PROBABILITÉS ET STATISTIQUES
UNIVERSITÉ DE GENÈVE

CHRISTOPHE PITTET

Table des matières

1. Les axiomes de Kolmogorov

Soit X un ensemble fini contenant n éléments. On note $|X| = n$ son cardinal.

Proposition 1.1. *(Nombre d'applications entre deux ensembles finis.) Soient X et Y des ensembles finis. Il existe*

$$|Y|^{|X|}$$

applications de X dans Y.

Démonstration. Soit $x \in X$. On peut lui associer $|Y|$ éléments de Y. On a donc en tout $|Y|^{|X|}$ choix possibles. $\square$

Définition 1.2. *Soit* Ω *un ensemble. Une* partie *de* Ω *est un sous-ensemble de* Ω*. On note* $\mathcal{P}(\Omega)$ *l'ensemble des parties de* Ω*.*

Exemple 1.3.

$$\Omega = \{A; B; C\},$$
$$\mathcal{P}(\Omega) = \{\emptyset; \{A\}, \{B\}; \{C\}; \{A; B\}; \{A; C\}; \{B; C\}; \Omega\}.$$

Exemple 1.4.

$$\Omega = \{0; 1\},$$
$$\mathcal{P}(\Omega) = \{\emptyset; \{0\}, \{1\}; \Omega\}.$$

Proposition 1.5. *(Un ensemble de cardinal* n *contient* 2^n *sous-ensembles.) Soit* Ω *un ensemble contenant* n *éléments. L'ensemble de ses parties* $\mathcal{P}(\Omega)$ *contient* 2^n *éléments :*

$$(|\Omega| = n) \implies (|\mathcal{P}(\Omega)| = 2^n).$$

Démonstration. Soit $A \subset \Omega$. On associe à A sa fonction caractéristique 1_A. C'est-à-dire :

$$1_A : \Omega \to \{0; 1\}$$
$$1_A(x) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{si } x \in \Omega \setminus A. \end{cases}$$

Il est évident que la donnée de la fonction 1_A permet de retrouver le sous-ensemble A. Il y a donc une correspondance bi-univoque (une bijection) entre les sous-ensembles de Ω et les applications de Ω dans $\{0; 1\}$. Le nombre de ces applications est $2^{|\Omega|}$. $\square$

Définition 1.6. *(Ensembles dénombrables.) Un ensemble* E *est dit dénombrable s'il existe une application surjective, de l'ensemble*

$$\mathbb{N} = \{1; 2; 3; \ldots; n; \ldots\}$$

des entiers naturels, dans E*.*

Tout ensemble fini est dénombrable. L'ensemble des entiers relatifs $\mathbb{Z}$ est dénombrable, l'ensemble $\mathbb{Q}$ des fractions rationnelles est dénombrble. Par contre, l'ensemble des suites infinies de 0 et de 1 n'est pas dénombrable (Cantor). Un produit cartésien fini d'ensembles dénombrables est dénombrable. Un sous-ensemble d'un ensemble dénombrable est dénombrable (évident).

Définition 1.7. *(σ-algèbre ou tribu.) Soit Ω un ensemble. Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$. On dit que la famille $\mathcal{A}$ de sous-ensembles de Ω est une* σ-algèbre *ou une* tribu *de Ω si les conditions suivantes sont vérifiées : l'ensemble tout entier Ω est un membre de la famille $\mathcal{A}$, si un sous-ensemble A de Ω est un membre de la famille $\mathcal{A}$ alors son complément $\Omega \setminus A$ est aussi un membre de la famille $\mathcal{A}$, si $A_1, \ldots, A_n, \ldots$ forment une sous-famille dénombrable de $\mathcal{A}$ alors la réunion de tous ces ensembles $A_1, \ldots, A_n, \ldots$ est un membre de $\mathcal{A}$. Formellement :*

(1) $\Omega \in \mathcal{A}$,

(2) $A \in \mathcal{A} \implies \Omega \setminus A \in \mathcal{A}$,

(3) $A_1, \ldots, A_n, \ldots \in \mathcal{A} \implies \bigcup_{n\geq 1} A_n \in \mathcal{A}$.

Exemple 1.8. *La plus grande tribu de Ω est évidement $\mathcal{A} = \mathcal{P}(\Omega)$. La plus petite tribu de Ω est évidement $\mathcal{A} = \{\emptyset; \Omega\}$.*

Théorème 1.9. *(Tribu engendrée par une famille.) Soit Ω un ensemble. Soit $\mathcal{E}$ une famille de sous-ensembles de Ω. Il existe une plus petite tribu $(\mathcal{E})$ de Ω contenant $\mathcal{E}$.*

Démonstration. Commençons par une remarque d'ordre psychologique : si $\mathcal{F}$ est une famille de sous-ensembles de Ω, il est ici avantageux de considérer $\mathcal{F}$ comme un sous-ensemble de $\mathcal{P}(\Omega)$. On réalise ainsi que l'on peut considérer le sous-ensemble $(\mathcal{E})$ de $\mathcal{P}(\Omega)$ défini comme l'intersection de tous les sous-ensembles $\mathcal{A}$ de $\mathcal{P}(\Omega)$ qui sont des tribus de Ω et qui contiennent le sous-ensemble $\mathcal{E}$ de $\mathcal{P}(\Omega)$:

$$(\mathcal{E}) = \bigcap_{\mathcal{E} \subset \mathcal{A}} \mathcal{A}.$$

La première chose à vérifier est qu'il existe au moins une tribu $\mathcal{A}$ contenant $\mathcal{E}$ (sans quoi on considèrerait une intersection d'ensembles qui n'existent pas !). C'est bien le cas puisque $\mathcal{P}(\Omega)$ est une tribu qui contient évidemment $\mathcal{E}$. Vérifions que $(\mathcal{E})$ est une tribu de Ω. Si $\mathcal{A}$ intervient dans cette intersection alors $\Omega \in \mathcal{A}$ puisque $\mathcal{A}$ est une tribu. Donc $\Omega \in (\mathcal{E})$. Si $E \in (\mathcal{E})$, alors $E \in \mathcal{A}$ pour tout $\mathcal{A}$ intervenant dans l'intersection. Comme un tel $\mathcal{A}$ est une tribu, $\Omega \setminus E \in \mathcal{A}$. Donc $\Omega \setminus E \in (\mathcal{E})$. Si $A_1, \ldots, A_n, \ldots \in (\mathcal{E})$, alors pour tout $\mathcal{A}$ intervenant dans l'intersection, $A_1, \ldots, A_n, \ldots \in \mathcal{A}$. Comme $\mathcal{A}$ est une tribu, $\bigcup_{n\geq 1} A_n \in \mathcal{A}$. Donc $\bigcup_{n\geq 1} A_n \in (\mathcal{E})$. Ce qui prouve que $(\mathcal{E})$ est une tribu de Ω. Comme tout $\mathcal{A}$ intervenant dans l'intersection contient $\mathcal{E}$, il en va de même de l'intersection $(\mathcal{E})$. Ainsi $(\mathcal{E})$ est une tribu contenant $\mathcal{E}$. C'est la plus petite car si une tribu contient $\mathcal{E}$, elle intervient dans l'intersection qui définit $(\mathcal{E})$. □

Exemple 1.10. *Soit $\Omega = \{1; 2; 3\}$. Soit $\mathcal{E} = \{\{1\}\}$ la famille de sous-ensembles de Ω qui ne possède qu'un membre : le singleton $\{1\}$. On vérifie que la plus petite tribu de Ω contenant $\mathcal{E}$ est :*

$$(\mathcal{E}) = \{\emptyset; \{1\}; \{2; 3\}, \Omega\}.$$

Définition 1.11. *(Les axiomes de Kolmogorov.) Soit Ω un ensemble. Soit $\mathcal{A}$ un tribu de Ω. Une* mesure de probabilités *sur $\mathcal{A}$ est une application*

$$\mathbb{P} : \mathcal{A} \to [0, 1]$$

$$A \mapsto \mathbb{P}(A),$$

qui vérifie les propriétés suivantes :

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

(2) si $A_1, \dots, A_n, \dots$ est une famille dénombrable d'éléments de $\mathcal{A}$ et que tous ces sous-ensembles de Ω sont disjoints, alors

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n).$$

Exemple 1.12. *(Probabilité uniforme sur un ensemble fini.) Soit Ω un ensemble fini. Soit $\mathcal{A} = \mathcal{P}(\Omega)$ la tribu maximale. La* probabilité uniforme *sur Ω est ainsi définie :*

$$\mathbb{P} : P(\Omega) \to [0, 1]$$

$$A \mapsto \frac{|A|}{|\Omega|}.$$

Exemple 1.13. *(Modélisation du jet d'une pièce.) Soit $\Omega = \{P; F\}$, $\mathcal{A} = \mathcal{P}(\Omega)$. Soit $0 \leq p \leq 1$. On définit une mesure de probabilité par les égalités*

$$\mathbb{P}(\emptyset) = 0,\ \mathbb{P}(P) = p,\ \mathbb{P}(F) = 1 - p,\ \mathbb{P}(\Omega) = 1.$$

Si $p = 1/2$ on dit que la pièce n'est pas biaisée.

Exemple 1.14. *(Modélisation du jet d'un dé.) Soit $\Omega = \{1; 2; 3; 4; 5; 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$. Soit*

$$0 \leq p_1, p_2, p_3, p_4, p_5, p_6 \leq 1,$$

tels que

$$\sum_{i=1}^{6} p_i = 1.$$

On définit la mesure de probabilité de $A \subset \Omega$ par la formule

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

Si $p_i = 1/6$ pour tout $1 \leq i \leq 6$, on dit que le dé n'est pas pipé. Dans ce cas on retrouve la mesure uniforme :

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

Théorème 1.15. *(Propriétés élémentaires d'une mesure de probabilité.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé (c'est-à-dire un triple qui vérifie les axiomes de Kolmogorov). Pour tout $A, B \in \mathcal{A}$, on a :*

(1) $\mathbb{P}(\Omega \setminus A) = 1 - \mathbb{P}(A)$.

(2) Si $A \subset B$ alors $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$.

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

(4) Soit $A_1, \ldots, A_i, \ldots, A_n$ des éléments de $\mathcal{A}$. Alors

$$\mathbb{P}\left(\cup_i A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots$$
$$+ (-1)^{n+1}\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n).$$

Démonstration. (1) Notons $A^c = \Omega \setminus A$. On a

$$A \cap A^c = \emptyset.$$

Donc

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

Ce qui prouve le premier point.

(2) Supposons $A \subset B$. Dans ce cas on a

$$B = A \cup (B \setminus A),$$

et comme la réunion est disjointe,

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A),$$

ce qui prouve le deuxième point.

(3) Pour prouver le troisième point, remarquons que la réunion

$$A \cup B = A \cup (B \setminus A)$$

est disjointe donc

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \cup (B \setminus A)) = \mathbb{P}(A) + \mathbb{P}(B \setminus A).$$

Mais

$$A \cap B \subset B$$

donc d'après le point précédent :

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

Ce que l'on peut écrire ainsi :

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

Finalement :

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

Ce qui prouve le troisième point.

(4) Le dernier point est connu sous le nom de formule de Poincaré ou principe d'inclusion exclusion. Lorsque $n = 2$ on retrouve le point précédent. On démontre cette formule par récurrence sur n. Il est intéressant de visualiser la formule lorsque $n = 3$ en dessinant un diagrame de Venn.

□

Corollaire 1.16. *(Formule d'inclusion-exclusion) Soit Ω un ensemble fini et soit $A_1, \ldots, A_i, \ldots, A_n$ une famille de n sous-ensembles de Ω. Alors :*

$$|\bigcup_i A_i| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \cdots$$
$$+ (-1)^{n+1} |A_1 \cap A_2 \cap \cdots \cap A_n|.$$

Démonstration. On choisit la mesure de probabilité uniforme sur Ω et on multiplie par $|\Omega|$. □

Proposition 1.17. *(Probabilités d'unions et d'intersections dénombrables.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.*

(1) Soit $A_1 \subset \cdots \subset A_n \subset \cdots \ldots$ une famille croissante d'éléments de $\mathcal{A}$. Alors

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

(2) Soit $B_1 \supset \cdots \supset B_n \supset \cdots \ldots$ une famille décroissante d'éléments de $\mathcal{A}$. Alors

$$\mathbb{P}\left(\bigcap_n B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Démonstration. Ecrivons A comme la réunion disjointe dénombrable

$$A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots$$

On obtient donc :

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A_1) + \sum_{n\geq 1} \mathbb{P}(A_{n+1} \setminus A_n) \\
&= \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{n} [\mathbb{P}(A_{i+1}) - \mathbb{P}(A_i))] \\
&= \lim_{n\to\infty} \mathbb{P}(A_n).
\end{aligned}$$

Afin de prouver le second point, définissons pour $n \in \mathbb{N}$,

$$A_n = \Omega \setminus B_n.$$

Comme B_n décroît, $A_n = \Omega \setminus B_n$ croît. On peut donc appliquer le premier point. On a :

$$\bigcap_n B_n = \Omega \setminus \bigcup_n A_n.$$

Ainsi :

$$\begin{aligned}
\mathbb{P}\left(\bigcap_n B_n\right) &= 1 - \mathbb{P}\left(\bigcup_n A_n\right) \\
&= 1 - \lim_{n\to\infty} \mathbb{P}(A_n) \\
&= 1 - \lim_{n\to\infty} (1 - \mathbb{P}(B_n)) \\
&= \lim_{n\to\infty} \mathbb{P}(B_n).
\end{aligned}$$

□

2. Probabilités conditionnelles

Dans un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, on interprète un élément de $\mathcal{A}$ comme un *évènement* et $\mathbb{P}(A)$ comme la probabilité que l'évènement A se réalise.

Définition 2.1. *(Probabilité conditionnelle.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $B \in \mathcal{A}$ *tel que* $\mathbb{P}(B) > 0$*. On définit la* probabilité conditionnelle *que* $A \in \mathcal{A}$ *se réalise sachant que* B *est réalisé comme*

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

Exemple 2.2. *Une famille comprend deux enfants. Quelle est la probabilité que les deux soient des garçons, sachant que l'un d'eux est un garçon ?*

(1) *Première réponse (sophiste) : comme on sait qu'un des enfants est un garçon et que l'autre enfant est soit un garçon soit une fille avec même probabilité, la réponse est* $1/2$.

(2) *Deuxième réponse basée sur les axiomes de Kolmogorov. On définit*

$$\Omega = \{GG; GF; FG; FF\}.$$

On suppose que

$$\mathbb{P}(GG) = \mathbb{P}(GF) = \mathbb{P}(FG) = \mathbb{P}(FF) = 1/4.$$

(On se permet d'omettre les parenthèses.) En d'autres termes, on considère la mesure uniforme sur l'univers Ω *des possibilités. D'après la définition de probabilité conditionnelle, la probabilité que les deux enfants soient des garçons, sachant que l'un d'eux est un garçon, se calcule ainsi :*

$$\begin{aligned}\mathbb{P}(GG|GG \cup GF \cup FG) &= \frac{\mathbb{P}(GG \cap (GG \cup GF \cup FG))}{\mathbb{P}(GG \cup GF \cup FG)} \\ &= \frac{\mathbb{P}(GG)}{\mathbb{P}(GG \cup GF \cup FG)} \\ &= \frac{1/4}{3/4} \\ &= 1/3.\end{aligned}$$

Proposition 2.3. *(Formule des probabilités totales.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit*

$$\Omega = \bigsqcup_n A_n$$

une partition dénombrable de Ω *par des éléments* A_n *de* $\mathcal{A}$*. Supposons que pour tout* n*,*

$$\mathbb{P}(A_n) > 0.$$

Alors, pour tout $B \in \mathcal{A}$*,*

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

Démonstration. On a :

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}\left(B \cap \bigsqcup_n A_n\right)\\ &= \mathbb{P}\left(\bigsqcup_n B \cap A_n\right)\\ &= \sum_n \mathbb{P}\left(B \cap A_n\right)\\ &= \sum_n \mathbb{P}(B|A_n)\mathbb{P}(A_n).\end{aligned}$$

□

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. En général, la réalisation d'un évènement $B \in \mathcal{A}$ influence la probabilité qu'un autre évènement $A \in \mathcal{A}$ se réalise. Dans le cas où la probabilité de A n'est pas affectée par la réalisation de A, c'est-à-dire lorsque

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

on en déduit que :

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

Dans la définition suivante, on étend cette idée aussi au cas où $\mathbb{P}(B) = 0$.

Définition 2.4. *(Événements indépendants.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. On dit que* $A \in \mathcal{A}$ *et* $B \in \mathcal{A}$ *sont* indépendants *si*

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

Soit

$$\{A_i \in \mathcal{A} : i \in I\}$$

une famille quelconque d'évènements. On dit que les évènements de cette famille sont mutuellement indépendants *si pour tout sous-ensemble fini* $J \subset I$*,*

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

On dit que les évènements de cette famille sont deux-à-deux indépendants *si pour toute paire d'indice* $i, j \in I$ *,*

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

Remarquons que si des évènements sont mutuellement indépendants alors ils sont deux-à-deux indépendants. Mais la réciproque n'est pas vraie.

Théorème 2.5. *(Formule de Bayes.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit*

$$\Omega = \bigsqcup_{i=1}^{n} A_i$$

une partition dénombrable de Ω *(c'est-à-dire que la réunion est disjointe) par des éléments* A_i *de* $\mathcal{A}$*. Supposons que pour tout* $1 \leq i \leq n$*,*

$$\mathbb{P}(A_i) > 0.$$

Soit $B \in \mathcal{A}$ *tel que* $\mathbb{P}(B) > 0$*. Alors pour tout* $1 \leq i \leq n$*,*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

Démonstration. On applique la formule des probabilités totale :

$$\begin{aligned}
\mathbb{P}(A_i|B) &= \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.
\end{aligned}$$

□

Exercice 2.6. *(Exemple d'application de la formule de Bayes.) Pour* $i = 1, 2$*, on considère une urne* U_i *qui contient* b_i *boules blanches et* n_i *boules noires. On choisit au hasard de manière équiprobable une des deux urnes. Dans l'urne choisie, on tire au hasard de manière équiprobable une boule. Sachant que la boule tirée est noire, quelle est la probabilité que l'urne choisie soit l'urne numéro i ?*

3. Combinatoire

L'analyse combinatoire est un outil indispensable de la théorie des probabilités. Nous passons en revue quelques formules omniprésentes.

Théorème 3.1. *(Nombre d'injections ou nombre de choix ordonnés.) Soit* $0 \leq k \leq n$*. Les nombres suivants sont égaux.*

1. *Le nombre de choix ordonnés sans répétitions de* k *objets parmi* n*.*
2. *Le nombre d'injections d'un ensemble de cardinal* k *dans un ensemble de cardinal* n*.*

(3) *Le nombre*

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

Démonstration. Choisir k éléments dans l'ensemble

$$\{1;\cdots;n\}$$

sans répétition et en tenant compte de l'ordre des k choix revient à définir une injection

$$i : \{1;\cdots;k\} \to \{1;\cdots;n\}.$$

Il y a n possibilités pour le choix de $i(1)$, il y a $n-1$ possibilités pour le choix de $i(2)$, et il y a $n+1-k$ possibilités pour le choix de $i(k)$. En tout il y a donc

$$n(n-1)\cdots(n-k+1)$$

injections. □

Corollaire 3.2. *(Nombres de permutations.) Soit n un entier. Les nombres suivants sont égaux.*

(1) *Le nombre de permutations de n objets.*

(2) *Le nombre de bijections d'un ensemble de cardinal n dans un ensemble de cardinal n.*

(3) *Le nombre*

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

Démonstration. Choisir $k=n$ dans le théorème. □

Exemple 3.3. *(Ranger des livres sur un rayon.) De combien de manière peut-on arranger 5 livres distincts sur un rayon ? Réponse : $5! = 120$.*

Théorème 3.4. *(Coefficients binomiaux).) Soit $0 \leq k \leq n$. Les nombres suivants sont égaux.*

(1) *Le nombre de choix (non-ordonnés) sans répétitions de k objets parmi n.*

(2) *Le nombre de sous-ensemble de cardinal k d'un ensemble de cardinal n.*

(3) *La valeur du coefficient binomial*

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

Démonstration. Afin de compter les sous-ensembles de cardinal k d'un ensemble de cardinal n, il suffit de considérer toutes les injections d'un ensemble de cardinal k dans un ensemble de cardinal n et de regrouper ces injections par classes. Deux injections sont dans la même classe si et seulement si elles ont même image. Combien y a-t-il d'injections dans une classe ? Réponse : $k!$. Donc en divisant le nombre

$$\frac{n!}{(n-k)!}$$

de ces injections par le cardinal

$$k!$$

d'une classe, on trouve le nombre

$$\frac{n!}{(n-k)!k!}$$

de sous-ensembles à k éléments dans un ensemble à n éléments. □

Exercice 3.5. *Une personne veut emprunter* 3 *livres pour ses vacances dans une bibliothèque qui contient* 1000 *livres. Combien cette personnes a-t-elle de choix possibles ?*

Proposition 3.6. *(Vers le triangle de Pascal.) Soit* $1 \leq k \leq n$ *des entiers. On a :*

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

Démonstration. Soit X un ensemble de cardinal $n \geq 1$. Fixons un élément $x \in X$. Les sous-ensembles de X de cardinal k se divisent en deux classes : ceux qui contiennent x et ceux qui ne le contiennent pas. Il y a

$$\binom{n-1}{k-1}$$

qui contiennent x. Il y en a

$$\binom{n-1}{k}$$

qui ne contiennent pas x. □

Théorème 3.7. *(Coefficients multinomiaux.) Soit* $n_1, n_2, \ldots, n_k$ *des entiers positifs ou nuls. Soit* $n = n_1 + n_2 + \cdots + n_k$*. Les nombres suivants sont égaux.*

(1) Le nombre de division d'une population de n *individus en* k *groupes numérotés, le premier groupe contenant* n_1 *individus, le second en contenant* n_2*, etc. le groupe numéro* k *contenant* n_k *individus.*

(2) Le nombre de partages $(A_1, \ldots, A_k)$ d'un ensemble à n éléments en k sous-ensembles A_i, $1 \leq i \leq k$ deux-à-deux disjoints tels que

$$|A_i| = n_i.$$

(3) Le nombre d'arrangements de n boules dont n_1 sont de la couleur numéro 1, n_2 sont de la couleur numéro 2, etc. et n_k sont de la couleur numéro k. (Les couleurs sont distinctes et les boules sont indistinguables si ce n'est par leur couleur.)

(4) Le coefficient multinomial

$$\binom{n}{n_1, \ldots, n_k} = \frac{n!}{n_1! \cdots n_k!}.$$

4. Variables aléatoires

Définition 4.1. *Soit $(\Omega, \mathcal{A})$ et $(E, \mathcal{B})$ deux ensembles muni de tribus. C'est-à-dire que $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une σ-algèbre inclue dans $\mathcal{P}(\Omega)$ et $\mathcal{B} \subset \mathcal{P}(E)$ est une σ-algèbre inclue dans $\mathcal{P}(E)$. Une* variable aléatoire *définie sur Ω et à valeurs dans E est une application*

$$X : \Omega \to E$$

qui respecte les tribus. Plus précisément :

$$\forall B \in \mathcal{B},\ X^{-1}(B) \in \mathcal{A}.$$

Exemple 4.2. *(Nombre de piles obtenus en lançant une pièce n fois.) Soit $\Omega = \{0; 1\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$, $E = \{0; \cdots; n\}$, $\mathcal{B} = \mathcal{P}(E)$.*

$$X : \{0; 1\}^n \to \{0; \cdots; n\}$$

$$(x_1, \ldots, x_n) \mapsto \sum_{i=1}^{n} x_i$$

Si on note 1 les piles et 0 les faces, on voit que si l'on interprète

$$(x_1, \ldots, x_n) \in \Omega$$

comme une réalisation de n lancés d'une pièce, alors l'évaluation de la variable aléatoire X sur cette réalisation

$$X(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i$$

s'interprète comme le nombre de piles obtenues lors des n lancés $(x_1, \ldots, x_n)$.

Définition 4.3. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et* $(E, \mathcal{B})$ *un ensemble muni d'une tribu. Soit*

$$X : \Omega \to E$$

une variable aléatoire. Soit $B \in \mathcal{B}$. *L'évènement*

$$(X \in B)$$

signifie intuitivement que la variable aléatoire X *prend une valeur qui est un élément de* B. *On le définit formellement comme la préimage de* B *par* X *:*

$$(X \in B) = X^{-1}(B) = \{x \in \Omega : X(x) \in B\}.$$

On définit la probabilité de l'évènement $(X \in B)$ *comme le nombre*

$$\mathbb{P}(X \in B) = \mathbb{P}(\{x \in \Omega : X(x) \in B\}).$$

Remarque 4.4. *Par définition d'une variable aléatoire*

$$\{x \in \Omega : X(x) \in B\} = X^{-1}(B) \in \mathcal{A}$$

et donc il est légitime d'appliquer la mesure de probabilité $\mathbb{P}$ *sur ce sous-ensemble de* Ω.

Exemple 4.5. *(Probabilité d'obtenir* k *piles lors de* n *lancés.) On reprend le modèle décrit ci-dessus et on le complète en définissant une mesure de probabilité. Soit* $\Omega = \{0; 1\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$,

$$\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$$

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

la probabilité uniforme.

$E = \{0; \cdots; n\}$, $\mathcal{B} = \mathcal{P}(E)$.

$$X : \{0; 1\}^n \to \{0; \cdots; n\}$$

$$(x_1, \ldots, x_n) \mapsto \sum_{i=1}^{n} x_i.$$

Le cardinal de

$$(X = k) = \{(x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i = k\}$$

est égal à :

$$\binom{n}{k}.$$

En effet, un élément de $(X = k)$ correspond à un choix de k éléments parmi n (les k éléments choisis sont les k coordonnées égales à 1, les $n-k$ autres coordonnées valant automatiquement 0). Donc :

$$\mathbb{P}(X = k) = \mathbb{P}\left(\left\{(x_1, \dots, x_n) : \sum_{i=1}^{n} x_i = k\right\}\right) = \frac{\binom{n}{k}}{2^n}.$$

Proposition 4.6. *(Densité de probabilités, cas discret.) Soit Ω un ensemble dénombrable. Il existe une correspondance naturelle entre :*

les esaces probabilisés $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ d'une part,

et

les fonctions $f : \Omega \to [0, 1]$ telles que

$$\sum_{x \in \Omega} f(x) = 1,$$

d'autre part.

Une telle fonction est appelée une densité de probabilités.

Démonstration. Supposons que $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est donné. On définit

$$f(x) = \mathbb{P}(\{x\}).$$

On a donc $f(x) \geq 0$ par définition d'une probabilité. Comme Ω est dénombrable, on a aussi par définition d'une probabilité :

$$\sum_{x \in \Omega} f(x) = \sum_{x \in \Omega} \mathbb{P}(\{x\}) = \mathbb{P}\left(\bigcup_{x \in \Omega} \{x\}\right) = \mathbb{P}(\Omega) = 1.$$

Réciproquement supposons $f : \Omega \to [0, 1]$ donnée et vérifiant

$$\sum_{x \in \Omega} f(x) = 1.$$

Soit $A \subset \Omega$. Définissons

$$\mathbb{P}(A) = \sum_{x \in A} f(x).$$

Il est évident que $\mathbb{P}$ prend ses valeurs dans $[0, 1]$. De plus comme déja vu :

$$\mathbb{P}(\Omega) = \sum_{x \in \Omega} f(x).$$

Selon l'hypothèse sur f on obtient donc bien $\mathbb{P}(\Omega) = 1$. La σ-additivité de $\mathbb{P}$ est un exercice. $\square$

Exercice 4.7. *Soit Ω un ensemble dénombrable. Soit $f : \Omega \to [0,1]$ telle que*

$$\sum_{x\in\Omega} f(x) = 1,$$

On définit

$$\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$$
$$\mathbb{P}(A) = \sum_{x\in A} f(x).$$

Vérifier que $(\Omega, P(\Omega), \mathbb{P})$ est un espace probabilisé.

Exercice 4.8. *(Densité de Poisson.) Soit $\lambda \geq 0$. Soit $\Omega = \mathbb{N} \cup \{0\}$. Pour tout $k \in \mathbb{N} \cup \{0\}$, posons*

$$f(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Vérifier que f est une densité de probabilités.

Proposition 4.9. *(Mesure produit, cas discret.) Pour $i = 1, 2$, soit $(\Omega_i, \mathcal{P}(\Omega_i), \mathbb{P}_i)$ un espace probabilisé. Il existe une unique mesure de probabilités $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$ telle que $(\Omega_1 \times \Omega_2, \mathbb{P}(\Omega_1 \times \Omega_2), \mathbb{P})$ soit un espace probabilisé vérifiant :*

$$\forall (x_1, x_2) \in \Omega_1 \times \Omega_2,\ \mathbb{P}(\{(x_1, x_2)\}) = \mathbb{P}_1(\{x_1\})\mathbb{P}_2(\{x_2\}).$$

Démonstration. Pour $i = 1, 2$, soit f_i la densité de probabilités de $\mathbb{P}_i$. Définissons

$$f : \Omega_1 \times \Omega_2 \to [0,1]$$
$$(x_1, x_2) \mapsto f_1(x_1)f_2(x_2).$$

On a :

$$\sum_{(x_1,x_2)} f(x_1, x_2) = \sum_{(x_1,x_2)} f_1(x_1)f_2(x_2) = \sum_{x_1} f_1(x_1) \sum_{x_2} f_2(x_2) = 1$$

On applique alors la Proposition 4.6 $\square$

Définition 4.10. *(Variables indépendantes.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(E_i, \mathcal{B}_i)$ une famille d'ensembles munis de tribus indexés par un ensemble I. Pour tout $i \in I$, soit*

$$X_i : \mathcal{A} \to E_i$$

une variable aléatoire. On dit que les variables sont deux à deux indépendantes si pour tout $i, j \in I$, $i \neq j$, pour tout $B_i \in \mathcal{B}_i$, pour tout $B_j \in \mathcal{B}_j$, les évènements

$$(X_i \in B_i),\ (X_j \in B_j)$$

sont indépendants, en d'autres termes si

$$\mathbb{P}(X_i^{-1}(B_i) \cap X_j^{-1}(B_j)) = \mathbb{P}(X_i^{-1}(B_i))\mathbb{P}(X_j^{-1}(B_j)).$$

On dit que les variables sont mutuellement indépendantes si pour tout sous-ensemble fini $F \subset I$, pour tout $i \in F$, pour tout $B_i \in \mathcal{B}_i$, pour tout les évènements

$$(X_i \in B_i)_{i \in F}$$

sont mutuellement indépendants, en d'autres termes si

$$\mathbb{P}\left(\bigcap_{i \in F} X_i^{-1}(B_i)\right) = \prod_{i \in F} \mathbb{P}(X_i^{-1}(B_i)).$$

Remarque 4.11. *Attention l'espace probabilisé est fixé et toutes les variables en question ont même source. Leurs buts peuvent être différents.*

Proposition 4.12. *(Loi de probabilité d'une variable aléatoire ou poussé avant d'une mesure de probabilités.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(E, \mathcal{B})$ un ensemble muni d'une tribu. Soit*

$$X : \Omega \to E$$

une variable aléatoire. On définit la loi de X par les conditions :

$$\mathbb{P}_X : \mathcal{B} \to [0,1]$$
$$B \mapsto (X^{-1}(B)).$$

Alors $(E, \mathcal{B}, \mathbb{P}_X)$ est un espace probabilisé.

Démonstration. Exercice.

□

Proposition 4.13. *(Variables de Bernoulli indépendantes.) Soit $\Omega = \{0;1\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$. Soit $1 \leq i \leq n$. Soit $0 \leq p_i \leq 1$. Soit*

$$\mathbb{P}_i : \mathcal{P}(\{0;1\}) \to [0,1]$$

la mesure de probabilité définie par la densité de probabilités

$$f_i : \{0;1\} \to [0,1]$$
$$f_i(0) = p_i,\ f_i(1) = 1 - p_i.$$

Soit

$$\mathbb{P} = \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n.$$

Le triple

$$(\{0;1\}^n, \mathcal{P}(\{0;1\}^n, \mathbb{P})$$

est un espace probabilisé. Soit

$$X_i : \{0;1\}^n \to \{0;1\}$$

$$X_i(x_1, \dots, x_n) = x_i$$

la projection sur la coordonnée numéro i.

(1) La loi de probabilité $\mathbb{P}_{X_i}$ de X_i vérifie

$$\mathbb{P}_{X_i}(X_i = 0) = p_i,\ \mathbb{P}_{X_i}(X_i = 1) = 1 - p_i.$$

(2) Les variables $X_1, \dots, X_n$ sont mutuellement indépendantes.

Démonstration. En appliquant la Proposition 4.9 plusieurs fois on voit que $\mathbb{P} = \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n$ est une mesure de probabilités sur $\mathcal{P}(\{0;1\}^n$. On a

$$\mathbb{P}_{X_i}(X_i = 0) = \mathbb{P}(X_i^{-1}(0)) = \mathbb{P}_1(\{0;1\}) \cdots \mathbb{P}_i(\{0\}) \cdots P_n(\{0;1\}) = p_i.$$

Afin de prouver l'indépendance mutuelle, soit $1 \leq m \leq n$ un entier. Considérons tout d'abord des événements du type suivant :

$$(X_{i_1} = x_{i_1}), \dots, (X_{i_m} = x_{i_m}).$$

En appliquant les définitions de mesure produit et de loi de probabilité d'une variable aléatoire, on obtient :

$$\begin{aligned}\mathbb{P}\left(\bigcap_{k=1}^{m}(X_{i_k} = x_{i_k})\right) &= \prod_{k=1}^{m} \mathbb{P}_{i_k}(\{x_{i_k}\}) \\ &= \prod_{k=1}^{m} \mathbb{P}_{X_{i_k}}(\{x_{i_k}\}) = \prod_{k=1}^{m} \mathbb{P}(X_{i_k} = x_{i_k}).\end{aligned}$$

On en déduit l'indépendance mutuelle sur tous les évènements. □

5. Lois de variables discrètes

Définition 5.1. *(Loi de Bernoulli) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(\{0;1\}, \mathcal{P}(\{0;1\}))$ un ensemble à deux éléments muni de sa tribu. Soit*

$$X : \Omega \to \mathbb{R}$$

une variable aléatoire. On dit que X suit une loi de Bernoulli $B(p,q)$ de paramètres $p, q \geq 0$ tels que $p + q = 1$ si

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

Définition 5.2. *(Loi binomiale.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $(\{0; 1; \cdots; n\}, \mathcal{P}(\{0; 1; \cdots; n\}))$ *un ensemble à* n *éléments muni de sa tribu. Soit*

$$X : \Omega \to \{0; 1; \cdots; n\}$$

une variable aléatoire. On dit que X *suit une loi binômiale* $B(n, p)$ *de paramètre* $0 \leq p \leq 1$ *si pour tout* $k \in \{0; 1; \cdots; n\}$

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

Exercice 5.3. *Soit* $X_1, \ldots, X_n$ *des variables aléatoires suivant une loi de Bernoulli de paramètres* p, q*. Si ces variables sont mutuellement indépendantes, montrer que*

$$S_n = X_1 + \cdots + X_n$$

suit une loi binomiale de paramètre p*.*

Définition 5.4. *(Loi de Poisson.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $(\mathbb{N} \cup \{0\}, \mathcal{P}(\mathbb{N} \cup \{0\})))$*. Soit*

$$X : \Omega \to \mathbb{N} \cup \{0\}$$

une variable aléatoire. On dit que X *suit une loi de Poisson de paramètre* $\lambda > 0$ *si pour tout* $n \in \mathbb{N} \cup \{0\}$

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

Exercice 5.5. *(Loi de Poisson comme approximation de la loi binomiale.) Soit* $0 \leq p \leq q \leq 1$ *tels que* $p + q = 1$*. Soit*

$$B(n, p)(k) = \binom{n}{k} p^k q^{n-k}$$

la densité de probabilité de la loi binomiale. Soit $\lambda > 0$ *supposons que* $p_n > 0$ *est une suite qui tend vers zéro telle que*

$$\lim_{n \to \infty} n p_n = \lambda.$$

Prouver que pour tout entier k *positif ou nul,*

$$\lim_{n \to \infty} B(n, p_n)(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Suggestion. Prouver tout d'abord l'égalité

$$\binom{n}{k} p^k q^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{(np)^k}{k!} \left(1 - \frac{np}{n}\right)^{n-k}.$$

Exercice 5.6. *La fabrication d'un objet dans une usine s'effectue avec* 4 *pour cent de défauts. Calculer la probabilité* P_k *qu'un lot de* 35 *objets choisis au hasard comprennent* k *objects défectueux* ($k = 0, 1, 2, 3$)*. Faire les calculs avec la loi exacte* $B(n, p)$ *et avec son approximation de Poisson de paramètre* $\lambda = np$.

6. Lois continues

Définition 6.1. *(Densité de probabilités, cas continu.) Soit* $f : \mathbb{R} \to [0, \infty[$ *une fonction continue par morceaux. On dit que* f *est une* densité de probabilités *si*

$$\int_{-\infty}^{\infty} f(t)dt = 1$$

Définition 6.2. *(Distribution de probabilités, cas continu.) Soit* $f : \mathbb{R} \to [0, \infty[$ *une densité de probabilités. La distribution de probabilité associée à* f *est la fonction*

$$F : \mathbb{R} \to [0, 1]$$
$$F(x) \int_{-\infty}^{x} f(t)dt.$$

Définition 6.3. *(Loi normale.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $(\mathbb{R}, \mathcal{B})$ *la droite réelle munie de la tribu des boréliens. Soit*

$$X : \Omega \to \mathbb{R}$$

une variable aléatoire. On dit que X *suit une loi normale* $N(\mu, \sigma)$ *de moyenne* $\mu \in \mathbb{R}$ *et d'écart type* $\sigma > 0$ *si pour tout* $x \in \mathbb{R}$,

$$\mathbb{P}(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt.$$

Exercice 6.4. *(Laplace.) Prouver que la fonction*

$$\mathbb{R} \to [0, \infty[$$
$$t \mapsto \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$$

est une densité de probabilités. C'est-à-dire :

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1.$$

Indications. Effectuer le changement de variable $y = \frac{x-\mu}{\sigma}$ *pour se ramener au calcul de*

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy.$$

Ensuite, calculer

$$\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}dy\right)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)}dxdy$$

en passant en coordonnées polaires :

$$x = r\cos(\theta),\ y = r\sin(\theta).$$

Définition 6.5. *(Loi exponentielle.) Soit $\lambda > 0$. On dit qu'une variable aléatoire X suit une* loi exponentielle *de paramètre λ si sa densité de probabilité est la fonction*

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & si\ x \geq 0, \\ 0 & si\ x < 0. \end{cases}$$

.

Exercice 6.6. *Soit X une variable aléatoire de loi exponentielle de paramètre λ. Exprimer par une formule sa loi de probabilité*

$$\mathbb{P}(X \leq x)$$

(on distinguera les cas $x < 0$ et $x \geq 0$.)

Définition 6.7. *(Loi uniforme.) Soit $a < b$. On dit qu'une variable aléatoire X suit une* loi uniforme *sur l'intervalle $[a, b]$ si sa densité de probabilité est la fonction*

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

.

Exercice 6.8. *Soit X une variable aléatoire de loi uniforme sur l'intervalle $[a, b]$. Exprimer par une formule sa loi de probabilité*

$$\mathbb{P}(X \leq x)$$

(on distinguera trois cas.)

7. Espérance, variance, covariance

Définition 7.1. *(Espérance mathématique.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit*

$$X : \Omega \to \mathbb{R}$$

une variable aléatoire à valeurs dans $\mathbb{R}$.

(1) *Supposons que l'image de X est un ensemble dénombrable. Numérotons les valeurs possibles de X : $y_1, \dots, y_n, \dots$. On définit* l'espérance *de X comme :*

$$\mathbb{E}[X] = \sum_i \mathbb{P}(X = y_i) y_i.$$

(C'est donc la moyenne des valeurs de X pondérées par leur probabilités d'occurence.)

(2) *Supposons que la loi de X est données par une distribution continue : il existe une densité de probabilité f telle que pour tout $x \in \mathbb{R}$,*

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

Dans ce cas, on définit l'espérance par la formule :

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t f(t) dt.$$

(C'est donc la moyenne des valeurs de X pondérées par leur densité de probabilités.)

Proposition 7.2. *(Propriétés de l'espérance.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient*

$$X, Y : \Omega \to \mathbb{R}$$

des v.a. à valeurs réelles. Soient $a, b \in \mathbb{R}$. On a :

(1) *S'il existe $c \in \mathbb{R}$ tel que $\mathbb{P}(X = c) = 1$ alors $\mathbb{E}[X] = c$.*

(2) $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$

(3) *Si X et Y sont indépendantes alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.*

Afin de prouver cette proposition, on introduit la notion de *distribution conjointe.*

Définition 7.3. *(Variable conjointe.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient*

$$X, Y : \Omega \to \mathbb{R}$$

des v.a. à valeurs réelles. La variable aléatoire conjointe *de X et Y est la variable aléatoire sur Ω à valeurs dans $\mathbb{R}^2$ ainsi définie :*

$$\Omega \to \mathbb{R}^2$$

$$\omega \mapsto (X(\omega), Y(\omega)).$$

Définition 7.4. *(Densité d'une variable conjointe.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soient*

$$X, Y : \Omega \to \mathbb{R}$$

des v.a. à valeurs réelles. La variable aléatoire conjointe de X *et* Y *admet une densité de probabilité s'il existe une fonction à valeurs positive* $f_{(X,Y)}$ *qui est intégrable et telle que pour tout* $x, y \in \mathbb{R}$,

$$\mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{(X,Y)}(s,t) ds dt.$$

Démonstration. Prouvons le premier point de la proposition.

$$\mathbb{E}[X] = \mathbb{P}(X = c)c = c.$$

Prouvons le second point dans le cas où les deux variables prennent un nombre dénombrable de valeurs. Décomposons ce qu'il faut prouver en deux étapes. Premièrement :

$$\mathbb{E}[aX] = \sum_i \mathbb{P}(X = x_i) a x_i = a \sum_i \mathbb{P}(X = x_i) x_i = a\mathbb{E}[X].$$

Deuxièmement :

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_{i,j} \mathbb{P}(X = x_i, Y = y_j)(x_i + y_j) \\
&= \sum_{i,j} \mathbb{P}(X = x_i, Y = y_j) x_i + \sum_{i,j} \mathbb{P}(X = x_i, Y = y_j) y_j \\
&= \sum_i \mathbb{P}(X = x_i) x_i + \sum_j \mathbb{P}(Y = y_j) y_j \\
&= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}$$

Prouvons le second point dans le cas où les deux variables admettent des distributions continues. Décomposons ce qu'il faut prouver en deux étapes. Premièrement :

$$\mathbb{E}[aX] = \int_{\mathbb{R}} a t f(t) dt = a \int_{\mathbb{R}} t f(t) dt = a\mathbb{E}[X].$$

Deuxièmement :

$$
\begin{aligned}
\mathbb{E}[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(X,Y)}(s,t)(s+t)dsdt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(X,Y)}(s,t)sdsdt + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(X,Y)}(s,t)tdsdt \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f_{(X,Y)}(s,t)dt\right)sds + \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f_{(X,Y)}(s,t)ds\right)tdt \\
&= \int_{-\infty}^{\infty} f_X(s)sds + \int_{-\infty}^{\infty} f_Y(t)tdt \\
&= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}
$$

La preuve du troisième point se déduit du fait que si X et Y sont indépendantes alors

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y).$$

Les détails font l'objet d'un exercice. □

Définition 7.5. *(Variance.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit*

$$X : \Omega \to \mathbb{R}$$

une variable aléatoire à valeurs dans $\mathbb{R}$. *Supposons que l'espérance de* X *existe. La* variance *de* X *est l'espérance de la variable aléatoire* $(X - \mathbb{E}[X])^2$ *:*

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \in [0, \infty].$$

Proposition 7.6. *(Formule pour la variance.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit*

$$X : \Omega \to \mathbb{R}$$

une variable aléatoire à valeurs dans $\mathbb{R}$. *Si* $E[X]$ *existe alors*

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

Démonstration. On a :

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + E[X]^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2.
\end{aligned}
$$

□

Remarque 7.7. *On a donc toujours :*

$$\mathbb{E}[X^2] \geq \mathbb{E}[X]^2.$$

C'est un cas particulier de l'inégalité de Cauchy-Schwartz (exercice).

Proposition 7.8. *(Propriétés de la variance.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit*

$$X : \Omega \to \mathbb{R}$$

une variable aléatoire à valeurs dans $\mathbb{R}$ *qui admet une variance. Soit* $a, b \in \mathbb{R}$. *On a :*

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

Démonstration.

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

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

□

Définition 7.9. *(Ecart-type.) Soit* X *une variable aléatoire qui admet une variance. L'*écart-type *ou la* déviation σ *de* X *est par définition la racine de la variance de* X *:*

$$\sigma = \sqrt{\mathrm{Var}(X)} \geq 0.$$

Proposition 7.10. *(Variable centrée réduite.) Soit* X *une variable aléatoire à valeurs réelles qui admet une variance. Notons* $\mu = \mathbb{E}[X]$ *son espérance et* σ *son écart-type. Si* $\sigma > 0$, *on définit la variable* centrée réduite

$$X^* = \frac{X - \mu}{\sigma}.$$

La variable X^* *est alors* centrée *:*

$$\mathbb{E}[X^*] = 0$$

et réduite *:*

$$\mathrm{Var}(X^*) = 1.$$

Réciproquement, si X^* *est une v.a. à valeurs réelles d'espérance nulle et d'écart-type égal à 1, si* $\sigma \geq 0$ *et* $\mu \in \mathbb{R}$ *sont données, alors la variable aléatoire*

$$X = \sigma X^* + \mu$$

satisfait

$$\mathbb{E}[X] = \mu$$

et

$$\mathrm{Var}(X) = \sigma^2.$$

Démonstration. On a :

$$\begin{aligned}\mathbb{E}[X^*] &= \mathbb{E}\left[\frac{X-\mu}{\sigma}\right] \\ &= \mathbb{E}\left[\frac{X}{\sigma}\right] - \mathbb{E}\left[\frac{\mu}{\sigma}\right] \\ &= 0.\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(X^*) &= \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma^2}\mathrm{Var}(X-\mu) \\ &= 1.\end{aligned}$$

D'autre part,

$$\mathbb{E}[X] = \mathbb{E}[\sigma X^* + \mu] = \mu,$$
$$\mathrm{Var}(X) = \mathrm{Var}(\sigma X^* + \mu) = \sigma^2 \mathrm{Var}(X^*) = \sigma^2.$$

□

Définition 7.11. *(Moment d'ordre 2.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et soit* X *une v.a. sur* Ω *à valeurs dans* $\mathbb{R}$*. On suppose que la loi de* X *admet une densité de probabilité* f_X*. On dit que* X *admet un moment d'ordre 2 si*

$$\int_{-\infty}^{\infty} t^2 f_X(t)dt$$

converge vers un nombre fini. Ou, dans le cas discret, si

$$\sum_i x_i^2 f_X(x_i)$$

converge vers un nombre fini.

Définition 7.12. *(Covariance.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soient*

$$X, Y : \Omega \to \mathbb{R}$$

deux variables aléatoires à valeurs dans $\mathbb{R}$ *qui chacune admet un moment d'ordre deux. La* covariance *de* X *et* Y *est l'espérance de la variable aléatoire* $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])$ *:*

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

Proposition 7.13. *(Formules pour la covariance.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soient*

$$X, Y : \Omega \to \mathbb{R}$$

deux variables aléatoires à valeurs dans $\mathbb{R}$ *qui chacune admet un moment d'ordre deux. Alors :*

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

(2) $\mathrm{cov}(X, X) = \mathrm{Var}(X)$,

(3) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{cov}(X, Y)$.

Démonstration. Pour prouver la première identité on calcule :

$$\begin{aligned} \mathrm{cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[(XY - X\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

Le deuxième identité découle de la première. Pour prouver la troisième on calcule :

$$\begin{aligned} \mathrm{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{cov}(X, Y). \end{aligned}$$

□

Corollaire 7.14. *(Indépendance implique covariance nulle.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soient*

$$X, Y : \Omega \to \mathbb{R}$$

deux variables aléatoires à valeurs dans $\mathbb{R}$ *qui chacune admet un moment d'ordre deux. Si* X *et* Y *sont indépendantes alors*

$$\mathrm{cov}(X, Y) = 0.$$

Démonstration. On a

$$\begin{aligned} \mathrm{cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= 0, \end{aligned}$$

où la deuxième égalité est vraie car X et Y sont indépendantes (pour des variables indépendantes, l'espérance est multiplicative). □

Exercice 7.15. *Donner un exemple de deux variables aléatoires qui ne sont pas indépendantes mais dont la covariance est tout de même nulle.*

Exemple 7.16. *(Moyenne et variance de la loi de Bernoulli.) Soit X une v.a. qui suit une loi de Bernoulli $B(p,q)$ de paramètres $p,q \geq 0$:*

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

On a donc

$$\mathbb{E}[X]=p$$

et

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

Exemple 7.17. *(Moyenne et variance de la loi binômiale.) Soit X une v.a. qui suit une loi binômiale $B(n,p)$ de paramètre $0 \leq p \leq 1$. Afin de déterminer son espérance et sa variance, on va utiliser le fait très important que la loi de X est identique à la loi de la somme*

$$S_n = X_1 + \cdots + X_n$$

de n variables aléatoires $X_1, \ldots, X_n$ mutuellement indépendantes suivant toute la même loi de Bernoulli de paramètres p,q. Ainsi :

$$\mathbb{E}[X]=\mathbb{E}[S_n]=\sum_{k=1}^{n}\mathbb{E}[X_k]=\sum_{k=1}^{n}p=np.$$

et

$$\mathrm{Var}(X)=\mathrm{Var}(S_n)=\sum_{k=1}^{n}\mathrm{Var}(X_k)=\sum_{k=1}^{n}pq=npq.$$

Exemple 7.18. *(Moyenne et variance de la loi de Poisson.) Soit X une v.a. qui suit une loi de Poisson de paramètre $\lambda > 0$. Ainsi, pour tout $n \in \mathbb{N} \cup \{0\}$*

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

On a :

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

Faisons un calcul préliminaire :

$$\begin{aligned}\sum_{n\geq 1}\frac{\lambda^{n-1}n}{(n-1)!} &= \sum_{n\geq 1}\frac{\frac{d}{d\lambda}\lambda^n}{(n-1)!}\\ &= \frac{d}{d\lambda}\sum_{n\geq 1}\frac{\lambda^n}{(n-1)!}\\ &= \frac{d}{d\lambda}\left(\lambda\sum_{n\geq 1}\frac{\lambda^{n-1}}{(n-1)!}\right)\\ &= \frac{d}{d\lambda}(\lambda e^\lambda)\\ &= e^\lambda(1+\lambda).\end{aligned}$$

Calculons :

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

Donc :

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

Exemple 7.19. *(Moyenne et variance de la loi de normale.) Soit $\mu \in \mathbb{R}$ et $\sigma > 0$. Soit X une v.a. de loi $N(\mu, \sigma)$, c'est-à-dire que pour tout $x \in \mathbb{R}$,*

$$\mathbb{P}(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}dt.$$

Alors

$$\mathbb{E}[X] = \mu,\ \mathrm{Var}(X) = \sigma^2.$$

Pour prouver cela, remarquons qu'il revient au même de prouver que la variable

$$X^* = \frac{X-\mu}{\sigma}$$

a une espérance nulle et une variance égale à 1*. Quelle est la loi de* X^* *? On a pour tout* $y \in \mathbb{R}$ *:*

$$\begin{aligned}
\mathbb{P}(X^* \leq y) &= \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq y\right) = \mathbb{P}(X \leq \sigma y + \mu) \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\sigma y + \mu} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}s^2} \sigma ds \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}s^2} ds.
\end{aligned}$$

C'est-à-dire que X^* *suit une loi normale de type* $N(0,1)$*. Ainsi*

$$\mathbb{E}[X^*] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} dt = 0.$$

D'autre part, en intégrant par partie on obtient :

$$\begin{aligned}
\sqrt{2\pi} \cdot \mathbb{E}[(X^*)^2] &= \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} dt \\
&= \int_{-\infty}^{\infty} t(t e^{-\frac{1}{2}t^2}) dt \\
&= -t e^{-\frac{1}{2}t^2} |_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt \\
&= 0 + \sqrt{2\pi} \\
&= \sqrt{2\pi}.
\end{aligned}$$

Ce qui prouve que

$$\mathbb{E}[(X^*)^2] = 1$$

et donc on obtient

$$\mathrm{Var}(X^*) = \mathbb{E}[(X^*)^2] - \mathbb{E}[X^*] = 1 - 0 = 1.$$

Exemple 7.20. *(Espérance et variance d'une loi exponentielle.) Soit X une v.a. qui suit une loi exponentielle de paramètre $\lambda > 0$. On a :*

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty t\lambda e^{-\lambda t} dt \\ &= (t\lambda)\left(-\frac{e^{-\lambda t}}{\lambda}\right)|_0^\infty + \int_0^\infty e^{-\lambda t} dt \\ &= \int_0^\infty e^{-\lambda t} dt \\ &= \frac{e^{-\lambda t}}{\lambda}|_\infty^0 \\ &= \lambda^{-1}.\end{aligned}$$

Toujours en intégrant par parties, on trouve :

$$\mathbb{E}[X^2] = 2\lambda^{-2}.$$

Ainsi

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

Exercice 7.21. *(Espérance et variance d'une loi exponentielle.) Soit $a < b$ des réels. Soit X une v.a. qui suit une loi uniforme sur l'intervalle $[a, b]$. Prouver que :*

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

8. Les inégalités de Markov et de Bienaymé-Tchebichev

Théorème 8.1. *(Inégalité de Markov.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit*

$$X : \Omega \to \mathbb{R}$$

une variable aléatoire dont le moment d'ordre 1 est fini : $\mathbb{E}[|X|] < \infty$. (Dans le cas discret cela signifie

$$\sum_{i \in \mathbb{N}} |x_i| \mathbb{P}(X = x_i) < \infty.$$

Dans le cas continu, cela signifie

$$\int_{-\infty}^{\infty} |t| f_X(t) dt < \infty.)$$

Alors, pour tout $a > 0$,

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

Démonstration. Considérons le cas discret.

$$\begin{aligned}
\mathbb{E}[|X|] &= \sum_{i\in\mathbb{N}} |x_i|\mathbb{P}(X = x_i) \\
&\geq \sum_{i:|x_i|\geq a} |x_i|\mathbb{P}(X = x_i) \\
&\geq \sum_{i:|x_i|\geq a} a\mathbb{P}(X = x_i) \\
&\geq a \sum_{i:|x_i|\geq a} \mathbb{P}(X = x_i) \\
&= a\mathbb{P}(|X| \geq a).
\end{aligned}$$

Le cas continu est analogue et fait l'objet d'un exercice. □

L'inégalité de Bienaymé-Tchebychev est une inégalité permettant de montrer qu'une variable aléatoire prendra avec une faible probabilité une valeur relativement lointaine de son espérance. Ce résultat s'applique dans des cas très divers, nécessitant la connaissance de peu de propriétés (seules l'espérance et la variance doivent être connues), et permet de démontrer la loi faible des grands nombres.

Théorème 8.2. *(Inégalité de Bienaymé-Tchebichev.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit*

$$X : \Omega \to \mathbb{R}$$

une variable aléatoire dont le moment d'ordre 2 *est fini :* $\mathbb{E}[|X|^2] < \infty$.

Alors, pour tout $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

Exercice 8.3. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit*

$$X : \Omega \to \mathbb{R}$$

une variable aléatoire dont le moment d'ordre 2 *est fini :* $\mathbb{E}[|X|^2] < \infty$. *Supposons que* $\mathbb{E}[X] = 0$. *Prouver que pour tout* $a > 0$,

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

Prouvons l'inégalité de Bienaymé-Tchebichev en nous ramenant à l'inégalité de Markov.

Démonstration. Définissons la variable aléatoire
$$Y = (X - \mathbb{E}[X])^2.$$
Comme $Y \geq 0$, on a :
$$\mathbb{E}[|Y|] = \mathbb{E}[Y] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$
existe bien car par hypothèse $\mathbb{E}[X^2] < \infty$ et l'inégalité de Cauchy-Schwartz montre que
$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|] \leq \sqrt{\mathbb{E}[X^2]} < \infty.$$
Il est donc légitime d'appliquer l'inégalité de Markov à Y. On en déduit ceci :
$$\begin{aligned}\mathbb{P}(|X - \mathbb{E}[X]| \geq a) &= \mathbb{P}(|X - \mathbb{E}[X]|^2 \geq a^2) \\ &= \mathbb{P}(|Y| \geq a^2) \\ &\leq \frac{\mathbb{E}[|Y|]}{a^2} \\ &= \frac{\mathrm{Var}(X)}{a^2}.\end{aligned}$$
□

9. Régression linéaire et coefficients de corrélation

Définition 9.1. *(Coefficients de corrélation.) Soient X et Y des v.a. à valeurs réelles qui admettent un moment d'ordre 2 et dont les variances sont non-nulles. Le* coeficient de corrélation *de X et Y est le nombre :*
$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$
où :
$$\begin{aligned}\sigma_X &= \sqrt{\mathrm{Var}(X)}, \\ \sigma_Y &= \sqrt{\mathrm{Var}(Y)}, \\ \mathrm{cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].\end{aligned}$$
Donc :
$$\rho_{X,Y} = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}\sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2]}}$$

Théorème 9.2. *(Corrélation maximale $\iff$ dépendance linéaire.) Soient X et Y des v.a. à valeurs réelles qui admettent un moment d'ordre 2 et dont les variances sont non-nulles.*

(1) $|\rho_{X,Y}| \leq 1$.

(2) $|\rho_{X,Y}| = 1$ *est équivalent à l'existence de $a, b, c \in \mathbb{R}$ tels que $(a, b, c) \neq (0, 0, 0)$ et tels que $\mathbb{P}(aX + bY + c = 0) = 1$.*

Démonstration. Le premier point est une conséquence immédiate de l'inégalité de Cauchy-Schwartz (exercice) :

$$\begin{aligned}|\mathrm{cov}(X,Y)| &= |\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]| \\ &\leq \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}.\end{aligned}$$

Prouvons le deuxième point. Supposons que

$$|\rho_{X,Y}| = 1.$$

Nous devons établir une relation de dépendance linéaire entre X et Y qui est vraie avec probabilité 1. Il est commode de centrer les variables. On pose

$$X_0 = X - \mathbb{E}[X],\ Y_0 = Y - \mathbb{E}[Y].$$

On définit le polynôme de variable λ :

$$P(\lambda) = \mathbb{E}[(X_0+\lambda Y_0)^2] = \mathbb{E}[Y_0^2]\lambda^2 + 2\mathbb{E}[X_0Y_0]\lambda + \mathbb{E}[X_0^2].$$

Le discriminant de P est

$$\Delta = b^2 - 4ac = (2\mathbb{E}[X_0Y_0])^2 - 4\mathbb{E}[Y_0^2]\mathbb{E}[X_0^2].$$

Or par hypothèse,

$$1 = \rho_{X,Y}^2 = \frac{\mathrm{cov}(X,Y)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} = \frac{\mathbb{E}[X_0Y_0]^2}{\mathbb{E}[Y_0^2]\mathbb{E}[X_0^2]}.$$

On en déduit que

$$\Delta = 0.$$

Donc P possède une unique racine λ_0 qui est double. Ainsi

$$0 = P(\lambda_0) = \mathbb{E}[(X_0+\lambda_0Y_0)^2] = \mathbb{E}[(X_0+\lambda_0Y_0)^2] - \mathbb{E}[X_0+\lambda_0Y_0]^2 = \mathrm{Var}(X_0+\lambda_0Y_0).$$

On peut appliquer l'inégalité de Tchebichev à la v.a. $X_0+\lambda_0Y_0$ puisque son moment d'ordre 2 est fini (égal à zéro) : pour tout $a>0$,

$$\mathbb{P}(|X_0+\lambda_0Y_0| \geq a) \leq \frac{\mathrm{Var}(X_0+\lambda_0Y_0)}{a^2} = 0.$$

Donc (exercice)

$$\mathbb{P}(|X_0+\lambda_0Y_0| = 0) = 1.$$

On conclu qu'avec probabilité 1,

$$(X-\mathbb{E}[X]) + \lambda_0(Y-\mathbb{E}[Y]) = 0,$$

ce qui établi la relation linéaire cherchée entre X et Y.

Réciproquement, supposons qu'il existe $a,b,c\in\mathbb{R}$ tels que

$$\mathbb{P}(aX+bY+c=0) = 1.$$

Nous devons prouver que
$$\rho_{X,Y} = \pm 1.$$
On commence par traiter le cas particulier suivant : on suppose qu'il existe $\alpha \neq 0$ et $\beta \in \mathbb{R}$ tels que
$$\mathbb{P}(X = \alpha Y + \beta) = 1.$$
Si tel est le cas, on a avec probabilité 1 :
$$X_0 = X - \mathbb{E}[X] = \alpha Y + \beta - \mathbb{E}[\alpha Y + \beta] = \alpha Y_0.$$
Donc
$$\mathrm{cov}(X, Y) = \mathbb{E}[X_0 Y_0] = \alpha \mathbb{E}[Y_0^2] = \alpha \mathrm{Var}(Y).$$
D'autre part,
$$\mathrm{Var}(X) = \mathrm{Var}(\alpha Y + \beta) = \alpha^2 \mathrm{Var}(Y).$$
Finalement :
$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\alpha \mathrm{Var}(Y)}{\sqrt{\alpha^2 \mathrm{Var}(Y) \mathrm{Var}(Y)}} = \pm 1.$$
Pour terminer la preuve, il suffit donc de prouver que l'on peut toujours se ramener au cas particulier traité.

Supposons donc qu'il existe $a, b, c \in \mathbb{R}$ tels que
$$\mathbb{P}(aX + bY + c = 0) = 1.$$

Premier cas : $c \neq 0$. Dans ce cas on a forcément $a \neq 0$ et $b \neq 0$. En effet, montrons que si on avait par exemple $a = 0$ alors on obtiendrait une contradiction (le cas $b = 0$ se traite de manière analogue). On aurait
$$\mathbb{P}(bY + c = 0) = 1.$$
Donc
$$0 = \mathrm{Var}(bY + c) = b^2 \mathrm{Var}(Y)$$
et comme par hypothèse, $\mathrm{Var}(Y) \neq 0$, on aurait $b = 0$. Donc
$$\mathbb{P}(c = 0) = 1,$$
ce qui est contraire à l'hypothèse $c \neq 0$. On peut donc diviser par a :
$$aX + bY + c = 0$$
est équivalent à
$$X = -\frac{b}{a} Y - \frac{c}{b}$$

où $-\frac{b}{a} \neq 0$.

Deuxième cas : $c = 0$. Ainsi,

$$\mathbb{P}(aX + bY = 0) = 1.$$

Comme on a supposé $(a, b, c) \neq (0, 0, 0)$ on a soit $a \neq 0$ soit $b \neq 0$. Supposons $a \neq 0$ (le cas $b \neq 0$ est analogue). En divisant par a, la condition

$$aX + bY = 0$$

devient

$$X = -\frac{b}{a}Y.$$

Pour achever la preuve il suffit donc de prouver que $-\frac{b}{a} \neq 0$. Mais si on avait $-\frac{b}{a} = 0$, alors avec probabilité 1 on aurait $X = 0$ et donc on aurait $\mathrm{Var}(X) = 0$ ce qui est contraire aux hypothèses. $\square$

Théorème 9.3. *(Régression linéaire.) Soient X et Y des v.a. à valeurs réelles qui admettent un moment d'ordre 2. On suppose que la variance de X est non-nulle. Alors la fonction*

$$\mathbb{R}^2 \to \mathbb{R}$$

$$(a, b) \mapsto \mathbb{E}[(Y - (aX + b))^2]$$

admet un minimum global au point $(a_0, b_0) \in \mathbb{R}^2$ où

$$a_0 = \frac{\mathrm{cov}(X, Y)}{\mathrm{Var}X},$$

$$b_0 = \mathbb{E}[Y] - \frac{\mathrm{cov}(X, Y)}{\mathrm{Var}X}\mathbb{E}[X]$$

Remarque 9.4. *On a :*

$$a_0 = \frac{\mathrm{cov}(X, Y)}{\mathrm{Var}X} = \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}.$$

On peut interpréter le théorème ci-dessus en disant que la variable aléatoire

$$a_0X + b_0$$

est la « meilleure » approximation linéaire (on devrait dire affine) en X de la variable aléatoire Y. Le qualificatif « meilleure » est à comprendre au sens de la méthode des moindres carrés. *La* régression linéaire *consiste à remplacer Y par son approximation linéaire $a_0X + b_0$.*

Démonstration. On commence par centrer les variables :

$$X_0 = X - \mathbb{E}[X],$$
$$Y_0 = Y - \mathbb{E}[Y].$$

On calcule :

$$\mathbb{E}[(Y - (aX + b))^2] = \mathbb{E}[(Y_0 + \mathbb{E}[Y] - (aX_0 + a\mathbb{E}[X] + b))^2] =$$
$$\mathbb{E}[(Y_0 - aX_0) + (\mathbb{E}[Y] - a\mathbb{E}[X] - b))^2] =$$
$$\mathbb{E}[(Y_0 - aX_0)^2] + 2\mathbb{E}[Y_0 - aX_0](\mathbb{E}[Y] - a\mathbb{E}[X] - b) + (\mathbb{E}[Y] - a\mathbb{E}[X] - b)^2 =$$
$$\mathbb{E}[(Y_0 - aX_0)^2] + (\mathbb{E}[Y] - a\mathbb{E}[X] - b)^2.$$

Pour tout a fixé, on voit que l'unique valeur de b qui minimise cette expression est

$$b(a) = \mathbb{E}[Y] - a\mathbb{E}[X].$$

Afin de trouver un minimum global nous cherchons a qui minimise

$$\mathbb{E}[(Y - (aX + b(a)))^2] = \mathbb{E}[(Y_0 - aX_0)^2].$$

Or

$$\begin{aligned}\mathbb{E}[(Y_0 - aX_0)^2] &= \mathbb{E}[Y_0^2] - 2a\mathbb{E}[Y_0X_0] + a^2\mathbb{E}[X_0^2] \\ &= \mathrm{Var}Y - 2a\mathrm{cov}(X, Y) + a^2\mathrm{Var}X \\ &= P(a)\end{aligned}$$

est un polynôme du second degré en a dont le coefficient du terme de degré 2 est strictement positif par hypothèse. Dérivons relativement à a afin de trouver l'unique minimum :

$$2a\mathrm{Var}X = 2\mathrm{cov}(X, Y),$$

donc le minimum est obtenu lorsque a prend la valeur

$$a_0 = \frac{\mathrm{cov}(X, Y)}{\mathrm{Var}X}.$$

□

10. Loi faible des grands nombres

La loi des grands nombres permet d'interpréter la probabilité comme une fréquence de réalisation, justifiant ainsi le principe des sondages, et présente l'espérance comme une moyenne. Plus formellement, elle signifie que la moyenne empirique, calculée sur les valeurs d'un échantillon, converge vers l'espérance lorsque la taille de l'échantillon tend vers l'infini.

Théorème 10.1. *(Loi faible des grands nombres.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$ une famille de variables aléatoires sur Ω à valeurs dans $\mathbb{R}$. On suppose que ces variables sont mutuellement indépendantes. On suppose aussi que chaque variable admet un moment d'ordre 2 :*

$$\forall n \in \mathbb{N},\ \mathbb{E}[X_n^2] < \infty.$$

Supposons qu'il existe $m \in \mathbb{R}$ tel que

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[X_k] = m.$$

Supposons que

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} \mathrm{Var} X_k = 0.$$

Alors pour tout $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{k=1}^{n} X_k - m \right| > \epsilon \right) = 0.$$

Corollaire 10.2. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$ une famille de variables aléatoires sur Ω à valeurs dans $\mathbb{R}$. On suppose que ces variables sont mutuellement indépendantes, qu'elles suivent une même loi telle que l'espérance*

$$\mathbb{E}[X_1] = m$$

est finie et telle que

$$\mathrm{Var} X_1 < \infty.$$

Alors pour tout $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{k=1}^{n} X_k - m \right| > \epsilon \right) = 0.$$

Le corollaire est une conséquence immédiate du corollaire.

Démonstration. Afin de prouver le théorème, fixons $\epsilon > 0$ et supposons que n est tel que

$$\left| \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[X_k] - m \right| \leq \epsilon/2.$$

Définissons la variable aléatoire

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

On a :

$$\left|\overline{X}_n - m\right| \leq \left|\overline{X}_n - \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k]\right| + \left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k] - m\right|.$$

Pour cet ϵ et ce n on a donc :

$$\left(\left|\overline{X}_n - \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k]\right| \leq \epsilon/2\right) \subset \left(\left|\overline{X}_n - m\right| \leq \epsilon\right).$$

En passant aux complémentaires on obtient l'inclusion :

$$\left(\left|\overline{X}_n - m\right| > \epsilon\right) \subset \left(\left|\overline{X}_n - \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k]\right| > \epsilon/2\right).$$

D'autre part, l'inégalité de Tchebichev s'applique et montre que

$$\mathbb{P}\left(\left|\overline{X}_n - \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k]\right| > \epsilon/2\right) \leq \frac{\mathrm{Var}\overline{X}_n}{(\epsilon/2)^2}.$$

Comme les variables sont indépendantes,

$$\mathrm{Var}\overline{X}_n = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}X_k.$$

On conclut que si $\epsilon > 0$ et $n \in \mathbb{N}$ sont tels que

$$\left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k] - m\right| \leq \epsilon/2,$$

alors

$$\mathbb{P}\left(\left|\overline{X}_n - m\right| > \epsilon\right) \leq \frac{4}{\epsilon^2}\frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}X_k.$$

On en déduit le théorème. □

11. Loi de Poisson résultant de lois exponentielles

Théorème 11.1. *(Caractérisation probabiliste de la loi exponentielle.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit*

$$T : \Omega \to]0, \infty[$$

une variable aléatoire à valeurs réelles strictement positives. Supposons que la loi de T *vérifie les conditions suivantes.*

(1) $\forall s, t \geq 0$, *la probabilité que* T *dépasse la valeur* $s + t$ *sachant que* T *dépasse la valeur* s *est égale à la valeur que* T *dépasse la valeur* t *:*

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

(2) *Pour tout* $t > 0$, $\mathbb{P}(T > t) > 0$.

(3) *La fonction*

$$[0, \infty[\to [0, 1]$$
$$t \mapsto \mathbb{P}(T > t)$$

est continue par morceaux (une hypothèse beaucoup moins forte, suffisante à entraîner les mêmes conclusions, est que $t \mapsto \mathbb{P}(T > t)$ *est mesurable).*

Alors il existe $a > 0$ *tel que* $\forall t \geq 0$,

$$\mathbb{P}(T \leq t) = 1 - e^{-at}.$$

Démonstration. Notons U la fonction

$$U : [0, \infty[\to]0, 1]$$
$$t \mapsto U(t) = \mathbb{P}(T > t).$$

On a donc

$$U(0) = \mathbb{P}(T > 0) = 1$$

puisque T est à valeurs strictement positives.

Soit $s, t \geq 0$. Comme $t \geq 0$, on a une inclusion évidente entre les événements suivants :

$$(T > s + t) \subset (T > s)$$

et donc

$$(T > s + t) \cap (T > s) = (T > s + t).$$

Ainsi :

$$\begin{aligned} U(t) &= \mathbb{P}(T > t) \\ &= \mathbb{P}(T > s + t | T > s) \\ &= \frac{\mathbb{P}\left((T > s + t) \cap (T > s)\right)}{\mathbb{P}(T > s)} \\ &= \frac{\mathbb{P}(T > s + t)}{\mathbb{P}(T > s)} \\ &= \frac{U(s + t)}{U(s)}. \end{aligned}$$

On conclut que la fonction U satisfait la condition suivante :

$$\forall s, t \geq 0, \, U(s + t) = U(s)U(t).$$

Exercice 11.2. *Soit* $U : [0, \infty[\to]0, 1]$ *une fonction continue par morceaux telle que*

$$\forall s, t \geq 0, \, U(s+t) = U(s)U(t).$$

Alors il existe $a \geq 0$ *tel que*

$$U(t) = e^{-at}.$$

En appliquant le résultat de l'exercice, on a donc l'existence de $a \geq 0$ tel que pour tout $t \geq 0$,

$$\mathbb{P}(T > t) = U(t) = e^{-at}.$$

Donc pour tout $t \geq 0$,

$$\mathbb{P}(T \leq t) = 1 - e^{-at}.$$

Le cas $a = 0$ est exclu car on aurait pour tout $t \geq 0$,

$$\mathbb{P}(T \leq t) = 0$$

ce qui impliquerait (nouvel exercice)

$$\mathbb{P}(0 < T < \infty) = 0$$

ce qui est absurde puisque

$$\mathbb{P}(0 < T < \infty) = 1.$$

□

Exemple 11.3. *Le temps d'attente* T *entre deux désintégrations d'atomes d'un matériau radioactif, l'intervalle* T *entre deux appels que reçoit une centrale téléphonique, sont des phénomènes qui peuvent être modélisés par une variable aléatoire* T *de loi exponentielle. En effet, l'expérience montre que dans un matériau radioactif, la probabilité de désintégration d'un atome par unité de temps reste constante au cours du temps (en vérité elle décroît très lentement comme on sait).*

Proposition 11.4. *(Convolution de deux densités.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soient*

$$X, Y : \Omega \to \mathbb{R}$$

deux variables aléatoires à valeurs réelles. Supposons que la loi de X *est définie par une densité de probabilité* f_X *et que la loi de* Y *est définie par une densité de probabilité* f_Y*. Si* X *et* Y *sont indépendantes, alors la loi de* $X + Y$ *est définie par la densité* f_{X+Y} *définie par la formule :*

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x) dx.$$

Remarque 11.5. *On note*

$$(f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x) dx$$

et la fonction obtenue $f_X * f_Y$ *est le* produit de convolution *de* f_X *avec* f_Y*.*

Démonstration. (La preuve que nous donnons ne fonctionne que sous des hypothèses trop restrictives de régularité sur f_X et f_Y que nous ne cherchons pas à expliciter mais qui serons vérifiées dans toutes les applications que nous allons considérer.) Soit $z \in \mathbb{R}$ fixé. Notons

$$\mathbf{1}_{\{(u,v):u+v\leq z\}}(x,y)$$

la fonction caractéristique de la portion du plan $\mathbb{R}^2$ d'équation

$$u + v \leq z.$$

Comme X et Y sont indépendantes

$$\begin{aligned}
\mathbb{P}(X+Y \leq z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(x) f_Y(y) \mathbf{1}_{\{(u,v):u+v\leq z\}}(x,y) dx dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_X(x) f_Y(y) dy \right) dx \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_Y(y) dy \right) f_X(x) dx \\
&= \int_{-\infty}^{\infty} F_Y(z-x) f_X(x) dx
\end{aligned}$$

Faisons les hypothèses de régularités suffisantes sur f_Y afin de pouvoir dériver ainsi :

$$\begin{aligned}
f_{X+Y}(z) &= \frac{d}{dz} \mathbb{P}(X+Y \leq z) \\
&= \frac{d}{dz} \int_{-\infty}^{\infty} F_Y(z-x) f_X(x) dx \\
&= \int_{-\infty}^{\infty} \frac{d}{dz} F_Y(z-x) f_X(x) dx \\
&= \int_{-\infty}^{\infty} f_Y(z-x) f_X(x) dx.
\end{aligned}$$

□

Exercice 11.6. *Etablir le résultat analogue dans le cas de variables aléatoires discrètes.*

Proposition 11.7. *(Loi d'une somme de variables exponentielles indépendantes.) Soient $X_1, \dots, X_n$ une famille de n variables aléatoires de loi exponentielle de même paramètre $a > 0$. C'est à dire que pour tout $1 \leq k \leq n$,*

$$\mathbb{P}(X_k \leq x) = 1 - e^{-ax}$$

Soit $S_n = X_1 + \cdots + X_n$ la variable aléatoire obtenue en sommant les variables données. Si les variables aléatoires $X_1, \dots, X_n$ sont mutuellement indépendantes alors

$$\begin{aligned}\mathbb{P}(S_n \leq x) &= \frac{a^n}{(n-1)!} \int_0^x s^{n-1} e^{-as} ds \\ &= 1 - e^{-ax} \sum_{k=1}^{n-1} \frac{(ax)^k}{k!}.\end{aligned}$$

Démonstration. Procédons par récurence sur n. Si $n = 1$ les égalités suivantes sont vraies :

$$\mathbb{P}(S_1 \leq x) = \mathbb{P}(X_1 \leq x) = a \int_0^x e^{-as} ds = 1 - e^{-ax}.$$

Supposons par induction que pour tout $t \geq 0$,

$$f_{S_n}(t) = \frac{a^n}{(n-1)!} t^{n-1} e^{-at},$$

et calculons pour $x \geq 0$

$$\begin{aligned}f_{S_{n+1}}(x) &= f_{S_n + X_1}(x) \\ &= \int_{-\infty}^{\infty} f_{X_1}(x-t) f_{S_n}(t) dt \\ &= \int_0^x f_{X_1}(x-t) f_{S_n}(t) dt \\ &= \int_0^x a e^{-a(x-t)} \frac{a^n}{(n-1)!} t^{n-1} e^{-at} dt \\ &= \frac{a^{n+1}}{(n-1)!} e^{-ax} \int_0^x t^{n-1} dt \\ &= \frac{a^{n+1}}{(n-1)!} e^{-ax} \frac{x^n}{n} \\ &= \frac{a^{n+1}}{n!} x^n e^{-ax},\end{aligned}$$

ce qui établit par récurrence la première égalité de la proposition. Afin de prouver la deuxième égalité, à savoir : pour tout $x \geq 0$,

$$\frac{a^n}{(n-1)!}\int_0^x s^{n-1}e^{-as}ds = 1 - e^{-ax}\sum_{k=1}^{n-1}\frac{(ax)^k}{k!},$$

remarquons que ces deux fonctions de x tendent vers 1 lorsque x tend vers $+\infty$:

$$\lim_{x\to\infty}\frac{a^n}{(n-1)!}\int_0^x s^{n-1}e^{-as}ds = \lim_{x\to\infty}\mathbb{P}(S_n \leq x) = 1,$$

et

$$\lim_{x\to\infty} 1 - e^{-ax}\sum_{k=1}^{n-1}\frac{(ax)^k}{k!} = 1 - \lim_{x\to\infty} e^{-ax}\sum_{k=1}^{n-1}\frac{(ax)^k}{k!} = 1.$$

Pour démontrer qu'elles sont égales il suffit donc de prouver que leurs dérivées coïncident. C'est bien le cas : pour tout $x \geq 0$, pour tout $a > 0$ pour tout entier $n \geq 1$,

$$\frac{a^n}{(n-1)!}x^{n-1}e^{-ax} = -\frac{d}{dx}\left(e^{-ax}\sum_{k=1}^{n-1}\frac{(ax)^k}{k!}\right).$$

(Vérification à faire en exercice.) □

Théorème 11.8. *(Loi de Poisson et nombre d'occurence d'évènements indépendants de même loi exponentielle.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient*

$$T_1, \ldots, T_n : \Omega \to]0, \infty[$$

une famille de n variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, de même loi exponentielle de paramètre $a > 0$. C'est-à-dire que pour tout $1 \leq k \leq n$,

$$\mathbb{P}(T_k \leq x) = 1 - e^{-ax}.$$

Soit $S_n = T_1 + \cdots + T_n$ la variable aléatoire obtenue en sommant les variables données. Pour tout $t > 0$, on définit la variable aléatoire

$$N_t : \Omega \to \mathbb{N} \cup \{0\}$$

$$N_t(\omega) = |\{k \in \mathbb{N} \cup \{0\} : S_k(\omega) \leq t\}|.$$

Si les variables $T_1, \ldots, T_n$ sont mutuellement indépendantes alors la variable N_t suit une loi de Poisson de paramètre $\lambda = at$. C'est-à-dire que pour tout entier $n \geq 0$,

$$\mathbb{P}(N_t = n) = e^{-at}\frac{(at)^n}{n!}.$$

Démonstration. Pour tout entier $k \geq 1$, pour tout $\omega \in \Omega$, pour tout $t > 0$,

$$0 \leq X_k(\omega).$$

Donc

$$S_{k-1}(\omega) \leq S_{k-1}(\omega) + X_k(\omega) = S_k(\omega).$$

Ainsi pour tout $t > 0$,

$$(S_k(\omega) \leq t) \implies (S_{k-1}(\omega) \leq t).$$

On en déduit que pour tout entier $n \geq 0$,

$$(N_t(\omega) = n) = [(S_n(\omega) \leq t) \setminus (S_{n+1}(\omega) \leq t)].$$

Donc

$$\begin{aligned}
\mathbb{P}(N_t(\omega) = n) &= \mathbb{P}(S_n(\omega) \leq t) - P(S_{n+1}(\omega) \leq t) \\
&= \left(1 - e^{-at} \sum_{k=1}^{n-1} \frac{(at)^k}{k!}\right) - \left(1 - e^{-at} \sum_{k=1}^{n} \frac{(at)^k}{k!}\right) \\
&= e^{-at} \frac{(at)^n}{n!}.
\end{aligned}$$

□

Remarque 11.9. *Ce théorème explique pourquoi le nombre de désintégrations radioactives dans un intervalle de temps de durée t ou le nombre d'appels dans un intervalle de temps de durée t sont modélisés par une variable aléatoire N_t qui suit une loi de Poisson.*

12. Le théorème central limite

Théorème 12.1. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Pour tout entier $n \geq 1$, soit*

$$X_n : \Omega \to \mathbb{R}$$

une variable aléatoire définie sur Ω à valeurs réelles. Supposons que les variables aléatoires $X_1, \ldots, X_n, \ldots$ ont toutes la même loi. On suppose que leur espérance (commune)

$$\mu = \mathbb{E}[X_n]$$

existe et est finie, on suppose que leur variance (commune) satisfait

$$0 < \sigma = \sqrt{\mathrm{Var} X_n} < \infty.$$

Soit

$$S_n = X_1 + \cdots + X_n,$$

la variable aléatoire sur Ω définie comme la somme des n premières variables aléatoires de la famille donnée. Si les variables aléatoires $X_1, \ldots, X_n, \ldots$ sont mutuellement indépendantes, alors $\forall a < b$,

$$\lim_{n\to\infty} \mathbb{P}\left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} < b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{\frac{-x^2}{2}}dx,$$

Remarque 12.2. *On a :*

$$\mathbb{E}[S_n] = n\mu.$$

Comme les variables sont mutuellement indépendantes, on a aussi :

$$\mathrm{Var} S_n = n\sigma^2.$$

Donc la variable centrée réduite

$$S_n^* = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var} S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

permet de reformuler la conclusion du théorème ainsi :

$$\lim_{n\to\infty} \mathbb{P}\left(a < S_n^* < b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{\frac{-x^2}{2}}dx.$$

Définition 12.3. *On définit $\Phi : \mathbb{R} \to]0,1[$ par la formule*

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

En d'autres termes, Φ est la fonction de répartition F_X d'une variable aléatoire X qui suit une loi normale de moyenne $\mu = 0$ et de d'écart-type $\sigma = 1$:

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

Exercice 12.4. *Prouver que pour tout $x \geq 0$,*

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

Avant de prouver le théorème central limite, nous en présentons une application dans la proposition suivante.

Proposition 12.5. *(Perturbation par une loi uniforme.) Soit $\mu \in \mathbb{R}$. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient*

$$Y_1, \ldots, Y_n : \Omega \to [-1, 1]$$

des variables aléatoires mutuellement indépendantes sur Ω, de loi uniforme sur $[-1,1]$. On définit pour tout entier $1 \leq k \leq n$ la variable aléatoire

$$X_k = \mu + Y_k.$$

Soit $S_n = \sum_{k=1}^n X_k$. *Soit* $0 \leq a$ *un nombre réel. Alors*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \leq a\right) = 2\Phi(a\sqrt{3}) - 1$$

Démonstration. Comme pour tout $1 \leq k \leq n$, on sait que $\mathbb{E}[Y_k] = 0$, on obtient :

$$\mathbb{E}[S_n] = \mathbb{E}[\sum_{k=1}^n X_k] = \sum_{k=1}^n \mathbb{E}[X_k] = \sum_{k=1}^n \mathbb{E}[\mu + Y_k] = n\mu.$$

Pour tout $1 \leq k \leq n$, on sait que

$$\mathrm{Var} X_k = \mathrm{Var} Y_k = \frac{(1-(-1)^2}{12} = \frac{1}{3}.$$

Ainsi, comme les variables sont mutuellement indépendantes,

$$\mathrm{Var} S_n = \frac{n}{3}.$$

Les évènements suivants sont donc égaux :

$$\left(\left|\frac{S_n}{n} - \mu\right| \leq a\right) = \left(\left|\frac{S_n - n\mu}{\sqrt{\frac{n}{3}}}\right| \leq a\sqrt{3n}\right) = \left(\left|\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var} S_n}}\right| \leq a\sqrt{3n}\right).$$

Les variables aléatoires $X_1, \ldots, X_n$ sont mutuellement indépendantes car les v.a. $Y_1, \ldots, Y_n$ le sont (exercice). On peut donc appliquer le théorème central limite :

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \leq a\right) &= \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var} S_n}}\right| \leq a\sqrt{3n}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-a\sqrt{3n}}^{a\sqrt{3n}} e^{\frac{-x^2}{2}} dx \\ &= 2\Phi(a\sqrt{3}) - 1.\end{aligned}$$

□

Remarque 12.6. *Une interprétation de la Proposition 12.5 est la suivante. On effectue une série de n mesures répétées* $X_1, \ldots, X_n$ *pour trouver une valeur cherchée* μ *(température, poids, vitesse, etc.). La mesure* X_k *est perturbée par une erreur* Y_k. *Ainsi*

$$X_k = \mu + Y_k.$$

Afin d'estimer la véritable valeur μ, *on calcule la moyenne des mesures :*

$$\frac{X_1 + \cdots X_n}{n} = \frac{S_n}{n}$$

Quelle est la probabilité que la valeur trouvée s'écarte de μ de moins de $a \geq 0$? Réponse :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \leq a\right).$$

Si on suppose que toutes les erreurs suivent la même loi et si l'on connait l'espérance et la variance des erreurs et si on sait que les erreurs sont mutuellement indépendantes, alors, si n est grand, on peut estimer cette probabilité à l'aide du théorème central limite. Dans la Proposition 12.5 on a supposé que les erreurs suivent toutes la loi uniforme et sont mutuellement indépendantes.

13. Fonctions caractéristiques

Définition 13.1. *(Fonction caractéristique.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire sur Ω à valeurs réelles. La* fonction caractéristique *de X est par définition la fonction*

$$\Phi_X : \mathbb{R} \to \mathbb{C}$$

$$t \mapsto \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)].$$

Exemple 13.2. *(Fonction caractéristique d'une variable de Bernoulli.) Soit X une variable aléatoire qui suit une loi de Bernoulli :*

$$\mathbb{P}(X = 1) = p, \, \mathbb{P}(X = 0) = q,$$

avec $p, q \geq 0$ et $p + q = 1$.

$$\Phi_X(t) = \mathbb{E}[e^{itX}] = pe^{it} + qe^{it0} = q + pe^{it}.$$

Proposition 13.3. *(Multiplicativité des fonctions caractéristiques.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X, Y des v.a. définies sur Ω à valeurs dans $\mathbb{R}$. Si X et Y sont indépendantes alors*

$$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t).$$

Démonstration.

$$\Phi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}e^{itY}] = \mathbb{E}[e^{itX}]\mathbb{E}[e^{itY}] = \Phi_X(t)\Phi_Y(t).$$

□

Exemple 13.4. *(Fonction caractéristique d'une variable binomiale.) Soit S_n une variable aléatoire de loi binomiale de type (n, p). En d'autres termes, pour tout entier $0 \leq k \leq n$*

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

où $0 \leq p \leq 1$. On sait que la loi de S_n est la loi de la somme de n variables aléatoire mutuellement indépendantes $X_1, \ldots, X_n$ qui suivent toutes la même loi de Bernoulli de paramètre p. Ainsi :

$$\Phi_{S_n}(t) = \Phi_{X_1}(t) \cdots \Phi_{X_n}(t) = (q + pe^{it})^n.$$

Proposition 13.5. *(Homogénéité des fonctions caractéristiques.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une v.a. définie sur Ω à valeurs dans $\mathbb{R}$. Soit $a, b \in \mathbb{R}$. On a :*

$$\Phi_{aX+b}(t) = e^{itb}\Phi_X(at).$$

Démonstration. $\Phi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = e^{itb}\mathbb{E}[e^{itaX}] = e^{itb}\Phi_X(at)$. □

Proposition 13.6. *(Un calcul formel : la fonction moment d'une loi à densité.) Soit X une variable aléatoire à valeurs réelles dont la loi est définie par une fonction de densité f_X. Soit $z \in \mathbb{C}$. Considérons l'expression formelle suivante*

$$M(z) = \mathbb{E}[e^{zX}] = \int_{-\infty}^{\infty} e^{zx} f_X(x)dx,$$

sans se soucier pour l'instant des questions de convergence. Soit $h \in \mathbb{C} \setminus \{0\}$. Toujours sans se soucier des questions de convergence, on a alors :

$$\frac{M(z+h) - M(z)}{h} = \mathbb{E}[Xe^{zX}] + \int_{-\infty}^{\infty} \left(\frac{e^{hx} - 1}{hx} - 1 \right) xe^{zx} f_X(x)dx.$$

Démonstration.

$$\begin{aligned}
\frac{M(z+h)-M(z)}{h} &= \int_{-\infty}^{\infty} \frac{e^{(z+h)x}-e^{zx}}{h} f_X(x)dx = \int_{-\infty}^{\infty} \frac{e^{hx}-1}{h} e^{zx} f_X(x)dx \\
&= \int_{-\infty}^{\infty} \frac{\sum_{n\geq 0} \frac{(hx)^n}{n!} - 1}{h} e^{zx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{\sum_{n\geq 1} \frac{(hx)^n}{n!}}{h} e^{zx} f_X(x)dx \\
&= \int_{-\infty}^{\infty} \left(\sum_{n\geq 1} \frac{h^{n-1}x^n}{n!} \right) e^{zx} f_X(x)dx = \int_{-\infty}^{\infty} \left(x + \sum_{n\geq 2} \frac{h^{n-1}x^n}{n!} \right) e^{zx} f_X(x)dx \\
&= \int_{-\infty}^{\infty} x e^{zx} f_X(x)dx + \int_{-\infty}^{\infty} \left(\sum_{n\geq 2} \frac{h^{n-1}x^n}{n!} \right) e^{zx} f_X(x)dx \\
&= \mathbb{E}[Xe^{zX}] + \int_{-\infty}^{\infty} \left(\sum_{n\geq 2} \frac{h^{n-1}x^n}{n!} \right) e^{zx} f_X(x)dx \\
&= \mathbb{E}[Xe^{zX}] + \int_{-\infty}^{\infty} \left(\sum_{n\geq 2} \frac{(hx)^{n-1}}{n!} \right) x e^{zx} f_X(x)dx \\
&= \mathbb{E}[Xe^{zX}] + \int_{-\infty}^{\infty} \left(\frac{1}{hx} \sum_{n\geq 2} \frac{(hx)^n}{n!} \right) x e^{zx} f_X(x)dx \\
&= \mathbb{E}[Xe^{zX}] + \int_{-\infty}^{\infty} \left(\frac{1}{hx} (e^{hx} - 1 - hx) \right) x e^{zx} f_X(x)dx \\
&= \mathbb{E}[Xe^{zX}] + \int_{-\infty}^{\infty} \left(\frac{e^{hx}-1}{hx} - 1 \right) x e^{zx} f_X(x)dx.
\end{aligned}$$

□

Théorème 13.7. *(Moments et fonction caractéristique.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* X *une v.a. définie sur* Ω *à valeurs dans* $\mathbb{R}$. *Supposons que pour tout entier* $0 \leq k \leq n$,

$$\mathbb{E}[|X^k|] < \infty.$$

Alors

$$\Phi_X(t) = \sum_{k=0}^{n} \frac{i^k \mathbb{E}[X^k]}{k!} t^k + o(t^n).$$

en particulier, la dérivée d'ordre k *de* Φ_X *en* 0 *vérifie*

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

Démonstration. On développe $\Phi_X(t)$ en 0 selon Taylor. On a $\Phi_X(0) = 1$. De plus pour tout entier $k \geq 1$, un calcul formel montre que la dérivée numéro k de Φ_X évaluée au point 0 est

$$\Phi_X^{(k)}(0) = i^k \mathbb{E}[X^k],$$

et le calcul est justifié si $\mathbb{E}[|X^k|]$ est fini. Regardons les détails lorsque $k = 1$. Soit $t \in \mathbb{R}$ et $s \in \mathbb{R} \setminus \{0\}$. On a :

$$\begin{aligned} \frac{\Phi_X(t+s) - \Phi_X(t)}{s} &= i\frac{M(it+is) - M(it)}{is} \\ &= i\mathbb{E}[Xe^{itX}] + \int_{-\infty}^{\infty} \left(\frac{e^{isx} - 1}{isx} - 1 \right) xe^{itx} f_X(x) dx. \end{aligned}$$

En faisant tendre s vers zéro, si $\mathbb{E}[|X|]$ est fini, on déduit que

$$\Phi'_X(t) = i\mathbb{E}[Xe^{itX}].$$

Si $\mathbb{E}[|X^k|]$ est fini, on déduit par le même raisonnement que

$$\Phi_X^{(k)}(t) = i^k \mathbb{E}[X^k e^{itX}].$$

$\square$

Lemme 13.8. *(La fonction moment d'une loi normale est $\mathbb{C}$-différentiable.) Soit X une variable aléatoire doit la loi est normale :*

$$X \sim N(0,1).$$

Soit $z \in \mathbb{C}$. Alors l'espérance

$$\mathbb{E}[Xe^{zX}]$$

est bien définie et la fonction

$$M : \mathbb{C} \to \mathbb{C}$$

$$z \mapsto \mathbb{E}[e^{zX}]$$

est $\mathbb{C}$-dérivable en z, c'est-à-dire que pour tout $z \in \mathbb{C}$, la limite

$$M'(z) = \lim_{h \in \mathbb{C}, h \to 0} \frac{M(z+h) - M(z)}{h}$$

existe (ce qui signifie en particulier que cette limite, qui est un nombre complexe, ne dépend pas de la manière dont le nombre complexe h tend vers 0) et

$$M'(z) = \mathbb{E}[Xe^{zX}].$$

Démonstration. On sait que pour tout $z \in \mathbb{C}$, pour tout $h \in \mathbb{C} \setminus \{0\}$:
$$\frac{M(z+h)-M(z)}{h} = \mathbb{E}[Xe^{zX}] + \int_{-\infty}^{\infty} \left(\frac{e^{hx}-1}{hx} - 1\right) xe^{zx} f_X(x)dx,$$
où
$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$
La décroissance de f_X lorsque $|x| \to \infty$ implique que pour tout entier $n \geq 0$,
$$\mathbb{E}[X^n e^{zX}]$$
est fini. En faisant tendre h vers 0, on obtient :
$$M'(z) = \mathbb{E}[Xe^{zX}].$$
□

Proposition 13.9. *(Fonction caractéristique d'une variable aléatoire suivant une loi normale.) Soit X une variable aléatoire doit la loi est normale :*
$$X \sim N(0,1).$$
Alors
$$\Phi_X(t) = e^{-\frac{t^2}{2}}.$$

Démonstration. Considérons la fonction moment
$$M : \mathbb{C} \to \mathbb{C}$$
$$z \mapsto \mathbb{E}[e^{zX}] = \int_{-\infty}^{\infty} e^{zx} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx.$$
Pour tout $x, t \in \mathbb{R}$,
$$tx - \frac{x^2}{2} = \frac{t^2}{2} - \frac{(x-t)^2}{2}.$$
Donc :
$$\begin{aligned} M(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{x^2}{2}} dx \\ &= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} dx \\ &= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \\ &= e^{t^2/2}. \end{aligned}$$
Définissons la fonction
$$h : \mathbb{C} \to \mathbb{C}$$

$$z \mapsto h(z) = M(z)e^{-z^2/2} - 1.$$

La fonction h est $\mathbb{C}$-différentiable car on a prouvé que M est $\mathbb{C}$-différentiable et il est facile de prouver que $z \mapsto e^{-z^2/2}$ l'est aussi. Mais le calcul précédent montre que la restriction de h à la droite réelle est identiquement nulle : pour tout $t \in \mathbb{R}$,

$$h(t) = M(t)e^{-t^2/2} - 1 = e^{t^2/2}e^{-t^2/2} - 1 = 0.$$

Une fonction $\mathbb{C}$-différentiable qui s'annule sur la droite réelle s'annule sur tout le plan complexe (si une fonction $\mathbb{C}$-différentiable est non nulle, ses zéros sont isolés). Donc pour tout $z \in \mathbb{C}$,

$$M(z) = e^{z^2/2}.$$

En particulier, si $z = it$ où $t \in \mathbb{R}$, on obtient :

$$\Phi_X(t) = \mathbb{E}[e^{itX}] = M(it) = e^{(it)^2/2} = e^{-t^2/2}.$$

$\square$

14. Preuve du théorème central limite

Exercice 14.1. *Prouver qu'il suffit de prouver le théorème central limite dans le cas particulier où les v.a. que l'on somme sont centrées réduites.*

Soit $X_1, \cdots, X_n$ une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ qui sont mutuellement indépendantes et qui suivent toutes la même loi. On suppose que les variables sont centrées réduites. On a donc pour tout entier $0 \leq k \leq n$,

$$\mathbb{E}[X_k] = 0,\ \mathbb{E}[X_k^2] = 1.$$

Comme les variables suivent toutes la même loi, elle ont toutes la même fonction caractéristique Φ. Comme les variables X_k admettent un moment d'ordre 2, la formule de Taylor à l'ordre 2 montre que

$$\Phi(t) = \sum_{k=0}^{2} \frac{i^k \mathbb{E}[X^k]}{k!} t^k + o(t^2) = 1 - \frac{t^2}{2} + o(t^2).$$

L'espérance de

$$S_n = X_1 + \cdots + X_n$$

est nulle et sa variance est n donc

$$S_n^* = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$$

est centrée réduite. Comme les variables $X_1, \cdots, X_n$ sont mutuellement indépendantes,

$$\Phi_{S_n^*}(t) = \Phi\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n.$$

Donc (exercice)

$$\lim_{n\to\infty} \Phi_{S_n^*}(t) = \lim_{n\to\infty} \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n = e^{-\frac{t^2}{2}}.$$

On conclut la preuve du théorème en appliquant le résultat suivant :

Théorème 14.2. *(Continuité des distributions.) Soit $F_1, \ldots, F_n, \ldots$ des fonctions de distributions de fonctions caractéristiques $\Phi_1, \ldots, \Phi_n, \ldots$. Si*

$$\Phi(t) = \lim_{t\to\infty} \Phi_n(t)$$

existe et est continue en 0 alors Φ est la fonction caractéristique d'un distribution F et pour tout $x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_n(x) = F(x).$$

15. Matrices stochastiques et mesures d'équilibre

Définition 15.1. *(Matrice stochastique.) Soit E un ensemble dénombrable. A chaque couple $(x, y) \in E \times E$, on associe un nombre $0 \leq P(x, y) \leq 1$. On dit que*

$$P = (P(x, y))_{(x,y)\in E\times E}$$

est une matrice stochastique si pour tout $x \in E$,

$$\sum_{y\in E} P(x, y) = 1.$$

Exercice 15.2. *(Produit de matrices stochastiques.) Soit P et Q des matrices stochastiques sur un ensemble dénombrable E. On définit le produit de P avec Q*

$$PQ = (PQ(x, z))_{(x,z)\in E\times E}$$

par la formule

$$PQ(x, z) = \sum_{y\in E} P(x, y)Q(y, z).$$

Prouver que PQ est une matrice stochastique.

Définition 15.3. *(Mesure sur un ensemble dénombrable.) Une mesure m sur un ensemble dénombrable E est une application*

$$m : \mathcal{P}(E) \to [0, \infty]$$

telle que

$$m(\emptyset) = 0$$

et telle que pour toute famille dénombrable $A_1, \ldots, A_n, \ldots$ de sous-ensembles disjoints de E,

$$m\left(\bigcup_n A_n\right) = \sum_n m(A_n).$$

N.B. : la mesure m peut prendre la valeur $+\infty$. Si $x \in E$, on se permet d'écrire $m(x)$ pour désigner $m(\{x\})$.

Définition 15.4. *(Mesure d'équilibre local.) On dit que m est P-*réversible *(on dit aussi que m est une* mesure d'équilibre local pour *P), si et seulement si par définition*

$$m(x)P(x, y) = m(y)P(y, x), \forall x, y \in E.$$

Définition 15.5. *(Mesure d'équilibre global.) On dit que m est P-*invariante *i.e. $mP = m$ (on dit aussi que m est* une mesure d'équilibre global*, ou simplement une mesure d'équilibre, pour P), si et seulement si par définition*

$$\sum_x m(x)P(x, y) = m(y), \forall y \in E.$$

Proposition 15.6. *(Equilibre local en tout point implique équilibre global.) Si m est P-réversible alors elle est P-invariante (en d'autres termes, l'équilibre local en tout point implique l'équilibre global).*

Démonstration.

$$\begin{aligned}
\sum_x m(x)P(x, y) &= \sum_x m(y)P(y, x) \\
&= m(y) \sum_x P(x, y) \\
&= m(y).
\end{aligned}$$

□

Exemple 15.7. *(L'équilibre global n'implique pas l'équilibre local.) Donnons un exemple d'une mesure m et d'une matrice stochastique P, telles que m est* *une mesure d'équilibre global pour P, mais n'est pas une mesure d'équilibre local pour P. On choisit $E = \{x, y, z\}$, m homogène,*

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

La matrice est bien stochastique puisque la somme de chaque ligne est 1. *Tout point absorbe autant qu'il résorbe : cela se voit sur le graphe pondéré orienté associé à P qui est formé d'un cycle orienté de longueur* 3 *dont les trois sommets ont une boucle et dont le poids de chaque arête orientée vaut* $1/2$. *Il y a donc un équilibre global. C'est exactement la condition $mP = m$. Par contre on a $P(x, y) = 1/2$ mais $P(y, x) = 0$ donc il ne peut y avoir réversibilité.*

Définition 15.8. *(Matrice stochastique irréductible.) Soit E un ensemble dénombrable ou fini, soit P une matrice stochastique sur E. On dit que P est irréductible si pour tout couple de points x, y de E, il existe un entier n tel que*

$$P^n(x, y) > 0.$$

Proposition 15.9. *(Une mesure P-invariante non-nulle ne s'annule nulle part si P est irréductible.) Soit E un ensemble dénombrable ou fini, soit P une matrice stochastique sur E. Supposons P irréductible. Supposons que m est une mesure invariante qui n'est pas identiquement nulle. Alors*

$$\forall y \in E,\ m(y) > 0.$$

Démonstration. Par hypothèse il existe $x \in E$ tel que $m(x) > 0$. Comme P est irréductible, il existe n tel que $P^n(x, y) > 0$. Comme $mP^n = m$, on obtient :

$$m(y) = \sum_z m(z)P^n(z, y) \geq m(x)P^n(x, y) > 0.$$

□

Proposition 15.10. *Soit E un ensemble dénombrable ou fini, soit P une matrice stochastique sur E. Supposons que m est une mesure invariante qui ne s'annule en aucun point. La matrice $\hat{P}$ sur E définie par les conditions*

$$m(y)\hat{P}(y, x) = m(x)P(x, y),\ \forall x, y \in E,$$

est stochastique. De plus, $m\hat{P} = m$.

Démonstration. Comme $mP = m$,

$$\sum_x \hat{P}(y, x) = \frac{1}{m(y)} \sum_x m(x)P(x, y) = \frac{1}{m(y)} m(y) = 1.$$

Afin de prouver que $m\hat{P} = m$, on remarque que

$$(m\hat{P})(x) = \sum_y m(y)\hat{P}(y,x) = \sum_y m(x)P(x,y) = m(x).$$

□

Exercice 15.11. *(Un système non-réversible.) Soit E un ensemble dénombrable ou fini, soit P une matrice stochastique sur E. Supposons P irréductible. Supposons que m est une mesure invariante qui n'est pas identiquement nulle.*

(1) Prouver que m est P-réversible si et seulement si $P = \hat{P}$. Indication : c'est évident.

(2) Prouver que $\hat{P}$ est irréductible.

(3) Soit $E = \{1; 2; 3\}$. On considère la matrice stochastique

$$P = \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 0 & 2/3 \\ 2/3 & 1/3 & 0 \end{pmatrix}.$$

Dessiner le graphe de P en orientant chaque arête et en indiquant son poids. Déterminer l'unique mesure de probabilité m sur E qui est P-invariante. Est-ce que m est P-réversible ?

16. Marche aléatoire simple sur un graphe

Définition 16.1. *(Sommets voisins dans un graphe, degré d'un sommet.) Soit G un graphe, sans boucle ni arêtes multiples. Soit X l'ensemble de ses sommets et A l'ensemble de ses arêtes (non-orientées). On dit que $x, y \in X$ sont* voisins *et on écrit $x \sim y$ ou ce qui est équivalent $y \sim x$ si et seulement s'il existe une arête entre x et y. On note d_x le* degré *du sommet x. C'est par définition le nombre de voisins de x. C'est aussi le nombre d'arêtes attachées au sommet x. On suppose que pour tout sommet x, on a $0 < d_x < \infty$.*

Définition 16.2. *(Marche aléatoire simple sur un graphe.) Soit G un graphe, sans boucle ni arêtes multiples. Soit X l'ensemble de ses sommets. On suppose que pour tout $x \in X$, le degré d_x est fini. La* matrice de transition *de G est la matrice sur X ainsi définie : $P(x,y) = 0$ si x et y ne sont pas voisins et*

$$P(x,y) = 1/d_x,$$

si y est voisin de x.

Proposition 16.3. *(La marche aléatoire simple admet une mesure réversible.) Soit G un graphe, sans boucle ni arêtes multiples. Soit X l'ensemble de ses sommets. Soit A l'ensemble de ses arêtes. On suppose que pour tout $x \in X$, le degré d_x est fini.*

(1) *La matrice de transition P de G est stochastique.*
(2) *Soit m la mesure sur X défini par les conditions*
$$m(x) = d_x, \forall x \in X.$$
Alors m est P-réversible et donc m est P-invariante.
(3) *Si X est fini alors les conditions $\forall x \in X$,*
$$\pi(x) = \frac{d_x}{2|A|},$$
définissent une mesure de probabilité π sur X qui est P-invariante.

Démonstration. Prouvons que P est stochastique :
$$\sum_y P(x,y) = \sum_{y \sim x} 1/d_x = 1.$$
Prouvons que la mesure m sur X définie par les conditions $m(x) = d_x$ est une mesure réversible sur X. Si x et y ne sont pas voisins $m(x)P(x,y) = 0 = m(y)P(y,x)$. Si x et y sont voisins,
$$m(x)P(x,y) = d_x/d_x = 1 = d_y/d_y = m(y)P(y,x).$$
La mesure m étant P-réversible (en d'autres termes, c'est une mesure d'équilibre local), elle est P-invariante (en d'autres termes c'est une mesure d'équilibre global). Si X est fini, on peut considérer la masse totale $\sum_x m(x)$ de m. Elle est égale à deux fois le nombre d'arêtes de G (lemme des poignées de mains - d'application interdite en ce moment). Donc
$$\pi(x) = \frac{d_x}{2|A|} = \frac{d_x}{\sum_y d_y}$$
définit une mesure de probabilité sur X qui est invariante. □

17. Chaînes de Markov homogènes

Définition 17.1. *(Chaîne de Markov.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soient $X_0, \ldots, X_n, \ldots$ des variables aléatoires sur Ω à valeurs dans un ensemble dénombrable E. On dit que les variables $(X_n)_{n\geq 0}$ forment une* chaîne de Markov *si pour toute collection finie $x_0, \ldots, x_{n+1}$ de points de E,*
$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0).$$

Terminologie. L'ensemble E est appelé l'ensemble des *états*. Les variables aléatoires $X_0, \ldots, X_n, \ldots$ forment un *processus stochastique à temps discret.* Une interprétation possible de cette définition est la suivante : l'état au temps $n+1$ du processus stochastique décrit par la chaîne de Markov $X_1, \ldots, X_n, \ldots$

dépend de son état au temps n mais la connaissance de ses états aux temps antérieurs à n n'apportent pas d'information supplémentaire quant à la probable évolution du système au temps $n+1$.

Définition 17.2. *(Chaîne homogène.) Une chaîne de Markov est dite* homogène *si pour tout $x, y \in E$, pour tout entier $n \geq 0$,*

$$\mathbb{P}(X_{n+1} = y | X_n = x) = \mathbb{P}(X_1 = y | X_0 = x).$$

On peut comprendre cette définition ainsi : la chaîne est homogène dans le temps, c'est-à-dire que si le processus se trouve dans un état x au temps n, son évolution probable au temps $n+1$ vers l'état y est équiprobable à son évolution probable au temps 1 vers l'état y, sachant que son état initial au temps 0 est l'état x. On ne considère dans ce cours que des chaînes homogènes.

Exercice 17.3. *(Marche aléatoire comme processus de Markov.) Soient*

$$X_0, \dots, X_n, \dots$$

des variables aléatoires indépendantes à valeurs dans l'ensemble des entiers $\mathbb{Z}$. On suppose que X_0 prend la valeur 0 avec probabilité 1. On suppose que les variables $X_1, \dots, X_n, \dots$ suivent toutes une même loi de Bernoulli de paramètres $p, q \geq 0$, $p + q = 1$ et sont à valeurs dans $\{-1; 1\}$: pour tout $n \geq 1$,

$$\mathbb{P}(X_n = 1) = p, \, \mathbb{P}(X_n = -1) = q.$$

Soit

$$S_n = \sum_{k=0}^{n} X_k.$$

Les variables $S_0, \dots, S_n, \dots$ forment une chaîne de Markov.

Définition 17.4. *(La matrice stochastique d'une chaîne de Markov.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_{n \geq 0}$ une chaîne de Markov homogène sur Ω à valeurs dans E. La matrice stochastique associée à $(X_n)_{n \geq 0}$ est définie par les égalités*

$$P(x, y) = \mathbb{P}(X_{n+1} = y | X_n = x), \, \forall x, y \in E.$$

N.B. Comme la chaîne est homogène, ces nombres ne dépendent pas du choix de l'entier $n \geq 0$.

Exercice 17.5. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_{n \geq 0}$ une chaîne de Markov homogène sur Ω à valeurs dans E. Prouver que pour tout entier $n \geq 0$,*

$$\mathbb{P}(X_n = y | X_0 = x) = P^n(x, y), \, \forall x, y \in E.$$

Suggestion : récurrence sur n en sommant sur tous les états possiblement visités avant le dernier pas.

18. Temps de premier retour, nombre moyen de visites

Définition 18.1. *(Temps de premier retour.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $(X_n)_{n\geq 0}$ *une chaîne de Markov sur* Ω *à valeurs dans* E*. Soit* $x \in E$*. La variable aléatoire*

$$T_x : \Omega \to \mathbb{N} \cup \{\infty\}$$

définie par la formule

$$T_x(\omega) = \inf\{n \in \mathbb{N} : X_n(\omega) = x\}$$

est le temps de premier retour *à l'état* x *de la chaîne* $(X_n)_{n\geq 0}$.

N.B. Dans la définition de T_x on considère la suite des évènements

$$X_1 = x, \dots, X_n = x, \dots$$

en commençant au temps 1 et non pas au temps zéro. Il est possible que l'ensemble $\{n \in \mathbb{N} : X_n(\omega) = x\}$ soit vide, dans ce cas $T_x = \infty$.

Définition 18.2. *(Etat récurrent.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $(X_n)_{n\geq 0}$ *une chaîne de Markov sur* Ω *à valeurs dans* E*. Un état* $x \in E$ *est dit* récurrent *pour la chaîne* $(X_n)_{n\geq 0}$ *si*

$$\mathbb{P}(T_x < \infty | X_0 = x) = 1.$$

Un état $x \in E$ est donc récurrent pour la chaîne si et seulement si la chaîne, initialisée dans un état x au temps 0, retrouve l'état x en un temps fini, ceci avec probabilité 1.

Définition 18.3. *(Nombre moyen de visites.) Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* $(X_n)_{n\geq 0}$ *une chaîne de Markov sur* Ω *à valeurs dans* E *dénombrable. Soit* $x \in E$*. Soit* $y \in E$*. On considère la variable aléatoire*

$$N_{x,y} : \Omega \to \{0; 1; \cdots; n; \dots\} \cup \{\infty\}$$

$$\omega \mapsto \sum_{n=0}^{T_x(\omega)-1} 1_{(X_n=y)}(\omega)$$

ainsi que son espérance, conditionnée par l'évènement $(X_0 = x)$ *:*

$$\mathbb{E}_x[N_{x,y}] = \sum_{n\in\mathbb{N}} \mathbb{P}(N_{x,y} = n | X_0 = x) n + \mathbb{P}(N_{x,y} = \infty | X_0 = x) \infty.$$

On note

$$\gamma_x : E \to [0, \infty]$$

$$y \mapsto \gamma_x(y) = \mathbb{E}_x[N_{x,y}]$$

la fonction qui associe à un état y son nombre moyen de visites par la chaîne (durant une excursion de temps $T_x - 1$, partant de x au temps zéro et de retour en x au temps T_x).

Théorème 18.4. *(Le nombre moyen de visites définit une mesure invariante.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov sur Ω à valeurs dans E dénombrable. On suppose que la chaîne est irréductible et qu'il existe un élément de E qui est récurrent. Alors tout point de E est récurrent. De plus, pour tout $x \in E$ fixé, la fonction*

$$\begin{aligned} \gamma_x : E &\to [0, \infty] \\ y &\mapsto \gamma_x(y) \end{aligned}$$

qui associe à tout état y son nombre moyen de visites de la chaîne

$$\gamma_x(y) = \mathbb{E}_x \left[\sum_{n=0}^{T_x - 1} 1_{(X_n = y)} \right]$$

vérifie les propriétés suivantes.

(1) $0 < \gamma_x(y) < \infty$,

(2) $\gamma_x(x) = 1$,

(3) *γ_x définit une mesure P-invariante sur E, c'est-à-dire que pour tout $z \in E$,*

$$\sum_{y \in E} \gamma_x(y) P(y, z) = \gamma_x(z),$$

ou, sous forme plus condensée,

$$\gamma_x P = \gamma_x.$$

Théorème 18.5. *(Unicité de la mesure invariante.) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov sur Ω à valeurs dans E dénombrable. On suppose que la chaîne est irréductible et récurrente. Alors, à une constante multiplicative près, il existe sur E une unique mesure strictement positive et P-invariante.*

Corollaire 18.6. *(Espérance du temps de premier retour d'une marche aléatoire simple.) Soit G un graphe fini, sans boucle ni arêtes multiples. Soit X l'ensemble de ses sommets. Soit A l'ensemble de ses arêtes. On suppose que pour tout $x \in X$, le degré d_x est fini et que $d_x > 0$. Si G est connexe alors pour tout sommet $x \in X$, l'espérance du temps de premier retour en x de la marche aléatoire simple sur G partant de x au temps 0 est donnée par la formule :*

$$\mathbb{E}_x[T_x] = \frac{2|A|}{d_x}.$$

Démonstration. Afin de prouver le corollaire on considère la chaîne de Markov associée à la marche aléatoire simple sur G. Plus précisément, on peut construire un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ sur lequel sont définies des variables aléatoires $X_0, \ldots, X_n, \ldots$ à valeurs dans X qui forment une chaîne de Markov homogène et dont la matrice stochastique est égale à la matrice de transition de G et dont la distribution initiale est uniforme. On a donc :

(1) $\forall x \in X$, $\mathbb{P}(X_0 = x) = \frac{1}{|X|}$ (distribution initiale uniforme),
(2) $P(x,y) = \mathbb{P}(X_{n+1} = y | X_n = x)$, $\forall x, y \in E$ (matrice stochastique associée à la chaîne égale à la matrice P de transition de G.)

Comme G est connexe, la chaîne est irréductible. Comme X est fini tout état est récurrent. D'après le théorème sur le nombre moyen de visites, pour tout $x \in X$, on sait que γ_x défini une mesure invariante pour P strictement positive. On a :

$$\begin{aligned} 0 < \sum_{y \in X} \gamma_x(y) &= \sum_{y \in X} \mathbb{E}_x \left[\sum_{n=0}^{T_x - 1} 1_{(X_n = y)} \right] \\ &= \mathbb{E}_x \left[\sum_{n=0}^{T_x - 1} \sum_{y \in X} 1_{(X_n = y)} \right] \\ &= \mathbb{E}_x \left[\sum_{n=0}^{T_x - 1} 1_{\Omega} \right] \\ &= \mathbb{E}_x \left[T_x \right]. \end{aligned}$$

Ainsi, pour tout $x \in X$,

$$\frac{\gamma_x}{\mathbb{E}_x \left[T_x \right]}$$

défini sur X une mesure de probabilité qui est P-invariante. Mais on sait d'autre part que la fonction

$$\begin{aligned} X &\to [0,1] \\ x &\mapsto \frac{d_x}{2|A|} \end{aligned}$$

défini aussi une mesure de probabilité P-invariante sur X. Par le théorème d'unicité des mesures invariantes,

$$\frac{\gamma_x(x)}{\mathbb{E}_x \left[T_x \right]} = \frac{d_x}{2|A|}.$$

Comme $\gamma_x(x) = 1$, on obtient

$$\mathbb{E}_x \left[T_x \right] = \frac{2|A|}{d_x}.$$

□

Exercice 18.7. *On place un cavalier dans le coin d'un échiquier vide. On déplace le cavalier en choisissant chaque coup de manière équiprobable parmi l'ensemble des déplacements permis. Combien de coups faut-il s'attendre à devoir jouer pour retrouver le cavalier sur sa case de départ ? Même question en partant d'une case proche du centre de l'échiquier. Suggestions : considérer le graphe dont les sommets sont les 64 cases de l'échiquier, lier deux cases par une arête si et seulement si un cavalier peut passer en un coup d'une case à l'autre. Vérifier que le graphe obtenu est connexe. Si x est la case de départ, calculer $\mathbb{E}_x[T_x]$ à l'aide du corollaire intitulé « Espérance du temps de premier retour d'une marche aléatoire simple ». Tester les résultats obtenus par à l'aide d'un ordinateur.*

19. Appendice

Définition 19.1. *(Application $\mathbb{C}$-différentiable, application holomorphe.) Soit $U \subset \mathbb{C}$ un ouvert du plan complexe. Soit $f : \mathbb{C} \to \mathbb{C}$ une fonction d'une variable complexe à valeurs complexes. Soit $a \in U$. On dit que f est* $\mathbb{C}$-différentiable *au point a si la limite*

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

existe. N.B. : si la norme du nombre complexe $h \in \mathbb{C} \setminus \{0\}$ est suffisamment petite alors $(a+h) \in U$. Le quotient $\frac{f(a+h)-f(a)}{h}$ est alors un nombre complexe bien défini. Si la limite de ces quotients existe on la note $f'(a)$. Cette condition d'existence de la limite est forte : la limite ne dépend pas de la manière dont h tend vers 0. Si f est $\mathbb{C}$-différentiable en tout point de l'ouvert U on dit que f est holomorphe *sur U.*

Exemple 19.2. *(Les fonction z^n sont holomorphes.) L'application*

$$\mathbb{C} \to \mathbb{C}$$

$$z \mapsto z^2$$

est holomorphe sur $\mathbb{C}$ tout entier : pour tout $a \in \mathbb{C}$ il est facile de prouver que

$$\lim_{h \to 0} \frac{(a+h)^2 - a^2}{h} = 2a.$$

De même, pour tout entier $n \geq 1$, la fonction $f(z) = z^n$ est holomorphe sur $\mathbb{C}$. Si $a \in \mathbb{C}$,

$$f'(a) = na^{n-1}.$$

Exemple 19.3. *(L'exponentielle complexe est holomorphe.) L'exponentielle complexe est la fonction définie par la série*

$$\begin{aligned} \mathbb{C} &\to \mathbb{C} \\ z &\mapsto \sum_{n=0}^{\infty} \frac{z^n}{n!}. \end{aligned}$$

On vérifie que cette série converge absolument en tout point et l'on pose

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

On vérifie que pour tout $a \in \mathbb{C}$,

$$\lim_{h \to 0} \frac{e^{a+h} - e^a}{h} = e^a.$$

Exemple 19.4. *(La conjugaison complexe n'est pas holomorphe.) La conjugaison complexe*

$$\begin{aligned} \mathbb{C} &\to \mathbb{C} \\ z = x + iy &\mapsto \overline{z} = x - iy \end{aligned}$$

n'est nulle part $\mathbb{C}$-différentiable. En effet si $a \in \mathbb{C}$,

$$\lim_{h \to 0} \frac{\overline{a+h} - \overline{a}}{h} = \lim_{h \to 0} \frac{\overline{h}}{h}.$$

En posant $h = te^{i\theta}$ avec $t \in \mathbb{R} \setminus \{0\}$ et $0 \leq \theta < \pi$ on voit que cette limite peut prendre toutes les valeurs du cercle unité dans $\mathbb{C}$.

Pour la preuve du théorème suivant voir par exemple le paragraphe 1.2. The Taylor Series dans le livre "Complex Analysis" de L. Alhfors.

Théorème 19.5. *(Série de Taylor pour une fonction holomorphe.) Soit $a \in U \subset \mathbb{C}$ un point dans un ouvert du plan complexe. Soit $f : U \to \mathbb{C}$ une fonction holomorphe. Soit $n \geq 1$ un entier. Le développement de f en série de Taylor*

$$f(z) = \sum_{n=0}^{\infty} \frac{f^n(a)}{n!}(z-a)^n$$

est valide dans tout disque de centre a inclus dans U.

Corollaire 19.6. *(Les zéros d'une fonction holomorphe non-nulle sont isolés.) Soit $U \subset \mathbb{C}$ un ouvert connexe du plan complexe. Soit $f : U \to \mathbb{C}$ une fonction holomorphe. Supposons que f n'est pas identiquement nulle. Soit $a \in U$ tel que $f(a) = 0$. Il existe $\epsilon > 0$ tel que si $|z - a| < \epsilon$ et si $f(z) = 0$ alors $z = a$.*