theta_1, \dots, \theta_n) = \mathbb{E} \left[\exp \left(\sum_{i=1}^n \theta_i X_i \right) \right].$$

En particulier si on écrit

$$\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \quad \text{et} \quad \vec{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}$$

alors on peut écrire

$$M_{\vec{X}}(\vec{\theta}) = \mathbb{E} [e^{\langle \vec{\theta}, \vec{X} \rangle}].$$

De même, si $M_{\vec{X}}(\vec{\theta})$ existe pour $\|\vec{\theta}\| < \delta$ alors elle caractérise la distribution conjointe de $X_1, \dots, X_n$ et de plus si $X_1, \dots, X_n$ sont indépendantes alors

$$M_{\vec{X}}(\vec{\theta}) = \prod_{i=1}^n M_{X_i}(\theta_i).$$

On remarque que l'on peut retrouver M_{X_i} en prenant $\theta_j = 0$ pour $i \neq j$ et $\theta_i = t$ puisqu'on retrouve

$$M_n(0, \dots, 0, t, 0, \dots, 0) = \mathbb{E} [e^{tX_i}] = M_{X_i}(t).$$

EXEMPLE 103. Soient X et Y deux variables aléatoires i.i.d $\mathcal{N}(\mu, \sigma^2)$. Montrons que $X + Y$ et $X - Y$ sont indépendantes et donnons leur loi. On peut calculer la fonction génératrices des moments conjoints

$$\begin{aligned} \mathbb{E} [e^{\theta_1(X+Y) + \theta_2(X-Y)}] &= \mathbb{E} [e^{(\theta_1 + \theta_2)X}] \mathbb{E} [e^{(\theta_1 - \theta_2)Y}] \\ &= \exp \left(\mu(\theta_1 + \theta_2) + \frac{(\theta_1 + \theta_2)^2 \sigma^2}{2} \right) \exp \left(\mu(\theta_1 - \theta_2) + \frac{(\theta_1 - \theta_2)^2 \sigma^2}{2} \right) \\ &= \exp (2\theta_1 \mu + \theta_1^2 \sigma^2) \exp (\sigma^2 \theta_2^2) \\ &= \mathbb{E} [e^{\theta_1 Z_1}] \mathbb{E} [e^{\theta_2 Z_2}] \end{aligned}$$

avec $Z_1 \sim \mathcal{N}(2\mu, 2\sigma^2) \sim X + Y$ et $Z_2 \sim \mathcal{N}(0, 2\sigma^2) \sim X - Y$ et on remarque aussi que $X + Y$ et $X - Y$ sont bien indépendantes.

7. THÉORÈMES LIMITES

7.1 Inégalités de probabilité

On commence par donner quelques inégalités très générales et très utiles.

THÉORÈME 7.1: Inégalité de Markov

Soit X une variable aléatoire avec $X \geq 0$ et $a > 0$ alors

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

Démonstration. Soit $a > 0$, on considère $\mathbb{1}_{X \geq a}$ alors on a

$$\mathbb{1}_{X \geq a} \leq \frac{X}{a}.$$

En effet, si $X < a$ alors l'inégalité devient $\frac{X}{a} \geq 0$ qui est vrai et si $X \geq a$, on obtient $1 \leq \frac{X}{a}$ qui est aussi vrai. En prenant l'espérance des deux côtés on obtient

$$\mathbb{E}[\mathbb{1}_{X \geq a}] \leq \mathbb{E}\left[\frac{X}{a}\right] \iff \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

□

THÉORÈME 7.2: Inégalité de Bienaymé–Tchebyshev

Soit X une variable aléatoire d'espérance μ et de variance σ^2 alors

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}.$$

Démonstration. On utilise l'inégalité de Markov pour écrire

$$\mathbb{P}(|X - \mu| \geq k) = \mathbb{P}(|X - \mu|^2 \geq k^2) \leq \frac{\mathbb{E}[|X - \mu|^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

□

Ces inégalités sont très utiles car elles permettent d'obtenir des bornes en ne connaissant que l'espérance ou la variance de variables aléatoires.

EXEMPLE 104. Une usine fabrique des composants et en fabrique 50 par semaine en moyenne. Bornons la probabilité que l'usine fasse 75 pièces la semaine prochaine.

Si on dénote X le nombre de composants par semaine alors on sait que $\mathbb{E}[X] = 50$ et ainsi

$$\mathbb{P}(X > 75) \leq \frac{\mathbb{E}[X]}{75} = \frac{50}{75} = \frac{2}{3}.$$

On sait que la variance de ce nombre de composants par semaine est 75. Bornons la probabilité que la production soit entre 40 et 60.

$$\mathbb{P}(40 < X < 60) = \mathbb{P}(-10 < X - 50 < 10) = \mathbb{P}(|X - 50| < 10).$$

Or par l'inégalité de Bienaymé–Tchebyshev, nous avons

$$\mathbb{P}(|X - 50| > 10) \leq \frac{\text{Var}(X)}{100} = \frac{75}{100} = \frac{3}{4}$$

et ainsi

$$\mathbb{P}(|X - 50| < 10) \geq \frac{1}{4}.$$

Il faut bien faire attention que ces bornes ne sont que des inégalités et pas des égalités.

EXEMPLE 105. Soit $X \sim \text{Unif}(0, 10)$ alors nous savons que $\mathbb{E}[X] = 5$ et $\text{Var}(X) = \frac{25}{3}$ et nous avons par l'inégalité de Bienaymé–Tchebyshev

$$\mathbb{P}(|X - 5| > 4) \leq \frac{25}{48} \approx 0.52.$$

Mais on peut aussi le calculer directement et

$$\mathbb{P}(|X - 5| > 4) = \mathbb{P}(X > 9) + \mathbb{P}(X \leq 1) = \int_9^{10} \frac{dx}{10} + \int_0^1 \frac{dx}{10} = \frac{1}{5} = 0.2.$$

On a bien sur $0.2 \leq 0.52$ mais la borne n'est pas du tout précise.

EXEMPLE 106. Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors on obtient

$$\mathbb{P}(|X - \mu| \geq 2\sigma) \leq \frac{\sigma^2}{4\sigma^2} = \frac{1}{4}.$$

Mais on a

$$\mathbb{P}(|X - \mu| > 2\sigma) = \mathbb{P}\left(\left|\frac{X - \mu}{\sigma}\right| > 2\right) = \mathbb{P}\left(\frac{X - \mu}{\sigma} > 2\right) + \mathbb{P}\left(\frac{X - \mu}{\sigma} < -2\right) = 2(1 - \Phi(2)) \approx 0.0456.$$

De même, la borne est très loin de la vérité.

Ces bornes sont cependant très utiles d'un point de vue théorique lorsque l'on ne connaît pas exactement les variables aléatoires.

THÉORÈME 7.3

Soit X une variable aléatoire telle que $\text{Var}(X) = 0$ alors $\mathbb{P}(X = \mathbb{E}[X]) = 1$.

Démonstration. Soient $n \geq 1$ et $\mu := \mathbb{E}[X]$, alors

$$\mathbb{P}\left(|X - \mu| > \frac{1}{n}\right) \leq \text{Var}(X)n = 0$$

Donc par la continuité des probabilités, on a

$$0 = \lim_{n \rightarrow \infty} \mathbb{P}\left(|X - \mu| > \frac{1}{n}\right) = \mathbb{P}\left(\lim_{n \rightarrow \infty} \left\{|X - \mu| > \frac{1}{n}\right\}\right) = \mathbb{P}(X \neq \mu).$$

□

L'inégalité de Bienaymé–Tchebyshev nous donne une borne pour un écart à la moyenne des deux côtés. Si on en veut qu'un, on peut bien sur faire

$$\mathbb{P}(X - \mu \geq k) \leq \mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

mais on peut faire mieux!

THÉORÈME 7.4: Inégalité unilatérale de Tchebyshev

Soit X une variable aléatoire d'espérance μ et de variance σ^2 alors pour tout $a > 0$,

$$\mathbb{P}(X \geq a + \mu) \leq \frac{\sigma^2}{a^2 + \sigma^2}, \quad \mathbb{P}(X \leq -a + \mu) \leq \frac{\sigma^2}{a^2 + \sigma^2}.$$

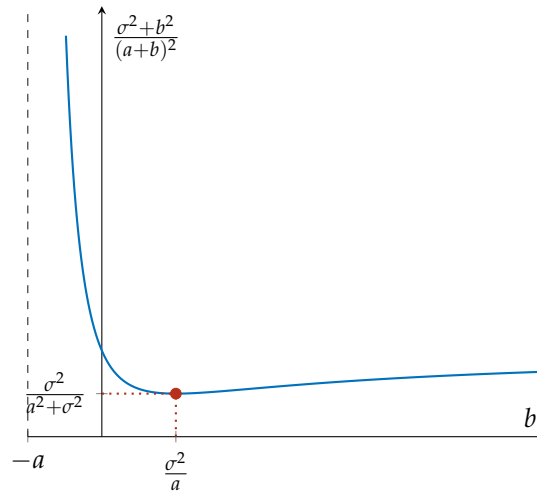
Démonstration. Posons $Y = X - \mu$ alors pour $a > 0$, on cherche, pour $b > 0$,

$$\mathbb{P}(Y \geq a) = \mathbb{P}(Y + b \geq a + b) = \mathbb{P}\left((Y + b)^2 \geq (a + b)^2\right) \leq \frac{\mathbb{E}[(Y + b)^2]}{(a + b)^2} = \frac{\sigma^2 + b^2}{(a + b)^2}.$$

On peut maintenant optimiser en b , si on dérive

$$\frac{d}{db} \left(\frac{\sigma^2 + b^2}{(a + b)^2} \right) = \frac{2b(a + b)^2 - 2(a + b)(\sigma^2 + b^2)}{(a + b)^4} = \frac{2b(a + b) - 2\sigma^2 - 2b^2}{(a + b)^3} = \frac{2(ab - \sigma^2)}{(a + b)^3}$$

et on voit que le seul point critique est donné par $x = \frac{\sigma^2}{a}$. On peut vérifier facilement que c'est un minimum.



Ainsi en choisissant ce b , on obtient

$$\mathbb{P}(Y \geq a) \leq \frac{\sigma^2 + \frac{\sigma^4}{a^2}}{\left(a + \frac{\sigma^2}{a}\right)^2} = \frac{a^2\sigma^2 + \sigma^4}{(a^2 + \sigma^2)^2} = \frac{\sigma^2}{a^2 + \sigma^2}.$$

□

EXEMPLE 107. On reprend l'exemple de l'usine qui produit en moyenne 50 composants par semaine avec une variance de 75. Par l'inégalité de Markov, on avait calculé

$$\mathbb{P}(X > 75) \leq \frac{2}{3} \approx 0.667.$$

Avec l'inégalité unilatérale de Tchebyshev on obtient

$$\mathbb{P}(X > 75) = \mathbb{P}(X > 50 + 25) \leq \frac{75}{75 + 25^2} = \frac{75}{700} = \frac{3}{28} \approx 0.107.$$

On voit que l'inégalité est beaucoup plus puissante dans ce cas.

THÉORÈME 7.5: Bornes de Chernoff

Soit X une variable aléatoire alors pour tout $a \in \mathbb{R}$,

$$\mathbb{P}(X \geq a) \leq e^{-at} \mathbb{E}[e^{tX}] \text{ pour tout } t > 0,$$

$$\mathbb{P}(X \leq a) \leq e^{-at} \mathbb{E}[e^{tX}] \text{ pour tout } t < 0.$$

Démonstration. On peut écrire pour $t > 0$,

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{at}) \leq e^{-at} \mathbb{E}[e^{tX}]$$

en utilisant l'inégalité de Markov. Si on choisit $t < 0$, alors on a

$$\mathbb{P}(X \leq a) = \mathbb{P}(e^{tx} \geq e^{ta}) \leq e^{-at} \mathbb{E}[e^{tX}]$$

par Markov encore.

□

EXEMPLE 108. Regardons les bornes de Chernoff pour une variable normale centrée réduite X . On sait que

$$\mathbb{E}[e^{tX}] = e^{\frac{t^2}{2}}.$$

Ainsi,

$$\mathbb{P}(X \geq a) \leq e^{-at} e^{\frac{t^2}{2}} = e^{\frac{t^2}{2} - at}.$$

On cherche le t qui minimise cette fonction avec $t > 0$ et on voit que cela correspond à $t = a$. Ainsi si $a > 0$,

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

De même, si $a < 0$ alors

$$\mathbb{P}(X \leq a) \leq e^{-\frac{a^2}{2}}.$$

EXEMPLE 109. Un joueur joue à un jeu : à chaque coup il gagne 1\$ avec probabilité $\frac{1}{2}$ et perd 1\$ avec probabilité $\frac{1}{2}$ et les coups sont indépendants. On cherche à borner son gain total sur n coups si X_i est la variable aléatoire donnée par $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$,

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq a\right) \leq ?$$

En utilisant la borne de Chernoff, on doit calculer la fonction génératrice des moments et on a

$$\mathbb{E}\left[e^{tX_i}\right] = \frac{1}{2}e^t + \frac{1}{2}e^{-t} = \cosh(t).$$

On peut essayer de borner cette fonction en écrivant

$$e^t = \sum_{k=0}^{+\infty} \frac{t^k}{k!}, \quad e^{-t} = \sum_{k=0}^{+\infty} (-1)^k \frac{t^k}{k!}$$

et ainsi en utilisant $(2\ell)! \geq 2^\ell \ell!$,

$$\cosh(t) = \frac{1}{2} \sum_{k=0}^{+\infty} \left(1 + (-1)^k\right) \frac{t^k}{k!} = \sum_{\ell=0}^{+\infty} \frac{t^{2\ell}}{(2\ell)!} \leq \sum_{\ell=0}^{+\infty} \frac{t^{2\ell}}{2^\ell \ell!} = \sum_{\ell=0}^{+\infty} \frac{\left(\frac{t^2}{2}\right)^\ell}{\ell!} = e^{\frac{t^2}{2}}.$$

Ainsi par indépendance,

$$\mathbb{E}\left[e^{t \sum_{i=1}^n X_i}\right] \leq e^{\frac{nt^2}{2}}.$$

En utilisant Chernoff, on peut alors borner

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq a\right) \leq e^{-at + \frac{nt^2}{2}}.$$

Ceci étant vrai pour tout $t > 0$, on choisit le meilleur et on voit que ceci est atteint pour $t = \frac{a}{n}$ et on obtient la borne

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq a\right) \leq e^{-\frac{a^2}{2n}}.$$

Par exemple, on voit que

$$\mathbb{P}\left(\sum_{i=1}^{10} X_i \geq 6\right) \leq e^{-\frac{36}{20}} \approx 0.1653.$$

Ceci est bien sur qu'une borne, on peut aussi directement calculer cette probabilité. Pour avoir la somme qui fait plus de 6, on doit nécessairement avoir au moins 8 gains de 1\$, on a donc trois possibilités 8, 9 ou 10 gains et on calcule donc

$$\mathbb{P}\left(\sum_{i=1}^{10} X_i \geq 6\right) = \frac{1}{2^{10}} \left(\binom{10}{8} + \binom{10}{9} + \binom{10}{10} \right) = \frac{45 + 10 + 1}{1024} = \frac{7}{120} \approx 0.055.$$

7.2 Loi des grands nombres

Nous allons voir notre premier théorème limite qui nous donne que la moyenne empirique d'une famille de variables aléatoires i.i.d converge vers l'espérance des variables aléatoires.

THÉORÈME 7.6: Loi faible des grands nombres

Soient (X_i) une suite de variables aléatoires i.i.d d'espérance μ alors pour tout $\varepsilon > 0$,

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0.$$

Démonstration. On suppose que $\text{Var}(X_i) = \sigma^2 < \infty$,

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n} \sum_{i=1}^n \mu = \mu.$$

On a aussi

$$\text{Var} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

Ainsi par l'inégalité de Bienaymé–Tchebyshev,

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

□

Même si on a utilisé la variance des variables aléatoires X_i , le théorème reste vrai sans cette hypothèse mais la preuve est plus compliquée. Ici, on parle de convergence en probabilité, on écrit parfois

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mu.$$

Il y a d'autre forme de convergence, par exemple on a le suivant

THÉORÈME 7.7: Loi forte des grands nombres

Soit (X_i) une suite de variables aléatoire i.i.d d'espérance μ alors

$$\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{} \mu \right) = 1.$$

Quelle est la différence entre pour tout $\varepsilon > 0$,

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0$$

et

$$\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{} \mu \right) = 1.$$



La première version nous dit que si N est assez grand alors avec grande probabilité $\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| < \varepsilon$. En revanche, cela ne nous dit pas que pour tout $n \geq N$, $\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| < \varepsilon$, il pourrait avoir de grands écarts entre $\frac{1}{n} \sum_{i=1}^n X_i$ et μ pour certains n , et même possiblement une infinité, mais ceci arriverait avec faible probabilité. La deuxième version nous donne que $\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right|$ ne peut être plus grand que ε qu'un nombre fini de fois avec probabilité 1.

Démonstration. La preuve dans le cas général est compliquée. On va supposer que $\mathbb{E}[X_i^4] < +\infty$ et ainsi $\mathbb{E}[X_i^2] < \infty$. Pour simplifier, sans perte de généralité, on suppose que $\mathbb{E}[X_i] = 0$ et on dénote $S_n = \sum_{i=1}^n X_i$. On a alors

$$\mathbb{E}[S_n^4] = \mathbb{E}\left[\sum_{i,j,k,\ell=1}^n X_i X_j X_k X_\ell\right].$$

Mais les i, j, k, ℓ peuvent être égaux, on a donc 5 types de termes de la forme

$$X_i^4, X_i^3 X_j, X_i^2 X_j^2, X_i^2 X_j X_k, X_i X_j X_k X_\ell.$$

Par indépendance, on a

$$\mathbb{E}[X_i^3 X_j] = \mathbb{E}[X_i^3] \mathbb{E}[X_j] = 0 = \mathbb{E}[X_i^2 X_j X_k] = \mathbb{E}[X_i X_j X_k X_\ell].$$

Il ne reste donc que les termes de la forme $X_i^2 X_j^2$ et X_i^4 . Si (i, j) est donné avec $i < j$ alors on a $\binom{4}{2} = 6$ termes qui donnent tous $X_i^2 X_j^2$ puisqu'on choisit deux des termes et on leur donne la valeur i et les autres la valeur j . Ainsi, on peut écrire

$$\mathbb{E}[S_n^4] = \sum_{i=1}^n \mathbb{E}[X_i^4] + 6 \sum_{i < j} \mathbb{E}[X_i^2 X_j^2] = \sum_{i=1}^n \mathbb{E}[X_i^4] + 3n(n-1) \mathbb{E}[X_i^2 X_j^2] = n \mathbb{E}[X_i^4] + 3n(n-1) \sigma^4.$$

On remarque que

$$0 \leq \text{Var}(X_i^2) = \mathbb{E}[X_i^4] - \mathbb{E}[X_i^2]^2 \iff \mathbb{E}[X_i^2]^2 = \sigma^4 \leq \mathbb{E}[X_i^4]$$

et ainsi on a la borne

$$\mathbb{E}[S_n^4] \leq (n + 3n(n-1)) \sigma^4$$

que l'on réécrit

$$\mathbb{E}\left[\frac{S_n^4}{n^4}\right] \leq \frac{\sigma^4}{n^3} + \frac{3\sigma^4}{n^2}.$$

On remarque donc que

$$\sum_{n=1}^{+\infty} \mathbb{E}\left[\frac{S_n^4}{n^4}\right] = \mathbb{E}\left[\sum_{n=1}^{+\infty} \frac{S_n^4}{n^4}\right] < \infty$$

où on a pu échanger la somme et l'intégrale car les variables aléatoires sont positives. Comme l'espérance de la variable aléatoire est finie alors on voit qu'elle est finie presque sûrement et ainsi

$$\mathbb{P}\left(\sum_{n=1}^{+\infty} \frac{S_n^4}{n^4} < +\infty\right) = 1$$

Mais la série est finie que si le terme général tend vers 0 et on obtient ainsi

$$\mathbb{P}\left(\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{} 0\right) = 1.$$

□

EXEMPLE 110. On suppose que la probabilité d'avoir un garçon est $\frac{1}{2}$ et on suppose que les couples font des enfants jusqu'à obtenir un garçon. On veut évaluer la proportion de garçons dans une génération. On dénote X le nombre d'enfants d'un couple pris au hasard alors on voit que $X \sim \text{Geo}\left(\frac{1}{2}\right)$ et ainsi $\mathbb{P}(X = k) = \frac{1}{2^k}$. On suppose que l'on a N couples et on dénote $X_1, \dots, X_N$ le nombre d'enfants respectifs. Alors la proportion de garçons s'écrit, comme on a un garçon par couple,

$$\frac{N}{X_1 + \dots + X_N}.$$

Par la loi forte des grands nombres on sait que

$$\mathbb{P} \left(\frac{N}{X_1 + \dots + X_N} \xrightarrow[N \rightarrow \infty]{} \frac{1}{\mathbb{E}[X]} \right) = 1$$

mais $\mathbb{E}[X] = 2$ d'où la proportion tend vers $\frac{1}{2}$. Il n'y a donc pas de déséquilibre!

7.3 Théorème central limite

Le deuxième théorème limite correspond à la limite des fluctuations d'une somme de variables aléatoires i.i.d autour de sa moyenne.

THÉORÈME 7.8: Théorème central limite

Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d d'espérance μ et de variance σ^2 alors la distribution de $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$ converge vers la distribution $\mathcal{N}(0, 1)$ dans le sens suivant, pour tout $x \in \mathbb{R}$,

$$\mathbb{P} \left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \leq x \right) \xrightarrow[n \rightarrow \infty]{} \int_{-\infty}^x \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du.$$

Ici on a convergence des fonctions de répartition. On appelle ceci la convergence en distribution, pour la prouver on va utiliser le théorème suivant dont on omettra la preuve.

THÉORÈME 7.9

Soit $(Z_i)_{i \geq 1}$ une suite de variable aléatoire de fonction de répartition F_{Z_i} et de fonction génératrice des moments M_{Z_n} . Soit Z une variable aléatoire de fonction de répartition F_Z et de fonction génératrice des moments M_Z alors si

$$M_{Z_n}(t) \xrightarrow[n \rightarrow \infty]{} M_Z(t) \text{ pour tout } t \implies F_{Z_n}(x) \xrightarrow[n \rightarrow \infty]{} F_Z(x) \text{ pour tout } x \text{ tel que } F_Z \text{ est continue.}$$

Ici, la fonction génératrice des moments de la limite est donnée par $e^{\frac{t^2}{2}}$, il suffit donc de montrer que $M_{Z_n}(t) \xrightarrow[n \rightarrow \infty]{} e^{\frac{t^2}{2}}$ pour tout $t \in \mathbb{R}$ et on obtiendra directement le résultat.

Démonstration. Pour simplifier, et sans perte de généralité on va supposer que $\mu = 0$ et $\sigma^2 = 1$. On veut donc montrer que $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ converge en distribution vers une loi normale centrée réduite. Dénotons $M(t) = \mathbb{E}[e^{tX_i}]$ alors on a

$$\mathbb{E} \left[e^{\frac{t}{\sqrt{n}} \sum_{i=1}^n X_i} \right] = \mathbb{E} \left[e^{\frac{tX_1}{\sqrt{n}}} \right]^n = M \left(\frac{t}{\sqrt{n}} \right)^n.$$

On veut donc montrer que

$$M \left(\frac{t}{\sqrt{n}} \right)^n \xrightarrow[n \rightarrow \infty]{} e^{\frac{t^2}{2}} \iff n \ln \left(M \left(\frac{t}{\sqrt{n}} \right) \right) \xrightarrow[n \rightarrow \infty]{} \frac{t^2}{2}.$$

Dénotons $L(t) = \ln(M(t))$ alors on a $L(0) = 0$, $L'(0) = \frac{M'(0)}{M(0)} = \frac{\mu}{1} = 0$ et on a

$$L''(0) = \frac{M(0)M''(0) - M'(0)^2}{M(0)^2} = M''(0) = \mathbb{E}[X_1^2] = 1.$$

Ainsi, on utilise la règle de de l'Hopital pour écrire

$$\lim_{n \rightarrow \infty} nL \left(\frac{t}{\sqrt{n}} \right) = \lim_{n \rightarrow \infty} \frac{L \left(\frac{t}{\sqrt{n}} \right)}{\frac{1}{n}} = \lim_{n \rightarrow \infty} \frac{t}{2n^{\frac{3}{2}}} \frac{L' \left(\frac{t}{\sqrt{n}} \right)}{\frac{1}{n^2}} = \lim_{n \rightarrow \infty} \frac{tL' \left(\frac{t}{\sqrt{n}} \right)}{2n^{-\frac{1}{2}}}.$$

On utilise la règle de de l'Hopital encore une fois,

$$\lim_{n \rightarrow \infty} nL\left(\frac{t}{\sqrt{n}}\right) = \lim_{n \rightarrow \infty} \frac{L''\left(\frac{t}{\sqrt{n}}n^{-\frac{3}{2}}\right)t^2}{2n^{-\frac{3}{2}}} = \lim_{n \rightarrow \infty} \frac{t^2}{2} L''\left(\frac{t}{\sqrt{n}}\right) = \frac{t^2}{2}$$

par continuité. $\square$

EXEMPLE 111. Un scientifique fait des mesures qu'on suppose être indépendantes et il prend la moyenne des mesures sur un grand échantillon. Il suppose que X_i a $\mathbb{E}[X_i] = d$ et $\mathbb{E}[(X_i - d)^2] = 4$. Combien de mesures doit-il réaliser pour être à 95% sûr d'être à 0.5 de d . La première méthode est d'utiliser l'inégalité de Bienaymé–Tchebyshev

$$\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = d, \quad \text{Var}\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \frac{4}{n}.$$

Ainsi,

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - d\right| > 0.5\right) \leq \frac{16}{n}.$$

On cherche donc n de telle façon que $\frac{16}{n} \leq 0.05$ qui nous donne $n = 320$.

On peut aussi réaliser l'approximation avec le théorème central limite en écrivant

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - d\right| < 0.5\right) = \mathbb{P}\left(\left|\frac{1}{2\sqrt{n}} \sum_{i=1}^n (X_i - d)\right| < \frac{\sqrt{n}}{4}\right) \approx 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1.$$

On cherche alors l'entier tel que $2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \geq 0.95$ et on obtient $n = 62$ ce qui est beaucoup moins que l'inégalité de Tchebyshev. Cependant, il y a un problème à cette méthode, on n'a pas une approximation quantitative donc on ne sait pas si $n = 62$ est assez grand pour faire l'approximation par le théorème central limite. Dans le cas de l'inégalité de Tchebyshev, on en est sûr! Une version quantitative du théorème central limite existe sous la forme suivante.

THÉORÈME 7.10

Soit (X_i) une suite de variables aléatoires i.i.d avec $\mathbb{E}[X] = \mu$, $\text{Var}(X) = \sigma^2$ et $\mathbb{E}[|X_i|^3] < \infty$ alors pour tout $a \in \mathbb{R}$, on a

$$\left|\mathbb{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu) < a\right) - \mathbb{P}(\mathcal{N}(0,1) < a)\right| \leq \frac{\mathbb{E}[|X_i|^3]}{2\sqrt{n}}.$$

EXEMPLE 112. On lance 10 dés équilibrés. On cherche la probabilité que la somme soit entre 30 et 40 en utilisant le théorème central limite. On dénote X_i le résultat du i -ième dé et alors $X = \sum_{i=1}^{10} X_i$. On calcule la moyenne et la variance

$$\mathbb{E}[X_i] = \frac{7}{2}, \quad \text{Var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2.$$

Or

$$\mathbb{E}[X_i^2] = \frac{91}{6}$$

et ainsi

$$\text{Var}(X_i) = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

On centre et on normalise

$$Z_{10} = \sqrt{\frac{12}{350}} \left(\sum_{i=1}^{10} X_i - \frac{70}{2}\right) \approx \mathcal{N}(0,1).$$

Ainsi,

$$\mathbb{P}(30 < X < 40) = \mathbb{P}\left(\sqrt{\frac{12}{350}}(30 - 35) < Z_{10} < \sqrt{\frac{12}{350}}(40 - 35)\right) \approx 2\Phi\left(5\sqrt{\frac{12}{350}}\right) - 1 \approx 0.6456.$$

EXEMPLE 113. Un enseignant dispose de 150 copies. On suppose que les temps de correction sont indépendants et de moyenne 20 minutes et d'écart-type 4 minutes. Approximons la probabilité que l'enseignant corrige 25 copies pendant les 7 premières heures. On dénote X_i le temps de correction de la i -ième copie, 7 heures donnent 420 minutes. On cherche donc

$$\mathbb{P} \left(\sum_{i=1}^{25} X_i \leq 420 \right)$$

avec $\mathbb{E}[X_i] = 20$ et $\text{Var}(X_i) = 16$. On centre et renormalise pour définir

$$Z_{25} = \frac{1}{\sqrt{25 \times 16}} \left(\sum_{i=1}^{25} X_i - 25 \times 20 \right) = \frac{1}{20} \left(\sum_{i=1}^{25} X_i - 500 \right).$$

Ainsi, on calcule,

$$\mathbb{P} \left(\sum_{i=1}^{25} X_i \leq 420 \right) = \mathbb{P} \left(Z_{25} \leq \frac{420 - 500}{20} \right) = \mathbb{P}(Z_{25} \leq -4) \approx 1 - \Phi(4) \approx 0.00003$$

On a aussi une version encore plus générale du théorème central limite où on relaxe l'hypothèse d'être identiquement distribuée.

THÉORÈME 7.11: Théorème central limite plus général

Soit (X_i) une suite de variables aléatoires indépendantes telle que $\mathbb{E}[X_i] = \mu_i$ et $\text{Var}(X_i) = \sigma_i^2$ alors si

- Il existe $M > 0$ tel que pour tout $i \in \mathbb{N}$, $\mathbb{P}(|X_i| < M) = 1$.
- $\sum_{i=1}^{\infty} \sigma_i^2 = +\infty$.

alors

$$\mathbb{P} \left(\frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq x \right) \xrightarrow{n \rightarrow \infty} \Phi(x) = \int_{-\infty}^x \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du.$$