Probabilités Élémentaires
2M231

Nicolas Broutin[1]
Sorbonne Université

Année 2018–2019
(version du 25 mars 2019)

1. très largement basé sur les notes de cours de LM231 rédigées par Raphael Krikorian en 2013–2014

Table des matières

1 Rappels de théorie des ensembles **5**
1.1 Opérations sur les ensembles . . . 5
1.2 Applications entre ensembles . . . 6
1.3 Dénombrement . . . 8
1.4 Dénombrabilité . . . 11

2 Espaces probabilisés **13**
2.1 Espaces probabilisés . . . 13
2.2 Tribus . . . 13
2.3 Probabilités . . . 14
2.3.1 Définition . . . 14
2.3.2 Exemples généraux . . . 14
2.3.3 Premières propriétés . . . 15
2.4 Probabilités sur un ensemble au plus dénombrable . . . 17
2.4.1 Construction des probabilités . . . 17
2.4.2 Probabilités uniformes et lien avec la combinatoire . . . 18
2.5 Evénements indépendants . . . 19
2.6 Jeu de pile ou face : le modèle de Bernoulli . . . 21
2.6.1 Jeu de n pile ou face : nombre fini de pièces . . . 21
2.6.2 Jeu infini de pile ou face : nombre infini de pièces . . . 23
2.7 Probabilités conditionnelles . . . 25

3 Variables aléatoires réelles **27**
3.1 Variables aléatoires réelles . . . 27
3.1.1 Définition et premières propriétés . . . 27
3.1.2 Loi d'une variable aléatoire réelle . . . 30
3.2 Variables aléatoires discrètes . . . 31
3.2.1 Loi d'une v.a. discrète et exemples . . . 31
3.2.2 Espérance d'une v.a. discrète . . . 33
3.2.3 Formule de transfert . . . 35
3.3 Espérance d'une variable aléatoire : cas général . . . 37
3.3.1 Espérance . . . 37
3.3.2 Variables aléatoires de carré intégrable . . . 38
3.3.3 Variance . . . 39
3.3.4 Inégalité de Bienaymé–Tchebychev . . . 40
3.4 Fonctions génératrices . . . 41
3.5 Variables aléatoires à densité . . . 43
3.5.1 Définition et exemples . . . 43
3.5.2 Espérance des v.a.r. admettant une densité . . . 44
3.5.3 Formule de transfert . . . 45
3.5.4 Variance des v.a. à densité . . . 45

3.5.5 Calcul de la loi d'une variable à densité . . . 46
3.6 Vecteurs aléatoires . . . 49
3.6.1 Loi d'un vecteur aléatoire . . . 49
3.6.2 Formules de transfert . . . 50
3.6.3 Loi de d'une somme de v.a.r. . . . 50
3.7 Variables aléatoires indépendantes . . . 51
3.7.1 Définition . . . 51
3.7.2 Cas des v.a. discrètes . . . 52
3.7.3 Cas des v.a admettant des densités . . . 53
3.7.4 Espérance des produits de v.a indépendantes . . . 53
3.7.5 Loi d'une somme de v.a.r. indépendantes . . . 54
3.7.6 Variance d'une somme de v.a.r. indépendantes . . . 56
3.7.7 Critères d'indépendance . . . 57

4 Convergence de suites de v.a. 59
4.1 Notions de convergences de variables aléatoires . . . 59
4.2 Lois des grands nombres . . . 61
4.2.1 Loi forte des grands nombres . . . 61
4.2.2 Loi faible des grands nombres . . . 61
4.2.3 Preuve de la loi forte des grands nombres – le cas $\mathcal{L}^4$ . . . 62
4.3 Théorème de la limite centrale . . . 64
4.3.1 Convergence en loi . . . 64
4.3.2 Fonctions caractéristiques . . . 65
4.3.3 Fonctions caractéristiques et convergence en loi . . . 67
4.3.4 Démonstration du théorème de la limite centrale . . . 69

Chapitre 1

Rappels de théorie des ensembles

La théorie mathématique des probabilités repose la théorie des ensembles, les dénombrements ou la combinatoire, et de manière cruciale sur la notion de dénombrabilité. Nous faisons ici les rappels élémentaires essentiels.

1.1 Opérations sur les ensembles

Ensemble, ensemble fini/infini, cardinal Un ensemble est intuitivement une collection d'éléments. Etant donnés un ensemble E et un élément a on écrit $a \in E$ si a est un élément de E. Il existe un unique ensemble ne contenant aucun élément ; on le note $\emptyset$. On dit qu'un ensemble est *fini* s'il ne contient qu'un nombre fini d'éléments et *infini* sinon. si A est un ensemble fini on appelle *cardinal* de A le nombre d'éléments de A et on note ce nombre entier $\#A$ ou card A. Si A est infini, on pose $\#A = \infty$.

Quantificateurs Pour décrire les ensembles, on a souvent besoin de quantificateurs "pour tout" $\forall$ et "il existe" $\exists$ qui permettent une écriture concise à laquelle il faut s'habituer.

Inclusion, complémentaire Si E et A sont deux ensembles on dit que F est inclus dans E ou que A est un sous-ensemble de E si tout élément de A est un élément de E et on écrit $A \subset E$. On peut alors définir le *complémentaire* de A dans E qui est l'ensemble des éléments de E qui n'appartiennent pas à A. On le notera dans ce cours $E - A$, ou $E \setminus A$, ou encore A^c ; cette dernière notation cesse d'être ambigüe si l'on suppose E fixé une fois pour toute, ce que nous ferons.

Ensemble des parties d'un ensemble Si E est un ensemble, l'ensemble constitué des sous-ensembles de E s'appelle l'ensemble des parties de E et se note $\mathcal{P}(E)$.

Union, intersection Si $(A_i)_{i\in I}$ est une collection d'ensembles inclus dans E, la réunion des A_i est l'ensemble $\bigcup_{i\in I} A_i$ des $a \in E$ pour lesquels *il existe* $i \in I$ tel que $a \in A_i$. De même l'intersection des A_i est l'ensemble $\bigcap_{i\in I} A_i$ des $a \in E$ pour lesquels $a \in A_i$ *pour tout* $i \in I$.

$$\bigcup_{i\in I} A_i = \{a \in E : \exists i \in I, a \in A_i\} \quad \text{et} \quad \bigcap_{i\in I} A_i = \{a \in E : \forall i \in I, a \in A_i\}.$$

On dit que deux ensembles sont *disjoints* si leur intersection est vide. On dit que les ensembles A_i, $i \in I$ constituent une *partition* de l'enemble E si i) ils sont non vides, ii) leur union sur $i \in I$ vaut E iii) ils sont disjoints deux à deux ($A_i \cap A_j = \emptyset$ si $i \neq j$) ; on dit aussi que E est union disjointe des A_i, $i \in I$.

On a les formules

$$\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c, \qquad \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c.$$

et

$$B \cap \left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} (B \cap A_i).$$

Produits d'ensembles Si $A_1, \ldots, A_n$ sont des ensembles on peut définir le produit cartésien de ces ensembles comme étant l'ensemble des n-uplets $(a_1, \ldots, a_n)$ où $a_1 \in A_1, \ldots, a_n \in A_n$. On note cet ensemble $A_1 \times \cdots \times A_n$. Quand les A_i sont finis son cardinal est le produit des cardinaux des A_i.

Exercice 1.1.1. *Soit $E = \mathbb{R}^{\mathbb{N}}$ l'ensemble des suites à valeurs dans $\mathbb{R}$.*
(i) Décrire en mots l'ensemble $A = \cap_{n\in\mathbb{N}}\{u \in E : u_n \in \{0,1\}\}$.
(ii) Décrire de manière formelle l'ensemble B des suites bornées inférieurement.

Solution. (i) L'ensemble A est l'ensemble des suites u telles que $u_n \in \{0,1\}$ pour tout $n \in \mathbb{N}$, c'est-à-dire l'ensemble des suites à valeurs dans $\{0,1\}$.
(ii) Une suite est bornée inférieurement s'il existe un $M \in \mathbb{R}$ tel que $u_n \geq M$ pour tout $n \in \mathbb{N}$. Par conséquent, $B = \cup_{M\in\mathbb{R}} \cap_{n\in\mathbb{N}} \{u \in E : u_n \geq M\}$. △

1.2 Applications entre ensembles

Injectivité, surjectivité, bijectivité Si A et B sont deux ensembles, une application associe à tout élément a de A un unique élément noté $f(a)$ de B. On dit que $f(a)$ est l'image de a par f. Un élément de B peut n'être l'image d'aucun élément de A ou au contraire être l'image de plusieurs éléments de A. Si $a \in A$, $b \in B$ et $f(a) = b$ on dit que a est un antécédent de a par f. On dit qu'une application est *injective* si tout élément de B est l'image d'au plus un élément de A, *surjective* si tout élément de B est l'image d'au moins un élément de A et *bijective* si elle est injective et surjective.

S'il existe une injection de A dans B on a $\#A \leq \#B$.
S'il existe une surjection de A sur B on a $\#A \geq \#B$.
S'il existe une bijection entre A et B on a $\#A = \#B$.

Ensemble des applications de A dans B On note B^A l'ensemble des applications de A dans B. Quand A et B son finis son cardinal vaut $(\#B)^{\#A}$.

Fonctions indicatrices, codages Si E est un ensemble fixé, l'ensemble des parties de E est en bijection avec l'ensemble des applications de E dans l'ensemble à deux éléments $\{0,1\}$. Cette bijection est la suivante : à tout ensemble $A \subset E$ on associe sa *fonction caractéristique* ou *fonction indicatrice* $\mathbf{1}_A : E \to \{0,1\}$ définie par $\mathbf{1}_A(e) = 1$ si $e \in A$ et $\mathbf{1}_A(e) = 0$ sinon. Réciproquement si f est une application de E dans $\{0,1\}$ l'ensemble A des $e \in E$ tels que $f(e) = 1$ est tel que $\mathbf{1}_A(\cdot) = f(\cdot)$. En particulier, cela démontre que quand E est fini le cardinal de $\mathcal{P}(E)$ est $2^{\#E}$

$$\#\mathcal{P}(E) = 2^{\#E}.$$

Cardinal et fonctions indicatrices Si $A \subset E$ on a

$$\#A = \sum_{x\in E} \mathbf{1}_A(x).$$

On a aussi $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$. Si $A_1, \dots, A_n$ sont des sous-ensembles de E on a

$$\mathbf{1}_{A_1 \cap \dots \cap A_n} = \prod_{i=1}^{n} \mathbf{1}_{A_i}.$$

Pré-image Si f est une application de E dans F on définit pour tout $B \subset F$ l'ensemble $f^{-1}(B)$ comme étant l'ensemble des $e \in E$ tels que $f(e) \in B$. (Cette définition a un sens même si f n'est pas inversible.) On dit que $f^{-1}(B)$ est la pré-image de B par f ; c'est l'ensemble des antécédents des éléments de B par f.

On a toujours

$$f^{-1}\Big(\bigcup_{i \in I} A_i\Big) = \bigcup_{i \in I} f^{-1}(A_i), \quad f^{-1}\Big(\bigcap_{i \in I} A_i\Big) = \bigcap_{i \in I} f^{-1}(A_i), \quad f^{-1}(A^c) = \big(f^{-1}(A)\big)^c.$$

Attention le comportement par image directe n'est pas aussi bon (voir Exercice 1.7 des feuilles de TD).

Exercice 1.2.1. *i) Montrer que si A, B sont deux sous-ensembles de E on a*

$$1 - \mathbf{1}_{A \cup B} = (1 - \mathbf{1}_A)(1 - \mathbf{1}_B),$$

et en déduire que

$$\#(A \cup B) = \#A + \#B - \#(A \cap B).$$

ii) En généralisant la formule précédente montrer que

$$\#(A_1 \cup \dots \cup A_n) = \sum_{p=1}^{n} (-1)^{p-1} \sum_{1 \le i_1 < \dots < i_p \le n} \#(A_{i_1} \cap \dots \cap A_{i_p}).$$

Solution. i) Pour tout ensemble $F \subset E$

$$\mathbf{1}_{F^c} = 1 - \mathbf{1}_F.$$

Donc

$$\begin{aligned} 1 - \mathbf{1}_{A \cup B} &= \mathbf{1}_{A^c \cap B^c} \\ &= \mathbf{1}_{A^c} \mathbf{1}_{B^c} \\ &= (1 - \mathbf{1}_A)(1 - \mathbf{1}_B). \end{aligned}$$

On a donc

$$\begin{aligned} \mathbf{1}_{A \cup B} &= \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \cdot \mathbf{1}_B \\ &= \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A \cap B}. \end{aligned}$$

Or, pour tout ensemble $F \subset E$

$$\#F = \sum_{e \in E} \mathbf{1}_F(e).$$

On a donc bien la conclusion.

ii) De façon générale,

$$1 - \mathbf{1}_{A_1 \cup \dots \cup A_n} = \prod_{i=1}^{n} (1 - \mathbf{1}_{A_i}),$$

et donc (on peut démontrer la formule du développement du membre de droite par récurrence ou de manière combinatoire)

$$\begin{aligned} 1 - \mathbf{1}_{A_1 \cup \cdots \cup A_n} &= 1 + \sum_{p=1}^{n} (-1)^p \sum_{1 \le i_1 < \ldots < i_p \le n} \mathbf{1}_{A_{i_1}} \cdots \mathbf{1}_{A_{i_p}} \\ &= 1 + \sum_{p=1}^{n} (-1)^p \sum_{1 \le i_1 < \ldots < i_p \le n} \mathbf{1}_{A_{i_1} \cap \cdots \cap A_{i_p}}, \end{aligned}$$

et en sommant sur $e \in E$ on obtient bien la formule annoncée.

1.3 Dénombrement

Cardinal d'une union disjointe finie. Si $A_1, \ldots, A_n$ sont des ensembles finis *disjoints deux à deux* tels que $A_1 \cup \cdots \cup A_n = E$ alors E et fini et

$$\#E = \sum_{i=1}^{n} \#A_i.$$

Cardinal d'un produit. Si $A_1, \ldots, A_n$ sont des ensembles finis le cardinal du produit $A_1 \times \cdots \times A_n$ est donné par

$$\#(A_1 \times \cdots \times A_n) = (\#A_1) \cdots (\#A_n).$$

Cardinal de l'ensemble des applications de A dans B. Si A et B sont des ensembles finis, l'ensemble des applications de A dans B est fini et à pour cardinal

$$\#(B^A) = (\#B)^{\#A}.$$

Nombre d'injections entre deux ensembles finis. Nombre d'arrangements Si A et B sont deux ensembles finis avec $\#A = p$, $\#B = n$, l'ensemble des applications injectives de A vers B a un cardinal égal à

$$\begin{cases} 0 \quad \text{si} \quad \#A > \#B \\ n(n-1)\cdots(n-p+1) \quad \text{si} \quad p \le n. \end{cases}$$

En effet, supposons $A = \{a_1, \ldots, a_p\}$; si $p > n$, il ne peut y avoir d'applications injective de A vers B, tandis que si $p \le n$, il y a n choix possibles pour la valeur $f(a_1)$, $n-1$ choix possibles pour la valeur de $f(a_2)$ (comme f est injective $f(a_2)$ ne peut pas prendre la même valeur que $f(a_1)$), ..., et $n-(p-1) = n-p+1$ choix possibles pour $f(a_p)$. C'est aussi le nombre de p-uplets (ordonnés), on dit aussi d'**arrangements**, $(e_1, \ldots, e_p)$ d'éléments distincts de E.

Nombre de bijections de A vers A. Nombre de permutations, factorielles Si A est un ensemble de cardinal n, une application de A vers A est bijective si et seulement si elle est injective et par conséquent le nombre de bijection de A vers A (on dit aussi le nombre de **permutations** de A) égale

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

Par convention $0! = 1$.

Cardinal de $\mathcal{P}(E)$. Si E est fini de cardinal n, le nombre de sous-ensembles de E est égal au nombre d'applications de E vers $\{0, 1\}$ et vaut donc

$$\#\mathcal{P}(E) = 2^n.$$

Nombre de sous-ensembles de cardinal p d'un ensemble à n éléments. Nombre de combinsiasons Si E est un ensemble fini de cardinal n, le nombre de sous-ensembles de E de cardinal exactement p (on parle aussi de **combinaison** égale

$$\binom{n}{p} = C_n^p = \frac{n(n-1)\cdots(n-p+1)}{p!} = \frac{n!}{p!(n-p)!}.$$

On pose $\binom{n}{k} = 0$ quand $k > n$ ou quand $n < 0$ et $\binom{n}{0} = 1$. Remarquer que $\binom{n}{k} = \binom{n}{n-k}$.

En effet, un sous-ensemble $\{a_1, \ldots, a_p\}$ de E peut être vu comme un p-uplet d'élément de E où l'on oublie l'ordre des éléments. Or, étant donnés p éléments de E on peut former $p!$ (nombre de bijections de $\{a_1, \ldots, a_p\}$ dans lui même) p-uplets. Ainsi, le nombre de sous-ensembles de cardinal p d'un ensemble à n éléments égale le nombre d'injection de l'ensemble $\{1, \ldots, p\}$ dans E (i.e le nombre de p-uplets de E) divisé par $p!$.

Binôme de Newton Considérons l'expression $(a+b)^n = (a+b)\cdots(a+b)$ (où a et b sont des nombres réels ou complexes ou des éléments d'un anneau commutatif). Quand on développe le produit, on obtient une somme de produits de a et de b et on voit que le coefficient de $a^p b^{n-p}$ est égal au nombre de façons de choisir p éléments (les a) parmi n. On a donc

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

Triangle de Pascal En utilisant la formule

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

on obtient pour tous entiers $0 \leq k \leq n$ la relation

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

(**Exercice.** Le vérifier de deux manière : par le calcul, puis de façon combinatoire).

Obtenir de nouvelles formules sur les coefficients binomiaux Voici quelques méthodes utiles pour obtenir de nouvelles formules sur les coefficients binomiaux. On considère le polynôme $(1+X)^n$. D'après la formule du binôme

$$(1+X)^n = \sum_{k=0}^{n} \binom{n}{k} X^k. \tag{1.1}$$

En faisant $X = 1$ dans cette expression on voit que

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

(**Exercice.** Comment démontrer cette formule de façon purement combinatoire ?).

De même, si on dérive chacun des membres de la formule (1.1) on obtient

$$n(1+X)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k X^{k-1},$$

c'est-à-dire en faisant $X = 1$

$$n2^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k.$$

En dérivant plusieurs fois la formule (1.1) on obtiendrait de la même manière d'autres expressions impliquant les coefficients binomiaux. (**Exercice.** Le vérifier de manière combinatoire).

On peut aussi écrire $(1+X)^{n+m} = (1+X)^n(1+X)^m$ et utiliser la formule du binôme :

$$\sum_{r=0}^{n+m} \binom{n+m}{r} X^r = \sum_{k=0}^{n} \binom{n}{k} X^k \sum_{l=0}^{n} \binom{m}{l} X^l. \tag{1.2}$$

Chacun des membres de l'équation précédente est un polynôme de degré $n+m$. Le coefficient du monôme de degré r de ce polynôme est $\binom{n+m}{r}$ (quand on regarde le membre de gauche) et également $\sum_{k+l=r} \binom{n}{k}\binom{m}{l}$. On a donc

$$\binom{n+m}{r} = \sum_{k=0}^{r} \binom{n}{k}\binom{m}{r-k}.$$

Exercice 1.3.1. *Une urne contient N boules noires et M boules blanches, toutes distinguables.*
i) On effectue n tirages sans remise. Quel est le nombre total de tels tirages ? Combien de tirages donnent x ($x \leq n$) boules noires ?
ii) On effectue n tirages avec remise. Quel est le nombre total de tels tirages ? Combien de tirages donnent x ($x \leq n$) boules noires ?

Solution. On note $\{1, \ldots, N\}$ l'ensemble des boules noires et $\{N+1, \ldots, N+M\}$ l'ensemble des boules blanches.

i) Un tirage sans remise est équivalent à la donnée d'une injection de $\{1, \ldots, n\}$ dans $\{1, \ldots, N+M\}$ (ou à une suite ordonnée, un n-uplet $(x_1, \ldots, x_n)$, $x_i \in \{1, \ldots, N+M\}$). Il y a donc $(N+M)\cdots(N+M-n+1)$ tirages sans remise.

Un tirage où x boules noires sont tirées est équivalent à la donnée d'un sous-ensemble A de $\{1, \ldots, n\}$ à x élements (si on pense au tirage comme à une expérience, A est l'ensemble des temps où le résultat de notre expérience est "boule noire") et de deux injections, une de A dans l'ensemble des boules noires, une seconde du complémentaire de A dans $\{1, \ldots, n\}$ dans l'ensemble des boules blanches : on a donc

$$\binom{n}{x} \cdot N(N-1)\cdots(N-x+1) \cdot M(M-1)\cdots(M-(n-x)+1)$$

choix possibles.

Remarquons que la proportion du nombre de tirages sans remise où x boules noires sortent dans l'ensemble des tirages sans remise est

$$\begin{aligned}
&\frac{\binom{n}{x} \times N(N-1)\cdots(N-x+1) \times M(M-1)\cdots(M-n+x+1)}{(N+M)\cdots(N+M-n+1)} \\
&= \frac{\binom{n}{x} \times \binom{N}{x} x! \times \binom{M}{n-x}(n-x)!}{\binom{N+M}{n} n!} \\
&= \frac{\binom{N}{x}\binom{M}{n-x}}{\binom{N+M}{n}}.
\end{aligned}$$

ii) Un tirage avec remise est équivalent à la donnée d'une application (pas nécessairement injective) de $\{1, \ldots, n\}$ vers $\{1, \ldots, N+M\}$ (ou encore d'un n-uplet $(e_1, \ldots, e_n)$ de $\{1, \ldots, N+M\}^n$) ; il y a donc $(N+M)^n$ choix possibles.

Un tirage où x boules noires sont tirées est équivalent à la donnée : d'un sous-ensemble A de $\{1, \ldots, n\}$ à x élements, d'une application (pas nécessairement injective) de A dans $\{1, \ldots, N\}$

(ou encore d'un x-uplet de $\{1,\dots,N\}^x$) et d'une application de $\{1,\dots,n\}-A$ dans $\{N+1,\dots,N+M\}$ (ou encore un $(n-x)$-uplet de $\{N+1,\dots,N+M\}$). Il y a donc

$$\binom{n}{x}\cdot N^x\cdot M^{n-x}$$

choix possibles. Remarquons que la proportion du nombre de tirages avec remise où x boules noires sortent dans l'ensemble des tirages avec remise est

$$\binom{n}{x}\frac{N^xM^{n-x}}{(N+M)^n}=\binom{n}{x}p^x(1-p)^{n-x},$$

où $p=N/(N+M)$. △

1.4 Dénombrabilité

Définition 1.4.1. *Un ensemble est dit dénombrable s'il est en bijection avec un sous-ensemble de l'ensemble $\mathbb{N}$ des entiers naturels.*

De façon plus concrète, un ensemble est dénombrable si on peut énumérer ses éléments.

Proposition 1.4.1. *Si A et B sont deux ensembles.*

a) S'il existe une injection de A dans B et si B est dénombrable alors A est dénombrable

b) S'il existe une surjection de A dans B et si A est dénombrable, alors B est dénombrable.

Théorème 1.4.1. *a) Si $A_1,\dots,A_n$ sont des ensembles dénombrables, le produit $A_1\times\cdots\times A_n$ est également dénombrable.*

b) Si $(A_i)_{i\in I}$ est une famille dénombrable (c'est-à-dire I est dénombrable) d'ensembles dénombrables (pour tout $i\in I$, A_i est dénombrable) alors la réunion $\bigcup_{i\in I}A_i$ est également dénombrable.

Preuve. a) On peut supposer $A_1=\ldots=A_n=\mathbb{N}$. Notons $p_1,\dots,p_n$ les n premiers nombres premiers (p est premier s'il est divisble uniquement par 1 et par p) et considérons l'application qui à $(l_1,\dots,l_n)\in\mathbb{N}^n$ associe le nombre $2^{l_1}\cdot 3^{l_2}\cdots p_n^{l_n}$ est une injection de $\mathbb{N}^n$ dans $\mathbb{N}$ car la décomposition en facteurs premiers d'un nombre est unique. La proposition 1.4.1 a) permet de conclure.

b) Considérons l'application de $\mathbb{N}\times\mathbb{N}$ dans $\bigcup_{i\in I}A_i$ qui au couple (n,m) associe le m-ième élément de l'ensemble A_{i_n} où i_n est le n-ième élément de I. C'est une surjection. La proposition 1.4.1 b) donne la conclusion. □

Corollaire 1.4.1. *L'ensemble des entiers relatifs $\mathbb{Z}$ et l'ensemble des nombres rationnels $\mathbb{Q}$ sont dénombrables.*

Preuve. L'ensemble $\mathbb{Z}$ est dénombrable car l'application de l'ensemble dénombrable $\{1,-1\}\times\mathbb{N}$ dans $\mathbb{Z}$ qui au couple (ϵ,n) associe le produit ϵn est une surjection. De même, $\mathbb{Q}$ est dénombrable car l'application de l'ensemble dénombrable $\mathbb{Z}\times(\mathbb{N}-\{0\})$ dans $\mathbb{Q}$ qui au couple (p,q) associe le rationnel p/q est une surjection. □

On peut démontrer que

Théorème 1.4.2. *L'ensemble des nombres réels $\mathbb{R}$ n'est pas dénombrable.*

Une des manières de montrer le Théorème 1.4.2 consiste à d'abord prouver que l'ensemble $\{0,1\}^{\mathbb{N}}$ des suites à valeurs dans $\{0,1\}$ n'est pas dénombrable, puis de montrer que $[0,1]$ n'est pas dénombrable. Voir Exercices 1.18 et 1.19 des feuilles de TD.

Corollaire 1.4.2. *L'ensemble des nombres irrationnels n'est pas dénombrable.*

Preuve. Car sinon, $\mathbb{R}$ qui est réunion de $\mathbb{Q}$ et de l'ensemble des nombres irrationnels serait dénombrable (comme union dénombrable d'ensembles dénombrables). □

Chapitre 2

Espaces probabilisés

2.1 Espaces probabilisés

Un espace probabilisé est la donnée

— d'un espace Ω que l'on appelle *l'espace des états*. Quand on modélise une situation concrète Ω est l'ensemble des états du système que l'on considère. Bien souvent cet espace est inaccessible à l'expérience ;

— d'un sous-ensemble $\mathcal{F}$ de $\mathcal{P}(\Omega)$ qui est *l'ensemble des évévements*. Dans une situation concrète c'est l'ensemble de tous les résultats d'expériences que l'on peut effectuer sur le système. En théorie des probabilités (donc quand on fait des mathématiques) cet ensemble $\mathcal{F}$ sera une *tribu* (on dit aussi une σ-algèbre), *cf.* définition 2.2.1 ;

— d'une *probabilité* $\mathbf{P}$: pour tout évévement $A \in \mathcal{F}$ le réel $\mathbf{P}(A)$ est le degré de vraisemblance de l'évévement A ; c'est un nombre réel compris entre 0 et 1. Mathématiquement, une probabilité est une application $\mathbf{P} : \mathcal{F} \to [0, 1]$ vérifiant les propriétés décrites dans la définition 2.3.1.

Nous précisons dans la suite les deux derniers points, mais avant tout, voici un exemples pour fixer les idées :

Example. Si on considère la modélisation d'un lancé d'une pièce, on peut choisir l'espace des états $\Omega = \{P, F\}$ pour P(ile) ou F(ace). Nous verrons que l'ensemble des événement devrait être $\mathcal{B} = \mathcal{P}(\mathcal{E})$, et on devrait au moins avoir intuitivement $\mathbf{P}(\{P\}) = \mathbf{P}(\{F\}) = 1/2$. Toujours de manière informelle, $\emptyset$ (le résultat du dé est dans $\emptyset$) est l'événement "il n'arrive rien" ni pile ni face , ce qui est bien evidemment impossible et on devrait avoir $\mathbb{P}(\emptyset) = 0$; de manière analogue, Ω est l'événement "quelque chose arrive" (le résulat du dé est dans Ω) soit pile soit face ce qui est bien entendu certain et on devrait avoir $\mathbb{P}(\Omega) = 1$. Nous parlerons plus en détails de la probabilité et de ses propriétés (imposées !) plus loin.

2.2 Tribus

Soit Ω un ensemble fixé (l'espace des états).

Définition 2.2.1. *Une* tribu *ou encore une* σ-algèbre *de* Ω *est un ensemble de parties de* Ω *(donc un sous-ensemble de* $\mathcal{P}(\Omega)$*, l'ensemble des parties de* Ω*) qui contient l'ensemble vide, est stable par passage au complémentaire et est stable par union dénombrable :*

— $\emptyset \in \mathcal{F}$

— *pour tout* $A \in \mathcal{F}$ *on a* $A^c \in \mathcal{F}$

— *pour toute famille (dénombrable)* $(A_i)_{i\in\mathbb{N}}$ *d'éléments de* $\mathcal{F}$ *l'union*

$$\bigcup_{i\in\mathbb{N}} A_i$$

est également dans $\mathcal{F}$.

Il est clair que Ω est toujours élément de la tribu (c'est le complémentaire de l'ensemble vide) et qu'une intersection dénombrable d'éléments de la tribu est encore dans la tribu (car $\cap_{i\in\mathbb{N}} A_i = (\cup_{i\in\mathbb{N}} A_i^c)^c$).

Exemples
1) Si Ω est un ensemble quelconque on peut toujours définir deux tribus :
la tribu triviale qui est $\mathcal{F} = \{\emptyset, \Omega\}$
la tribu totale qui est $\mathcal{F} = \mathcal{P}(\Omega)$.
2) Si $\Omega = \{1, 2, 3\}$ le sous-ensemble de $\mathcal{P}(\Omega)$, $\mathcal{F} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ est une tribu de Ω.
3) **(Exercice :)** *Si Ω est un ensemble le sous-ensemble de $\mathcal{P}(\Omega)$ constitué des ensembles qui sont dénombrables ou dont le complémentaire est dénombrable est une tribu.*

Remarque. On peut se demander pourquoi on ne prends pas toujours $\mathcal{F} = \mathcal{P}(\Omega)$: en fait, si Ω n'est pas dénombrable, alors $\mathcal{P}(\Omega)$ n'est pas adapté parce qu'on ne peut alors pas toujours définir de probabilité $\mathbb{P}$ qui satisfait les propriétés naturelles. Voir la remarque suivant la définition de la probabilité.

2.3 Probabilités

2.3.1 Définition

Définition 2.3.1. *Si Ω est un ensemble et $\mathcal{F}$ est une tribu de Ω, une probabilité $\mathbf{P}$ est une application de $\mathcal{F}$ dans $[0, 1]$ telle $\mathbb{P}(\Omega) = 1$ et telle que pour toute famille dénombrable $(A_i)_{i\in\mathbb{N}}$ d'évévements de $\mathcal{F}$* **disjoints 2 à 2** *on a*

$$\mathbf{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i=0}^{\infty} \mathbf{P}(A_i). \tag{2.1}$$

où l'égalité précédente signifie la chose suivante : la probabilité $\mathbf{P}(\cup_{i\in\mathbb{N}} A_i)$ est égale à la limite de la suite croissante de nombres réels $\sum_{i=0}^{N} \mathbf{P}(A_i)$ quand N tend vers l'infini. (Cette limite existe toujours car la suite en question est croissante et bornée.)

Remarque **1)** A priori, rien de garantit qu'il existe effectivement des probabilités, et en général nous admettrons l'existence. Dans ce chapitre, nous montrons qu'on peut effectivement faire certaines des constructions de façon plus ou moins explicite en détaillant Ω. Cependant, il faut avoir à l'esprit qu'on veut à terme plutôt s'affranchir de Ω et reposer sur l'existence pour se focaliser sur le calcul des probabilités.
2) L'intérêt d'autoriser la stabilité par unions (intersections) dénombrables dans la définition d'une tribu permet de construire à partir d'évévements simples des évévements beaucoup plus intéressants que ceux qu'on obtiendrait en ne supposant que la stabilité par unions (intersections) finies. En revanche, si on autorisait la stabilité par unions (intersections) quelconques on ne pourrait pas construire beaucoup de probabilités. La stabilité par unions (intersections) dénombrable est donc le bon compromis.

2.3.2 Exemples généraux

Mesures de Dirac Sur tout ensemble Ω muni d'une tribu $\mathcal{F}$ il est possible de construire des mesures de la façon suivante : pour tout $\alpha \in \Omega$ définissons l'application $\delta_\alpha : \mathcal{F} \to [0, 1]$ qui à un ensemble $A \in \mathcal{F}$ associe le réel 1 si $\alpha \in A$ et 0 sinon. Cette application δ_α est une mesure de probabilité que l'on appelle la mesure de Dirac au point α. Vérifions rapidement que c'est

bien une mesure : déjà $\delta_\alpha(\Omega) = 1$ puisque $\alpha \in \Omega$; par ailleurs si $A_i \in \mathcal{F}$, $i \geq 0$ est une famille dénombrable d'ensembles de la tribu disjoints deux à deux on a

$$\delta_\alpha\left(\bigcup_{i\geq 0} A_i\right) = \sum_{i\geq 0} \delta_\alpha(A_i),$$

car :

– soit α appartient à $\cup_{i\geq 0} A_i$; mais alors il existe un $i \geq 0$ pour lequel $\alpha \in A_i$ et cet indice i est unique car les A_i sont disjoints deux à deux. L'égalité précédente se réduit à $1 = 1$;

– soit α n'appartient pas à $\cup_{i\geq 0} A_i$ et de ce fait n'appartient à aucun des A_i : l'égalité se réduit à $0 = 0$.

Sommes pondérées de probabilités Si $\mathbb{P}_1, \ldots, \mathbb{P}_n$ sont des probabilités sur $\mathcal{F}$ ($\mathcal{F}$ étant une tribu) et si $\lambda_1, \ldots, \lambda_n$ sont des réels positifs tels que $\lambda_1 + \cdots + \lambda_n = 1$ alors $\mathbb{P} := \lambda_1\mathbb{P}_1 + \cdots + \lambda_n\mathbb{P}_n$ est aussi une probabilité. On peut vérifier que ceci s'étant à une famille dénombrable de probabilités $(\mathbb{P}_n)_{n\in\mathbb{N}}$ si on a aussi $(p_n)_{n\in\mathbb{N}}$ avec $p_n \geq 0$ pour tout $n \in \mathbb{N}$ et $\sum_n p_n = 1$.

Ces deux exemples élémentaires permettent déjà de construire des probabilités assez complexes : en effet, quelque soit l'ensemble Ω et la tribu $\mathcal{F}$ sur Ω, $\sum_{i\geq 0} p_i\delta_{x_i}$ est une probabilité prouvu que $x_i \in \Omega$, $p_i \geq 0$ pour tout $i \in \mathbb{N}$ et $\sum_i p_i = 1$.

2.3.3 Premières propriétés

Mentionnons tout d'abord deux propriétés immédiates des probabilités :

Proposition 2.3.1. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.*
a) Si $A \in \mathcal{F}$,

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

b) $\mathbb{P}(\emptyset) = 0$
c)(Positivité) Si $A, B \in \mathcal{F}$ vérifient $A \subset B$ alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.
d) Si $A, B \in \mathcal{F}$ alors

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

Preuve. a) Il suffit décrire Ω comme l'union disjointe finie $\Omega = A \cup A^c$: comme $\mathbb{P}$ est une probabilité $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$.
b) suit de la formule précédente et du fait que $\mathbb{P}(\Omega) = 1$.
c) On écrit B comme l'union disjointe $B = A \cup (B \cap A^c)$ et $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$. Comme $\mathbb{P}(B \cap A^c) \geq 0$ on a bien $\mathbb{P}(B) \geq \mathbb{P}(A)$.
d) De l'union disjointe $A \cup B = A \cup (B \cap A^c)$ on déduit $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$. Mais de l'union disjointe $B = (B \cap A^c) \cup (B \cap A)$ on obtient $\mathbb{P}(B) = \mathbb{P}(B \cap A^c) + \mathbb{P}(A \cap B)$. De ces deux égalités on déduit la formule d) □

Définition 2.3.2. *Une suite d'ensembles $(A_i)_{i\in\mathbb{N}}$ est dite croissante (respectivement décroissante) si pour tout $i \in \mathbb{N}$ on a $A_i \subseteq A_{i+1}$ (respectivement $A_i \supseteq A_{i+1}$). Si $(A_i)_{i\in\mathbb{N}}$ est croissante, on définit sa limite $\lim_i A_i$ comme $\cup_i A_i$; de manière similaire, si A_i est décroissante on définit $\lim_i A_i = \cap_i A_i$.*

La preuve des propriétés qui suivent n'est pas difficile mais, à la différence de la démonstration des propriétés précédentes, ne pourrait se faire sans autoriser des unions *dénombrables* :

Proposition 2.3.2. *a) Si A_i, $i \in \mathbb{N}$ est une famille croissante d'éléments de $\mathcal{F}$ dont l'union est A alors $A \in \mathcal{F}$ et la suite $\mathbb{P}(A_n)$ (qui est croissante bornée) converge vers $\mathbb{P}(A)$:*

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A);$$

b) Si A_i, $i \in \mathbb{N}$ est une famille décroissante d'éléments de $\mathcal{F}$ dont l'union est A alors $A \in \mathcal{F}$ et la suite $\mathbb{P}(A_n)$ (qui est décroissante positive) converge vers $\mathbb{P}(A)$:

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A);$$

c) Si A_i, $i \in \mathbb{N}$ est une famille dénombrable d'ensembles appartenant à $\mathcal{F}$ on a toujours (même si les A_i ne sont pas disjoints deux à deux)

$$\mathbb{P}\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) \leq \sum_{i=0}^{\infty} \mathbb{P}(A_i),$$

(où le membre de droite de l'inégalité précédente qui est la limite de la suite croissante peut éventuellement être infini).

Preuve. a) Définissons les ensembles B_n, $n \geq 0$ de la façon suivante : $B_0 = A_0$, et pour $n \geq 1$, $B_n = A_n \cap A_{n-1}^c$. Les B_n constituent une famille dénombrable d'ensembles disjoints deux à deux d'éléments de $\mathcal{F}$ et on peut donc écrire

$$\sum_{k=0}^{\infty} \mathbb{P}(B_k) = \mathbb{P}\Big(\bigcup_{k\in\mathbb{N}} B_k\Big).$$

c'est-à-dire

$$\lim_{N\to\infty} \sum_{k=0}^{N} \mathbb{P}(B_k) = \mathbb{P}\Big(\bigcup_{k\in\mathbb{N}} B_k\Big),$$

ou encore, puisque les B_k sont disjoints deux à deux

$$\lim_{N\to\infty} \mathbb{P}\Big(\bigcup_{k=0}^{N} B_k\Big) = \mathbb{P}\Big(\bigcup_{k\in\mathbb{N}} B_k\Big),$$

Mais

$$\bigcup_{k=0}^{N} B_k = A_N, \qquad \bigcup_{k=0}^{\infty} B_k = A$$

ce qui établit la preuve de a).

b) Il suffit de passer au complémentaire et d'utiliser a)

c) Pour $\omega \in \Omega$ définissons l'entier $\nu(\omega)$ comme étant le plus petit entier $k \geq 0$ pour lequel $\omega \in A_k$. L'ensemble C_n des $\omega \in \Omega$ pour lesquels $\nu(\omega) = n$ est l'ensemble

$$C_n = \{\omega \in \Omega, \nu(\omega) = n\} = A_n \cap (A_{n-1} \cup \cdots \cup A_0)^c$$

qui est clairement dans $\mathcal{F}$. Les ensembles C_n sont de toute évidence disjoints deux à deux et leur union pour $n \geq 0$ est $\cup_{n\in\mathbb{N}} A_n$ car pour tout ω dans $\cup_{n\in\mathbb{N}} A_n$ il existe un n tel que $\nu(\omega) = n$ c'est-à-dire il existe un n tel que $\omega \in C_n$. On a donc

$$\mathbb{P}\Big(\bigcup_{n\geq 0} A_n\Big) = \mathbb{P}\Big(\bigcup_{n\geq 0} C_n\Big) = \sum_{n=0}^{\infty} \mathbb{P}(C_n),$$

et comme $\mathbb{P}(C_n) \leq \mathbb{P}(A_n)$ (puisque $C_n \subset A_n$) on obtient la conclusion du c). □

Les deux propriétés précédentes a) et b) sont des propriétés de *continuité* des probabilités, puisque $\cup_i A_i = \lim_i A_i$ (resp. $\cap_i A_i = \lim_i A_i$) si la suite A_i est croissante (resp. decroissante).

2.4 Probabilités sur un ensemble au plus dénombrable

2.4.1 Construction des probabilités

Les espaces probabilisés les plus simples sont ceux où l'espace des états Ω est fini. On choisit en général comme tribu $\mathcal{F}$ l'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω (qui est bien une tribu). C'est ce que nous ferons (car le cas où $\mathcal{F}$ est une tribu plus petite que $\mathcal{P}(\Omega)$ s'y ramène). Cela étant, il reste à définir la probabilité. Supposons donc que $\mathbb{P}$ soit une probabilité sur $\mathcal{F} = \mathcal{P}(\Omega)$; tout ensemble $A \in \mathcal{F} = \mathcal{P}(\Omega)$ est fini (car inclus dans Ω qui est fini) et est par conséquent l'union (finie donc dénombrable) des singletons $\{a\}$ où a décrit A :

$$A = \bigcup_{a \in A} \{a\}.$$

Comme cette union est disjointe et finie on a

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

Supposons que $\Omega = \{\omega_1, \dots, \omega_n\}$ et notons $p_i = \mathbb{P}(\{\omega_i\})$; on a donc $p_i \in [0,1]$. D'autre part on a

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

Si on choisit $A = \Omega$ on voit que les p_i, qui sont dans $[0,1]$, vérifient

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

En conclusion :

Proposition 2.4.1. *Dans le cas où $\Omega = \{\omega_1, \dots, \omega_n\}$est fini, une probabilité $\mathbb{P}$ sur $\mathcal{F} = \mathcal{P}(\Omega)$ est déterminée par ses valeurs sur les singletons de Ω. Réciproquement si on se donne n nombres réels positifs $p_1, \dots, p_n$ dont la somme vaut 1 ($p_1 + \cdots + p_n = 1$) alors, l'application $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ qui à $A \in \mathcal{P}(\Omega)$ associe le réel (dans $[0,1]$)*

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

est une probabilité.

Preuve. Nous avons démontré la première partie de la proposition, démontrons la réciproque. Il suffit de vérifier que pour toute famille A_k, $k \in \mathbb{N}$, $A_k \subset \Omega$ où les A_k sont deux à deux disjoints, la formule (2.1) est vérifiée. Comme Ω est fini, il en est de même de $\mathcal{F} = \mathcal{P}(\Omega)$ et il suffit donc de considérer le cas où la famille $(A_k)_{k \in \mathbb{N}}$ est finie et est $A_0, \dots, A_N$, $A_k \neq A_l$ si $k \neq l$. Si on note I_k l'ensemble des indices $i \in \{1, \dots, n\}$ pour lesquels $\omega_i \in A_k$, les ensembles $I_0 \dots, I_N$ sont deux à deux disjoints et leur union I est l'ensemble des indices $i \in \{1, \dots, n\}$ pour lesquels $\omega_i \in \bigcup_{k=0}^{N} A_k$. On a donc

$$\mathbb{P}\left(\bigcup_{k=0}^{N} A_k\right) = \sum_{i \in I} p_i = \sum_{k=0}^{N} \sum_{i \in I_k} p_i = \sum_{k=0}^{N} \mathbb{P}(A_k),$$

ce qui termine la preuve. □

Remarque. On verifie aisément que la proposition 2.4.1 s'étend au cas où Ω est dénombrable ; dans ce cas, on peut aussi toujours supposer que $\mathcal{F} = \mathcal{P}(\Omega)$. En effet, toute la preuve repose sur le fait que pour tout $A \in \mathcal{F}$, on a $A = \cup_{a \in A}\{a\}$ est une union disjointe dénombrable ce qui est bien le cas si Ω est dénombrable.

2.4.2 Probabilités uniformes et lien avec la combinatoire

Un cas important est celui où tous les p_i, $1 \leq i \leq n$ précédents sont égaux. Comme leur somme doit valoir 1 ceci signifie que $p_1 = \cdots = p_n = \frac{1}{n}$. On dit dans ce cas que la probabilité $\mathbb{P}$ est *uniforme*. On a alors, pour tout sous-ensemble A de Ω

$$\mathbb{P}(A) = \sum_{i:\omega_i \in A} p_i = \#\{i \in \{1, \ldots, n\}, \omega_i \in A\} \times \frac{1}{n}$$

soit

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

Ainsi, quand on travaille avec une probabilité uniforme sur un ensemble fini, déterminer la probabilité d'un évévement revient à calculer son cardinal : on voit apparaîitre le lien avec la combinatoire. Un

Lancer de deux dés On se propose de modéliser le lancer de deux dés (numérotés de 1 à 6). On appèlera p_i la probabilité pour chaque dé, quand on le lance, d'obtenir la face i et on suppose de façon implicite que le lancer de chacun des deux dés est indépendant de l'autre (dans un sens qui pour l'instant n'est pas bien défini).

La modélisation du problème se fait de la façon suivante.

— *Espace des états :* on pose $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$ l'ensemble des couples (i, j) avec $1 \leq i \leq 6$, $1 \leq j \leq 6$.

— *Espace des événements :* on posera $\mathcal{F} = \mathcal{P}(\Omega)$.

— *Probabilité :* Si on suppose que quand on lance un dé, la probabilité de sortie de la face i égale p_i il est naturel de penser que la probabilité p_{ij} de sortie du couple (i, j) (i pour le 1er dé et j pour le second) est égal au produit $p_i p_j$ de la probabilité d'obtenir i sur le premier dé par la probabilité d'obtenir j sur le second. Naturellement, les p_i vérifient $p_i \geq 0$ et $\sum_{i=1}^{6} p_i = 1$. On décide donc de définir, ayant à l'esprit la Proposition 2.4.1, la probabilité $\mathbb{P}$ sur $\mathcal{F}$ par $\mathbb{P}(\{(i, j)\}) = p_i p_j$. Pour que cela soit possible il faut vérifier que les $p_i p_j$ sont positifs, ce qui est évident, et que $\sum_{(i,j)\in\Omega} p_i p_j = 1$. Mais cela aussi est clair car

$$\sum_{1 \leq i,j \leq 6} p_i p_j = \left(\sum_{i=1}^{6} p_i\right) \times \left(\sum_{j=1}^{6} p_j\right) = 1 \times 1 = 1.$$

Avec cette modélisation, on peut ainsi déterminer la probabilité pour que la somme des deux dés après lancer vaille 7. Cet événement, appelons-le A, s'écrit de façon mathématique

$$A = \{(i, j) \in \Omega : i + j = 7\}$$

ou de façon explicite

$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

La probabilité de A vaut

$$\mathbb{P}(A) = \sum_{1 \leq i,j \leq 6 : i+j=7} p_i p_j.$$

Si on suppose que pour tout $1 \leq i \leq 6$, $p_i = 1/6$ on obtient $\mathbb{P}(A) = \#A/\#\Omega = 6/36 = 1/6$. De la même façon l'événement B : "après le lancer des deux dé la face du premier dé est un nombre pair", est mathématiquement l'ensemble

$$B = \{(i, j) : i = 2, 4, 6\}.$$

Si on suppose que pour tout $1 \leq i \leq 6$, $p_i = 1/6$ on obtient $\mathbb{P}(B) = \#B/\#\Omega = 18/36 = 1/2$.

La conjonction de ces deux événements, “la somme des deux dés vaut 7” et “le résultat que donne le premier dé est un nombre pair” est mathématiquement l'ensemble $A \cap B = \{(2,5),(4,3),(6,1)\}$. Sa probabilité (si on suppose que pour tout $1 \leq i \leq 6$, $p_i = 1/6$) égale $\#(A \cap B)/\#(\Omega) = 3/36 = 1/12$.

Exercice 2.4.1. *On tire cinq cartes d'un jeu de 32 cartes. Quelle est la probabilité d'obtenir un full c'est-à-dire deux cartes de même valeur et trois autres cartes de même valeur. On supposera chaque tirage équiprobable.*

Exercice 2.4.2. *(cf. exercice 1.3.1) Une urne contient n boules noires et b boules blanches.*
a) On effectue N tirages sans remises. Quelle est la probabilité d'obtenir x boules noires ?
b) Même question si les tirages sont avec remises.

On supposera les tirages équiprobables. On prendra soin de bien définir l'espace probabilisé sur lequel on travaille.

Solution. a) On pose $T = \{1,2,\ldots,n\}$ l'ensemble des temps et $E = \{1,2,\ldots,N,N+1,\ldots,N+M\}$ l'ensemble des boules où {1,2,..., N} représente l'ensemble des boules noires et $\{N+1,\ldots,N+M\}$ celui des boules blanche. On travaille avec Ω l'ensemble des injections de T dans E et $\mathcal{F} = \mathcal{P}(\Omega)$ la tribu totale sur Ω. On pose $\mathbb{P}$ la probabilité uniforme sur $(\Omega, \mathcal{P}(\Omega))$ (l'unique probabilité sur $(\Omega, \mathcal{F})$ telle que $\mathbb{P}(\{\omega\}) = 1/\#\Omega$ quelque soit $\omega \in \Omega$). L'événement “obtenir x boules noires” est l'ensemble $A_x == \{\omega \in \Omega : \#\{1 \leq i \leq n : \omega(i) \leq N\}\}$ (les ω sont des applications, $\omega(i)$ est un nombre entre 1 et $N+M$). Alors, d'après les calculs effectués à l'exercice 1.3.1 :

$$\mathbb{P}(A_x) = \frac{\binom{N}{x}\binom{M}{n-x}}{\binom{N+M}{n}}.$$

b) On travaille avec l'ensemble $\Omega' = E^T$ des applications de T dans E. On pose $\mathcal{F}' = \mathcal{P}(\Omega')$ et on prend $\mathbb{P}'$ la probabilité uniforme. L'événement obtenir x boules noires est $A'_x = \{\omega \in \Omega' : \#\{1 \leq i \leq n : \omega(i) \leq N\}\}$. Alors, d'après les calculs effectués à l'exercice 1.3.1 :

$$\mathbb{P}'(A'_x) = \frac{\#A_x}{\#\Omega} = \binom{n}{x} p^x (1-p)^{n-x},$$

où $p = \frac{N}{N+M}$. △

2.5 Evénements indépendants

Définition 2.5.1. *Une famille d'évévements $(A_i)_{i\in I}$ est dite indépendante (ou encore les événements de la famille $(A_i)_{i\in I}$ sont dits mutuellement indépendants) si pour toute sous famille finie $A_{i_1},\ldots,A_{i_n}$ on a*

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_n}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_n}).$$

Remarque. Il faut bien faire attention au faits suivants :
(i) Il n'est évidemment pas suffisant que $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ quelque soient $i \neq j$ dans I pour que la famille $(A_i)_{i\in I}$ soit indépendante.
(ii) Il n'est pas non plus suffisant que $\mathbb{P}(\cap_{i\in I} A_i) = \prod_{i\in I} \mathbb{P}(A_i)$, même si I est un ensemble fini. Pour s'en convaincre, considerer la famille $(A, A, \emptyset)$: en effet $\mathbb{P}(A \cap A \cap \emptyset) = \mathbb{P}(\emptyset) = 0 = \mathbb{P}(A)\mathbb{P}(A)\mathbb{P}(\emptyset)$, mais la famille n'est pas forcément indépendante puisque $\mathbb{P}(A \cap A) \neq \mathbb{P}(A)\mathbb{P}(A)$ si $\mathbb{P}(A) = 1/2$ par exemple.
(iii) Il faut bien faire la différence entre des événements indépendants et des événements disjoints : en effet, si $A \cap B = \emptyset$, A et B ne peuvent être indépendants que si $0 = \mathbb{P}(\emptyset) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ c'est-à-dire si $\mathbb{P}(A) = 0$ ou $\mathbb{P}(B) = 0$.

Exercice 2.5.1. *Démontrer que si A, B sont deux évévements indépendants les quatre familles (A, B), (A, B^c), (A^c, B), (A^c, B^c) sont chacune indépendantes.*

Solution. Montrons par exemple que (A^c, B) est indépendante :

$$\mathbb{P}(A^c \cap B) = \mathbb{P}((E - A) \cap B) = \mathbb{P}(B - (A \cap B)) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

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

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B)(1 - \mathbb{P}(A)) = \mathbb{P}(B)\mathbb{P}(A^c).$$

Par symétrie, (A, B^c) est indépendante. Mais alors, en posant $A_0 = A$ et $B_0 = B^c$ on obtient que $(A_0^c, B_0) = (A^c, B^c)$ est aussi indépendante. △

Exercice 2.5.2. *Trouver un exemple où chacune des familles (A, B), (B, C), (C, A) est indépendante mais pas la famille (A, B, C).*

Solution. On peut par exemple choisir une expérience avec deux pièces : $\Omega = \{0, 1\} \times \{0, 1\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, et $\mathbb{P}$ la probabilité uniforme. Posons $A = \{1\} \times \{0, 1\} = \{(1, 0), (1, 1)\}$, $B = \{0, 1\} \times \{1\} = \{(0, 1), (1, 1)\}$ et $C = \{(0, 0), (1, 1)\}$: A est l'événement que la première pièce soit Face (1), B est l'événement que la deuxième pièce soit Face et C est l'événement que les deux pièces soient identiques. Alors $A \cap B = \{(1, 1)\} = A \cap C = B \cap C = \{(1, 1)\}$ et comme $\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 1/2$, $\mathbb{P}(C) = 1/2$ et $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = 1/4$, les familles (A, B), (B, C) et (A, C) sont chacune indépedantes. Par contre $A \cap B \cap C = \{(1, 1)\}$ et donc $\mathbb{P}(A \cap B \cap C) = 1/4$ alors que $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 1/8$. △

Un résultat important qui permet de construire des familles d'événements indépendants est le suivant ; nous avons avant tout besoin de la definition suivante, qui aussi sera utile par la suite, notamment pour construire des tribus non-triviales et utiles sur des ensembles non-denombrables.

Proposition 2.5.1. *Soit Ω un ensemble et $\mathcal{S}$ un sous-ensemble de parties de Ω (un sous-ensemble de $\mathcal{P}(\Omega)$) sans structure particulière. Il existe une unique tribu $\mathcal{F}$ qui contient $\mathcal{S}$ et qui est minimale pour cette propriété c'est-à-dire :*
i) ($\mathcal{S} \subset \mathcal{F}$) : pour tout $A \in \mathcal{S}$ on a $A \in \mathcal{F}$
ii) (minimalalité) : si $\mathcal{F}'$ est une autre tribu telle que $\mathcal{S} \subset \mathcal{F}'$ alors $\mathcal{F} \subset \mathcal{F}'$.
On appelle $\mathcal{F}$ la tribu engendrée *par la partie $\mathcal{S}$ et on la note $\mathcal{F} = \sigma(\mathcal{S})$.*

Preuve. Considérons l'ensemble $\mathbb{T}$ des tribus $\mathcal{C}$ de Ω tel que $\mathcal{S} \subset \mathcal{C}$. Cet ensemble $\mathbb{T}$ est non vide puisqu'il contient la tribu $\mathcal{P}(\Omega)$ et puisque $\mathcal{S} \subset \mathcal{P}(\Omega)$. Notons $\mathcal{F}$ l'intersection des $\mathcal{C}$ quand $\mathcal{C}$ décrit $\mathbb{T}$. C'est un sous-ensemble de $\mathcal{P}(\Omega)$ qui contient $\mathcal{S}$ mais c'est également une tribu comme il est facile de vérifier (nous recommandons au lecteur de vérifier ce point). Ainsi, $\mathcal{F}$ est une tribu contenant $\mathcal{S}$ et appartient donc à $\mathbb{T}$. Comme $\mathcal{F}$ est l'intersection des $\mathcal{C}$ décrivant $\mathbb{T}$ on a pour toute tribu $\mathcal{C}$ contenant $\mathcal{S}$ l'inclusion $\mathcal{F} \subset \mathcal{C}$: cela signifie que $\mathcal{F}$ est la plus petite tribu contenant $\mathcal{S}$. □

Théorème 2.5.1. *Soient $(A_i)_{i \in I}$ une famille d'événements indépendants et $I = I_1 \cup \cdots \cup I_p$ une partition de I. Notons pour $1 \leq k \leq p$, $\mathcal{B}_k$ la tribu engendrée par les événements A_i, $i \in I_k$. Si $B_1, \ldots B_p$ sont des événements tels que $B_k \in \mathcal{B}_k$, alors la famille d'événements $(B_k)_{1 \leq k \leq p}$ est indépendante.*

Preuve. Nous illustrons la preuve dans le cas où $p = 2$. Démontrons déjà que pour tout événement A_i, $i \in I_1$ et tout événement B_2 dans $\mathcal{B}_2$ on a

$$\mathbb{P}(A_i \cap B_2) = \mathbb{P}(A_i)\mathbb{P}(B_2). \tag{2.2}$$

Pour cela, notons $\mathcal{C}_2$ l'ensemble des événements B_2 pour lesquels cette relation est satisfaite pour tout $i \in I_1$. On constate déjà que $\mathcal{C}_2$ est une tribu (**exercice**). Ensuite, on observe que d'après

l'hypothèse d'indépendance des A_i, $i \in I$, les événements A_j, $j \in I_2$ appartiennent à $\mathcal{C}_2$. Par définition de la tribu engendrée et sa minimalité, cela signifie que $\mathcal{B}_2 \subset \mathcal{C}_2$. Par conséquent, la relation (2.2) est vraie pour tout $B_2 \in \mathcal{B}_2$.

Démontrons à présent que pour tout événement $B_1 \in \mathcal{B}_1$ et tout événement B_2 dans $\mathcal{B}_2$ on a

$$\mathbb{P}(B_1 \cap B_2) = \mathbb{P}(B_1)\mathbb{P}(B_2). \tag{2.3}$$

Notons $\mathcal{C}_1$ l'ensemble des événements B_1 pour lesquels cette relation est satisfaite pour tout $B_2 \in \mathcal{B}_2$. On constate comme précédemment que $\mathcal{C}_1$ est une tribu (**exercice**) et qu'elle contient tous les A_i, $i \in I_1$. La définition du fait que $\mathcal{B}_1$ est la tribu engendrée par les A_i, $i \in I_1$ montre que $\mathcal{B}_1 \subset \mathcal{C}_1$ et donc la relation (2.3) est satisfaite pour tout $B_1 \in \mathcal{B}_1$ et tout $B_2 \in \mathcal{B}_2$. □

Exercice 2.5.3. *Refaire à l'exercice 2.5.1 à l'aide du théorème 2.5.1 (ça doit être trivial !).*

2.6 Jeu de pile ou face : le modèle de Bernoulli

2.6.1 Jeu de n pile ou face : nombre fini de pièces

Espace des états On se propose de modéliser un jeu où l'on lance n fois une pièce (Pile/Face). De façon équivalente un expérimentateur réalise n expériences, le résultat de chaqu'une d'entre elles pouvant être positif (1) ou négatif (0). Il est naturel de décrire le jeu ou l'expérience précédentes de la façon suivante : on choisit comme espace des états l'ensemble Ω de toutes les suites de longueur n constituées de 0 ou de 1. Une telle suite est donc un n-uplet $\omega = (\omega_1, \ldots, \omega_n)$ chaque ω_i, $1 \leq i \leq n$ appartenant à l'ensemble à deux éléments $\{0, 1\}$. Ainsi

$$\Omega = \{0, 1\}^n,$$

et a 2^n éléments et est fini

Tribu des événements Comme Ω est fini, nous choisirons comme tribu $\mathcal{F}_n = \mathcal{P}(\Omega)$, l'ensemble des parties de Ω (qui a donc 2^{2^n} mais cela n'a pas d'importance). Cette tribu nous permet de décrire des évévements. Par exemple l'évévement (A) "obtenir k Pile lors des n lancers" est décrit par l'ensemble $A \in \mathcal{F}_n$ (Pile=1, Face=0)

$$A = \Big\{\omega = (\omega_1, \ldots, \omega_n), \sum_{i=1}^{n} \omega_i = k.\Big\}.$$

L'évévement (B) "on tire au moins un Pile" est décrit par l'ensemble

$$B = \{\omega = (\omega_1, \ldots, \omega_n), \exists i \in \{1, \ldots, n\} \quad \omega_i = 1\}.$$

L'évévement "(A) et (B)" est décrit par l'intersection $A \cap B$, l'événement "non A" est décrit par A^c, l'évévement "A ou B" par $A \cup B$ etc.

Probabilité Le choix de la probabilité sur notre ensemble est dicté par le jeu ou l'expérience que l'on modélise. Ainsi, on ne modélisera pas de la même façon un jeu ou pile et face ont les mêmes chances de sortir qu'un jeu où pile a deux fois plus de chance de sortir que face. Ici, comme Ω est fini, il suffit pour définir $\mathbb{P}$ de manière unique de fixer la probabilité des singletons, c'est-à-dire dans notre cas $\mathbb{P}((\epsilon_1, \epsilon_2, \ldots, \epsilon_n))$ pour tous les $(\epsilon_1, \epsilon_2, \ldots, \epsilon_n) \in \Omega$; notons qu'il faut le faire de manière à ce que

$$\sum_{(\epsilon_1, \ldots, \epsilon_n) \in \Omega} \mathbb{P}(\{(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)\}) = 1,$$

ce que nous devrons vérifier.

Pour l'instant, il faut d'abord trouver un choix en accord avec l'expérience pour $\mathbb{P}(\{(\epsilon_1, \epsilon_2, \dots, \epsilon_n)\})$. Pour cela, nous voudrions bien sûr que, conformément à l'expérience (le résultat de la pièce i n'influence pas les résultats des autres pièces !) les événements

$$C_{i,\epsilon} = \{\omega \in \Omega : \omega_i = \epsilon\} = \{\text{le résultat de la } i\text{ème pièce est } \epsilon\},$$

soient indépendants (ce que l'on veut dire ici, c'est que pour tout $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ fixé, les événements C_{i,ϵ_i} sont indépendants, mais évidemment $C_{1,0}$ et $C_{1,1}$ ne sont pas indépendants...). Or

$$\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\} = C_{1,\epsilon_1} \cap C_{2,\epsilon_2} \cap \dots \cap C_{n,\epsilon_n},$$

et cette volonté de rendre la famille $(C_{i,1})_{1 \le i \le n}$ indépendante impose que

$$\mathbb{P}(\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}) = \mathbb{P}(C_{1,\epsilon_1}) \times \mathbb{P}(C_{2,\epsilon_2}) \times \dots \times \mathbb{P}(C_{n,\epsilon_n}).$$

Si nous avons une pièce unique que l'on lance n fois, la probabilité d'obtenir face est la même pour tous les lancers et on peut raisonablement fixer $\mathbb{P}(C_{i,1}) = p$, $\mathbb{P}(C_{i,0}) = 1 - p$ pour un réel $p \in [0, 1]$ fixé, et ce quelque soit $i \in \{1, 2, \dots, n\}$. On a donc, pour tout $i \in \{1, 2, \dots, n\}$ et $\epsilon \in \{0, 1\}$,

$$\mathbb{P}(C_{i,\epsilon}) = p^\epsilon (1 - p)^{1-\epsilon},$$

ce qui vaut bien p si $\epsilon = 1$ et $(1 - p)$ sinon. Nous sommes donc conduits à fixer

$$\mathbb{P}(\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}) = p^{\#\{i:\epsilon_i=1\}}(1 - p)^{\#\{i:\epsilon_i=0\}} = p^{\#\{i:\epsilon_i=1\}}(1 - p)^{n-\#\{i:\epsilon_i=1\}},$$

ce qui fixe bien la probabilité de tous les singletons, et fixe la probabilité de façon unique pourvu que $\mathbb{P}(\{\epsilon_1, \dots, \epsilon_n\}) \ge 0$ pour tout $(\epsilon_1, \dots, \epsilon_n) \in \Omega$, ce qui est clair, et que la somme vaille 1, ce que nous vérifions immédiatement.

On remarque que $\#\{i : \epsilon_i = 1\} = \epsilon_1 + \epsilon_2 + \dots + \epsilon_n$. Pour $0 \le r \le n$, on pose

$$\Omega_r = \{\omega \in \Omega : \omega_1 + \omega_2 + \dots + \omega_n = r\}$$

pour l'ensemble des suites de Ω qui contiennent r pièces qui valent 1. Les ensembles Ω_r, $0 \le r \le n$ forment une partition de Ω. On a

$$\begin{aligned}
\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega_0 \sqcup \dots \sqcup \Omega_n} \mathbb{P}(\{\omega\}) &= \sum_{r=0}^{n} \sum_{\omega \in \Omega_r} \mathbb{P}(\{\omega\}) \\
&= \sum_{r=0}^{n} \sum_{\omega \in \Omega_r} p^r (1 - p)^{n-r} \\
&= \sum_{r=0}^{n} p^r (1 - p)^{n-r} \sum_{\omega \in \Omega_r} 1 \\
&= \sum_{r=0}^{n} p^r (1 - p)^{n-r} \#\Omega_r \\
&= \sum_{r=0}^{n} p^r (1 - p)^{n-r} \binom{n}{r} = 1,
\end{aligned}$$

par le binome de Newton. On a utilisé le fait que $\#\Omega_r = \binom{n}{r}$ comme il faut précisément choisir les r positions des "1", c'est-à-dire un sous-ensemble de l'ensemble des positions $\{1, 2, \dots, n\}$ de cardinal r. D'après la Proposition 2.4.1 on peut donc définir une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F}_n)$ telle que

$$\mathcal{P}(\{(\omega_1, \dots, \omega_n)\}) = p^{\omega_1 + \dots + \omega_n}(1 - p)^{n-(\omega_1 + \dots + \omega_n)}.$$

On appelera l'espace probabilisé ainsi construit le modèle (fini) de Bernoulli de paramètre p $(0 \le p \le 1)$.

Mentionnons une propriété importante du modèle de Bernoulli, qui ressort des caluls précédents :

Proposition 2.6.1. *Dans le modèle de Bernoulli de paramètre p, l'événement*

$$\Omega_k = \{(\omega_1, \ldots, \omega_n) : \omega_1 + \cdots + \omega_n = k\}$$

admet pour probabilité

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

Preuve. On a en effet

$$\begin{aligned}\mathbb{P}(\Omega_k) &= \sum_{(\omega_1,\ldots,\omega_n)\in\Omega_k} p^{\omega_1+\cdots+\omega_n}(1-p)^{n-(\omega_1+\cdots+\omega_n)} \\ &= p^k(1-p)^{n-k}\#\Omega_k.\end{aligned}$$

Or, le cardinal de $\#\Omega_k = \binom{n}{k}$, le nombre de façon de choisir k éléments parmi n. □

Exercice 2.6.1. *Calculer dans chacun des cas $p = 1/2$ ou $p = 2/3$ les probabilités des évévements A et B définis précédemment.*

Exemple 2.6.1. *Une application moins triviale des tribus engendré. On remarque que, conformément à l'intuition, les événements*

$$A = \left\{\omega \in \Omega : \sum_{\{i:1\le 2i\le n\}} \omega_i = k\right\} \qquad \text{et} \qquad B = \left\{\omega \in \Omega : \sum_{\{j:1\le 2j+1\le n\}} \omega_j = \ell\right\}$$

sont indépendants quelque soient k et ℓ. En effet, A ne concerne que les pièces d'indices pairs et B que les pièces d'indices impairs : ici, on a $A \in \sigma(C_{2i,1} : 2i \le n)$ et $B \in \sigma(C_{2i+1,1} : 2i+1 \le n)$ et les suites d'indices pairs et impairs sont évidemment disjointes, ce qui nous permet d'appliquer le théorème 2.5.1 pour dire que A et B sont effectivement indépendants.

2.6.2 Jeu infini de pile ou face : nombre infini de pièces

L'espace des états Nous présentons dans ce paragraphe la modélisation du jeu de pile ou face où l'on joue une infinité de fois. Il est naturel d'introduire comme espace des états l'ensemble Ω des suites $\omega = (\omega_1, \omega_2, \ldots)$ où les ω_i valent 0 ou 1 : en effet, Ω doit au moins être assez riche pour décrire le résultat de chacune des pièces. On a ainsi $\Omega = \{0,1\}^{\mathbb{N}-\{0\}}$; cet ensemble n'est pas dénombrable (voir section 1.4), et il ne tombe donc pas sous le coup de la construction des probabilités sur les ensembles au plus dénombrables. En particulier c'est un problème plus délicat de trouver une tribu raisonnable sur cet ensemble ; $\mathcal{P}(\Omega)$ ne convient pas (trop grand).

Pourtant, on aimerait par exemple pouvoir décrire un évévement du type : "en moyenne pile sort deux fois plus souvent que face" qui de façon ensembliste est l'ensemble des $\omega = (\omega_1, \omega_2, \ldots) \in \Omega$ pour lesquels la limite quand n tend vers l'infini de la suite

$$\frac{1}{n}\sum_{i=1}^{n} \omega_i$$

existe et vaut 2/3. Un moment de réflexion montre que cet évévement n'appartient à aucune des tribus $\mathcal{F}_n$ qui modélisent un jeu de n pile/face[1].

1. On peut toujours considérer un jeu de n pile/face comme un cas particulier d'un jeu infini de pile/face : Il suffit d'associer à toute suite $\omega = (\omega_1, \ldots, \omega_n)$ de $\{0,1\}^n$ la suite $\tilde{\omega} \in \{0,1\}^{\mathbb{N}} - \{0\}$ définie par $\tilde{\omega}_i = \omega_i$ si $1 \le i \le n$ et $\tilde{\omega}_i = 0$ si $i \ge n+1$

Construction de la tribu des événements Une tribu raisonable doit au moins contenir tous les événements

$$C_{i,\epsilon} = \{\omega \in \Omega : \omega_i = \epsilon\},$$

où i décrit $\mathbb{N} - \{0\}$ et ϵ décrit $\{0, 1\}$, qui disent juste que le résultat de la pièce numéro i est ϵ. Nous définirons la tribu $\mathcal{F}$ sur Ω de la façon suivante : la tribu $\mathcal{F}$ est la tribu engendrée (au sens de la proposition 2.5.1) par tous les $C_{i,\epsilon}$, $i \in \mathbb{N}$, $\epsilon \in \{0, 1\}$. Par ailleurs, on pourrait vérifier que les événements (cruciaux !) tels que

$$A = \big\{\omega \in \Omega : \text{le nombre de } i \in \mathbb{N} \text{ tels que } \omega_i = 1 \text{ est infini}\big\}$$

ou encore

$$B = \left\{\omega \in \Omega : \frac{1}{n}\sum_{i=1}^{n} \omega_i \text{ converge}\right\}$$

qui ne sont dans aucun des $\mathcal{F}_n$ sont effectivement dans $\mathcal{F}$. Dans le cas de A c'est assez clair car

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} C_{k,1}.$$

Le cas de B est plus délicat (il ne faut que des unions et intersection dénombrables dans la description !).

Construction de la probabilité Il reste à présent à construire une probabilité sur $\mathcal{F}$ ce qui est assez délicat. Si on joue avec une pièce qui donne Pile (resp. Face) avec probabilité p (resp. $1 - p$) il est naturel d'attribuer à tout évévement $C_{i,\epsilon}$ la probablité p si $\epsilon = 1$ et 0 sinon : il est également naturel de demander que la probabilité d'un évévement de la forme [2]

$$\{\omega \in \Omega : \omega_{i_1} = \epsilon_1, \dots, \omega_{i_n} = \epsilon_n\} = C_{i_1,\epsilon_1} \cap \cdots \cap C_{i_r,\epsilon_n}$$

soit égale à $p^r(1 - p)^{n-r}$ où r est le nombre de $1 \leq k \leq n$ pour lesquels $\epsilon_k = 1$; on remarquera que $r = \sum_{i=1}^{n} \epsilon_i$. Il n'est en revanche pas du tout clair que l'on puisse attribuer à tout évévement de la tribu $\mathcal{F}$ une probabilité qui soit *compatible* avec ces choix. En fait c'est possible :

Théorème 2.6.1. *Il existe une unique mesure de probabilité* $\mathbf{P}$ *définie sur* $(\Omega, \mathcal{F})$ *telle que pour tous* $n \geq 1$, $i_1, \dots, i_n \in \mathbb{N} - \{0\}$, $\epsilon_1, \dots, \epsilon_n \in \{0, 1\}$ *on ait*

$$\mathbf{P}(C_{i_1,\epsilon_1} \cap \cdots \cap C_{i_n,\epsilon_n}) = p^{\epsilon_1+\cdots+\epsilon_n}(1 - p)^{n-(\epsilon_1+\cdots+\epsilon_n)}.$$

Exemple 2.6.2. *On peut étendre l'exemple du jeu de Bernoulli fini : toute famille* $(A_k)_k$ *d'événements qui "ne regardent que des ensembles de pièces disjoints" sont indépendants par le théorème 2.5.1 : en effet, le fait que* A_k *"ne regarde que les pièces dans* I_k*" se traduit par le fait que* $A_k \in \sigma(C_{i,1} : i \in I_k)$ *et le fait que les* $(I_k)_k$ *sont deux-à-deux disjoints nous permet de conclure. Par exemple, les événements*

$$A = \left\{\omega \in \Omega : \left(\frac{1}{n}\sum_{i=0}^{n-1} \omega_{2i}\right)_n \textit{ converge}\right\} \quad \textit{et} \quad B = \left\{\omega \in \Omega : \exists n_0 \in \mathbb{N}, \frac{1}{n}\sum_{i=0}^{n-1} \omega_{2i+1} \geq \frac{3}{4} \textit{ si } n \geq n_0\right\}$$

sont indépendants.

2. Cet évévement décrit l'expérience suivante : au temps $i_1, \dots, i_n$, on observe $\epsilon_1, \dots, \epsilon_n$ et on ne précise pas ce qui se passe aux autres temps

2.7 Probabilités conditionnelles

Définition 2.7.1. *Si A et Bsont deux évenements de la tribu $\mathcal{F}$et si $\mathbb{P}(B) > 0$ on définit la probabilité de A sachant B comme étant*

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

Remarque. Deux événements A et B de probabilité non nulle sont indépendants précisément si $\mathbb{P}(A|B) = \mathbb{P}(A)$ et $\mathbb{P}(B|A) = \mathbb{P}(B)$: c'est-à-dire que le fait de savoir que A arrive n'influence pas la probabilité que B arrive et réciproquement.

Exemple 2.7.1. *On considère une classe de 100 élèves : il y a 40 filles, 60 garçons, 20 gaucher(e)s et 80 droitier(e)s, et 15 garçons gauchers. Soit $\mathbb{P}$ la probabilité uniforme (on choisit un élève au hasard). On note G et D les événements être gaucher / droitier, ainsi que ♂ / ♀ pour garçon / fille. Alors :*

$$\mathbb{P}(G|♀) = \frac{\mathbb{P}(G, ♀)}{\mathbb{P}(♀)} = \frac{5/100}{40/100} = \frac{1}{8} \qquad \textit{et} \qquad \mathbb{P}(G|♂) = \frac{\mathbb{P}(G, ♂)}{\mathbb{P}(♂)} = \frac{15/100}{60/100} = \frac{1}{4}.$$

En fait tout évenement $B \in \mathcal{F}$ définit une nouvelle probabilité sur $(\Omega, \mathcal{F})$ dès que $\mathbb{P}(B) > 0$:

Proposition 2.7.1. *Si $B \in \mathcal{F}$ est tel que $\mathbb{P}(B) > 0$, l'application $\mathbb{P}_B : \mathcal{F} \to [0, 1]$ qui à $A \in \mathcal{F}$ associe $\mathbb{P}_B(A) = \mathbb{P}(A|B)$ est une probabilité.*

La preuve de cette proposition est laissée en exercice au lecteur.

Supposons à présent que nous ayons une partition de Ω en évévements $B_1, \ldots, B_r$ (c'est-à-dire que les évévements $B_1, \ldots, B_r$ sont disjoints deux à deux et que leur union soit égale à Ω) et que pour tout i on ait $\mathbb{P}(B_i) > 0$. On peut donc définir r probabilités $\mathbb{P}(\cdot\ |B_i)$.

Théorème 2.7.1 (Formule des causes)**.** *Sous les hypothèses précédentes*

$$\mathbb{P}(A) = \sum_{i=1}^{r} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Preuve. Il suffit de remarquer que A est l'union disjointe des $(A \cap B_i)$ et donc

$$\mathbb{P}(A) = \sum_{k=1}^{r} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{r} \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(B_i)} \cdot \mathbb{P}(B_i)$$

ce qui est la formule annoncée. □

Exercice 2.7.1. *Chacune des trois boîtes posées sur la table contient exactement deux pièces de monnaie. Dans une des trois boîtes il y a deux pièces d'or, dans une autre il y a deux pièces d'argent, dans la troisième se trouvent une pièce d'or et une pièce d'argent.*

On choisit une boîte au hasard (probabilité uniforme), on y choisit ensuite une pièce au hasard, et il se trouve que c'est une pièce d'or. Sachant cela, quelle est la probabilité (conditionnelle) que l'autre pièce se trouvant dans la même boite soit également en or ?

Solution. On a les boites avec deux pièces d'or $B_1 = \{O, O\}$, celle avec une pièce d'or et une pièce d'argent $B_2 = \{O, A\}$ et enfin la dernière avec deux pièces d'argent $B_3 = \{A, A\}$. Soit OR_1 l'événement d'on a tiré une première pièce en or et OR_2 pour que la deuxième pièce soit en or. Alors, comme $\mathrm{OR}_1 \cap \mathrm{OR}_1$ est évidemment l'événement d'avoir choisit la boite B_1, on a

$$\begin{aligned}\mathbb{P}(\mathrm{OR}_2|\mathrm{OR}_1) = \frac{\mathbb{P}(\mathrm{OR}_2, \mathrm{OR}_1)}{\mathbb{P}(\mathrm{OR}_1)} &= \frac{\mathbb{P}(B_1)}{\mathbb{P}(B_1) \times \mathbb{P}(\mathrm{OR}_1|B_1) + \mathbb{P}(B_2)\mathbb{P}(\mathrm{OR}_1|B_2) + \mathbb{P}(B_3)\mathbb{P}(\mathrm{OR}_1|B_3)} \\ &= \frac{\frac{1}{3}}{\frac{1}{3} \times 1 + \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 0} = \frac{2}{3}.\end{aligned}$$

On peut retrouver ce résultat rapidement : comme le choix des boites est uniforme et le nombre de pièce le même dans chaque boite, la première pièces est une pièce uniforme parmi les 6 pièces. Il y a trois possibilités pour qu'elle soit en or ; mais pour que la deuxième soit aussi en or, il faut avoir choisit une pièce de la boite B_1, et il y a deux possibilités : au final $\mathbb{P}(\mathrm{OR}_2|\mathrm{OR}_1) = 2/3$. Il faut cependant être prudent avec ce genre de raisonnement et vérifier chaque étape. △

Dans la pratique un problème courant est de calculer $\mathbb{P}(B_i|A)$ connaissant les $\mathbb{P}(A|B_j)$.

Théorème 2.7.2 (Formule de Bayes)**.** *Sous les hypothèses précédentes :*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{i=1}^{r} \mathbb{P}(A|B_j)\mathbb{P}(B_j)}.$$

Preuve. Il suffit d'écrire

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

et d'utiliser la formule des causes. □

Exercice 2.7.2. *Une maladie touche* 1 *personne sur 1000. On dispose d'un test de dépistage imparfait. On suppose qu'un test donné à une personne malade est positif dans* 99 *cas sur* 100*, et qu'un test donné à une personne saine est négatif dans* 999 *cas sur* 1000*. Un patient est testé positif. Quelle est la probabilité qu'il soit effectivement malade ? Conclusion ?*

Solution. On choisit un ensemble d'état assez riche : $\{M, S\} \times \{P, N\}$ où M et S représentent malade et sain, alors que P et N représentent test positif et test négatif. La probabilité est $\mathbb{P}$ est telle que $\mathbb{P}(M) = 1/1000$ et $\mathbb{P}(P|M) = 99/100$, $\mathbb{P}(N|S) = 99/100$. On en déduit :

$$\mathbb{P}(M|P) = \frac{\mathbb{P}(P|M)\mathbb{P}(M)}{\mathbb{P}(P|M)\mathbb{P}(M) + \mathbb{P}(P|S)\mathbb{P}(S)} = \frac{\frac{99}{100} \times \frac{1}{1000}}{\frac{99}{100} \times \frac{1}{1000} + \frac{1}{1000} \times \frac{999}{1000}} = \frac{990}{990 + 999} \approx 0.5.$$

Conclusion : ce test est donc complètement inutile : on obtient sensiblement le même résultat en jetant une pièce (indépendante du patient !) et en déclarant positif si on tombe sur face ! △

Chapitre 3

Variables aléatoires réelles

3.1 Variables aléatoires réelles

3.1.1 Définition et premières propriétés

Dans toute la suite on suppose que $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilité.

Définition 3.1.1. *Une variable aléatoire réelle (en abrégé v.a.r. mais nous écrirons plus simplement v.a.) est une application $X : \Omega \to \mathbb{R}$ telle que pour tout intervalle I de $\mathbb{R}$ l'ensemble $X^{-1}(I)$ des $\omega \in \Omega$ tels que $X(\omega) \in I$, appartient à la tribu $\mathcal{F}$.*

Remarque. (1) La raison pour laquelle on impose ce concept de variable aléatoire est le suivant : une fonction n'est une "observable" intéressante de l'experience aléatoire que si on peut mesurer la probabilité des valeurs qu'elle prend, et il faut donc que les $\{\omega \in \Omega : X(w) \in A\}$ soient des événements pour un assez grand nombre de sous-ensembles A de $\mathbb{R}$; la restriction aux intervalles s'avère un choix judicieux et pratique.
(2) Evidemment, si $\mathcal{F} = \mathcal{P}(\Omega)$ alors toute fonction $X : \Omega \to \mathbb{R}$ est une variable aléatoire puisque $X^{-1}(I) \subseteq \Omega$ quelque soit I.

Exemple 3.1.1. *Dans l'exemple du lancer de deux dé, on a $\Omega = \{1, 2, \ldots, 6\}^2$. La fonction $X : \Omega \to \mathbb{R}$ donnée par $X(i, j) \to i$ donne le résultat du premier dé. C'est une v.a. puisque $\mathcal{F} = \mathcal{P}(\Omega)$. De manière similaire, la fonction $Y : \Omega \to \mathbb{R}$ donnée par $Y(i, j) = i + j$ qui rend la somme des deux dé est aussi une variable aléatoire.*

Proposition 3.1.1. *Si $X : \Omega \to \mathbb{R}$ est une application alors les trois propositions suivantes sont équivalentes*
(1) X est une variable aléatoire ;
(2) pour tout intervalle I de la forme $I =]-\infty, a]$, $X^{-1}(I)$ appartient à $\mathcal{F}$;
(3) pour tout intervalle I de la forme $]a, b[$, $X^{-1}(I)$ appartient à $\mathcal{F}$.

Preuve. Que (1) implique (2) est évident. Démontrons donc que (2) implique (3). On constate que tout intervalle $]a, b]$ peut s'écrire sous la forme $]-\infty, a]^c \cap]-\infty, b]$ et que $]a, b[= \bigcup_{n \geq 1}]a, b - \frac{1}{n}]$. On a donc (en utilisant le fait que les opérations d'intersection, union, passage au complémentaire se comportent bien par images inverses)

$$X^{-1}(I) = \bigcup_{n \geq 1} \Big(X^{-1}(]-\infty, a]) \Big)^c \cap X^{-1}\Big(\Big] -\infty, b - \frac{1}{n} \Big] \Big).$$

Comme les ensembles apparaissant dans le membre de droite de l'égalité sont des éléments de $\mathcal{F}$ et que $\mathcal{F}$ est une tribu, on en déduit que le membre de gauche est également dans $\mathcal{F}$.

La démonstration du fait que (3) implique (1) se démontre de manière analogue. Démontrons par exemple que pour $a \leq b$, $X^{-1}([a,b])$ est dans $\mathcal{F}$. On observe que $[a,b] = \bigcap_{n\geq 1}]a - \frac{1}{n}, b + \frac{1}{n}[$ et donc

$$X^{-1}([a,b]) = \bigcap_{n\geq 1} X^{-1}\Big(\Big]a - \frac{1}{n}, b + \frac{1}{n}\Big[\Big)$$

est dans $\mathcal{F}$. Les autres cas se traitent de façon analogue. □

Notation Dans la suite du cours, quand X est une v.a. et $A \subset \mathbb{R}$ nous noterons $\{X \in A\}$ ou $[X \in A]$ ou $(X \in A)$ l'ensemble $\{\omega \in \Omega : X(\omega) \in A\}$.

Limites supérieures et inférieures de suites numériques. Un petit rappel concernant les limites supérieures (et inférieures) d'une suite. D'abord, bien sûr, une suite numérique $(u_n)_{n\in\mathbb{N}}$ ne converge pas toujours et la notion de limite est sujet à vérification. Les limites supérieures et inféfieures, par contre, existent toujours : Si on considère $\overline{u}_k := \sup_{n\geq k} u_n$ existe pour tout $k \in \mathbb{N}$, et est une suite décroissante (quand k augmente, on prend le suprémum sur des ensembles $\{n \in \mathbb{N} : n \geq k\}$ qui sont décroissants), par conséquent la limite de $(\overline{u}_k)_{k\in\mathbb{N}}$ est bien définie (potentiellement égale à $-\infty$) et on a $\lim_{k\to\infty} \overline{u}_k = \inf_{k\in\mathbb{N}} \overline{u}_k$ du fait de la décroissance ; on appelle cette quantité la limite supérieure et on la note

$$\limsup_{n\to\infty} u_n := \inf_{k\in\mathbb{N}} \sup_{n\geq k} u_n.$$

Par définition, on a $\limsup_n u_n > \ell$ si quelque soit $k \in \mathbb{N}$, il existe un $n \geq k$ tel que $u_n > \ell$; en d'autres termes, il y a une infinité de $n \in \mathbb{N}$ tels que $u_n > \ell$. De manière symétrique, on définit la limite inférieure par

$$\liminf_{n\to\infty} u_n := \sup_{k\in\mathbb{N}} \inf_{n\geq k} u_n.$$

Par définition, on a $\liminf_n u_n < \ell$ si quelque soit $k \in \mathbb{N}$ il existe un $n \geq k$ tel que $u_n < \ell$; c'est-à-dire que $u_n < \ell$ pour un nombre infini de valeurs de $\mathbb{N}$.

On vérifie aisément (**exercice !**) que la suite (u_n) converge vers une limite finie si et seulement si $\liminf_n u_n = \limsup_n u_n \notin \{-\infty, +\infty\}$.

La proposition qui suit permet de construire des v.a. :

Proposition 3.1.2. *a) Si X_n, $n \geq 1$ est une famille de v.a. alors $Z = sup_{n\geq 1} X_n$ (resp. $Z = \inf_{n\geq 1} X_n$) est une v.a.*
b) Si X_n, $n \geq 1$ est une famille de v.a alors $Z = \limsup_{n\geq 1} X_n$ (resp. $Z = \liminf_{n\geq 1} X_n$). En particulier si pour tout $\omega \in \Omega$, $X_n(\omega)$ converge vers $X(\omega)$ alors X est une v.a.
c) Si $X_1, \ldots, X_n$ sont des v.a et $f : \mathbb{R}^n \to \mathbb{R}$ est une application continue alors $Z = f(X_1, \ldots, X_n)$ est une v.a. En particulier $X_1 + X_2$, λX, ($\lambda \in \mathbb{R}$), $X_1 \cdot X_2$ sont des v.a.

Preuve. a) Soit ω tel que $\sup_{n\geq 1} X_n(\omega) > a$. Alors par définition du sup, il existe un n pour lequel $X_n(\omega) > a$ et ω est donc dans l'union $\cup_{n\geq 1}\{X_n > a\}$. Réciproquement si $\omega \in \cup_{n\geq 1}\{X_n > a\}$ alors il existe n tel que $X_n(\omega) > a$ et *a fortiori* $\sup_n X_n(\omega) > a$. Nous avons donc démontré que les deux ensembles $\{Z > a\}$ et $\cup_{n\geq 1}\{X_n > a\}$ sont égaux. Mais ce dernier ensemble est une union dénombrable d'éléments de la tribu $\mathcal{F}$ (car chaque X_i est une v.a). Ainsi pour tout a l'évévement $\{\sup_{n\geq 1} X_n(\omega) > a\}$ est dans $\mathcal{F}$ et il en est de même de son complémentaire $\{\sup_{n\geq 1} X_n(\omega) \leq a\}$. La proposition 3.1.1 permet de conclure.

b) Par définition

$$\limsup_{n\to\infty} X_n(\omega) = \inf_{p\in\mathbb{N}} \sup_{k\geq p} X_k(\omega)$$

(resp. $\liminf_{n\to\infty} X_n(\omega) = \sup_{p\in\mathbb{N}} \inf_{k\geq p} X_k(\omega)$). Il suffit donc d'appliquer deux fois a). La dernière assertion résulte du fait que $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ si et seulement si $\liminf X_n(\omega) = \limsup X_n(\omega) = X(\omega)$.

c) Si I est un intervalle ouvert de $\mathbb{R}$, $Z^{-1}(I)$ est l'ensemble des $\omega \in \Omega$ tels que $(X_1(\omega), \ldots, X_n(\omega)) \in f^{-1}(I)$. Comme f est continue, $f^{-1}(I)$ est un ensemble *ouvert* de $\mathbb{R}^n$ et, par conséquent, est une union dénombrable de pavés ouverts c'est-à-dire d'ensembles P de la forme $]a_1, b_1[\times \cdots \times]a_n, b_n[$ (nous admettrons ce fait qui n'est par ailleurs pas très difficile à démontrer). Par conséquent l'ensemble des $\omega \in \Omega$ tels que $(X_1(\omega), \ldots, X_n(\omega)) \in f^{-1}(I)$ est une union dénombrable d'ensembles de la forme $\{\omega \in \Omega, (X_1(\omega), \ldots, X_n(\omega) \in]a_1, b_1[\times \cdots \times]a_n, b_n[\}$ c'est-à-dire d'ensembles de la forme $\{\omega \in \Omega, X_1(\omega) \in]a_1, b_1[, \ldots, X_n(\omega) \in]a_n, b_n[\} = X_1^{-1}(]a_1, b_1[) \cap \cdots \cap X_n^{-1}(]a_n, b_n[)$ qui sont clairement dans $\mathcal{F}$. □

En particulier

Proposition 3.1.3. *a) Si $X : \Omega \to \mathbb{R}$ est une v.a. et $f : \mathbb{R} \to \mathbb{R}$ est une application continue, alors la fonction $Y : \Omega \to \mathbb{R}$ définie par $Y = f \circ X$ (c'est-à-dire $Y(\omega) = f(X(\omega))$ pour tout $\omega \in \Omega$) est encore une v.a. On la note $Y = f(X)$.*
b) Si X, Y sont deux v.a. l'application $Z = \max(X, Y)$ est une v.a.
c) Si X et Y sont deux v.a., $aX + bY$ est également une v.a.

Exemple 3.1.2. *Si $A \in \mathcal{F}$, sa fonction indicatrice $\mathbf{1}_A : \Omega \to \mathbb{R}$ (qui prend la valeurs 1 si $\omega \in A$ et 0 sinon) est une v.a.r. En effet :*

$$\mathbf{1}_A^{-1}(I) = \begin{cases} \emptyset & si\ \{0,1\} \cap I = \emptyset \\ A & si\ \{0,1\} \cap I = \{1\} \\ A^c & si\ \{0,1\} \cap I = \{0\} \\ \Omega & si\ \{0,1\} \cap I = \{0,1\}. \end{cases}$$

Réciproquement, si une v.a. ne prend que les valeurs 0 et 1 elle est la fonction indicatrice de $A = \mathbf{1}_A^{-1}(\{1\}) \in \mathcal{F}$. On notera aussi que si $A \notin \mathcal{F}$ alors $\mathbf{1}_A$ n'est pas une variable aléatoire !

Dans le cas où l'ensemble $X(\Omega)$ des valeurs prises par X est un ensemble fini ou dénombrable de $\mathbb{R}$, la caractérisation des variables aléatoires est plus simple.

Définition 3.1.2 (Variable aléatoire discrète)**.** *On dit qu'une variable aléatoire est discrète si elle prends ses valeurs dans un ensemble au plus dénombrable.*

Proposition 3.1.4. *Si $X : \Omega \to \mathbb{R}$ est à valeurs dans un ensemble fini ou dénombrable E alors X est une variable aléatoire si et seulement si pour tout $e \in E$, $X^{-1}(\{e\}) \in \mathcal{F}$.*

Preuve. Supposons que pour tout $e \in E$, $X^{-1}(\{e\}) \in \mathcal{F}$ et soit I un intervalle de $\mathbb{R}$. Comme X est à valeurs dans E, on a

$$X^{-1}(I) = \bigcup_{e \in I \cap E} X^{-1}(\{e\})$$

et comme $E \cap I \subset E$ est dénombrable on déduit que $X^{-1}(I)$ est dans $\mathcal{F}$.

Réciproquement, si $X^{-1}(I)$ est dans $\mathcal{F}$ pour tout intervalle, on a en prenant $I = \{e\}$, $e \in E$ que $X^{-1}(\{e\}) \in \mathcal{F}$. □

Exemple 3.1.3. *Revisitons l'exemple du jeu infini de Pile/Face : $\Omega = \{0,1\}^{\mathbb{N}}$ et $\mathcal{F}$ est la tribu engendrée par les ensembles $C_{i,\epsilon_i} = \{\omega = (\omega_0, \ldots) \in \Omega, \omega_i = \epsilon_i\}$. Pour $n \in \mathbb{N}$ l'application $X_n : \{0,1\}^{\mathbb{N}} \to \{0,1\}$ qui à $\omega = (\omega_0, \omega_1, \ldots)$ associe ω_n est une variable aléatoire. Il suffit en effet de vérifier que pour $\epsilon = 0$ ou $\epsilon = 1$ l'ensemble des ω pour lesquels $\omega_n = \epsilon$ appartient à $\mathcal{F}$. Or, cet ensemble est le cylindre $C_{n,\epsilon}$ qui par définition est dans $\mathcal{F}$. En fait, la tribu $\mathcal{F}$ a été construite de façon que toutes les applications $X_n : \Omega \to \{0,1\}$ ($n \geq 0$) soient des variables aléatoires (c'est d'ailleurs la plus petite tribu ayant cette propriété).*

Exercice 3.1.1. *On considère le jeu infini de Pile/Face $(\Omega, \mathcal{F}, \mathbb{P})$ et on garde les notations de l'exemple précédent. Définissons pour tout $\omega \in \Omega$, l'entier $\nu(\omega)$ comme étant le plus petit entier k pour lequel $X_k(\omega) = 1$ (en d'autres termes $\nu(\omega)$ est le premier temps où on tire Pile). Démontrer que ν est une variable aléatoire.*

Solution. A priori ν est une application de Ω dans $\mathbb{N} \cup \{+\infty\}$; $\mathbb{N} \cup \{+\infty\}$ est dénombrable. Si $\nu^{-1}(\{0\}) = \{\nu = 0\} = \{X_0 = 1\}$ est dans $\mathcal{F}$ car X_0 est une v.a. et pour tout $n \in \mathbb{N}$, si $n \geq 1$ alors

$$\begin{aligned}\nu^{-1}(\{n\}) = \{\omega \in \Omega : \nu(\omega) = n\} &= \{\omega \in \Omega : \omega_0 = \cdots = \omega_{n-1} = 0, \omega_n = 1\} \\ &= \{X_0 = 0\} \cap \cdots \cap \{X_{n-1} = 0\} \cap \{X_n = 1\};\end{aligned}$$

c'est une intersection finie déléments de $\mathcal{F}$ (car, puisque chaque X_i est une v.a., les ensembles $\{X_i = 0\}$, $1 \leq i \leq n-1$ et $\{X_n = 1\}^c$ sont dans $\mathcal{F}$). De plus,

$$\nu^{-1}(+\infty) = \{\omega \in \Omega : \omega_n = 0, n \in \mathbb{N}\} = \bigcap_{n \in \mathbb{N}} \{X_n = 0\},$$

qui est dans $\mathcal{F}$ comme intersection dénombrable d'éléments de $\mathcal{F}$. Par conséquent, $\nu^{-1}(\{e\}) \in \mathcal{F}$ quelque soit $e \in \mathbb{N} \cup \{+\infty\}$ et ν est bien une variable aléatoire. △

3.1.2 Loi d'une variable aléatoire réelle

Soit $X : \Omega \to \mathbb{R}$ une v.a. On sait que pour tout intervalle I de $\mathbb{R}$ l'ensemble $X^{-1}(I)$ est un évévement (appartient à $\mathcal{F}$). Il est donc possible de parler de la probabilité $\mathbb{P}(X \in I)$ de l'évévement $\{X \in I\}$.

Définition 3.1.3. *(Loi d'une variable aléatoire) Soit X une v.a.r. L'application qui à tout intervalle I de $\mathbb{R}$ associe $\mathbb{P}(X \in I)$ s'appelle la loi de X.*

Compléments En fait la loi d'une v.a.r. a une signification plus profonde. Appelons $\mathcal{B}$ le sous-ensemble de $\mathcal{P}(\mathbb{R})$ défini comme étant la tribu engendrée par les intervalles de $\mathbb{R}$. Cette tribu s'appelle la *tribu borélienne* (ou de Borel) ; les éléments de la tribu borélienne s'appellent les boréliens de $\mathbb{R}$. On a alors le résultat suivant :

Proposition 3.1.5. *Si $X : \Omega \to \mathbb{R}$ est une v.a.r, alors pour tout borélien $A \in \mathcal{B}$, $X^{-1}(A) \in \mathcal{F}$.*

Preuve. Si on note $\mathcal{C}$ l'ensemble des $A \subset \mathbb{R}$ tels que $X^{-1}(A) \in \mathcal{F}$, il est facile de voir que $\mathcal{C}$ est une tribu ; par ailleurs, il est clair que $\mathcal{C}$ contient les intervalles. Comme la tribu borélienne est la tribu engendrée par les intervalles, c'est-à-dire la plus petite tribu de $\mathbb{R}$ contenant les intervalles, on a $\mathcal{B} \subset \mathcal{C}$. Mais alors, pour tout borélien A de $\mathbb{R}$, $X^{-1}(A) \in \mathcal{F}$. □

La proposition suivante montre que toute v.a.r. définit naturellement une probabilité sur $\mathbb{R}$ muni de sa tribu borélienne.

Proposition 3.1.6. *L'application $\mu_X : \mathcal{B} \to [0,1]$ qui à tout borélien de $\mathbb{R}$ associe le réel $P(X \in A)$ de $[0,1]$ est une probabilité sur $(\mathbb{R}, \mathcal{B})$. On appelle cette probabilité la* loi *de la v.a. X.*

Preuve. Il suffit de démontrer que si $(A_i)_{i \in \mathbb{N}}$ est une famille dénombrable de boréliens de $\mathbb{R}$ disjoints deux à deux alors

$$\mathbb{P}\Big(X \in \bigcup_{i \in \mathbb{N}} A_i\Big) = \sum_{i=0}^{\infty} P(X \in A_i),$$

ce qui est clair car l'évévement $\{X \in \cup_{i \in \mathbb{N}} A_i\}$ est l'union dénombrable *disjointe* des évévements $\{X \in A_i\}$. Enfin la condition $P(X \in \mathbb{R}) = 1$ achève la preuve. □

Il faut retenir que la loi d'une v.a est une *probabilité* sur $\mathbb{R}$ (muni de sa tribu borélienne). Ceci illustre le fait qu'il est possible de construire de nombreuses mesures de probablilités sur $\mathbb{R}$ muni de sa tribu borélienne.

Définition 3.1.4 (Fonction de répartition)**.** *Si X est une v.a.r. on définit sa fonction de répartition $F_X : \mathbb{R} \to [0,1]$ par*

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x]).$$

Voici quelques propriétés des fonctions de répartition.

Proposition 3.1.7. *Si $F_X : \mathbb{R} \to [0,1]$ est la fonction de répartition d'une v.a.r. X*

(i) la fonction $F_X : \mathbb{R} \to [0,1]$ est croissante : si $x_1 \leq x_2$ alors $F_X(x_1) \leq F_X(x_2)$.
(ii) on a $\lim_{x\to-\infty} F_X(x) = 0$ et $\lim_{x\to\infty} F_X(x) = 1$.
(iii) F_X est continue à droite en tout point.

Preuve. (i) La croissance est évidente : si $x_1 \leq x_2$ alors $(-\infty, x_1] \subseteq (-\infty, x_2]$ et par conséquent, on a $\{X \in (-\infty, x_1]\} \subseteq \{X \in (-\infty, x_x]\}$, ce qui implique $F_X(x_1) \leq F_X(x_2)$.
(ii) La suite d'intervalles $(-\infty, x]$, $x \in \mathbb{R}$, tend en croissant vers $\mathbb{R}$ lorsque $x \to \infty$, donc $\{X \leq x\} \uparrow \{X \in \mathbb{R}\}$ et il s'ensuit que $F_X(x) = \mathbb{P}(X \in (-\infty, x])$ tend en croissant[1] vers $\mathbb{P}(X \in \mathbb{R}) = 1$. De manière similaire, comme la suite $(-\infty, -x]$, $x \in \mathbb{R}$, decroit vers $\emptyset$, on obtient $F_X(x) \to \mathbb{P}(X \in \emptyset) = 0$.
(iii) Pour $\epsilon \downarrow 0$ (c'est-à-dire $\epsilon \to 0$ par valeurs positives), on a $F_X(x+\epsilon) = \mathbb{P}(X \in (-\infty, x+\epsilon)) \downarrow \mathbb{P}(X \in (-\infty, x]) = F_X(x)$ car $\{X \leq x+\epsilon\} \downarrow \{X \leq x\}$. □

Exemple 3.1.4. *Si X est une v.a.r. prenant seulement deux valeurs 0 ou 1 et $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$ (X est une v.a. de Bernoulli de paramètre p) alors sa fonction de répartition $F_X(t) = \mathbb{P}(X \leq t)$ vaut 0 si $x < 0$, $1 - p$ si $0 \leq x < 1$ et 1 si $x \geq 1$.*

L'importance de la fonction de répartition tient à cette proposition :

Proposition 3.1.8. *La fonction de répartition caractérise la loi ; en d'autres termes, deux variables aléatoires X et Y on la même fonction de répartition ($F_X(t) = F_Y(t)$) si et seulement si X et Y ont même loi.*

Preuve. D'abord, si X et Y on même loi, alors F_X et F_Y sont clairement égales puisque, quelque soit $s \in \mathbb{R}$, $F_X(s) = \mathbb{P}(X \leq s) = \mathbb{P}(X \in (-\infty, s]) = \mathbb{P}(Y \in (-\infty, s]) = F_Y(s)$, où on a utilisé l'égalité des lois avec l'intervalle $I = (-\infty, s]$.

Réciproquement, supposons que $F_X = F_Y$. Si I est un intervalle de la forme $(\infty, b]$, alors on a $\mathbb{P}(X \in I) = F_X(b) = F_Y(b) = \mathbb{P}(Y \in I)$. Si I est de la forme $(a, b]$ alors on a $\mathbb{P}(X \in I) = F_X(b) - F_X(a) = F_Y(b) - F_Y(a) = \mathbb{P}(Y \in I)$. Pour les intervalles de la forme $[a, b]$ on utilise les propriétés de continuité la probabilité : comme $[a, b] = \cap_{n\geq 0}(a - 1/n, b]$ est une suite décroissante on a

$$\mathbb{P}(X \in [a, b]) = \lim_{n\to\infty} \mathbb{P}(X \in (a - 1/n, b]) = \lim_{n\to\infty} \mathbb{P}(Y \in (a - 1/n, b]) = \mathbb{P}(Y \in [a, b]).$$

Les autres cas se traitent de la même manière,s par continuité. □

Remarque. Attention : il faut bien comprendre que deux variables aléatoires X et Y peuvent avoir la même loi sans pour autant être égales. (Ici, comme X est Y sont des fonctions, $X = Y$ signifie que $X(\omega) = Y(\omega)$ quelque soit $\omega = \Omega$.) En fait, elles peuvent même être toujours différentes tout en ayant la même loi : par exemple, soit X à valeurs dans $\{-1, 1\}$ telle que $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ (ce qu'on peut réaliser, par exemple, avec $\Omega = \{-1, 1\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}$ la probabilité uniforme sur Ω et $X : \Omega \to \Omega$ l'identité, telle que $X(\omega) = \omega$). Si on définit $Y = -X$, c'est-à-dire, $Y(\omega) = -X(\omega)$ pour tout $\omega \in \Omega$ alors quelque soit $\omega \in \Omega$, on a $X(\omega) \neq Y(\omega)$. Cependant, X et Y ont même loi puisque $\mathbb{P}(Y = 1) = \mathbb{P}(X = -1) = 1/2 = \mathbb{P}(Y = -1)$.

3.2 Variables aléatoires discrètes

3.2.1 Loi d'une v.a. discrète et exemples

Proposition 3.2.1. *Si X est une variable aléatoire discrète (à valeur dans $E \subseteq \mathbb{R}$ dénombrable), alors la loi de X est complètement déterminée par la famille (dénombrable) de réels $(\mathbb{P}(X = e) : e \in E)$.*

1. Formellement, il faudrait prendre des familles croissantes dénombrables (cf. la propostion 2.3.2) donc n'importe quelle suite croissante $x_n \to \infty$; ici, la monotonie de F_X nous dit que ce n'est pas nécessaire et que la limite ne dépend pas de la suite

Preuve. Il suffit de montrer que pour tout intervalle I de $\mathbb{R}$, $\mathbb{P}(X \in I)$ est déterminée par $(\mathbb{P}(X = e) : e \in E)$. Soit I un intervalle. Comme E est dénombrable, $I \cap E$ est dénombrable. De plus, comme X est à valeurs dans E,

$$\mathbb{P}(X \in I) = \mathbb{P}(X \in I \cap E) = \mathbb{P}\Big(X \in \bigsqcup_{e \in E \cap I} \{e\}\Big) = \sum_{e \in E \cap I} \mathbb{P}(X = e),$$

puisque $\sqcup_{e \in E \cap I}\{e\}$ est une union dénombrable d'ensembles deux-à-deux disjoints. □

Remarque. Une conséquence importante : en pratique, calculer la loi d'une v.a. discrète revient à calculer les $\mathbb{P}(X = e)$ pour $e \in E = X(\Omega)$ l'ensemble des valeurs prises par X.

Compléments. Dans ce cas, la loi μ_X de X est la probabilité sur $(\mathbb{R}, \mathcal{B})$ définie par

$$\mu_X = \sum_{e \in E} \mathbb{P}(X = e)\delta_e$$

où δ_e est la mesure de Dirac (pour $A \in \mathcal{B}$, $\delta_e(A) = 1$ si $e \in A$ et 0 sinon). En effet pour tout borélien (ou tout intervalle) A

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A \cap E) = \sum_{e \in A \cap E} \mathbb{P}(X = e),$$

et cette somme n'est rien d'autre que

$$\sum_{e \in E} \mathbb{P}(X = e)\delta_{\mathbb{E}}(A).$$

Loi de Bernoulli C'est la loi d'une v.a.r prenant pour valeur 0 ou 1 et $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$. On dit que $p \in [0, 1]$ est le paramètre de la loi. La fonction de répartition de X est donnée par

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - p & \text{si } x \in [0, 1) \\ 1 & \text{si } x \geq 1. \end{cases}$$

Loi géométrique On dit qu'une v.a $X : \Omega \to \mathbb{N}$ (à valeurs dans $\mathbb{N}$) suit une loi géométrique de paramètre a $(0 < a < 1)$ si

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

On remarquera que l'on a bien $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$ puisque $\sum_{k \geq 0} a^k = 1/(1 - a)$.

La v.a ν de l'exercice 3.1.1 suit une loi géométrique de paramètre 1/2. En effet

$$\begin{aligned} \nu^{-1}(n) &= \{\omega = (\omega_0, \omega_1, \ldots) \in \Omega, \omega_0 = 0, \ldots, \omega_{n-1} = 0, \omega_n = 1\} \\ &= C_{0,0} \cap \cdots \cap C_{n-1,0} \cap C_{n,1} \end{aligned}$$

et d'après le théorème 2.6.1, $\mathbb{P}(\nu = n) = 2^{-n} \cdot \frac{1}{2}$.

Loi binomiale On dit qu'une variable aléatoire Z à valeurs dans $\{0, \ldots, n\}$ suit une loi binomiale (n, p) si

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

où $\binom{n}{k}$ est le coefficient binomial

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

On a bien (formule du binôme de Newton) $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = (p + 1 - p)^n = 1$.

Exemple 3.2.1. *Jouons n fois au jeu de pile/face où pile sort avec probabilité p et face avec probabilité $1-p$ et notons Z la variable aléatoire : Z est le nombre de pile qui sortent (après avoir joué n fois). Si on note X_i les variables aléatoires $X_i(\omega) = \omega_i$ ($\omega = (\omega_1, \dots, \omega_n)$) on a*

$$Z = X_1 + \cdots + X_n.$$

C'est bien une variable aléatoire $Z : \mathcal{F} \to \mathbb{N}$ à valeurs dans l'ensemble fini $\{0, \dots, n\}$ ($\mathcal{F} = \mathbb{P}(\Omega)$) et

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

La v.a. Z suit donc une loi binomiale (n, p).

Loi de Poisson Une variable aléatoire $Z : \Omega \to \mathbb{N}$ suit une loi de Poisson de paramètre $\lambda > 0$ si

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

On vérifie encore que $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$ (cf. le développement en série de e^{λ}).

Exercice 3.2.1. *Soit X_n une v.a suivant une loi binomiale (n, p_n). Montrer que si $\lim_{n\to\infty} np_n = \lambda$ on a pour tout $k \in \mathbb{N}$*

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

(On dit que X_n converge en loi vers une loi de Poisson de paramètre λ)

Solution. On a, par définition, tout $k \in \mathbb{N}$ et pour $n \geq k$,

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n!}{(n-k)!k!} \cdot \frac{\lambda^k}{n^k} \cdot (1-p_n)^{n-k}.$$

Cependant, comme $p_n \sim \lambda/n$ lorsque $n \to \infty$, puisque $\ln(1-x) \sim x$ lorsque $x \to 0$, on a

$$(1-p_n)^{n-k} = \exp((n-k)\ln(1-p_n)) \sim \exp(-(n-k)p_n) \sim \exp(-(n-k)\frac{\lambda}{n}) \sim e^{-\lambda},$$

et

$$\frac{n!}{(n-k)!n^k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} = 1 \times \left(1 - \frac{1}{n}\right) \times \cdots \times \left(1 - \frac{k-1}{n}\right) \to 1,$$

puisque k est fixé que que chacun des termes tendent vers 1. Ceci prouve bien que

$$\mathbb{P}(X_n = k) \to \frac{\lambda^k}{k!} e^{-\lambda},$$

lorsque $n \to \infty$. △

3.2.2 Espérance d'une v.a. discrète

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire ne prenant qu'un nombre fini de valeurs $x_1, \dots x_r$. On définit l'espérance de X comme le barycentre des x_i, $1 \leq i \leq r$, pondérées par les probabilités $\mathbb{P}(X = x_i)$, c'est-à-dire ble nombre réel

$$\mathbb{E}(X) = \sum_{i=1}^{r} x_i \cdot \mathbb{P}(X = x_i).$$

Remarquons que si X prend ses valeurs dans un ensemble infini dénombrable la quantité

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i \cdot \mathbb{P}(X = x_i),$$

qui semble être un bon candidat pour la définition de l'espérance peut ne pas exister car la série peut ne pas converger. Pour garantir cette convergence il suffit de demander que la série précédente soit absolument convergente.

On adoptera donc la définition suivante :

Définition 3.2.1. *Si X est une v.a.r. prenant un nombre fini ou dénombrable de valeurs dans $E \subset \mathbb{R}$, telle que*

$$\sum_{x \in E} |x| \mathbb{P}(X = x) < \infty \tag{3.1}$$

on définit l'espérance de X comme étant le nombre réel

$$\mathbb{E}(X) = \sum_{x \in E} x \mathbb{P}(X = x).$$

Quand la condition (3.1) est vérifiée on dit que X est dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

Remarque. 1) Quand X est positive, et dans ce cas seulement, si la série $\sum_{x \in E} |x| \mathbb{P}(X = x) = \infty$ diverge on pose $\mathbb{E}(X) = \infty$.
2) L'espérance représente intuitivement la *valeur moyenne* prise par la variable aléatoire X.

Exemple 3.2.2. *On considère U une variable aléatoire uniforme sur $\{1, 2, \ldots, n\}$. Alors $\mathbb{P}(U = i) = 1/n$ pour tout $i \in \{1, 2, \ldots, n\}$ et*

$$\mathbb{E}U = \sum_{i=1}^{n} i \times \frac{1}{n} = \frac{n(n+1)}{2} \cdot \frac{1}{n} = \frac{n+1}{2}.$$

Exemple 3.2.3. *Soit N une variable aléatoire binomiale (n, p). Alors, $\mathbb{E}N$ existe puisque N ne prend qu'un nombre fini de valeurs. De plus, en posant $k - 1 = i$ à la 5e ligne et en utilisant le binome de Newton pour resommer le développement de $(1 + (1 - p))^{n-1}$, on obtient :*

$$\begin{aligned}
\mathbb{E}N &= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \cdot \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \\
&= np(1 - p + p)^{n-1} \\
&= np.
\end{aligned}$$

Théorème 3.2.1. *L'espérance des v.a.r. prenant un nombre fini ou dénombrable de valeurs vérifie les conditions suivantes :*
a) si $A \in \mathcal{F}$ on a $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$.
b) Si X, Y sont des v.a.r. positives prenant un nombre fini ou dénombrable de valeurs telles que $X \leq Y$ (c'est-à-dire pour tout $\omega \in \Omega$ $X(\omega) \leq Y(\omega)$) alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

c) Si X, Y sont des v.a dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ prenant un nombre fini ou dénombrable de valeurs et $a, b \in \mathbb{R}$ on a (linéarité de l'espérance)

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

d) Si X est une v.a positive prenant un nombre fini ou dénombrable de valeurs telle que $\mathbb{E}(X) = 0$ alors X est nulle $\mathbb{P}$-presque sûrement c'est-à-dire que l'ensemble des $\omega \in \Omega$ pour lesquels $X(\omega) > 0$ a une probabilité nulle.

Remarque. De manière générale, si l'ensemble des $\omega \in \Omega$ qui ne satisfont pas une propriété $\mathcal{P}$ est de probabilité nulle, on dit que la propriété $\mathcal{P}$ est vraie $\mathbb{P}$-presque sûrement, ou presque sûrement quand il n'y a pas d'ambiguité sur $\mathbb{P}$.

Preuve. Le point a) est immédiat. Démontrons le point b). Notons $(x_i)_i$ l'ensemble des valeurs prises par X et $(y_j)_j$ celles prises par Y. Définissons $A_i = X^{-1}(x_i)$, $B_j = Y^{-1}(y_j)$ qui sont dans la tribu des événements $\mathcal{F}$. Les $(A_i)_i$ constituent une partition de Ω tout comme les $(B_j)_j$. Par conséquent les $(A_i \cap B_j)_{i,j}$ constituent également une partition de Ω. Si $A_i \cap B_j \neq \emptyset$ alors $x_i \leq y_j$ car $X(\omega) \leq Y(\omega)$ pour $\omega \in A_i \cap B_j$; sinon $\mathbb{P}(A_i \cap B_j) = 0$. Donc dans tous les cas $\mathbb{P}(A_i \cap B_j)x_i \leq \mathbb{P}(A_i \cap B_j)y_j$. A présent écrivons

$$\begin{aligned}
\mathbb{E}(X) &= \sum_i \mathbb{P}(A_i)x_i \\
&= \sum_{i,j} \mathbb{P}(A_i \cap B_j)x_i \\
&\leq \sum_{i,j} \mathbb{P}(A_i \cap B_j)y_j \\
&\leq \sum_j \mathbb{P}(B_j)y_j = \mathbb{E}(Y).
\end{aligned}$$

Démontrons maintenant le point c). En utilisant la décomposition précédente on voit que

$$\begin{aligned}
\mathbb{E}(aX + bY) &= \sum_{i,j} \mathbb{P}(A_i \cap B_j)(ax_i + by_j) \\
&= a\sum_{i,j} \mathbb{P}(A_i \cap B_j)x_i + b\sum_{i,j} \mathbb{P}(A_i \cap B_j)y_j \\
&= a\sum_i \mathbb{P}(A_i)x_i + b\sum_{i,j} \mathbb{P}(B_j)y_j \\
&= a\mathbb{E}(X) + b\mathbb{E}(Y).
\end{aligned}$$

Enfin, pour le point d) nous remarquons (avec les notations précédentes) que si $\mathbb{E}(X) = 0$ on a $\sum_i \mathbb{P}(A_i)x_i = 0$; mais comme X est positive, chacun des termes de la somme précédentes est positif ou nulle et donc pour tout i, $\mathbb{P}(A_i)x_i = 0$. Cela démontre que pour tout i pour lequel $x_i > 0$ on doit avoir $\mathbb{P}(A_i) = 0$. Ainsi $\mathbb{P}(\bigcup_{i, x_i \neq 0} A_i) = 0$. Cela démontre le point d. □

3.2.3 Formule de transfert

Il est important dans la pratique de savoir caluler des espérances de v.a aléatoires de la forme $Y = f(X)$ où $X : \Omega \to \mathbb{R}$ est une v.a et $f : \mathbb{R} \to \mathbb{R}$ est une fonction (disons continue, mais pour le cas des variables discrètes ça a peu d'importance).

Proposition 3.2.2. *Si X prend un nombre fini ou dénombrable de valeurs dans $E \subset \mathbb{R}$ et si*

$$\sum_{e \in E} |f(e)|\mathbb{P}(X = e) < \infty$$

alors l'espérance de la variable aléatoire $Y = f(X)$ est donnée par

$$\mathbb{E}(f(X)) = \sum_{e \in E} f(e)\mathbb{P}(X = e).$$

Preuve. Supposons que $X : \Omega \to \mathbb{R}$ prenne ses valeurs dans un ensemble fini $E \subset \mathbb{R}$ et soit $f : E \to \mathbb{R}$. Comme $Y = f(X)$, la v.a Y ne prend qu'un nombre fini de valeurs e' qui sont dans $E' = f(E)$. Par définition de l'espérance

$$\mathbb{E}(Y) = \sum_{e' \in E'} e'\mathbb{P}(Y = e').$$

Pour chaque $e' \in E'$, notons $A_{e'}$ l'ensemble des $e \in E$ tels que $f(e) = e'$ ($A_{e'} = f^{-1}(\{e'\})$ et constatons que E est l'union disjointe des $A_{e'}$, $e' \in E'$. On a donc,

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{e' \in E'} e'\mathbb{P}(Y = e') \\
&= \sum_{e' \in E'} e'\mathbb{P}(f(X) = e') \\
&= \sum_{e' \in E'} e'\mathbb{P}(X \in f^{-1}(e')) \\
&= \sum_{e' \in E'} e' \sum_{e \in f^{-1}(e')} \mathbb{P}(X = e) \\
&= \sum_{e' \in E'} \sum_{e \in f^{-1}(e')} f(e)\mathbb{P}(X = e) \\
&= \sum_{e \in E} f(e)\mathbb{P}(X = e) \\
&= \mathbb{E}(X).
\end{aligned}$$

Un examen de la preuve que nous venons de présenter montre qu'elle s'étend au cas où X prend un nombre infini dénombrable de valeurs à condition que l'hypothèse de la proposition soit vérifiée. □

Exercice 3.2.2. *Soit N une variable aléatoire binomiale (n, p). Calculer $\mathbb{E}[N(N-1)]$.*

Solution. Si on veut utiliser la définition de l'espérance alors il faut calculer la somme

$$\sum_{k \in \mathbb{Z}} k\mathbb{P}(N(N-1) = k),$$

ce qui ne paraît pas une bonne idée puisque les pondérations $\mathbb{P}(N(N-1) = k)$ semblent difficiles à calculer. On utilise donc la formule de transfert de manière à sommer sur les valeurs prises par N et non celles prises par $N(N-1)$. En utilisant les même technique que pour calculer $\mathbb{E}N$, on

pose $i = k - 2$

$$\begin{aligned}
\mathbb{E}[N(N-1)] &= \sum_{k=0}^{n} k(k-1) \times \binom{n}{k} p^k (1-p)^k \\
&= \sum_{k=2}^{n} k(k-1) \times \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} n(n-1) \frac{(n-2)!}{(n-k)!(k-2)!} p^{k-2} (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(n-k)!(k-2)!} p^{k-2} (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{i=0}^{n-2} \binom{n-2}{i} p^i (1-p)^{n-i},
\end{aligned}$$

et en resommant par le binome de Newton on obtient $\mathbb{E}[N(N-1)] = n(n-1)p^2$. △

3.3 Espérance d'une variable aléatoire : cas général

3.3.1 Espérance

En s'inspirant de la définition de l'espérance que nous avons exposée dans la section précédente pour les v.a ne prenant qu'un nombre fini ou dénombrable de valeurs, il est possible de définir, sous certaines conditions, l'espérance de v.a.r. en général. La construction de l'espérance dans le cas général est cependant un peu fastidieuse et nous la laisserons donc de côté.

Théorème 3.3.1 (Espérance : cas positif)**.** *A toute variable aléatoire réelle* **positive** X *il est possible d'associer un élément de* $[0, \infty[\cup\{\infty\}$ *que l'on appelle l'espérance de* X *et que l'on note* $\mathbb{E}(X)$ *et qui vérifie les propriétés suivantes :*
a) si $A \in \mathcal{F}$ *on a* $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$.
b) Si X, Y *sont des v.a positives telle que* $X \leq Y$ *(c'est-à-dire pour tout* $\omega \in \Omega$ $X(\omega) \leq Y(\omega)$*) alors* $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
c) Si X, Y *sont des v.a positives et* $a, b \in \mathbb{R}$ *on a (linéarité de l'espérance)*

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

d) Si X *est une v.a positive telle que* $\mathbb{E}(X) = 0$ *alors* X *est nulle* $\mathbb{P}$*-presque sûrement c'est-à-dire que l'ensemble des* $\omega \in \Omega$ *pour lesquels* $X(\omega) > 0$ *a une probabilité nulle.*

Remarque. 1) Notons que $\mathbb{E}(a) = a$ si a est une constante.
2) Il est facile de voir que les conditions précédentes imposent que l'espérance coïncide avec celle que nous avons définie pour les v.a. prenant un nombre fini ou dénombrable de valeurs.

Remarque On dit qu'une propriété P_ω qui dépend de $\omega \in \Omega$ est vraie $\mathbb{P}$-presque sûrement si l'ensemble des $\omega \in \Omega$ pour lesquels P_ω est fausse est de $\mathbb{P}$-probabilité nulle ($\mathbb{P}(\{\omega : P_\omega \quad \text{fausse}\}) = 0$). On suppose ici implicitement que $\{\omega \in \Omega : P_\omega\}$ est un événement pour que sa probabilité existe.

Définition 3.3.1. *On dit qu'une v.a.r. est dans* $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ *(on note aussi* $\mathcal{L}^1(\Omega, \mathbb{P})$ *ou* $\mathcal{L}^1(\mathbb{P})$ *ou simplement* $\mathcal{L}^1$*) si l'espérance de la v.a.r.* **positive** $|X|$ *est finie :* $\mathbb{E}(|X|) < \infty$*. Dans ce cas on dit que la v.a.r.* X *est (*$\mathbb{P}$*-)intégrable.*

Si on pose $X_+ = \max(0, X)$ et $X_- = -\min(0, X)$, on a $X_+ \geq 0$, $X_- \geq 0$

$$X = X_+ - X_-$$

et

$$|X| = X_+ + X_-,$$

et en particulier

$$0 \leq X_+ \leq |X|, \qquad 0 \leq X_- \leq |X|.$$

Donc si $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ on a aussi $X_+, X_- \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 3.3.2. *Si $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ on a pose*

$$\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_-).$$

Théorème 3.3.2 (Espérance : cas $\mathcal{L}^1$)**.** *a) si $A \in \mathcal{F}$ on a $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$.*
b) Si X, Y sont des v.a dans $\mathcal{L}^1(\Omega, \mathbb{P})$ telles que $X \leq Y$ (c'est-à-dire pour tout $\omega \in \Omega$ $X(\omega) \leq Y(\omega)$) alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
c) Si X, Y sont des v.a dans $\mathcal{L}^1(\Omega, \mathbb{P})$ et $a, b \in \mathbb{R}$ on a (linéarité de l'espérance)

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

d) Si $X \in \mathcal{L}^1(\Omega, \mathcal{F})$ on a toujours $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$ et on a égalité si et seulement si $\mathbb{P}$-ps $X \geq 0$ ou $\mathbb{P}$-ps $X \leq 0$.

3.3.2 Variables aléatoires de carré intégrable

Commençons par une définition.

Définition 3.3.3. *On dit qu'une v.a. X est dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ si son carré est $\mathbb{P}$-intégrable,*

$$\mathbb{E}(|X|^2) < \infty.$$

Une des propriétés importantes de ces espaces $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ est qu'ils sont stables par combinaisons linéaires :

Théorème 3.3.3. *L'espace $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ est un $\mathbb{R}$-espace vectoriel c'est-à-dire que si $a, b \in \mathbb{R}$ et $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ on a aussi $aX + bY \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. En outre, la propriété suivante est toujours vérifiée (inégalité de Minkowski) :*

$$\mathbb{E}(|X + Y|^2)^{1/2} \leq \mathbb{E}(|X|^2)^{1/2} + \mathbb{E}(|Y|^2)^{1/2}.$$

Une propriété très utile des espaces $\mathcal{L}^2$ est la propriété de Cauchy–Schwarz (qui permet entre autres choses de démontrer l'inégalité de Minkowski).

Théorème 3.3.4 (Cauchy–Schwarz)**.** *Si X et Y sont des v.a.r dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ alors le produit XY est dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ et on a*

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|X|^2)^{1/2}\mathbb{E}(|Y|^2)^{1/2}$$

avec égalité si et seulement si X et Y sont colinéaires.

Preuve. L'application $\langle \cdot | \cdot \rangle : \mathcal{L}^1 \times \mathcal{L}^1 \to \mathbb{R}$ définie par $\langle X|Y \rangle \mapsto \mathbb{E}(XY)$ est bilinéaire symétrique et positive $\langle X|X \rangle = \mathbb{E}(X^2) \geq 0$. Par conséquent, quelque soit $X, Y \in \mathcal{L}^1$ et $\lambda \in \mathbb{R}$,

$$0 \leq \langle X + \lambda Y | X + \lambda Y \rangle = \langle X|X \rangle + 2\lambda \langle X|Y \rangle + \lambda^2 \langle Y|Y \rangle.$$

Par conséquent le membre de droite, qui est un polynôme de $\mathbb{R}[\lambda]$ de degré deux, est positif pour toute valeur de λ, et son discriminant est donc négatif ou nul :

$$4\langle X|Y \rangle^2 - 4\langle X|X \rangle \cdot \langle Y|Y \rangle \leq 0,$$

ce qui implique $\langle X|Y \rangle^2 \leq \langle X|X \rangle \cdot \langle Y|Y \rangle$, c'est-à-dire ce qu'il fallait montrer. □

Proposition 3.3.1. *Si X est une v.a. dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ alors X est aussi dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.*

Preuve. Par l'inégalité de Cauchy–Schwarz,

$$\mathbb{E}(|X|) = \mathbb{E}(|X| \times 1) \leq \mathbb{E}(|X|^2)^{1/2} \cdot \mathbb{E}(1^2)^{1/2} = \mathbb{E}(|X|^2)^{1/2} < \infty. \qquad \square$$

3.3.3 Variance

Nous pouvons à présent définir la variance d'une v.a.r. de carré intégrable. Si X est dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, alors par la proposition 3.3.1 X est aussi dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathbb{E}(X) < \infty$ est donc bien définie. De plus, la v.a $X - \mathbb{E}(X)$ est également dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ par le théorème 3.3.3 puisque c'est une somme de deux v.a de $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ (une v.a constante est toujours dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$).

Définition 3.3.4. *La variance d'une v.a.r. dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ est la quantité*

$$\mathbf{Var} X := \mathbb{E}(|X - \mathbb{E}(X)|^2).$$

La racine carrée σ de ce nombre s'appelle l'écart-type de X.

La variance est donc la moyenne (l'espérance) des carrés des écarts de X par rapport à $\mathbb{E}(X)$. Elle mesure en quelque sorte le caractère plus ou moins diffus de la variable aléatoire X.

Proposition 3.3.2. *Si $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, on a $\mathbf{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.*

Preuve. If suffit d'écrire

$$\begin{aligned}\mathbf{Var}(X) = \mathbb{E}(|X - \mathbb{E}(X)|^2) &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \qquad \square\end{aligned}$$

Exemple 3.3.1. *Soit X une variable aléatoire de loi de Bernoulli de paramètre $p \in]0, 1[$. Alors,*

$$\mathbf{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2 = p(1 - p).$$

Exemple 3.3.2. *Soit N une variable aléatoire de loi binomiale de paramètre $n \geq 1$ et $p \in]0, 1[$. Alors, par la proposition 3.3.2 et la linéarité de l'espérance,*

$$\begin{aligned}\mathbf{Var}(N) = \mathbb{E}(N^2) - \mathbb{E}(N)^2 &= \mathbb{E}(N(N-1)) + \mathbb{E}(N) - \mathbb{E}(N)^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= np(1-p),\end{aligned}$$

où la deuxième ligne découle du calcul de $\mathbb{E}(N(N-1))$ de l'exercice 3.2.2. On observe que $\mathbf{Var}(N) = n\mathbf{Var}(X)$ où X est Bernoulli de paramètre p. Nous verrons plus tard que ce n'est pas un hasard.

Exemple 3.3.3. *Soit X une variable aléatoire uniforme sur $\{1, 2, \ldots, n\}$. Alors $\mathbb{P}(X = i) = 1/n$ pour tout $i \in \{1, 2, \ldots, n\}$ et*

$$\mathbb{E}(X^2) = \sum_{i=1}^{n} i^2 \mathbb{P}(X = i) = \frac{1}{n} \sum_{i=1}^{n} i^2 = \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.$$

et $\mathbb{E}(X) = (n+1)/2$, par conséquent

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

Remarque. De façon plus générale on peut définir pour tout $p \geq 1$ l'espace $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ des v.a X dont le *moment d'ordre p* est fini c'est-à-dire telles que $\mathbb{E}(|X|^p) < \infty$. Ce sont encore des espaces vectoriels et la quantité $\mathbb{E}(|X|^p)^{1/p}$ définit une (semi-)norme sur $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ qui en fait un espace de Banach (un espace vectoriel normé complet).

3.3.4 Inégalité de Bienaymé–Tchebychev

L'intérêt de considérer les moments d'une v.a X réside dans les propositions suivantes qui quantifient, en fonction des moments la probabilité qu'une variable aléatoire soit grande ou loin de sa moyenne.

Proposition 3.3.3 (Inégalité de Markov)**.** *Si X est une v.a dans $\mathcal{L}^1(\Omega, \mathbb{P})$ on a pour tout $\lambda > 0$*

$$\mathbb{P}(|X| > \lambda) \leq \frac{\mathbb{E}(|X|)}{\lambda}.$$

Preuve. La v.a $|X|$ peut sécrire

$$|X| = |X| \cdot \mathbf{1}_{\{|X|>\lambda\}} + |X| \cdot \mathbf{1}_{\{|X|\leq\lambda\}},$$

et par additivité et positivité de l'espérance on a

$$\mathbb{E}(|X|) \geq \mathbb{E}(|X| \cdot \mathbf{1}_{\{|X|>\lambda\}}).$$

Or,

$$|X| \cdot \mathbf{1}_{\{|X|>\lambda\}} \geq \lambda \cdot \mathbf{1}_{\{|X|>\lambda\}},$$

et par conséquent

$$\mathbb{E}(|X|) \geq \lambda \mathbb{E}(\mathbf{1}_{\{|X|>\lambda\}}),$$

c'est-à-dire

$$\mathbb{E}(|X|) \geq \lambda \mathbb{P}(\{|X|\rangle\lambda\},$$

ce qui est la conclusion de la proposition. □

La proposition précédente est une version quantitative du fait que la probabilité que X prenne de grandes valeurs a tendance à être petite. Si on a des informations sur les moments d'ordre supérieurs l'estimation précédente est meilleure.

Proposition 3.3.4. *Si X est une v.a dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ on a pour tout $\lambda > 0$*

$$\mathbb{P}(|X| > \lambda) \leq \frac{\mathbb{E}(|X|^2)}{\lambda^2}.$$

Preuve. Il suffit de remarquer que $\{X > \lambda\} = \{X^2 > \lambda^2\}$ et d'appliquer la proposition précédente à la v.a $Y = X^2$. □

Appliquée à la v.a $Y = X - \mathbb{E}(X)$ la proposition précédente donne le théorème de Bienaymé–Tchebychev qui permet d'obtenir les probabilités des déviations importantes de la v.a X par rapport à sa moyenne.

Théorème 3.3.5 (Inégalité de Bienaymé–Tchebychev)**.** *Si X est une v.a dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ on a pour tout $\lambda > 0$*

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda) \leq \frac{\mathbf{Var}(|X|)}{\lambda^2}.$$

Si on note $\sigma = \sqrt{\mathbf{Var}(X)}$ l'écart type on a donc

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda\sigma) < \frac{1}{\lambda^2},$$

(d'où le nom d'écart type donné à σ)

3.4 Fonctions génératrices

Pour les variables aléatoires à valeurs dans $\mathbb{N}$, de nombreux calculs sont simplifiés par l'utilisation des fonctions génératrices ; c'est un outil très utile notamment pour calculer les moments d'ordre p d'une v.a à valeurs dans $\mathbb{N}$.

Définition 3.4.1. *La fonction génératrice d'une v.a X à valeurs dans $\mathbb{N}$ est la fonction définie par*

$$\phi_X(t) = \mathbb{E}(t^X) = \sum_{k\in\mathbb{N}} t^k \mathbb{P}(X = k).$$

Proposition 3.4.1. *Si X est une variable aléatoire à valeurs dans $\mathbb{N}$ alors sa fonction génératrice ϕ_X est une série entière de rayon de convergence au moins 1.*

Preuve. Pour t dans $[-1, 1]$, on a

$$\sum_{k\in\mathbb{N}} |t^k \mathbb{P}(X = k)| \leq \sum_{k\in\mathbb{N}} \mathbb{P}(X = k) = 1,$$

et donc on a convergence absolue de la série sur le disque $\{z \in \mathbb{C} : |z| \leq 1\}$. □

L'un des intérêts de ϕ_X réside dans la proposition suivante :

Proposition 3.4.2. *On a toujours*

$$\lim_{t\to 1, t<1} \frac{d}{dt}\phi_X(t) = \mathbb{E}(X)$$

et de façon plus générale

$$\lim_{t\to 1, t<1} \frac{d^p}{dt^p}\phi_X(t) = \mathbb{E}(X(X-1)\cdots(X-p+1)).$$

Preuve. Afin de simplifier la démonstration nous supposerons que X ne prend qu'un nombre fini de valeurs dans $\{0, 1, \ldots N\}$. Il suffit de calculer

$$\frac{d^p}{dt^p}\left(\sum_{k=0}^{N} t^k \mathbb{P}(X = k)\right) = \sum_{k=0}^{N} k(k-1)\cdots(k-p+1)t^{k-p}\mathbb{P}(X = k)$$

En faisant $t = 1$ on obtient le résultat d'après la formule de transfert.

Dans le cas général où X prend ses valeurs dans $\mathbb{N}$ on peut procéder de la façon suivante : pour $0 \leq t < 1$,

$$\frac{d^p}{dt^p}\left(\sum_{k=0}^{\infty} t^k \mathbb{P}(X = k)\right) = \sum_{k=0}^{\infty} k(k-1)\cdots(k-p+1)t^{k-p}\mathbb{P}(X = k),$$

car les séries sont uniformément convergentes. La série du membre de droite converge quand $t \to 1^-$ vers $\sum_{k=0}^{\infty} k(k-1)\cdots(k-p+1)\mathbb{P}(X = k)$ d'après le théorème de convergence monotone et cette quantité égale $\mathbb{E}(X\cdots(X-p+1))$ d'après le théorème de transfert. □

Le calcul de la variance est alors clair puisque, si $\phi''_X(1) < \infty$,

$$\begin{aligned}\mathbf{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 \\ &= \phi''_X(1) + \phi'_X(1) - (\phi'_X(1))^2.\end{aligned}$$

Loi géométrique Dans ce cas X prend ses valeurs dans $\mathbb{N}$ et $\mathbb{P}(X = k) = (1-a)a^k$ $(0 \leq a < 1)$. Ainsi

$$\phi_X(t) = (1-a)\sum_{k=0}^{\infty} t^k a^k = (1-a)\frac{1}{1-ta},$$

pourvu que $0 \leq t < a^{-1}$. On a

$$\phi'_X(t) = a\frac{1-a}{(1-ta)^2}, \qquad \phi''_X(t) = 2a^2\frac{1-a}{(1-ta)^3},$$

et donc

$$\mathbb{E}(X) = \frac{a}{1-a}, \qquad \mathbf{Var}(X) = \frac{a}{(1-a)^2}.$$

Loi binomiale Si la v.a X suit une loi (p, n) elle prend ses valeurs dans $\{0, 1, \ldots, n\}$ et $\mathbb{P}(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$. On a

$$\phi_X(t) = \sum_{k=0}^{n}\binom{n}{k}t^k p^k(1-p)^{n-k} = (tp + 1 - p)^n,$$

si bien que

$$\phi'_X(t) = pn(tp + 1 - p)^{n-1}, \qquad \phi''_X(t) = p^2 n(n-1)(tp + 1 - p)^{n-2},$$

et donc

$$\mathbb{E}(X) = np, \qquad \mathbf{Var}(X) = np(1-p).$$

Loi de Poisson Si la v.a. X suit une loi de Poisson de paramètre λ, elle prend ses valeurs dans $\mathbb{N}$ et $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$. Il vient

$$\phi_X(t) = e^{-\lambda}\sum_{k=0}^{\infty} t^k \lambda^k \frac{1}{k!} = e^{-\lambda}e^{t\lambda} = e^{(t-1)\lambda}.$$

Ainsi,

$$\phi'_X(t) = \lambda e^{(t-1)\lambda}, \qquad \phi''_X(t) = \lambda^2 e^{(t-1)\lambda},$$

et donc

$$\mathbb{E}(X) = \lambda, \qquad \mathbf{Var}(X) = \lambda.$$

Mentionons enfin une propriété très utile des fonctions génératrices : une fonction génératrice caractérise de façon unique la loi de la variable aléatoire qui la définit.

Proposition 3.4.3. *Soient X et Y deux v.a.r. à valeurs dans $\mathbb{N}$ et supposons que pour tout $0 \leq t < 1$ (ou même pour tout $t \in]a, b[\subset [0, 1[$) on ait $\phi_X(t) = \phi_Y(t)$. Alors, X et Y ont même loi : pour tout $k \in \mathbb{N}$, on a $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$.*

Preuve. Si deux séries entières qui convergent absolument sur $[0, 1]$ coincident, alors elles sont égales, et en particulier elles ont les même coefficients : en effet, comme on a convergence abolue, on peut dériver la série entière terme à terme :

$$\phi_X^{(k)}(t) = \sum_{n\in\mathbb{N}} \mathbb{P}(X = n)\frac{d}{dt}t^n = \sum_{n\in\mathbb{N}} \mathbb{P}(X = n)n(n-1)\ldots(n-k+1)t^{n-k}.$$

Dans la somme de droite, tous les termes pour $n \leq k-1$ s'annulent et on a,

$$\phi_X^{(k)}(t) = \sum_{n\geq k} \mathbb{P}(X = n)n(n-1)\cdots(n-k+1)t^{n-k},$$

et, en évaluant en $t = 0$, $\phi_X^{(k)}(0) = \mathbb{P}(X = k)k!$. Ainsi, si ϕ_X et ϕ_Y coincident sur $[0, 1]$, toutes les dérivées en zero sont égales et donc, quelque soit $k \in \mathbb{N}$, on a $\mathbb{P}(X = k) = \mathbb{P}(Y = k)$. $\square$

3.5 Variables aléatoires à densité

3.5.1 Définition et exemples

Définition 3.5.1. *On dit que la variable aléatoire $X : \Omega \to \mathbb{R}$ admet une densité continue (resp. continue par morceaux* etc.*) s'il existe une fonction positive continue (resp. continue par morceaux* etc.*) $\rho_X : \mathbb{R} \to [0, \infty[$ telle que $\int_{-\infty}^{\infty} \rho_X(t)dt = 1$ et telle que pour tout intervalle I*

$$\mathbb{P}(X \in I) = \int_I \rho_X(t)dt.$$

Faisons une remarque importante : si une v.a. X admet une densité ρ_X alors sa *fonction de répartition*

$$F_X(x) = \mu_X(]-\infty, x]) = \int_{-\infty}^{x} \rho_X(t)dt$$

est continue.

De la définition précédente on déduit facilement la proposition suivante.

Proposition 3.5.1. *Une v.a.r. X admet une densité ρ continue par morceaux si et seulement si sa fonction de répartition F_X est continue et dérivable par morceaux. On a alors en tout point où F_X est dérivable $F'_X(x) = \rho_X(x)$.*

Compléments. Si on note μ_X la loi de X ($\mu_X(A) = \mathbb{P}(X \in A)$ pour $A \in Bor(\mathbb{R})$) on a donc $\mu_X(I) = \int_I \rho_X(t)dt$ pour tout intervalle I et $F_X(x) = \mu_X(]-\infty, x]) = \int_{-\infty}^{x} \rho_X(t)dt$. Il existe donc des variables aléatoires n'admettant pas de densité : par exemple une v.a X à valeurs dans $\mathbb{R}$ ne prenant que deux valeurs 0 ou 1 et telle que $\mathbb{P}(X = 0) = p$ avec $p = 1/2$ ne peut posséder de densité car sa fonction de répartition $F_X(x)$ vaut 0 si $x < 0$, $1/2$ si $0 \leq x < 1$ et 1 si $1 \leq x$: elle est discontinue en 0 et en 1 (mais bien continue à droite).

Loi uniforme La variable aléatoire $X : \Omega \to \mathbb{R}$ suit une loi uniforme sur l'intervalle $[a, b]$ si sa densité est donnée par

$$\rho_X(x) = \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}.$$

On a bien $\rho_X(t) \geq 0$ pour tout t et $\int_{\mathbb{R}} \rho_X(t)dt = 1$. Cette loi est caractérisée par

$$\begin{aligned}
\mathbb{P}(X \in [c, d]) &= \int_{[c,d]} \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}(x)dx \\
&= \frac{1}{b-a} \int_{\mathbb{R}} \mathbf{1}_{[c,d]}(x) \cdot \mathbf{1}_{[a,b]}(x)dx \\
&= \frac{1}{b-a} \int_{\mathbb{R}} \mathbf{1}_{[c,d]\cap[a,b]}(x)dx \\
&= \frac{1}{b-a} \text{longueur}([a, b] \cap [c, d])
\end{aligned}$$

Loi exponentielle de paramètre θ La v.a admet une densité ρ_X exponentielle de paramètre θ si

$$\rho_X(x) = \theta e^{-\theta x} \mathbf{1}_{[0,\infty[}(x).$$

La fonction de répartition est

$$F_X(x) = \int_{-\infty}^{x} \rho_X(t)dt = (1 - e^{-\theta x})\mathbf{1}_{[0,\infty[}(x),$$

et converge bien vers 1 en ∞. Intuitivement, la loi uniforme sur l'intervalle $[a, b]$ modélise une expérience où la probabilité d'un point de tomber dans un intervalle de taille $2\Delta x$, $]x - \Delta x, x + \Delta x[\subset [a, b]$ ne dépend pas de x (et est linéaire en Δx).

Loi normale $\mathcal{N}(\mu, \sigma^2)$ C'est la loi de densité

$$\rho_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}.$$

Il n'est pas complètement évident que $\int_{\mathbb{R}} \rho_X(x)dx = 1$ (ce qui est indispensable pour que ρ_X soit une densité). Ceci résulte, après le changement de variable $u = (x-\mu)/\sigma$ de l'égalité (voir plus loin pour une preuve)

$$\int_{-\infty}^{\infty} e^{-u^2/2} du = \sqrt{2\pi}.$$

La loi normale $\mathcal{N}(0,1)$, qui a pour densité

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

est dite loi normale centrée réduite.

3.5.2 Espérance des v.a.r. admettant une densité

Dans la section précédente 3.2.2 nous avons vu comment étendre la notion d'espérance définie pour des v.a.r prenant un nombre fini ou dénombrable de valeurs à des v.a.r. positives ou $\mathcal{L}^1$. Dans cette section nous donnons une formule qui permet d'exprimer l'espérance d'une v.a.r. admettant une densité en fonction de la densité. Nous *admettrons* ainsi le théorème important suivant :

Théorème 3.5.1. *Soit X une v.a .r. X qui admet une densité ρ_X.*
a) Si X est positive on peut toujours écrire

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\rho_X(x)dx.$$

b) Si X est de signe quelconque, on a $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si

$$\int_{-\infty}^{\infty} |x|\rho_X(x)dx < \infty \quad \textit{et dans ce cas} \qquad \mathbb{E}(X) = \int_{-\infty}^{\infty} x\rho_X(x)dx.$$

Loi uniforme sur $[a,b]$ Soit X une v.a. uniforme sur l'intervalle $[a,b]$, qui a densité $\rho(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$. Alors $\mathbb{E}X$ existe bien et

$$\mathbb{E}X = \int_{\mathbb{R}} x\rho(x)dx = \int_{\mathbb{R}} x\frac{\mathbf{1}_{[a,b]}(x)}{b-a}dx = \frac{1}{b-a}\int_a^b xdx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2},$$

ce qui est conforme à l'intuition : la position en moyenne d'un point jeté au hasard sur l'intervalle $[a,b]$ sera située au milieu de l'intervalle $[a,b]$.

Loi exponentielle de paramètre θ La v.a admet une densité

$$\rho_X(x) = \theta e^{-\theta x}\mathbf{1}_{[0,\infty[}(x).$$

On a

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x\theta e^{-\theta x}\mathbf{1}_{[0,\infty[}(x)dx \\
&= \int_0^{\infty} x\theta e^{-\theta x}dx \\
&= \frac{1}{\theta}\int_0^{\infty} ue^{-u}du \\
&= \frac{1}{\theta},
\end{aligned}$$

(après changement de variable et intégration par parties).

Loi normale centrée réduite Soit X une v.a suivant une loi normale centrée réduite : Une telle loi admet une densité

$$\rho_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Comme

$$\mathbb{E}(|X|) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x| e^{-x^2/2} dx$$

est finie X est dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. L'espérance de X est donc définie et vaut

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Il s'agit de l'intégrale d'une fonction intégrable, impaire sur un intervalle symétrique : cette intégrale est donc nulle, c'est-à-dire $\mathbb{E}(X) = 0$.

3.5.3 Formule de transfert

Etant donnée une v.a.r. X admettant une densité ρ_X et une fonction continue (ou continue par morceaux) on se propose de déterminer l'espérance de la v.a.r. $Y := f(X)$. Nous *admettrons* le résultat suivant :

Théorème 3.5.2. *Si $X : \Omega \to \mathbb{R}$ est une v.a admettant une densité ρ_X et $f : \mathbb{R} \to \mathbb{R}$ est une fonction continue (ou continue par morceaux) alors la v.a $Y = f(X)$ est $\mathbb{P}$-intégrable si et seulement si l'intégrale $\int_{-\infty}^{\infty} |f(x)| \rho_X(x) dx$ est finie et dans ce cas l'espérance de $Y = f(X)$ est donnée par*

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

Exercice 3.5.1. *Calculer $\mathbb{E}(Y)$ où $Y = X^2$ est le carré d'une v.a suivant une loi normale centrée réduite.*

Solution. On a

$$\begin{aligned}
\mathbb{E}(X^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x (x e^{-x^2/2}) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \frac{d}{dx}(-e^{-x^2/2}) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \\
&= 1
\end{aligned}$$

(on a effectué une intégration par parties). △

Compléments : Dans le cas général des v.a de la forme $Y = f(X)$ où $X : \Omega \to \mathbb{R}$ est une v.a et $f : \mathbb{R} \to \mathbb{R}$ est continue, la forme générale de la formule de transfert fait intervenir la *loi* de X :

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x) d\mu_X(x)$$

toutes les fois où l'intégrale converge.

3.5.4 Variance des v.a. à densité

Si X est une v.a admettant une densité ρ_X, alors d'après la formule de transfert, X est dans $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ si et seulement si

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \rho_X(x) dx,$$

est finie.

Loi uniforme La variable aléatoire $X : \Omega \to \mathbb{R}$ suit une loi uniforme sur l'intervalle $[a, b]$ si sa densité est donnée par

$$\rho_X(x) = \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}(x).$$

On a a déjà montré à l'exercice 3.5.2 que

$$\mathbb{E}(X) = \frac{a+b}{2}.$$

Pour calculer la variance, on calcule d'abord $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 \cdot \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}(x)dx = \frac{1}{b-a}\int_a^b x^2 dx = \frac{1}{b-a} \times \frac{b^3-a^3}{3} = \frac{a^2+b^2+ab}{3},$$

puisque $b^3 - a^3 = (b-a)(b^2+ab+a^2)$, et donc

$$\begin{aligned}\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X) &= \frac{a^2+b^2+ab}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{(b-a)^2}{12}.\end{aligned}$$

Loi exponentielle de paramètre θ La v.a admet une densité

$$\rho_X(x) = \theta e^{-\theta x}\mathbf{1}_{[0,\infty[}(x).$$

Le moment d'ordre 2 s'obtient comme l'espérance

$$\mathbb{E}(X^2) = \int_0^\infty x^2\theta e^{-\theta x}dx = \frac{1}{\theta}\int_0^\infty ue^{-u}du = \frac{2}{\theta^2},$$

(après changement de variable et deux intégrations par parties). On a donc

$$\mathbf{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}.$$

Loi normale $\mathcal{N}(\mu, \sigma^2)$ On sait (*cf.* l'exercice de la section 3.5.5) que si Y suit une loi $\mathcal{N}(\mu, \sigma^2)$ alors elle est de la forme $\sigma X + \mu$ où X suit une loi normale $\mathcal{N}(0, 1)$ de densité

$$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

On sait que pour une telle loi, $\mathbb{E}(X) = 0$ et $\mathbf{Var}(X) = 1$. Par conséquent, $\mathbb{E}(Y) = \sigma\mathbb{E}(X) + \mu = \mu$ et $\mathbf{Var}(Y) = \sigma^2\mathbf{Var}(X) = \sigma^2$.

3.5.5 Calcul de la loi d'une variable à densité

Le problème qui nous intéresse dans cette section est le suivant : Etant donnée une v.a X dont on connait la densité ρ_X, déterminer la densité, si elle existe de la v.a $Y = f(X)$, où f est une fonction continue de $\mathbb{R}$ dans $\mathbb{R}$. D'après la formule de transfert, si on connait la densité ρ_Y de Y alors on peut exprimer $\mathbb{E}(\phi(Y))$ quelque soit ϕ continue, pourvu que $\mathbb{E}(|\phi(Y)|) < \infty$:

$$\mathbb{E}(\phi(Y)) = \int \phi(y)\rho_Y(y)dy. \tag{3.2}$$

C'est en particulier vrai dès que ϕ est bornée puisque $\mathbb{E}(\phi(Y)) \leq \|\phi\|_\infty$. On peut voir les fonctions ϕ comme des fonctions "test" qui nous apportent un peu d'information sur Y, à travers l'espérance $\mathbb{E}(\phi(Y))$. Il se trouve que si (3.2) est vraie pour suffisamment de fonctions ϕ, alors on connaît Y !

Le résultat qui va nous permettre d'aborder cette question est le théorème suivant que nous admettrons :

Théorème 3.5.3. *Si X est une v.a telle que pour toute fonction continue bornée $\phi : \mathbb{R} \to \mathbb{R}$ on a*

$$\mathbb{E}(\phi(X)) = \int_{-\infty}^{\infty} \phi(x)\rho_X(x)dx,$$

alors X admet ρ_X pour densité.

Pour calculer la loi de $Y = f(X)$, on calcule sa densité en exprimant $\mathbb{E}[\phi(Y)]$ sous forme d'une intégrale dont l'intégrand est de la forme $\phi(y)\rho(y)$. On aura alors montré que ρ est la densité de Y. Plus précisément, supposons que Y admette une densité ρ_Y. On doit alors avoir pour toute fonction continue bornée $\phi : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}(\phi(Y)) = \int_{-\infty}^{\infty} \phi(y)\rho_Y(y)dy.$$

Mais $\phi(Y) = \phi(f(X)) = \phi \circ f(X)$ et on a donc,

$$\mathbb{E}(\phi(Y)) = \mathbb{E}(\phi \circ f(X)) = \int_{-\infty}^{\infty} \phi \circ f(x)\rho_X(x)dx.$$

Supposons que ϕ soit une bijection dérivable de $\mathbb{R}$ dans $\mathbb{R}$ envoyant $\mathbb{R}$ sur $\mathbb{R}$. La formule classique de changement de variable montre que ($x = f^{-1}(y)$, $dx = 1/f'(f^{-1}(y))dy$)

$$\int_{-\infty}^{\infty} \phi \circ f(x)\rho_X(x)dx = \int_{-\infty}^{\infty} \phi(y)\frac{1}{|f'(f^{-1}(y))|}\rho_X(f^{-1}(y))dy.$$

En conclusion, pour toute fonction ϕ continue de $\mathbb{R} \to \mathbb{R}$

$$\int_{-\infty}^{\infty} \phi(y)\rho_Y(y)dy = \int_{-\infty}^{\infty} \phi(y)\frac{1}{|f'(f^{-1}(y))|}\rho_X(f^{-1}(y))dy$$

et il est naturel de penser que

$$\rho_Y(y) = \frac{1}{|f'(f^{-1}(y))|}\rho_X(f^{-1}(y)). \tag{3.3}$$

Justifions le fait que c'est effectivement le cas. Le calcul précédent montre que pour toute fonction continue bornée $\phi : \mathbb{R} \to \mathbb{R}$ on a

$$\mathbb{E}(\phi(Y)) = \int_{-\infty}^{\infty} \phi(y)\frac{1}{|f'(f^{-1}(y))|}\rho_X(f^{-1}(y))dy.$$

Mais d'après le théorème 3.5.3 cela implique que Y admet une densité et que celle-ci est donnée par la formule (3.3).

On pourrait démontrer de la même manière :

Théorème 3.5.4. *Soient X une v.a de densité ρ_X prennant ses valeurs dans un intervalle I (fini ou infini) et $f : I \to J$ est une application de classe C^1 pas nécessairement bijective mais telle que tout point $y \in J$ ait un nombre fini d'antécédents. Alors, la v.a $Y = f(X)$ admet une densité ρ_Y dont l'expression est donnée par*

$$\rho_Y(y) = \sum_{x \in f^{-1}(y)} \frac{\rho_X(x)}{|f'(x)|} \cdot \mathbf{1}_J.$$

L'expression précédente peut prendre la valeur ∞ mais la fonction positive ρ_Y restera d'intégrale 1.

Appliquons ce qui précède à un exemple.

Exercice 3.5.2. *Soit U une v.a. uniforme sur $(0, 1)$. Déterminer la loi de $X = -\log U$.*

Solution. La v.a. U a pour densité $\mathbf{1}_{[0,1]}(u)$. Soit ϕ une fonction continue bornée. Alors, avec $f := x \mapsto -\log x$, la fonction $\phi \circ f$ est continue bornée et donc, par la formule de transfert, et avec le changement de variable $-\log u = x$, $u = e^{-x}$, $dx = -du/u$

$$\mathbb{E}[\phi(X)] = \mathbb{E}[\phi(-\alpha \log U)] = \int_0^1 \phi(-\alpha \log u)du = \int_0^\infty \phi(x)e^{-x}dx = \int_{\mathbb{R}} \phi(x)e^{-x}\mathbf{1}_{\mathbb{R}^+}(x)dx,$$

ce qui prouve que la densité de $X = -\log U$ est $e^{-x}\mathbf{1}_{\mathbb{R}^+}(x)$ c'est-à-dire que X est exponentielle de paramètre 1. △

Exercice 3.5.3. *Si X suit une loi normale centrée réduite, déterminer la loi de $Y = \sigma X + \mu$.*

Solution. On a $Y = f(X)$ où $f(x) = ax + b$. C'est une bijection C^1 de $\mathbb{R}$ sur $\mathbb{R}$. La formule du théorème 3.5.4 montre donc que la denstité de Y est

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2},$$

c'est-à-dire que Y suit une loi $\mathcal{N}(\mu, \sigma^2)$. △

Exercice 3.5.4. *Supposons que X admette une densité ρ_X. Déterminer la densité, si elle existe de $Y = X^2$. Application au cas où X suit une loi normale $\mathcal{N}(0, 1)$.*

Solution. On a $Y = f(X)$ où $f(x) = x^2$ est une bijection de $I_- =]-\infty, 0[$ sur $]0, \infty[$ et de $I_+ =]0, \infty[$ sur $]0, \infty[$ (f est une fonction continue strictement décroissante sur $I_- =]-\infty, 0]$ et strictement croissante sur $I_+ = [0, \infty[$.) Pour toute fonction $\phi : \mathbb{R} \to \mathbb{R}$ continue et bornée

$$\mathbb{E}(\phi(Y)) = \mathbb{E}(\phi \circ f(X)) = \int_{-\infty}^{\infty} \phi(f(x))\rho_X(x)dx.$$

Ecrivons,

$$\int_{-\infty}^{\infty} \phi(f(x))\rho_X(x)dx = \int_{-\infty}^{0} \phi(x^2)\rho_X(x)dx + \int_0^\infty \phi(x^2)\rho_X(x)dx$$

et effectuons dans chacune des intégrales du membre de droite le changement de variable $y = x^2$:

$$\int_{-\infty}^{0} \phi(x^2)\rho_X(x)dx = \int_0^\infty \phi(y)\rho(-\sqrt{y})\frac{dy}{2\sqrt{y}},$$
$$\int_0^\infty \phi(x^2)\rho_X(x)dx = \int_0^\infty \phi(y)\rho(\sqrt{y})\frac{dy}{2\sqrt{y}},$$

si bien que

$$\begin{aligned}\mathbb{E}(\phi(Y)) &= \int_0^\infty \phi(y)\Big(\rho(-\sqrt{y}) + \rho(\sqrt{y})\Big)\frac{1}{2\sqrt{y}}dy \\ &= \int_{\mathbb{R}} \phi(y)\Big(\rho(-\sqrt{y}) + \rho(\sqrt{y})\Big)\frac{1}{2\sqrt{y}} \cdot \mathbf{1}_{]0,\infty[}(y)dy\end{aligned}$$

et comme cette formule est vraie pour toute fonction $\phi : \mathbb{R} \to \mathbb{R}$ continue bornée, on peut conclure que Y admet une densité ρ_Y égale à

$$\rho_Y(y) = \Big(\rho(-\sqrt{y}) + \rho(\sqrt{y})\Big)\frac{1}{2\sqrt{y}} \times \mathbf{1}_{]0,\infty[}(y).$$

Il faut bien faire attention à ne pas oublier le terme $\mathbf{1}_{]0,\infty[}(y)$, qui est crucial puisqu'on doit avoir $\int_{\mathbb{R}} \rho_Y(y)dy = 1$!

Si X suit une loi normale $\mathcal{N}(0, 1)$ sa densité est $\rho_X(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ et la densité de $Y = X^2$ vaut

$$\rho_Y(y) = \frac{1}{\sqrt{2\pi}}\frac{e^{-y}}{\sqrt{y}}\mathbf{1}_{]0,\infty[}(y),$$

ce qui caractérise la loi de Y. △

3.6 Vecteurs aléatoires

Un vecteur aléatoire est un n-uplet de variables aléatoires réelles $X = (X_1, \dots, X_n)$.

3.6.1 Loi d'un vecteur aléatoire

Définition 3.6.1. *(Loi jointe) Si $X_1, \dots, X_n$ sont des v.a, on appelle loi du vecteur aléatoire $X = (X_1, \dots, X_n)$, ou encore la loi jointe des $X_1, X_2, \dots, X_n$, la donnée de tous les quantités*

$$\mathbb{P}((X_1, X_2, \dots, X_n) \in I_1 \times I_2 \times \cdots \times I_n) = \mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n)$$

pour tous intervalles $I_1, \dots, I_n$ de $\mathbb{R}$.

Définition 3.6.2. *(Loi marginale) Si $X = (X_1, X_2, \dots, X_n)$ est vecteur aléatoire on appelle loi marginale de X_i la loi de la variable X_i, c'est-à-dire la donnée de $\mathbb{P}(X_i \in I)$ pour tous les intervalles I de $\mathbb{R}$.*

Remarque. La donnée de toutes les lois marginales des $X_1, X_2, \dots, X_n$ ne suffit pas à caractériser la loi du vecteur $X = (X_1, \dots, X_n)$. Pour s'en convaincre, voir l'exemple suivant.

Exemple 3.6.1. *Soit U une variable aléatoire uniforme sur $[-1, 1]$. Alors, $-U$ est aussi uniforme sur $[-1, 1]$: si ϕ est continue bornée alors*

$$\mathbb{E}[\phi(-U)] = \int_{[-1,1]} \phi(-u)\frac{du}{2} = \int_{[-1,1]} \phi(v)\frac{dv}{2}.$$

Les vecteurs (U, U) et $(U, -U)$ ont donc les mêmes marginales ; par contre, ils n'ont clairement pas la même loi puisque si $I = [1/2, 1]$, $\mathbb{P}(U \in I, U \in I) = 1/4$ mais $\mathbb{P}(U \in I, -U \in I) = 0$.

Dans la suite nous distinguerons deux cas, suivant que les X_i prennent des valeurs dans un ensemble fini ou dénombrable ou sont des v.a.r. ayant une densité.

Cas des v.a. discrètes Si chacune des v.a. X_i, prend ses valeurs dans un ensemble E_i fini ou dénombrable ($i = 1, \dots, n$), la loi du vecteur $X = (X_1, \dots, X_n)$ est déterminée par les nombres

$$p(e_1, \dots, e_n) = \mathbb{P}(X_1 = e_1, \dots, X_n = e_n),$$

où $(e_1, \dots, e_n) \in E_1 \times \cdots \times E_n$. La loi marginale de X_i est alors donnée par

$$\mathbb{P}(X_i = x) = \mathbb{P}\Big(\bigsqcup_{e_1,\dots,e_{i-1},e_i,\dots,e_n} \{X_1 = e_1, \dots, X_{i-1} = e_{i-1}, X_i = x, X_{i+1} = e_{i+1}, \dots, X_n = e_n\}\Big)$$
$$\sum_{e_1,\dots,e_{i-1},e_{i+1},\dots,e_n} p(e_1, \dots, e_{i-1}, x, e_{i+1}, \dots, e_n).$$

Cas des v.a. à densité Si le vecteur $X = (X_1, \dots, X_n)$ admet une densité, c'est-à-dire qu'il existe une fonction $\rho(x_1, \dots, x_n)$ telle que pour tous intervalles $I_1, \dots, I_n$

$$\mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n) = \int_{I_1 \times \cdots \times I_n} \rho(x_1, \dots, x_n) dx_1 \cdots dx_n. \tag{3.4}$$

Dans ce cas, chacune des v.a. X_i admet aussi une densité ρ_i telle que pour tout intervalle I de $\mathbb{R}$ on a

$$\rho_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} \rho(x_1, \dots, x_i, \dots x_n) dx_1 \cdots dx_{i-1} \cdot dx_{i+1} \cdots dx_n,$$

(et donc on intègre sur toutes le variables sauf x_i.)

3.6.2 Formules de transfert

Etant donné un vecteur aléatoire $X = (X_1, \dots, X_m)$ et

$$\begin{aligned} f : \mathbb{R}^m &\to \mathbb{R}^p \\ (x_1, \dots, x_m) &\mapsto (f_1(x_1, \dots, x_m), \dots, f_p(x_1, \dots, x_m)) \end{aligned}$$

une application continue, on cherche à calculer l'espérance de $Y = f(X)$ c'est-à-dire que l'on veut calculer le vecteur $\mathbb{E}(Y) = (\mathbb{E}(Y_1), \dots, \mathbb{E}(Y_p))$ où $Y_j = f_j(X_1, \dots, X_m)$. Le problème se ramène donc au suivant : étant donné $X = (X_1, \dots, X_m)$ un vecteur aléatoire calculer $\mathbb{E}(\phi(X_1, \dots, X_m))$ où ϕ est une application de $\mathbb{R}^m$ dans $\mathbb{R}$. Ce problème se résout de la même façon que dans le cas $m = 1$ (variables aléatoires). Nous donnons les résultats sans démonstration (les preuves sont identiques à celles du cas $m = 1$).

Cas des vecteurs aléatoires discrets Si les $E_1, \dots, E_n$ sont tous dénombrables et si $X = (X_1, \dots, X_m)$ prend ses valeurs dans $E_1 \times \cdots \times E_m$ on a

$$\mathbb{E}(\phi(X_1, \dots, X_m)) = \sum_{e_1 \in E_1, \dots, e_m \in E_m} \phi(e_1, \dots, e_m) \mathbb{P}(X_1 = e_1, \dots, X_m = e_m).$$

Cas des vecteurs aléatoires à densité Si le vecteur aléatoire $X = (X_1, \dots, X_m)$ admet une densité $\rho_(x_1, \dots, x_m)$ on a

$$\mathbb{E}(\phi(X_1, \dots, X_m)) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \phi(x_1, \dots, x_m) \rho_X(x_1, \dots, x_m) dx_1 \cdots dx_m.$$

Remarque. 1) On notera que si toutes les variables aléatoires $X_1, \dots, X_n$ sont discrètes alors le vecteur $(X_1, \dots, X_n)$ est lui aussi toujours discret : en effet, si $E_1, \dots, E_n$ sont dénombrables alors le produit cartesien $E_1 \times E_2 \times \cdots \times E_n$ est aussi dénombrable.

2) Par contre, si les v.a. $X_1, X_2, \dots, X_n$ sont toutes à densité alors le vecteur $(X_1, \dots, X_n)$ peut très bien ne pas admettre de densité. Voir l'exemple suivant.

Exemple 3.6.2. *Soit X une variable aléatoire uniforme sur $[0, 1]$. Donc X admet une densité. On considère le vecteur (X, X), c'est à dire la fonction de Ω dans $\mathbb{R}^2$ qui à $\omega \in \Omega$ associe $(X(\omega), X(\omega))$. Quelque soit $\omega \in \Omega$, ce vecteur est dans la diagonale $\Delta = \{(x, x) : x \in [0, 1]\}$, et on se dit, même intuitivement qu'il ne devrait pas y avoir de densité pour (X, X). Pour le vérifier, on peut supposer l'existence d'une densité ρ pour aboutir à une absurdité. Soit A pavé $[0, 1]^2$ tel que $A \cap \Delta = \varnothing$. Alors*

$$\mathbb{P}((X, X) \in A) = \int_A \rho(x_1, x_2) dx_1 dx_2 = 0.$$

On en déduit que $\rho(x_1, x_2) = 0$ pour tout $(x_1, x_2) \notin \Delta$ c'est-à-dire pour tout $x_1 \neq x_2$. Mais alors

$$\int_{[0,1]^2} \rho(x_1, x_2) dx_1 dx_2 = 0!$$

Le vecteur aléatoir (X, X) n'admet donc pas de densité sur $\mathbb{R}^2$.

3.6.3 Loi de d'une somme de v.a.r.

Théorème 3.6.1. *Soient X et Y deux v.a.r.*

a) Si les v.a.r. X, Y prennent leurs valeurs dans des ensembles finis ou dénombrables et si p est la loi du vecteur aléatoire (X, Y), alors la loi p_{X+Y} de $X + Y$ vérifie pour tout e

$$p_{X+Y}(e) = \sum_{e_1 + e_2 = e} p(e_1, e_2).$$

b) Si le vecteur aléatoire (X, Y) admet pour densité ρ alors la v.a.r. $X + Y$ admet pour densité ρ_{X+Y} :

$$\rho_{X+Y}(x) = \int_{\mathbb{R}} \rho(x - y, y)dy.$$

Preuve. a) Si les ensembles de valeurs prises par X et Y sont au plus dénombrables, alors

$$\mathbb{P}(X + Y = e) = \mathbb{P}\Big(\bigsqcup_{e_1+e_2=e} \{X = e_1, Y = e_2\}\Big) = \sum_{e_1+e_2=e} \mathbb{P}(X_1 = e_1, X_2 = e_2),$$

puisque l'union est disjointe et dénombrable. Comme par définition $\mathbb{P}(X_1 = e_1, X_2 = e_2) = p(e_1, e_2)$ on a bien le résultat annoncé.
b) Faisons maintenant la preuve dans le second cas. Introduisons le vecteur aléatoire $Z = (X + Y, Y)$ et calculons sa loi. Pour toute fonction $\phi : \mathbb{R}^2 \to \mathbb{R}$ continue bornée,

$$\mathbb{E}(\phi(Z)) = \mathbb{E}(\phi(X + Y, Y)) = \mathbb{E}(\psi(X, Y)),$$

où $\psi : \mathbb{R}^2 \to \mathbb{R}$ égale à $\psi(x, y) = \phi(x + y, y)$. Comme (X, Y) admet une densité ρ on a d'après la formule de transfert

$$\begin{aligned}\mathbb{E}(\psi((X, Y))) &= \int_{\mathbb{R}^2} \psi(x, y)\rho(x, y)dxdy \\ &= \int_{\mathbb{R}^2} \phi(x + y, y)\rho(x, y)dxdy \\ &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(x + y, y)\rho(x, y)dx\Big)dy.\end{aligned}$$

Effectuons le changement de variables (y étant fixé) $u = x + y$, $x = u - y$ dans l'intégrale du milieu :

$$\mathbb{E}(\psi((X, Y))) = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} \phi(u, y)\rho(u - y, y)du\Big)dy$$

et donc,

$$\mathbb{E}(\phi(Z)) = \int_{\mathbb{R}^2} \phi(u, v)\rho(u - v, v)dudv.$$

Cela étant vrai pour toute fonction ϕ continue bornée on peut dire que $Z = (U, V) = (X + Y, Y)$ admet une densité donnée par $\rho_Z(u, v) = \rho(u - v, v)$. On en déduit que, quelque soit l'intervalle I de $\mathbb{R}$,

$$\mathbb{P}(X + Y \in I) = \mathbb{P}(X + Y \in I, Y \in \mathbb{R}) = \int_I \int_{\mathbb{R}} \rho(u - v, v)dvdu$$

et donc que la densité de $X + Y$ est donnée par

$$\rho_{X+Y}(x) = \int_{-\infty}^{\infty} \rho(x - v, v)dv,$$

comme annoncé. □

3.7 Variables aléatoires indépendantes

3.7.1 Définition

La notion d'indépendance de n v.a.r. est la suivante :

Définition 3.7.1. *Une suite $X_1, \ldots, X_n$ de variables aléatoires est dite indépendante si pour tous intervalles $I_1, \ldots, I_n$ de $\mathbb{R}$,*

$$\mathbb{P}(X_1 \in I_1, \ldots, X_n \in I_n) = \mathbb{P}(X_1 \in I_1) \cdots \mathbb{P}(X_n \in I_n).$$

Remarque. 1) En d'autres termes, une famille $(X_1, X_2, \ldots, X_n)$ est indépendante si et seulement si quelque soient $I_1, I_2, \ldots, I_n$ intervalles de $\mathbb{R}$ les événements $\{X_1 \in I_1\}, \ldots, \{X_n \in I_n\}$ sont indépendants.
2) On remarquera que dans le cas des variables aléatoires, on n'a pas besoin de considérer toutes les sous-familles ! En effet, le fait de devoir autoriser tous les n-uplets d'intervalles $(I_1, \ldots, I_n)$ rend ceci déjà possible : si on fixe, $I_i = \mathbb{R}$ l'événement correspondant $\{X_i \in I_i\} = \Omega$ n'intervient pas dans la probabilité !

Définition 3.7.2. *Une famille $(X_i)_{i\in A}$ de v.a est dite indépendante si toute sous famille finie $(X_i)_{i\in J}$, $J \subset A$ fini est indépendante.*

Nous admettrons le théorème suivant :

Théorème 3.7.1. *Soit $(X_i)_{i\in A}$ une famille de variables aléatoires indépendantes et $(f_i)_{i\in A}$ une famille de fonctions continues. Alors la famille $(f_i(X_i))_{i\in A}$ est indépendante.*

Exemple 3.7.1. *L'exemple suivant sera crucial lorsque nous reparlerons de fonctions génératrices. Si X et Y sont indépendantes et $s \in [0, 1]$ alors les variables aléatoires s^X et s^Y sont indépendantes.*

3.7.2 Cas des v.a. discrètes

Si l'on suppose que les X_i sont à valeurs dans E_i ensembles fini ou dénombrables la définition précédente se simplifie car on a la proposition suivante dont la preuve est facile :

Proposition 3.7.1. *La famille $(X_1, \ldots, X_n)$ est indépendante si et seulement si pour tout $(e_1, \ldots, e_n) \in E_1 \times \cdots \times E_n$ on a*

$$\mathbb{P}(X_1 = e_1, \ldots, X_n = e_n) = \mathbb{P}(X_1 = e_1) \cdots \mathbb{P}(X_n = e_n).$$

Définition 3.7.3. *Si on note $p_{X_i}(e) := \mathbb{P}(X_i = e)$ et $p(e_1, \ldots, e_n) = \mathbb{P}(X_1 = e_1, \ldots, X_n = e_n)$ on dit que p est la loi du vecteur aléatoire $(X_1, \ldots, X_n)$. Ainsi, si $X_1, \ldots, X_n$ sont indépendantes on a*

$$p(e_1, \ldots, e_n) = p_{X_1}(e_1) \cdots p_{X_n}(e_n).$$

Donnons un exemple important.

Exemple 3.7.2. *Dans le jeu de Pile/Face infini $\Omega = \{0, 1\}^{\mathbb{N}}$ muni de la tribu $\mathcal{F}$ engendrée par les évévements élémentaires (les cylindres) $C_{i,e} = X_i^{-1}(e)$, $e \in \{0, 1\}$, $i \in \mathbb{N}$ (où les X_i sont définies par $X_i(\omega) = \omega_i$) et de la probabilité $\mathcal{P}$ décrite en 2.6.2. Pour toute sous famille $X_{i_1}, \ldots, X_{i_n}$ de $X_1, X_2, \ldots,$ on a comme précédemment*

$$\begin{aligned}
\mathbb{P}(X_{i_1} = \epsilon_1, \ldots, X_{i_n} = \epsilon_n) &= \mathbb{P}(\{\omega : \omega_1 = \epsilon_1, \ldots, \omega_n = \epsilon_n\}) \\
&= \mathbb{P}(\{\omega : \omega \in C_{i_1,\epsilon_1} \cap \cdots \cap C_{i_n,\epsilon_n}) \\
&= p^{\sum_{l=1}^n \epsilon_l}(1-p)^{n-\sum_{l=1}^n \epsilon_l}
\end{aligned}$$

par définition de la probabilité $\mathbb{P}$. Par ailleurs,

$$\begin{aligned}
\mathbb{P}(X_{i_1} = \epsilon_1) \cdots \mathbb{P}(X_{i_n} = \epsilon_n) &= p^{\epsilon_1}(1-p)^{1-\epsilon_1} \cdots p^{\epsilon_n}(1-p)^{1-\epsilon_n} \\
&= p^{\sum_{l=1}^n \epsilon_l}(1-p)^{n-\sum_{l=1}^n \epsilon_l} \\
&= \mathbb{P}(X_{i_1} = \epsilon_1, \ldots, X_{i_n} = \epsilon_n).
\end{aligned}$$

On a donc bien pour tous $\epsilon_1, \ldots, \epsilon_n$ l'égalité $\mathbb{P}(X_{i_1} = \epsilon_1, \ldots, X_{i_n} = \epsilon_n) = \mathbb{P}(X_{i_1} = \epsilon_1) \cdots \mathbb{P}(X_{i_n} = \epsilon_n)$ ce qui démontre l'indépendance de toute sous-famille $X_{i_1}, \ldots, X_{i_n}$. Nous avons donc démontré que la famille $X_1, X_2, \ldots$ est indépendante.

Remarque. Le résultat précédent est beaucoup plus profond qu'il n'apparaît. La vérification de l'indépendance des v.a X_i est facile mais la construction de le mesure $\mathbb{P}$, qui est justement celle qui rend la famille $(X_i)_i$ indépendante, est loin d'être triviale.

Il existe en fait un résultat plus général : Etant donnée une famille $(X_\alpha)_{\alpha\in A}$ de v.a à valeurs dans $\mathbb{R}$ définies sur un espace probabilisé $(\Omega', \mathcal{F}', \mathbb{P}')$, il existe un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et des v.a Y_α, $\alpha \in A$ telles que pour tout $\alpha \in A$, Y_α a même loi que X_α et telles que la famille $(Y_\alpha)_{\alpha\in A}$ soit indépendante.

Ainsi, la construction de v.a de lois données qui sont indépendantes est toujours possible : c'est ce qui justifie les énoncé d'exercice du type "Soient X et Y deux variables aléatoires de loi *** et indépendantes...".

3.7.3 Cas des v.a admettant des densités

On a la proposition suivante

Proposition 3.7.2. *Si les $X_1, \dots, X_n$ sont des v.a indépendantes admettant des densités ρ_{X_1}, $\dots$, ρ_{X_n} alors pour tous intervalles $I_1, \dots, I_n$ de $\mathbb{R}$ on a*

$$\mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n) = \int_{I_1\times\cdots\times I_n} \rho(x_1, \dots, x_n) dx_1 \cdots dx_n$$

où

$$\rho(x_1, \dots, x_n) = \rho_{X_1}(x_1) \cdots \rho_{X_n}(x_n).$$

On dit que ρ est la densité du vecteur aléatoire $X = (X_1, \dots, X_n)$. Réciproquement, si la densité du vecteur $(X_1, X_2, \dots, X_n)$ s'écrit $\rho_{X_1(x_1)} \cdots \rho_{X_n}(x_n)$ alors les v.a. $(X_1, \dots, X_n)$ sont indépendantes.

Preuve. Si les v.a. sont indépendantes, alors, pour tous intervalles $I_1, \dots, I_n$,

$$\begin{aligned}
\mathbb{P}(X \in I_1 \times \cdots \times I_n) &= \mathbb{P}(X_1 \in I_1) \cdots \mathbb{P}(X_n \in I_n) \\
&= \left(\int_{I_1} \rho_{X_1}(x_1) dx_1\right) \cdots \left(\int_{I_n} \rho_{X_n}(x_n) dx_n\right) \\
&= \int_{I_1\times\cdots\times I_n} \rho_{X_1}(x_1) \cdots \rho_{X_n}(x_n) dx_1 \cdots dx_n.
\end{aligned}$$

Réciproquement, si on suppose que la densité a la forme produit, alors l'égalité précédente montre que $\mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n) = \mathbb{P}(X_1 \in I_1) \cdots \mathbb{P}(X_n \in I_n)$ quelque soient les intervalles $I_1, \dots, I_n$ et donc que les v.a. sont indépendantes. □

3.7.4 Espérance des produits de v.a indépendantes

Le théorème fondamental de cette section est le suivant :

Théorème 3.7.2. *Si $X_1, \dots, X_n$ est une famille de v.a indépendantes et dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ le produit $Y = X_1 \cdots X_n$ est également une v.a dans $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ et son espérance égale le produit des espérance des X_i :*

$$\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n).$$

Remarque. La simplification de la formule précédente est cruciale ! En effet, le membre de gauche est une somme n-uple si les variables sont discrètes et une intégrale n-uple si le vecteur $(X_1, \dots, X_n)$ est à densité, et par conséquent difficile à calculer. Le membre de droite est simplement un produit de sommes simples (pour des variables discrètes) ou un produit d'intégrales simples (pour les variables à densité) et est bien plus facile à calculer !

Preuve. Donnons d'abord une preuve de ce résultat quand les X_i sont à valeurs dans un ensemble fini ou dénombrable. Aussi, nous ne traitons que le cas $n = 2$ le cas général se faisant de la même façon. Supposons que X_1, X_2 sont à valeurs dans un ensemble fini E. Notons $f : \mathbb{R}^2 \to \mathbb{R}$ l'application définie par $f(x_1, x_2) = x_1 \cdot x_2$. On a d'après la formule de transfert

$$\begin{aligned}
\mathbb{E}(f(X_1 \cdot X_2)) &= \sum_{(e_1,e_2)\in E\times E} f(e_1, e_2)\mathbb{P}(X_1 = e_1, X_2 = e_2) \\
&= \sum_{(e_1,e_2)\in E\times E} e_1 e_2 \mathbb{P}(X_1 = e_1)\mathbb{P}(X_2 = e_2) \\
&= \left(\sum_{e_1\in E} e_1\mathbb{P}(X_1 = e_1)\right)\left(\sum_{e_2\in E} e_2\mathbb{P}(X_2 = e_2)\right) \\
&= \mathbb{E}(X_1)\mathbb{E}(X_2).
\end{aligned}$$

Dans le cas où les variables sont toutes admettent des densités continues, alors comme les variables sont indépendantes le vecteur $(X_1, \dots, X_n)$ a une densité égale au produit des densités des X_i et :

$$\begin{aligned}
\mathbb{E}(X_1 \cdot X_2 \cdots X_n) &= \int_{\mathbb{R}^n} x_1 x_2 \cdots x_n \rho_{X_1}(x_1)\rho_{X_2}(x_2)\cdots\rho_{X_n}(x_n)dx_1 dx_2 \cdots dx_n \\
&= \left(\int_{\mathbb{R}} x_1\rho_{X_1}(x_1)dx_1\right)\cdots\left(\int_{\mathbb{R}} x_n\rho_{X_n}(x_n)dx_n\right) \\
&= \mathbb{E}(X_1)\cdots\mathbb{E}(X_n).
\end{aligned}$$

Le résultat reste vrai dans le cas général où les X_i sont arbitraires et dans $\mathcal{L}^1$, mais la preuve dépasse le cadre du cours. □

Remarque. Attention, la réciproque du résultat précédent est fausse : si deux v.a X, Y sont telles que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ on ne peut pas conclure que X, Y sont indépendantes. En revanche, le résultat suivant est vrai :

Théorème 3.7.3. *Si $X_1, \dots, X_n$ sont des v.a telles que pour toutes fonctions continues bornées $\phi_1, \dots, \phi_n$ de $\mathbb{R} \to \mathbb{R}$ on a*

$$\mathbb{E}(\phi_1(X_1)\cdots\phi_n(X_n)) = \mathbb{E}(\phi_1(X_1))\cdots\mathbb{E}(\phi_n(X_n)),$$

alors la famille de v.a $X_1, \dots, X_n$ est indépendante.

3.7.5 Loi d'une somme de v.a.r. indépendantes

Nous pouvons à présent calculer la densité d'une suite de v.a.r. indépendantes en utilisant les résultats des sous-sections précédentes et le théorème 3.6.1

Théorème 3.7.4. *Soient X et Y deux v.a.r. indépendantes*

(a) Si X et Y sont des v.a.r. prenant leurs valeurs dans des ensembles finis ou dénombrables et admettent respectivement pour lois p_X et p_Y alors la loi p_{X+Y} de $X + Y$ vérifie pour tout e

$$p_{X+Y}(e) = \sum_{e_1+e_2=e} p_X(e_1)p_Y(e_2).$$

(b) Si X et Y sont des v.a.r. admettant respectivement pour densités ρ_X et ρ_Y alors la v.a.r. $X + Y$ admet pour densité ρ_{X+Y} :

$$\rho_{X+Y}(x) = \int_{\mathbb{R}} \rho_X(x - y)\rho_Y(y)dy.$$

Exemple 3.7.3. *Soient G_1 et G_2 deux v.a. indépendantes de même loi géométrique de paramètre θ : $\mathbb{P}(G_1 = k) = \mathbb{P}(G_2 = k) = \theta(1-\theta)^k$ pour $k \in \mathbb{N}$. Alors*

$$\mathbb{P}(G_1 + G_2 = n) = \sum_{k=0}^{n} \mathbb{P}(G_1 = k)\mathbb{P}(G_2 = n-k) = \sum_{k=0}^{n} \theta(1-\theta)^k \theta(1-\theta)^{n-k} = (n+1)\theta^2(1-\theta)^n.$$

Exemple 3.7.4. *Soient X et Y deux v.a. indépendantes, de loi de Poisson de paramètres respectifs λ et μ. Pour calculer la loi de $X+Y$, il suffit d'identifier $\mathbb{P}(X+Y=n)$ pour chaque $n \in \mathbb{N}$: soit donc $n \in \mathbb{N}$, comme X et Y sont indépendantes,*

$$\begin{aligned}
\mathbb{P}(X+Y=n) &= \sum_{k=0}^{n} \mathbb{P}(X=k) \cdot \mathbb{P}(Y=n-k) \\
&= \sum_{k=0}^{n} e^{-\lambda} \frac{\lambda^k}{k!} \cdot e^{-\mu} \frac{\mu^{n-k}}{(n-k)!} \\
&= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda^k \mu^{n-k} \\
&= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!},
\end{aligned}$$

par le binome de Newton. La v.a. $X+Y$ a donc une loi de Poisson de paramètre $\lambda+\mu$.

Exemple 3.7.5. *Soient E_1 et E_2 deux v.a. indépendantes de même loi exponentielle de paramètre λ, c'est-à-dire de densité $\rho(x) = \lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}^+}(x)$. On a*

$$\begin{aligned}
\rho_{E_1+E_2}(x) &= \int_{\mathbb{R}} \rho(y)\rho(x-y)dy \\
&= \int_{\mathbb{R}} \lambda e^{-\lambda y}\mathbf{1}_{\mathbb{R}^+}(y)\lambda e^{-\lambda(x-y)}\mathbf{1}_{\mathbb{R}^+}(x-y)dy \\
&= \int_{\mathbb{R}} \lambda^2 e^{-\lambda x} dy \mathbf{1}_{[0,x]}(y)\mathbf{1}_{\mathbb{R}^+}(x) \\
&= \lambda^2 x e^{-\lambda x}\mathbf{1}_{\mathbb{R}^+}(x).
\end{aligned}$$

Les fonctions génératrices peuvent être remarquablement efficaces pour identifier la loi d'une somme de variables aléatoires indépendantes. Prenons quelques exemples.

Exemple 3.7.6. *On peut reprendre l'exemple 3.7.4 à l'aide des fonctions génératrices. Soient X et Y deux v.a. de indépendantes de loi de Poisson de paramètres respectifs λ et μ.*

$$\phi_{X+Y}(t) = \mathbb{E}[t^{X+Y}] = \mathbb{E}[t^X t^Y] = \mathbb{E}[t^X]\mathbb{E}[t^Y] = e^{\lambda(t-1)}e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)},$$

on reconnait la fonction génératrice d'une variable aléatoire de Poisson de paramètre $\lambda+\mu$. Comme la fonction génératrice caractérise la loi, $X+Y$ est une v.a. de Poisson de paramètre $\lambda+\mu$.

Exemple 3.7.7. *Soient X et Y deux variables aléatoires indépendantes, toutes les deux de loi normale de paramètres $\mu = 0$ et $\sigma^2 = 1$. On cherche la loi de $Z = X+Y$. La variable aléatoire Z a une densité donnée par*

$$\begin{aligned}
\rho_Z(z) &= \int_{\mathbb{R}} \rho_X(x)\rho_Y(z-x)dx \\
&= \int_{\mathbb{R}} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \cdot \frac{e^{-(z-x)^2/2}}{\sqrt{2\pi}} dx.
\end{aligned}$$

On veut simplifier ce qu'il y a a dans l'exponentielle : on a

$$\begin{aligned}-\frac{x^2}{2}-\frac{(z-x)^2}{2} &= -\frac{x^2}{2}-\frac{z^2}{2}-\frac{x^2}{2}-xy\\ &= -\frac{z^2}{4}-\left(x-\frac{z}{2}\right)^2.\end{aligned}$$

Par conséquent,

$$\rho_Z(z)=\frac{e^{-z^2/4}}{\sqrt{2\pi}}\int_{\mathbb{R}}\frac{e^{-(x-z/2)^2}}{\sqrt{2\pi}}dx.$$

On reconnaît presque des densités de gaussiennes centrées, et on les réécrit de manière à avoir exactement des formules du type $e^{-s^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}$ (qui s'intègrent donc à 1 sur tout $\mathbb{R}$!). En faisant le changement de variable $y=x-z/2$, z étant fixé on a $dy=dz$ et $y\in\mathbb{R}$:

$$\begin{aligned}\rho_Z(z) &= \frac{e^{-z^2/4}}{\sqrt{2\pi 2}}\int_{\mathbb{R}}\frac{e^{-(x-z/2)^2/(2\cdot\frac{1}{2})}}{\sqrt{2\pi\frac{1}{2}}}dx\\ &= \frac{e^{-z^2/4}}{\sqrt{2\pi 2}}\int_{\mathbb{R}}\frac{e^{-y^2/(2\cdot\frac{1}{2})}}{\sqrt{2\pi\frac{1}{2}}}dy\\ &= \frac{e^{-z^2/4}}{\sqrt{2\pi 2}},\end{aligned}$$

et par conséquent, on se rend compte que la densité de Z est celle d'une variable de loi $\mathcal{N}(0,2)$.

Exercice 3.7.1. *Soit $X_1,X_2,\dots,X_n$ des variables aléatoires indépendantes, toute de loi de Bernoulli de paramètre $p\in[0,1]$. On définit $N=X_1+\cdots+X_n$. Montrer que N est une variable aléatoire de loi binomialde de paramètre n et p.*

Solution. On utilise la fonction génératrice. Pour tout $1\le i\le n$, on a $\phi_{X_i}(s)=1-p+ps$. De plus, comme les variables aléatoires $X_1,\dots,X_n$ sont indépendantes,

$$\phi_N(s)=\mathbb{E}(s^N)=\mathbb{E}(s^{X_1+\cdots+X_n})=\mathbb{E}(s^{X_1}\cdot\dots\cdot s^{X_n})=\mathbb{E}(s_1^X)\dots\mathbb{E}(s_n^X)=(1-p+ps)^n,$$

où on a utilisé le fait que si $X_1,\dots,X_n$ sont indépendantes alors $f_1(X_1),f_2(X_2),\dots,f_n(X_n)$ sont aussi indépendantes pour toutes fonctions $f_1,\dots,f_n$. On reconnait alors la fonction génératrice d'une variable aléatoire binomiale de paramètre n et p. Comme la fonction génératrice caractérise la loi, N est bien une variable aléatoire binomiale(n,p). △

3.7.6 Variance d'une somme de v.a.r. indépendantes

Si $(X_i)_{i\in\mathbb{N}}$ est une famille de v.a.r. on a toujours

$$\mathbb{E}(X_1+\cdots+X_n)=\mathbb{E}(X_1)+\cdots+\mathbf{E}(X_n).$$

parce que l'espérance est linéaire. La variance n'est elle pas linéaire, et en général on n'a pas d'équivalent pour $\mathbf{Var}(X_1+X_2+\cdots+X_n)$. Cependant, si la famille $(X_1,\dots,X_n)$ est indépendante on peut calculer facilement la variance de $X_1+\cdots+X_n$.

Théorème 3.7.5. *Soit $(X_i)_{i\in\mathbb{N}}$ une famille de v.a.r. Alors,*

$$\mathbf{Var}(X_1+\cdots+X_n)=\sum_{i=1}^n\mathbf{Var}(X_i)+2\sum_{1\le i<j\le n}\mathbf{Cov}(X_i,X_j),$$

où $\mathbf{Cov}(X_i,X_j)=\mathbb{E}(X_iX_j)-\mathbb{E}(X_i)\mathbb{E}(X_j)$ est la covariance de X_i et X_j. Dans le cas où les variables sont indépendantes, on a :

$$\mathbf{Var}(X_1+\cdots+X_n)=\mathbf{Var}(X_1)+\cdots+\mathbf{Var}(X_n).$$

Preuve. Soit $S := X_1 + \cdots + X_n$; il suffit de calculer $\mathbf{Var}(S) = \mathbb{E}(S^2) - (\mathbb{E}(S))^2$. On a

$$\mathbb{E}[S^2] = \mathbb{E}\Big[\sum_{1\le i,j\le n} X_i \cdot X_j\Big] = \sum_{1\le i,j\le n} \mathbb{E}(X_iX_j).$$

D'autre part,

$$(\mathbb{E}[S])^2 = \Big(\sum_{i=1}^n \mathbb{E}(X_i)\Big)^2 = \sum_{1\le i,j\le n} \mathbb{E}(X_i)\cdot\mathbb{E}(X_j).$$

On en déduit que

$$\begin{aligned}
\mathbf{Var}(S) &= \mathbb{E}(S^2) - (\mathbb{E}(S))^2 \\
&= \sum_{1\le i,j\le n} \Big\{\mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)\Big\} \\
&= \sum_{i=1}^n \Big\{\mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2\Big\} + \sum_{1\le i\ne j\le n} \Big\{\mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)\Big\} \\
&= \sum_{i=1}^n \mathbf{Var}(X_i) + 2\sum_{1\le i<j\le n} \mathbf{Cov}(X_i, X_j).
\end{aligned}$$

Dans le cas où les variables $X_1, \ldots, X_n$ sont indépendantes, alors $\mathbf{Cov}(X_i, X_j) = \mathbb{E}[X_iX_j] - \mathbb{E}(X_i)\mathbb{E}(X_j) = 0$ quelque soient i et j et on a alors :

$$\mathbf{Var}(S) = \sum_{i=1}^n \mathbf{Var}(X_i). \qquad \square$$

Exercice 3.7.2. *a) Calculer l'espérance et la variance d'une v.a.r. suivant une loi de Bernoulli* $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$.
b) On suppose que les v.a.r. $(X_i)_{i\in\mathbb{N}}$ *sont indépendantes et suivent chacune une loi de Bernoulli de paramètre* p. *Calculer l'espérance et la variance de* $X_1 + \cdots + X_n$.
c) Comparer avec l'espérance et la variance d'une loi binomiale (n, p).

Remarque. 1) On se rend facilement compte que les variables aléatoires n'ont pas besoin d'être indépendantes pour que $\mathbf{Var}(X_1 + \cdots + X_n)$ soit égal à $\mathbf{Var}(X_1) + \cdots + \mathbf{Var}(X_n)$. En effet, il suffit que pour toute paire de variables X_i, X_j, on ait $\mathbb{E}(X_iX_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$, ce qui est vrai dès que les $X_1, X_2, \ldots, X_n$ sont indépendantes deux-à-deux, ce qui est une contrainte beaucoup plus faible que l'indépendance de la famille $(X_1, \ldots, X_n)$.
2) Encore une fois : si les variables X et Y sont indépendantes, alors $\mathbf{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$; cependant, la réciproque est bien sûr fausse ! Il ne suffit pas que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ pour que les variables X et Y soient indépendantes. Voir le théorème 3.7.3.

3.7.7 Critères d'indépendance

Dans la pratique il est important de déterminer si une famille de v.a est indépendante. Un cas qui se présente fréquemment est le suivant : on suppose donnée une famille de v.a indépendantes $X_1, X_2, \ldots$ et on construit, à partir des X_i, de nouvelles v.a $Y_1, Y_2, \ldots$. Par exemple, on peut définir $Y_1 = X_1$, $Y_2 = X_1 + X_2$, et de manière générale $Y_n = X_1 + \cdots + X_n$; on pourrait aussi définir les Y_i par $Y_1 = X_1 + X_2$, $Y_2 = X_2 + X_3$, $Y_n = X_n + X_{n+1}$ etc. ou faire des choses plus compliquées. On se propose alors de savoir si la famille ainsi construite est indépendante. Le théorème qui suit permet dans certains cas de répondre à cette question.

Théorème 3.7.6. *Soient $(X_i)_{i\in\mathbb{N}}$ une famille de v.a indépendantes et (J_i) $(i = 1, 2, \ldots)$ des sous ensembles finis de $\mathbb{N}$ qui forment une partition de $\mathbb{N}$ (i.e les J_i sont non vides, $\mathbb{N} = \cup_{i\geq 1} J_i$ et $J_i \cap J_j = \emptyset$ si $i \neq j$). Supposons données des applications (continues, continues par morceaux,...) f_i de $\mathbb{R}^{\#J_i} \to \mathbb{R}$ et posons $Y_i = f_i(X_{i_1}, \ldots, X_{i_{\#J_i}})$ (où $i_1 < \cdots < i_{\#J_i}$ sont les éléments de J_i). Alors, la famille de v.a $Y_1, Y_2, \ldots$ est indépendante.*

Ainsi, si la famille de v.a.r. $X_1, X_2, \ldots$ est indépendante, il en est de même de $X_1 + X_2, X_3 + X_4, \ldots, X_{2n-1} + X_{2n}, \ldots$. En revanche, la suite de v.a.r. $X_1 + X_2, X_2 + X_3, X_3 + X_4, \ldots$ ne sera en général pas indépendante.

Exemple 3.7.8. *Considérons jeu de n pile ou face et les variables $X_1, X_2, \ldots, X_n$: on a $\Omega = \{0, 1\}^n$ et pour $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$, on a $X_i(\omega) = \omega_i$. On suppose que $\mathbb{P}$ est la probababilité uniforme sur Ω ; lest varaibles $X_1, \ldots, X_n$ sont alors indépedendantes. Soit $m \in \mathbb{N}$ et pour $0 < r < m$ l'ensemble $J_r = \{1 \leq k \leq n : n = r \mod m\}$. Alors les J_r sont disjoints par définition de la division euclidienne et les variables aléatoires*

$$f_0\Big(\sum_{i\in J_0} X_i\Big), \ldots, f_{m-1}\Big(\sum_{i\in J_{m-1}} X_i\Big)$$

sont indépendantes quelque soit les fonctions $f_1, f_2, \ldots, f_m$.

Pour être plus concret : les variables

$$S_1 = \sum_{1\leq i\leq n, i \text{ pair}} X_i \qquad \text{et} \qquad S_2 = \sum_{1\leq i\leq n, i \text{ impair}} X_i$$

sont indépendantes.

Chapitre 4

Convergence de suites de v.a.

Soient $(X_i)_{i\geq 1}$ une famille de v.a indépendantes et introduisons les v.a. $S_n = X_1 + \cdots + X_n$ pour $n \geq 1$. Si par exemple les v.a. X_i ont même loi, l'intuition que nous avons des probabilités nous incite à penser que les moyennes

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

convergent quand n tend vers l'infini vers l'espérance $\mathbb{E}(X_1)$ de X_1. En effet, si par exemple les X_i suivent une loi de Bernoulli de paramètre $\frac{1}{2}$ et forment une famille indépendante de v.a — les X_i modélisent donc un jeu infini de Pile ou Face où les tirages sont indépendants — l'expérience ou l'intuition indique que

$$\frac{1}{n}(X_1 + \cdots + X_n)$$

convergent vers le nombre (ou la v.a constante) $1/2$ qui n'est rien d'autre que l'espérance de X_1. La première difficulté que nous rencontrons est de donner un sens à la convergence précédente.

4.1 Notions de convergences de variables aléatoires

Rappelons que les suites variables aléatoires sont des fonctions de Ω dans $\mathbb{R}$. Une version naturelle de convergence pour les suites de fonctions est la convergence simple (ponctuelle) : une suite de fonctions f_n converge simplement vers une fonction f sur un domaine D si pour tout $x \in D$ on a $f_n(x) \to f(x)$ lorsque $n \to \infty$. La convergence presque sûre est une version probabiliste qui relaxe la convergence en tout point x en une convergence en presque tout point dans le sens suivant :

Définition 4.1.1 (Convergence presque sûre)**.** *On dit que la suite de v.a. Y_n converge $\mathbb{P}$-presque sûrement vers la v.a. Y si l'ensemble (il s'agit en fait d'un événement) des $\omega \in \Omega$ pour lesquels la suite $(Y_n(\omega))_n$ converge vers $Y(\omega)$ est de probabilité 1. C'est-à-dire :*

$$\mathbb{P}(\lim_n Y_n = Y) = \mathbb{P}\big(\{\omega \in \Omega : \lim_n Y_n(\omega) = Y(\omega)\}\big) = 1.$$

Même si elle est particulièrement naturelle et simple à énoncé cette notion de convergence peut être difficile à manipuler puisqu'il faut considérer un événement complexe $\{\lim_n Y_n = Y\}$. On peut essayer plutôt de regarder la probabilité que Y_n diffère de Y, puis prendre la limite de ses probabilités. Ceci conduit à la définition suivante :

Définition 4.1.2 (Convergence en probabilité)**.** *On dit que la suite de v.a $(Y_n)_n$ converge en probabilité vers la v.a Y si pour tout $\epsilon > 0$ on a*

$$\lim_{n \to infty} \mathbb{P}(|Y_n - Y| > \epsilon) = 0.$$

La convergence en probabilité est bien plus simple : il s'agit de regarder la probabilité d'un événement assez simple $|Y_n - Y| > \epsilon$ et de quantifier sa probabilité lorsque $n \to \infty$. On peut alors se demander si les deux notions sont équivalentes.

Proposition 4.1.1. *Si une famile de variables aléatoire $(X_n)_{n\geq 1}$ converge $\mathbb{P}$-presque sûrement vers une variable aléatoire X alors $(X_n)_{n\geq 1}$ converge en probabilité vers X.*

Preuve. Soit $(X_n)_{n\geq 1}$ qui converge presque sûrement vers X, c'est-à-dire que l'événement

$$C = \{\omega \in \Omega : X_n(\omega) \to X(\omega)\}$$

est tel que $\mathbb{P}(C) = 1$. Cependant, une suite numérique u_n converge vers une limite ℓ si et seulement si pour tout $\epsilon > 0$ il existe un $n_0 = n_0(\epsilon)$ tel que pour tout $n \geq n_0$, $|u_n - \ell| < \epsilon$. Autrement dit, $u_n \to \ell$ si pour tout $\epsilon > 0$, $|u_n - \ell| > \epsilon$ pour seulement un nombre fini de valeurs de n (qui peut dépendre de ϵ). Soit donc

$$A_n(\epsilon) = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}$$

On vérifie aisément que

$$A(\epsilon) = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon \text{ pour une infinité de valeurs de } n\} = \bigcap_{k\geq 1} \bigcup_{n\geq k} A_n(\epsilon).$$

Si $X_n(\omega) \to X(\omega)$ alors $\omega \notin A(\epsilon)$ pour tout $\epsilon > 0$ et donc $\mathbb{P}(A(\epsilon)) = 0$ dès que $\mathbb{P}(C) = 1$. D'autre part, comme $A(\epsilon) \subseteq A(\epsilon')$ si $\epsilon > \epsilon'$,

$$\mathbb{P}(C^c) = \mathbb{P}(\exists \epsilon > 0 : A(\epsilon)) = \mathbb{P}\Big(\bigcup_{\epsilon>0} A(\epsilon)\Big) = \mathbb{P}\Big(\bigcup_{i\geq 1} A(1/i)\Big) \leq \sum_{i\geq 1} \mathbb{P}(A(1/i))$$

et donc $\mathbb{P}(C^c) = 0$ si $\mathbb{P}(A(\epsilon)) = 0$ pour tout $\epsilon > 0$. Nous avons donc montré que $\mathbb{P}(C) = 1$ si et seulement si $\mathbb{P}(A(\epsilon)) = 0$ pour tout $\epsilon > 0$.

La suite d'événements $B_k(\epsilon) = \bigcup_{n\geq k} A_k(\epsilon)$ est une suite décroissante d'événements. Par conséquent,

$$\mathbb{P}(A(\epsilon)) = \mathbb{P}\Big(\bigcap_{k\geq 1} \bigcup_{n\geq k} A_n(\epsilon)\Big) = \mathbb{P}\Big(\bigcap_{k\geq 1} B_k(\epsilon)\Big) = \lim_{k\to\infty} \mathbb{P}(B_k(\epsilon)).$$

Si X_n converge presque sûrement vers X alors $\mathbb{P}(A_n(\epsilon)) = 0$ pour tout $\epsilon >$) ce qui implique, comme $A_n(\epsilon) \subseteq B_n(\epsilon)$,

$$\mathbb{P}(A_n(\epsilon)) \leq \mathbb{P}(B_n(\epsilon)) \to \mathbb{P}(A(\epsilon)) = 0,$$

qui qui montre que X_n converge aussi en probabilité. □

Proposition 4.1.2. *La convergence en probabilité n'implique pas la convergence presque sûre.*

Preuve. Il s'agit de trouver un contre-exemple. Pour $k \geq 1$ on considère la partition de $[0, 1[$ en 2^k intervalles $A_{k,i} = [i2^{-k}, (i+1)2^{-k})$, $i = 0, 1, 2 \ldots, 2^k - 1$. On se donne une suite de variables aléatoire U de loi uniforme dans $[0, 1]$. Pour $n \geq 1$, il existe une unique paire (k, i) avec $n = 2^{k-1} + i$ avec $0 \leq i < 2^k$ et on définit $X_n = \mathbf{1}_{U\in A_{k,i}}$. Alors X_n converge en probabilité vers 0 : en effet, pour tout $\epsilon > 0$,

$$\mathbb{P}(|X_n| > \epsilon) = \mathbb{P}(X_n = 1) = \mathbb{P}(U \in A_{k,i}) = 2^{-k} \leq \frac{1}{n} \xrightarrow[n\to\infty]{} 0.$$

Cependant, pour tout k, on a évidemment

$$\sum_{i=0}^{2^k-1} X_i = \sum_{i=0}^{2^k-1} \mathbf{1}_{U\in A_{k,i}} = \mathbf{1}_{U\in[0,1[} = 1,$$

et par conséquent, pour tout n_0 il existe $n \geq n_0$ tel que $X_n = 1$. En d'autres terme, quelque soit ω, la suite $X_n(\omega)$ ne peut pas converger vers 0 ! Il s'ensuit que $\mathbb{P}(\lim_n X_n = 0) = 0$. □

Exemple 4.1.1. *Voici un deuxième exemple de suite de variables aléatoires qui converge en probabilités mais pas presque sûrement. Soit $(X_n)_{n\geq 1}$ une famille indépendante de variables aléatoires telles que X_n suive une loi de Bernoulli de paramètre $1/n$:*

$$\mathbb{P}(X_n = 1) = \frac{1}{n} \qquad \text{et} \qquad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$

Alors X_n converge vers 0 en probabilité : pour tout $\epsilon \in]0,1[$, $\mathbb{P}(|X_n| > \epsilon) = \mathbb{P}(X_n = 1) = 1/n \to 0$. Cependant, avec $A_n(\epsilon) = \{|X_n| > \epsilon\}$ et $B_k(\epsilon) = \cup_{n\geq k} A_n(\epsilon)$, quelque soit $0 < \ell < 1$, en utilisant l'indépendance des (X_i),

$$\mathbb{P}(B_k(\epsilon)^c) = \mathbb{P}\Big(\bigcap_{n\geq k} A_n(\epsilon)^c\Big) \leq \mathbb{P}\Big(\bigcap_{n=k}^{k+\ell} A_n(\epsilon)^c\Big) = \prod_{n=k}^{k+\ell} \mathbb{P}(A_n(\epsilon)^c) = \prod_{n=k}^{k+\ell} \frac{n-1}{n} = \frac{k-1}{k+\ell}.$$

Comme $\ell \geq 1$ est arbitraire, on a

$$\mathbb{P}(B_k(\epsilon)^c) \leq \lim_{\ell\to\infty} \frac{k-1}{k+\ell} = 0.$$

Par conséquent, $\mathbb{P}(B_k(\epsilon)) = 1$ pour tout $k \geq 1$. Mais alors

$$\begin{aligned}
\mathbb{P}(|X_n| > \epsilon \text{ pour une infinité de valeurs de } n) &= \mathbb{P}\Big(\bigcap_{k\geq 1} B_k(\epsilon)\Big) \\
&= 1 - \mathbb{P}\Big(\bigcup_{k\geq 1} B_k(\epsilon)^c\Big) \\
&\geq 1 - \sum_{k\geq 1} \mathbb{P}(B_k(\epsilon)^c) = 1,
\end{aligned}$$

et donc avec probabilité 1, X_n ne converge pas vers 0.

4.2 Lois des grands nombres

4.2.1 Loi forte des grands nombres

Le théorème fondamental de ce chapitre et qui est à la base de la théorie mathématique des probabilités est *la loi forte des grands nombres* qui est

Théorème 4.2.1. *Soit $X_1, \ldots, X_n, \ldots$ une famille de v.a indépendante où les X_i ont même loi et sont dans $\mathcal{L}^1(\Omega, \mathbb{P})$. Alors, la suite de v.a*

$$\frac{1}{n}(X_1 + \cdots + X_n)$$

converge $\mathbb{P}$-p.s vers la v.a constante $\mathbb{E}(X_1)$.

La démonstration de ce théorème fondamental dans cette généralité sort du cadre de ce cours, mais nous allons en donner une preuve sous des hypothèses plus faibles.

4.2.2 Loi faible des grands nombres

Avant de montrer une convergence presque sûre, nous allons affaiblir l'énoncé et ne montrer qu'une convergence en probabilité : la *loi faible des grands nombres* s'obtient très facilement à l'aide de l'inégalité de Bienaymé-Tchebyshev.

Théorème 4.2.2. *Si $(X_i)_{i\geq 1}$ est une famille de v.a indépendante et si les X_i ont même loi et sont dans $\mathcal{L}^2$ alors la suite S_n/n converge en probabilité vers $\mathbb{E}(X_1)$.*

Preuve. Nous savons déjà d'après la linéarité de l'espérance que

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)$$

et comme les variables sont indépendantes, on a aussi

$$\mathbf{Var}(S_n) = \mathbf{Var}(X_1) + \cdots + \mathbf{Var}(X_n).$$

Puisque les X_i ont même loi, on obtient

$$\mathbb{E}(S_n) = n\mathbb{E}(X_1) \qquad \text{et} \qquad \mathbf{Var}(S_n) = nVar(X_1).$$

Le fait que la variance de la somme des n v.a $X_1, \ldots, X_n$ se comporte comme n et non pas comme n^2 (c'est ici où intervient l'hypothèse d'indépendance) est l'observation fondamentale. Appliquons en effet l'inégalité de Bienaymé–Tchebychev :

$$\mathbb{P}(|S_n - n\mathbb{E}(X_1)| > \lambda) \leq \frac{n\mathbf{Var}(X_1)}{\lambda^2},$$

ce qui peut s'écrire

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| > \frac{\lambda}{n}\right) \leq \frac{n\mathbf{Var}(X_1)}{\lambda^2}.$$

Posons à présent $\lambda = n\epsilon$ où ϵ est un réel positif :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| > \epsilon\right) \leq \frac{\mathbf{Var}(X_1)}{n\epsilon^2}.$$

Nous voyons donc que pour tout $\epsilon > 0$ on a

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(X_1)\right| > \epsilon\right) = 0,$$

ce qui prouve le résultat. □

4.2.3 Preuve de la loi forte des grands nombres – le cas $\mathcal{L}^4$

Nous allons maintenant montrer une version de la loi forte des grands nombres pour les variables aléatoires $(X_i)_{i\geq 1}$ qui sont dans $\mathcal{L}^4$, c'est-à-dire que $\mathbb{E}(X_i^4) < \infty$:

Théorème 4.2.3. *Soit $X_1, \ldots, X_n, \ldots$ une famille de v.a.r. indépendante où les X_i ont même loi et sont dans $\mathcal{L}^4(\Omega, \mathbb{P})$. Alors, la suite de v.a*

$$\frac{1}{n}(X_1 + \cdots + X_n)$$

converge $\mathbb{P}$-p.s vers la v.a. constante $\mathbb{E}(X_1)$.

Preuve. i) Remarquons que l'hypothèse $\mathbb{E}(X_i^4) < \infty$ entraine que

$$\forall k = 0, 1, 2, 3, 4, \ \mathbb{E}(|X_i|^k) < \infty. \tag{4.1}$$

Pour $k = 0$ c'est évident. On a d'après l'inégalité de Cauchy–Schwarz (cf. 3.3.4)

$$\mathbb{E}(X_i^2 \cdot \mathbf{1}) \leq \mathbb{E}(X_i^4)^{1/2}\mathbb{E}(\mathbf{1}^2)^{1/2} = \mathbb{E}(X_i^4)^{1/2} < \infty.$$

On a donc (4.1) pour $k = 4$ et $k = 2$. En appliquant à nouveau Cauchy-Schwarz on a $\mathbb{E}(|X_i \cdot \mathbf{1}|) \leq \mathbb{E}(X_i^2)^{1/2}\mathbb{E}(\mathbf{1}^2)^{1/2} < \infty$; ainsi (4.1) est vraie pour $k = 1$. Enfin, $\mathbb{E}(|X_i|^3) = \mathbb{E}(X_i^2|X_i|) \leq \mathbb{E}(X_i^4)^{1/2}\mathbb{E}(X_i^2)^{1/2} < \infty$. On a donc bien établi (4.1) pour $k = 0, 1, 2, 3, 4$.

ii) Posons $\bar{X}_k = X_k - \mathbb{E}(X_k)$ et $\bar{S}_n = \bar{X}_1 + \cdots + \bar{X}_k$. La famille de v.a $(\bar{X}_i)_i$ est indépendante et les $\bar{X}_i$ sont de même loi et d'espérance nulle : $\mathbb{E}(\bar{X}_i) = 0$. Vérifions que

$$\forall k = 0, 1, 2, 3, 4,\ \mathbb{E}(|\bar{X}_i|^k) < \infty. \tag{4.2}$$

Pour cela, il suffit de constater que $\mathbb{E}(|\bar{X}|^k) \leq \mathbb{E}\big((|X| + \mathbb{E}(|X|))^k\big)$ et que cette dernière quantité est une combinaison linéaire de terme de la forme $\mathbb{E}(|X_i|^l)\mathbb{E}(|X|)^m$ pour $l+m = k$, $0 \leq l, m \leq k$. On utilise alors pour conclure (4.1).

iii) Posons $\bar{S}_n = \sum_{i=1}^n \bar{X}_i$ et remarquons que

$$\bar{S}_n^4 = \left(\sum_{i=1}^n \bar{X}_i\right)^4 = \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \bar{X}_{i_1}\bar{X}_{i_2}\bar{X}_{i_3}\bar{X}_{i_4}$$

et donc

$$\mathbb{E}(\bar{S}_n^4) = \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} \mathbb{E}(\bar{X}_{i_1}\bar{X}_{i_2}\bar{X}_{i_3}\bar{X}_{i_4}). \tag{4.3}$$

On constate à présent que si les indices i_r, $r = 1, 2, 3, 4$ sont distincts deux à deux on a d'après l'indépendance des $\bar{X}_i$ et le fait que $\mathbb{E}(\bar{X}_i) = 0$

$$\mathbb{E}(\bar{X}_{i_1}\bar{X}_{i_2}\bar{X}_{i_3}\bar{X}_{i_4}) = \mathbb{E}(\bar{X}_{i_1})\mathbb{E}(\bar{X}_{i_2})\mathbb{E}(\bar{X}_{i_3})\mathbb{E}(\bar{X}_{i_4}) = 0$$

De la même manière si un indice est différent des trois autres, $\mathbb{E}(\bar{X}_{i_1}\bar{X}_{i_2}\bar{X}_{i_3}\bar{X}_{i_4}) = 0$; en effet si par exemple cet indice est i_1, $\bar{X}_{i_1}$ est indépendant de $\bar{X}_{i_2}\bar{X}_{i_3}\bar{X}_{i_4}$ et donc

$$\mathbb{E}(\bar{X}_{i_1}\bar{X}_{i_2}\bar{X}_{i_3}\bar{X}_{i_4}) = \mathbb{E}(\bar{X}_{i_1})\mathbb{E}(\bar{X}_{i_2}\bar{X}_{i_3}\bar{X}_{i_4}) = 0 \times \mathbb{E}(\bar{X}_{i_2}\bar{X}_{i_3}\bar{X}_{i_4}) = 0.$$

Cela montre que les seuls termes qui contribuent à la somme (4.3) sont les indices tels que $\#\{i_1, i_2, i_3, i_4\}$ égale 1 ou 2. Il est alors facile de voir que

$$\mathbb{E}(\bar{S}_n^4) = \binom{4}{2} \sum_{1 \leq i < j \leq n} \mathbb{E}(\bar{X}_i^2\bar{X}_j^2) + \sum_{1 \leq i \leq n} \mathbb{E}(\bar{X}_i^4)$$

et donc en utilisant le fait qu'il y a $n(n-1)/2$ termes dans la première somme du membre de droite de cette inégalité et que $\mathbb{E}(\bar{X}_i^2\bar{X}_j^2) = \mathbb{E}(\bar{X}_i^2)\mathbb{E}(\bar{X}_j^2) = \mathbb{E}(\bar{X}_1^2)^2$ (les v.a.r. $\bar{X}_i^2$ et $\bar{X}_j^2$ sont indépendantes si $i \neq j$) on a

$$\mathbb{E}(\bar{S}_n^4) \leq 3n^2\mathbb{E}(\bar{X}_1^2)^2 + n\mathbb{E}(\bar{X}_1^4).$$

Par conséquent,

$$\mathbb{E}\left(\left(\frac{\bar{S}_n}{n}\right)^4\right) = \frac{\mathbb{E}(\bar{S}_n^4)}{n^4} \leq \frac{3\mathbb{E}(\bar{X}_1^2)}{n^2} + \frac{\mathbb{E}(\bar{X}_1^4)}{n^3}.$$

iv) La dernière inégalité montre que

$$\sum_{n=1}^{\infty} \mathbb{E}\left(\left(\frac{\bar{S}_n}{n}\right)^4\right) < \infty.$$

Mais on peut démontrer (c'est le théorème de convergence monotone) que

$$\mathbb{E}\left(\sum_{n=1}^{\infty}\left(\frac{\bar{S}_n}{n}\right)^4\right) = \sum_{n=1}^{\infty} \mathbb{E}\left(\left(\frac{\bar{S}_n}{n}\right)^4\right) < \infty.$$

On a donc

$$\mathbb{E}\left(\sum_{n=1}^{\infty}\left(\frac{\bar{S}_n}{n}\right)^4\right) < \infty.$$

Mais si l'espérance d'une v.a.r. positive est finie, cette v.a.r. est finie $\mathbb{P}$-presque sûrement. Par conséquent, $\mathbb{P}$-p.s. la somme

$$\sum_{n=1}^{\infty}\left(\frac{\bar{S}_n}{n}\right)^4$$

converge et en particulier

$$\frac{\bar{S}_n}{n}$$

converge $\mathbb{P}$-p.s. vers 0. Mais, si on se rappelle la définition de $\bar{S}_n$, cela est équivalent au fait que

$$\lim_{n\to\infty}\frac{S_n}{n} = \mathbb{E}(X_1)$$

$\mathbb{P}$-presque sûrement. □

4.3 Théorème de la limite centrale

L'idée générale du théorème de la limite centrale est d'obtenir des information plus précise sur les variables aléatoires $S_n = X_1 + X_2 + \cdots + X_n$. En effet, en restant informel, les lois des grands nombres disent que S_n est proche de $n\mathbb{E}(X_1)$. Mais alors, que peut-on dire de $S_n - n\mathbb{E}(X_1)$? Est-ce qu'on peut savoir son ordre de grandeur ? Et quand quel sens ?

Nous donnons une première version de ce théorème :

Théorème 4.3.1. *Soient $X_1, \ldots, X_n, \ldots$ une famille de v.a.r. indépendantes et supposons que les X_i soient de même loi et dans $\mathcal{L}^2(\Omega, \mathbb{P})$ (c'est-à-dire de carré intégrable, $\mathbb{E}(X_i^2) < \infty$). Notons $\mu = \mathbb{E}(X_1)$ et $\sigma^2 = \mathbf{Var}(X_1)$ (comme les X_i sont de même loi $\mathbb{E}(X_i) = \mu$ et $\mathbf{Var}(X_i) = \sigma$ pour tout i). Alors, pour tout intevalle I de $\mathbb{R}$*

$$\lim_{n\to\infty}\mathbf{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \in I\right) = \int_I \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx.$$

Une autre façon d'énoncer ce théorème est de dire que

$$\lim_{n\to\infty}\mathbf{P}\left(\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right) \in I\right) = \int_I \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx.$$

En d'autres termes, $S_n/n - \mu$ converge d'après la loi forte des grands nombres vers 0, la déviation des moyennes S_n par rapport à l'espérance, "renormalisée" par le facteur $\sqrt{n}/\sigma$, converge dans un certain sens vers une loi gaussienne normalisée (espérance nulle, variance égale à 1).

Avant de passer à la preuve de ce théorème, nous devons faire quelques rappels et introduire quelques notions utiles.

4.3.1 Convergence en loi

Définition 4.3.1. *On dit qu'une suite de v.a. $(Y_n)_{n\in\mathbb{N}}$ converge en loi vers une v.a Y si et seulement si pour toute fonction continue bornée $f : \mathbb{R} \to \mathbb{R}$*

$$\lim_{n\to\infty}\mathbb{E}(f(Y_n)) = \mathbb{E}(f(Y)).$$

L'intérêt de cette notion réside dans la proposition suivante (admise) :

Théorème 4.3.2. *Soient $(Y_n)_{n\in\mathbb{N}}$ une suite de v.a. et Y une v.a. et notons $F_{Y_n} : \mathbb{R} \to [0,1]$ et $F_Y : \mathbb{R} \to [0,1]$ les fonctions de répartition des v.a. Y_n, Y : c'est-à-dire*

$$F_{Y_n}(t) = \mathbb{P}(Y_n \leq t) \qquad \text{et} \qquad F_Y(t) = \mathbb{P}(Y \leq t).$$

La suite $(Y_n)_{n\in\mathbb{N}}$ converge en loi vers Y si et seulement si en tout point t_0 où F_Y est continue à gauche

$$\lim_{n\to\infty}F_{Y_n}(t_0) = F_Y(t_0).$$

Un corollaire utile du théorème précédent est le suivant

Corollaire 4.3.1. *Si la suite de v.a. $(Y_n)_{n\geq 0}$ converge en loi vers une v.a. Y qui admet une densité ρ_Y alors pour tout intervalle I de $\mathbb{R}$*

$$\lim_{n\to\infty} \mathbb{P}(Y_n \in I) = \int_I \rho_Y(y)dy$$

Preuve. En effet, dans ce cas F_Y est continue en tout point. □

4.3.2 Fonctions caractéristiques

Définition 4.3.2. *Si Y est une v.a.r, la fonction caractéristique de Y est la fonction continue $\varphi_Y : \mathbb{R} \to \mathbb{R}$ définie par*

$$\varphi_Y(t) = \mathbb{E}(e^{itY}),$$

(où $i = \sqrt{-1} \in \mathbb{C}$).

Remarque. i) Pour t fixé la v.a e^{itY} est bornée par 1 (puisque Y est à valeurs réelles) et est donc intégrable.
ii) La fonction caractéristique d'une v.a ne dépend que de la *loi* de cette v.a.

On peut préciser le résultat de continuité précédent (admis) :

Proposition 4.3.1. *Si Y est une v.a.r intégrable, alors la fonction caractéristique de Y est de classe C^1 (dérivable et de dérivée continue) et*

$$\varphi'_Y(t) = \mathbb{E}\big((iY)e^{itY}\big).$$

De même, si Y est dans $\mathcal{L}^p(\Omega, \mathbb{P})$ la fonction caractéristique de Y est de classe C^p et on a

$$\frac{d^p}{dt^p}\varphi_Y(t) = \mathbb{E}\big((iY)^p e^{itY}\big).$$

Exercice 4.3.1. *Montrer que si $Z = aY + b$ alors*

$$\varphi_Z(t) = e^{itb}\varphi_Y(ta).$$

Calculons à présent les fonctions caractéristiques de certaines lois classiques.

cas des v.a. discrètes Si Y prend un nombre fini de valeurs $y_1, \ldots, y_r$ et si on note $p_r = \mathbb{P}(Y = y_r)$ on a

$$\begin{aligned}\varphi_Y(t) &= \mathbb{E}(e^{itY}) \\ &= \sum_{k=1}^{r} e^{ity_k}\mathbb{P}(Y = y_k) \\ &= \sum_{k=1}^{r} (e^{it})^{y_k}\mathbb{P}(Y = y_k)\end{aligned}$$

et on reconnait (si Y est à valeurs entières) la fonction génératrice de Y au point e^{it}. Le calcul des fonctions caractéristiques de v.a discrètes est exactement le même que celui que nous avons effectué dans un chapitre précédent. De manière générale, si Y prend un nombre dénombrable de valeurs $y_1, y_2, \ldots$ alors

$$\sum_{k\geq 1} e^{iy_k}\mathbb{P}(Y = y_k)$$

est absolument convergente puisque $|e^{iy_k}| = 1$ pour tout $k \geq 1$. On dit que la variable aléatoire complexe e^{iY} est dans L^1 : $\mathbb{E}(|e^{iY}|) = 1$. On a alors

$$\varphi_Y(t) = \mathbb{E}(e^{iY}) = \sum_{k\geq 1} e^{iy_k}\mathbb{P}(Y = y_k) = \phi_Y(e^{it}).$$

cas des v.a. admettant une densité Si Y admet la densité ρ_Y alors

$$\begin{aligned}\varphi_Y(t) &= \mathbb{E}(e^{itY}) \\ &= \int_{-\infty}^{\infty} e^{ity}\rho_Y(y)dy.\end{aligned}$$

On habituellement $\hat{\rho}_Y(t) = \int_{-\infty}^{\infty} e^{ity}\rho_Y(y)dy$ et on dit que la fonction $\hat{\rho}_Y$ est la *transformée de Fourier* de la fonction ρ_Y. Encore une fois, on vérifie aisément que l'intégrale est bien définie : en effet, elle est absolument convergente puisque $|e^{ity}\rho_Y(y)| = \rho_Y(y)$ et $\int \rho_Y(y)dy = 1$.

Exemple 4.3.1. *Fonctions caractéristique d'une gaussienne Rappelons que si Z est une v.a suivant une loi gaussienne $\mathcal{N}(\mu, \sigma)$ on peut l'écrire sous la forme $Z = \sigma Y + \mu$ où Y suit une loi gaussienne normalisée $\mathcal{N}(0, 1)$ de densité*

$$\rho(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}.$$

On a donc

$$\varphi_Y(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ity}e^{-y^2/2}dy.$$

On a (cf. exercice) :

$$\varphi_Y(t) = e^{-t^2/2}.$$

Ainsi, la fonction d'une caractéristique d'une v.a suivant une loi gaussienne $\mathcal{N}(\mu, \sigma)$ est

$$\varphi_Z(t) = e^{it\mu-\sigma^2(t^2/2)}.$$

Exercice 4.3.2. *Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$. On souhaite montrer que sa fonction caractéristique est donnée par $\varphi(t) = \phi_X(t) = e^{-t^2}$. On suit le chemin suivant :*
a) *Montrer que $\varphi'(t) = -t\varphi(t)$ pour tout $t \in \mathbb{R}$.*
b) *En déduire $\varphi(t)$.*

Solution. **a)** La variable aléatoire X est intégrable et donc φ est de classe C^1 ; de plus, par la proposition 4.3.1,

$$\begin{aligned}\varphi'(t) &= \mathbb{E}((iX)e^{itX}) \\ &= \int_{\mathbb{R}} ixe^{itx}\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx \\ &= \int_{\mathbb{R}} ie^{itx}\frac{d}{dx}\left(\frac{-e^{-x^2/2}}{\sqrt{2\pi}}\right)dx \\ &= \int_{\mathbb{R}} i^2te^{itx}\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx \\ &= -t\varphi(t),\end{aligned}$$

où on a utilisé une intégration par parties.
b) On déduit de la question précédente que $\varphi(t) = \varphi(0)\exp(-\int_0^t sds) = \varphi(0)e^{-t^2/2}$, et comme $\varphi(0) = \mathbb{E}(1) = 1$, $\varphi(t) = e^{-t^2/2}$. △

Remarque. On pourrait être tenté de démontrer que $\varphi(t) = e^{-t^2/2}$ en faisant un changement de variable : en effet,

$$\begin{aligned}\varphi(t) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{itx}e^{-x^2/2}dx \\ &= e^{-t^2/2}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{t^2/2+itx-x^2/2}dx \\ &= e^{-t^2/2}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-(x-it)^2/2}dx,\end{aligned}$$

puisque $t^2/2 + itx - x^2/2 = -(x - it)^2/2$. Cependant, ce changement de variable $y = x - it$ est totalement interdit ! En effet, un changement de variable est juste une autre paramétrisation du même ensemble de points : ici, si on pose $y = x - it$, on aura forcément y qui est complexe non-réel ! On n'aurait donc plus du tout une intégrale du même type. Vous verrez en analyse complexe que l'on peut justifier ce genre de calculs (mais il y a des conditions !)

De même que pour les fonctions génératrices, la fonction caractéristique d'une somme de variables aléatoires indépendante est particulièrement facile à calculer :

Théorème 4.3.3. *Si les v.a.r. $X_1, \ldots, X_n$ forment une famille indépendante pour tout $t \in \mathbb{R}$*

$$\varphi_{X_1+\cdots+X_n}(t) = \varphi_{X_1}(t) \cdots \varphi_{X_n}(t).$$

Preuve. Il suffit de constater que, comme les variables aléatoires $e^{iX_1}, \ldots, e^{iX_n}$ sont indépendantes,

$$\varphi_{X_1+\cdots+X_n}(t) = \mathbb{E}(e^{it(X_1+\cdots+X_n)}) = \mathbb{E}(e^{itX_1} \cdots e^{itX_n}) = \mathbb{E}(e^{itX_1}) \cdots \mathbb{E}(e^{itX_n}) = \varphi_{X_1}(t) \cdots \varphi_{X_n}(t),$$

comme on l'a annoncé. □

Exemple 4.3.2. *Soit X_1 et X_2 deux variables aléatoires indépendantes telles que $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Alors $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. En effet, comme X_1 et X_2 sont indépendantes,*

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t) = e^{it\mu_1 - \sigma_1^2 t^2/2} \cdot e^{it\mu_2 - \sigma_2^2 t^2/2} = e^{it(\mu_1+\mu_2) - (\sigma_1^2+\sigma_2^2)t^2/2},$$

et nous reconnaissons la fonction caractéristique d'une variable aléatoire de loi $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Comme la fonction caractérise la loi, on a bien $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

4.3.3 Fonctions caractéristiques et convergence en loi

Les fonctions caractéristiques jouent un rôle important dans les problèmes où inteviennent des convergences en loi. Les deux théorèmes qui suivent illustrent ce fait.

Théorème 4.3.4. *La loi d'une v.a.r. (resp. d'un vecteur aléatoire) est déterminée par sa fonction caractéristique : si Y et Z sont deux v.a.r. telles que pour tout $t \in \mathbb{R}$*

$$\varphi_Y(t) = \varphi_Z(t),$$

alors la loi de Y et la loi de Z sont les mêmes : pour tout intervalle I de $\mathbb{R}$

$$\mathbb{P}(Y \in I) = \mathbb{P}(Z \in I).$$

En particulier, elles ont la même fonction de répartition.

La notion de fonction caractéristique est très utile pour donner un critère utile de convergence en loi. On a ainsi le théorème important suivant (admis) :

Théorème 4.3.5. *La suite de v.a.r. (resp. de vecteurs aléatoires) $(Y_n)_{n\in\mathbb{N}}$ converge en loi vers Y si et seulement si pour tout $t \in \mathbb{R}$*

$$\lim_{n\to\infty} \varphi_{Y_n}(t) = \varphi_Y(t).$$

Exercice 4.3.3. *Soit $\lambda > 0$. Soit $(X_n)_{n\geq 1}$ une suite de v.a. telle que X_n soit binomiale de paramètres n et λ/n. Montrer que X_n converge en loi vers une variable aléatoire dont on précisera la loi.*

Solution. Soit φ_n la fonction caractéristique de X_n. Alors

$$\varphi_n(t) = \mathbb{E}(e^{itX_n}) = \phi_{X_n}(e^{it/n}) = \left(1 + \frac{\lambda}{n}(e^{it} - 1)\right)^n.$$

Par conséquent, comme $|e^{it} - 1| \leq 2$, pour n assez grand pour que $2\lambda/n < 1/2$, on se trouve dans le domaine de convergence absolue de $\log(1 + x)$, et donc

$$\log \varphi_n(t) = n \log\left(1 + \frac{\lambda}{n}(e^{it} - 1)\right) \sim \lambda(e^{it} - 1),$$

et par continuité de exp (attention ici, l'argument est complexe, mais l'exponentielle a un rayon de convergence infini, donc pas de problème) on obtient, pour tout $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \varphi_n(t) = e^{\lambda(e^{it}-1)},$$

qui est la fonction caractéristique d'une variable aléatoire Z de loi de Poisson de paramètre λ puisque $\varphi_Z(t) = \phi_Z(e^{it}) = e^{\lambda(e^{it}-1)}$. D'après le théorème 4.3.5, X_n/n converge en loi vers une variable aléatoire de Poisson de paramètre λ. △

Exercice 4.3.4. a) *Soit Z une variable aléatoire exponentielle de paramètre λ. Calculer sa fonction caractéristique.*
b) *Soit $\lambda > 0$ et soit $(X_n)_{n\geq 1}$ une suite de v.a. telle que X_n suive une loi géométrique de paramètre λ/n, c'est-à-dire $\mathbb{P}(X_n = k) = \frac{\lambda}{n}(1-\lambda/n)^{k-1}$. Montrer que X_n/n converge en loi vers une variable aléatoire dont on précisera la loi.*

Solution. **a)** D'après la formule de transfert, on a

$$\begin{aligned}\varphi_Z(t) &= \int_0^\infty e^{itx}\lambda e^{-\lambda x}dx \\ &= \lambda \int_0^\infty e^{(it-\lambda)x}dx.\end{aligned}$$

On a ici affaire à l'intégrale d'une fonction complexe : cependant, il s'agit d'intégrer une exponentielle, que l'on peut réécrire comme une serie entière qui converge absolument sur tout $\mathbb{C}$. Par conséquent, on peut intégrer terme à terme, ce qui justifie que, même lorsque α est complexe $\int e^{\alpha x} = e^{\alpha x}/\alpha$, pourvu bien sûr que $\alpha \neq 0$. Il s'ensuit que

$$\varphi_Z(t) = \lambda \left[\frac{e^{(it-\lambda)x}}{it - \lambda}\right]_0^\infty = \frac{\lambda}{\lambda - it},$$

puisque, comme $\lambda > 0$, $\lim_{x\to\infty} |e^{(it-\lambda)x}| = \lim_{x\to\infty} e^{-\lambda x} = 0$.
b) Soit $\varphi_n(t)$ la fonction caractéristique de X_n/n. Alors

$$\begin{aligned}\varphi_n(t) = \mathbb{E}(e^{itX_n/n}) &= \phi_{X_n}(e^{it/n}) \\ &= \frac{\lambda/n}{1 - e^{it/n}(1 - \lambda/n)} \\ &= \frac{\lambda/n}{1 - (1 + it/n + o(1/n))(1 - \lambda/n)} \\ &= \frac{\lambda/n}{\lambda/n - it/n + O(1/n^2)} \\ &\sim \frac{\lambda}{\lambda - it} = \varphi_Z(t)\end{aligned}$$

Par conséquent, d'après le théorème 4.3.5, X_n/n converge en loi vers Z. On aurait pu aussi directement montrer la convergence de la fonction de répartition de X_n/n. △

Exercice 4.3.5. *Soit* $(\mu_n)_{n\geq 1}$ *et* $(\sigma_n^2)_{n\geq 1}$ *deux suites telles que* $\lim_n \mu_n = \mu \in \mathbb{R}$ *et* $\lim_n \sigma_n^2 = \sigma^2 > 0$. *Montrer que si* $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ *alors* X_n *converge en loi vers* $Z \sim \mathcal{N}(\mu, \sigma^2)$.

Solution. La fonction caractéristique de X_n est $\varphi_n(t) = \exp(i\mu_n t - \sigma_n^2 t^2/2)$. Par continuité de l'exponentielle sur tout $\mathbb{C}$, on obtient

$$\lim_{n\to\infty} \varphi_n(t) = \exp(i\mu t - \sigma^2 t^2/2),$$

qui est la fonction caractéristique d'une variable aléatoire de loi $\mathcal{N}(\mu, \sigma^2)$. D'après le théorème 4.3.5, $X_n \to Z$ où Z a une loi $\mathcal{N}(\mu, \sigma^2)$. △

4.3.4 Démonstration du théorème de la limite centrale

Passons à présent à la démonstration du théorème de la limite centrale (ou "Central Limit Theorem")

Soient donc $X_1, \ldots, X_n, \ldots$ une famille indépendante de v.a qui sont de même loi et de carré intégrable. Nous posons $\mu = \mathbb{E}(X_1)$ et $\sigma = \mathbf{Var}(X_1)$. Notons $S_n = X_1 + \cdots + X_n$, $\Sigma_n = (X_1 - \mu) + \cdots + (X_n - \mu)$ et posons

$$Z_n = \frac{S_n - n\mathbb{E}(X_1)}{\sigma\sqrt{n}} = \frac{\Sigma_n}{\sigma\sqrt{n}}.$$

La formule de l'exercice du début de la section 4.3.2 montre que

$$\varphi_{Z_n}(t) = \varphi_{\Sigma_n}\left(\frac{t}{\sigma\sqrt{n}}\right),$$

et comme Σ_n est la somme des v.a. indépendantes $X_i - \mu$, $1 \leq i \leq n$ et que ces v.a ont même loi :

$$\varphi_{\Sigma_n}(t) = \varphi_{X_1-\mu}(t)^n.$$

Si on pose $\varphi(t) = \varphi_{X_1-\mu}(t)$ on a donc

$$\varphi_{Z_n}(t) = \varphi\left(\frac{t}{\sigma\sqrt{n}}\right)^n.$$

Puisque la v.a est de carré intégrable la fonction φ est de classe C^2. Par ailleurs,

$$\varphi(0) = \mathbb{E}(1), \quad \varphi'(0) = i\mathbb{E}(X_1 - \mu), \quad \varphi''(0) = -\mathbb{E}((X_1 - \mu)^2),$$

c'est-à-dire

$$\varphi(0) = 1, \quad \varphi'(0) = 0, \quad \varphi''(0) = -\sigma^2.$$

D'après la formule de Taylor

$$\varphi(t) = 1 - \frac{\sigma^2}{2}t^2 + o(t^2),$$

et donc pour t fixé

$$\varphi_{Z_n}(t) = \left(1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o((t/(\sigma\sqrt{n}))^2)\right)^n,$$

ou encore

$$\varphi_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + o(\frac{1}{n})\right)^n.$$

Pour n suffisamment grand, $1 - (t^2)/(2n) + o(1/n)$ est dans la boule de centre 1 et de rayon 1/2 et on peut écrire

$$\varphi_{Z_n}(t) = \exp\left(n\log\left(1 - \frac{t^2}{2n} + o(\frac{1}{n})\right)\right),$$

où log est la détermination principale du logarithme dans le plan complexe (qui admet le même développement en série que le logarithme réel). On a donc

$$\varphi_{Z_n}(t) = \exp\left(-\frac{t^2}{2} + o(1)\right)\Bigg) = e^{-(t^2/2)} + o(1),$$

et partant,

$$\lim_{n\to\infty} \varphi_{Z_n}(t) = e^{-t^2/2}.$$

On reconnait dans le membre de droite la fonction caractéristique d'une v.a gaussienne normalisée et le théorème 4.3.5 montre que Z_n converge en loi vers une loi gausienne normalisée.