

Chapitre 05 – Probabilités (1)

Ce gros chapitre va nous permettre de généraliser les probabilités faites en sup. Malgré la première partie, il faut retenir trois points importants :

- l'univers, on s'en moque (pour parler poliment)
- une variable aléatoire n'est pas une variable, et n'est pas aléatoire : c'est une fonction sur l'univers qui montre le **résultat** d'une expérience aléatoire
- ce sont les variables aléatoires qui créent les événements

1 Probabilités sur un univers discret

Remarque 0.a. En sup, une probabilité est définie sur toutes les parties d'un ensemble fini appelé univers, vérifiant $\mathbb{P}(\Omega) = 1$ et, pour $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

C'est une définition très robuste. Si on veut définir une probabilité sur un ensemble dénombrable, on pourra aussi se contenter de ces hypothèses.

En revanche, si l'univers n'est pas dénombrable, les choses se compliquent : il existerait des parties de cet ensemble **qui n'ont pas de probabilité**... d'où la nécessité de définir $\mathbb{P}$ sur quelque chose de plus restreint que $\mathcal{P}(\mathbb{R})$. Ce quelque chose, c'est ce que l'on appellera tribu ou σ -algèbre.

Il faut avoir en tête que les univers infinis non dénombrables, ça existe, et très facilement : si on prend une suite de variables iid $(X_n)_{n \in \mathbb{N}}$ où chaque X_n est à valeurs dans $\mathbb{N}$, on a besoin d'un univers de la taille de $\mathbb{N}^{\mathbb{N}}$ pour les construire.

1.1 Événements et variables aléatoires

Définition 1.

1. On appelle tribu (ou sigma-algèbre) sur un **univers** Ω une partie $\mathcal{A}$ de $\mathcal{P}(\Omega)$ telle que :

- (i) $\Omega \in \mathcal{A}$
- (ii) $\forall A \in \mathcal{A}, \bar{A} = \Omega \setminus A \in \mathcal{A}$
- (iii) pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{A}$, $\bigcup_{n=0}^{+\infty} A_n \in \mathcal{A}$

2. Les éléments de $\mathcal{A}$ sont appelés événements (ou parties mesurables).

3. Deux événements A et B sont dits incompatibles lorsque $A \cap B = \emptyset$.

4. Un système complet d'événements est une suite $(A_n)_{n \in \mathbb{N}}$ d'événements deux à deux incompatibles dont la réunion est égale à Ω .

Remarque 1.a. Il faut remarquer immédiatement que $\mathcal{P}(\Omega)$ vérifie, bien sûr, les axiomes d'une tribu. Le but est de considérer d'autres tribus.

Proposition 2. Si $\mathcal{A}$ est une tribu sur Ω , $\emptyset \in \mathcal{A}$ et $\bigcap_{n=0}^{+\infty} A_n \in \mathcal{A}$ pour toute suite $(A_n)_{n \in \mathbb{N}}$ de $\mathcal{A}^{\mathbb{N}}$.

Définition 3.

1. Si $\mathcal{A}$ est une tribu sur un ensemble Ω , une variable aléatoire discrète sur $(\Omega, \mathcal{A})$ est une application X sur Ω , d'image finie ou dénombrable, telle que : $\forall x \in X(\Omega), X^{-1}(\{x\}) \in \mathcal{A}$.

2. On définit, si $B \subset X(\Omega)$, l'événement $\{X \in B\} = \{\omega \in \Omega, X(\omega) \in B\}$.

3. Si $X(\Omega) \subset \mathbb{R}$, X est dite réelle. On définit alors les événements $X < a$, $X \leq a$, etc.

4. Si A est un événement, la **fonction indicatrice** de A , notée $\mathbb{1}_A$, est la variable aléatoire définie par $\mathbb{1}_A(\omega) = 1$ si $\omega \in A$ et 0 sinon.

Exemple 3.a.

1. Vous avez 20 dés à 6 faces et vous lancez les dés. La somme obtenue est notée S . Alors S est une variable aléatoire et on note $S = \sum_{k=1}^{20} X_k$, ce qui signifie que pour tout ω dans Ω , $S(\omega) = \sum_{k=1}^{20} X_k(\omega)$.

2. Vous avez un dé à 20 faces et 20 dés à 6 faces. On lance le dé à 20 faces, on note N son résultat, et on lance autant de dés à 6 faces que le résultat obtenu. On note S la somme des valeurs obtenues. Alors N est une variable aléatoire, S aussi. Si on note $X_1, \dots, X_{20}$ le résultat du lancer de chacun des 20 dés, alors,

pour tout ω dans Ω , $S(\omega) = \sum_{k=1}^{N(\omega)} X_k(\omega)$.

3. Si on tire une permutation σ au hasard (vocabulaire à préciser !) et que l'on note A_i l'événement « le point i est fixé par la permutation σ », si F est le nombre de points fixes de σ , alors
- σ est une variable aléatoire à valeurs dans $\mathcal{S}_n$: on note $\sigma(\omega)$ le résultat de son tirage. **Attention aux objets !** $\sigma(\omega) \in \mathcal{S}_n$! Ainsi, pour dire que i est un point fixe de $\sigma(\omega)$, on écrit que $\sigma(\omega)(i) = i$.
 - Les A_i sont bien des événements : $A_i = \{\omega \in \Omega, \sigma(\omega)(i) = i\}$.
 - On a alors $F = \sum_{i=1}^n \mathbb{1}_{A_i}$, ce qui signifie que pour tout ω , $F(\omega) = \sum_{i=1}^n \mathbb{1}_{A_i}(\omega)$.

Remarque 3.b. Il faut retenir ce principe fondamental sur les probabilités : les objets fondamentaux sont les **variables aléatoires**, l'univers et la tribu ne seront **jamais à définir**. De même, les événements sont créés par les variables aléatoires.

Remarque 3.c. Un théorème fondamental, dû à Kolmogorov, assure que, pour une suite de lois données, il existe une tribu sur laquelle vit une suite de variables aléatoires indépendantes suivant cette loi.

Proposition 4 (et définition).

1. Si X est une variable aléatoire discrète à valeurs dans E , ensemble dénombrable, si f est une fonction de E dans F , alors on note $f(X)$ la variable aléatoire définie par : $\forall \omega \in \Omega, f(X)(\omega) = f(X(\omega))$. On définit ainsi une variable aléatoire discrète.
2. Si X et Y sont deux variables aléatoires discrètes, à valeurs respectives dans E_1 et E_2 , alors pour toute $g : E_1 \times E_2 \rightarrow G$, $g(X, Y)$ est une variable aléatoire discrète.
3. En particulier, (X, Y) est une variable aléatoire discrète.
4. On généralise au cas de n variables aléatoires.

Exemple 4.a. C'est exactement ce que l'on a fait précédemment dans les exemples.

Exercice 4.b. On lance une pièce (une infinité de fois). Montrer que « n'obtenir jamais pile », et « obtenir pile une infinité de fois » sont des événements.

1.2 Probabilité

Définition 5.

1. Si $\mathcal{A}$ est une tribu sur Ω , une probabilité sur $(\Omega, \mathcal{A})$ est une application $\mathbb{P}$ de $\mathcal{A}$ dans $[0, 1]$ vérifiant : $\mathbb{P}(\Omega) = 1$ et pour toute suite $(A_n)_{n \geq 0}$ d'événements deux à deux incompatibles, $\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$ (σ -additivité).
2. On dit alors que $(\Omega, \mathcal{A}, \mathbb{P})$ est un **espace probabilisé**.
3. Si $\mathbb{P}(A) = 1$ (resp 0), A est dit presque sûr (resp. négligeable).
4. On dit qu'une suite d'événements $(A_n)_{n \in \mathbb{N}}$ est un système quasi-complet lorsque ces événements sont deux à deux incompatibles et que la somme de leurs probabilités vaut 1.

Proposition 6. Soient A, B , deux événements, $(A_n)_{n \in \mathbb{N}}$ une suite d'événements.

1. $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
3. Une probabilité est croissante pour l'inclusion.
4. (Continuité croissante) Si pour tout $n, A_n \subset A_{n+1}$, alors $\mathbb{P}(A_n) \xrightarrow{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=0}^{+\infty} A_k\right)$.
5. De manière générale, $\mathbb{P}\left(\bigcup_{k=0}^n A_k\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=0}^{+\infty} A_k\right)$.
6. (Continuité décroissante) Si pour tout $n, A_{n+1} \subset A_n$, alors $\mathbb{P}(A_n) \xrightarrow{n \rightarrow +\infty} \mathbb{P}\left(\bigcap_{k=0}^{+\infty} A_k\right)$.
7. De manière générale, $\mathbb{P}\left(\bigcap_{k=0}^n A_k\right) \xrightarrow{n \rightarrow +\infty} \mathbb{P}\left(\bigcap_{k=0}^{+\infty} A_k\right)$.
8. (Sous-additivité) $\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) \leq \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$.
9. Une réunion dénombrable d'événements négligeables est négligeable.
10. Une intersection finie ou dénombrable d'événements presque sûrs est presque sûre.

Démonstration.

1. On utilise que $\mathbb{P}(A \sqcup \bar{A}) = \mathbb{P}(A) + \mathbb{P}(\bar{A})$.
2. On utilise que $A \cup B = (A \cap B) \sqcup B$.
3. On utilise que si $A \subset B$, alors $B = A \sqcup (B \setminus A)$.
4. Là, on détaille. On note, pour k dans $\mathbb{N}$, $B_k = A_k \setminus A_{k-1}$, avec comme convention que $B_0 = A_0$. On a alors, par récurrence immédiate, pour tout n dans $\mathbb{N}$,

$$\bigcup_{k=0}^n B_k = \bigcup_{k=0}^n A_k = A_n,$$

d'où $\bigcup_{k=0}^{+\infty} B_k = \bigcup_{k=0}^{+\infty} A_k$: la première inclusion est évidente car $B_k \subset A_k$ et la seconde l'est car $A_k \subset \bigcup_{i=0}^k B_i$.

Mais, par sigma additivité, on en déduit que

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k=0}^{+\infty} A_k\right) &= \mathbb{P}\left(\bigcup_{k=0}^{+\infty} B_k\right) \\ &= \sum_{k=0}^{+\infty} \mathbb{P}(B_k) \\ &= \lim_{n \rightarrow +\infty} \sum_{k=0}^n \mathbb{P}(B_k) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=0}^n B_k\right) \text{ par } \sigma\text{-additivité.} \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}(A_n), \end{aligned}$$

ce qui est exactement ce que l'on souhaitait.

5. C'est la proposition précédente, mais en notant $A'_n = \bigcap_{k=0}^n A_k$. On a alors $A'_n \subset A'_{n+1}$.
6. On utilise la continuité croissante (car $\overline{A_k} \subset \overline{A_{k+1}}$) et les événements contraires. On sait que

$$\mathbb{P}(A_n) = 1 - \mathbb{P}(\overline{A_n}) \xrightarrow{n \rightarrow +\infty} 1 - \mathbb{P}\left(\bigcup_{k=0}^{+\infty} \overline{A_k}\right) = \mathbb{P}\left(\bigcap_{k=0}^{+\infty} A_k\right).$$

7. S'en déduit bien.
8. Idem : on sait que $\mathbb{P}\left(\bigcup_{k=0}^{+\infty} A_k\right) = \sum_{k=0}^{+\infty} \mathbb{P}(B_k) \leq \sum_{k=0}^{+\infty} \mathbb{P}(A_k)$.
9. Pour les événements négligeables, c'est la conséquence de la sous-additivité. Pour les presque-sûrs, on passe au complémentaire.

□

Remarque 7. Si Ω est discret, une distribution de probabilité sur Ω est une famille $(a_\omega)_{\omega \in \Omega}$ de réels de $[0, 1]$ telle que $\sum_{\omega \in \Omega} a_\omega = 1$. On définit alors naturellement une probabilité sur $\mathcal{P}(\Omega)$.

Proposition 8 (Probabilités totales). Si $(A_n)_{n \in \mathbb{N}}$ est un sqce alors : $\forall B, \mathbb{P}(B) = \sum_{n=0}^{+\infty} \mathbb{P}(B \cap A_n)$.

Démonstration. Cela vient simplement de la σ additivité !

□

Définition 9. Si X est une variable aléatoire discrète sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, on appelle loi de X l'application $\mathbb{P}_X$:

$$\begin{cases} \mathcal{P}(X(\Omega)) \rightarrow [0, 1] \\ U \mapsto \mathbb{P}(X \in U) = \mathbb{P}(X^{-1}(U)) \end{cases}$$

Remarque 10.

1. $\mathbb{P}_X$ est une probabilité sur $(X(\Omega), \mathcal{P}(X(\Omega)))$, entièrement déterminée par $\{P(X = x)\}_{x \in X(\Omega)}$.
2. En pratique, on n'utilise **JAMAIS** la notation $\mathbb{P}_X$.
3. Déterminer la **loi** d'une variable aléatoire X , c'est déterminer un ensemble fini ou dénombrable E tel que $X(\Omega) \subset E$ et, pour tout x dans E , $\mathbb{P}(X = x)$.
4. Si deux variables X et Y suivent la même loi, on note $X \sim Y$.
5. Si $X \sim Y$ et que $f : X(\Omega) \rightarrow F$, alors $f(X) \sim f(Y)$.

Définition 11 (Lois usuelles). Soit $p \in]0, 1[$, $\lambda > 0$.

Loi	Paramètre	Notation	$X(\Omega)$	$\mathbb{P}(X = k)$	Espérance	Variance
Uniforme		$X \sim \mathcal{U}(E)$	E	$1/\text{Card}(E)$		
Bernoulli	p	$X \sim \mathcal{B}(p)$	$\{0, 1\}$	$1 - p$ ($k = 0$); p ($k = 1$)	p	$p(1 - p)$
Binomiale	(n, p)	$X \sim \mathcal{B}(n, p)$	$[0, n]$	$\binom{n}{k} p^k (1 - p)^{n-k}$	np	$np(1 - p)$
Géométrique	p	$X \sim \mathcal{G}(p)$	$\mathbb{N}^*$	$p(1 - p)^{k-1}$	$1/p$	$(1 - p)/p^2$
Poisson	λ	$X \sim \mathcal{P}(\lambda)$	$\mathbb{N}$	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ

Exercice 11.a.

1. Vérifier que les lois définissent bien des lois
2. $\mathbb{P}(X \text{ est pair})$ lorsque X suit une loi géométrique/de Poisson.
3. Loi des événements rares (exo classique) Soient $\lambda > 0$ et $(p_n)_{n \in \mathbb{N}} \in [0, 1]^{\mathbb{N}}$ telle que $\lim_{n \rightarrow +\infty} np_n = \lambda$. Pour tout entier naturel n , soit X_n une variable aléatoire de loi binomiale de paramètres (n, p_n) . Alors,

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{E}^{-\lambda} \frac{\lambda^k}{k!}$$

Remarque 12. Si $X \sim \mathcal{G}(p)$, alors $\mathbb{P}(X > k) = (1 - p)^k$.

Définition 13.

1. La loi du couple (X, Y) est appelée loi conjointe.
Elle est déterminée par $\{P((X, Y) = (x, y)) = P((X = x) \cap (Y = y)), (x, y) \in X(\Omega) \times Y(\Omega)\}$.
2. Les lois de X et de Y sont appelées lois marginales.

Remarque 14. Il faut pouvoir retrouver les lois marginales à l'aide des probabilités totales :

$$\mathbb{P}(X = x_0) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x_0, Y = y) \text{ et } \mathbb{P}(Y = y_0) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y_0).$$

Remarque 14.a. La loi conjointe permet de retrouver les lois marginales, pas l'inverse. On peut prendre l'exemple d'une urne avec deux boules noires, deux boules blanches, et un tirage de deux boules, l'un avec et l'autre sans remise.

2 Conditionnement et indépendance

2.1 Conditionnement

Définition 15. Si A et B sont des événements et si $P(B) \neq 0$, la probabilité conditionnelle de A sachant B est définie par : $P_B(A) = P(A | B) = \frac{P(A \cap B)}{P(B)}$. On définit bien une probabilité.

Proposition 16 (Le trio des formules). On suppose que toutes les probabilités conditionnelles suivantes sont définissables.

1. (Probabilités composées) $\mathbb{P}(A_1 \cap \dots \cap A_n) = P(A_1) P_{A_1}(A_2) P_{A_1 \cap A_2}(A_3) \dots P_{A_1 \cap \dots \cap A_{n-1}}(A_n)$.
2. (Probabilités totales) Si $(A_n)_{n \in \mathbb{N}}$ est un s(q)ce, $\forall B \in \mathcal{A}$, $P(B) = \sum_{n=0}^{+\infty} P(B \cap A_n)$.
3. (Formule de Bayes) $\mathbb{P}_A(B) = \frac{P(A \cap B)}{P(A)} = \frac{P_B(A)P(B)}{P(A)}$.

Remarque 16.a.

1. Ces formules sont celles que l'on utilise naturellement dans les petites classes, lorsqu'on utilise un arbre de probabilités.
2. La formule des probabilités totales à retenir est celle sans les probabilités conditionnelles.

Exercice 16.b. La loi géométrique est la seule loi sans mémoire. En d'autres termes, si X est une variable aléatoire à valeurs dans $\mathbb{N}$, on a l'équivalence entre « X suit une loi géométrique » et « $\forall (n, k) \in \mathbb{N}^2, \mathbb{P}_{X>n}(X > n + k) = \mathbb{P}(X > k)$ ».

Définition 17. Soient X une variable aléatoire discrète, A un événement. La loi de X conditionnée à l'événement A est la donnée, pour tout x de E , de $\mathbb{P}_A(X = x)$.

Exercice 17.a. Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$. X suit une loi de Poisson de paramètre $\lambda > 0$ et pour tout n dans $\mathbb{N}$, la loi de Y conditionnée par $X = n$ est une loi binomiale de paramètres n et $p \in [0, 1]$.

1. Donner la loi de (X, Y) .
2. Reconnaître la loi de Y .
3. Soit $Z = X - Y$. Donner la loi de Z .
4. X et Y sont-elles indépendantes ? (cf. section suivante)

2.2 Indépendance

Définition 18. Soient $A, B, A_1, \dots, A_n$ des événements, $X, Y, (X_k)_{k \in \mathbb{N}}$ des variables aléatoires discrètes.

1. A et B sont dits indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ ou (quand $\mathbb{P}(B) > 0$) à $\mathbb{P}_B(A) = \mathbb{P}(A)$.
2. $A_1, A_2, \dots, A_n$ sont dits mutuellement indépendants si : $\forall I \subset \llbracket 1, n \rrbracket, \mathbb{P}(\bigcap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i)$.
3. X et Y sont indépendantes lorsque : $\forall A \subset X(\Omega), \forall B \subset Y(\Omega), \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$.
On note $X \perp\!\!\!\perp Y$.
4. $(X_1, \dots, X_n)$ sont mutuellement indépendantes si pour tous $A_1 \subset X_1(\Omega), \dots, A_n \subset X_n(\Omega)$,
$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i).$$
5. Les variables aléatoires $(X_k)_{k \in \mathbb{N}}$ sont dites mutuellement indépendantes si pour tout n dans $\mathbb{N}$, $(X_0, \dots, X_n)$ sont mutuellement indépendantes.
6. Des variables aléatoires sont dites i.i.d. (indépendantes et identiquement distribuées) si elles sont indépendantes et suivent la même loi.

Proposition 19.

1. $X \perp\!\!\!\perp Y \iff \forall (x, y) \in X(\Omega) \times Y(\Omega), P(X = x, Y = y) = P(X = x)P(Y = y)$.
2. L'indépendance mutuelle implique l'indépendance deux à deux ; pas l'inverse.
3. Si $A_1, A_2, \dots, A_n$ sont mutuellement indépendants alors $A_1, A_2, \dots, \overline{A_n}$ le sont aussi.
4. (Lemme des coalitions) Si $X_1, \dots, X_n$ sont des v.a.d. indépendantes alors $f(X_1, \dots, X_k)$ et $g(X_{k+1}, \dots, X_n)$ sont deux v.a.d. indépendantes.

Démonstration. On ne montre que la première propriété, le reste se démontre de la même manière !

Soient $\varphi : E \rightarrow G$ et $\psi : F \rightarrow H$. Montrons que $\varphi(X)$ et $\psi(Y)$ sont indépendantes.

Soient $M \subset G$ et $N \subset H$.

Notons $A = \varphi^{-1}(M)$ et $B = \psi^{-1}(N)$, les images réciproques respectives de M par φ et N par ψ . Alors, comme E et F sont finis, A et B sont aussi finis.

Ainsi,

$$\begin{aligned} \mathbb{P}(\varphi(X) \in M, \psi(Y) \in N) &= \mathbb{P}(X \in A, Y \in B) \\ &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \text{ par indépendance de } X \text{ et } Y \\ &= \mathbb{P}(\varphi(X) \in M)\mathbb{P}(\psi(Y) \in N), \end{aligned}$$

d'où l'indépendance des deux variables aléatoires ! □

Proposition 20. Soient $X_1, \dots, X_n$ des v.a.i.d. telles que $X_1 \sim \mathcal{B}(p)$. Alors $X_1 + \dots + X_n \sim \mathcal{B}(n, p)$.

Démonstration. Démontrons par récurrence sur n la proposition.

Initialisation. Pour $n = 1$, si $X_1 \sim \mathcal{B}(p)$, alors $X_1 \sim \mathcal{B}(1, p)$.

Hérédité. On suppose la propriété vraie pour un certain rang. Soient $X_1, \dots, X_n, X_{n+1}$ $n + 1$ variables aléatoires indépendantes, suivant toutes une loi de Bernoulli de paramètre p . Alors

- $\forall i \in \llbracket 1, n \rrbracket$, $X_i \in \{0, 1\}$, donc $X_1 + \dots + X_{n+1} \in \llbracket 0, n + 1 \rrbracket$.
- Soit $k \in \llbracket 0, n + 1 \rrbracket$. Notons

$$S_n = X_1 + \dots + X_n, \quad S_{n+1} = X_1 + \dots + X_{n+1} = S_n + X_{n+1}.$$

Alors

$$\begin{aligned} \mathbb{P}(S_{n+1} = k) &= \mathbb{P}(S_{n+1} = k, X_{n+1} = 0) + \mathbb{P}(S_{n+1} = k, X_{n+1} = 1) \\ &= \mathbb{P}(S_n + X_{n+1} = k, X_{n+1} = 0) + \mathbb{P}(S_n + X_{n+1} = k, X_{n+1} = 1) \\ &= \mathbb{P}(S_n = k, X_{n+1} = 0) + \mathbb{P}(S_n = k - 1, X_{n+1} = 1) \\ &= \mathbb{P}(S_n = k)\mathbb{P}(X_{n+1} = 0) + \mathbb{P}(S_n = k - 1)\mathbb{P}(X_{n+1} = 1) \text{ par indépendance.} \end{aligned}$$

Or,

— si $k = 0$, $\mathbb{P}(S_n = k - 1) = 0$, donc

$$\mathbb{P}(S_{n+1} = 0) = \mathbb{P}(S_n = 0)\mathbb{P}(X_{n+1} = 0) = (1 - p)^n \times (1 - p) = (1 - p)^{n+1} = \binom{n+1}{0} p^0 (1 - p)^{n+1-0},$$

— si $k = n + 1$, $\mathbb{P}(S_n = k) = 0$, donc

$$\mathbb{P}(S_{n+1} = n + 1) = \mathbb{P}(S_n = n)\mathbb{P}(X_{n+1} = 1) = p^n \times p = \binom{n+1}{n+1} p^{n+1} (1 - p)^{n+1-(n+1)},$$

— si $k \in \llbracket 1, n \rrbracket$,

$$\begin{aligned} \mathbb{P}(S_{n+1} = k) &= \mathbb{P}(S_n = k)\mathbb{P}(X_{n+1} = 0) + \mathbb{P}(S_n = k - 1)\mathbb{P}(X_{n+1} = 1) \\ &= \binom{n}{k} p^k (1 - p)^{n-k} \times (1 - p) + \binom{n}{k-1} p^{k-1} (1 - p)^{n-(k-1)} \times p \\ &= \left[\binom{n}{k} + \binom{n}{k-1} \right] p^k (1 - p)^{n+1-k} \\ &= \binom{n}{k+1} p^k (1 - p)^{n+1-k}, \end{aligned}$$

par la formule du triangle de Pascal.

□

3 Moments d'une variable aléatoire

3.1 Espérance d'une v.a.d. réelle ou complexe

Définition 21.

1. L'espérance d'une v.a.d. X à valeurs dans $[0, +\infty]$ est définie par : $\mathbb{E}(X) = \sum_{x \in X(\Omega)} xP(X = x)$. (avec la convention $xP(X = x) = 0$ si $x = +\infty$ et $\mathbb{P}(X = +\infty) = 0$)
2. On dit qu'une v.a.d. X à valeurs dans $\mathbb{C}$ est d'espérance finie, et on note $X \in L^1$, lorsque la famille $(xP(X = x))_{x \in X(\Omega)}$ est sommable. $\mathbb{E}(X)$ est alors la somme de cette famille.
3. On dit que X est centrée lorsque $\mathbb{E}(X) = 0$.

Proposition 22. Si $X \in L^1$ et X est réelle, presque sûrement positive, alors $\mathbb{E}(X) \geq 0$, avec égalité ssi X est presque sûrement nulle.

Proposition 23. Si X est à valeurs dans $\mathbb{N} \cup \{+\infty\}$, alors $\mathbb{E}(X) = \sum_{n \geq 1} \mathbb{P}(X \geq n)$.

Démonstration. Toutes les quantités étant dans $[0, +\infty]$, nous pouvons utiliser le théorème de sommation par paquets. On écrit que

$$\begin{aligned}\sum_{n \geq 1} \mathbb{P}(X \geq n) &= \sum_{n \geq 1} \sum_{k=n}^{+\infty} \mathbb{P}(X = k) \\ &= \sum_{1 \leq n \leq k} u_{n,k} \text{ en notant } u_{n,k} = \mathbb{P}(X = k). \\ &= \sum_{k \geq 1} \sum_{n=1}^k \mathbb{P}(X = k)\end{aligned}$$

par sommation par paquets

$$= \sum_{k \geq 0} \sum_{n=1}^k \mathbb{P}(X = k) = \mathbb{E}(X).$$

D'où la formule désirée. □

Proposition 24 (Espérances usuelles). Cf. Définition 11.

Démonstration.

1. Si $X \sim \mathcal{B}(p)$,

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

2. Si $X \sim \mathcal{U}([a, b])$, alors, pour tout k dans $[a, b]$, $\mathbb{P}(X = k) = \frac{1}{b-a+1}$, donc

$$\mathbb{E}(X) = \sum_{k=a}^b k \cdot \mathbb{P}(X = k) = \sum_{k=a}^b k \frac{1}{b-a+1} = \frac{1}{b-a+1} \sum_{k=a}^b k = \frac{1}{b-a+1} \frac{(a+b)(b-a+1)}{2} = \frac{a+b}{2}.$$

3. Si $X \sim \mathcal{B}(n, p)$,

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

4. Si $X \sim \mathcal{G}(p)$, alors

$$\mathbb{E}(X) = \sum_{k \geq 1} \mathbb{P}(X \geq k) = \sum_{k \geq 1} (1-p)^{k-1} = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

5. Si $X \sim \mathcal{P}(\lambda)$, alors

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

□

Remarque 25. Il est important de remarquer que $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

Proposition 26 (Formule de transfert). $f(X) \in L^1$ si et seulement si la famille $(f(x)\mathbb{P}(X = x))_{x \in X(\Omega)}$ est sommable. Dans ce cas, $\mathbb{E}(f(X)) = \sum_{x \in X(\Omega)} f(x)\mathbb{P}(X = x)$.

Démonstration. On montre déjà l'équivalence des sommabilités. On est toujours dans le cas où X est à valeurs dans un ensemble dénombrable. On note $Z = |f(X)|$. On a alors $X(\Omega) = \bigsqcup_{z \in Z(\Omega)} \underbrace{\{x \in X(\Omega), |f(x)| = z\}}_{E_z}$.

$$\begin{aligned} \sum_{x \in X(\Omega)} |f(x)|\mathbb{P}(X = x) &= \sum_{z \in Z(\Omega)} \sum_{x \in E_z} |f(x)|\mathbb{P}(X = x) \text{ par sommation par paquets} \\ &= \sum_{z \in Z(\Omega)} \sum_{x \in E_z} z\mathbb{P}(X = x) \\ &= \sum_{z \in Z(\Omega)} z\mathbb{P}(X \in E_z) \\ &= \sum_{z \in Z(\Omega)} z\mathbb{P}(Z = z). \end{aligned}$$

Ainsi $\sum_{x \in X(\Omega)} |f(x)|\mathbb{P}(X = x) < +\infty$ si et seulement si $\sum_{z \in Z(\Omega)} z\mathbb{P}(Z = z) < +\infty$. Ceci assure que $f(X) \in L^1$ si et seulement si la famille $(f(x)\mathbb{P}(X = x))_{x \in X(\Omega)}$ est sommable. Dans ce cas, l'interversion sans les valeurs absolues est licite et on obtient la formule désirée. □

Remarque 26.a. On va le voir, mais la formule de transfert est à la base de la preuve de beaucoup d'autres propositions.

Proposition 27.

1. Linéarité de l'espérance : si X et Y sont deux v.a.d. d'espérance finie, si $(\lambda, \mu) \in \mathbb{C}^2$, alors $\lambda X + \mu Y$ admet une espérance et $\mathbb{E}(\lambda X + \mu Y) = \lambda \mathbb{E}(X) + \mu \mathbb{E}(Y)$.
2. Si $X \leq Y$ presque sûrement, $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
3. Si $|X| \leq Y$ et $\mathbb{E}(Y) < +\infty$ alors X est d'espérance finie et $|\mathbb{E}(X)| \leq \mathbb{E}(Y)$.

Démonstration.

1. La variable aléatoire $\lambda X + \mu Y$ est une fonction de la variable aléatoire (X, Y) . Ainsi, par la formule de transfert, et sous réserve de justification,

$$\begin{aligned} \mathbb{E}(\lambda X + \mu Y) &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} (\lambda x + \mu y)\mathbb{P}(X = x, Y = y) \\ &= \lambda \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} x\mathbb{P}(X = x, Y = y) + \mu \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} y\mathbb{P}(X = x, Y = y) \text{ par linéarité de la somme} \\ &= \lambda \sum_{x \in X(\Omega)} x\mathbb{P}(X = x) + \mu \sum_{y \in Y(\Omega)} y\mathbb{P}(Y = y) \\ &= \lambda \mathbb{E}(X) + \mu \mathbb{E}(Y). \end{aligned}$$

La justification de la sommabilité de la première famille se fait aisément :

$$\begin{aligned}
 \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} |(\lambda x + \mu y) \mathbb{P}(X = x, Y = y)| \\
 &\leq \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} (|\lambda x| + |\mu y|) \mathbb{P}(X = x, Y = y) \\
 &\leq |\lambda| \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} |x| \mathbb{P}(X = x, Y = y) + |\mu| \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} |y| \mathbb{P}(X = x, Y = y) \\
 &\leq |\lambda| \sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) + |\mu| \sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y) \\
 &\leq |\lambda| \mathbb{E}(|X|) + |\mu| \mathbb{E}(|Y|) < +\infty.
 \end{aligned}$$

2. Cela vient de la linéarité : si $X \leq Y$ presque sûrement, $Y - X \geq 0$ presque sûrement, ce qui assure que $\mathbb{E}(Y - X) \geq 0$ et, par linéarité, que $\mathbb{E}(Y) \geq \mathbb{E}(X)$.
3. Cela vient de l'inégalité triangulaire :

$$|\mathbb{E}(X)| = \left| \sum_{x \in X(\Omega)} x \mathbb{P}(X = x) \right| \leq \sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) = \mathbb{E}(|X|).$$

□

Remarque 28. On retrouve l'espérance de la loi binomiale.

Exercice 28.a. Démontrer l'inégalité de Jensen pour l'espérance : si X est une variable aléatoire réelle, si φ est une fonction convexe et X une variable aléatoire telle que X et $\varphi(X)$ sont dans L^1 , alors $\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X))$.

Exercice 28.b. Déterminer le nombre moyen de points fixes d'une permutation tirée uniformément au hasard.

Proposition 29. Si X et Y sont deux v.a.r. indépendantes d'espérances finies, alors XY aussi et $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. On adapte au cas de n v.a.r. mutuellement indépendantes.

Démonstration. C'est encore une formule de transfert, sur le couple (X, Y) et la fonction $f : (x, y) \mapsto xy$. On écrit, sous réserve de justification, que

$$\begin{aligned}
 \mathbb{E}(XY) &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} xy \mathbb{P}(X = x, Y = y) \\
 &= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} xy \mathbb{P}(X = x) \mathbb{P}(Y = y) \text{ par indépendance} \\
 &= \sum_{x \in X(\Omega)} x \mathbb{P}(X = x) \sum_{y \in Y(\Omega)} \mathbb{P}(Y = y) \\
 &= \mathbb{E}(X) \mathbb{E}(Y).
 \end{aligned}$$

La justification de la sommabilité se fait de la même manière, en prenant les valeurs absolues des quantités précédentes. □

Remarque 29.a. ATTENTION!!!! La réciproque est très fautive !

En effet, soit X uniforme sur $-1, 0, 1$ et Y la fonction indicatrice de $X = 0$, c'est-à-dire que Y vaut 1 si $X = 0$ et 0 sinon.

Alors

- $XY = 0$ car si $X \neq 0$, $Y = 0$, donc $\mathbb{E}(XY) = 0$,
- $\mathbb{E}(X) = 0$, donc $\mathbb{E}(X)\mathbb{E}(Y) = 0$.

Mais X et Y ne sont pas indépendantes : en effet, $\mathbb{P}(X = 0, Y = 1) = \mathbb{P}(X = 0) \neq \mathbb{P}(X = 0) \times \mathbb{P}(Y = 1)$.

Proposition 30 (Inégalité de Markov). Si X est une v.a.d. réelle **positive** (presque sûrement) d'espérance finie, alors pour tout $a > 0$, $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$.

Démonstration. On va utiliser les fonctions indicatrices, plus précisément que $\mathbb{P}(X \geq a) = \mathbb{E}(\mathbb{1}_{X \geq a})$. Alors on a, pour tout ω dans Ω ,

$$a\mathbb{1}_{X \geq a}(\omega) \leq X(\omega)\mathbb{1}_{X \geq a}(\omega) \leq X(\omega),$$

la première inégalité se faisant par disjonction de cas et la seconde venant de la positivité de X . Par croissance de l'espérance, il vient $\mathbb{E}(a\mathbb{1}_{X \geq a}) \leq \mathbb{E}(X)$, d'où l'inégalité de Markov. $\square$

Exercice 30.a. Montrer que, si $e^{tX} \in L^1$ pour tout t , alors pour tout t , $\mathbb{P}(X \geq \varepsilon) \leq e^{-t\varepsilon}\mathbb{E}(e^{tX})$.

3.2 Moments d'ordre 2 – Variance d'une v.a.d. réelle

Proposition 31 (Inégalité de Cauchy-Schwarz). Soient X et Y deux v.a.d. réelles. Si X^2 et Y^2 sont dans L^1 , alors XY aussi et $|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$, avec égalité ssi X et Y sont presque sûrement colinéaires. Ainsi, $X \in L^2 \Rightarrow X \in L^1$.

Démonstration. Le fait que XY soit dans L^1 est évident, car $|XY| \leq \frac{X^2 + Y^2}{2}$.

Ensuite, on suppose X non presque sûrement nulle (sinon le résultat est immédiat) et on considère la fonction $P(t) = \mathbb{E}((tX + Y)^2)$. Alors P est bien définie (en développant le carré, on a des variables L^1), positive, et, pour tout t dans $\mathbb{R}$,

$$P(t) = \mathbb{E}(X^2)t^2 + 2\mathbb{E}(XY)t + \mathbb{E}(Y^2),$$

qui est une expression polynomiale de degré 2, étant donné que $\mathbb{E}(X^2) \neq 0$. Ce polynôme est positif sur $\mathbb{R}$, donc de discriminant négatif ou nul, ce qui assure que $\mathbb{E}(XY)^2 - \mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0$.

Pour le cas d'égalité, on a les équivalences :

$$\begin{aligned} \mathbb{E}(XY) = \mathbb{E}(X^2)\mathbb{E}(Y^2) &\Leftrightarrow \exists t_0 \in \mathbb{R}, P(t_0) = 0 \\ &\Leftrightarrow \exists t_0 \in \mathbb{R}, \mathbb{E}((t_0X + Y)^2) = 0 \\ &\Leftrightarrow \exists t_0 \in \mathbb{R}, Y = -t_0X \text{ presque sûrement.} \end{aligned}$$

$\square$

Remarque 31.a. Plutôt que de dire « X^2 est dans L^1 », on peut dire que X est dans L^2 ...

Définition 32. Si X et Y sont deux v.a.d. réelles avec X^2 et Y^2 dans L^1 , on définit :

1. la variance de X , $\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
2. la covariance de X et Y , $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
3. l'écart type de X , $\sigma(X) = \sqrt{\mathbb{V}(X)}$ (et l'on dit que X est réduite lorsque $\sigma(X) = 1$)

Proposition 33 (Variances usuelles). Cf. Définition 11.

Démonstration.

1. Par la formule de Koenig-Huygens, $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. Or, on a déjà vu que $\mathbb{E}(X) = \frac{n}{2}$. Par la formule de transfert,

$$\mathbb{E}(X^2) = \sum_{k=0}^n k^2 \mathbb{P}(X = k) = \frac{1}{n+1} \sum_{k=0}^n k^2 = \frac{1}{n+1} \frac{n(n+1)(2n+1)}{6} = \frac{n(2n+1)}{6},$$

donc

$$\mathbb{V}(X) = \frac{n(2n+1)}{6} - \frac{n^2}{4} = \frac{2n(2n+1) - 3n^2}{12} = \frac{n(4n+2-3n)}{12} = \frac{n(n+2)}{12}.$$

2. Si $X \sim \mathcal{B}(p)$, $\mathbb{E}(X) = p$ et $\mathbb{E}(X^2) = p$ car $X^2 = X$, donc

$$\mathbb{V}(X) = p - p^2 = p(1-p).$$

3. (en cours, on ne va pas faire le calcul, mais vu qu'on peut le faire, faisons-le ici !) Toujours par la formule

de transfert,

$$\begin{aligned}
 \mathbb{E}(X^2) &= \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \sum_{k=1}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \sum_{k=1}^n kn \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
 &= \sum_{\ell=k-1}^{n-1} (\ell+1)n \binom{n-1}{\ell} p^{\ell+1} (1-p)^{n-1-\ell} \\
 &= \sum_{\ell=0}^{n-1} \ell n \binom{n-1}{\ell} p^{\ell+1} (1-p)^{n-1-\ell} + n \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell+1} (1-p)^{n-1-\ell} \\
 &= A + B.
 \end{aligned}$$

Déjà,

$$B = np \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{n-1-\ell} = np.$$

Ensuite,

$$\begin{aligned}
 A &= n \sum_{\ell=0}^{n-1} \ell \binom{n-1}{\ell} p^{\ell+1} (1-p)^{n-1-\ell} \\
 &= n \sum_{\ell=1}^{n-1} \ell \binom{n-1}{\ell} p^{\ell+1} (1-p)^{n-1-\ell} \\
 &= n \sum_{\ell=1}^{n-1} (n-1) \binom{n-2}{\ell-1} p^{\ell+1} (1-p)^{n-1-\ell} \\
 &= \sum_{j=\ell-1}^{n-2} n(n-1) \binom{n-2}{j} p^{j+2} (1-p)^{n-2-j} \\
 &= n^2 p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} = n(n-1)p^2,
 \end{aligned}$$

donc

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

donc

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = n^2 p^2 + np(1-p) - n^2 p^2 = np(1-p),$$

d'où le résultat désiré !

4. On calcule, par la formule de transfert,

$$\begin{aligned}
 \mathbb{E}(X^2) &= \sum_{k=1}^{+\infty} k^2 p q^{k-1} \\
 &= \sum_{k=1}^{+\infty} (k-1+1)^2 p q^{k-1} \\
 &= \sum_{k=1}^{+\infty} (k-1)^2 p q^{k-1} + 2p \sum_{k=1}^{+\infty} (k-1) q^{k-1} + \sum_{k=1}^{+\infty} p q^{k-1} \\
 &= q \mathbb{E}(X^2) + 2 \frac{q}{p} + 1 \\
 &= q \mathbb{E}(X^2) + \frac{2q+p}{p}
 \end{aligned}$$

donc, comme $p = 1 - q$,

$$p\mathbb{E}(X^2) = \frac{q+1}{p}, \text{ i.e. } \mathbb{E}(X^2) = \frac{q+1}{p^2}$$

Ainsi,

$$\mathbb{V}(X) = \frac{q+1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

5. On calcule

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

Donc $\mathbb{V}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

□

Proposition 34.

1. La covariance est une forme bilinéaire symétrique positive sur l'espace vectoriel L^2 .
2. $\mathbb{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j)$.
3. Si X et Y sont indépendantes ou juste décorrélées, alors $\text{Cov}(X, Y) = 0$ et donc $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$. On généralise à n variables.
4. $\mathbb{V}(X) = 0$ si et seulement si X est presque sûrement constante.
5. $\forall (a, b) \in \mathbb{R}^2, \mathbb{V}(aX + b) = a^2 \mathbb{V}(X)$.
6. si $\sigma(X) > 0$ alors $\frac{X - \mathbb{E}(X)}{\sigma(X)}$ est centrée réduite.

Démonstration.

1. Assez clair.
2. Idem
3. On a :

$$\begin{aligned} \mathbb{V}(X + Y) &= \mathbb{E}((X + Y)^2) - \mathbb{E}(X + Y)^2 \\ &= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)^2 \text{ par indépendance de } X \text{ et } Y \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \\ &= \mathbb{V}(X) + \mathbb{V}(Y), \end{aligned}$$

d'où le résultat !

4. On calcule

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

par linéarité de l'espérance. Donc

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

donc

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

5. Déjà,

$$\mathbb{E}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}(\mathbb{E}(X) - \mu) = \frac{1}{\sigma}(\mu - \mu) = 0,$$

donc $\frac{X - \mu}{\sigma}$ est bien centrée. Ensuite,

$$\mathbb{V}\left(\frac{X - \mu}{\sigma}\right) = \mathbb{V}\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2}\mathbb{V}(X) = \frac{\sigma^2}{\sigma^2} = 1,$$

donc $\frac{X - \mu}{\sigma}$ est centrée réduite.

□

Remarque 34.a. Comme la proposition $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$ est une conséquence de $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ lorsque $X \perp\!\!\!\perp Y$, on peut aussi préciser que « $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$ » **n'entraîne pas** que $X \perp\!\!\!\perp Y$. (prendre le même contre-exemple que précédemment)

Remarque 34.b. On a une inégalité de Cauchy-Schwarz pour la covariance : soient X et Y deux v.a.r. Alors $\text{Cov}(X, Y)^2 \leq \mathbb{V}(X)\mathbb{V}(Y)$, avec égalité si et seulement si X est presque sûrement égale à une fonction affine de Y , i.e. ssi il existe α et β tels que $X = \alpha Y + \beta$ presque sûrement.

Proposition 35.

1. (inégalité de Bienaymé-Tchebycheff) Soit X une v.a.d. réelle d'espérance et de variance finies.

Alors pour tout $a > 0$, $\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathbb{V}(X)}{a^2}$.

2. (loi faible des grands nombres) Soit $(X_k)_{k \in \mathbb{N}^*}$ une suite de v.a.d. réelles i.i.d., d'espérance finie μ et de variance finie σ^2 . On note, pour n dans $\mathbb{N}^*$, $S_n = \sum_{k=1}^n X_k$. Alors pour tout $\varepsilon > 0$, $\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0$.

On dit que la suite $\left(\frac{S_n}{n}\right)_{n \in \mathbb{N}^*}$ converge en probabilité vers μ .

Démonstration.

1. Soit $\varepsilon > 0$. Alors

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) &= \mathbb{P}((X - \mathbb{E}(X))^2 \geq \varepsilon^2) \text{ car } \varepsilon > 0 \\ &\leq \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{\varepsilon^2} \text{ par l'inégalité de Markov} \\ &= \frac{\mathbb{V}(X)}{\varepsilon^2}. \end{aligned}$$

2. Soit $n \in \mathbb{N}^*$.

Par linéarité de l'espérance,

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\mathbb{E}(S_n) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = \mu$$

Ainsi,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) &= \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| \geq \varepsilon\right) \\ &\leq \frac{\mathbb{V}\left(\frac{S_n}{n}\right)}{\varepsilon^2} \text{ par l'inégalité de Bienaymé-Tchebycheff.} \end{aligned}$$

Mais

$$\begin{aligned}\mathbb{V}\left(\frac{S_n}{n}\right) &= \frac{1}{n^2} \mathbb{V}(S_n) \\ &= \frac{1}{n^2} \mathbb{V}\left(\sum_{k=1}^n X_k\right) \\ &= \frac{1}{n^2} \times n \sum_{k=1}^n \mathbb{V}(X_k) \text{ par indépendance} \\ &= \frac{1}{n} \mathbb{V}(X_1) \text{ car les } X_k \text{ suivent la même loi,}\end{aligned}$$

donc

$$\mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{\mathbb{V}(X_1)}{n\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0,$$

d'où le résultat !

□