Notes de cours de l'ISIMA, première année
http://www.isima.fr/~leborgne

Introduction aux probabilités

Gilles Leborgne

16 mars 2016

Table des matières

Entre crochet [...] [square bracket] on donne la traduction anglaise.

1 Préambule

1.1 Les différentes probabilités

Les probabilités servent à décrire une expérience dont le résultat est impossible à prévoir avec certitude. On distingue :

1- probabilité expérimentale = connaissance a posteriori. Exemple : dans une population, on prend en compte tous les habitants, et on en déduit le pourcentage (= proportion) d'hommes.

2- Probabilité théorique = connaissance a priori. Exemple, on lance une pièce équilibrée, et la probabilité d'obtenir pile est $\frac{1}{2}$ (une chance sur deux). La probabilité théorique est en général différente de la probabilité expérimentale, mais on montre que si on recommence cette expérience du lancer un très grand nombre de fois, alors les probabilités théoriques et expérimentales s'accordent.

2'- Probabilité déductive. Exemple, on fait un sondage sur une "petite" partie de la population, d'où une vraie connaissance a posteriori sur ce sondage, et on essaie d'en déduire un résultat théorique pour toute la population. Ce cas de probabilité déductive rentre dans le cadre des probabilités théoriques.

La démarche des probabilités est : si on "reproduit" un très grand nombre de fois une "expérience aléatoire" (exemple un lancer de pile ou face), on peut en déduire des propriétés (exemple pièce équilibrée ou pas, et plus précisément donner le pourcentage moyen de pile obtenues). (Il est clair que si on ne lance la pièce que deux ou trois fois, on ne pourra pas quantifier cette information de manière "fiable".)

D'où l'importance de la "loi des grand nombres" et du "théorème central limite" qui sont démontrés à la fin du poloycopié.

1.2 Difficultés liées au vocabulaire

Les difficultés liées au vocabulaire sont souvent liées à des motivations historiques, et sont souvent sources de confusion, d'incompréhension, de découragement...

Ainsi certaines expressions qui se veulent "illustratives" semblent être employées "à contresens".

Un exemple flagrant est l'expression : "une variable aléatoire" (v.a.r.) : qui n'est ni variable, ni aléatoire... : une variable aléatoire réelle X est une fonction à valeurs réelles $X : \Omega \to \mathbb{R}$ bien déterminée. (L'ensemble de définition Ω est l'ensemble qui contient tous les résultats possibles d'une expérience aléatoire.)

Exemple : on prend deux dés (à 6 faces numérotées de 1 à 6). On les lance. L'ensemble des résultats possibles est l'ensemble $\Omega = \{(i,j) \in [1,6]_{\mathbb{N}} \times [1,6]_{\mathbb{N}}\}$ des couples résultats des lancers. (Exemple : le premier donne 2 et le second dé donne 5, donc le résultat est $(2,5) \in \Omega$.) Et on s'intéresse à la somme des résultats. On note $X : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ la fonction somme ; $X(a,b) = a + b$. Cette fonction X est parfaitement définie : ce n'est pas une variable (c'est une fonction), et elle n'a rien d'aléatoire (elle est parfaitement définie). Mais X sera appelée "variable aléatoire" parce que le couple (i,j) auquel on s'intéresse lorsqu'on calcule $X(i,j)$, est issue d'une expérience aléatoire (le lancer de deux dés).

Ainsi si on obtient $(1,1)$ (le double 1) alors $X(1,1) = 2$ (parfaitement déterminé), et si un second lancer donne le résultat $(4,5)$ alors $X(4,5) = 9$ (parfaitement déterminé). Comme les éléments $(1,1)$ et $(4,5)$ de Ω sont dits "aléatoires" (i.e. résultats issus d'une expérience aléatoire), la fonction X est appelée variable aléatoire... De fait, ce n'est pas cette fonction X qui nous intéressera directement, mais sa "loi" de probabilité P_X, i.e. la probabilité que, pour cette fonction, on obtienne un résultat donné après un lancer de dés (exemple $P_X(2) = \frac{1}{36}$ car il n'y a qu'un couple sur 36 qui donne somme = 2).

De plus la valeur d'une v.a.r X est prise en ω et est notée $x = X(\omega)$... (usuellement la valeur d'une fonction f est prise en x et est notée $y = f(x)$...)

Autres exemples de vocabulaire qui peut s'avérer déroutant : loi marginale (paragraphe 2.10), "probabilité conditionnelle" (voir remarque 3.3 page 34)...

Tout ceci peut décourager... jusqu'à ce qu'on "s'imprègne" des définitions.

N.B. : la théorie des probabilités est basée sur la théorie des ensembles pour laquelle on emploie des synonymes relatifs aux noms déjà connu. Exemples :
- un sous-ensemble = un événement ;
- l'ensemble vide = l'événement impossible ;
- l'ensemble tout entier = l'événement certain ;
- le complémentaire d'un sous-ensemble = l'événement contraire ;
- un élément d'un ensemble = un événement élémentaire = une éventualité...

En Français, un événement s'écrit aussi évènement (depuis 1990), et se prononce dans les deux cas évènement.
Dans le jeu de "pile ou face", les mots pile et face sont tous deux féminins (une pile et une face).

1.3 Quelques rappels fonctionnels

1.3.1 Complémentaire

Si E est un ensemble, on note $\mathcal{P}(E)$ l'ensemble des sous-ensembles de E. Donc :

$$A \in \mathcal{P}(E) \quad \Longleftrightarrow \quad A \subset E.$$

En particulier l'ensemble vide, noté $\emptyset$, est dans $\mathcal{P}(E)$.
Si $A \subset E$ on note $A^C = E - A$ son complémentaire :

$$A^C = E - A = \{x \in E : x \notin A\}.$$

Proposition 1.1 *Si $(A_i)_{i\in I}$ est une famille de sous-ensembles de E (avec I ensemble quelconque), on a :*

$$(\bigcup_{i\in I} A_i)^C = \bigcap_{i\in I} A_i^C, \qquad (\bigcap_{i\in I} A_i)^C = \bigcup_{i\in I} A_i^C. \tag{1.1}$$

(Faire un dessin avec $I = \{1,2\}$.)

Preuve. $x \in (\bigcup_{i\in I} A_i)^C \Leftrightarrow x \notin \bigcup_{i\in I} A_i \Leftrightarrow \forall i \in I,\, x \notin A_i \Leftrightarrow \forall i \in I,\, x \in A_i^C \Leftrightarrow x \in \bigcap_{i\in I} A_i^C$.
$x \in (\bigcap_{i\in I} A_i)^C \Leftrightarrow x \notin \bigcap_{i\in I} A_i \Leftrightarrow \exists i \in I,\, x \notin A_i \Leftrightarrow \exists i \in I,\, x \in A_i^C \Leftrightarrow x \in \bigcup_{i\in I} A_i^C$. ▪

1.3.2 Fonction réciproque et fonction inverse

On se donne deux ensembles E et F. Soit $f : E \to F$ une fonction. La fonction réciproque de f est la fonction f^{-1} définie par :

$$f^{-1} : \begin{cases} \mathcal{P}(F) & \to \mathcal{P}(E) \\ \quad B & \mapsto f^{-1}(B) \overset{\text{déf}}{=} \{x \in E : f(x) \in B\} \qquad (\subset E), \end{cases} \tag{1.2}$$

et pour $B \subset E$, le sous-ensemble $f^{-1}(B)$ de E est appelé l'image réciproque de B par f.

Exemple 1.2 Exemple fondamental : voir § 1.3.3 et remarque 1.6. ▪

Abus de notation. Quand $B = \{y\}$ (cas B est un singleton), on note :

$$f^{-1}(\{y\}) \overset{\text{noté}}{=} f^{-1}(y). \tag{1.3}$$

Et si f est bijective, alors $f^{-1}(y)$ est un singleton $\{x\}$, et on note :

$$f^{-1} : \begin{cases} F & \to E \\ y & \mapsto x = f^{-1}(y) \quad \text{quand} \quad y = f(x), \end{cases} \tag{1.4}$$

et f^{-1} est appelée fonction inverse de f (quand f est bijective).

Remarque 1.3 L'utilisation de la même notation f^{-1} pour (1.2) (définie sur les ensembles) et pour (1.4) (définies sur les points) ne doit pas prêter à confusion : le contexte lève les ambiguïtés.

En probabilité, la notation f^{-1} est surtout utilisée sous la forme ensembliste (1.2) : une probabilité est une mesure d'ensembles.

Donc, si on ne fait pas allusion à la bijectivité (comme dans ce polycopié en général), quand on utilise f^{-1}, c'est de la fonction réciproque (1.2) dont il s'agit (pas de la fonction inverse). ▪

Proposition 1.4 *La fonction réciproque $f^{-1} : \mathcal{P}(F) \to \mathcal{P}(E)$ commute avec les opérations de complémentation, d'union et d'intersection : si $B \subset F$, si I est un ensemble fini non vide et $B_i \subset F$ pour tout $i \in I$, alors :*

$$\begin{aligned} &f^{-1}(B^C) = (f^{-1}(B))^C, \\ &f^{-1}(\bigcap_{i\in I} B_i) = \bigcap_{i\in I}(f^{-1}(B_i)), \\ &f^{-1}(\bigcup_{i\in I} B_i) = \bigcup_{i\in I}(f^{-1}(B_i)). \end{aligned} \tag{1.5}$$

Preuve. Montrons $(1.5)_1$. Il faut montrer $f^{-1}(B^C)\bigcap(f^{-1}(B)) = \emptyset$ et $f^{-1}(B^C)\bigcup(f^{-1}(B)) = E$. On a $F = B \cap B^C$ est partition de F, donc $\left\{ \begin{array}{l} \{x \in E : f(x) \in B^C\}\bigcap\{x \in E : f(x) \in B\} = \emptyset, \\ \{x \in E : f(x) \in B^C\}\bigcup\{x \in E : f(x) \in B\} = E. \end{array} \right\}$

Montrons $(1.5)_2$.
$\subset$: si $f^{-1}(\bigcap_i B_i) = \emptyset$ c'est trivial. Sinon, soit $x \in f^{-1}(\bigcap_i B_i)$: donc $f(x) \in \bigcap_i B_i$, donc $f(x) \in B_i$ pour tout i, donc $x \in f^{-1}(B_i)$ pour tout i, donc $x \in \bigcap_i f^{-1}(B_i)$.
$\supset$: si $\bigcap_i(f^{-1}(B_i)) = \emptyset$ c'est trivial. Sinon, soit $x \in \bigcap_i(f^{-1}(B_i))$: donc $x \in f^{-1}(B_i)$ pour tout i, donc $f(x) \in B_i$ pour tout i, donc $f(x) \in \bigcap_i B_i$, donc $x \in f^{-1}(\bigcap_i B_i)$.

Montrons $(1.5)_3$.
$\subset$: si $f^{-1}(\bigcup_i B_i) = \emptyset$ c'est trivial. Sinon, soit $x \in f^{-1}(\bigcup_i B_i)$: donc $f(x) \in \bigcup_i B_i$, donc il existe i tel que $f(x) \in B_i$, donc il existe i tel que $x \in f^{-1}(B_i)$, donc $x \in \bigcup_i(f^{-1}(B_i))$.
$\supset$: si $\bigcup_i(f^{-1}(B_i)) = \emptyset$ c'est trivial. Sinon, soit $x \in \bigcup_i(f^{-1}(B_i))$: donc il existe i t.q. $x \in f^{-1}(B_i)$, donc il existe i t.q. $f(x) \in B_i$, donc $f(x) \in \bigcup_i B_i$, donc $x \in f^{-1}(\bigcup_i B_i)$. ▪

On rappelle que, pour $A \subset E$:

$$f(A) \stackrel{\text{déf}}{=} \{y \in F \text{ t.q. } \exists x \in A,\ y = f(x)\}. \tag{1.6}$$

En particulier $f(E) =^{\text{noté}} \text{Im}(f)$.

Proposition 1.5 *Si $f : E \to F$ et $g : F \to G$, alors :*

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}. \tag{1.7}$$

Preuve. $g \circ f : E \to G$. Donc $(g \circ f)^{-1} : G \to E$. Pour $C \subset G$ on a $(g \circ f)^{-1}(C) = \{e \in E : (g \circ f)(e) \in C\} = \{e \in E : g(f(e)) \in C\} = \{e \in E : f(e) \in g^{-1}(C)\} = \{e \in E : e \in f^{-1}(g^{-1}(C))\}$ ▪

1.3.3 Fonction indicatrice

Pour A un sous-ensemble d'un ensemble E, on note 1_A la fonction indicatrice de A (ou fonction caractéristique, suivant les auteurs), i.e. la fonction $1_A : E \to \mathbb{R}$ définie par :

$$1_A(x) = \left\{ \begin{array}{ll} 1 & \text{si } x \in A, \\ 0 & \text{si } x \notin A. \end{array} \right. \tag{1.8}$$

(Faire un dessin.)

N.B. : la fonction indicatrice de A est aussi appelée caractéristique en analyse, suivant les auteurs. En probabilité la fonction caractéristique sera la transformée de Fourier de la mesure.

Remarque 1.6 Ainsi $1_A^{-1}(\{x\}) = \emptyset$ pour tout $x \neq 0, 1$, et $1_A^{-1}(\{1\}) = A = 1_A^{-1}(\mathbb{R}^*)$, et $1_A^{-1}(\{0\}) = A^C = 1_A^{-1}(\mathbb{R}-\{1\})$, et $1_A^{-1}(\{0,1\}) = E = 1_A^{-1}(\mathbb{R})$... ▪

Proposition 1.7 *Si $A, A_1, A_2 \subset E$ alors :*

$$1_{A^C} = 1 - 1_A, \qquad 1_{A_1 \cap A_2} = 1_{A_1} 1_{A_2}, \qquad 1_{A_1 \cup A_2} = 1_{A_1} + 1_{A_2} - 1_{A_1 \cap A_2}. \tag{1.9}$$

Faire un dessin. Si $f : E \to F$ est une fonction, si $B \subset F$, alors $f^{-1}(B) \subset E$ et :

$$1_{f^{-1}(B)} = 1_B \circ f. \tag{1.10}$$

(La fonction $1_{f^{-1}(B)}$ est définie sur E et la fonction 1_B est définie sur F.)

Preuve. (1.9) : immédiat. Pour (1.10) : on a $\{x \in E : f(x) \in B\} = \{x \in E : x \in f^{-1}(B)\}$, d'où pour $x \in E$ on a :

$$(1_B \circ f)(x) = 1_B(f(x)) = \left\{ \begin{array}{l} 1 \ \text{ si } f(x) \in B \\ 0 \ \text{ sinon} \end{array} \right\} = \left\{ \begin{array}{l} 1 \ \text{ si } x \in f^{-1}(B) \\ 0 \ \text{ sinon} \end{array} \right\} = 1_{f^{-1}(B)}(x). \tag{1.11}$$

■

Proposition 1.8 *Pour $A \subset E$ on a :*

$$A \subset f^{-1}(f(A)), \tag{1.12}$$

l'inclusion inverse étant fausse en général.

Preuve. Soit $x \in A$, et soit $y = f(x)$. Dnnc $x \in f^{-1}(\{y\})$, avec $f^{-1}(\{y\}) \subset f^{-1}(f(A))$, donc $x \in f^{-1}(f(A))$.

La réciproque est fausse : soit $f : \mathbb{R} \to \mathbb{R}$ donnée par $f = 1_{[0,1]}$. En particulier $f(]-1,2[) = \{0,1\}$, et $f^{-1}(f(]-1,2[)) = f^{-1}(\{0,1\}) = \mathbb{R} \supsetneq]-1,2[$. ■

1.3.4 Mesure (ou masse) de Dirac sur les ensembles

Pour $a \in E$, la mesure de Dirac (ou masse de Dirac) δ_a est la fonction $\delta_a : \left\{ \begin{array}{rl} \mathcal{P}(E) & \to \mathbb{R} \\ A & \mapsto \delta_a(A) \end{array} \right.$ définie par :

$$\delta_a(A) \stackrel{\text{déf}}{=} \left\{ \begin{array}{l} 1 \text{ si } a \in A, \\ 0 \text{ si } a \notin A, \end{array} \right. \qquad \text{donc} \quad = 1_A(\{a\}). \tag{1.13}$$

Donc δ_a est définie sur les ensembles. En particulier, $\delta_a(\{x\}) = \left\{ \begin{array}{l} 1 \text{ si } x{=}a \\ 0 \text{ si } x{\neq}a \end{array} \right.$.

1.3.5 Mesure (ou masse) de Dirac sur les fonctions

Puis on définit la mesure de Dirac $\widetilde{\delta}_a$ sur les fonctions en commençant par la définition sur les fonctions caractéristiques : pour $A \subset E$ on définit :

$$\widetilde{\delta}_a(1_A) \stackrel{\text{déf}}{=} \delta_a(A), \qquad \text{puis} \qquad \widetilde{\delta}_a(1_A) \stackrel{\text{noté}}{=} \delta_a(1_A). \tag{1.14}$$

Attention : on abuse des notations : on utilisera le même symbole δ_a pour désigner la mesure d'un ensemble et la mesure d'une fonction. Le contexte lève les ambiguïtés de notation.

Puis, par linéarité, on définit la mesure de Dirac d'une fonction en escalier [simple function] : si $f = \sum_{i=1}^n c_i 1_{A_i}$ alors :

$$\delta_a(f) = \delta_a(\sum_{i=1}^n c_i 1_{A_i}) \stackrel{\text{déf}}{=} \sum_{i=1}^n c_i \delta_a(A_i) = \sum_{i=1}^n c_i 1_{A_i}(a), \qquad \text{donc} \qquad \delta_a(f) = f(a). \tag{1.15}$$

D'où, pour toute fonction $f : E \to \mathbb{R}$ et tout $a \in E$, voir cours "intégrale de Lebesgue" :

$$\delta_a(f) = f(a). \tag{1.16}$$

1.3.6 Abus de notation fonctionnelle

Soit $id : \mathbb{R} \to \mathbb{R}$ l'application identité :

$$id : x \to id(x) = x, \tag{1.17}$$

et soit $f : x \in \mathbb{R} \to f(x) \in \mathbb{R}$ une fonction donnée. La fonction produit :

$$id\, f : x \to (id\, f)(x) = id(x) f(x) = x f(x), \tag{1.18}$$

est notée :

$$id\, f \overset{\text{noté}}{=} xf. \tag{1.19}$$

Ainsi $xf : \mathbb{R} \to \mathbb{R}$ est définie par :

$$(xf)(x) \overset{\text{déf}}{=} x f(x). \tag{1.20}$$

Sous-entendu, quand on utilise cette notation, le nom de la variable est x (attention aux notations !). Sans abus de notation, xf est la fonction produit $id f$.

Remarque 1.9 Attention : x joue ici deux rôles, celui usuel de variable et celui abusif de fonction (abusif mais pratique). Le contexte doit lever toute ambiguïté. ▪

Exemple 1.10 ωf est la fonction $\omega \to \omega f(\omega)$.
Sans abus de notation, c'est la fonction produit $id f : \omega \to id(\omega) f(\omega) = \omega f(\omega)$. ▪

Et on continue à abuser des notations en notant :

$$(id - x_0 1_{\mathbb{R}}) f \overset{\text{noté}}{=} (x - x_0) f, \tag{1.21}$$

i.e. $(x-x_0)f$ est la fonction de $x \to ((x-x_0)f)(x)$ donnée par :

$$((x-x_0)f)(x) \overset{\text{déf}}{=} (x-x_0) f(x).$$

Exemple : la fonction $X - \overline{X}$, qui à la fonction X retranche sa valeur moyenne, est un abus de notation usuel : on aurait dû noter $X - \overline{X} 1_{\mathbb{R}}$ cette fonction, différence des fonctions X et $\overline{X} 1_{\mathbb{R}}$ (fonction constante). On a donc $(X - \overline{X})(x) = X(x) - \overline{X}$.

Cela paraît trivial, mais la confusion entre un réel et une fonction conduit à certaines incompréhensions comme la confusion entre une fonction et ses valeurs. Dans le doute (et dans les démonstrations), il faut revenir aux notations non abusives.

1.3.7 Dénombrement de fonctions

Soit $\mathcal{F}(E; F)$ l'ensemble des fonctions $f : E \to F$.

Proposition 1.11 *Si E et F sont des ensembles finis non vides, de cardinal $|E| = m$ et $|F| = n$, alors $\mathcal{F}(E; F)$ est un ensemble fini de cardinal $|\mathcal{F}(E; F)| = n^m$.*
En particulier $|\mathcal{F}(\{1, 2, ..., m\}; \{a, b\})| = 2^m$ quand $a \neq b$.

Preuve. Notons $E = \{a_1, ..., a_m\}$, $F = \{b_1, ..., b_n\}$, et $f : E \to F$. Pour $f(a_1)$ on a n choix possibles. Pour $f(a_2)$ on a n choix possibles... Pour $f(a_m)$ on a n choix possibles. Au total $n * n * ... * n = n^m$ comme annoncé. ▪

Exemple 1.12 Avec un alphabet de n lettres (de cardinal 26 en français), combien de "mots" (ayant un sens ou pas) de k lettres peut-on former ?

Réponse. Si F est l'ensemble des n lettres de l'alphabet (de cardinal 26 en français), on a n possibilités pour la 1ère lettre, n possibilités pour la 2ème lettre,..., donc au total n^k possibilités (soit 26^k en français). C'est le cardinal de $\mathcal{F}(\{1, 2, ..., k\}; F)$ (le cardinal de $\mathcal{F}(\{1, 2, ..., k\}; \{a, b, ..., z\})$ en français). ▪

1.3.8 Cardinal de $\mathcal{P}(\Omega)$

Proposition 1.13 *Soit Ω est un ensemble fini de cardinal $|\Omega| = n \geq 0$. Le cardinal de $\mathcal{P}(\Omega) = 2^n$.*

Preuve. On donne 2 démonstrations usuelles de $\mathcal{P}(\Omega) = 2^n$.

1- Par récurrence. Vrai pour $\Omega = \emptyset$ (de cardinal n=0) puisque $\mathcal{P}(\Omega) = \{\emptyset\}$. Vrai pour $\Omega = \{a\}$ (de cardinal n=1) puisque $\mathcal{P}(\Omega) = \{\emptyset, \{a\}\}$. Supposons que ce soit vrai pour $\Omega = \{a_1, ..., a_n\}$ de cardinal n. Alors pour $\Omega = \{a_1, ..., a_{n+1}\}$ de cardinal n+1, l'ensemble $\mathcal{P}(\Omega)$ contient tous les sous-ensembles ne contenant pas a_{n+1}, au nombre de 2^n, et tous ceux contenant a_{n+1}, qui sont les précédents auquel on a ajouté a_{n+1}, donc au nombre de 2^n. Au total $2^n + 2^n = 2^{n+1}$.

2- Calcul direct, à l'aide des combinaison $C_n^k = \binom{n}{k}$, cf. § suivant. $\mathcal{P}(\Omega)$ contient $\emptyset$ (donc $1 = C_n^0$ élément), contient les singletons (donc $n = C_n^1$ éléments), les sous-ensembles de 2 éléments (donc C_n^2 éléments),.... , donc au total $\sum_{k=0}^n C_n^k = (1+1)^n = 2^n$. ▪

1.4 Quelques rappels combinatoires

1.4.1 Arrangements $A_n^k = (n)_k$

On rappelle que "n factoriel" [n factorial] est l'entier $n! = n \times (n-1) \times ... \times 2 \times 1$ pour tout $n \in \mathbb{N}^*$, et qu'on pose $0! = 1$. (Motivation pour définir $0! = 1$: voir la formule du binôme de Newton ci-dessous équation (1.25).)

Question 1- : soit n objets $x_1, ..., x_n$ deux à deux distincts (avec $n \geq 1$). Combien y-a-t-il de manières de les arranger entre eux, i.e. combien de n-uplets ordonnés [permutations] $(x_{\varphi(1)}, ..., x_{\varphi(n)})$ peut-on former ? Autrement dit, combien y-a-t-il de bijections φ entre $[1, n]_{\mathbb{N}}$ et lui-même ?

Réponse : $n!$.

Démonstration par récurrence : c'est vrai pour $n = 1$. Puis supposons que ce soit vrai pour n objets. Prenons un ensemble de n+1 objets. Il y a n+1 possibilités pour le choix du premier objet, et il reste alors n objets, avec $n!$ possibilités de les arranger par hypothèse de récurrence. Donc au total $(n+1)n! = (n+1)!$ possibilités.

Remarque 1.14 $A_n^n = n!$ est le nombre de bijections entre un ensemble de n éléments et un autre ensemble de n éléments. Ou encore le nombre de bijections entre $\{1, ..., n\}$ et $\{x_1, ..., x_n\}$. Ou encore le nombre de bijections entre $\{x_1, ..., x_n\}$ et lui même, une telle bijection étant de la forme $(x_1, ...x_n) \rightarrow (x_{\varphi(1)}, ..., x_{\varphi(n)})$ avec φ bijection entre $\{1, ..., n\}$ et $\{1, ..., n\}$. ▪

Question 2- : soit n objets $x_1, ..., x_n$ deux à deux distincts (avec $n \geq 1$). Soit $k \in [1, n]_{\mathbb{N}}$. Combien de k-uplets ordonnés [permutations] $(x_{\varphi(1)}, ..., x_{\varphi(k)})$ peut-on former ? Autrement dit, combien y-a-t-il d'injections φ [permutations] de $[1, k]_{\mathbb{N}}$ dans $[1, n]_{\mathbb{N}}$?

Réponse : c'est le nombre d'arrangements de "k objets parmi n" [number of ordered arrangements of k objects taken from n] :

$$A_n^k \stackrel{\text{déf}}{=} \frac{n!}{(n-k)!} = n \times (n-1) \times ... \times (n-k+1). \tag{1.22}$$

En particulier $A_n^n = n!$.

English $A_n^k = (n)_k =$ "n down k" = "n lower k" = $P(n, k)$ (permutations).

Démonstration par récurrence. C'est vrai pour $n = 1$ et $k = 1$ puisque $A_1^1 = \frac{1!}{(1-1)!} = 1$. Si $n = 2$ et $k = 1$, alors $2 = A_2^1$ est le nombre de manières de choisir 1 objet parmi 2, et pour $k = 2$, $1 = A_2^2$ est le nombre de manières de choisir 2 objets parmi 2.

Supposons que la formule est établie pour $n-1 \geq 1$ et tout $k \in [1, n-1]_{\mathbb{N}}$. Passons à n objets. Si $k = 1$ alors $n = A_n^1$ est le nombre de manières de choisir 1 objet parmi n. Supposons que la formule soit vrai pour k objets avec $1 \leq k \leq n-1$. Passons à k+1 objets. Il y a n manières de choisir le premier, et il reste $n-1$ objets parmi lesquels on doit en prendre k (en tenant compte de l'ordre), donc A_{n-1}^k manières de les arranger. Donc au total $nA_{n-1}^k = n\frac{(n-1)!}{(n-1-k)!} = \frac{n!}{(n-(k+1))!} = A_n^{k+1}$.

En particulier pour $k = n$ on retrouve le cas de la question 1- : $A_n^n = n!$. (Pour $k = 0$ la question 2- n'a pas de sens.)

Exemple 1.15 (Tiercé dans l'ordre.) Parmi 20 chevaux, combien de triplets (ordonnés) peut-on former ? Réponse : $A_{20}^3 = \frac{20!}{17!} = 18 * 19 * 20 = 6840$.

Exercice 1.16 On dispose des lettres B E E I L R T. En les tirant succesivement au hasard, avec quelle probabilité forme-t-on LIBERTE ?

Réponse. L : 1/7, I : 1/6, B : 1/5, E : 2/4, R :1/3, T : 1/2, E : 1/1. Donc probabilité $P = \frac{2}{7!}$. En termes d'arrangements : il y a 7! arrangements possibles (nombre de mots possibles). Et il 2 possibilités pour tirer E (deux mots possibles indifférenciables).

Remarque 1.17 An arrangement where order is important is called a permutation. An arrangement where order is not important is called combination.

1.4.2 Combinaisons $C_n^k = \binom{n}{k}$ et coefficients binomiaux

Question 3- : soit $n \in \mathbb{N}^*$ et soit $k \in [1, n]_{\mathbb{N}}$. Combien y-a-t-il de manières de choisir k objets parmi n, sans tenir compte de l'ordre ?

Autrement dit, si $E = \{x_1, ..., x_n\}$ est l'ensemble des n objets, combien y-a-t-il de sous-ensembles de E contenant k objets ?

Autrement dit, combien y-a-t-il de partitions $U_1 \bigcup U_2$ de E tel que $|U_1| = k$?

Réponse : c'est le nombre de combinaisons [combinations] de k objets parmi n [number of ways of selecting k objects from n, or number of subpopulations of size k in a population of size n] :

$$C_n^k \overset{\text{noté}}{=} \binom{n}{k} \overset{\text{déf}}{=} \frac{n!}{k!(n-k)!} = \frac{n \times (n-1) \times ... \times (n-k+1)}{k!} \qquad (= \frac{A_n^k}{k!}), \tag{1.23}$$

et C_n^k est un coefficient binomial (voir la formule du binôme de Newton ci-dessous équation (1.25)).
($C_n^k = \binom{n}{k}$ = "n choose k" = $C(n, k)$.)

Démonstration avec les arrangements. A_n^k est le nombre de k-uplets ordonnés, et pour un choix de k éléments, il y a $k!$ manières de les arranger. Donc $k!\, C_n^k = A_n^k$.

Démonstration par récurrence. C'est vrai pour $n = 1$ et $k = 1$ puisque $C_1^1 = \frac{1!}{1!0!} = 1$. Pour $n = 2$ et $k = 1$, on a $2 = C_2^1$ est le nombre de manières de choisir 1 objet parmi 2, et pour $k = 2$, $1 = C_2^2$ est le nombre de manières de choisir 2 objets parmi 2.

Supposons que la formule est établie pour $n \geq 1$ et tout $k \in [1, n]_{\mathbb{N}}$. Passons à $n+1$ objets. C'est vrai pour $k = 1$ puisqu'il y a $n+1$ singletons et $C_{n+1}^1 = \frac{(n+1)!}{1!n!} = n+1$. Supposons que ce soit vrai pour k avec $1 \leq k \leq n$. Passons à $k+1$ objets. Regardons les sous-ensembles ne contenant pas x_1 : on doit choisir $k+1$ objets parmi n, soit C_n^{k+1} possibilités. Regardons les sous-ensembles contenant x_1 : il reste k objets à prendre parmi n, soit C_n^k possibilités. Au total $C_n^k + C_n^{k+1} = C_{n+1}^{k+1}$, voir lemme suivant.

Remarque 1.18 Comme $C_n^k =^{\text{noté}} \binom{n}{k} = \frac{n!}{k!(n-k)!}$, il est immédiat que :

$$C_n^k = C_n^{n-k} = \binom{n}{k} = \binom{n}{n-k}, \qquad \text{et} \qquad C_{n+k}^n = C_{n+k}^k = \binom{n+k}{k} = \binom{n+k}{n}.$$

Lemme 1.19 *(Formule du triangle de Pascal.) Pour $n \in \mathbb{N}^*$ et $k \in [1, n]_{\mathbb{N}}$ on a :*

$$C_n^k + C_n^{k+1} = C_{n+1}^{k+1}, \tag{1.24}$$

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

Preuve. $\dfrac{n!}{k!(n-k)!} + \dfrac{n!}{(k+1)!(n-k-1)!} = \dfrac{n!(k+1) + n!(n-k)}{(k+1)!(n-k)!} = \dfrac{n!(n+1+k-k)}{(k+1)!(n+1-(k+1))!}$.

Exemple 1.20 (Tiercé dans le désordre.) Parmi 20 chevaux, combien d'ensembles de 3 chevaux peut-on formés ? Réponse : $C_{20}^3 = \frac{20!}{3!17!} = \frac{18.19.20}{2.3} = 1140$.

Question 4- : n lancers d'une pièce avec résultat p=pile ou f=face. Combien de résultats contiennent exactement k fois p, pour $k \in [1,n]_{\mathbb{N}}$.

Réponse : C_n^k.

Par récurrence. Pour $n=1$, l'ensemble des résultats possibles est $\{p,f\}$. Et pour $k=0$ et $k=1$ on a $C_1^0 = 1 = C_1^1$: la formule est vraie.

Supposons que la formule est établie pour $n \geq 1$ et tout $k \in [0,n]_{\mathbb{N}}$. Passons à $n+1$ objets. C'est vrai pour $k=0$ car $C_{n+1}^0 = 1$ (le seul résultat est $(f,f,,...,f)$). C'est vrai pour $k=1$ car $C_{n+1}^1 = n+1$ (les tirages ou p ne sort qu'une fois, soit $(p,f,f,...)$, $(f,p,f,f,...)$,... , $(f,...,f,p)$).

Supposons que ce soit vrai pour k avec $k<n$. Passons à $k+1$. Si le premier lancer donne p alors il reste k p à tirer sur les $n-1$ lancers suivant, soit C_{n-1}^k possibilités. Si le premier lancer donne f alors il reste $k+1$ p à tirer sur les $n-1$ lancers suivant, soit C_{n-1}^{k+1} possibilités. Donc au total $C_{n-1}^k + C_{n-1}^{k+1} = C_n^{k+1}$, cf (1.24).

Exemple 1.21 Dans un alphabet à 2 lettres a et b, C_n^k est le nombre de mots de longueur n qui contiennent exactement k fois la lettre b (exemple $n=8$ et $k=3$: mot comme *aababbaa*). C'est le cas de la question 4 où pile et face ont été remplacés par a et b. ▪

Exercice 1.22 Soit un groupe de n personnes. Combien de paires mathématiques (non ordonnées appelées couples dans la vie courante) peut-on former ? Combien de couples mathématiques (paires ordonnées) peut-on former ?

Réponse. Paires : C_n^2. Ou directement : avec la "1ère" personne on peut en faire $(n-1)$, puis avec la "2ième" $(n-2)$ différentes, ..., au total $1+2+...+(n-2)+(n-1) = \frac{n(n-1)}{2}$ (cf. matrice $n*n$ symétrique, on compte les éléments "supérieurs stricts").

Couples : la 1ère personne peut en faire $n-1$, la seconde $(n-1)$, ..., la n-ième $(n-1)$: au total $n(n-1)$ (cf. matrice $n*n$, on compte les éléments "hors diagonale"). ▪

Exercice 1.23 Jeu de 52 cartes. On tire au hasard 4 cartes. Quelle chance a-t-on d'avoir exactement 2 rois ?

Réponse. Une main = 4 cartes. Nombres de mains = C_{52}^4. Nombre de main avec 2 rois = $C_4^2 \times C_{48}^2$ (2 rois parmi 4, et 2 autres parmi 48). Donc probabilité = $\frac{C_4^2 C_{48}^2}{C_{52}^4} = \frac{4!}{2!2!}\frac{48!}{2!46!}\frac{4!48!}{52!} = 6 \times 47 \times 24 \times \frac{6}{49\times50\times51\times13} \simeq 0.025 \simeq 2.5\%$. ▪

Exemple 1.24 Voir plus loin (2.30) page 30. ▪

1.4.3 Formule du binôme (Newton)

Proposition 1.25 *(Formule du binôme.) Pour $x,y \in \mathbb{R}$ (ou $\in \mathbb{C}$) et $n \in \mathbb{N}^*$, on a :*

$$(x+y)^n = \sum_{k=0}^{n} C_n^k x^k y^{n-k}, \qquad \textit{en particulier} \qquad 2^n = \sum_{k=0}^{n} C_n^k. \tag{1.25}$$

Preuve. Pour $(1.25)_1$: par récurrence sur n. Vrai pour $n=1$ car $C_1^0 = 1 = C_1^1$. Puis

$$\begin{aligned}
(x+y)^{n+1} = x(x+y)^n + y(x+y)^n &= \sum_{k=0}^{n} C_n^k x^{k+1} y^{n-k} + \sum_{k=0}^{n} C_n^k x^k y^{n+1-k} \\
&= x^{n+1} + \sum_{k=0}^{n-1} C_n^k x^{k+1} y^{n-k} + y^{n+1} + \sum_{k=1}^{n} C_n^k x^k y^{n+1-k} \\
&= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} (C_n^{k-1} + C_n^k) x^k y^{n+1-k} \\
&= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} C_{n+1}^k x^k y^{n+1-k} = \sum_{k=0}^{n+1} C_{n+1}^k x^k y^{n+1-k}.
\end{aligned}$$

Puis $x=y=1$ donne $(1.25)_2$. ▪

Exemple 1.26 Vers la loi binomiale.

$(x+y)^2 = (x+y)(x+y) = x^2+xy+yx+y^2$: pour tirer x et y, on peut tirer d'abord x puis y (résultat xy dans la formule), ou tirer d'abord y puis x (résultat yx dans la formule), donc au total deux possibilités (ici $2xy = C_2^1 xy$).

$(x+y)^3 = (x^2+xy+yx+y^2)(x+y) = x^3+x^2y+xyx+xy^2+yx^2+yxy+y^2x+y^3$: pour tirer 2 fois x et 1 fois y, on peut tirer d'abord deux fois x et une fois y (résultat x^2y dans la formule), ou tirer x puis y puis x (résultat xyx dans la formule), ou tirer y puis deux fois x (résultat yx^2 dans la formule), donc au total trois possibilités ($3x^2y = C_3^1x^2y = C_3^2x^2y$).

$(x+y)^4 = (x^3+x^2y+xyx+xy^2+yx^2+yxy+y^2x+y^3)(x+y) = ... + x^2y^2 + ... + xyxy + ... + xy^2x + ... + yx^2y + ... + yxyx + ... + y^2x^2 + ...$: pour tirer 2 fois x et 2 fois y, on a $6 = C_4^2$ possibilités.

1.4.4 Triangle de Pascal et formules pour les coefficients binomiaux

Soit $n \geq 1$ et $k \in [0,n]_{\mathbb{N}}$. On a :

$$C_{n+1}^{k+1} = \frac{n+1}{k+1} C_n^k. \tag{1.26}$$

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

Pour trouver les coefficients de $(x+y)^{n+1}$ à partir des coefficients de $(x+y)^n$, on utilise le triangle de Pascal (faire dessin) grâce à la formule (1.24).

On déduit de (1.24), pour $k \leq n$: $C_k^k + C_{k+1}^k + ... + C_n^k = C_{n+1}^{k+1}$, soit, pour $k \leq n$:

$$\sum_{j=k}^{n} C_j^k = C_{n+1}^{k+1}, \qquad \text{soit} \qquad \sum_{j=k}^{n} \binom{j}{k} = \binom{n+1}{k+1}. \tag{1.27}$$

Démonstration par récurrence : Pour $n=1$ et $k=1$ on a $C_1^1 = C_2^2 = 1$, et pour $n=1$ et $k=0$ on a $C_0^0 + C_1^0 = 1+1 = C_2^1$.

Supposons que ce soit vrai pour $n \geq 1$, et tous les $k \in [0,n]_{\mathbb{N}}$. Passons à $n+1$.

Pour $k = n+1$ on a $\sum_{j=n+1}^{n+1} C_j^{n+1} = C_{n+1}^{n+1}$.

Pour $k \leq n$ on a $\sum_{j=k}^{n+1} C_j^k = \sum_{j=k}^{n} C_j^k + C_{n+1}^k = C_{n+1}^{k+1} + C_{n+1}^k = C_{n+2}^{k+1}$, cf. (1.24).

Exercice 1.27 Notons $(1+x+...+x^r)^n = \alpha_{n,0} + \alpha_{n,1} + ... + \alpha_{n,r-1}x^{r-1} + \alpha_{n,r}x^r + ...$, pour $n \geq 1$ et $r \geq 0$. Montrer que $\alpha_{n,j} = C_{j+n-1}^{n-1}$ pour $0 \leq j \leq r$.

Réponse. Pour $n=1$ on lit directement $\alpha_{1,r} = 1$ pour tout $r \geq 0$.

Puis récurrence : on suppose que c'est vrai pour n donné et tout $r \geq 0$. Passons à $n+1$.

On a $(1+x+...+x^r)^{n+1} = (1+x+...+x^r)^n(1+x+...+x^r)$, donc :

$$(1+x+...+x^r)^{n+1} = (\alpha_{n,0} + \alpha_{n,1}x + ... + \alpha_{n,r-1}x^{r-1} + \alpha_{n,r}x^r + ...)(1+x+...+x^{r-1}+x^r),$$

polynôme dont le coefficient devant x^j pour $j \leq r$ vaut immédiatement $\alpha_{n+1,j} = \alpha_{n,0} + \alpha_{n,1} + ... + \alpha_{n,j} = C_{n-1}^{n-1} + C_n^{n-1} + ... + C_{j+n-1}^{n-1} = C_{j+n}^n$, cf. (1.27).

1.4.5 Coefficients multinomiaux

Question 5- : soit $\ell \geq 1$ et soit $E = \{a_1, ..., a_\ell\}$ un ensemble de ℓ objets.

Exemple 1.28 E est une alphabet constitué de ℓ lettres, avec $\ell = 26$ en France.

Soit $n \geq 1$ et soit $k_1, ..., k_\ell \in [0,\ell]_{\mathbb{N}}$ tels que $k_1 + ... + k_\ell = n$ (exemple : on s'intéresse au "mots" de longueur n écrits à l'aide de l'alphabet E).

Combien y-a-t-il de "mots" de longueur n contenant k_i fois la lettre a_i avec $k_1 + ... + k_m = n$? (Autrement dit quel est le nombre de partitions $U_1 \cup ... \cup U_\ell$ de E^n où U_i contient k_i fois a_i ?)

Exemple 1.29 Les mots de longueur $n=1$ qui contiennent une fois k_1 sont au nombre de 1 (le seul mot a_1).

Le nombre de mots de longueur 2 tels que $k_1 = 2$ est 1 (le seul mot a_1a_1). Le nombre de mots de longueur 2 tels que $k_1 = 1$ et $k_2 = 1$ est 2 (les mots a_1a_2 et a_2a_1).

Réponse :

$$\frac{n!}{k_1!...k_\ell!}, \tag{1.28}$$

appelé coefficient multinomial, voir formule (1.30) du multinôme de Newton, ci-dessous.

Démonstration 1- rapide. Soit $A \subset E^n$ l'ensemble des mots de longueur n contenant k_i fois la lettre a_i pour $i = 1, ..., \ell$, avec $k_i \in [0, n]_{\mathbb{N}}$ et $k_1 + ...k_\ell = n$. Son cardinal vaut, cf. (1.23) :

$$\begin{aligned} |A| &= C_n^{k_1} C_{n-k_1}^{k_2} C_{n-(k_1+k_2)}^{k_3} ... C_{n-(k_1+...+k_{\ell-1})}^{k_\ell} \\ &= \frac{n!}{k_1!(n-k_1)!} \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} ... = \frac{n!}{k_1!k_2!...k_\ell!}, \end{aligned} \tag{1.29}$$

= le nombre de combinaisons d'avoir k_1 fois a_1 dans un ensemble de n objets, multiplié par le nombre de combinaisons d'avoir k_2 fois a_2 dans l'ensemble de $n-k_1$ objets restants, multiplié par...

Démonstration 2- par récurrence. Pour $\ell = 1$ (une lettre), donc un seul mot $a_1....a_1$ de longueur n, donc $k_1 = n$, et donc $\frac{n!}{k_1!} = 1$.

Pour $\ell = 2$ (deux lettres), $k_1 + k_2 = n$, formule du binôme : $C_n^{k_1}$ mots contenant k_1 fois la lettre a_1, et donc k_2 fois la lettre a_2, et $C_n^{k_1} = \frac{n!}{k_1!k_2!}$.

Supposons que ce soit vrai pour $\ell \geq 1$. Passons à ℓ+1 (alphabet de ℓ+1 lettres).

Pour $n = 1$ (mot de longueur 1), si $k_1 = 1$ alors les autres $k_i = 0$ (on a le seul mot a_1), et $\frac{n!}{k_1!k_2!...k_\ell!} = \frac{1!}{1!0!...0!} = 1$. Idem si c'est k_i qui est non nul, pour $i \in [1, \ell]_{\mathbb{N}}$.

Supposons que de soit vrai pour $n \geq 1$, et passons à n+1. au moins un des k_i est non nul. Quitte à renuméroter, supposons $k_1 \neq 0$. La formule du binôme donne $C_{n+1}^{k_1}$ manière de placer k_1. Il reste ℓ lettres à placer parmi $n+1-k_1$, soit $\frac{(n+1-k_1)!}{k_2!...k_{\ell+1}!}$ par hypothèse de récurrence. Donc au total $C_{n+1}^{k_1} \frac{(n+1-k_1)!}{k_2!...k_{\ell+1}!} = \frac{[n+1)!}{k_1!...k_{\ell+1}!}$.

Exemple 1.30 Jeu "des chiffres et des lettres". $m = 26$, $n = 9$, $k_1 + ... + k_m = n$, où les $k_i \in \mathbb{N}$. Le nombre de mots de longueur 9 qui contiennent k_1 fois a_1, ..., k_m fois a_m est $\frac{9!}{k_1!...k_m!}$. ▪

1.4.6 Formule du multinôme (Newton)

Corollaire 1.31 *(Formule du multinôme.) Pour $x, y \in \mathbb{R}$ (ou $\in \mathbb{C}$), pour $n \in \mathbb{N}$ et pour $m \in \mathbb{N}^*$, on a :*

$$(x_1 + ... + x_m)^n = \sum_{\substack{(k_1,...,k_m) \in \mathbb{N}^m \\ k_1+...+k_m=n}} \frac{n!}{k_1!...k_m!} x_1^{k_1} ... x_m^{k_m}. \tag{1.30}$$

Preuve. Par récurrence. C'est vrai pour $m = 1$, ce pour tout n : trivial.

C'est vrai pour $m = 2$, ce pour tout n : formule du binôme.

Supposons que ce soit vrai pour m, ce pour tout n. Passons à $m + 1$:

$$\begin{aligned} &(x_1 + ... + x_{m+1})^n \\ &\quad = (x_1 + ... + (x_m + x_{m+1}))^n \\ &\quad = \sum_{k_1+...+k_{m-1}+K=n} \frac{n!}{k_1!...k_{m-1}!K!} x_1^{k_1} ... x_{m-1}^{k_{m-1}} (x_m + x_{m+1})^K \\ &\quad = \sum_{k_1+...+k_{m-1}+K=n} \frac{n!}{k_1!...k_{m-1}!K!} x_1^{k_1} ... x_{m-1}^{k_{m-1}} \sum_{k_m+k_{m+1}=K} \frac{K!}{k_m!k_{m+1}!} x_m^{k_m} x_{m+1}^{k_{m+1}} \\ &\quad = \sum_{k_1+...+k_{m-1}+K=n,} \sum_{k_m+k_{m+1}=K} \frac{n!}{k_1!...k_{m-1}!K!} \frac{K!}{k_m!k_{m+1}!} x_1^{k_1} ... x_{m-1}^{k_{m-1}} x_m^{k_m} x_{m+1}^{k_{m+1}}. \end{aligned}$$

▪

1.5 Gaussienne et sa masse

On rappelle qu'une gaussienne est une fonction de type, pour $a > 0$, $m \in \mathbb{R}$ et $c \neq 0$:

$$g(x) = c\, e^{-a(x-m)^2}. \tag{1.31}$$

Et la gaussienne réduite est la fonction $x \to e^{-\frac{x^2}{2}}$.

Proposition 1.32 *On a (masse de la gaussienne réduite = aire sous sa courbe) :*

$$\int_{\mathbb{R}} e^{-\frac{x^2}{2}}\, dx = \sqrt{2\pi}, \tag{1.32}$$

et plus généralement, pour $a > 0$ et $m \in \mathbb{R}$:

$$\int_{\mathbb{R}} e^{-a(x-m)^2}\, dx = \sqrt{\frac{\pi}{a}}. \tag{1.33}$$

En particulier $\displaystyle\int_{\mathbb{R}} e^{-\pi x^2}\, dx = 1$.

(Et $\displaystyle\int_{\mathbb{R}} e^{-\frac{x^2}{4\varepsilon}}\, dx = 2\sqrt{\varepsilon\pi}$ pour $\varepsilon > 0$, en posant $a = \frac{1}{4\varepsilon}$. Et on montre, voir cours de distribution que la fonction $x \to \frac{1}{2\sqrt{\varepsilon\pi}} e^{-\frac{x^2}{4\varepsilon}}$, de masse unité, tend vers la masse de Dirac δ_0 quand $\varepsilon \to 0$, au sens des distributions. On s'en servira pour la démonstration du théorème central limite.)

Preuve. L'astuce consiste à passer en coordonnées polaires, car $r \to re^{-r^2}$ a une primitive, à savoir $r \to -\frac{1}{2}e^{-r^2}$. Avec $r = \sqrt{x^2+y^2}$ on a :

$$\begin{aligned}(\int_{\mathbb{R}} e^{-ax^2}\, dx)^2 &= \iint_{\mathbb{R}^2} e^{-ax^2} e^{-ay^2}\, dxdy = \int_{r=0}^{\infty}\int_{\theta=0}^{2\pi} e^{-ar^2} r dr d\theta \\ &= (\int_{r=0}^{\infty} re^{-ar^2} dr)(\int_{\theta=0}^{2\pi} d\theta) = [\frac{-1}{2a} e^{-ar^2}]_0^{\infty} 2\pi = \frac{\pi}{a}.\end{aligned}$$

(Puis en remplaçant a par $\frac{1}{4\varepsilon}$, $\sqrt{\frac{\pi}{a}}$ devient $\sqrt{\frac{\pi}{\frac{1}{4\varepsilon}}} = 2\sqrt{\varepsilon\pi}$.)

Et pour $\int_{\mathbb{R}} e^{-a(x-m)^2}\, dx$ on fait le changement de variable $y = x - m$, donc $dy = dx$. ∎

1.6 Rappels sur les séries

On aura besoin des séries $\sum_{\mathbb{N}^*} u_n$ pour calculer des valeurs moyennes (espérances), on encore lorsque l'hypothèse $u_n \longrightarrow_{n\to\infty} 0$ ne sera pas suffisante alors que l'hypothèse $\sum_{\mathbb{N}^*} |u_n| < \infty$ sera suffisante (penser à $\sum_{\mathbb{N}^*} \frac{1}{n} = \infty$).

1.6.1 Définitions et premières propriétés

On note K le corps $\mathbb{R}$ des réels ou $\mathbb{C}$ des complexes. Un élément de K est appelé un scalaire. (On aura besoin de se placer dans $\mathbb{C}$ quand on utilisera les séries ou intégrales de Fourier, cf. le théorème central limite.) On a choisi ici $\mathbb{N}^*$ comme ensemble dénombrable. Tout autre ensemble dénombrable convient (ensemble en bijection avec $\mathbb{N}^*$).

Définition 1.33 Soit $(u_n)_{\mathbb{N}^*}$ une suite de scalaires (une suite de réels ou de complexes).

• $\sum_{n\in\mathbb{N}^*} u_n \overset{\text{noté}}{=} \sum_{\mathbb{N}^*} u_n$ est la série de terme u_n.

• La série partielle, ou somme partielle, d'ordre n est la somme finie $S_n = \sum_{i=1}^n u_i$.

• $(S_n)_{\mathbb{N}^*}$ est la suite des sommes partielles.

• La série $\sum_{\mathbb{N}^*} u_n$ est convergente dans K ssi la suite $(S_n)_{\mathbb{N}^*}$ est convergente dans K (donc converge vers une limite finie). Sinon elle est divergente.

• Quand $(S_n)_{\mathbb{N}^*}$ est convergente dans K, le scalaire $S = \lim_\infty S_n$ est noté $S = \sum_{\mathbb{N}^*} u_n$ et est appelé la somme de la série. Donc $\displaystyle\sum_{n\in\mathbb{N}^*} u_n \overset{\text{déf}}{=} \lim_{n\to\infty}(\sum_{i=1}^n u_i)$ quand la limite existe.

• La série $\sum_{\mathbb{N}^*} u_n$ est absolument convergente (ou sommable) ssi $\sum_{\mathbb{N}^*} |u_n| < \infty$.

(“Absolument convergente” = “sommable” = “Lebesgue intégrable pour la mesure de comptage”, voir cours d'intégration)

• La série $\sum_{\mathbb{N}^*} u_n$ est semi-convergente ssi elle est convergente mais n'est pas absolument convergente, i.e. ssi $\sum_{\mathbb{N}^*} u_n \in K$ et $\sum_{\mathbb{N}^*} |u_n| = \infty$.

• Extension du vocabulaire : dans $\mathbb{R}$, quand $\sum_{\mathbb{N}^*} u_n$ est t.q. $u_n \geq 0$ pour tout n, la suite $(S_n)_{\mathbb{N}^*}$ est alors croissante, et on dit que la série converge, au sens où elle converge dans $\overline{\mathbb{R}}_+$: elle converge dans $\mathbb{R}_+$ ou vers $+\infty$.

Proposition 1.34 *1- Si $\displaystyle\sum_{\mathbb{N}^*} u_n$ converge, alors $u_n \underset{n\to\infty}{\longrightarrow} 0$.*

2- Si $\displaystyle\sum_{\mathbb{N}^*} u_n$ et $\displaystyle\sum_{\mathbb{N}^*} v_n$ convergent, si $\lambda \in \mathbb{R}$, alors $\displaystyle\sum_{\mathbb{N}^*}(\lambda u_n + v_n)$ converge, et $\displaystyle\sum_{\mathbb{N}^*}(\lambda u_n + v_n) = \lambda\sum_{\mathbb{N}^*} u_n + \sum_{\mathbb{N}^*} v_n$.

3- Si $\sum_{\mathbb{N}^*} |u_n| < \infty$ (série absolument convergente), alors $\sum_{\mathbb{N}^*} u_n$ converge. La réciproque est fausse.

Preuve. 1- $S_n = \sum_{i=1}^n u_i \in K$ et soit $S = \sum_{\mathbb{N}^*} u_n \in K$ (la série converge). Autrement dit $S_n \longrightarrow_{n\to\infty} S$. Donc pour $n \geq 2$ on a $u_n = S_n - S_{n-1} \underset{n\to\infty}{\longrightarrow} S - S = 0$.

2- Soit $S_n = \sum_{i=1}^n u_i$, $S = \sum_{\mathbb{N}^*} u_i$, et $T_n = \sum_{i=1}^n v_i$, $T = \sum_{\mathbb{N}^*} v_i$. Alors $|(\lambda S+T)-(\lambda S_n+T_n)| \leq |\lambda|\,|S - S_n| + |T - T_n| \longrightarrow_{n\to\infty} \lambda 0 + 0 = 0$, donc la suite $(\lambda S_n + T_n)_{\mathbb{N}^*} = (\sum_{i=1}^n(\lambda u_i + v_i))_{\mathbb{N}^*}$ est convergente de limite $\lambda S + T$.

3- On suppose $\sum_{\mathbb{N}^*} |u_n| \in \mathbb{R}_+$. Soit $m \leq n$. On a $|S_n - S_m| = |\sum_{i=m}^n u_i| \leq \sum_{i=m}^n |u_i| \leq \sum_{i=m}^\infty |u_i| \longrightarrow_{m\to\infty} 0$, donc $(S_n)_{\mathbb{N}^*}$ est de Cauchy dans K complet. Donc $(S_n)_{\mathbb{N}^*}$ est convergente dans K.

La réciproque est fausse : prendre $u_n = \frac{(-1)^n}{n}$: on a $\sum_{\mathbb{N}^*} u_n = \sum_{\mathbb{N}^*} \frac{(-1)^n}{n} = \ln(2)$, voir exercice 1.42, mais $\sum_{\mathbb{N}^*} |u_n| = \sum_{\mathbb{N}^*} \frac{1}{n}$ diverge car $\sum_{i=1}^n \frac{1}{n} \geq \int_1^n \frac{1}{x}\,dx = \ln(n)$, faire un dessin. ▪

1.6.2 Séries absolument convergentes

On note, dans $\mathbb{R}$:

$$u_n^+ = \max(0, u_n) \quad (\geq 0) \qquad \text{et} \qquad u_n^- = -\min(0, u_n) \quad (\geq 0), \tag{1.34}$$

avec donc :

$$u_n = u_n^+ - u_n^- \qquad \text{et} \qquad |u_n| = u_n^+ + u_n^-. \tag{1.35}$$

Proposition 1.35 *Si $\displaystyle\sum_{n\in\mathbb{N}^*} |u_n| < \infty$ (série absolument convergente), alors pour tout sous-ensemble $I \subset \mathbb{N}^*$ on a $\displaystyle\sum_{i\in I} |u_i| < \infty$ (toute série partielle est absolument convergente), avec :*

$$\sum_{n\in\mathbb{N}^*} |u_n| = \sum_{i\in I} |u_i| + \sum_{j\in\mathbb{N}^*-I} |u_j|, \qquad \textit{et} \qquad \sum_{n\in\mathbb{N}^*} u_n = \sum_{i\in I} u_i + \sum_{j\in\mathbb{N}^*-I} u_j. \tag{1.36}$$

En particulier :

$$\sum_{\mathbb{N}^*} |u_n| = \sum_{\mathbb{N}^*} u_n^+ + \sum_{\mathbb{N}^*} u_n^-, \qquad \textit{et} \qquad \sum_{\mathbb{N}^*} u_n = \sum_{\mathbb{N}^*} u_n^+ - \sum_{\mathbb{N}^*} u_n^-. \tag{1.37}$$

Preuve. Soit $v_i = u_i$ si $i \in I$, et $v_i = 0$ sinon. Alors $\sum_{i=1}^n |v_i| \leq \sum_{i=1}^n |u_i| \leq \sum_{\mathbb{N}^*} |u_i|$. Donc la suite $(T_n = \sum_{i=1}^n |v_i|)_{\mathbb{N}^*}$ est une suite croissante bornée, donc convergente, donc la série $\sum_{i\in\mathbb{N}^*} |v_i| = \sum_{i\in I} |u_i|$ est convergente.

Idem pour $w_i = u_i$ si $i \in \mathbb{N}^*-I$, et $w_i = 0$ sinon.

Et $u_i = v_i + w_i$ (immédiat) et $|u_i| = |v_i| + |w_i|$ (immédiat), d'où $\sum_{i=1}^n |u_i| = \sum_{i=1}^n |v_i| + \sum_{i=1}^n |w_i|$, d'où $\sum_{\mathbb{N}^*} |u_i| = \sum_{\mathbb{N}^*} |v_i| + \sum_{\mathbb{N}^*} |w_i|$. D'où $\sum_{i=1}^n u_i = \sum_{i=1}^n v_i + \sum_{i=1}^n w_i$ donne $\sum_{\mathbb{N}^*} u_i = \sum_{\mathbb{N}^*} v_i + \sum_{\mathbb{N}^*} w_i$, cf. proposition 1.34,3-.

C'est en particulier vrai pour $I \subset \mathbb{N}^*$ t.q. $i \in I \Leftrightarrow u_i \geq 0$. ▪

Proposition 1.36 *(Sommation par paquets.) Soit $\sum_{\mathbb{N}^*}|u_n|<\infty$ (une série absolument convergente), soit $\mathcal{I}\subset\mathbb{N}^*$ (sous-ensemble fini ou dénombrable), et soit $I_i\subset\mathbb{N}^*$ pour tout $i\in\mathcal{I}$, t.q. $\bigcup_{i\in\mathcal{I}} I_i=\mathbb{N}^*$ est une partition (finie ou dénombrable) de $\mathbb{N}^*$. Notons $S_i=\sum_{j\in I_i}u_j\in\mathbb{R}$ pour tout $i\in\mathcal{I}$. Alors on a (cas des séries absolument convergentes) :*

$$S=\sum_{i\in\mathcal{I}}S_i,\qquad \text{soit}\qquad \sum_{n\in\mathbb{N}^*}u_n=\sum_{i\in\mathcal{I}}(\sum_{j\in I_i}u_j). \tag{1.38}$$

C'est en particulier vrai si la sommation se fait dans un ordre quelconque : $\sum_{n\in\mathbb{N}^*}u_n=\sum_{i\in\mathbb{N}^*}u_{\varphi(i)}$ pour toute bijection $\varphi:\mathbb{N}^*\to\mathbb{N}^*$.

C'est faux si $\sum_{\mathbb{N}^*}u_n$ est semi-convergente.

Preuve. Si $\mathcal{I}$ est fini on peut prendre $\mathcal{I}=[1,n]_{\mathbb{N}}$, et on développe la démonstration de la proposition 1.35 (récurrence).

Supposons $\mathcal{I}$ infini. Soit $j\in\mathbb{N}^*$ et soit $i_j\in\mathbb{N}^*$ t.q. $j\in I_{i_j}$ (possible car $\bigcup_{i\in\mathbb{N}^*}I_i$ est une partition). Soit $m\in\mathbb{N}^*$ et soit $J_m=\bigcup_{j=1}^m I_{i_j}$ (union finie). En particulier $[1,m]\subset J_m$.

Soit $K_m=\mathbb{N}^*-J_m$. En particulier si $n\in K_m$ alors $n>m$. Donc $\sum_{n\in K_m}|u_n|\le\sum_{n>m}|u_n|$.

La série étant absolument convergente, pour $\varepsilon>0$ donné, il existe $M\in\mathbb{N}^*$ t.q. pour tout $n\ge M$ on a $\sum_{i>n}|u_n|<\varepsilon$.

Donc $\sum_{n\in K_M}|u_n|\le\sum_{n>M}|u_n|<\varepsilon$. Avec $\sum_{j\in J_M}|u_j|+\sum_{k\in K_M}|u_k|=\sum_{\mathbb{N}^*}|u_n|$, cf. proposition 1.35. Donc $\left|\sum_{\mathbb{N}^*}|u_n|-\sum_{J_N}|u_n|\right|<\varepsilon$. Donc $\sum_{j\in J_n}|u_j|\big|\longrightarrow_{n\to\infty}\sum_{\mathbb{N}^*}|u_n|$.

En particulier vrai en remplaçant u_n par u_n^+ et par u_n^-. D'où (1.38).

Contre-exemple pour une série semi-convergente : voir exercice 1.44. ∎

Définition 1.37 Une famille $(a_{mn})_{(m,n)\in\mathbb{N}^{*2}}$ est une suite double (l'espace $\mathbb{N}^*\times\mathbb{N}^*$ est dénombrable). Et $\sum_{m,n\in\mathbb{N}^*}a_{mn}$ est une série double.

Proposition 1.38 *(Fubini–Tonelli). Soit $\sum_{\mathbb{N}^*\times\mathbb{N}^*}a_{mn}$ une série double. Si pour tout $n\in\mathbb{N}^*$ on a $\sum_{m\in\mathbb{N}^*}|a_{mn}|<\infty$ et si $\sum_{n\in\mathbb{N}^*}(\sum_{m\in\mathbb{N}^*}|a_{mn}|)<\infty$, alors pour tout $m\in\mathbb{N}^*$ on a $\sum_{n\in\mathbb{N}^*}|a_{mn}|<\infty$ et $\sum_{n\in\mathbb{N}^*}(\sum_{m\in\mathbb{N}^*}|a_{mn}|)=\sum_{m\in\mathbb{N}^*}(\sum_{n\in\mathbb{N}^*}|a_{mn}|)$. Idem en inversant les rôles de m et n.*

C'est faux si on oublie les valeurs absolues.

Preuve. Voir polycopié “Intégrale de Lebesgue”. ∎

1.6.3 Séries semi-convergentes

Proposition 1.39 *Si $\sum_{\mathbb{N}^*}u_n$ est une série réelle semi-convergente, alors pour tout $s\in\mathbb{R}$, il existe alors une bijection $k:\mathbb{N}^*\to\mathbb{N}^*$ telle $\sum_{\mathbb{N}^*}u_{k(n)}=s$ (on peut réordonner la série pour avoir n'importe quel réel s comme résultat de la somme).*

(Pour un exemple voir exercice 1.44.)

Preuve. Soit $U_+=\{u_i>0\}$ et $U_-=\{u_i\le 0\}$.

1- On a U_+ et U_- infinis, sinon $\sum_{\mathbb{N}^*}u_n$ est absolument convergente (immédiat). On note φ et ν les injections $\mathbb{N}^*\to\mathbb{N}^*$ croissantes t.q. $U_+=\{u_{\varphi(n)}\}$ et $U_-=\{u_{\nu(n)}\}$. Quitte à changer $(u_n)_{\mathbb{N}^*}$ en $(-u_n)_{\mathbb{N}^*}$, on suppose $\varphi(1)=1$ (et donc le premier terme u_1 est positif).

2- Soit $S_+=\sum_{\mathbb{N}^*}u_{\varphi(n)}$ et $S_-=\sum_{\mathbb{N}^*}u_{\nu(n)}$: montrons que $S_+=+\infty$ et $S_-=-\infty$.

Sinon, quitte à changer les u_n en les $-u_n$, on a $S_+=\sum_{\mathbb{N}^*}u_{\varphi(n)}<\infty$. Pour $n\in\mathbb{N}^*$ soit j_n t.q. $\varphi(j_n)\le\nu(n)\le\varphi(j_n+1)$; on a $\sum_{i=1}^{j_n}u_{\varphi(i)}+\sum_{i=1}^{\nu(n)}(-u_{\nu(i)})\le\sum_{i=1}^{\nu(n)}|u_i|\le\sum_{i=1}^{j_n+1}u_{\varphi(i)}+\sum_{i=1}^{n}(-u_{\nu(i)})$.

Donc si $S_->-\infty$ on a $\sum_{i=1}^{\nu(n)}|u_i|=S_++S_-<\infty$, contraire à l'hypothèse $\sum_{\mathbb{N}^*}u_n$ semi-convergente. Donc $S_-=-\infty$. Donc $S_+=+\infty$: sinon $\sum_{\mathbb{N}^*}u_n=-\infty$, car $\sum_{i=1}^{\nu(n)}u_i\le\sum_{i=1}^{j_n+1}u_{\varphi(i)}+\sum_{i=1}^{n}u_{\nu(i)}\le S_+-\sum_{i=1}^{n}u_{\nu(i)}\to-\infty$.

3- D'où la construction de la bijection k : on part de u_1. Soit s un réel. Supposons $u_1\le s$ (cas $u_1\ge s$ similaire). On ajoute les termes successifs $u_{\varphi(i)}$ et on s'arrête dès que la somme devient $>s$, ce qui est possible car $S_+ = \infty$. Puis on ajoute les termes successifs $u_{\nu(i)}$ et on s'arrête dès que la somme devient $< s$, ce qui est possible car $S_- = -\infty$. Et on itère. On a ainsi construit notre bijection k :

$$(x_1, x_2, ...) = (x_{\varphi(1)}, ..., x_{\varphi(p_1)}, x_{\nu(1)}, ..., x_{\nu(n_1)}, x_{\varphi(p_1+1)}, ..., x_{\varphi(p_2)}, x_{\nu(n_1+1)}, ...),$$

suite alternée de paquets positifs et négatifs.

4- Vérifions que $\sum_{\mathbb{N}^*} u_{k(n)} = s$. Comme $\sum_{\mathbb{N}^*} u_n$ est convergente, on a $u_n \to 0$. Soit $\varepsilon > 0$ et soit N t.q. pour tout $n \geq N$ on a $|u_{\varphi(n)}| < \varepsilon$ et $|u_{\nu(n)}| < \varepsilon$. Par construction de la bijection, on a $|\sum_{i=1}^n u_{k(i)} - s| < \varepsilon$ pour $n \geq N$. ∎

Proposition 1.40 *Soit une application* $\varphi : \begin{Bmatrix} \mathbb{N}^* & \to \mathbb{N}^* \\ i & \to \varphi(i) \end{Bmatrix}$ *strictement croissante avec* $\varphi(1) = 1$.

Soit $v_n = u_{\varphi(n)} + ... + u_{\varphi(n+1)-1} = \displaystyle\sum_{i=\varphi(n)}^{\varphi(n+1)-1} u_i$ *(un "paquet de termes consécutifs").*

• *1 : si* $\sum_{\mathbb{N}^*} u_n$ *converge alors* $\sum_{\mathbb{N}^*} v_n$ *converge et* $\sum_{\mathbb{N}^*} v_n = \sum_{\mathbb{N}^*} u_n$ *(ici on ne réordonne pas la série).*

Et la réciproque est fausse en général. Elle est vraie, i.e. $\sum_{\mathbb{N}^*} v_n$ *converge implique* $\sum_{\mathbb{N}^*} u_n$ *converge, par exemple dans les cas suivants :*

• *20 : la série est absolument convergente (proposition 1.36).*

• *21 : s'il existe* $M \in \mathbb{N}^*$ *t.q.* $\varphi(n+1) - \varphi(n) \leq M$ *pour tout* $n \in \mathbb{N}^*$ *("chaque paquet contient au plus* M *termes"), si* $u_n \underset{n\to\infty}{\longrightarrow} 0$, *et si* $\sum_{\mathbb{N}^*} v_n$ *converge, alors* $\sum_{\mathbb{N}^*} u_n$ *converge et* $\sum_{\mathbb{N}^*} u_n = \sum_{\mathbb{N}^*} v_n$.

• *22 : si* $\displaystyle\sum_{n=\varphi(i)}^{\varphi(i+1)-1} |u_n| \underset{i\to\infty}{\longrightarrow} 0$, *et si* $\sum_{\mathbb{N}^*} v_n$ *converge, alors* $\sum_{\mathbb{N}^*} u_n$ *converge et* $\sum_{\mathbb{N}^*} u_n = \sum_{\mathbb{N}^*} v_n$.

• *23 : si à l'intérieur de "chaque paquet"* $I_i = \{n \in [\varphi(i), \varphi(i+1) - 1]_{\mathbb{N}}\}$ *les* u_n *ont même signe, et si* $\sum_{\mathbb{N}^*} v_n$ *converge, alors* $\sum_{\mathbb{N}^*} u_n$ *converge et* $\sum_{\mathbb{N}^*} u_n = \sum_{\mathbb{N}^*} v_n$.

Preuve. • 1 : soit $S_n = \sum_{i=1}^n u_i$. Alors $\sum_{i=1}^n v_i = \sum_{i=1}^{\varphi(n+1)-1} u_i = S_{\varphi(n+1)-1}$. Comme (S_n) est convergente vers S, il en est de même de toute sous-suite extraite. En particulier $(S_{\varphi(n+1)-1})_{n\in\mathbb{N}^*}$ converge vers S.

La réciproque est fausse : par exemple prendre $u_n = (-1)^n$ et $\varphi(i) = 2i - 1$: on a $v_k = 0$ pour tout $k \geq 1$ donc $\sum_{\mathbb{N}^*} v_n = 0$ alors que $\sum_{\mathbb{N}^*} u_n$ est divergente.

• 21 : Notons $T = \sum_{\mathbb{N}^*} v_n$. Soit $N \in \mathbb{N}^*$ et $i \in \mathbb{N}^*$ t.q. $\varphi(i) \leq N \leq \varphi(i+1) - 1$. Alors :

$$|\sum_{n=1}^N u_n - S| = |\sum_{n=1}^{\varphi(i)-1} u_n - T + \sum_{n=\varphi(i)}^{\varphi(i+1)} u_n| \leq |\sum_{n=1}^{i} v_i - T| + \sum_{n=\varphi(i)}^{\varphi(i+1)-1} |u_n|. \tag{1.39}$$

Le premier terme tend vers 0 par définition de T, et le second est une somme d'au plus M termes qui tend vers 0 (on a $\varphi(i+1) - \varphi(i) \leq M$ et $u_n \to 0$, donc $\sum_{n=\varphi(i)+1}^{\varphi(i+1)} |u_n| \leq M\varepsilon$ dès que i est assez grand pour avoir $|u_n| \leq \varepsilon$ quand $n \geq \varphi(i)$).

• 22 : on a (1.39), et la conclusion est immédiate.

• 23 : $\sum_{\mathbb{N}^*} v_n$ existe, donc $v_n \to 0$, et $|v_n| = |\sum_{n=\varphi(i)}^{\varphi(i+1)-1} u_n| = \sum_{n=\varphi(i)}^{\varphi(i+1)-1} |u_n|$ car le signe est constant dans chaque paquet, donc $\sum_{n=\varphi(i)}^{\varphi(i+1)-1} |u_n| \to 0$. On conclut avec • 22. ∎

1.6.4 Exemples

Exercice 1.41 Soit $(u_n)_{\mathbb{N}^*}$ une suite t.q. $u_n = (-1)^{n+1}|u_n|$ pour tout n (suite alternée commençant par un terme positif). Montrer que si $(|u_n|)_{\mathbb{N}^*}$ est décroissante et $u_n \underset{n\to\infty}{\longrightarrow} 0$ alors $S_2 \leq S_{2n} \leq S_{2n+2} \leq S_{2n+1} \leq S_{2n-1} \leq S_1$ pour tout n, et donc que $\sum_{\mathbb{N}^*} u_n$ est convergente.

Réponse. 1- On a $u_{2n+1} \geq 0$ et $u_{2n} \leq 0$ pour tout n. D'où $S_{2n+2} - S_{2n} = u_{2n+1} + u_{2n+2} = u_{2n+1} - (-u_{2n+2})$ avec $|u_{2n+1}| \geq |-u_{2n+2}|$ car $(|u_n|)$ est décroissante, d'où $S_{2n+2} - S_{2n} \geq 0$, pour tout n. Et $S_{2n+1} - S_{2n-1} = u_{2n} + u_{2n+1} = -(-u_{2n} - u_{2n+1})$ avec $|u_{2n}| \leq |-u_{2n+1}|$ car $(|u_n|)$ est décroissante, d'où $S_{2n+1} - S_{2n-1} \leq 0$, pour tout n. Et $S_{2n+2} - S_{2n+1} = u_{2n+2} \leq 0$ pour tout n.

2- D'où 21- $(S_{2n})_{\mathbb{N}^*}$ est croissante et bornée donc convergente, 22- $(S_{2n-1})_{\mathbb{N}^*}$ est décroissante et bornée donc convergente, 23- $S_{2n} - S_{2n-1} = u_{2n} \to 0$ car $u_n \to 0$, et donc 24- $\lim S_{2n} = \lim S_{2n+1}$, donc $= \lim S_n$. ∎

Exercice 1.42 Montrer que $\displaystyle\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n} = \ln(2)$.

Réponse. $\ln(1+x) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{x^n}{n}$ pour $x \in]-1,1]$ (développement en séries entières). ▪

Exercice 1.43 Montrer que $\displaystyle\sum_{n=1}^{\infty}\frac{1}{n(2n-1)} = 2\ln(2)$. On notera $\gamma = \lim_{N\to\infty}\big((\sum_{n=1}^{N}\frac{1}{n}) - \ln(N)\big)$ = la constante d'Euler, i.e. $\gamma = \lim_{N\to\infty}\big((\sum_{n=1}^{N}\frac{1}{n}) - (\int_1^N \frac{dx}{x})\big)$, faire un dessin.

Réponse. $\dfrac{1}{n(2n-1)} = \dfrac{2}{2n-1} - \dfrac{1}{n}$ (décomposition en éléments simples). D'où (sommes finies) :

$$\sum_{n=1}^{N}\frac{1}{n(2n-1)} = \sum_{n=1}^{N}\frac{2}{2n-1} - \sum_{n=1}^{N}\frac{1}{n} = 2\sum_{n=1}^{N}\frac{1}{2n-1} + 2\sum_{n=1}^{N}\frac{1}{2n} - \sum_{n=1}^{N}\frac{2}{2n} - \sum_{n=1}^{N}\frac{1}{n} = 2\sum_{n=1}^{2N}\frac{1}{n} - 2\sum_{n=1}^{N}\frac{1}{n}.$$

Avec $\displaystyle\sum_{n=1}^{N}\frac{1}{n} = \ln(N)+\gamma+o(1)$ quand $N\to\infty$. D'où $\displaystyle\sum_{n=1}^{N}\frac{1}{n(2n-1)} = 2(\ln(2N)-\ln(N)+o(1)) = 2\ln(2)+o(1)$ au voisinage de $N=\infty$. ▪

Exercice 1.44 Donner un exemple d'une série $\sum_{\mathbb{N}^*} u_n$ convergente et d'une bijection $k : \mathbb{N}^* \to \mathbb{N}^*$ telles que $\sum_{n\in\mathbb{N}^*} u_n \neq \sum_{n\in\mathbb{N}^*} u_{k(n)}$ (cas particulier de la proposition 1.39).

Réponse. Soit $(u_n)_{\mathbb{N}^*}$ la suite alternée $u_n = \frac{(-1)^{n+1}}{n}$. On a $\displaystyle\sum_{n=1}^{\infty} u_n = \ln(2)$, voir exercice 1.42. On modifie l'ordre de la suite (u_n) en posant :

$$v_1 = +1,\ v_2 = \frac{-1}{2},\ v_3 = \frac{-1}{4},\ v_4 = \frac{+1}{3},\ v_5 = \frac{-1}{6},\ v_6 = \frac{-1}{8},\ v_7 = \frac{+1}{5},\ v_8 = \frac{-1}{10},\ v_9 = \frac{-1}{12}, ...,$$

soit donc pour tout $k\in\mathbb{N}^*$:

$$v_{3k-2} = \frac{+1}{2k-1} = u_{2k-1},\ v_{3k-1} = \frac{-1}{4k-2} = u_{4k-2},\ v_{3k} = \frac{-1}{4k} = u_{4k}.$$

Donc :

$$2(v_{3k-2}+v_{3k-1}+v_{3k}) = 2(\frac{1}{4k-2}+\frac{-1}{4k}) = \frac{1}{2k-1}+\frac{-1}{2k} = u_{2k-1}+u_{2k}. \tag{1.40}$$

Donc $2\sum_{i=1}^{\infty} v_i = \sum_{i=1}^{\infty} u_i$: donc $\sum_{i=1}^{\infty} v_i \neq \sum_{i=1}^{\infty} u_i$.

(Ou encore $2(v_{3k-2}+v_{3k-1}+v_{3k}) = \dfrac{1}{2}\dfrac{1}{k(2k-1)}$ donne $2\displaystyle\sum_{i=1}^{\infty} v_i = \frac{1}{2}\sum_{n=1}^{\infty}\frac{1}{n(2n-1)} = \ln(2)$ avec l'exercice 1.43. ▪

Exercice 1.45 Une somme finie est associative et commutative. Donner un exemple ou c'est faux pour une somme infinie.

Réponse. Soit $u_n = (-1)^n\ln(1+\frac{1}{n})$. Avec l'exercice 1.41, la série $\sum_{\mathbb{N}^*} u_n$ est convergente car (u_n) est alternée et de terme général décroissant (immédiat) t.q. $u_n \to 0$ (immédiat).

On a $\ln(1+\frac{1}{n}) = \ln(\frac{n+1}{n}) = \ln(n+1) - \ln(n)$. Et :

$$\sum_{n=1}^{\infty} u_n = \sum_{n=1}^{\infty}(-1)^n(\ln(n+1)-\ln(n)) = (\ln(1)-\ln(2)) + (\ln(3)-\ln(2)) + (\ln(3)-\ln(4)) + ...$$
$$\neq\ -2\ln(2) + 2\ln(3) + ... = \sum_{n=1}^{\infty}(-1)^{n+1}2\ln(n),$$

non égalité car la série $\sum_{n=1}^{\infty}(-1)^{n+1}2\ln(n)$ est divergente (terme général $2\ln(n)\to\infty$). ▪

2 Espace probabilisé

Les probabilités "permettent de mettre à jour les régularités qui se cachent derrière ces phénomènes apparemment gouvernés par le hasard", citation prise dans [18].

2.1 Univers et événements, premières définitions

Une expérience [an experiment] est dite aléatoire [random] si, répétée dans les "mêmes" conditions, elle peut donner des résultats différents.

Exemple 2.1 Lancers répétés d'une pièce, durée de vie de lampes parmi un lot de 1000 lampes, durée d'attente à une caisse de supermarché, défaut on non d'une pièce à la sortie d'une chaîne de fabrication... ▟

Définition 2.2 Soit Ω l'ensemble de tous les résultats possibles (les observations possibles) d'une expérience aléatoire [the results of an experiment, or observations]. Cet ensemble Ω est appelé l'univers (ou univers des possibles, ou ensemble fondamental) [sample space].

Exemple 2.3 On lance une pièce [we toss a coin, we flip a coin, we play heads or tails]. Si on note $p =$ pile [tails] et $f =$ face [heads], on a $\Omega = \{p, f\}$. Et on cherchera à déterminer la probabilité d'obtenir pile ou face (probabilité dans Ω), probabilité notée $P(\omega)$ quand $\omega = p$ ou f. Par exemple, pour une pièce équilibrée on a $P(p) = P(f) = \frac{1}{2}$, soit $P(\omega) = \frac{1}{2}$ pour $\omega \in \Omega$. ▟

Exemple 2.4 Pour deux lancers successifs d'une pièce, on a, l'ordre étant pris en compte, $\Omega = \{(p,p),(p,f),(f,p),(f,f)\} =^{\text{noté}} \{pp, pf, fp, ff\}$: considéré alors comme l'ensemble des mots de longueur 2 contenant les lettres p et f (l'ordre est pris en compte).

Et si l'ordre ne compte pas (un lancer de deux pièces), on a $\Omega = \{\{p,p\},\{p,f\},\{f,f\}\}$. ▟

Exemple 2.5 Pour la mesure d'un bruit pendant le temps $[a,b]$, on a $\Omega = L^2([a,b];\mathbb{R})$ l'ensemble des fonctions d'énergie finie. ▟

Exemple 2.6 On mesure le temps d'attente à un feu rouge : $\Omega = \mathbb{R}_+$. ▟

Exemple 2.7 Résultat d'un tir sur une cible : $\Omega = \mathbb{R}^2$. ▟

Exemple 2.8 Prix d'une action en bourse : $\Omega = C([t_1,t_2],\mathbb{R}_+)$ (fonctions continues) sur un intervalle de temps $[t_1,t_2]$. ▟

On notera $|\Omega|$ le cardinal de l'univers Ω (on s'en sert dans le cas Ω ensemble fini) :

$$|\Omega| = \text{card}(\Omega). \tag{2.1}$$

Définition 2.9 Un élément ω de Ω (un résultat possible) est appelé une éventualité, ou un événement élémentaire, ou une issue [a simple event, or indecomposable event, or sample point].

Donc Ω est l'ensemble des éventualités (ensemble des événements élémentaires).

Définition 2.10 Un sous-ensemble $A \subset \Omega$ est appelé un événement (ou événement aléatoire) [an event, or compound event, or decomposable event, or aggregate of sample points]. Un événement est donc un ensemble d'éventualités.

Dans un espace probabilisable, on verra que l'ensemble des événements constitue la tribu (la σ-algèbre).

Définition 2.11 L'événement Ω et appelé événement certain (on est certain que le résultat est dans Ω par définition de l'univers Ω).

L'événement $\emptyset$ est appelé événement impossible (après un tirage on a au moins un résultat).

L'événement contraire de A est l'événement $A^C = \Omega - A$ (complémentaire de A dans Ω).

Remarque 2.12 Cette définition d'un événement suggère que l'ensemble des événements est $\mathcal{P}(\Omega) =$ l'ensemble de tous les sous-ensembles de Ω. On affinera cette définition au paragraphe suivant : on ne considèrera que les événements mesurables, événement qu'on pourra alors mesurer (avec une mesure de probabilité). ▟

Définition 2.13 Après une expérience donnant un résultat ω, on dit qu'un événement A est réalisé si $\omega \in A$.

Deux événements A et B sont incompatibles si $A \cap B = \emptyset$ (dans ce cas, un résultat ne peut pas être à la fois dans A et dans B).

Exemple 2.14 (Alternative simple.) On lance une pièce de monnaie, et on s'intéresse uniquement au résultat pile=p ou face=f : $\Omega = \{p, f\}$ est l'ensemble des résultats possibles. Les éventualités sont $\{p\}$ et $\{f\}$. Les événements sont $\emptyset$, $\{p\}$, $\{f\}$ et Ω. Et si le résultat d'un lancer est pile, les événements $\{p\}$ et Ω sont réalisés, alors que les événements $\emptyset$ et $\{f\}$ ne le sont pas.

Les événements $\{p\}$ et $\{f\}$ sont incompatibles. ▟

Exemple 2.15 On lance une pièce deux fois (en tenant compte de l'ordre de sortie). L'univers est $\Omega = \{p, f\}^2 = \{(p, p), (p, f), (f, p), (f, f)\}$ = l'ensemble des 4 résultats possibles de l'expérience. Et $\{(f, f)\}$ est l'événement constitué du singleton (f, f) qui est une éventualité. ▟

Exemple 2.16 On lance une pièce n fois (en tenant compte de l'ordre de sortie). L'univers est $\Omega = \{p, f\}^n$ l'ensemble des 2^n résultats possibles de l'expérience. Chaque éventualité $\omega \in \Omega$ est une suite finie $\omega = (\omega_1, ..., \omega_n)$ où $\omega_i \in \{p, f\}$ pour tout $i = 1, ..., n$. Autrement dit, une éventualité ω un est mot de longueur n ne contenant que les lettres p, f, et un événement est un ensemble de mots. ▟

Exemple 2.17 (Alternative généralisée.) On lance un dé [we throw a dice] : $\Omega = \{1, 2, 3, 4, 5, 6\}$ = l'ensemble des résultats possibles, et les événements possibles sont les sous-ensembles de Ω, i.e. tout ensemble $A \in \mathcal{P}(\Omega)$, où $\mathcal{P}(\Omega)$ est l'ensemble des ensembles de Ω. Par exemple 4 est une éventualité, et $\{4\}$ et $\{1, 2, 5\}$ sont des événements. ▟

2.2 Tribu et événement : définitions mathématiques

(Voir polycopié "Intégrale de Lebesgue".)

Quand Ω est fini ou dénombrable, on prendra généralement $\mathcal{P}(\Omega)$ comme ensemble des événements, i.e. on pourra mesurer la probabilité de tout sous-ensemble de Ω.

En revanche, quand Ω n'est pas fini ou dénombrable (exemple $\Omega = \mathbb{R}_+$), $\mathcal{P}(\Omega)$ pourra être trop grand : un sous-ensemble A de Ω pourrait être trop "pathologique" pour pouvoir être mesuré de manière usuelle (i.e. avec la "longueur usuelle" = la mesure de Lebesgue).

On va définir "le plus grand sous-ensemble possible" de $\mathcal{P}(\Omega)$, la tribu ou σ-algèbre [σ-field or σ-algebra], qui exclu les sous-ensembles "pathologiques". Une telle tribu sera stable par les opérations d'union, d'intersection, de complémentation.

Définition 2.18 Une tribu ou σ-algèbre [σ-algebra, σ-field] $\mathcal{T}$ sur Ω est un sous-ensemble de $\mathcal{P}(\Omega)$ tel que :

1- $\emptyset, \Omega \in \mathcal{T}$,

2- Si $A \in \mathcal{T}$ alors son complémentaire $A^C = \Omega - A$ est dans $\mathcal{T}$,

3- $\mathcal{T}$ est stable pour toute union ou intersection finie ou dénombrable. I.e., si I est un ensemble fini ou dénombrable, si $(A_i)_{i \in I}$ est une famille d'éléments de $\mathcal{T}$, alors $\bigcup\limits_{i \in I} A_i \in \mathcal{T}$ et $\bigcap\limits_{i \in I} A_i \in \mathcal{T}$.

Remarque 2.19 Le point 2- car on voudra : si on peut déterminer la probabilité $P(A)$ de A alors nécessairement on peut déterminer la probabilité de A^C : c'est $P(A^C) = 1 - P(A)$.

Le point 3- car on voudra : si $(A_n)_{\mathbb{N}}$ est suite croissante dans $\mathcal{T}$ (ensembles emboités), alors $\lim\limits_{n \to \infty} P(A_n) = P(\lim\limits_{n \to \infty} A)$, et on a donc besoin de $\lim\limits_{n \to \infty} A = \bigcup\limits_{n \in \mathbb{N}} A_n \in \mathcal{T}$. Idem pour une suite décroissante. ▟

Remarque 2.20 La définition fait apparaître des redondances : il suffit de supposer 1- $\emptyset \in \mathcal{T}$, 2- idem, 3- $\mathcal{T}$ est stable pour toute union finie ou dénombrable. Les autres propriétés découlent de 2-. ▟

Définition 2.21 Un ensemble Ω muni d'une tribu $\mathcal{T}$ (i.e. sur lequel on se donne une tribu) est appelé un espace mesurable et est noté $(\Omega, \mathcal{T})$, ou simplement Ω si la tribu est implicite.

Définition 2.22 Et dans un espace mesurable $(\Omega, \mathcal{T})$, un sous-ensemble $A \subset \Omega$ tel que $A \in \mathcal{T}$ est appelé un événement (ou événement réalisable, ou événement aléatoire).

N.B. : on a donc restreint le sens donné à la définition 2.10 de manière à exclure les sous-ensembles "pathologiques" de Ω.

Exemple 2.23 Pour Ω quelconque, $\mathcal{T} = \{\emptyset, \Omega\}$ est toujours une tribu, appelée la tribu grossière. Dans la suite, on ne considèrera pas cette tribu. ▪

Exemple 2.24 Si $A \subset \Omega$ avec $A \neq \emptyset$ et $A \neq \Omega$, l'ensemble $\mathcal{T} = \{\emptyset, A, A^C, \Omega\}$ est une tribu. En effet, les conditions 1-, 2- et 3- sont trivialement satisfaites.

Cette tribu sera utilisée pour l'espérance conditionnelle $E(Y|X)$ d'une variable aléatoire par une autre. ▪

Exemple 2.25 Pour Ω quelconque, $\mathcal{T} = \mathcal{P}(\Omega)$ est toujours une tribu, appelée la tribu discrète.

C'est le cas des exemples de lancers de pièces et de dés du paragraphe précédent, ou de statistiques sur une population. ▪

Exemple 2.26 Dans $\Omega = \mathbb{R}$, on définit la tribu borélienne $\mathcal{T}_\mathbb{R}$ comme étant la tribu engendrée par tous les intervalles ouverts (et donc aussi par tous les intervalles fermés, par complémentation). Ainsi les événements pourront être mesurés avec la mesure de Lebesgue (ou plus généralement avec les mesures “de densité”).

Cette tribu borélienne est très grande, à tel point que, pour un ingénieur, elle ne se différencie pas vraiment de $\mathcal{P}(\Omega)$ au sens où on ne connaît pas de sous-ensemble explicite de $\mathbb{R}$ appartenant à $\mathcal{P}(\mathbb{R}) - \mathcal{T}_\mathbb{R}$: il faut faire appel à l'axiome du choix pour montrer qu'il existe des sous-ensembles de $\mathbb{R}$ qui n'appartiennent pas à la tribu borélienne de $\mathbb{R}$ (voir polycopié “Intégrale de Lebesgue”).

Donc à retenir : on introduit les tribus pour que dans la suite le raisonnement mathématique ne soit pas mis en défaut (on exclut les sous-ensembles pathologiques) ; et dans la pratique ingénieur, quand on “verra” un sous-ensemble de Ω, il sera dans la tribu (ce sera un événement). ▪

Exemple 2.27 Si $\Omega = \mathbb{R}$ et si on s'intéresse aux “valeurs ponctuelles”, on considèrera $\mathcal{T} = P(\Omega)$. Mais on peut également prendre $\mathcal{T} = \mathcal{T}_\mathbb{R}$.

Dans les deux cas on pourra considérer les mesures discrètes comme les mesures δ_x de Dirac par exemple, cf. (1.13), que l'on prenne A dans $\mathcal{P}(\mathbb{R})$ ou dans $\mathcal{T}_\mathbb{R}$ ($\subsetneq \mathcal{P}(\mathbb{R})$). ▪

2.3 Tribu engendrée

Soit $(A_i)_{i\in I}$ une famille de sous-ensembles de Ω.

Définition 2.28 La plus petite tribu contenant tous les A_i (i.e. l'intersection de toutes les tribus contenant tous les A_i) est appelée la tribu engendrée par les A_i, et est notée $\sigma((A_i)_{i\in I})$.

Exemple 2.29 La tribu $\mathcal{T} = \{\emptyset, A, A^C, \Omega\}$ est la tribu engendrée par A, tribu notée $\sigma(A)$. ▪

Exemple 2.30 La tribu borélienne $\mathcal{T}_\mathbb{R}$ est par définition engendrée par les intervalles de $\mathbb{R}$. ▪

Définition 2.31 Si $f : (\Omega, \mathcal{T}) \to \mathbb{R}$ est une fonction mesurable (i.e. t.q. $f^{-1}(]a,b[) \in \mathcal{T}$ pour tout $a, b \in \mathbb{R}$, voir plus loin), alors on appelle tribu engendrée par f la tribu notée $\sigma(f)$ engendrée par tous les ensembles “images réciproques” $A = f^{-1}(]a,b[)$, pour tout $a, b \in \mathbb{R}$.

Exercice 2.32 Montrer que pour $A \in \mathcal{T}$ et $f = 1_A$, on a $\sigma(1_A) = \{\emptyset, A, A^C, \Omega\} = \sigma(A)$.

Réponse. On a $1_A(x) = 1$ si $x \in A$, et $= 0$ sinon. En particulier $f(\Omega) = \{0, 1\}$. Donc :
si $]a,b[\supset \{0,1\}$, alors $f^{-1}(]a,b[) = \Omega$,
si $]a,b[\cap \{0,1\} = \emptyset$, alors $f^{-1}(]a,b[) = \emptyset$,
si $]a,b[\ni 1$ et $]a,b[\not\ni 0$, alors $f^{-1}(]a,b[) = A$,
si $]a,b[\ni 0$ et $]a,b[\not\ni 1$, alors $f^{-1}(]a,b[) = A^C$. ▪

2.4 Probabilité

But : connaître le “pourcentage de chance” (= la probabilité) qu'un événement a de se réaliser.

2.4.1 Définition des mesures (d'ensembles)

Soit $(\Omega, \mathcal{T})$ un ensemble mesurable (un ensemble muni d'une tribu).

Définition 2.33 Une application :

$$\mu : \begin{cases} \mathcal{T} & \to \mathbb{R}_+ \\ A & \to \mu(A) \end{cases} \tag{2.2}$$

est une mesure dans l'espace mesurable $(\Omega, \mathcal{T})$ ssi :

1. $\mu(\emptyset) = 0$.
2. - μ est σ-additive [countable additive], i.e. si I est un ensemble fini ou dénombrable, si $(A_i)_{i\in I} \in \mathcal{T}^I$ est une famille (au plus dénombbrable) d'événements 2 à 2 disjoints, alors la mesure de l'union est la somme des mesures :

$$\text{si} \quad \forall i,j \in I \quad \text{t.q.} \quad i \neq j \quad \text{on a} \quad A_i \cap A_j = \emptyset \qquad \text{alors} \qquad \mu(\bigcup_{i\in I} A_i) = \sum_{i\in I} \mu(A_i). \tag{2.3}$$

   Et $\mu(A)$ est alors appelée la mesure de l'ensemble A (quand $A \in \mathcal{T}$).
3. On ajoute souvent la propriété (sauf en probabilités) : μ doit être σ-finie, i.e. Ω est réunion finie ou dénombrable d'ensembles $A_i \subset \mathcal{T}$ de mesure finie (voir cependant la proposition 4.23 de la mesure image où cette propriété ne sera pas supposée).

Exemple 2.34 Mesure μ_ℓ de Lebesgue sur $\mathbb{R}$: $\mu_\ell(]a,b[) = b-a$ pour $a \leq b$, faire un dessin. ▪

Définition 2.35 Un espace mesurable $(\Omega, \mathcal{T})$ muni d'une mesure μ est noté $(\Omega, \mathcal{T}, \mu)$ et est appelé un espace mesuré.

Exercice 2.36 Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, et A, B deux événements. Montrer :

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B), \tag{2.4}$$

écriture équivalente à :

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{2.5}$$

Réponse. Faire un dessin. Notons $A' = A - (A\cap B)$. On a $A' \cap (A \cap B) = \emptyset$, donc $\mu(A) = \mu(A') + \mu(A\cap B)$, donc $\mu(A') = \mu(A) - \mu(A \cap B)$. De même, avec $B' = B - (A \cap B)$, $\mu(B') = \mu(B) - \mu(A \cap B)$. Et $A \cup B = A' \cup (A \cap B) \cup B'$, les ensembles A', B' et $A \cap B$ étant disjoints 2 à 2. D'où $\mu(A \cup B) = \mu(A') + \mu(A \cap B) + \mu(B') = \mu(A) + \mu(B) - \mu(A \cap B)$. ▪

Exercice 2.37 Si $(A_i)_{i\in[1,n]}$ est une famille d'événements, $n \in \mathbb{N}^*$, montrer la formule du crible :

$$\mu(\bigcup_{i=1}^n A_i) = \sum_i \mu(A_i) - \sum_{i<j} \mu(A_i \cap A_j) + \sum_{i<j<k} \mu(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n+1} \mu(A_1 \cap \ldots \cap A_n). \tag{2.6}$$

Réponse. Par récurrence. ▪

2.4.2 Première propriété

Rappel : une suite $(A_n)_{n\in\mathbb{N}^*}$ dans $\mathcal{P}(\Omega)$ est dite croissante (resp. décroissante) ssi pour tout $n \in \mathbb{N}^*$ on a $A_n \subset A_{n+1}$ (resp. $A_{n+1} \subset A_n$). Et une suite $(A_n)_{n\in\mathbb{N}^*}$est dite monotone ssi elle est soit croissante, soit décroissante.

De plus, si $(A_n)_{\mathbb{N}^*}$ est croissante, on note :

$$\bigcup_{k=1}^{\infty} A_n \overset{\text{noté}}{=} \sup_{n\in\mathbb{N}^*} A_n \overset{\text{noté}}{=} \lim_{n\to\infty} A_n, \tag{2.7}$$

et si $(A_n)_{\mathbb{N}^*}$ est décroissante, on note :

$$\bigcap_{k=1}^{\infty} A_n \overset{\text{noté}}{=} \inf_{n\in\mathbb{N}^*} A_n \overset{\text{noté}}{=} \lim_{n\to\infty} A_n. \tag{2.8}$$

Proposition 2.38 *Si $(\Omega, \mathcal{T}, \mu)$ est un espace mesuré, si $(A_n)_{\mathbb{N}^*}$ est une suite croissante d'ensembles mesurables, alors :*

$$\mu(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mu(A_n) \quad \text{dans } [0,\infty]. \tag{2.9}$$

On peut passer à la limite sous le signe μ pour les suites croissante. On dit que μ est croissante.

Idem si $(A_n)_{n\in\mathbb{N}^*}$ est décroissante telle que $\mu(A_1) < \infty$ (toujours vrai pour une mesure finie).

Preuve. Exercice (ou voir polycopié intégration). ∎

Remarque 2.39 Si (A_n) est décroissante, l'hypothèse $\mu(A_1) < \infty$ est nécessaire : sinon prendre $A_n = [n, \infty[$, où $\mu(A_n) = \infty$ et $\lim A_n = \emptyset$ qui donne $\lim(\mu(A_n)) \neq \mu(\lim A_n)$. ∎

2.4.3 Probabilité = mesure de probabilité

Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré.

Définition 2.40 La mesure μ est bornée ssi $\mu(\Omega) < \infty$ (masse totale finie).

Définition 2.41 La mesure μ est une "mesure de probabilité" (ou plus simplement une "probabilité") ssi c'est une mesure bornée de masse totale $1 = 100\%$:

$$\mu \overset{\text{notée}}{=} P : \left\{ \begin{array}{ll} \mathcal{T} & \to [0,1] \\ A & \mapsto P(A) \end{array} \right\} \quad \text{mesure telle que} \quad P(\Omega) = 1. \tag{2.10}$$

Ainsi l'événement Ω a pour probabilité $1 = \frac{100}{100} = 100\%$, i.e. 100% de "chance d'être réalisé" (un résultat est nécessairement dans Ω par définition de Ω : Ω est l'événement certain).

Définition 2.42 $(\Omega, \mathcal{T}, P)$ est alors un espace probabilisé.

"Toute expérience aléatoire se décrit mathématiquement par la donnée d'un espace probabilisé".

Définition 2.43 Un événement $A \in \mathcal{T}$ tel que $P(A) = 1$ est dit presque sûr.

(Il est sûr "à un ensemble négligeable près relativement à la mesure P", voir cours d'intégration : l'ensemble $B = \Omega - A$ vérifie $P(B) = \emptyset$, i.e. B est invisible pour la mesure B, et la mesure P ne voit pas de différence entre A et Ω.)

Notation abusive. Quand $A = \{a\}$ est un singleton, on note :

$$P(\{a\}) \overset{\text{noté}}{=} P(a). \tag{2.11}$$

2.5 Equiprobabilité

Définition 2.44 Dans le cas $\Omega = \{\omega_i : i \in I\}$ ensemble fini, quand $P(\{\omega_i\}) = \frac{1}{|\Omega|}$ est indépendant de i (même probabilité pour tous les singletons), on dit que P est une probabilité équiprobable ou probabilité uniforme (sur Ω) : elle est donc définie par, pour tout $A \subset \Omega$:

$$P(A) = \frac{|A|}{|\Omega|}, \tag{2.12}$$

pourcentage du nombre d'éléments de A dans Ω.

Exemple 2.45 Lancer d'une pièce équilibrée : $\Omega = \{p, f\}$, $|I| = 2$, et $P(p) = P(f) = \frac{1}{2}$. (Si la pièce n'est pas équilibrée, la probabilité associée n'est pas équiprobable.) ∎

Exercice 2.46 On lance une pièce équilibrée deux fois de suite. Donner la probabilité d'obtenir "deux faces", ou "deux piles", ou "une face et une pile".

Réponse. Ici $\Omega = \{(p,p), (p,f), (f,p), (f,f)\}$ est l'ensemble des résultats possibles équiprobables. On a $|\Omega| = 4$, donc $P(f,f) = P(f,p) = P(p,f) = P(p,p) = \frac{1}{4}$. La probabilité d'obtenir deux faces ou deux piles est donc $\frac{1}{4}$ et la probabilité d'obtenir une face et une pile est $P(\{p,f\}) = P(p,f) + P(f,p) = \frac{1}{2}$. ∎

Exemple 2.47 On s'intéresse au résultat du lancer d'un dé équilibré : l'ensemble des résultats est $\Omega = \{1,2,3,4,5,6\} =^{\text{noté}} [1,6]_{\mathbb{N}}$, avec, pour tout $i \in \Omega$, $P(i) = \frac{1}{6} = p_i$.

La probabilité d'obtenir l'événement $\{1,3\}$ est $P(\{1,3\}) = P(\{1\} \cup \{3\}) = P(1) + P(3) = \frac{2}{6}$ (probabilité d'obtenir 1 ou 3).

Exemple 2.48 Contre-exemple avec un dé équilibré. On lance un dé deux fois et on s'intéresse à la somme des résultats des lancers. On note $X : [1,6]_{\mathbb{N}}^2 \to \mathbb{R}$ la fonction somme $X(a,b) = a+b$ quand $a,b \in [1,6]_{\mathbb{N}}$. On note $\Omega_X = [2,12]_{\mathbb{N}}$ l'ensemble des sommes possibles (l'univers des résultats). Pour $n \in \Omega_X$, on note $P_X(n)$ la probabilité que la somme soit n. On suppose les dés équilibrés, i.e. on suppose que $P_X(2) = \frac{1}{36} = P_X(12)$ (1 façon d'avoir le résultat), $P_X(3) = \frac{2}{36} = P_X(11)$ (2 façons d'avoir le résultat), $P_X(4) = \frac{3}{36} = P_X(10)$ (3 façons d'avoir le résultat), $P_X(5) = \frac{4}{36} = P_X(9)$ (4 façons d'avoir le résultat), $P_X(6) = \frac{5}{36} = P_X(8)$ (5 façons d'avoir le résultat), $P_X(7) = \frac{6}{36}$ (6 façons d'avoir le résultat). La probabilité d'obtenir l'événement $\{6,7,8\}$ est $\frac{2*5+6}{36} = 8/18$. On vérifie que $P_X(\Omega_X) = \frac{2*(1+2+3+4+5)+6}{36} = 1$.

Ici la probabilité P_X n'est pas équiprobable : $P_X(2) \neq P_X(3)$.

Ici tout se passe comme si on avait un dé à 11 faces (numérotées de 2 à 12), dé qui n'est pas équilibré.

Exercice 2.49 Soit $n \geq 2$. Dans un groupe de n personnes toutes nées la même année (365 jours), quelle est la probabilité qu'au moins deux personnes soient nées le même jour (on suppose que les naissances sont réparties régulièrement dans l'année = équiprobabilité) ? Quel nombre minimal n de personnes donne une probabilité $\geq \frac{1}{2}$? Et $\geq 99\%$?

Réponse. On prend comme univers l'ensemble $\Omega = [1,365]_{\mathbb{N}}^n$ de toutes les dates d'anniversaire possibles, i.e. l'ensemble des n-uplets $\vec{\omega} = (a_1, ..., a_n)$, où $a_i \in [1,365]_{\mathbb{N}}$ est un jour possible de naissance de la personne i. Donc $|\Omega| = 365^n$.

Intéressons-nous à la probabilité "toutes les personnes sont nées un jour different". Le nombre de n-uplets $(a_1, ..., a_n)$ t.q. $a_i \neq a_j$ pour tout $i \neq j$ est $A_{365}^n = 365*364*...*(365-n+1)$ (nombre d'arrangements = nombre d'injections de $[1,n]_{\mathbb{N}}$ dans $[1,365]_{\mathbb{N}}$). Donc la probabilité pour que pour que toutes les personnes soient nées un jour différent est $q = \frac{A_{635}^n}{|\Omega|} = \frac{365}{365} * \frac{364}{365} * ... * \frac{(365-n+1)}{365}$. Et la probabilité pour qu'au moins deux personnes soient nées le même jour est $p = 1 - q$.

On trouve $p > 50\%$ pour $n \geq 23$, et $p > 99\%$ pour $n \geq 57$.

(Si $n = 2$ on a $q = \frac{364}{365}$, donc $p = \frac{1}{365}$.)

Exemple 2.50 On lance un dé (équilibré) 4 fois de suite. Quelle probabilité a-t-on d'obtenir au moins une fois un 6 ?

Réponse. 5/6 chance d'obtenir un autre numéro lors d'un lancer, donc $(5/6)^4$ chance d'obtenir un autre numéro lors de 4 lancers, donc $1 - (5/6)^4 \simeq 52/100$ de chance d'obtenir au moins un 6.

Exemple 2.51 On lance deux dés (équilibrés) n fois de suite. Quel est le n minimum pour avoir plus d'une chance sur deux d'avoir un double 6 ?

Réponse. 35/36 chance d'obtenir un autre double lors d'un lancer, donc $(35/36)^n$ chance d'obtenir un autre double lors de n lancers, donc $1 - (35/36)^n$ de chance au moins un double 6 après n lancers. Donc $n = 25$ pour une probabilité $\geq \frac{1}{2}$.

Exercice 2.52 Suite. 1- Probabilité qu'une personne soit née le 14 juillet ? Née un autre jour ?

2- Probabilité que deux personnes soient nées le 14 juillet ? Toutes les deux nées un autre jour ? Une seule sur deux le 14 juillet ?

3- Probabilité qu'au moins une personne parmi n soit née le 14 juillet ? Pour quelle valeur de n cette probabilité est $\geq 50\%$?

Réponse. 1- Probabilité qu'une personne soit née le 14 juillet : $\frac{1}{365}$. Non née e 14 juillet : $\frac{364}{365}$.

2- Les 2 nées le 14 juillet : $(\frac{1}{365})^2$. Les 2 non nées un autre jour : $(\frac{364}{365})^2$. Une seule née le 14 juillet : $\frac{1}{365}\frac{364}{365} + \frac{364}{365}\frac{1}{365}$. On vérifie que $(\frac{1}{365})^2 + (\frac{364}{365})^2 + 2\frac{1}{365}\frac{364}{365} = 1$.

3- Aucune personne parmi n née le 14 juillet : $(\frac{364}{365})^n$. Donc pour qu'au moins une personne soit née le 14 juillet : $1 - (\frac{364}{365})^n$. Et pour $n \geq 253$ cette probabilité est $\geq 50\%$.

Exercice 2.53 On voit 3 portes. Un objet est caché derrière une des portes de manière équiprobable. On vous demande de choisir une des portes. Puis, parmi les deux portes restantes, on retire une porte derrière laquelle l'objet ne se trouve pas. On vous redemande de choisir une porte (parmi les deux restantes). Avez-vous plus de chance de gagner l'objet si vous persistez dans votre choix de porte ou si vous changer de porte ?

Réponse. 1ère réponse : à l'issue du premier choix, il y avait 1 chance sur 3 de gagner l'objet, donc 2 chances sur 3 de ne pas le gagner. La seconde manipulation est a posteriori. Si on ne change pas de porte, on a toujours 1 chance sur 3, donc si on change de porte, on a 2 chances sur 3 de gagner l'objet (la porte choisie pèsera toujours $\frac{1}{3}$, alors qu'après avoir retiré une porte "vide", "la" porte restante pèsera $\frac{2}{3}$: elle correspond aux 2 portes non choisies).

2ème réponse : univers des possibilités initiales : $\Omega_1 = \{(1,0,0),(0,1,0),(0,0,1)\}$. On ne va pas se servir de cet univers Ω_1. On retire une porte vide : l'univers des nouvelles possibilités : $\Omega_2 = \{(1,-,0)(1,0,-),(-,1,0),(0,1,-),(-,0,1),(0,-,1)\}$, où $-$ désigne la porte manquante (l'univers a changé), et $|\Omega_2| = 6$. Un choix dans Ω_2 est équiprobable. Hypothèse : on garde la porte i ; dans Ω_2 il y a exactement 2 éventualités où 1 est en i, donc la probabilité d'avoir l'objet derrière i est $= \frac{2}{6} = \frac{1}{3}$. Et probabilité $\frac{4}{6} = \frac{2}{3}$ que l'objet soit derrière une autre porte.

3ème réponse : après avoir vu les probabilités conditionnelles. Portes numérotées $1,2,3$. On fait deux choix (un à chaque question). Univers $\Omega_3 = \{1,2,3\}^2$ l'ensemble des couples (i,j) pour $1 \le j \le 3$, de cardinal $|\Omega_3| = 9$. À la première étape, pour $i = 1,2,3$ (fixé) la probabilité des couples $(i,*)$, où $*$ est quelconque dans $\{1,2,3\}$, est de $\frac{1}{3}$ (équiprobabilité). À la deuxième étape on note $A = \{(1,2),(1,3),(2,1),(2,3),(3,1),(3,2)\}$ le sous-ensemble de Ω_3 (les couples) correspondant aux 2 portes restantes (après avoir enlevé une porte). Le cardinal de A est $|A| = 6$, nombres de couples possibles. Soit $B_1 = \{(1,1),(1,2),1,3)\}$ l'évènement "on garde la 1ère porte". Le cardinal de $A \cap B_1$ est $|A \cap B_1| = 2$. Donc $P(B_1|A) = \frac{P(A\cap B_1)}{P(A)} = \frac{\frac{2}{9}}{\frac{6}{9}} = \frac{1}{3}$. On vérifie que $P(B_1|A) + P(B_2|A) + P(B_3|A) = 1$ (formule de Bayes), soit $P(A\cap B_1) + P(A\cap B_2) + P(A\cap B_3) = P(A)$; en effet les B_i sont disjoints et leur union est Ω_3, et donc $P(A\cap B_1) + P(A\cap B_2) + P(A\cap B_3) = P((A\cap B_1) \cup (A\cap B_2) \cup (A\cap B_3)) = P(A\cap\Omega_3) = P(A)$. ▪

2.6 Probabilités discrètes et continues

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé.

2.6.1 Probabilités discrètes

Définition 2.54 Si l'univers Ω est fini ou dénombrable et si la tribu considérée est $\mathcal{T} = \mathcal{P}(\Omega)$, alors une probabilité P sur l'ensemble mesurable $(\Omega, \mathcal{T})$ est dite probabilité discrète. Et $(\Omega, \mathcal{T}, P)$ est un espace probabilisé discret.

Soit I un ensemble fini ou dénombrable. Soit l'ensemble $\Omega = \{\omega_i : i \in I\}$, où donc les ω_i sont deux à deux distincts.

Proposition 2.55 *Soit P et Q deux probabilités discrètes sur Ω.*
On a $P = Q$ ssi $P(\{\omega_i\}) = Q(\{\omega_i\})$ pour tout $i \in I$.

Preuve. Si $P = Q$ alors $P(\{\omega_i\}) = Q(\{\omega_i\})$ pour tout $i \in I$.

Réciproquement. Supposons $P(\{\omega_i\}) = Q(\{\omega_i\})$ pour tout $i \in I$, et il s'agit de montrer que $P(A) = Q(A)$ pour tout $A \in \mathcal{P}(\Omega)$. Soit $J \subset I$ et $A_J = \bigcup_{j\in J}\{\omega_j\}$, Donc A_J est une union disjointe, donc $P(A_J) = \sum_{j\in J} P(\{\omega_i\})$, et $Q(A_J) = \sum_{j\in J} Q(\{\omega_i\})$, donc $P(A_J) = Q(A_J)$ pour tout $A_J \in \mathcal{P}(\Omega)$. Donc $P = Q$. ▪

Soit P une probabilité discrète sur Ω. Pour $\omega_i \in \Omega$ on note :

$$p_i = P(\{\omega_i\}) \overset{\text{noté}}{=} P(\omega_i), \tag{2.13}$$

la probabilité d'obtenir ω_i.

Proposition 2.56 *On a :*

$$\sum_{i\in I} p_i = 1 \quad (= P(\Omega)) \qquad \text{et} \qquad P = \sum_{i\in I} p_i \delta_{\omega_i}, \tag{2.14}$$

où δ_{ω_i} est la masse de Dirac en ω_i. Et donc, pour $J \subset I$, et $A_J = \{\omega_i : i \in J\} \subset \Omega$ (événement), on a $P(A_J) = \sum_{i\in J} p_i$.

Preuve. P étant une probabilité, $1 = P(\Omega) = P(\bigcup_{i\in I}\{\omega_i\}) = \sum_{i\in I} P(\{\omega_i\}) = \sum_{i\in I} p_i$.

Soit $Q = \sum_{i\in I} p_i \delta_{\omega_i}$. Montrons que Q est une probabilité. On a $Q(\emptyset) = \sum_{i\in I} p_i \delta_{\omega_i}(\emptyset) = 0$. Soit $A_J = \{\omega_i : i \in J\} \in \mathcal{P}(\Omega)$ où $J \subset I$. On a $Q(A_J) = \sum_{i\in I} p_i \delta_{\omega_i}(A_J) = \sum_{i\in J} p_i$. En particulier

$Q(\Omega) = 1$. Et si $A, B \in \mathcal{P}(\Omega)$ vérifient $A \cap B = \emptyset$, avec $A = \bigcup_{j \in J}(\{\omega_j\})$ et $B = \bigcup_{k \in K}(\{\omega_k\})$, où donc $J \cap K = \emptyset$, alors $Q(A \cup B) = Q(\bigcup_{j \in J \cup K}(\omega_j)) = \sum_{j \in J \cup K} p_j = \sum_{j \in J} p_j + \sum_{j \cup K} p_j = Q(A) + Q(B)$. Et Q est σ-finie puisque $Q(\Omega) = \sum_{i \in I} Q(\{\omega_i\}) = \sum_{i \in I} p_i$ avec I au plus dénombrable. Donc Q est une probabilité.

Et $Q(\{\omega_i\} = p_i = P(\{\omega_i\})$ pour tout $i \in I$. Donc $Q = P$. ∎

Exemple 2.57 Lance d'un dé équilibré, $p_i = \frac{1}{6}$ pour tout i. On a $P(\{2, 5\}) = \sum_{i=1}^{6} p_i \delta_i(\{2, 5\}) = p_1 * 0 + p_2 * 1 + p_3 * 0 + p_4 * 0 + p_5 * 1 + p_6 * 0 = p_2 + p_5 = \frac{2}{6} = \frac{1}{3}$ est la probabilité d'obtenir 2 ou 5 (une chance sur trois d'obtenir 2 ou 5). ∎

2.6.2 Probabilités continues sur $\mathbb{R}$

Soit $\mathcal{T}_\mathbb{R}$ la tribu borélienne de $\mathbb{R}$ (la tribu engendrée par les intervalles de $\mathbb{R}$). Soit $\mu_\ell = d\omega$ la mesure de Lebesgue dans $\mathbb{R}$, voir cours d'intégration, i.e., si $a < b$:

$$\mu_\ell(]a, b[) = b - a \overset{\text{noté}}{=} \int_{\omega=a}^{b} d\omega.$$

Soit $\Omega \subset \mathbb{R}$ avec $\Omega \in \mathcal{T}_\mathbb{R}$ (mesurable), et soit $\mathcal{T}_\Omega$ la tribu $\mathcal{T}_\mathbb{R}$ restreinte à Ω :

$$\mathcal{T}_\Omega = \{A \cap \Omega : A \in \mathcal{T}_\mathbb{R}\}.$$

On vérifie immédiatement que $\mathcal{T}_\Omega$ est bien une tribu : 1- $\emptyset$ et Ω sont dans $\mathcal{T}_\Omega$; 2- si $B = A \cap \Omega$ avec $A \in \mathcal{T}_\mathbb{R}$ alors $A^C \in \mathcal{T}_\mathbb{R}$, donc $A^C \cap \Omega \in \mathcal{T}_\Omega$, i.e. le complémentaire $(A \cap \Omega)^C \cap \Omega$ de B dans Ω est bien dans $\mathcal{T}_\Omega$; 3- $\bigcap_I (A_i \bigcap \Omega) = (\bigcap_I A_i) \bigcap \Omega) \in \mathcal{T}_\Omega$ et $\bigcup_I (A_i \bigcap \Omega) = (\bigcup_I A_i) \bigcap \Omega) \in \mathcal{T}_\Omega$, pour I au plus dénombrable. Ainsi $(\Omega, \mathcal{T}_\Omega)$ est un espace mesuré.

Définition 2.58 Soit Ω un ouvert dans $\mathbb{R}$, soit $\mathcal{T}_\Omega$ la tribu borélienne restreinte à Ω, et soit P un probabilité sur $(\Omega, \mathcal{T}_\Omega)$. La probabilité P est dite continue sur Ω ssi il existe une fonction $p : \Omega \to \mathbb{R}_+$ positive intégrable sur Ω pour la mesure de Lebesgue, dite densité, telle que pour $]a, b[\in \Omega$ où $a < b$:

$$P(]a, b[) = \int_{\omega=a}^{b} p(\omega)\, d\omega, \qquad \text{où donc} \qquad \int_\Omega p(\omega)\, d\omega = 1. \tag{2.15}$$

On dit aussi que P est une mesure diffuse, de masse unité sur Ω.

On note :

$$dP(\omega) = p(\omega)\, d\omega, \qquad dP = p\, d\omega. \tag{2.16}$$

Ainsi :

$$P([a, b]) = \int_{\omega=a}^{b} p(\omega)\, d\omega \overset{\text{noté}}{=} \int_{\omega=a}^{b} dP(\omega) \overset{\text{noté}}{=} \int_a^b dP. \tag{2.17}$$

Exemple 2.59 Densité uniforme sur $\Omega = [a, b]$: $p(\omega) = \frac{1}{b-a}$. ∎

Exemple 2.60 Tir gaussien centré sur $\mathbb{R}$: $p(\omega) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\omega^2}{2\sigma^2}}$. ∎

2.6.3 Probabilités continues sur $\mathbb{R}^d$

Définition 2.61 Soit $\mathcal{T}_{\mathbb{R}^d}$ la tribu borélienne de $\mathbb{R}^d$, soit $\Omega \in \mathcal{T}_{\mathbb{R}^d}$, soit $\mathcal{T}_\Omega$ la tribu $\mathcal{T}_{\mathbb{R}^d}$ restreinte à Ω, et soit P une probabilité sur $\mathcal{T}_\Omega$.

On note $\vec{\omega} = (\omega_1, ..., \omega_d)$ un point dans Ω.

On note $d\omega = d\omega_1 ... d\omega_d = d\Omega$ la mesure de Lebesgue sur $\mathbb{R}^d$, où donc $d\omega(]a_1, b_1[\times ... \times]a_d, b_d[) = \prod_{i=1}^{d}(b_i - a_i) =$ volume du pavé $]a_1, b_1[\times ... \times]a_d, b_d[$, où $a_i < b_i$ pour tout i.

La probabilité P est dite continue sur Ω ssi P est définie par une densité p, i.e. ssi il existe $p : \Omega \to \mathbb{R}_+$ une fonction positive intégrable sur Ω pour la mesure de Lebesgue telle que, pour tout $A \in \mathcal{T}_\Omega$:

$$P(A) = \int_{\vec{\omega} \in A} p(\vec{\omega})\, d\Omega, \qquad \text{où donc} \qquad \int_{\vec{\omega} \in \Omega} p(\vec{\omega})\, d\Omega = 1. \tag{2.18}$$

On dit aussi que P est une mesure diffuse de masse unité.

Et on dit que P est donné par :

$$dP(\vec{\omega}) = p(\vec{\omega})\,d\omega = p(\vec{\omega})\,d\Omega, \qquad dP = p\,d\omega = p\,d\Omega. \tag{2.19}$$

Exemple 2.62 Cas $\mathbb{R}^2$. Densité uniforme sur $\Omega = [a,b] \times [c,d]$: $p(\omega) = \frac{1}{(b-a)(c-d)}$. ▪

Exemple 2.63 Tir gaussien centré dans $\mathbb{R}^2$: $p(\vec{\omega}) = \frac{1}{\sigma^2 2\pi} e^{-\frac{\omega_1^2+\omega_2^2}{2\sigma^2}}$. ▪

2.6.4 Approximation d'une probabilité continue par une probabilité discrète

Soit $P = p\,d\omega$ une probabilité continue de densité p sur $(\Omega, \mathcal{T})$, où Ω est un ouvert borné dans $\mathbb{R}^d$. On note $\overline{\Omega}$ la fermeture de Ω.

On "maille" (on fait un maillage de) $\overline{\Omega}$ en $\bigcup_{i=1}^n \overline{\Omega_i}$ où $n \in \mathbb{N}$ et les Ω_i sont des ouverts de $\mathbb{R}^d$ vérifiant $\bigcap_{i=1}^n \Omega_i = \emptyset$ et $\bigcup_{i=1}^n \overline{\Omega}_i = \overline{\Omega}$. (On utilise souvent abusivement l'expression : on "partitionne" (on fait une partition de) Ω, ce qui n'est pas la définition usuelle de partition en mathématiques).

Notons :

$$p_i = \int_{\Omega_i} p(\vec{\omega})\,d\omega = \text{la "masse" de } p \text{ sur } \Omega_i, \tag{2.20}$$

les p_i vérifiant donc $\sum_{i=1}^n p_i = 1$ (masse totale $1 = 100\% = \frac{100}{100}$).

Pour $i \in [1,n]_{\mathbb{N}}$, on choisit $\vec{\omega}_i \in \Omega_i$ un point fixé dans Ω_i, et on note :

$$\tilde{P} = \sum_{i=1}^n p_i \delta_{\vec{\omega}_i}. \tag{2.21}$$

Alors $\tilde{P}$ est une probabilité discrète sur $\bigcup_{i \in [1,n]_{\mathbb{N}}} \{\vec{\omega}_i\}$ (immédiat), probabilité qui "approche" P.

Interprétation. Soit $|\Omega_i| = \int_{\Omega_i} d\omega$ le volume de Ω_i. Soit :

$$q_i = \frac{p_i}{|\Omega_i|} \qquad (= \frac{\int_{\Omega_i} p(\omega)\,d\omega}{\int_{\Omega_i} d\omega}).$$

On approche p par la fonction en escalier :

$$q = \sum_i q_i 1_{\Omega_i},$$

fonction constante sur chaque Ω_i. (En éléments finis, on dit que q est une approximation P_0 de p.)

On a :

$$\int_\Omega q\,d\omega = \sum_i q_i \int_\Omega 1_{\Omega_i}\,d\omega = \sum_i q_i |\Omega_i| = \sum_i p_i = \sum_i \int_{\Omega_i} p(\omega)\,d\omega = \int_\Omega p(\omega)\,d\omega = 1,$$

et q est une densité de probabilité : notons :

$$dQ = q\,d\omega. \tag{2.22}$$

Q n'est pas une mesure de probabilité discrète puisque $Q(\{\vec{\omega}_i\}) = 0$ pour tout point $\vec{\omega}_i \in \Omega_i$: c'est une probabilité continue de densité $\widetilde{p}$ discontinue (en escalier). Une probabilité discrète a été obtenue juste avant, cf. (2.21).

Remarque 2.64 Donc attention au vocabulaire : ici Q est bien une probabilité continue : elle est donnée par la densité q qui est une fonction "en escalier". Donc probabilité continue (ici Q) n'est pas synonyme de densité continue (ici q est en escalier). ▪

2.6.5 Autres probabilités

Il y a d'autres types de probabilités : par exemple sur $\mathbb{R}$, soit p une densité positive telle que $\int_{\mathbb{R}} p(\omega)\,d\omega = \frac{1}{2}$, et soit la probabilité donnée par $P = p\,d\omega + \frac{1}{2}\delta_0$. Ici P est la somme d'une mesure diffuse et d'une mesure ponctuelle.

On appelle d'ailleurs probabilité mixte sur $\mathbb{R}$ une probabilité telle que $\exists J \subset \mathbb{N}$, $p_1, ..., p_J \in \mathbb{R}_+$ et $p : \mathbb{R} \to \mathbb{R}_+$ intégrable telle que $P(A) = \sum_{i \in J} p_i \delta_{\omega_i}(A) + \int_A p(x)\,dx$, où $\sum_{i \in J} p_i + \int_{\mathbb{R}} p(x)\,dx = 1$.

2.7 Fonction de répartition d'une probabilité

[Cumulative distribution function.]

2.7.1 Définition

Rappel. Si f est une fonction définie dans un voisinage d'un réel t à valeurs dans $\mathbb{R}$, on note :

$$f(t-) \stackrel{\text{déf}}{=} \lim_{\substack{s\to t\\ s<t}} f(s) \stackrel{\text{noté}}{=} \lim_{s\to t-} f(s) \tag{2.23}$$

la limite à gauche en t. De même $f(t+) \stackrel{\text{déf}}{=} \lim_{\substack{s\to t\\ s>t}} f(s) \stackrel{\text{noté}}{=} \lim_{s\to t+}$ est la limite à droite en t.

Définition 2.65 Cadre : espace probabilisé $(\mathbb{R}, \mathcal{T}_{\mathbb{R}}, P)$. On appelle fonction de répartition de la probabilité P la fonction F définie par, pour $t \in \mathbb{R}$:

$$F(t) \stackrel{\text{déf}}{=} P(]-\infty, t]). \tag{2.24}$$

Si on se place sur l'espace probabilisé $(\Omega, \mathcal{T}_\Omega, P)$ avec Ω ouvert de $\mathbb{R}$, on continue à utiliser (2.24) avec la convention P nulle à l'extérieur de Ω.

Proposition 2.66 *La fonction de répartition F d'une probabilité P est une fonction qui vérifie (faire un dessin) :*
(i) $F(-\infty) = 0$ et $F(+\infty) = 1$.
(ii) $F(t) = F(s) + P(]s,t])$ pour tout $s < t$.
(iii) $F : \mathbb{R} \to [0,1]$ est croissante, en particulier $F(t-) \le F(t)$ pour tout t.
(iv) $F(t-) = P(]-\infty, t[)$.
(v) $P(t) = F(t) - F(t-)$.
(vi) $F : \mathbb{R} \to [0,1]$ est continue à droite, i.e. $F(t) = F(t+)$ pour tout t.
(vii) La fonction de répartition détermine complètement P.

Preuve. (i) Comme $P(\emptyset) = 0$ et $P(\mathbb{R}) = 1$, avec (2.9) on déduit $F(-\infty) = 0$ et $F(\infty) = 1$.

(ii) On a $]-\infty,t] =]-\infty,s] \cup]s,t]$, avec $]-\infty,s] \cap]s,t] = \emptyset$, et P est une probabilité, d'où $P(]-\infty,t]) = P(]-\infty,s]) + P(]s,t])$.

(iii) Le (ii) implique F croissante car $P \ge 0$. Et F est croissante et bornée (par 1), donc la limite $F(t-)$ existe pour tout t, et F croissante donne $F(t-) \le F(t)$.

(iv) On applique (2.9) : avec $A_n =]-\infty, t-\frac{1}{n}[$ suite croissante, $F(t-) = \lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P(]-\infty,t[)$

(v) $]-\infty,t] =]-\infty,t[\cup \{t\}$, d'où $F(t) = F(t-) + P(t)$.

(vi) On a $F(u) = F(t) + P(]t,u])$ pour tout $u > t$, avec $\lim_{u\to t+}]t,u] = \emptyset$, et donc $F(t+) = F(t) + 0$, cf. (2.9).

(vii) $P(]s,t]) = F(t) - F(s)$ pour tout $s < t$, donc si on connaît F, alors on connaît P sur tout intervalle ouvert, donc on connaît P sur toute la tribu (borélienne). ∎

2.7.2 Fonction de répartition en escalier

Cas particulier P probabilité discrète : notant $\Omega = \{\omega_i : i \in I,\ \text{t.q. } \omega_i < \omega_j\ \forall i < j\}$ avec $I \subset \mathbb{N}^*$, I fini ou dénombrable, et avec $P(\{x\}) = 0$ si $x \notin \Omega$ et $P(\Omega) = 1$, alors la fonction de répartition est une fonction en escalier, donnée par

$$F(\omega_j) = \sum_{i \le j} P(\{\omega_i\}). \tag{2.25}$$

Immédiat.

Exercice 2.67 Montrer que si $F : \mathbb{R} \to [0,1]$ est une fonction croissante qui est continue à droite (i.e. $F(t) = F(t+)$ pour tout t) avec $F(-\infty) = 0$ et $F(\infty) = 1$, alors il existe une unique probabilité P sur les boréliens de $\mathbb{R}$ telle que F soit la fonction de répartition de P.

Réponse. On a nécessairement $P(]-\infty,t]) = F(t)$. On vérifie que P ainsi défini est une probabilité. ∎

2.7.3 Fonction de répartition absolument continue

Soit P une probabilité de densité $p : \Omega \to \mathbb{R}_+$, où p est Lebesgue-intégrable positive de masse 1. Ainsi :

$$F(x) = \int_{-\infty}^{x} p(t)\, dt. \tag{2.26}$$

Donc F est continue.

Et même absolument continue, i.e. primitive d'une fonction intégrable, et p est appelée la dérivée généralisée de F_X (c'est la dérivée classique si F est dérivable). Voir cours intégrale de Lebesgue.

2.8 Introduction à l'échantillonnage

Le cadre de l'échantillonnage est : l'ensemble $\Omega =^{\text{noté}} S = \{a_1, ..., a_n\}$ est appelé une population (finie) de taille $n = |S|$; la tribu est $\mathcal{T} = \mathcal{P}(S)$ (l'ensemble des sous-populations) ; et P est une probabilité discrète. Les éléments a_i sont appelés les individus.

2.8.1 Introduction à l'échantillonnage 1

On considère l'expérience d'extraire un élément de S. Et on répète l'expérience r fois. L'issue de l'expérience est souvent l'un des types suivant :

Définition 2.68
1- Tirage avec remise, on tient compte de l'ordre.
À l'issue de chaque tirage on remet l'élément tiré dans la population : l'ensemble des résultats de l'expérience est :

$$\Omega_1(n,r) = S^r = S \times S \times ... \times S \quad (r\text{-fois}), \qquad |\Omega_1(n,r)| = n^r, \tag{2.27}$$

et un résultat de l'expérience est appelé un échantillon (ordonné) de taille r avec replacement. C'est aussi le nombre de "mots" de longueur r dans un alphabet de n lettres. Voir plus loin les lois binomiale paragraphe 6.3 et multinomial paragraphe 6.5.

2- Tirage sans remise, on tient compte de l'ordre.
À l'issue de chaque tirage on ne remet pas l'élément tiré dans la population. Ici nécessairement $r \le n$. L'ensemble des résultats de l'expérience est :

$$\Omega_2(n,r) = \{(a_{i_1}, ..., a_{i_r}) \in S^r,\ a_{i_j} \ne a_{i_k},\ \forall j \ne k\}, \qquad |\Omega_2(n,r)| = A_n^r, \tag{2.28}$$

et un résultat de l'expérience est appelé un échantillon (ordonné) de taille r sans replacement. Voir plus loin la loi hypergéométrique paragraphe 6.6.

3- Tirage sans remise, on ne tient pas compte de l'ordre.
À l'issue de chaque tirage on ne remet pas l'élément tiré dans la population. Ici nécessairement $r \le n$. Et on ne tient pas compte de l'ordre du tirage :

$$\Omega_3(n,r) = \{\{a_{i_1}, ..., a_{i_r}\} \subset S^r,\ a_{i_j} \ne a_{i_k},\ \forall j \ne k\}, \qquad |\Omega_3(n,r)| = C_n^r = \binom{n}{r}. \tag{2.29}$$

Autrement dit, on tire une sous-population de r éléments (on ne tient pas compte de l'ordre). Voir plus loin la loi hypergéométrique paragraphe 6.6.

Dans ce dernier cas, une sous-population est souvent appelé un échantillon...

Et un sondage est une étude sur une sous-population.

4- Tirage avec remise, on ne tient pas compte de l'ordre.
On remet l'élément tiré dans la population et on ne tient pas compte de l'ordre du tirage. On note $i_j \in [0,n]_{\mathbb{N}}$ le nombre de répétitions de l'individu a_j, avec par convention "$i_j = 0$ quand a_j n'est pas tiré". L'univers est (constitué des sous-ensembles de r éléments) :

$$\Omega_4(n,r) = \{\{a_1, ..., a_n\},\ i_1 + ... + i_n = r\}, \qquad |\Omega_4(n,r)| = C_{n+r-1}^r = \binom{n+r-1}{r}. \tag{2.30}$$

(Si $i_j = 0$ alors a_j n'est pas pas l'événement $\{a_1, ..., a_n\}$, et si $i_j > 0$ alors $\{a_1, ..., a_n\}$ contient i_j fois a_j).) Démonstration : voir exercices 2.72 et 2.73.

Et $\Omega_4(n,r)$ est l'espace (improprement appelé) des sous-populations de S de taille $r \in \mathbb{N}^*$ avec répétition.

4-bis- Statistique de Bose–Einstein. Autre interprétation de 4-.

Soit S un ensemble de n cases a_1,...,a_n (la population). On doit placer r boules (indistinguables) dans les n cases (mettre une boule dans une case revient à tirer une case au hasard). Une case peut être vide, et on peut mettre plusieurs boules dans une case (une même case peut être tirée plusieurs fois).

Ou encore on doit placer r particules (indistinguables) dans n états d'énergie (Bose–Einstein).

On retrouve $\Omega_4(n,r)$ et donc $|\Omega_4(n,r)| = C^r_{n+r-1}$.

Exemple 2.69 Deux boules p et f et deux tirages : correspond à $S = \{p,f\}$, $n = 2$, et $r = 2$.

Ordonné avec remise : $\Omega_1(2,2) = S^2 = \{(p,p),(p,f),(f,p),(f,f)\}$, et $|\Omega_1(2,2)| = 2^2 = 4$.

Ordonné sans remise : $\Omega_2(2,2) = \{(p,f),(f,p)\}$, et $|\Omega_2(2,2)| = A^2_2 = 2$.

Non ordonné sans remise : $\Omega_3(2,2) = \{\{p,f\}\}$, et $C^2_2 = 1$.

Non ordonné avec remise : $\Omega_4(2,2) = \{\{p\},\{p,f\},\{f\})\}$, et $C^2_{2+2-1} = C^2_3 = 3$. ▟

Exemple 2.70 Trois boules a_1, a_2, a_3 et deux tirages : correspond à $S = \{a_1, a_2, a_3\}$, $n = 3$, et $r = 2$.

Ordonné avec remise : $\Omega_1(3,2) = S^2 = \{(a_1,a_1),(a_1,a_2),...,(a_3,a_3)\}$ de cardinal $3^2 = 9$.

Ordonné sans remise : $\Omega_2(3,2) = \{(a_1,a_2),(a_1,a_3),(a_2,a_1),(a_2,a_3),(a_3,a_1),(a_3,a_2)\}$ de cardinal $A^2_3 = \frac{3!}{1!} = 6$.

Non ordonné sans remise : $\Omega_3(3,2) = \{\{a_1,a_2\},\{a_1,a_3\},\{a_2,a_3\}\}$ de cardinal $C^2_3 = \frac{3!}{2!1!} = 3$.

Non ordonné avec remise : $\Omega_4(3,2) = \{\{a_1\},\{a_1,a_2\},\{a_1,a_3\},\{a_2\},\{a_2,a_3\},\{a_3\}\}$ de cardinal $C^2_{3+2-1} = C^4_2 = \frac{4!}{2!2!} = 6$. ▟

Exercice 2.71 Un panier contient 5 boules noires et 3 boules rouges, les boules étant indistinguables au touché. On fait un premier tirage, on ne remet pas la boule tirée, et on fait un second tirage. Quelle est la probabilité : 1- la première boule tirée est noire et la seconde boule tirée est noire ? 2- la première boule tirée est noire et la seconde boule tirée est rouge ? 3- la première boule tirée est rouge et la seconde boule tirée est noire ? 4- la première boule tirée est rouge et la seconde boule tirée est rouge ? 5- les boules tirées sont rouge et noire ?

Réponse. Première réponse. Pour avoir un boule noire suivie d'une boule noire : boule noire au 1er tirage : 5 chances sur 8, et boule noire au 1er tirage : 4 chances sur 7, d'où $P(N,N) = \frac{5}{8}\frac{4}{7} = \frac{20}{56}$. De même $P(N,R) = \frac{5}{8}\frac{3}{7} = \frac{15}{56}$, $P(R,N) = \frac{3}{8}\frac{5}{7} = \frac{15}{56}$, $P(R,R) = \frac{3}{8}\frac{2}{7} = \frac{6}{56}$. On vérifie que $P(N,N) + P(N,R) + P(R,N) + P(R,R) = 1$.

D'où $P(\{N,R\}) = P((N,R) \cup (R,N)) = P(N,R) + P(R,N) = \frac{30}{56}$.

Deuxième réponse : population $S = \{N_1,...,N_5,R_1,...R_3\} = \{a_i : i = 1,...,8\}$. On est dans le cas 2- avec $\Omega_2(8,2) = \{(a_i,a_j),\ i \neq j \in [1,8]_{\mathbb{N}}\}$. Et donc $|\Omega_2(8,2)| = A^2_8 = 56$ (nombre d'issues possibles). D'où $P(N,N) = \sum_{i=1}^{5}\sum_{\substack{j=1\\ j\neq i}}^{5} P(N_i,N_j) = \sum_{i=1}^{5}\sum_{\substack{j=1\\ j\neq i}}^{5} \frac{1}{56} = 20\frac{1}{56}$. De même $P(N,R) = \sum_{i=1}^{5}\sum_{k=1}^{3} P(N_i,R_k) = 15\frac{1}{56}$... ▟

Exemple 2.72 (Statistique de Bose–Einstein.) Soit S un ensemble de n cases a_1,...,a_n. On doit placer r boules (indistinguables) dans les n cases. Une case peut être vide, et on peut mettre plusieurs boules dans une case. Ou encore on doit placer r particules (indistinguables) dans n états d'énergie (Bose–Einstein). Exemple : si $n = 1$ on a une seule case, et les r boules vont dans la seule case, avec donc $\Omega_4(1,r) = 1$ (pour $r \geq 0$).

Si on note B une boule, une configuration est de type $B|BB|||B|BB$... où les barres verticales sont les séparations des cases. Et on a $r+n-1$ objets "les r boules et les $n-1$ séparations". Pour les r boules on a donc C^r_{n+r-1} manières de les distribuer dans les cases. Pour les $n-1$ séparations on a donc $C^{n-1}_{n+r-1} = C^r_{n+r-1}$ manières de séparer les boules C^r_{n+r-1}. ▟

Exercice 2.73 Soit $\Omega_4(n,r)$ donné en $(2.30)_1$. Montrer par récurrence que $|\Omega_4(n,r)| = C^{n-1}_{r+n-1} = \frac{(r+n-1)!}{r!(n-1)!}$ pour $n \geq 1$ et $r \geq 1$.

Réponse. On prend l'interprétation 4-bis-. Pour $n = 1$ (une seule case), alors toutes les boules sont nécessairement dans cette case : $\Omega_4(1,r) = \{i_1 \in \mathbb{N} : i_1 = r\}$ est le singleton $\{r\}$ (pour $r \geq 1$). Donc $|\Omega_4(1,r)| = 1 = C^0_r$. Donc le résultat est vrai pour $n = 1$.

Récurrence : hypothèse : soit $n \geq 1$ et supposons $|\Omega_4(n,r)| = C^{n-1}_{r+n-1}$ pour tout $r \geq 1$. Soit $\omega = (i_1,...,i_n,i_{n+1}) \in \Omega_4(n+1,r)$.

Si $i_{n+1}=0$ alors le nombre de ω possible est $|\Omega_4(n,r)|=C_{r+n-1}^{n-1}$ par hypothèse de récurrence.
Si $i_{n+1}=1$ alors le nombre de ω possible est $|\Omega_4(n,r-1)|=C_{r+n}^{n-1}$ par hypothèse de récurrence.
...
Si $i_{n+1}=r-1$ alors le nombre de ω possible est $|\Omega_4(n,1)|=C_n^{n-1}$ par hypothèse de récurrence.
Si $i_{n+1}=r$ alors le nombre de ω possible est $1=C_{n-1}^{n-1}$ puisque toutes les boules sont dans la dernière case.

Donc $|\Omega_4(n+1,r)|=C_{n-1}^{n-1}+C_n^{n-1}+...+C_{r+n-1}^{n-1}=C_{r+n-1+1}^{n+1-1}=C_{r+n}^n$, cf. (1.27). ∎

2.8.2 * Introduction à l'échantillonnage 2

On part toujours d'une population $S=\{a_1,...,a_n\}$ de taille $n=|S|$, et on fait r tirages. On considère cette expérience en termes de fonctions (observation d'une suite de r tirages) :

$$\omega : \begin{cases} \{1,...,r\} & \to S \\ \quad i & \mapsto \omega(i) \overset{\text{noté}}{=} \omega_i \text{ est l'élément de } S \text{ obtenu au } i\text{-ème tirage.} \end{cases} \tag{2.31}$$

Soit $\mathcal{F}(\{1,...,r\};S)$ l'ensemble des fonctions de $[1,r]_{\mathbb{N}}$ à valeurs dans S, identifiable à l'ensemble des suites finies $\omega=(\omega_i)_{i=1,...,r}$ de taille r. Une expérience est donc un élément $\omega\in\mathcal{F}(\{1,...,r\};S)$.

1- Un échantillon de taille r avec replacement est l'ensemble $\Omega_1(n,r)=\mathcal{F}(\{1,...,r\};S)$,

2- Un échantillon de taille r sans replacement est un élément de l'ensemble $\Omega_2(n,r)$ des injections, sous-ensemble de Ω_1.

3- Une sous-population de taille r est un élément de l'ensemble $\Omega_3(n,r)$ des classes de fonctions : identification des injections f et g à l'aide de la relation d'équivalence sur l'ensemble d'arrivée donnée par les permutations : $f\simeq g \Leftrightarrow \exists\sigma$ une permutation t.q. $f(i)=g(\sigma(i))$ pour tout i. Et quand on parlera de $\omega\in\Omega_3$, on aura extrait un élément de la classe d'équivalence ω (axiome du choix), élément qu'on notera abusivement ω.

2.9 Espace produit et produit de probabilités

Définition 2.74 Soit $(\Omega_1,\mathcal{T}_1,P_1)$ et $(\Omega_2,\mathcal{T}_2,P_2)$ deux espaces probabilisés. On définit l'espace produit $(\Omega,\mathcal{T},P)$ par :

1- $\Omega=\Omega_1\times\Omega_2$,

2- $\mathcal{T}$ est la tribu sur Ω engendrée par les éléments de $\mathcal{T}_1\times\mathcal{T}_2$ (i.e. la plus petite tribu contenant tous les pavés $A_1\times A_2\in\mathcal{T}_1\times\mathcal{T}_2$),

3- P est la probabilité qui restreinte aux événements de type $B=A_1\times A_2$ ("rectangles") est donnée par :

$$P(A_1\times A_2)=P_1(A_1)P_2(A_2) \tag{2.32}$$

("aire du rectangle = produit des longueurs des côtés").

Proposition 2.75 *P ainsi défini est bien une mesure de probabilité.*

Preuve. $P(\Omega_1\times\Omega_2)=P_1(\Omega_1)P_2(\Omega_2)=1*1=1$. Puis voir cours d'intégration. ∎

On note :

$$P=P_1\otimes P_2. \tag{2.33}$$

Exemple 2.76 On lance une pièce deux fois. On note $Z=\{\text{face},\text{pile}\}$ l'ensemble des résultats possibles d'un tirage. On définit la probabilité p sur Z par $p(\text{pile})=p$ et $p(\text{face})=q$. On définit P la probabilité sur $\Omega=Z\times Z$ (produit cartésien) par $P(x,y)=p(x)p(y)$ quand $(x,y)\in Z^2$. Donc P donné par

×	p	q
p	pp	pq
q	qp	qq

On vérifie que $pp+pq+qp+qq=p(p+q)+q(p+q)=p+q=1$. ∎

Exemple 2.77 Tir gaussien sur une cible plane : on suppose que la densité de probabilité de tirer en x est $p_1(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ et en y par $p_2(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$. Ainsi la densité de probabilité de tirer en un point (x,y) de la cible est $p(x,y)=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$. Ainsi la probabilité de tirer dans un domaine D de $\mathbb{R}^2$ est $P(D)=\int_D\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}\,dxdy$. En particulier, la probabilité pour que le tir soit à moins d'une distante $r\le R$ du centre de la cible est (ici D est le disque de centre 0 et de rayon R) $P(r\le R)=\dfrac{1}{2\pi}\displaystyle\int_{\theta=0}^{2\pi}\int_{r=0}^{R}e^{-\frac{r^2}{2}}\,rdrd\theta=[-e^{-\frac{r^2}{2}}]_{r=0}^R=(1-e^{\frac{-R^2}{2}})$. ∎

Exercice 2.78 Exercice 2.71 avec les probabilités produits. Un panier contient cinq boules noires et trois boules rouges. On fait un premier tirage, on ne remet pas la boule tirée, et on fait un second tirage. Quelle est la probabilité pour que les deux boules obtenues soient noires ? Soient noires ? Soient rouge et noire ?

Réponse. Pour les boules noires : considérer qu'on fait deux tirages indépendants, avec deux paniers distincts, l'un Ω_1 qui contient 5 boules noires et 3 boules rouges, et l'autre Ω_2 qui contient 4 boules noires et 3 boules rouges. Ici $p_1(N) = \frac{5}{8}$ pour le premier panier et $p_2(N) = \frac{4}{7}$ pour le second panier. D'où $P(N,N) = p_1(N)p_2(N) = \frac{20}{56}$.

Pour les boules rouges, $p_1(R) = \frac{3}{8}$ pour le premier panier et $q_2(R) = \frac{2}{7}$ pour le second panier. D'où $P(R,R) = p_1(R)q_2(R) = \frac{6}{56}$.

Pour le couple (N,R) on obtient $\frac{5}{8}\frac{3}{7} = \frac{15}{56}$ et pour le couple (R,N) on obtient $\frac{3}{8}\frac{5}{7} = \frac{15}{56}$, et donc $P(\{R,N\}) = \frac{15}{56} + \frac{15}{56} = \frac{30}{56}$. ▪

2.10 Cas $\Omega \subset \mathbb{R}^d$ et lois marginales

On considère le cas $\Omega \subset \mathbb{R}^d$. Soit $\pi_i : \mathbb{R}^d \to \mathbb{R}$ la projection sur la i-ème composante :

$$\pi_i(\vec{x}) = x_i \qquad \text{quand} \qquad \vec{x} = (x_1, ..., x_d). \tag{2.34}$$

On a :

$$\pi_i^{-1}([a,b]) = \mathbb{R}^{i-1} \times [a,b] \times \mathbb{R}^{d-i},$$

la "bande" de $\mathbb{R}^n$ de i-ème côté limité par a et b.

Exemple 2.79 Cas $\Omega = \mathbb{R}^2$, $\pi_i : \mathbb{R}^2 \to \mathbb{R}$, et :

$$\pi_1^{-1}([a,b]) = [a,b] \times \mathbb{R}, \qquad \pi_2^{-1}([c,d]) = \mathbb{R} \times [c,d],$$

sont les bandes "verticales" et "horizontales". Faire un dessin. ▪

Exemple 2.80 Soit $\Omega = \{p,f\}^2 = \{(p,p),(p,f),(f,p),(f,f)\} \subset \mathbb{R}^2$, $\pi_i : \Omega \to \mathbb{R}$:

$$\pi_1^{-1}(p) = \{(p,p),(p,f)\}, \qquad \pi_1^{-1}(f) = \{(f,p),(f,f)\},$$

et $\pi_2^{-1}(p) = \{(p,p),(f,p)\}$, $\pi_2^{-1}(f) = \{(p,f),(f,f)\}$. ▪

On considère $\Omega = \mathbb{R}^d$ muni de sa tribu borélienne $\mathcal{T}_{\mathbb{R}^d}$. Soit P une probabilité sur $\mathcal{T}_{\mathbb{R}^d}$ qui fait de $(\Omega, \mathcal{T}_{\mathbb{R}^d}, P)$ un espace probabilisé.

Définition 2.81 La i-ème loi marginale de $P : \mathcal{T}_{\mathbb{R}^d} \to [0,1]$ est la fonction $P_{\pi_i} : \mathcal{T}_{\mathbb{R}} \to [0,1]$ donnée par :

$$P_{\pi_i} \stackrel{\text{déf}}{=} P \circ \pi_i^{-1}. \tag{2.35}$$

(C'est un cas particulier des lois de probabilités $P_X = P \circ X^{-1}$ d'une v.a.r. X, voir plus loin.)

Exemple 2.82 Tir gaussien sur une cible plane, cf. example 2.77 : on s'intéresse aux tirs dans la bande verticale $[a,b]$:

$$P_{\pi_1}([a,b]) = P([a,b] \times \mathbb{R}) = \int_{x=a}^{b} (\int_{y\in\mathbb{R}} p(x,y)\,dy)dx \qquad \text{ici } = \int_{x=a}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx,$$

et P_{π_1} est donc la loi normale centrée. Idem pour P_{π_2}. ▪

Proposition 2.83 *Les P_{π_i} sont des probabilités.*

Preuve. On a $P_{\pi_i}(\mathbb{R}) = P(\pi_i^{-1}(\mathbb{R})) = P(\mathbb{R}^d) = 1$, et $P_{\pi_i}(\bigcup_{i\in I} A_i) = P(\mathbb{R}^{i-1} \times (\bigcup_{i\in I} A_i) \times \mathbb{R}^{d-i})$, et si les $(A_i)_{i\in I}$ sont disjoints 2 à 2 dans $\mathbb{R}$, il en est de même des $(\mathbb{R}^{i-1} \times A_i \times \mathbb{R}^{d-i})_{i\in I}$ dans $\mathbb{R}^n$ (les bandes sont parallèles et distinctes). ▪

Exercice 2.84 Dans $\mathbb{R}^2$. Montrer que si $P = P_1 \otimes P_2$, cf. (2.33), alors $P_{\pi_i} = P_i$.

Réponse. $(P \circ \pi_1^{-1})(]a,b[) = P(]a,b[\times\mathbb{R}) = (P_1 \otimes P_2)(]a,b[\times\mathbb{R}) = P_1(]a,b[)P_2(\mathbb{R}) = P_1(]a,b[)$... ▪

Exercice 2.85 Dans $\mathbb{R}^2$. Donner un exemple de probabilité P t.q., notant $P_{\pi_i} =^{\text{déf}} P\circ\pi_i^{-1}$, on a $P \neq P_{\pi_1}\otimes P_{\pi_2}$.

Réponse. Suite de l'exemple 2.80, et $P:\Omega\to\mathbb{R}$ donné par :

P	p	f
p	0	$\frac12$
f	$\frac12$	0

(soit $P(p,p)=0$, $P(p,f)=\frac12$, $P(f,p)=\frac12$, $P(f,f)=0$). Donc $P_{\pi_1}(p)=(P\circ\pi_1^{-1})(p)=P(\{(p,p),(p,f)\})=0+\frac12=\frac12$, et $P_{\pi_1}(f)=(P\circ\pi_1^{-1})(f)=P(\{(f,p),(f,f)\})=\frac12+0=\frac12$, et de même $P_2(p)=P_2(f)=\frac12$. Et $P_{\pi_1}\otimes P_{\pi_2}$ est donnée par

$P_{\pi_1}\otimes P_{\pi_2}$	p	f
p	$\frac14$	$\frac14$
f	$\frac14$	$\frac14$

, donc $P\neq P_{\pi_1}\otimes P_{\pi_2}$. ▪

3 Probabilité conditionnelle $P_{|A}$

3.1 Exemple

On reprend ici l'exemple donné par Feller [3] [conditional probability].

Soit Ω un ensemble de N personnes. Soit H le sous-ensemble de Ω constitué des hommes, et soit Dal le sous-ensemble de Ω constitué des daltoniens [color-blind]. On a :

$$P(H)=\frac{|H|}{|\Omega|},\qquad P(Dal)=\frac{|Dal|}{|\Omega|},$$

pourcentages (probabilités) respectifs dans la population Ω. Et la probabilité, dans Ω, de choisir au hasard une personne qui est à la fois homme et daltonien est le pourcentage :

$$P(Dal\cap H)=\frac{|Dal\cap H|}{|\Omega|}.$$

Question : parmi les hommes, quel est le pourcentage de daltoniens ?

Réponse : le nombre d'hommes daltonien est $|Dal\cap H|$, donc le pourcentage (la probabilité) cherché est :

$$\frac{|Dal\cap H|}{|H|} \overset{\text{noté}}{=} P_{|H}(Dal) \overset{\text{noté}}{=} P(Dal|H). \tag{3.1}$$

Définition 3.1 $P_{|H}(Dal)$ est appelée probabilité d'être dans Dal sachant qu'on se restreint à H, ou encore probabilité conditionnelle de Dal sachant H (probabilité d'être dans Dal sous la condition d'avoir été préalablement choisi dans H).

Donc on a :

$$P_{|H}(Dal)=\frac{\frac{|Dal\cap H|}{|\Omega|}}{\frac{|H|}{|\Omega|}}\qquad (=\frac{P(Dal\cap H)}{P(H)}), \tag{3.2}$$

égalité qui nous donnera la définition générique.

Remarque 3.2 Interprétation : se restreindre à la population H c'est donc : poser $P_{|H}(H)=1=100\%$ (tout l'intérêt est porté sur l'ensemble H) et $P_{|H}(H^C)=0$ (aucun intérêt pour H^C) ; autrement dit, on a "redimensionné" le poids des ensembles : le nouveau poids de H est $1=100\%$, et le nouveau poids de H^C est 0, quand on se sert de l'instrument de mesure $P_{|H}$. ▪

Remarque 3.3 Le vocabulaire utilisé, probabilité conditionnelle, n'est pas très explicite. Feller écrit à ce propos : "... it is unfortunate that its great simplicity is somewhat obscured by a singularly clumsy terminology...". ▪

Exercice 3.4 Soit Ω = population française. On a $P(H)\simeq 48\%$ et $P_{|H}(Dal)\simeq 8\%$, pourcentage de daltoniens parmi les hommes. Et on a $P_{|H^C}(Dal)\simeq 0.5\%$, pourcentage de femmes qui sont daltoniennes.

1- Si on tire au hasard une personne dans la population française, qu'elle est la probabilité de tirer un homme daltonien ?

2- Donner la probabilité d'être daltonien en France.

3- Calculer $P_{|Dal}(H)$, la probabilité, parmi le sous-ensemble des daltoniens, de tirer un homme.

Réponse. 1- On cherche $P(H \cap Dal)$. On a $P(H \cap Dal) = P_{|H}(Dal)P(H) \simeq \frac{8}{100}\frac{48}{100} = \frac{3,84}{100} = 3,84\%$.

2- On cherche $P(Dal)$. On a $Dal = (Dal \cap H) \cup (Dal \cap H^C)$ union disjointe, donc $P(Dal) = P_{|H}(Dal)P(H) + P_{|H^C}(Dal)P(H^C)$. Avec $P(Dal \cap H) = P_{|H}(Dal)P(H)$ et $P(Dal \cap H^C) = P_{|H^C}(Dal)P(H^C)$ Donc $P(Dal) \simeq \frac{8}{100}\frac{48}{100} + \frac{0.5}{100}\frac{52}{100} \simeq 4.1\%$.

3- $P_{|Dal}(H) = \frac{P(H \cap Dal)}{P(Dal)} \simeq \frac{\frac{3,84}{100}}{\frac{4,1}{100}} \simeq 94\%$. ▪

Remarque 3.5 Autre approche : la probabilité d'être dans H est $P(H) \in \mathbb{R}$. Et la probabilité d'être à la fois dans H et dans Dal est $P(H \cap Dal) \in \mathbb{R}$. Et il existe un réel α tel que :

$$P(H \cap Dal) = \alpha P(H),$$

avec $\alpha \in [0,1]$ car $P(H \cap Dal) < P(H)$, puisque $(H \cap Dal) \subset H$. Ce réel α est noté $\alpha = P_{|H}(Dal)$. C'est le rapport $P_H(Dal) = \frac{P(H \cap Dal)}{P(H)}$. ▪

Remarque 3.6 La notation $P_{|H}$ n'est pas la restriction usuelle de P au sous-ensemble H. La restriction usuelle $P_{|H\,\text{usuelle}} : S \in \mathcal{P}(H) \to P_{|H\,\text{usuelle}}(S) \in \mathbb{R}$ a pour ensemble de définition $\mathcal{P}(H)$.

Alors que $P_H : A \in \mathcal{P}(\Omega) \to P_H(A) \in \mathbb{R}$ a pour ensemble de définition $\mathcal{P}(\Omega)$, cf. (3.2). ▪

3.2 Définition d'une probabilité conditionnelle $P_{|A}$

Soit $(\Omega, \mathcal{T}_\Omega, P)$ un espace probabilisé. S'inspirant de (3.2) :

Définition 3.7 Pour $A \in \mathcal{T}_\Omega$ tel que $P(A) \neq 0$, la probabilité conditionnelle sachant A est :

$$P_{|A} : \begin{cases} \mathcal{T}_\Omega & \to [0,1], \\ B & \mapsto P_{|A}(B) \overset{\text{déf}}{=} \dfrac{P(B \cap A)}{P(A)} \overset{\text{noté}}{=} P(B|A). \end{cases} \tag{3.3}$$

Et $P_{|A}(B)$ est appelé probabilité conditionnelle de B sachant A (probabilité pour que la propriété décrite par B se réalise sachant que le choix du tirage est restreint à A).

En particulier :

$$P_{|\Omega} = P, \qquad P_{|A}(A) = 1 = P_{|A}(\Omega), \qquad P_{|A}(A^C) = 0, \qquad P_{|A}(B) = P_{|A}(B \cap A).$$

On a attribué une probabilité de 100% à A, et une probabilité nulle à l'extérieur de A, et seule la partie $B \cap A$ de B nous intéresse (i.e. $B \cap A^C$ ne nous intéresse pas), faire un dessin.

Notations. $P_{|A} =^{\text{noté}} P(\cdot|A) =^{\text{noté}} P(\cdot/A)$. Donc, pour tout $B \in \mathcal{T}_\Omega$:

$$P_{|A}(B) = P(B|A) = P(B/A) \qquad (= \frac{P(B \cap A)}{P(A)}). \tag{3.4}$$

Exemple 3.8 Paragraphe 3.1. ▪

Exercice 3.9 Au cours d'une épidémie (donc un très grand nombre de malades), on constate que, sur quinze malades, deux personnes avaient été vaccinées. Transcrire cette donnée en termes de probabilité conditionnelle.

Réponse. Soit V l'ensemble des vaccinés et M l'ensemble des malades.

Proportion de vaccinés parmi les malades : $P_{|M}(V) = \frac{2}{15} = \frac{P(V \cap M)}{P(M)} = P(V|M) = P(V/M)$. ▪

Exercice 3.10 On lance 2 dés. La somme est 5. Calculer la probabilité qu'un des résultats soit 2.

Réponse. L'univers est $\Omega = [1,6]_{\mathbb{N}}^2$. On se restreint à $A = \{(1,4),(2,3),(3,2),(4,1)\}$ (la somme est 5). Donc $P(A) = \frac{|A|}{|\Omega|} = \frac{4}{36}$. Soit $B = \{(2,i),(j,2) : i,j \in [1,6]\}$ l'événement "un des résultat est 2". Donc $B \cap A = \{(2,3),(3,2)\}$, et donc $P(A \cap B) = \frac{2}{36}$. Donc $P_{|A}(B) = \frac{\frac{2}{36}}{\frac{4}{36}} = \frac{1}{2}$: on a une chance sur 2. ▪

Exercice 3.11 Jeu de 52 cartes. 1- Tirages avec remise : calculer la probabilité que la 2ème carte tirée soit un as rouge, sachant que la 1ère carte tirée est un as noir.

2- Tirage sans remise : même question.

Réponse. Soit A l'événement "la 1ère carte tirée est un as-noir" (et la 2ème carte quelconque). Soit B l'événement "la 2ème carte tirée est un as-rouge" (et la 1ère carte quelconque).

1- Avec remise. Les tirages sont indépendants, donc $P(B|A) = \frac{2}{52} = P(B)$ (il y a deux as rouges). Calcul exhaustif : l'univers est l'ensemble Ω des couples de cartes possibles. $|\Omega| = 52^2$. Donc $P(A) = \frac{2*52}{52^2} = \frac{2}{52}$ et $P(B) = \frac{52*2}{52^2} = \frac{2}{52} = P(A)$. Et $A \cap B$ est l'ensemble des couples (as-noir,as-rouge) possibles de cardinal 4. D'où $P(A \cap B) = \frac{4}{52^2}$. $P_{|A}(B) = \frac{\frac{4}{52^2}}{\frac{2}{52}} = \frac{2}{52} = P(B)$.

2- Sans remise. 1ère carte tirée noire : il reste 51 cartes contenant les 2 as-rouge. Donc $P(B|A) = \frac{2}{51}$. Calcul exhaustif : l'univers Ω a pour cardinal $|\Omega| = A_{52}^2 = \frac{52!}{50!} = 52 * 51$. Et $P(A) = \frac{2*51}{52*51} = \frac{2}{52}$. Et $P(A \cap B) = \frac{2*2}{52*51}$. Donc $P_{|A}(B) = \frac{\frac{2*2}{52*51}}{\frac{2}{52}} = \frac{2}{51}$.

Ici $P(B) = \frac{50*2+1*2}{52*51} = \frac{2}{52}$ (car $50 * 2$ couples (non-rouge,rouge) et deux couples (rouge-rouge)). Donc $P_{|A}(B) \neq P(B)$ (les événements A et B ne sont pas indépendants). ▪

Exercice 3.12 Urne avec 10 boules Blanches, 20 boules Noires et 30 boules Rouge. Calculer la probabilité pour que la 1ère boule soit blanche et la 2èmes soit noire.

Réponse. Soit A l'événement la 1ère boule est blanche et B l'événement la 2ère boule est noire. On veut $P(A \cap B)$. On a $P(A \cap B) = P(B|A)P(A)$ avec $P(A) = \frac{10}{60} = \frac{1}{6}$ et $P(B|A) = \frac{20}{59}$. Donc $P(A \cap B) = \frac{1}{6}\frac{20}{59}$. ▪

Exercice 3.13 QCM avec 10 questions, chaque question ayant 2 réponses possibles. On gagne si on a au plus 1 question fausse. On perd sinon (au moins 2 réponses fausses). En supposant qu'on réponde au hasard :

1- Calculer la probabilité de gagner.

2- Calculer la probabilité de perdre.

3- Calculer la probabilité de perdre à la 5ième question sachant qu'on a perdu.

Réponse. Il y a 2^{10} réponses possibles.

1- Soit G l'événement "gagné" : soit toutes les réponses sont bonnes (1 possibilité), soit une des réponses est fausse (10 possibilités). Donc $P(G) = \frac{11}{2^{10}} = \frac{11}{1024} \simeq 1\%$.

2- Donc $P(G^C) = 1 - \frac{11}{2^{10}} \simeq 99\%$.

3- Soit A l'événement : on a perdu à la 5ième question : donc 2 réponses fausses sur 5, donc $C_5^2 = 10$ cas possibles pour les 5 premières questions et 2^5 réponses possibles pour les 5 dernières. Donc $P(A) = \frac{10*2^5}{2^{10}} = \frac{10}{2^5} \simeq 16\%$. On cherche $P(A|G^C)$: comme $A \cap G^C = A$, on a $P(A|G^C) = \frac{P(A \cap G^C)}{P(G^C)} = \frac{P(A)}{P(G^C)} \simeq 16\%$. ▪

3.3 Premières propriétés

Proposition 3.14 *Soit $A \in \mathcal{T}_\Omega$ tel que $P(A) \neq 0$.*

1- $P_{|A} = P(\cdot|A)$ est bien une probabilité sur Ω, i.e. $P_{|A} : \mathcal{T}_\Omega \to \mathbb{R}$ vérifie la définition 2.41.

2- Si $A \cap B = \emptyset$, alors $P_{|A}(B) = P(B|A) = 0$: si les événements sont incompatibles (i.e. disjoints), la probabilité conditionnelle de B sachant A est nulle. En particulier $P_{|A}(A^C) = 0$.

3- On a toujours, cf. (3.3) :

$$P_{|A}(B) \geq P(A \cap B). \tag{3.5}$$

4- Pour tout $B \in \mathcal{T}_\Omega$, on a :

$$P_{|A}(B^C) = 1 - P_{|A}(B), \tag{3.6}$$

autrement dit, $P_{|A}(B^C) + P_{|A}(B) = 1$ (car $P_{|A}$ est une probabilité).

5- Si $B \subset A$, on a $P_{|A}(B) = \frac{P(B)}{P(A)}$; et si $B \supset A$, on a $P_{|A}(B) = 1$.

Preuve. Immédiat. ▪

Exercice 3.15 Montrer que si $B_1 \cap B_2 = \emptyset$ alors $P_{|A}(B_1 \cup B_2) = P_{|A}(B_1) + P_{|A}(B_2)$, et que dans le cas général $P_{|A}(B_1 \cup B_2) = P_{|A}(B_1) + P_{|A}(B_2) - P_{|A}(B_1 \cap B_2)$.

Réponse. Soit le démontrer directement à l'aide de (3.3), soit appliquer la proposition précédente : $P_{|A}$ est une probabilité. ▪

Proposition 3.16 *Soit $A \in \mathcal{T}_\Omega$ tel que $P(A) \neq 0$, et soit $\mathcal{T}_A$ la tribu de Ω restreinte à A, i.e. $\mathcal{T}_A = \{B \cap A : B \in \mathcal{T}_\Omega\}$.*

$P_{|A} = P(\cdot|A)$ restreinte à $\mathcal{T}_A$ est une probabilité sur $(A, \mathcal{T}_A)$.

$P_{|A}$ restreinte à $\mathcal{T}_A$ est proportionnelle à P : si $C \in \mathcal{T}_A$ alors $P_{|A}(C) = \frac{1}{P(A)}P(C)$.

Preuve. Immédiat avec (3.3). ▪

Exercice 3.17 On lance un dé deux fois de suite. Donner la probabilité que la somme donne l'une des deux valeurs 4 ou 8, sachant qu'au premier lancer on a obtenu 1.

Réponse. Univers $\Omega = [1,6]_{\mathbb{N}}^2$ (les résultats des lancers).

Notre condition : être dans l'événement $A = \{(1,1),...,(1,6)\}$ (au premier lancer on a obtenu 1). On a $P(A) = \frac{6}{36} = \frac{1}{6}$.

Et l'événement B "somme $=$ 4 ou 8" est $B = \{(1,3),(2,2),(3,1),(2,6),(3,5),...,(6,2)\}$. On a $A \cap B = \{(1,3)\}$ et donc $P(A \cap B) = \frac{1}{36}$.

D'où $P_{|A}(B) = \frac{\frac{1}{36}}{\frac{6}{36}} = \frac{1}{6}$. ▪

Exercice 3.18 Loto simplifié : 3 boules différentes $1,2,3$. On procède à un premier tirage, puis sans remettre la boule tirée, on précède à un second tirage. Calculer la probabilité d'obtenir au second tirage la boule 2 sachant qu'au premier tirage on a obtenu la boule 1.

Réponse. 1- Réponse rapide (sans utiliser les probabilités conditionnelles) : on veut calculer $P(2|1)$. Au second tirage, il y a une chance sur 2 de tirer la boule 2 sachant que la boule 1 a déjà été tirée : $P(2|1) = \frac{1}{2}$.

2- Réponse détaillée (probabilités conditionnelles). L'univers Ω est :

$$\Omega = \{(1,2),(1,3),(2,1),(2,3),(3,1),(3,2)\}, \qquad |\Omega| = A_3^2 = 6.$$

L'événement "la première boule tirée est 1" est l'ensemble $A = \{(1,2),(1,3)\}$ de probabilité $P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6} = \frac{1}{3}$. L'événement "la seconde boule tirée est 2" est l'ensemble $B = \{(1,2),(3,2)\}$. D'où $A \cap B = \{(1,2)\}$ et $P(A \cap B) = \frac{|A\cap B|}{|\Omega|} = \frac{1}{6}$. D'où $P(B|A) = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2}$ est la probabilité conditionnelle cherchée.

3- Mise en garde au sujet de la réponse rapide 1- : cette réponse ne doit pas laisser supposer que l'univers est $S = \{1,2,3\}$ l'ensemble des 3 boules. Si on s'était placé dans S, pour $P_{|A}$ l'événement A serait $A = \{1\}$. Et l'événement B est $B = \{2\}$. Donc $A \cap B = \emptyset$, donc $P_{|A}(B) = 0$: c'est visiblement absurde. On ne peut donc pas se contenter de travailler dans S : pour pouvoir considérer $A \cap B$ il est indispensable de se placer dans $\Omega \subset S \times S$ qui est l'ensemble des résultats de l'expérience "on fait deux tirages". ▪

Exercice 3.19 (Cf. exercice 2.71 page 31.) Un panier contient cinq boules noires et trois boules rouges. On fait un premier tirage, on ne remet pas la boule tirée, et on fait un second tirage. Quelle est la probabilité d'avoir une boule noire au second tirage sachant qu'on a eu une boule noire au premier tirage ?

Réponse. 1- Réponse rapide. On note A l'événement "on a obtenu une boule noire au premier tirage". On note B l'événement "on a obtenu une boule noire au second tirage". Sachant qu'on a tiré une boule noire au premier tirage, il reste 4 boules noires sur les 7 : on a $P(B|A) = \frac{4}{7}$. C'est le résultat cherché.

2- Réponse détaillée. Soit $Z = \{N,R\}$ l'ensemble des résultats possibles d'un tirage. On a $\Omega = Z^2 = \{(N,N),(N,R),(R,N),(R,R)\}$ l'ensemble des résultats possibles après les deux tirages (puisqu'après le premier tirage on est sûr d'avoir encore les 2 couleurs).

On note A l'événement "on a obtenu une boule noire au premier tirage" : on a $A = \{(N,N),(N,R)\}$, donc avec $P(A) = \frac{5}{8}\frac{4}{7} + \frac{5}{8}\frac{3}{7} = \frac{35}{56}$, cf exercice 2.71.

On note B l'événement "on a obtenu une boule noire au second tirage" : on a $B = \{(N,N),(R,N)\}$.

On a donc $A \cap B = \{(N,N)\}$, et $P(A \cap B) = \frac{5}{8}\frac{4}{7} = \frac{20}{56}$, cf exercice 2.71. Et donc $P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{20}{35} = \frac{4}{7}$. C'est le résultat cherché.

3- Mise en garde au sujet de la réponse rapide 1- : cette réponse ne doit pas laisser supposer qu'on s'est placé dans $Z = \{N,R\}$ l'ensemble des résultats possibles d'un seul tirage : noire ou rouge. Si on c'était placé dans Z, l'événement A serait alors $A = \{N\}$. Et l'événement B étant également $B = \{N\}$, on aurait $B = A$, donc $P(A \cap B) = P(A)$ et donc $P(B|A) = \frac{P(A)}{P(A)} = 1$, i.e. on est certain d'obtenir une seconde boule noire : c'est visiblement absurde. On ne peut donc pas se contenter de travailler dans Z : pour pouvoir considérer $A \cap B$ il est indispensable de se placer dans $\Omega \subset Z \times Z$ qui est l'ensemble des résultats de l'expérience "on a fait deux tirages". ▪

3.4 Relation entre $P_{|B}(A)$ et $P_{|A}(B)$

Proposition 3.20 *Si $A, B \in \mathcal{T}$ sont des événements de probabilité non nulle, on a :*

$$P_{|A}(B)P(A) = P_{|B}(A)P(B) \qquad (= P(A \cap B)), \tag{3.7}$$

encore noté :

$$P(B|A)P(A) = P(A|B)P(B) \qquad (= P(A \cap B)). \tag{3.8}$$

Soit encore, première formule de Bayes, dans les cas où on ne divise pas par 0 :

$$\frac{P_{|B}(A)}{P_{|A}(B)} = \frac{P(A)}{P(B)}, \qquad \text{ou encore} \qquad P_{|B}(A) = P_{|A}(B)\frac{P(A)}{P(B)}, \tag{3.9}$$

encore noté :

$$\frac{P(A|B)}{P(B|A)} = \frac{P(A)}{P(B)}, \qquad \text{ou encore} \qquad P(A|B) = P(B|A)\frac{P(A)}{P(B)}. \tag{3.10}$$

Preuve. On a $P_{|B}(A)P(B) = P(A \cap B) = P(B \cap A) = P_{|A}(B)P(A)$. ▪

Exercice 3.21 Suite de l'exemple 3.9. Le tiers d'une population a été vacciné contre une maladie. Au cours d'une épidémie, on constate que, sur 15 malades, il y a 2 personnes vaccinées.

11- Comparer la proportion de malades parmi les vaccinés et la proportion globale de malade.

12- Comparer la proportion de gens malades et vaccinés et la proportion globale de malade.

13- Comparer la proportion de malade parmi les non-vaccinés et la proportion globale de malade.

On suppose de plus que sur 100 personnes vaccinées, 8 sont malades.

21- Quelle est la proportion de gens dans la population qui sont malades ?

22- Quelle est la proportion de gens dans la population qui sont à la fois vacciné et malade ?

23- Quelle est la proportion de gens malade parmi les non vaccinés ?

Réponse. Soit V la population des vaccinés et M celle de malades. Les hypothèses sont, $P(V) = \frac{1}{3}$ (un tiers de la population est vaccinée) et $P_{|M}(V) = \frac{2}{15} \simeq 13\%$ (proportion de vaccinés parmi les malades), cf. exercice 3.9.

11- Avec (3.10) on a $P_{|V}(M) = \frac{P_{|M}(V)}{P(V)}P(M) = \frac{\frac{2}{15}}{\frac{1}{3}}P(M) = \frac{2}{5}P(M) = 40\%P(M)$.

12- $P(V \cap M) = P_{|M}(V)P(M) = \frac{2}{15}P(M) \simeq 13\%P(M)$.

13- $P_{|V^C}(M) = \frac{P(M \cap V^C)}{P(V^C)} = \frac{P(M) - P(M \cap V)}{1 - P(V)} = \frac{3}{2}P(M)(1 - \frac{2}{15}) = \frac{13}{10}P(M) = 1.3P(M)$: il est plus probable de faire partie des malades quand on n'est pas vacciné.

On suppose de plus que sur 100 personnes vaccinées, 8 sont malades : $P_{|V}(M) = \frac{8}{100} = 8\%$ (proportion de malade parmi les vaccinés).

21- Avec (3.10) on a $P(M) = P(V)\frac{P_{|V}(M)}{P_{|M}(V)} = \frac{1}{3}\frac{2}{25}\frac{15}{2} = \frac{1}{5}$, i.e. 20% de la population est malade.

22- Puis $P(M \cap V) = P_{|V}(M)P(V) = \frac{2}{25}\frac{1}{3} = \frac{2}{75} \simeq 2,66\%$ (bien égal à $P_{|M}(V)P(M) = \frac{2}{15}\frac{1}{5}$) : moins de 3% de la population est à la fois malade et vacciné.

23- $P_{|V^C}(M) = \frac{13}{10}P(M) = \frac{13}{10}\frac{1}{5} = \frac{13}{50} = 26\%$, à comparer au $8\% = P_{|V}(M)$. ▪

Exercice 3.22 En Belgique, il y a 6 millions de belges qui parlent le néerlandais et il y a 4 millions de wallons. On sait que 7% des wallons parlent néerlandais. Parmi les habitants qui parlent le néerlandais, combien sont wallons ?

(Chiffres approximatifs de l'an 2000, donnés pour faire un exercice, et on ne tient pas compte des $\simeq$1.5% de germanophones et des $\simeq$1.5% d'autres communautés.)

Réponse. On note Ω l'ensemble des belges, Ω_W l'ensemble des wallons, Ω_N l'ensemble des belges qui parlent néerlandais. On a $7\% = P(\Omega_N|\Omega_W)$, proportion parmi les wallons de ceux qui parlent néerlandais. D'où, cf. (3.10), $P(\Omega_W|\Omega_N) = 7\%\frac{P(\Omega_W)}{P(\Omega_N)} = \frac{7}{100}\frac{\frac{4}{|\Omega|}}{\frac{6}{|\Omega|}} = \frac{14}{3}\% \simeq 4.7\%$. ▪

Exercice 3.23 Pris dans Jolion [4]. Deux machines M_1 et M_2 produisent respectivement 100 et 200 objets. M_1 produit 5% de pièces défectueuses et M_2 en produit 6%. Les objets sont mélangés. Quelle est la probabilité pour qu'un objet défectueux ait été fabriqué par la machine M_1 ? Et par la machine M_2 ?

Réponse. Il s'agit de calculer $P(M_1|A)$ et $P(M_2|A)$ où A est l'ensemble de pièces défectueuses (on tire un objet défectueux : quelle est son origine ?). Le nombre total de pièces produites est 300, d'où $P(M_1) = \frac{1}{3}$

et $P(M_2) = \frac{2}{3}$ sont les probabilités qu'une pièce ait été fabriquée par M_1 et M_2. Et la proportion de pièces défectueuses est :

$$P(A) = \frac{100\frac{5}{100} + 200\frac{6}{100}}{300} = \frac{17}{300}.$$

Et par hypothèse on a $P(A|M_1) = \frac{5}{100}$ et $P(A|M_2) = \frac{6}{100}$. On obtient :

$$P(M_1|A) = \frac{P(A|M_1)P(M_1)}{P(A)} = \frac{\frac{5}{100}\frac{1}{3}}{\frac{17}{300}} = \frac{5}{17} \simeq 0.29.$$

De même :

$$P(M_2|A) = \frac{P(A|M_2)P(M_2)}{P(A)} = \frac{\frac{2}{3}\frac{6}{100}}{\frac{17}{300}} = \frac{12}{17} \simeq 0.71.$$

On vérifie que la somme donne bien 1.

3.5 Règle de multiplication (formule de probabilités composées)

Corollaire 3.24 *Règle de multiplication. Si A, B, C sont des événements t.q. $P(A \cap B) \neq 0$, on a :*

$$P(A \cap B \cap C) = P(A)P_{|A}(B)P_{|A\cap B}(C), \tag{3.11}$$

soit :

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B). \tag{3.12}$$

Et plus généralement, quand $P(A_1 \cap ... \cap A_{n-1}) \neq 0$:

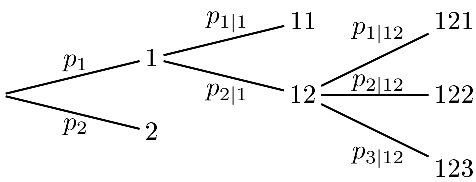
$$P(A_1 \cap ... \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)...P(A_n|A_1 \cap ... \cap A_{n-1}). \tag{3.13}$$

Preuve. Si $P(A\cap B) \neq 0$, alors $P(A) \neq 0$ (car $P(A) \geq P(A \cap B)$), et $P(B|A)$ et $P(C|A\cap B)$ ont un sens, et (3.12) a un sens.

On a $P((A \cap B) \cap C) = P(A \cap B)P(C|A \cap B)$ et $P(A \cap B) = P(A)P(B|A)$.

Puis récurrence laissée au lecteur.

Représentation sous la forme d'un arbre. Dans une population S on étudie le comportement en probabilité des individus.



Et par exemple $P(123) = P(1)P(2|1)P(3|12)$ est la probabilité d'aboutir à 123 connaissant les probabilités conditionnelles précédentes.

3.6 Partition et formule de Bayes (probabilités totales et probabilités des causes)

Si $A, B \in \mathcal{T}$, comme $A = (A \cap B) \cup (A \cap B^C)$ (partition), on a :

$$P(A) = P(A \cap B) + P(A \cap B^C), \tag{3.14}$$

et donc, si $P(B) \neq 0, 1$:

$$P(A) = P_{|B}(A)P(B) + P_{B^C}(A)P(B^C) \qquad (= P(A|B)P(B) + P(A|B^C)P(B^C)). \tag{3.15}$$

De manière plus générale :

Théorème 3.25 *Soit I un ensemble fini ou dénombrable. Si $\bigcup_{i\in I} B_i$ est une partition de Ω, alors pour tout événement A t.q. $P(A) \neq 0$ on a :*

$$P(A) = \sum_{i\in I} P(A \cap B_i). \tag{3.16}$$

D'où, supposant de plus $P(B_i) \neq 0$ pour tout i :

$$P(A) = \sum_{i\in I} P_{|B_i}(A)P(B_i) \qquad (= \sum_{i\in I} P(A|B_i)P(B_i) = \sum_{i\in I} P(A \cap B_i)), \tag{3.17}$$

appelé théorème (ou formule) des probabilités totales.

Et donc, pour tout $i \in I$:

$$P_{|A}(B_i) = \frac{P_{|B_i}(A)P(B_i)}{\sum_{j\in I} P_{|B_j}(A)P(B_j)} \qquad (= \frac{P_{|B_i}(A)P(B_i)}{P(A)}), \tag{3.18}$$

appelée formule de Bayes. Soit $P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j\in I} P(A|B_j)P(B_j)} \qquad (= \frac{P(A|B_i)P(B_i)}{P(A)})$.

Et pour un événement C :

$$\text{si} \quad \forall i \in I \quad P(B_i \cap C) \neq 0 \quad \text{alors} \quad P_{|C}(A) = \sum_{i\in I} P_{|B_i\cap C}(A)\ P_{|C}(B_i), \tag{3.19}$$

soit $P(A|C) = \sum_{i\in I} P(A|B_i \cap C)\ P(B_i|C)$.

Preuve. On a $A = \bigcup_{i\in I}(A \cap B_i)$ avec $(A \cap B_i) \cap (A \cap B_j) = \emptyset$ quand $i \neq j$, car $(B_i)_{i\in I}$ est une partition de Ω, d'où $P(A) = \sum_{i\in I} P(B_i \cap A) = \sum_{i\in I} P(A|B_i)P(B_i)$.

Puis $P(B_i|A) = P(A|B_i)\frac{P(B_i)}{P(A)}$.

Puis $\sum_I P(A|C\cap B_i)\ P(B_i|C) = \sum_I \frac{P(A\cap C\cap B_i)}{P(C\cap B_i)} \frac{P(C\cap B_i)}{P(C)} = \sum_I \frac{P(A\cap C\cap B_i)}{P(C)} = \frac{P(A\cap C)}{P(C)}$, car $(B_i)_I$ est une partition de Ω. ▪

Interprétation : probabilité des causes. Soit A un événement conséquence de n "causes" B_i disjointes deux à deux de probabilité $P(B_i)$. Sachant que A s'est réalisé, quelle est la probabilité que B_i en soit la cause ? La formule de Bayes donne le calcul à réaliser :

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)} \qquad (= \frac{P(A|B_i)P(B_i)}{P(A)}).$$

3.7 Exemples

Exemple 3.26 On fait 3 lancers d'une pièce (pile ou face). Ici $\Omega = \{p, f\}^3$, donc $|\Omega| = 2^3 = 8$. Soit A l'événement : on obtient plus de piles que de faces :

$A = \{(p, p, p), (p, p, f), (p, f, p), (f, p, p)\}$.

Soit B_1 l'événement l'événement "le premier lancer est pile" :

$B_1 = \{(p, p, p), (p, p, f), (p, f, p), (p, f, f)\}$,

et soit l'événement B_2 "le premier lancer est face" :

$B_2 = \{(f, p, p), (f, p, f), (f, f, p), (f, f, f)\}$.

On a $B_1 \cup B_2$ partition de Ω.

Soit u la probabilité conditionnelle d'être dans A sachant que le premier lancer donne pile, i.e. $u = P(A|B_1) = \frac{P(A\cap B_1)}{P(B_1)}$. Et soit v la probabilité conditionnelle d'être dans A sachant que le premier lancer donne face, i.e. $v = P(A|B_2) = \frac{P(A\cap B_2)}{P(B_2)}$.

Ici $A \cap B_1 = \{(p, p, f), (p, p, p), (p, f, p)\}$ et $A \cap B_2 = \{(f, p, p)\}$

Et le théorème nous dit que :

$$P(A) = uP(B_1) + vP(B_2),$$

ce qu'on vérifie : $\frac{P(A\cap B_1)}{P(B_1)}P(B_1) + \frac{P(A\cap B_2)}{P(B_2)}P(B_2) = P(A \cap B_1) + P(A \cap B_2) = P(A)$. ▪

Exemple 3.27 (Grâce à la formule de Bayes, connaître une probabilité à partir des probabilités conditionnelles). On dispose de 3 urnes U_1, U_2, U_3 qu'on choisit avec des probabilités p_1, p_2, p_3, les urnes contenant des boules rouges R en proportion $\alpha_1, \alpha_2, \alpha_3$.

Les hypothèses sont donc $P(U_i) = p_i$ et $P(R|U_i) = \alpha_i$.

Après avoir choisi une urne, on tire une boule b (avec équiprobabilité des tirages). But : connaître la probabilité "b est rouge", i.e. connaître $P(R)$ la probabilité de tirer une boule rouge.

La formule de Bayes (3.17) donne :

$$P(R) = \sum_{i=1}^3 \alpha_i p_i. \tag{3.20}$$

▪

Exercice 3.28 Donner les détails du calcul de l'exemple précédent : commencer par donner l'univers Ω et les événements en question.

Réponse. Calcul détaillé : Soit $E = \{U_1, U_2, U_3\}$ l'ensemble des urnes et soit $F = \{R, N\}$ l'ensemble des couleurs qui nous intéresse, à savoir "Rouge" et "Non rouge". L'univers est $\Omega = E \times F$, l'ensemble des couples "(urne,couleur)", muni d'une probabilité P. Traduisons les hypothèses permettant d'exprimer cette probabilité.

On note $\mathcal{U}_i = \{U_i\} \times F$ l'événement "l'urne U_i est choisie" (quelle soit la couleur), et $\mathcal{R} = E \times \{R\}$ l'événement "la couleur rouge est obtenue" (quelle que soit l'urne).

Les hypothèses réécrites sont $P(\mathcal{U}_i) = p_i$ et $P(\mathcal{R}|\mathcal{U}_i) = \alpha_i$ (ce sont les bonnes hypothèses pour pouvoir donner un sens à $P(\mathcal{R} \cap \mathcal{U}_i)$). Et la formule de Bayes donne :

$$P(\mathcal{R}) = \sum_{i=1}^{3} \alpha_i p_i,$$

réécriture non abusive de (3.20) : $\mathcal{R}$ au lieu de R.

Exemple 3.29 (Probabilité des causes.) Suite de l'exemple précédent. On se place dans le cas : la boule tirée est rouge.

Qu'elle est la probabilité d'avoir choisi l'urne U_i ?

Ici on veut connaître $P(U_i|R)$ (probabilité d'avoir choisi l'urne U_i sachant que la boule est rouge). On a :

$$P(U_i|R) = P(R|U_i)\frac{P(U_i)}{P(R)} = \frac{\alpha_i p_i}{\sum_j \alpha_j p_j}.$$

3.8 Exercices

Exercice 3.30 Soit un lot de 10000 pièces, dont 9000 sont bonnes et 1000 sont fausses. Ces pièces sont reparties dans trois boîtes : B_1 = 5000 pièces dont $n_1 = 200$ fausses, B_2 =3000 pièces dont $n_2 = 100$ sont fausses, et B_3 =2000 pièces dont $n_3 = 700$ sont fausses.

1- Avant de mettre les pièces dans une boîte, donner la probabilité de piocher une pièce fausse.

2- On a mis les pièces dans les boîtes. Donner la probabilité pour qu'en tirant une pièce, après avoir choisi une boîte, elle soit fausse.

Réponse. 1- Réponse rapide : 11- Soit Z l'ensemble des pièces, et F l'ensemble des pièces fausses, $F \subset Z$. Avant de mettre les pièces dans une boîte, la probabilité de piocher une pièce fausse est $\frac{|F|}{|Z|} = \frac{1000}{10000} = 10\%$.

12- Puis on met les pièces dans les boîtes. Par hypothèse, $P(F|B_1) = \frac{200}{5000}$, $P(F|B_2) = \frac{100}{3000}$, $P(F|B_3) = \frac{700}{2000}$, avec $P(B_i) = \frac{1}{3}$ (choix équiprobable), et donc (théorème de Bayes des probabilités totales) :

$$P(F) = \sum_{i=1}^{3} P(F|B_i)P(B_i) = \frac{1}{3}(\frac{2}{50} + \frac{1}{30} + \frac{7}{20}) \simeq 14.1\%.$$

On a plus de chance de tirer une pièce fausse si on les met dans 3 boîtes comme indiqué que si on les laisse dans une seule boîte.

2- Réponse détaillée : par définition, $P(F|B_i) = \frac{P(F \cap B_i)}{P(B_i)}$. Que veut dire $F \cap B_i$?

Pour donner un sens, il faut que les événements F et B_i appartiennent à un même univers.

On définit $\Omega_1 = \{b_1, b_2, b_3\}$ l'ensemble des boîtes.

On définit $\Omega_2 = \{f, v\}$, f pour faux et v pour vrai (les deux résultats possibles lors d'un tirage).

D'où l'univers $\Omega = \Omega_1 \times \Omega_2 = \{(b_1, f), (b_1, v), (b_2, f), (b_2, v), (b_3, f), (b_3, v)\}$.

Pour $i = 1, 2, 3$, on note $B_i = \{(b_i, f), (b_i, v)\}$ = l'événement la boîte i.

Soit $F = \{(b_1, f), (b_2, f), (b_3, f)\}$ l'événement des pièces fausses.

Donc $B_i \cap F$ a un sens, et $B_i \cap F = \{(b_i, f)\}$

21- Avant de mettre les pièces dans les boîtes, notons P_0 la probabilité d'avoir une pièce fausse : $P_0(F) = \frac{1000}{10000} = 10\%$.

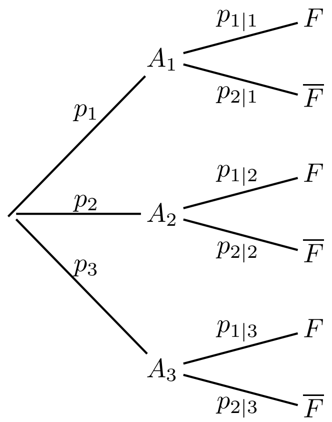
22- Après avoir mis les pièces dans les boîtes, notons P la probabilité d'avoir une pièce fausse. Les hypothèses sont : $P(F|B_1) = \frac{200}{5000}$, $P(F|B_2) = \frac{100}{3000}$, $P(F|B_3) = \frac{700}{2000}$. On suppose que le choix d'une boîte est équiprobable, soit $P(B_i) = \frac{1}{3}$. Donc $P(B_1 \cap F)) = P_{|B_1}(F)P(B_i) = \frac{200}{5000}\frac{1}{3}$. Idem pour B_2 et B_3. Donc $P(F) = \sum_{i=1}^{3} P(B_i \cap F)) = \sum_{i=1}^{3} P_{|B_i}(F)P(B_i) = \frac{1}{3}(\frac{2}{50} + \frac{1}{30} + \frac{7}{20}) \simeq 14.1\%$.

Exercice 3.31 On observe : on a une panne dans 1% des cas. On observe : une panne est détectée dans 98% des cas. On observe : s'il n'y a pas de panne, l'alerte est donnée dans 5% des cas (fausse alerte). Si l'alerte est donnée, qu'elle est la probabilité de panne ?

Réponse. Hypothèses. Soit A l'événement "panne" : $P(A) = 0.01$. Donc A^C est l'événement "pas de panne", $P(A^C) = 0.99$. Soit B l'événement alerte. On a $P_{|A}(B) = 0.98$ et $P_{|A^C}(B) = 0.05$.

On veut $P_{|B}(A)$. On a $P_{|B}(A) = \frac{P_{|A}(B)P(A)}{P_{|A}(B)P(A)+P_{|A^C}(B)P(A^C)} = \frac{0.98*0.01}{0.98*0.01+0.05*0.99} \simeq 17\%$. (C'est faible.)

Exemple 3.32 Soit 3 maladies exclusives A_1, A_2 et A_3, en proportion $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{4}$ et $p_3 = \frac{1}{4}$, donnant de la fièvre F avec les probabilité $p_{1|1} = P(F|A_1) = 0.7$, $p_{1|2} = P(F|A_2) = 0.8$, $p_{1|3} = P(F|A_1) = 0.9$, voir le tableau suivant :



où $\overline{F}$ désigne l'absence de fièvre.

Calculer les $p_{2|i}$ et calculer la probabilité de fièvre en cas d'une des ces trois maladies.

Réponse. On a $p_{2|i} = 1 - p_{1|i}$ car $F \cap \overline{F} = \emptyset$ et $p_{2|i} + p_{1|i} = \frac{P(F\cap A_i)}{P(A_i)} + \frac{P(\overline{F}\cap A_i)}{P(A_i)} = 1$.
Puis $P(F) = \sum_i P(F \cap A_i) = \sum_i P(F|A_i)P(A_i) = 0.7\frac{1}{2} + 0.8\frac{1}{4} + 0.9\frac{1}{4} = 0.775$.

3.9 Prévalence, sensibilité et spécificité d'un test

Vocabulaire médical. Soit une population Ω, et soit M la sous-population d'individus infectés par un virus, ou plus généralement ayant un problème pouvant être testé (M pour malade).

Définition 3.33 La prévalence est $P(M) = \frac{|M|}{|\Omega|}$ = le pourcentage d'individus infectés par le virus rapporté à la population totale.

Soit un test de détection du virus ; on note T_+ l'événement "le test est positif".

Définition 3.34 La sensibilité du test est la probabilité pour une personne infectée d'être testée positive :

$$\text{sensibilité} = \mathit{Se} = P_{|M}(T_+) = P(T_+|M) \qquad (= \frac{P(M\cap T_+)}{P(M)}). \tag{3.21}$$

La spécificité du test est la probabilité pour une personne non infectée d'être testée négative :

$$\text{spécificité } = \mathit{Sp} = P_{|M^C}(T_+^C) = P(T_+^C|M^C) \qquad (= \frac{P(M^C\cap T_+^C)}{P(M^C)}). \tag{3.22}$$

Interprétation.
Un test diagnostique parfait a une sensibilité et une spécificité de 1.
Un test avec une bonne sensibilité sera positif chez presque toutes les personnes infectées.
Un test avec une bonne spécificité sera négatif chez presque toutes les personnes non infectées.

Définition 3.35
La valeur prédictive positive (VPP) est la probabilité d'être malade quand on est testé positif :

$$VPP = P_{|T_+}(M) = P(M|T_+). \tag{3.23}$$

La valeur prédictive négative (VPN) est la probabilité d'être non malade quand on est négatif :

$$VPN = P_{|T_+^C}(M^C) = P(M^C|T_+^C). \tag{3.24}$$

Tableau des intersections : soit, relativement au test :

VP = probabilité vrai positif $P(T_+ \cap M)$ (infecté testé positif),
VN = probabilité vrai négatif $P(T_+^C \cap M^C)$ (non infecté testé négatif),
FP = probabilité faux positif $P(T_+ \cap M^C)$ (non infecté testé positif),
FN = probabilité faux négatif $P(T_+^C \cap M)$ (infecté testé négatif),

$\cap$	M	M^C
T_+	VP	FP
T_+^C	FN	VN

Alors on a par ligne, puisque $M \cap M^C = \emptyset$:

$$P(T_+) = VP + FP, \qquad P(T_+^C) = FN + VN, \tag{3.25}$$

et par colonnes, puisque $T_+ \cap T_+^C = \emptyset$:

$$P(M) = VP + FN, \qquad P(M^C) = FP + VN. \tag{3.26}$$

D'où :

$$Se = \frac{VP}{VP+FN}, \qquad Sp = \frac{VN}{VN+FP}, \tag{3.27}$$

et :

$$VPP = \frac{VP}{VP+FP}, \qquad VPN = \frac{VN}{VN+FN} \tag{3.28}$$

Exercice 3.36 Dans une population on a une proportion de $10^{-4} = \frac{1}{10000} = 0,01\%$ grippés. On dispose d'un test de détection caractérisé par :

1- sa sensibilité de 99%,

2- sa spécificité de $99,9\%$,

Sachant que le test est positif, qu'elle est la probabilité pour que l'individu soit effectivement grippé ? Le test est-il bon ? Quelle est la probabilité que le test soit positif ?

Réponse. But : trouver $P_{|T_+}(M) = P(M|T_+)$.

Hypothèses : On a $P(M) = 0,01\%$, donc $P(M^C) = 99,99\%$, on a $Se = P(T_+|M) = 99\%$, et on a $Sp = P(T_+^C|M^C) = 99,9\%$, donc $P_{|M^C}(T_+) = 0,1\%$. D'où :

$$P_{|M^C}(T_+) = \frac{P(M \cap T_+)}{P(T_+)} = \frac{P(T_+|M)P(M)}{P(T_+|M)P(M) + P(T_+|M^C)P(M^C)}$$
$$= \frac{0,99 \times 10^{-4}}{0,99 \times 10^{-4} + 0,001 \times 0,9999} \simeq 0,09 = 9\%.$$

Le test est "mauvais" : $P_{|T_+}(M) = P(M|T_+)$ est trop grand.

Et $P(T_+) = P_{|M}(T_+)P(M) + P_{|M^C}(T_+)P(M^C) = 0,99 \times 10^{-4} + 0,001 \times 0,9999 = 0,0011 = 0,11\%$ est la probabilité d'avoir un test positif. C'est "grand relativement à" $0,01\%$ = le nombre de grippés. ▪

Exercice 3.37 (Pris dans Suquet [15].)

1- Un test sanguin a une probabilité de 95% de détecter un virus donné (et donc 5% des cas ne sont pas détectés). Quelle est sa sensibilité ?

2- Le test donne un faux résultat positif pour 1% des personnes. Quelle est la spécificité ?

3- Si 0.5% de la population est porteuse du virus, quelle est la probabilité qu'une personne ait le virus sachant qu'elle a un test positif ? Le test est-il fiable ?

4- Si 50% de la population est porteuse du virus, quelle est la probabilité qu'une personne ait le virus sachant qu'elle a un test positif ? Le test est-il fiable ?

5- Pour un critère de fiabilité de 95%, quel pourcentage de la population doit être touché par le virus pour que le critère soit jugé fiable ?

Réponse. Soit $V = \{$personnes qui ont réellement le virus$\}$ et soit $T_+ = \{$personnes testées positives$\}$.

1- $95\% = P_{|V}(T_+) = Se$ = sensibilité : parmi les personnes qui ont le virus, 95% ont un test positif.

2- On donne $1\% = P_{|V^C}(T_+)$: parmi les personnes qui n'ont pas le virus, 1% a un test positif.

Donc $P_{|V^C}(T_+^C) = 1 - P_{|V^C}(T_+) = 99\% = Sp$ = spécificité.

3- But : trouver $P_{|T_+}(V) = P(V|T_+)$ (parmi les personnes dont le test est positif, quelles sont celles qui ont le virus). Soit $\alpha = P(V)$ et donc $P(V^C) = (1-\alpha)$. On applique la formule de Bayes :

$$P_{|T_+}(V) = \frac{P_{|V}(T_+)P(V)}{P(T_+)} = \frac{P_{|V}(T_+)P(V)}{P_{|V}(T_+)P(V) + P_{|V^C}(T_+)P(V^C)} = \frac{0.95 * \alpha}{0.95 * \alpha + 0.01 * (1-\alpha)},$$

$\alpha = 0.5\% = 0.005$ de la population est porteuse du virus : $P_{|T_+}(V) \simeq 32\%$. Conclusion : le test n'est pas fiable : si la personne est malade, elle n'a qu'une chance sur 3 d'être testée positive.

4- $\alpha = 50\% = 0.5$ de la population est porteuse du virus : $PP_{|T_+}(V) = \frac{0.95*.5}{0.95*.5+0.01*0.5} \simeq 99\%$, et le test est fiable si le critère de fiabilité demande plus 98% de réussite.

5- On cherche α t.q. $PP_{|T_+}(V) \geq 0.95$. Donc $\alpha \geq 0.95 * \alpha + 0.01 * (1-\alpha)$, Donc $\alpha \geq \frac{0.01}{0.06}$. Donc si au moins 17% de la population est porteuse du virus, le test est jugé fiable. ▪

3.10 Indépendance de 2 événements

Définition 3.38 Les événements A et B sont dits indépendants ssi :

$$P(A \cap B) = P(A)P(B). \tag{3.29}$$

(En particulier, si $P(A) \neq 0$ et $P(A) \neq 1$, alors A n'est pas indépendant de lui-même.)

Proposition 3.39 *Si A et B sont indépendants, alors B et A sont indépendants.*
Si $P(A) = 0$ alors A et B sont indépendants (idem si $P(B) = 0$).
Quand $P(A) \neq 0$, les événements A et B sont indépendants ssi :

$$P_{|A}(B) = P(B) \qquad (= P(B|A)), \tag{3.30}$$

i.e. la probabilité de B n'est pas conditionnée par A : la probabilité d'avoir la propriété B est la même dans A que dans tout l'univers.

Preuve. A et B jouent des rôles symétriques.
Si $P(A) = 0$, comme $A \cap B \subset A$ on a aussi $P(A \cap B) = 0$, donc (3.29) est vérifiée.
Si $P(A) \neq 0$ et $P(A \cap B) = P(A)P(B)$ (les événements sont indépendants) alors, avec (3.3), $P_{|A}(B) = \frac{P(A)P(B)}{P(A)} = P(B)$. Réciproquement, si $P_{|A}(B) = P(B)$, alors $\frac{P(A \cap B)}{P(A)} = P(B)$, donc les événements sont indépendants, cf. (3.29). ▪

Exemple 3.40 Daltonisme, cf. exercice 3.4 : *Dal* et H ne sont pas indépendants : $P_{|H}(Dal) = 8\%$ alors que $P(CB) = 4,1\%$. ▪

Proposition 3.41 *Si les événements A et B sont indépendants ssi A^C et B sont indépendants. Ainsi :*

$$\begin{aligned} P(A \cap B) = P(A)P(B) \quad &\iff \quad P(A^C \cap B) = P(A^C)P(B) \\ &\iff \quad P(A^C \cap B^C) = P(A^C)P(B^C) \end{aligned} \tag{3.31}$$

Et donc si $P(A^C) \neq 0$, on a aussi :

$$P_{|A}(B) = P(B) \quad \iff \quad P_{|A^C}(B) = P(B) \qquad (= P(B|A^C)). \tag{3.32}$$

Preuve. $\Rightarrow$: si A et B sont indépendants, alors $P(A^C \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(B)P(A^C)$. Puis on remplace B par B^C.
$\Leftarrow$: on applique $(A^C)^C = A$. ▪

Exercice 3.42 Montrer que les tirages de l'exemple 3.18 ne sont pas indépendants.

Réponse. 1- Réponse rapide. Il y a une chance sur trois de tirer une boule quelconque : probabilité $P(\{2\}) = \frac{1}{3}$. Or exercice 3.18 on a vu que $P(2|1) = \frac{1}{2}$ donc $\neq \frac{1}{3} = P(2)$: les tirages ne sont pas indépendants.

2- Réponse détaillée. On reprend les notations de l'exercice 3.18. $P(A) = \frac{1}{3} = P(B)$, et $P(A \cap B) = \frac{1}{6}$, d'où $P(A)P(B) = \frac{1}{9} \neq \frac{1}{6} = P(A \cap B)$: les tirages ne sont pas indépendants. ▪

Exercice 3.43 On lance un dé une fois. Trouver une probabilité pour laquelle les événements $A = \{1, 2\}$ et $B = \{2, 3\}$ sont indépendants, et une autre pour laquelle ils ne le sont pas.

Réponse. Dé équilibrée : probabilité P équiprobable, i.e. donnée par $P(i) = \frac{1}{6}$ pour un lancer. Donc $P(A) = \frac{1}{3} = P(B)$. Et $P(A \cap B) = P(\{2\}) = \frac{1}{6} \neq P(A)P(B)$: les événements A et B ne sont pas indépendants. Cette probabilité ne répond pas à la question.
Dé déséquilibré : probabilité P donnée par $P(1) = P(2) = \frac{1}{2}$, et $P(i) = 0$ pour $i = 3, 4, 5, 6$. On a $P(A) = \frac{1}{2} + \frac{1}{2} = 1$, $P(B) = \frac{1}{2} + 0 = \frac{1}{2}$, $P(A \cap B) = P(2) = \frac{1}{2} = P(A)P(B)$. ▪

Exercice 3.44 Jeu de 52 cartes. $A = \{\text{dame}\}$, $B = \{\text{coeur}\}$.
1- On suppose le jeu non truqué. A et B sont-ils indépendants ?
2- Jeu truqué : soit $V = \{\text{valet de pique}\}$. Soit la probabilité P donnée par $P(V) = \frac{1}{2}$ et $P(\{x\}) = \frac{1}{2}\frac{1}{51} = \frac{1}{102}$ la probabilité de toute autre carte. A et B sont-ils indépendants ?

Réponse. 1- $P(A) = \frac{4}{52}$, $P(B) = \frac{13}{52}$, $P(A \cap B) = \frac{1}{52}$; on a donc $P(A)P(B) = P(A \cap B)$: indépendance.
2- $P(A) = \frac{4}{102}$, $P(B) = \frac{13}{102}$, $P(A \cap B) = \frac{1}{102}$. on a donc $P(A)P(B) \neq P(A \cap B)$: non indépendance. ▪

Proposition 3.45
1- Si $A \subset B$ et $P(A) \neq 0$ et $P(B) \neq 1$, alors A et B sont dépendants.
2- Soit A, B tels que $A \cap B = \emptyset$ (événements incompatibles) et $P(A) \neq 0$ et $P(B) \neq 0$: ils ne sont jamais indépendants (on a $B \subset A^C$).
3- Un événement A est indépendant de lui-même ssi c'est l'événement impossible ou l'événement certain. (Donc attention au vocabulaire.)

Preuve. 1- $A \subset B$ donne $A \cap B = A$. Donc ici $P(A \cap B) = P(A)$; et $P(A) = P(A \cap B) = P(A)P(B)$ ssi $P(A) = 0$ ou $P(B) = 1$, cas exclus.
2- Cas $A \cap B = \emptyset$. On a $P(A \cap B) = 0 \neq P(A)P(B)$ quand $P(A) \neq 0$ et $P(B) \neq 0$. Ou encore : $A \subset B^C$, $P(A) \neq 0$ et $P(B^C) \neq 1$, et on applique le point précédent.
3- $P(A) = P(A \cap A) = P(A)^2$ ssi $P(A) = 0$ ou 1. ▪

Exercice 3.46 On lance un dé. Soit A l'événement le lancer est pair, et soit B l'événement le lancer est ≤ 3. Les événements A et B sont-ils indépendants ?

Réponse. $A = \{2,4,6\}$, $P(A) = \frac{1}{2}$. $B = \{1,2,3\}$, $P(B) = \frac{1}{2}$. $A \cap B = \{2\}$, $P(A \cap B) = \frac{1}{6}$. Et $\frac{1}{6} \neq \frac{1}{2}\frac{1}{2}$. Donc A et B ne sont pas indépendants.
Ou encore $P_{|A}(B)$ est la probabilité d'être ≤ 3 sachant qu'on est pair, soit $P_{|A}(B) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$ qui est $\neq P(B) = \frac{1}{2}$.
Ou encore $P_{|A^C}(B)$ est la probabilité d'être ≤ 3 sachant qu'on est impair, soit $P_{|A^C}(B) = \frac{P(B \cap A^C)}{P(A^C)} = \frac{\frac{2}{6}}{\frac{1}{2}} = \frac{2}{3}$ qui est $\neq P(B) = \frac{1}{2}$. ▪

3.11 Indépendance de n événements

Soit I un ensemble quelconque.

Définition 3.47 Une famille $(A_i)_{i \in I}$ d'événements est (mutuellement) indépendante ssi :

pour toute sous-famille finie $(A_{i_k})_{k=1,...,n}$, où $n \leq |I|$, on a :

$$P(A_{i_1} \cap ... \cap A_{i_n}) = P(A_{i_1})...P(A_{i_n}). \tag{3.33}$$

(À comparer à l'indépendance de v.a.r., cf. (5.16) plus loin.)

Exemple 3.48 Pour une famille (A, B, C) de trois événements, ils sont donc indépendants ssi les 4 conditions suivantes sont réalisées :

$$P(A \cap B \cap C) = P(A)P(B)P(C), \quad \underline{\text{et}} \tag{3.34}$$

$$P(A \cap B) = P(A)P(B), \quad P(B \cap C) = P(B)P(C), \quad P(A \cap C) = P(A)P(C). \tag{3.35}$$

Voir remarques suivantes 3.49 et 3.50. ▪

Remarque 3.49 N.B. : si la famille est indépendante, alors les A_i sont deux à deux indépendants (immédiat : prendre toutes les sous-familles de 2 éléments).
Mais la réciproque est fausse : si les éléments sont 2 à 2 indépendants, la famille n'est pas nécessairement indépendante.
Exemple : soit $\Omega = \{1,2,3,4\}$ et tirage équiprobable ; soit les événements $A = \{1,2\}$, $B = \{1,3\}$ et $C = \{1,4\}$, avec donc avec $P(A) = \frac{|A|}{|\Omega|} = \frac{1}{2} = P(B) = P(C)$: on a bien (3.35), i.e. $P(A \cap B) = P(B \cap C) = P(C \cap A) = P(\{1\}) = \frac{1}{4}$ (les événements sont 2 à 2 indépendants), mais on n'a pas (3.34) car $P(A \cap B \cap C) = P(\{1\}) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$. ▪

Remarque 3.50 N.B. : si (3.34) est vraie, on n'a pas nécessairement (3.35). Comme A, B, C jouent le même rôle, cela revient à dire qu'on peut avoir (3.34) et $P(A \cap B) \neq P(A)P(B)$, i.e. on peut avoir (3.34) alors que les événements A et B ne sont pas indépendants. Exemple :

Tirage équiprobable à 4 boules numérotées $1,2,3,4$, et 2 tirages successifs sans remise où on tient compte de l'ordre. L'univers est :

$$\Omega = \{(1,2),(1,3),(1,4),...,(4,1),(4,2),(4,3)\},$$

de cardinal $|\Omega| = 12$ ($= A_4^2 = \frac{4!}{2!}$ cf. (2.28)). Soit A l'événement "la première boule tirée est 1 ou 2" :

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

de cardinal $|A| = 6$. D'où $P(A) = \frac{|A|}{|\Omega|} = \frac{1}{2}$. (Dessiner les points dans $\mathbb{N} \times \mathbb{N}$.)

Soit B l'événement "la deuxième boule tirée est 3 ou 4" :

$$B = \{(1,3),(2,3),(4,3),(1,4),(2,4),(3,4)\},$$

de cardinal $|B| = 6$. D'où $P(B) = \frac{|B|}{|\Omega|} = \frac{1}{2}$. En particulier $P(A)P(B) = \frac{1}{4}$.

On a $A \cap B = \{(1,3),(2,3),(1,4),(2,4)\}$ d'où $P(A \cap B) = \frac{4}{12} = \frac{1}{3} \neq \frac{1}{4} = P(A)P(B)$: les événements A et B ne sont pas indépendants.

Et soit C l'événement "les boules tirées sont 1 et 2 ou bien 1 et 3 : $C = \{(1,2),(2,1),(1,3),(3,1)\}$. On a $P(C) = \frac{|C|}{|\Omega|} = \frac{1}{3}$ et $P(A)P(B)P(C) = \frac{1}{12}$ avec $A \cap B \cap C = \{(1,3)\}$ donc $P(A \cap B \cap C) = \frac{1}{12} = P(A)P(B)P(C)$. On a bien (3.34) sans avoir (3.35). ▪

Exemple 3.51 $(\omega_i)_{i=1,...,N}$ suite de N tirages de "pile ou face" : $\omega_i = p$ ou f. Ici l'univers est $\Omega = (\{p,f\})^N$. Une éventualité est un $\vec{\omega} = (\omega_1,...,\omega_N) = (\omega_i)_{i=1,...,N} \in \Omega$ (une suite finie de longueur N). Soit $(\vec{\omega}_k)_{k\in\mathbb{N}^*}$ une suite (dénombrable) d'éventualités. La famille $(\{\vec{\omega}_k\})_{k\in\mathbb{N}^*}$ constitue une famille indépendante d'événements. ▪

Proposition 3.52 *Si $(A_i)_{i\in I}$ est une famille d'événements (mutuellement) indépendants, alors la famille des complémentaires $(A_i^C)_{i\in I}$ est une famille d'événements (mutuellement) indépendants.*

Preuve. Récurrence. Hypothèse : si $(A_i)_{i=1,...,n}$ famille indépendante alors $(A_i^C)_{i=1,...,n}$ famille indépendante.

Vrai pour $n = 2$, cf proposition 3.41. Supposons que ce soit vrai pour un $n \in \mathbb{N}^*$.

Soit $(A_i)_{i=1,n+1}$ famille indépendante. Par hypothèse, toute sous famille $(A_{i_1},...,A_{i_k})$ de taille $k \le n$ de $(A_i)_{i=1,n+1}$ vérifie $(A_{i_1}^C,...,A_{i_k}^C)$ est indépendante.

Considérons la famille $(A_i)_{i=1,...,n+1}$ supposée indépendante. En particulier A_{n+1} et $(\bigcup_{i\le n} A_i)$ sont indépendants, donc A_{n+1}^C et $(\bigcup_{i\le n} A_i)^C$ sont indépendants, cf. proposition 3.41, soit $P(A_{n+1}^C \cap (\bigcup_{i\le n} A_i)^C = P(A_{n+1}^C)P((\bigcup_{i\le n} A_i)^C)) = P(A_{n+1}^C)P(\bigcap_{i\le n} A_i^C)) = P(A_{n+1}^C)P(A_1^C)...P(A_n^C)$ (hypothèse de récurrence). On a vérifié (3.33). ▪

3.12 Lemme du zéro-un de Borel–Cantelli

3.12.1 lim sup et lim inf

Définition 3.53 Si $(A_n)_{n\in\mathbb{N}^*}$ est une suite de sous-ensembles de Ω, on note :

$$\limsup_{n\to\infty}(A_n) = \bigcap_{n\in\mathbb{N}^*}(\bigcup_{k\ge n} A_k) \quad \text{et} \quad \liminf_{n\to\infty}(A_n) = \bigcup_{n\in\mathbb{N}^*}(\bigcap_{k\ge n} A_k), \tag{3.36}$$

les sous-ensembles de Ω appelés limite supérieure et limite inférieure de la suite $(A_n)_{\mathbb{N}^*}$.

Soit, pour $n \in \mathbb{N}^*$:

$$B_n = \bigcup_{k\ge n} A_k \quad \text{et} \quad C_n = \bigcap_{k\ge n} A_k. \tag{3.37}$$

Proposition 3.54 *Les suites $(B_n)_{\mathbb{N}^*}$ et $(C_n)_{\mathbb{N}^*}$ sont respectivement décroissantes et croissantes, et les limsup et liminf d'une suite $(A_n)_{n\in\mathbb{N}^*}$ existent toujours dans $\mathcal{P}(\Omega)$:*

$$\limsup_{n\to\infty}(A_n) = \lim_{n\to\infty}(\bigcup_{k\ge n} A_k) = \bigcap_{\mathbb{N}^*} B_n \quad \text{et} \quad \liminf_{n\to\infty}(A_n) = \lim_{n\to\infty}(\bigcap_{k\ge n} A_k) = \bigcup_{\mathbb{N}^*} C_n. \tag{3.38}$$

Preuve. $B_n = \bigcup_{k\geq n} A_k = A_k \bigcup(\bigcup_{k\geq n+1} A_k) = A_k \bigcup B_{n+1} \supset B_{n+1}$, donc (B_n) est décroissante, donc convergente dans $\mathcal{P}(\Omega)$ de limite $\bigcap_{\mathbb{N}^*} B_n = \lim_{n\to\infty} B_n$. Et $\limsup_{n\to\infty}(A_n) =^{\text{déf}} \bigcap_{\mathbb{N}^*} B_n$.

$C_n = \bigcap_{k\geq n} A_k = A_k \bigcap(\bigcap_{k\geq n+1} A_k) = A_k \bigcap C_{n+1} \subset C_{n+1}$, donc (C_n) est croissante, donc convergente dans $\mathcal{P}(\Omega)$ de limite $\bigcup_{\mathbb{N}^*} C_n = \lim_{n\to\infty} C_n$. Et $\liminf_{n\to\infty}(A_n) =^{\text{déf}} \bigcup_{\mathbb{N}^*} C_n$. ▟

Interprétation :

$$\begin{aligned}\limsup_{n\to\infty}(A_n) &= \bigcap_{n\in\mathbb{N}^*} B_n = \{\omega : \forall n \in \mathbb{N}^*,\ \omega \in B_n\}\\ &= \{\omega : \forall n \in \mathbb{N}^*,\ \exists k \geq n,\ \omega \in A_k\}\\ &= \{\omega : \omega \text{ appartient à une infinité de } A_k\}.\end{aligned} \tag{3.39}$$

Et :

$$\begin{aligned}\liminf_{n\to\infty}(A_n) &= \bigcup_{n\in\mathbb{N}^*} C_n = \{\omega : \exists n \in \mathbb{N}^*,\ \omega \in C_n\}\\ &= \{\omega : \exists n \in \mathbb{N}^*,\ \forall k \geq n,\ \omega \in A_k\}\\ &= \{\omega : \exists n \in \mathbb{N}^*,\ \omega \text{ appartient à tous } A_k \text{ pour } k \geq n\},\\ &= \{\omega : \text{il n'existe qu'un nombre fini de } A_k \text{ t.q. } \omega \notin A_k\}.\end{aligned} \tag{3.40}$$

Proposition 3.55 *On a :*

$$\limsup_{n\to\infty}(A_n) \supset \liminf_{n\to\infty}(A_n). \tag{3.41}$$

Preuve. Il s'agit de montrer que $\bigcap_{n\in\mathbb{N}^*}(B_n) \supset \bigcup_{m\in\mathbb{N}^*}(C_m)$, i.e. $\forall n \in \mathbb{N}^*$, $B_n \supset \bigcup_{m\in\mathbb{N}^*}(C_m)$, i.e. $\forall n \in \mathbb{N}^*$, $\forall m \in \mathbb{N}^*$, $B_n \supset C_m$, i.e. $\forall n \in \mathbb{N}^*$, $\forall m \in \mathbb{N}^*$, $\bigcup_{k\geq n} A_k \supset \bigcap_{\ell\geq m} A_\ell$.

Comme $\bigcap_{\ell\geq m} A_\ell \subset \bigcap_{\ell\geq 1} A_\ell$, il suffit de vérifier : $\forall n \in \mathbb{N}^*$, $\bigcup_{k\geq n} A_k \supset \bigcap_{\ell\in\mathbb{N}^*} A_\ell$, ce qui est trivial car, $\forall n \in \mathbb{N}^*$, $\bigcap_{\ell\in\mathbb{N}^*} A_\ell \subset A_n \subset \bigcup_{k\geq n} A_k$. ▟

Exemple 3.56 Soit $(a_n)_{\mathbb{N}^*}$ la suite de réels définie par $a_{2n} = 1 + \frac{1}{2n}$ et $a_{2n+1} = 0$.

Soit $A_n = \{a_n\}$ (singleton).

Alors $B_n = \{0, 1 + \frac{1}{2n}, 1 + \frac{1}{2(n+1)}, ...\}$, donc $\limsup(A_n) = \bigcap_{\mathbb{N}^*} B_n = \{0\}$.

Et $C_n = \emptyset$, donc $\liminf(A_n) = \emptyset$.

N.B. : ne pas confondre avec les limsup et liminf des suites : ici la limite sup de la suite $(a_n)_{\mathbb{N}^*}$ est $\limsup(a_n) = \inf_{n\in\mathbb{N}^*}(\sup_{k\geq n} a_n) = 1$ (la plus grande valeur d'adhérence), et la limite inf de la suite $(a_n)_{\mathbb{N}^*}$ est $\liminf(a_n) = \sup_{n\in\mathbb{N}^*}(\inf_{k\geq n} a_n) = 0$ (la plus petite valeur d'adhérence). ▟

Proposition 3.57 *Si la suite $(A_n)_{\mathbb{N}^*}$ est monotone (croissante ou décroissante) dans $\mathcal{P}(\Omega)$, alors* $\limsup(A_n) = \lim(A_n) = \liminf(A_n)$.

Preuve. Si (A_n) est croissante, i.e. $A_n \subset A_{n+1}$ pour tout n, notons $A_\infty = \lim_{n\to\infty} A_n = \bigcup_{\mathbb{N}^*} A_n$, cf. (2.7). Et donc $B_n = \bigcup_{k\geq n} A_k = A_\infty$, et $C_n = \bigcap_{k\geq n} A_k = A_n$. Donc $\bigcap_{\mathbb{N}^*} B_n = A_\infty = \bigcup C_n$.

Si (A_n) est décroissante, i.e. $A_n \supset A_{n+1}$ pour tout n, notons $A_\infty = \lim_{n\to\infty} A_n = \bigcap_{\mathbb{N}^*} A_n$, cf. (2.8). Et donc $B_n = \bigcup_{k\geq n} A_k = A_n$, et $C_n = \bigcap_{k\geq n} A_k = A_\infty$. Donc $\bigcap_{\mathbb{N}^*} B_n = A_\infty = \bigcup C_n$. ▟

Proposition 3.58 *On a (relation entre* lim sup *et* lim inf*) :*

$$(\limsup_{n\to\infty}(A_n))^C = \liminf_{n\to\infty}(A_n^C). \tag{3.42}$$

Preuve. $\liminf_{n\to\infty}(A_n^C) = \bigcup_{\mathbb{N}^*}(\bigcap_{k\geq n} A_k^C) = \bigcup_{\mathbb{N}^*}((\bigcup_{k\geq n} A_k)^C) = \bigcup_{\mathbb{N}^*}(B_n^C) = (\bigcap_{\mathbb{N}^*} B_n)^C = (\limsup_{n\to\infty} A_n)^C$. ▟

Exemple 3.59 Suite de l'exemple 3.56 : donc $(\limsup_{n\to\infty}(A_n))^C = \mathbb{R} - \{0\}$.

Et $A_{2n}^C = \mathbb{R} - \{1 + \frac{1}{2n}\}$ et $A_{2n+1}^C = \mathbb{R} - \{0\}$. Donc $\bigcap_{k\geq n} A_k^C = \mathbb{R} - \{0, 1 + \frac{1}{2n}, 1 + \frac{1}{2(n+1)}, ...\}$, donc $\liminf(A_n^C) = \mathbb{R} - \{0\}$. ▟

3.12.2 Lemme du zéro-un

Théorème 3.60 *(dit lemme du zéro-un de Borel–Cantelli.)*

Soit $(\Omega, \mathcal{T}, \mu)$ un espace mesuré, et soit $(A_n)_{\mathbb{N}} \in \mathcal{T}^{\mathbb{N}}$ une suite d'ensembles mesurables.

1- On a :

$$\text{si} \qquad \sum_{n\in\mathbb{N}} \mu(A_n) < \infty \qquad \text{alors} \qquad \mu(\limsup_{n\to\infty} A_n) = 0, \tag{3.43}$$

i.e. si la série de terme générale "la mesure de A_n" est convergente alors la mesure de la limite supérieure est nulle, et

1'- Cas particulier $\mu = P$ mesure de probabilité : si $\sum_{n\in\mathbb{N}} P(A_n) < \infty$ alors la probabilité qu'une infinité de A_n se réalise est nulle ; ou encore, avec une probabilité égale à 1 au plus un nombre fini d'événements A_n se réalise.

Et dans (3.43) il n'est pas suffisant de supposer $\lim_{n\to\infty} \mu(A_n) \to 0$ pour que : avec une probabilité égale à 1, au plus un nombre fini d'événements A_n se réalise.

2- Si P est une mesure de probabilité, et si $(A_n)_{\mathbb{N}}$ est une suite d'événements indépendants, on a :

$$\text{si} \qquad \sum_{n\in\mathbb{N}} P(A_n) = \infty \qquad \text{alors} \qquad P(\limsup_{n\to\infty} A_n) = 1, \tag{3.44}$$

i.e., si les A_n sont indépendants et si $\sum_{n\in\mathbb{N}} P(A_n) = \infty$, alors une infinité d'événements A_n sont nécessairement réalisés. On peut réécrire (3.44) comme, la famille $(A_n)_{\mathbb{N}}$ étant indépendante :

$$\text{si} \qquad \sum_{n\in\mathbb{N}} P(A_n) = \infty \qquad \text{alors} \qquad \forall n \in \mathbb{N}, \quad P(\bigcup_{k\ge n} A_k) = 1. \tag{3.45}$$

Preuve. 1- Soit $B_n = \bigcup_{k\ge n} A_k$. On a $\mu(B_n) \le \sum_{k\ge n} \mu(A_k)$ reste d'une série convergente, donc $\mu(B_n) \underset{n\to\infty}{\longrightarrow} 0$.

Et $(B_n)_{n\in\mathbb{N}}$ est une suite décroissante, donc $(\mu(B_n))_{\mathbb{N}}$ est une suite décroissante ; avec $\mu(B_1) < \infty$, donc $\lim_{n\to\infty} \mu(B_n) = \mu(\lim_{n\to\infty} B_n)$, cf. proposition 2.38 page 24. Donc $0 = \lim_{n\to\infty} \mu(B_n) = \mu(\bigcap_{n\in\mathbb{N}} B_n) = \mu(\limsup A_n)$. D'où (3.43).

Et il n'est pas suffisant de supposer $\mu(A_k) \to_{k\to\infty} 0$: prendre $\Omega = [0,1]$ et μ mesure de Lebesgue, $A_0 = [0,1] =^{\text{noté}} A_{f_0}$, puis $A_1 = [0, \frac{1}{2}]$, $A_2 = [\frac{1}{2}, 1] =^{\text{noté}} A_{f_1}$, puis $A_3 = [0, \frac{1}{3}]$, $A_4 = [\frac{1}{3}, \frac{2}{3}]$, $A_5 = [\frac{2}{3}, 1] =^{\text{noté}} A_{f_2}$, puis... (faire un dessin). On a $f_0 = 0$, $f_1 = f_0 + 2 = 2$, $f_2 = f_1 + 3 = 5$, ..., $f_n = f_{n-1} + n+1$ (donc suite croissante), on a $B_n = [0,1]$ pour tout n, d'où $\mu(\limsup A_n) = 1$, bien que $|A_{f_k}| = \frac{1}{k}$ et donc $\lim_{k\to\infty} \mu(A_k) \to 0$. (Ici $\sum_{n\in\mathbb{N}} \mu(A_n) \ge \sum_{n\in\mathbb{N}} \mu(A_{f_n}) \ge \sum_{n\in\mathbb{N}^*} \frac{1}{n} = \infty$.)

2- $P(\bigcup_{k\ge n} A_k) = 1 - P(\bigcap_{k\ge n} A_k^C) = 1 - \prod_{k=n}^{\infty} P(A_k^C)$, car les A_k^C sont indépendants (puisque les A_k le sont, cf. proposition 3.52), soit $P(B_n) = 1 - \prod_{k=n}^{\infty} (1 - P(A_k)) \ge 1 - \prod_{k=n}^{\infty} e^{-P(A_k)}$ car $1 - x \le e^{-x}$. Donc $P(B_n) \ge 1 - e^{-\sum_{k=n}^{\infty} P(A_k)} = 1$ car $\sum_{k=n}^{\infty} P(A_k) = \infty$ (c'est le reste d'une série positive divergente). D'où (3.44), donc (3.45).

Puis soit A_i tel que $P(A_i) \ne 0$, et supposons que A_i ne soit réalisé qu'un nombre fini de fois. Soit N sa dernière réalisation. Alors pour tout $n > N$ on a $1 = P(A_i \cup B_n) = P(A_i) + P(B_n)$, donc $P(B_n) = 1 - P(A_i)$ ne converge pas vers 1. Contraire au résultat, donc A_i est nécessairement réalisé une infinité de fois. ∎

Exemple 3.61 Répétition dans un jeu de pile ou face.

On lance 2 fois une pièce équilibrée. L'univers est $\Omega = \{p, f\}^2$. Et pour un résultat $\vec{\omega} \in \Omega$, par exemple $\vec{\omega} = (p,p)$, on a $P(\vec{\omega}) = \frac{1}{4}$.

On répète une infinité de fois cette expérience. Soit $(\vec{\omega}_n)_{n\in\mathbb{N}^*} \in \Omega^{\mathbb{N}^*}$ une suite infinie de résultats indépendants, avec $\vec{\omega}_k =^{\text{noté}} (\omega_{k_1}, \omega_{k_2}) \in \Omega$. Soit $A_n = \{\vec{\omega}_n\}$ (le singleton $(\omega_{n_1}, \omega_{n_2})$). Soit $B_n = \bigcup_{k\ge n} A_n = \bigcup_{k\ge n} \{\vec{\omega}_k\} = \bigcup_{k\ge n} \{(\omega_{k_1}, \omega_{k_2})\} \subset \Omega$. Et le lemme donne $P(B_n) \to_{n\to\infty} 1$, car les ω_k sont indépendants et $\sum_{n=1}^{\infty} \frac{1}{4} = \infty$.

Ainsi, de manière presque certaine (avec une probabilité 1), on va retrouver tout événement $\vec{\omega}_n$ (ici l'un des 4 événements possibles) une infinité de fois dans la suite$(\vec{\omega}_n)_{n\in\mathbb{N}^*}$.

De même, étant donné un mot M de longueur donnée, en tapant au hasard une infinité de fois des mots de même longueur, on est certain d'avoir tapé le mot M, et ce une infinité de fois. ▪

Exercice 3.62 Exercice sur les suites : dans la démonstration du théorème 3.60 on a la suite $f_n = f_{n-1}+n+1$ pour $n\geq 1$. Donner f_n en fonction de n.

Réponse. $f_1 = f_0+2$, $f_2 = f_1+3$, ..., $f_n = f_{n-1}+n+1$. On ajoute pour obtenir : $f_1+...+f_n = f_0+...+f_{n-1}+1+2+...+(n+1)$, et il reste $f_n = f_0+\frac{(n+1)(n+2)}{2}$. ▪

4 Variable aléatoire réelle et loi d'une v.a.r.

[real-valued random variable]

4.1 Définition

4.1.1 Rappel : fonction mesurable

Voir polycopié "intégrale de Lebesgue".

Soit $(E,\mathcal{T}_E)$ un espace mesurable et soit $(\mathbb{R},\mathcal{T}_\mathbb{R})$ l'espace des réels muni de sa tribu borélienne.

Définition 4.1 Une fonction $f : E\to\mathbb{R}$ est mesurable ssi pour tout $a\in\mathbb{R}$ on a $f^{-1}(]-\infty,a])\in\mathcal{T}_E$, i.e. ssi l'image réciproque de tout intervalle $]-\infty,a]$ est dans la tribu $\mathcal{T}_E$ de E

(Définition équivalente : une fonction $f : E\to\mathbb{R}$ est mesurable ssi pour tout I borélien de $\mathbb{R}$ on a $f^{-1}(I)\in\mathcal{T}_E$.)

Exemple 4.2 Si $\mathcal{T}_E = \mathcal{P}(E)$ alors toute fonction $f : E\to\mathbb{R}$ est mesurable. ▪

Remarque 4.3 La notion de fonction mesurable n'est donc pertinente que quand $\mathcal{T}_E \subsetneq \mathcal{P}(E)$, comme dans le cas pour $(E,\mathcal{T}_E) = (\mathbb{R},\mathcal{T}_\mathbb{R})$ où $\mathcal{T}_\mathbb{R}$ est la tribu borélienne, voir polycopié "intégrale de Lebesgue".

Dans la pratique ingénieur usuelle, "toutes" les fonctions $f : (\mathbb{R},\mathcal{P}(\mathbb{R}))\to\mathbb{R}$ sont mesurables, cf. exemple 2.26 page 22. (Ce n'est pas le cas mathématiquement, voir exemple suivant.) ▪

Exemple 4.4 Soit E muni d'une tribu $\mathcal{T}_E$ quelconque. La fonction $f = 1_A$ est mesurable ssi $A\in\mathcal{T}_E$. En effet :

si $a<0$ alors $1_A^{-1}(]-\infty,a]) = \emptyset$,

si $a\in[0,1[$ alors $1_A^{-1}(]-\infty,a]) = A^C$,

si $a\geq 1$ alors $1_A^{-1}(]-\infty,a]) = E$,

et $A^C\in\mathcal{T}_E$ ssi $A\in\mathcal{T}_E$. ▪

Définition 4.5 Plus généralement : soit $(E,\mathcal{T}_E)$ un ensemble mesurable, soit $(F,\mathcal{O}_F)$ un ensemble muni d'une topologie $\mathcal{O}_F$. Une fonction $f : E\to F$ est mesurable ssi pour tout $U\in\mathcal{O}_F$ (ouvert de F) on a $f^{-1}(U)\in\mathcal{T}_E$ (son image réciproque est mesurable dans E).

Proposition 4.6 et définition. *Soit $\mathcal{T}_F$ la tribu engendrée par la topologie $\mathcal{O}_F$.*

Si $f : (E,\mathcal{T}_E)\to(F,\mathcal{T}_F)$ est mesurable, alors posant :

$$f^{-1}(\mathcal{T}_F) = \{A = f^{-1}(B) : B\in\mathcal{T}_F\},$$

le sous-ensemble $f^{-1}(\mathcal{T}_F) =^{\text{noté}} \sigma(f)$ de $\mathcal{P}(E)$ est une tribu, sous-tribu de $\mathcal{T}_E$, appelée tribu engendrée par f.

Preuve. On vérifie facilement les conditions de la définition 2.18 à l'aide de (1.5). ▪

Exemple 4.7 Pour $A\in\mathcal{T}_E$ et $f = 1_A : E\to\mathbb{R}$ la fonction indicatrice de A, on a $\sigma(1_A) = 1_A^{-1}(\mathcal{T}_\mathbb{R}) = \{\emptyset, A, A^C, E\}$, l'ensemble des ensembles pertinents pour 1_A. ▪

4.1.2 Variable aléatoire réelle (v.a.r.)

Définition 4.8 En probabilité, une fonction mesurable $f : \Omega \to \mathbb{R}$ est appelée une variable aléatoire réelle (v.a.r.).

Donc, si $(\Omega, \mathcal{T}_\Omega, P)$ un espace probabilisé, alors f est une v.a.r. ssi f est mesurable, i.e. ssi, pour tout $a \in \mathbb{R}$ on a $f^{-1}(]-\infty, a[) \in \mathcal{T}_\Omega$.

(Donc, en pratique ingénieur, une v.a.r. est une fonction.)

Définition 4.9 Cas particulier : une v.a.r. est dite déterministe ssi f est une fonction constante (une seule possibilité de résultat, soit une seule issue).

Et en probabilité on note souvent $f =^{\text{noté}} X$ une v.a.r. :

$$f \overset{\text{noté}}{=} X : \begin{cases} \Omega & \to \mathbb{R}, \\ \omega & \mapsto x = X(\omega). \end{cases}$$

Remarque 4.10 Une variable aléatoire est utilisée pour donner une "valeur" à un résultat ω. ▪

Définition 4.11 Une v.a.r. X est dite discrète ssi $X(\Omega)$ est discret (i.e. est de cardinal fini ou dénombrable).

Une v.a.r. X est dite continue ssi $\mathrm{Im}(X) = X(\Omega)$ est un intervalle ouvert non vide de $\mathbb{R}$.

Exemple 4.12 Jeu de pile ou face, $X(\text{pile}) = a$ et $X(\text{face}) = b$, avec a la valeur du "gain" quand on a obtenu pile, et avec b la valeur du "gain" quand on a obtenu face. On voudra connaître la probabilité de gagner a, i.e. on voudra connaître $P_X(a) = P(\{\omega : X(\omega) = a\} =^{\text{noté}} P(X = a)$, probabilité pour que la valeur de X soit a. ▪

Exemple 4.13 On lance deux dés, résultat $\vec{\omega} = (\omega_1, \omega_2)$, on s'intéresse à la somme $S = X(\vec{\omega}) = \omega_1 + \omega_2$, On voudra connaître la probabilité $P_X(4)$ d'obtenir la somme 4. Ici :

$P_X(4) = P(\{\vec{\omega} : X(\vec{\omega}) = 4\} = P((1,3)) + P((2,2)) + P((3,1)) = \frac{3}{36} = \frac{1}{12} =^{\text{noté}} P(X = 4)$, probabilité pour que la valeur de X soit 4 (pour des dés équilibrés). ▪

4.1.3 Notations ensemblistes $\{X \in I\}$ et $\{X = x\}$

Définition 4.14 Si $X : \Omega \to \mathbb{R}$ est une v.a.r. et $I \in \mathcal{T}_\mathbb{R}$, on note :

$$\{X \in I\} \overset{\text{déf}}{=} X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\} \qquad (\subset \Omega) \tag{4.1}$$

l'ensemble des éventualités $\omega \in \Omega$ t.q. la valeur $X(\omega)$ prise par ω est dans I.

Et $\{X \in I\}$ est l'événement de Ω appelé "X prend ses valeurs dans I".

Définition 4.15 Et dans le cas particulier $I = \{x\}$, on note :

$$\{X = x\} \overset{\text{déf}}{=} \{X \in \{x\}\} \qquad (= X^{-1}(\{x\}) = \{\omega \subset \Omega : X(\omega) = x\}), \tag{4.2}$$

appelé l'événement "X prend la valeur x".

Remarque 4.16 Cette notation $\{X = x\}$ ne doit pas créer de confusion : X est bien une fonction et non une valeur. Et $\{X = x\}$ est un ensemble : c'est le sous-ensemble $X^{-1}(\{x\})$ de Ω. ▪

Exemple 4.17 $A \in \mathcal{T}_\Omega$ et $X = 1_A$. Alors

$\{1_A = 1\} = A$,

$\{1_A = 0\} = A^C = \{1_A([-1, \frac{1}{2}])$,

$\{1_A \in [2,3]\} = \emptyset$... ▪

Exemple 4.18 Alternative simple pile ou face : soit $X : \Omega = \{p, f\} \to \mathbb{R}$ donnée par $X(f) = +1$ et $X(p) = -1$ (gain de 1 euro si c'est face, et perte de 1 euro si c'est pile).

Cas d'une pièce équilibrée : $P(X{=}1) = P(f) = \frac{1}{2}$ est la probabilité de gagner 1 euros ; et $P(X{=}-1) = P(p) = \frac{1}{2}$ est la probabilité de perdre 1 euros ; et $P(X \in [2,3]) = P(\emptyset) = 0$. ▪

Exemple 4.19 On lance deux dés et on s'intéresse à la somme. Ici l'univers est $\Omega = [1,6]_{\mathbb{N}}^2$ (l'ensemble des couples (n,m) résultats des lancers). Soit $S : \Omega \to \mathbb{R}$ la fonction somme sur Ω, $S(n,m) = n+m$. Par exemple :

$$\{S=11\} = S^{-1}(11) = \{(n,m) \in [1,6]_{\mathbb{N}}^2 : n{+}m = 11\} = \{(5,6),(6,5)\}. \tag{4.3}$$

Ainsi, pour un dé équilibré :

$$P(S=11) = P(\{(5,6),(6,5)\}) = P(\{(5,6)\}) + P(\{(6,5)\}) = \frac{2}{36}, \tag{4.4}$$

soit une chance sur dix-huit d'obtenir 11. Et on notera $P(S=11) =^{\text{noté}} P_S(11)$. ▪

4.2 Loi (de probabilité) d'une variable aléatoire : les probabilités images

[Law of a random variable, or, image measure of a random variable.]
Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé.

4.2.1 Loi d'une variable aléatoire

Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{T})$ (i.e. soit $X : \Omega \to \mathbb{R}$ une fonction mesurable).

Définition 4.20 La loi de la v.a.r. X (la distribution de la v.a.r. X) est la fonction $P_X : \mathcal{T}_{\mathbb{R}} \to [0,1]$ définie sur la tribu borélienne de $\mathbb{R}$ par :

$$P_X \stackrel{\text{déf}}{=} P \circ X^{-1}, \tag{4.5}$$

i.e., pour tout I intervalle de $\mathbb{R}$ (ou tout I borélien de $\mathbb{R}$) :

$$P_X(I) \stackrel{\text{déf}}{=} P(X^{-1}(I)) = P(X \in I) \qquad (= P(\{\omega \in \Omega : P(\omega) \in I\})). \tag{4.6}$$

C'est la probabilité de l'événement "X prend ses valeurs dans I".

Et dans le cas particulier $I = \{x\}$ (singleton) on note :

$$P_X(x) \stackrel{\text{déf}}{=} P_X(\{x\}) \qquad \text{donc} \qquad = P(X=x) = P(\{\omega \in \Omega : P(\omega) = x\}). \tag{4.7}$$

Exemple 4.21 Suite de l'exemple 4.19 (lancer de deux dés, somme S) : $P_S(11) = \frac{2}{36}$. ▪

Exemple 4.22 Soit $X = 1_A$ où $A \in \mathcal{T}_\Omega$. Alors :

$$P_{1_A}(\{1\}) = P(A). \tag{4.8}$$

En effet $P_{1_A}(\{1\}) = P(1_A \in \{1\}) = P(\{\omega : 1_A(\omega) = 1\}) = P(A)$. ▪

4.2.2 Proposition de la mesure image

Pour montrer que P_X est une probabilité, on commence par la proposition de la mesure image :
Soit $(E, \mathcal{T}_E)$ un espace mesurable, soit $(F, \mathcal{O}_F)$ un espace topologique, soit $\mathcal{T}_F$ la tribu engendrée par $\mathcal{O}_F$. Soit $\varphi : E \to F$ une fonction mesurable.
Soit $\mu : \mathcal{T}_E \to \mathbb{R}_+$ une mesure sur l'espace mesurable $(E, \mathcal{T}_E)$.

Proposition 4.23 et définition. *Alors la fonction $\nu : \mathcal{T}_F \to \mathbb{R}_+$ définie par :*

$$\nu \stackrel{\text{déf}}{=} \mu \circ \varphi^{-1} \tag{4.9}$$

est une mesure sur $(F, \mathcal{T}_F)$. De plus, si μ est une mesure finie (telle que $\mu(E) < \infty$) alors ν est également finie.

Cette mesure ν est appelée mesure image de μ par φ. Elle est donc donnée par, pour tout $B \in \mathcal{T}_F$:

$$\nu(B) = \mu(\varphi^{-1}(B)) \qquad (= \mu(\{\omega : \varphi(\omega) \in B\}) = \mu(\varphi \in B)), \tag{4.10}$$

encore noté :

$$\int_{x \in B} d\nu(x) = \int_{\omega \in \varphi^{-1}(B)} d\mu(\omega), \qquad \text{ou bien} \qquad \int_B d\nu = \int_{\varphi^{-1}(B)} d\mu. \tag{4.11}$$

Preuve. Il s'agit pour ν de vérifier les points 1-et 2-, mais pas le point 3-, de la définition 2.33 page 23.

1- On a $\nu(\emptyset) = \mu(\varphi^{-1}(\emptyset)) = \mu(\emptyset) = 0$ car μ est une mesure.

2- Soit $(B_i)_{i\in\mathbb{N}}$ une famille dans $\mathcal{T}_F$ telle que $B_i \cap B_j = \emptyset$ pour tout $i \neq j$. Vérifions $\nu(\bigcup_i B_i) = \sum_i \nu(B_i)$.

Posons $A_i = \varphi^{-1}(B_i)$; $A_i \in \mathcal{T}_E$ car φ est mesurable. On a $\nu(B_i) = \mu(A_i)$ par définition de ν.

Pour $i \neq j$, ayant $B_i \cap B_j = \emptyset$ on a $A_i \cap A_j = \emptyset$ (proposition 1.5).

D'où, μ étant une mesure, $\sum_i \nu(B_i) = \sum_i \mu(A_i) = \mu(\bigcup_i A_i)$, d'où $\sum_i \nu(B_i) = \mu(\bigcup_i \varphi^{-1}(B_i)) = \mu(\varphi^{-1}(\bigcup_i B_i))$ (proposition 1.5), donc $= \nu(\bigcup_i B_i)$.

Cas μ mesure finie : ayant $\varphi^{-1}(F) \subset E$, on a $\nu(F) = \mu(\varphi^{-1}(F)) \leq \mu(E) < \infty$. ∎

Exemple 4.24 $\mu = dx$ mesure de Lebesgue sur $\mathbb{R}$ et $\varphi = 1_{\mathbb{R}_+} : \mathbb{R} \to \mathbb{R}$. Soit $\nu = \mu \circ \varphi^{-1}$. Pour $B \in \mathcal{T}_\mathbb{R}$ on a $1_{\mathbb{R}_+}^{-1}(B) =$ soit $\emptyset$, soit $\mathbb{R}_-^*$, soit $\mathbb{R}_+$, soit $\mathbb{R}$. Donc $\nu(B) = \mu(1_{\mathbb{R}_+}^{-1}(B)) = 0$ ou $= \infty$. Noter que ν n'est pas σ-finie (le point 3- de la définition 2.33 n'est pas vérifié). ∎

4.2.3 Probabilité image P_X (ou loi de X, ou distribution de X)

Corollaire 4.25 *Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé, et soit $(\mathbb{R}, \mathcal{T}_\mathbb{R})$ l'espace mesurable $\mathbb{R}$ muni de sa tribu borélienne $\mathcal{T}_\mathbb{R}$. Si $X : \Omega \to \mathbb{R}$ est une v.a.r., alors la fonction :*

$$P_X = P \circ X^{-1} \tag{4.12}$$

est une (mesure de) probabilité sur l'espace mesurable $(\mathbb{R}, \mathcal{T}_\mathbb{R})$.

Preuve. P étant une mesure, P_X était également une mesure, cf. proposition 4.23. Et on a $P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$ car P est une probabilité. Donc P_X est une probabilité (mesure de masse 1). ∎

Définition 4.26 La mesure image P_X (de la probabilité P par la v.a.r. X) est appelée la loi de probabilité de la v.a.r. X, encore appelée la loi de X, ou la distribution de X.

Donc, cf. notation (4.11) :

$$\int_{x\in B} dP_X(x) = \int_{\omega \in X^{-1}(B)} dP(\omega), \qquad \text{ou bien} \qquad \int_B dP_X = \int_{X^{-1}(B)} dP. \tag{4.13}$$

Exercice 4.27 Soit $X : \Omega \to \mathbb{R}$ une fonction constante : donc il existe $c \in \mathbb{R}$ t.q. pour tout $\omega \in \Omega$ on a $X(\omega) = c =^{\text{noté}} \overline{X}$ (la valeur de X). Autrement dit $X = \overline{X} 1_\Omega$.

Montrer que la loi de X est la loi discrète qui est la mesure de Dirac en $\overline{X}$, i.e. montrer :

$$X = \overline{X} 1_\mathbb{R} \qquad \Longrightarrow \qquad P_X = \delta_{\overline{X}}\,, \tag{4.14}$$

soit encore $P_X(\overline{X}) = 1$, et $P_X(x) = 0$ pour tout $x \neq \overline{X}$; soit encore, pour $I \in \mathcal{T}_\mathbb{R}$:

$$\begin{cases} P_X(I) = 1 & \text{si } \overline{X} \in I, \\ P_X(I) = 0 & \text{si } \overline{X} \notin I. \end{cases}$$

Réponse. On a $P_X(I) = P(X^{-1}(I)) = P(\{\omega : X(\omega) \in I\})$, avec ici $\mathrm{Im}(X) = \{\overline{X}\}$; donc pour I donné, si $I \cap \mathrm{Im}(X) \supset \{\overline{X}\}$, i.e. si $\overline{X} \in I$ alors $P_X(I) = P(\Omega) = 1$, et si $\overline{X} \notin I$ alors $P_X(I) = P(\emptyset) = 0$. ∎

Exemple 4.28 Voir exemple 2.48 page 25. ∎

Exemple 4.29 Voir exemple 4.21. ∎

4.2.4 Fonction de répartition de la loi d'une variable aléatoire

On reprend le paragraphe 2.7 où la probabilité considérée est maintenant P_X.

Définition 4.30 On appelle fonction de répartition de la variable aléatoire X [cumulative distribution function of a real-valued random variable X, or distribution function of X] la fonction $F_X : \mathbb{R} \to \mathbb{R}$ de répartition de la probabilité P_X :

$$F_X(t) = P_X(]-\infty, t]) \qquad (= P(X^{-1}(]-\infty, t]))). \tag{4.15}$$

(Intervalle fermé en t.) Autrement dit $F_X(t) = P(X \leq t)$, soit $F_X(t) = P(\{\omega : X(\omega) \leq t\})$.

4.2.5 Variable aléatoire de loi discrète

Définition 4.31 La v.a.r. X est dite de loi discrète ssi P_X est une probabilité discrète, i.e. de la forme :

$$P_X = \sum_{i\in I} p_i \delta_{x_i}, \tag{4.16}$$

où I est un ensemble fini ou dénombrable, où $x_i \neq x_j$ pour tout $i \neq j$, où $p_i \geq 0$ pour tout i, et où $\sum_{i\in I} p_i = 1$. En particulier $P_X(x_i) = p_i$ est la probabilité que X prenne la valeur x_i.

Proposition 4.32 *Cas $\Omega = \{\omega_1, ..., \omega_n, ...\}$ ensemble fini ou dénombrable, et $X : \Omega \to \mathbb{R}$ une fonction mesurable : c'est nécessairement une v.a.r. de loi discrète.*

Preuve. $\mathrm{Im}(X) \subset \{X(\omega_1), ..., X(\omega_n), ...\} = \{x_1, ..., x_k, ...\} =^{\text{noté}} (x_k)_{k\in K}$ où $K = |\mathrm{Im}(X)|$ (le cardinal) est donc fini ou dénombrable.

On pose $p_k = P_X(x_k) = P(X^{-1}(x_k)) = P(\{\omega_i : X(\omega_i) = x_k\})$, et on a imédiatement $P_X = \sum_{k\in K} p_k \delta_{x_k}$. ∎

Exemple 4.33 On lance deux dés équilibrés : on a $\Omega = [1,6]_{\mathbb{N}} \times [1,6]_{\mathbb{N}}$. Soit $S(i,j) = i+j$. Alors sur $\mathbb{R}$ on a $P_S = \sum_{i=2}^{12} p_i \delta_i$ où $p_2 = P_S(2) = P(1,1) = \frac{1}{36}$, $p_3 = P_S(3) = P\{(1,2),(2,1)\} = \frac{2}{36}$, ... La loi de la somme est une loi discrète. ∎

Exemple 4.34 Voir exercice 4.27. ∎

Proposition 4.35 *Soient $N+1$ réels $a_0, a_1, ..., a_N \in \mathbb{R}$ t.q. $a_{i-1} < a_i$ pour tout $i = 1, ..., N$, et $[a_0, a_N[= \bigcup_{i=1}^N [a_{i-1}, a_i[$ (partition). Soit P une probabilité sur $\Omega = [a_0, a_N[$. On note :*

$$p_i = P([a_{i-1}, a_i[). \tag{4.17}$$

Soit $x_1, ..., x_N \in \mathbb{R}$ et soit $X : \Omega \to \mathbb{R}$ la fonction en escalier donnée par :

$$X = \sum_{i=1}^N x_i 1_{[a_{i-1}, a_i[}. \tag{4.18}$$

Alors la loi de X est la loi discrète :

$$P_X = \sum_{i=1}^N p_i \delta_{x_i}, \qquad \text{avec} \qquad P_X(x_i) = p_i. \tag{4.19}$$

Preuve. On a $X(\omega) = \sum_{i=1}^N x_i 1_{[a_{i-1},a_i[}(\omega)$, donc $X(\omega) = x_i$ ssi $\omega \in [a_{i-1}, a_i[$, donc $X^{-1}(x_i) = [a_{i-1}, a_i[$. Donc $P_X(x_i) = P(X^{-1}(x_i)) = P([a_{i-1}, a_i[) = p_i$.

Et si $x \neq x_i$ pour tout i alors $P_X(x) = 0$. D'où $P_X = \sum_{i=1}^N P([a_{i-1}, a_i[)\delta_{x_i}$. ∎

Exercice 4.36 Montrer : la fonction de répartition $F_X : \mathbb{R} \to \mathbb{R}$ d'une v.a.r. discrète est la fonction croissante en escalier donnée par, avec (4.16) :

$$F_X = \sum_{i\in I} p_i 1_{[x_i, \infty[}$$

continue à droite, donc donnée par $F_X(x) = \sum_{i\in I} p_i 1_{[x_i,\infty[}(x)$. Faire un dessin : on ajoute une marche de hauteur p_i en chaque point x_i.

Réponse. On a $1_{]-\infty,x]}(x_i) = 1_{[x_i,\infty[}(x)$ (vaut 1 ssi $x_i \leq x$ et vaut 0 sinon).

Donc $F_X(x) = P_X(]-\infty, x]) = \sum_{i\in I} p_i \delta_{x_i}(]-\infty, x]) = \sum_{i\in I} p_i 1_{]-\infty,x]}(x_i) = \sum_{i\in I} p_i 1_{[x_i,\infty[}(x)$. ∎

4.2.6 Variable aléatoire continue et sa loi

Définition 4.37 La v.a.r. $X : \Omega \to \mathbb{R}$ est une loi de probabilité P_X continue ssi F_X est absolument continue, i.e. ssi "$dP_X(x) = p_X(x)\,dx$", i.e. ssi :

$$P_X([a,b]) = \int_a^b p_X(x)\,dx, \tag{4.20}$$

où $p : \mathbb{R} \to \mathbb{R}_+$ est une fonction positive (une densité) de masse unité : $\displaystyle\int_{\mathbb{R}} p_X(x)\,dx = 1$.

En particulier on a dans ce cas :

$$P_X([a,b]) = P_X(]a,b[) = P_X(]a,b]) = P_X([a,b[). \tag{4.21}$$

Exemple 4.38 Loi normale centrée réduite sur $\mathbb{R}$: $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. ♣

Exemple 4.39 Loi uniforme sur un intervalle $[\alpha, \beta]$ (avec $\alpha < \beta$) : $p(x) = \frac{1}{\beta-\alpha}$. ♣

Exemple 4.40 p fonction en escalier : $p(x) = \sum_{i=1}^n \frac{p_i}{x_i - x_{i-1}} 1_{]x_{i-1},x_i[}$ (avec $x_{i-1} < x_i$ pour tout i), où les $p_i \geq 0$. (Avec donc p nulle en dehors de $[x_0, x_n]$), et où nécessairement $\sum p_i = 1$ puisque $\int_{\mathbb{R}} p(x)\,dx = 1$ s'écrit $\sum_{i=1}^n \frac{p_i}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} dx = 1$.) Ici p (la densité) n'est pas continue, mais F_X est absolument continue : on est dans le cadre de la définition 4.37. ♣

5 Dépendance et indépendance de variables aléatoires

5.1 Notations $f \otimes g$, $f(X)$ et $f - c$

Notation : Si $f : A \to \mathbb{R}$ et $g : B \to \mathbb{R}$ sont des fonctions, alors $f \otimes g : A \times B \to \mathbb{R}$ est la fonction de deux variables, dit fonction à variables séparées, définie par, pour tout $(a,b) \in A \times B$:

$$(f \otimes g)(a,b) = f(a)g(b). \tag{5.1}$$

Et $f \otimes g$ est appelé produit tensoriel de f et g.

(Une fonction $h : A \times B \to \mathbb{R}$ est dite à variables séparées ssi il existe $f : A \to \mathbb{R}$ et $g : B \to \mathbb{R}$ t.q. $h = f \otimes g$.)

Notation : Si $f : \mathbb{R} \to \mathbb{R}$ et $X : \Omega \to \mathbb{R}$ est une v.a.r, on note $f \circ X =^{\text{noté}} f(X)$. Donc :

$$f(X) \overset{\text{déf}}{=} f \circ X : \begin{cases} \Omega & \to \mathbb{R} \\ \omega & \to f(X)(\omega) = f(X(\omega)). \end{cases} \tag{5.2}$$

Attention : cette notation est source de confusion : X n'est pas une variable dans (5.2), c'est bien une fonction. Dans les démonstrations, on utilisera $f \circ X$ pour lever les ambiguïtés.

Abus de notation : pour $c \in \mathbb{R}$ (une constante), on fait l'abus de notation usuel consistant à noter c la fonction constante $c1_\Omega$:

$$c1_\Omega \overset{\text{noté}}{=} c. \tag{5.3}$$

Donc on fait jouer à c deux rôles : celui d'un réel et celui d'une fonction, le contexte devant lever toute ambiguïté. Ainsi, si $f : \Omega \to \mathbb{R}$ est une fonction :

$$f - c1_\Omega \overset{\text{noté}}{=} f - c. \tag{5.4}$$

Attention : cet abus de notation (5.3) est source de confusion. Dans les démonstrations, on utilisera $c1_\Omega$ pour lever les ambiguïtés.

5.2 Dépendance et indépendance de 2 variables aléatoires

Soient 2 variables aléatoires réelles $X_1, X_2 : \Omega \to \mathbb{R}$ définies sur même un espace probabilisé $(\Omega, \mathcal{T}, P)$. On note, pour tous les intervalles $I_i \subset \mathbb{R}$, $i = 1, 2$ (ou pour tous les boréliens) :

$$\begin{aligned} P(X_1 \in I_1, X_2 \in I_2) &\overset{\text{déf}}{=} P((X_1 \in I_1) \cap (X_2 \in I_2)) \\ &= P(\{\omega \in \Omega : X_1(\omega) \in I_1 \text{ et } X_2(\omega) \in I_2\}), \end{aligned} \tag{5.5}$$

la probabilité que $X_1 \in I_1$ et $X_2 \in I_2$ soient simultanément réalisés.

Notations :

$$\begin{aligned} P(X_1 \in I_1, X_2 \in I_2) &\overset{\text{noté}}{=} P((X_1, X_2) \in I_1 \times I_2) \\ &\overset{\text{noté}}{=} P_{(X_1,X_2)}(I_1 \times I_2) \\ &\overset{\text{noté}}{=} P_{(X_1,X_2)}(I_1, I_2). \end{aligned} \tag{5.6}$$

Remarque 5.1 Dans la suite on notera $\vec{X} = (X_1, X_2) : \Omega \to \mathbb{R}^2$ le "vecteur aléatoire réel" (la fonction mesurable à valeurs vectorielles) donné par $\vec{X}(\omega) = (X_1(\omega), X_2(\omega))$. ▪

Exemple 5.2 On lance deux fois une pièce équilibrée. Soit $g : \{p, f\} \to \mathbb{R}$ la fonction "gain", $g(p) = a$ et $g(f) = b$. Pour $i = 1, 2$ on note X_i le gain du i-ème lancer. On a $P(X_1{\in}\{a\}, X_2{\in}\{a\}) =^{\text{noté}} P(X_1{=}a, X_2{=}a) =^{\text{noté}} P_{(X_1,X_2)}(a, a)$ la probabilité du gain a au premier et au second lancer, qui vaut 1/4 pour une pièce équilibrée et lancers "indépendants".

Détails : ici $\Omega = \{(p, p), (p, f), (f, p), (f, f)\}$ est l'univers des résultats des deux lancers, et si on note $\pi_i : \vec{\omega} = (\omega_1, \omega_2) \in \Omega \to \pi_i(\omega) = \omega_i$ la projection usuelle, on a $X_i : \Omega \to \{a, b\}$ donné par $X_i = g \circ \pi_i$, soit $X_i(\vec{\omega}) = g(\omega_i) = a$ ou b gain du i-ème lancer. ▪

Définition 5.3 Les v.a.r. X_1 et X_2 définies sur Ω sont dites indépendantes ssi, pour tous les intervalles $I_i \subset \mathbb{R}$, $i = 1, 2$:

$$P(X_1 \in I_1, X_2 \in I_2) = P(X_1 \in I_1) P(X_2 \in I_2). \tag{5.7}$$

I.e. la probabilité d'être dans l'intersection est égale au produit des probabilités.

Autrement dit, X_1 et X_2 sont indépendantes ssi leurs lois vérifient, pour tout I_1, I_2 :

$$P_{(X_1,X_2)}((I_1, I_2)) = P_{X_1}(I_1) P_{X_2}(I_2), \tag{5.8}$$

i.e. ssi (cf. (5.1)) :

$$P_{(X_1,X_2)} = P_{X_1} \otimes P_{X_2}. \tag{5.9}$$

Donc X_1 et X_2 sont indépendantes ssi la loi $P_{(X_1,X_2)}$ du couple (X_1, X_2) est le produit tensoriel $P_{X_1} \otimes P_{X_2}$ des lois.

Exemple 5.4 On lance deux fois un dé équilibré (sans tricher), X_i donnant le gain du i-ème lancer, cf. exercice précédent. Pour $x, y \in \{a, b\}$, on a $P(X_1{=}x, X_2{=}y) = P(X_1{=}x) P(X_2{=}y)$, et les lois sont indépendantes. ▪

Exercice 5.5 On considère deux tirages successifs de boules numérotés 1 à 3, sans replacement de la première boule tirée. X_i donnant le résultat du i-ème tirage. Montrer que X_1 et X_2 ne sont pas indépendantes.

Réponse. Ici $\Omega = \{(m, n) \in [1, 3]^2 \text{ t.q. } m \neq n\} = \{((1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$ est l'univers des résultats, et $|\Omega| = A_3^2 = 6$. Et, pour des tirages non biaisés, la probabilité P sur Ω est donnée par $P((m, n)) = \dfrac{1}{\text{card}(\Omega)} = \dfrac{1}{6}$ si $(m, n) \in \Omega$ et $= 0$ sinon.

$X_1 = \pi_1$ est la projection sur la première variable : $X_1(m, n) = m$, et

$X_2 = \pi_2$ est la projection sur la seconde variable : $X_2(m, n) = n$.

$P(X_1{=}x) = \dfrac{\text{card}\{(m, n) \in \Omega : \ m{=}x\}}{\text{card}(\Omega)} = \dfrac{2}{6} = \dfrac{1}{3}$ pour $x \in [1, 3]_{\mathbb{N}}$, et

$P(X_2{=}y) = \dfrac{\text{card}\{(m, n) \in \Omega : \ n{=}y\}}{\text{card}(\Omega)} = \dfrac{2}{6} = \dfrac{1}{3}$ pour $y \in [1, 3]_{\mathbb{N}}$.

Ainsi $P_{X_1} = P_{X_2}$ (les lois de X_1 et X_2 sont identiques).

Et $P(X_1{=}x, X_2{=}y)) = \frac{\mathrm{card}\{(m,n)\in\Omega:\ m=x,\ n=y\}}{\mathrm{card}(\Omega)} = \frac{1}{6}$ pour tout $(x,y) \in \Omega$.

Et $P(X_1{=}x)P(X_2{=}y) = \frac{1}{9} \neq \frac{1}{6} = P(X_1{=}x, X_2{=}y))$ pour tout $(x,y) \in [1,3]_{\mathbb{N}}$: (5.7) n'est pas vérifiée : les lois ne sont pas indépendantes.

D'ailleurs : $P(X_1{=}x|X_2{=}y) = P_{|X_2=y}(X_1{=}x) = \frac{1}{2}$ si $x \neq y$ et $= 0$ si $x = y$ (quand $x, y \in [1,6]_{\mathbb{N}}$), alors que $P(X_1{=}x) = \frac{1}{3}$ pour tout $x \in [1,3]_{\mathbb{N}}$, voir (3.30), donc $P_{|X_2=y}(X_1{=}x) \neq P(X_1{=}x)$. ▪

Proposition 5.6 *Si X et Y sont des v.a.r. indépendantes, alors pour toutes fonctions $f, g : \mathbb{R} \to \mathbb{R}$ mesurables, les v.a.r. $f(X) = f \circ X$ et $g(Y) = g \circ Y$ (cf. (5.2)) sont indépendantes :*

$$P_{(X,Y)} = P_X \otimes P_Y \quad \Longrightarrow \quad P_{(f(X),g(Y))} = P_{f(X)} \otimes P_{g(Y)}. \tag{5.10}$$

En particulier (toujours dans le cas X et Y indépendantes), pour tout $a, b \in \mathbb{R}$ et tout $\alpha, \beta \in \mathbb{R}^*$, la v.a.r. $\alpha X - a 1_\Omega =^{\text{noté}} \alpha X - a$ et la v.a.r. $\beta Y - b 1_\Omega =^{\text{noté}} \beta Y - b$ (cf. (5.4)) sont indépendantes :

$$P_{(X,Y)} = P_X \otimes P_Y \quad \Longrightarrow \quad P_{(\alpha X - a, \beta Y - b)} = P_{\alpha X - a} \otimes P_{\beta Y - b}. \tag{5.11}$$

Preuve. Si $I \in \mathcal{T}_{\mathbb{R}}$ on a $(f \circ X \in I) = \{\omega : f(X(\omega)) \in I\} = \{\omega : X(\omega) \in f^{-1}(I)\}$, avec $f^{-1}(I) \in \mathcal{T}_{\mathbb{R}}$ car f est mesurable. Donc $(f(X) \in I) = (X \in f^{-1}(I))$.

Donc si X et Y sont indépendantes et si $I, J \in \mathcal{T}_{\mathbb{R}}$, alors :

$$P(f \circ X \in I, g \circ Y \in J) = P(X \in f^{-1}(I), Y \in g^{-1}(J)) = P(X \in f^{-1}(I))P(Y \in g^{-1}(J)),$$

cf. (5.7). Comme $(X \in f^{-1}(I)) = \{\omega : X(\omega) \in f^{-1}(I)\} = \{\omega \in X^{-1}(f^{-1}(I)) = (f \circ X)^{-1}(I)\}$, idem pour Y, on déduit :

$$P_{(f\circ X, g\circ Y)}(I, J) = P_{f\circ X}(I) P_{g\circ Y}(J),$$

et les v.a.r. $f \circ X$ et $g \circ Y$ sont indépendantes.

Avec $f(x) = \alpha x - a$, on obtient $(f \circ X)(\omega) = \alpha X(\omega) - a$ donc $f(X) =^{\text{déf}} f \circ X = \alpha X - a 1_\Omega$, d'où (5.11). ▪

Notation. Soit $\varphi : \mathbb{R}^2 \to \mathbb{R}$ une v.a.r. (tribus boréliennes de $\mathbb{R}^2$ et de $\mathbb{R}$). Si $X, Y : \Omega \to \mathbb{R}$ sont deux v.a.r. sur un même espace probabilisé $(\Omega, \mathcal{T}, P)$, on note $\varphi(X,Y) : \Omega \to \mathbb{R}$ la v.a.r. :

$$\varphi(X,Y) \overset{\text{déf}}{=} \varphi \circ (X,Y), \qquad \text{donc} \qquad \varphi(X,Y)(\omega) = \varphi(X(\omega), Y(\omega)), \quad \forall \omega \in \Omega. \tag{5.12}$$

Proposition 5.7 *Si X et Y sont v.a.r. indépendantes et si $\varphi : \mathbb{R}^2 \to \mathbb{R}$ est mesurable, alors :*

$$P_{\varphi(X,Y)} = (P_X \otimes P_Y) \circ \varphi^{-1}, \qquad \text{quand } X \text{ et } Y \text{ indépendantes}, \tag{5.13}$$

i.e. $P_{\varphi(X,Y)}$ est la mesure image par φ de $(P_X \otimes P_Y)$.

Donc, toujours avec X et Y indépendantes, pour tout fonction f mesurable telle que $f \circ \varphi$ est $P_X \otimes P_Y$ intégrable, on a :

$$P_{\varphi(X,Y)}(f) = (P_X \otimes P_Y)(f \circ \varphi), \tag{5.14}$$

encore noté :

$$\int_{z\in\mathbb{R}} f(z)\, dP_{\varphi(X,Y)}(z) = \iint_{(x,y)\in\mathbb{R}^2} f(\varphi(x,y))\, dP_X(x) dP_Y(y). \tag{5.15}$$

Preuve. Pour $B \in \mathcal{T}_{\mathbb{R}}$, on a :

$$\begin{aligned} P_{\varphi\circ(X,Y)}(B) &\overset{\text{déf}}{=} P((\varphi \circ (X,Y))^{-1}(B)) = P(((X,Y)^{-1} \circ \varphi^{-1})(B)) \\ &= P((X,Y)^{-1}(\varphi^{-1}(B)) = P_{(X,Y)}(\varphi^{-1}(B)), \end{aligned}$$

et X et Y étant indépendantes on a $P_{(X,Y)} = P_X \otimes P_Y$, cf. (5.9), d'où (5.13).

(Exemple : $P_{\varphi\circ(X,Y)}(\mathbb{R}) = (P_X \otimes P_Y)(\varphi^{-1}(\mathbb{R})) = (P_X \otimes P_Y)(\mathbb{R}^2) = P_X(\mathbb{R})P_Y(\mathbb{R}) = 1$.)

Pour passer aux fonctions, on se sert de (1.10) : (5.13) donne :

$$P_{\varphi(X,Y)}(1_B) = (P_X \otimes P_Y)(1_{\varphi^{-1}(B)}) = (P_X \otimes P_Y)(1_B \circ \varphi),$$

d'où (5.14) pour f mesurable (limite d'une combinaison linéaire $\sum_i c_i 1_{B_1}$). ▪

5.3 Dépendance et indépendance de plusieurs v.a.r.

Soient n variables aléatoires réelles $X_1, ..., X_n$ définies sur un espace probabilisé $(\Omega, \mathcal{T}, P)$, avec $n \in \mathbb{N}^*$. On note, pour des intervalles $I_i \subset \mathbb{R}$, $i = 1, ..., n$ (ou pour tous les boréliens) :

$$\begin{aligned} P(X_1 \in I_1, ..., X_n \in I_n) &\stackrel{\text{déf}}{=} P((X_1 \in I_1) \cap ... \cap (X_n \in I_n)) \\ &= P(\{\omega \in \Omega : \forall i = 1, ..., n,\ X_i(\omega) \in I_i\} \\ &\stackrel{\text{noté}}{=} P((X_1, ..., X_n) \in (I_1 \times ... \times I_n)) \\ &\stackrel{\text{noté}}{=} P_{(X_1,...,X_n)}(I_1 \times ... \times I_n) \end{aligned}$$

la probabilité que tous les résultats $(X_i \in I_i) = \{\omega : X_i(\omega) \in I_i\}$ soient simultanément réalisés.

Définition 5.8 Les variables aléatoires réelles $X_1, ..., X_n$ sont dites indépendantes ssi, pour tous les intervalles $I_i \subset \mathbb{R}$, $i = 1, ..., n$:

$$P(X_1 \in I_1, ..., X_n \in I_n) = P(X_1 \in I_1)...P(X_n \in I_n), \tag{5.16}$$

i.e. la probabilité d'être dans l'intersection est égale au produit des probabilités, i.e. ssi :

$$P_{(X_1,...,X_n)}(I_1 \times ... \times I_n) = P_{X_1}(I_1)...P_{X_n}(I_n), \tag{5.17}$$

Donc, les v.a.r. sont indépendantes ssi :

$$P_{(X_1,...,X_n)} = P_{X_1} \otimes ... \otimes P_{X_n}, \tag{5.18}$$

(Quand $n = 2$, on retrouve (5.8).)

Proposition 5.9 *Les variables aléatoires réelles $X_1, ..., X_n$ sont indépendantes ssi les événements $\{X_1 \in I_1\}$, ..., $\{X_n \in I_n\}$ sont (mutuellement) indépendants pour tout $I_1, ..., I_n \in \mathcal{T}_\mathbb{R}$ (cf. (3.33)).*

Preuve. Supposons (5.16) satisfaite (pour tous les intervalles). Notons $A_i = X_i^{-1}(I_i) \in \Omega$ (les événements $\{X_i \in I_i\}$). Il s'agit de montrer que ces événements sont mutuellement indépendants, cf (3.33), i.e. que $P(A_{i_1}, ..., A_{i_k}) = P(A_{i_1})...P(A_{i_k})$ pour toute sous-famille finie $(i_1, ..., i_k)$ de $\{1, ..., n\}$.

Quitte à renuméroter les intervalles, c'est vrai pour $k = n$ (équation (5.16)).

Prenant $I_n = \mathbb{R}$, ayant $\{X_n \in \mathbb{R}\} = \Omega$, on en déduit que $P(X_1 \in I_1, ..., X_{n-1} \in I_{n-1}) = P_{X_1}(I_1)...P_{X_{n-1}}(I_{n-1})$. Quitte à renuméroter les intervalles, cela reste vrai pour tout sous-famille finie $(i_1, ..., i_{n-1})$. Puis, idem pour toute sous-famille de taille $\leq n-2$.

La réciproque est immédiate. ∎

5.4 V.a.r. indépendantes et identiquement distribuées (i.i.d.)

[i.i.d. = independant identically distributed random sequence]

Définition 5.10 Une famille $(X_i)_{i \in I}$ de v.a.r. toutes définies sur un même espace probabilisé $(\Omega, \mathcal{T}_\Omega, P)$ est dite famille de v.a.r. indépendantes et identiquement distribuées (i.i.d.) si les X_i sont des v.a.r. indépendantes qui ont toutes la même loi : $P_{X_j} = P_{X_1}$ pour tout $j \in I$.

Exemple 5.11 Soit X_i la v.a.r. donnant le résultat du i-ème lancer d'une pièce équilibrée : $P_{X_i}(p) = \frac{1}{2} = P_{X_i}(f)$ pour tout i. La famille $(X_i)_{\mathbb{N}^*}$ est une famille de v.a.r. i.i.d..

Détails : n lancers, $\Omega = \{p, f\}^n$ l'univers, $X_i = \pi_i$ la i-ème projection usuelle, soir $X_i(\vec{\omega}) = \omega_i$ quand $\vec{\omega} = (\omega_1, ..., \omega_n)$. Alors $P_{\vec{X}} = P_{X_1} \otimes ... \otimes P_{X_n}$ où $\vec{X} =^{\text{déf}} (X_1, ..., X_n) : \Omega \to \mathbb{R}^n$. ∎

5.5 * Indépendance de tribus

On aura besoin de cette notion lors de l'étude de l'espérance conditionnelle d'une v.a.r. relativement à une autre.

5.5.1 Indépendance de v.a.r. versus indépendance d'événements

Il y a ainsi une très grosse différence dans la vérification de "n événements sont indépendants" et de "n v.a.r. sont indépendantes", voir définition 3.47 équation (3.33) :

* dans le cas des variables aléatoires, on a vu que si l'équation (5.16) est satisfaite, alors cette équation est automatiquement satisfaite pour toute sous-famille $(I_i)_{i=1,...,k}$ de taille $k \le n$ (il suffit de prendre $I_i = \mathbb{R}$ pour les $i > k$),

* alors que ce n'est pas le cas pour les événements, cf. remarque 3.50.

Considérons le cas très particulier des n v.a.r. 1_{A_1}, ..., 1_{A_n} (fonctions indicatrices avec $A_i \in \mathcal{T}_\Omega$ pour tout i), et comparons l'indépendance de ces v.a.r. 1_{A_i} et l'indépendance des événements A_i.

Proposition 5.12 *Les n événements A_i sont indépendants ssi les n v.a.r. 1_{A_i} sont indépendantes.*

Preuve. Pour $d\in[1,n]_\mathbb{N}$ et pour $I_1,...,I_d\in\mathcal{T}_\mathbb{R}$, on s'intéresse aux probabilités des ensembles :

$$A_1 \cap ... \cap A_d \qquad \text{et} \qquad (1_{A_1}\in I_1, ..., 1_{A_d}\in I_d). \tag{5.19}$$

$\Rightarrow$: supposons les n événements A_i indépendants ; donc les A_i^C sont indépendants, ainsi que toutes les combinaisons A_i, A_j^C. On a $1_{A_i}^{-1}(I_i)\in\{\emptyset, A_i, A_i^C, \Omega\}$ (uniquement 4 possibilités). Et on a $P(1_{A_j}\in I_j) = P(\{\omega : 1_{A_j}(\omega) \in I_j\} = \emptyset, A_j, A_j^C, \mathbb{R}$ suivant les cas $0, 1\notin I_j$, $1\in I_j$ et $0\notin I_j$, $0\in I_j$ et $1\notin I_j$, $0, 1\in I_j$. Donc $P(1_{A_1}\in I_1, ..., 1_{A_n}\in I_n) = P(1_{A_1}^{-1}(I_1) \cap ... \cap 1_{A_n}^{-1}(I_n)) = P(1_{A_1}^{-1}(I_1))...P(1_{A_n}^{-1}(I_n)) = P(1_{A_1}\in I_1)...P(1_{A_n}\in I_n)$: les v.a.r. sont indépendantes.

$\Leftarrow$: supposons les n v.a.r. $(1_{A_i})_{i=1,...,n}$ indépendantes. On a $1_{A_i}^{-1}(\{1\}) = A_i$ pour tout i. Donc $P(A_{i_1})...P(A_{i_d}) = P(1_{A_{i_1}}^{-1}(\{1\}))...P(1_{A_{i_d}}^{-1}(\{1\})) = P(1_{A_{i_1}}^{-1}(\{1\})\cap...\cap 1_{A_{i_d}}^{-1}(\{1\})) = P(A_{i_1}\cap...\cap A_{i_d})$, ce pour tout injection $i : [1,d]_\mathbb{N} \to [1,n]_\mathbb{N}$: on a vérifié (3.33). ▪

5.5.2 Indépendance de tribus et indépendance de v.a.r.

On a vu que $\mathcal{T}(1_A) = \{\emptyset, A, A^C, \mathbb{R}\}$ est la tribu engendrée par A, i.e. la tribu obtenue à partir de la fonction 1_A, cf. exemple 2.29 et exercice 2.32.

Définition 5.13 Dans l'espace probabilisé $(\Omega, \mathcal{T}, P)$, une famille de sous-tribus $(\mathcal{T}_i)_{i=1,...,n}$ de $\mathcal{T}$ est dite indépendante ssi pour tout $A_i \in \mathcal{T}_i$, $i = 1, ..., n$, on a :

$$P(A_1 \cap ... \cap A_n) = P(A_1)...P(A_n). \tag{5.20}$$

La proposition 5.12 précédente s'énonce donc également :

Proposition 5.14 *Les v.a.r. 1_{A_i} sont indépendantes ssi les tribus $\mathcal{T}(1_{A_i})$ sont indépendantes.*

Plus généralement :

Proposition 5.15 *n v.a.r. X_i sont indépendantes ssi les n tribus $\mathcal{T}(X_i)$ sont indépendantes.*

L'étude de l'indépendance des fonctions X_i peut ainsi être ramenée à l'étude d'indépendance de familles d'ensembles.

Preuve. C'est vrai pour les fonctions indicatrices, et une v.a.r. est une limite d'une suite de fonctions étagées (somme de fonctions indicatrices), voir cours d'intégration, et on dispose de la proposition 5.12 et de (1.5) page 7. ▪

6 Exemples de lois

6.1 Loi uniforme

Sur un ensemble fini Ω c'est la loi constante :

$$P(\omega) = \frac{1}{|\Omega|}, \tag{6.1}$$

donc $P = \frac{1}{|\Omega|} \sum_{\omega\in\Omega} \delta_\omega$.

Cas $\Omega = [a,b]$ intervalle fini de $\mathbb{R}$: c'est la loi de densité constante, pour tout $\omega \in \Omega$:

$$p(\omega) = \frac{1}{|b-a|}, \tag{6.2}$$

donc $P(A) = \int_a^b 1_A(t)\, \frac{1}{|b-a|} dt$, pour $A \in \mathcal{T}_{\mathbb{R}}$ la tribu borélienne.

Exercice 6.1 Une urne contient n jetons numérotés de 1 à n indifférenciables au touché. On en extrait p. On note X_i la v.a.r. qui donne le numéro du i-ème jeton extrait, $1 \le i \le p$. On pose $S = X_1 + ... + X_p$. Les tirages se font <u>avec</u> remise.

11. Donner l'univers et son cardinal.
12. Montrer que la variable X_i, suit une loi uniforme à définir.
13. Montrer que les variables X_i et X_j sont indépendantes pour $i \neq j$.
14. Calculer $E(X_i)$, $E(S)$, $E(X_i^2)$, $V(X)$ et $V(S)$ (définitions plus loin...).

Réponse. 11. Soit $J = \{j_1, ..., j_n\}$ l'ensemble des jetons, le jeton j_i correspondant au numéro i. L'univers est $\Omega = [1,n]^p$ (ensemble des résultats des tirages) et $|\Omega| = n^p$. La v.a.r. $X_i : \Omega \to \mathbb{R}$ est donnée par $X_i(\vec{\omega}) = \omega_i$, où $\vec{\omega} = (\omega_1, ..., \omega_p) \in [1,n]^p$ est le résultat d'un tirage. Et $n = |\mathrm{Im} X_i|$ cardinal de l'image de X_i.

12. L'hypothèse d'indifférentiabilité au touché (une chance sur n au i-ème tirage) donne, pour $i = 1, ..., p$ et $k = 1, ..., n$:

$$P_{X_i}(k) = \frac{1}{|\mathrm{Im} X_i|} = \frac{1}{n} \qquad (= P(X_i = k) = P(\omega_i = k) = P((X_i^{-1}(k))).$$

Résultat indépendant de k : donc la loi P_{X_i} de X_i est la loi uniforme $\frac{1}{n}$. Ou encore, pour appliquer (6.1) :

$$P_{X_i}(k) = P(\{\vec{\omega} \text{ t.q. } \omega_i = k\}) = \frac{1 * n^{p-1}}{P(\Omega)} = \frac{1}{n}, \qquad k = 1, ..., n,$$

car 1 seule possibilité pour ω_i et n possibilités pour les ω_j où $j \neq i$.

13. Pour $j \neq i$ et $h, k \in [1,n]_{\mathbb{N}}$, on a :

$$P((X_i{=}k) \cap (X_j{=}h)) = P(\{\omega : \omega_i = k \text{ et } \omega_j = h\} = \frac{n^{p-2}}{P(\Omega)} = \frac{1}{n^2} = P(X_i{=}k) P(X_j = h).$$

Donc les variables sont indépendantes.

14. Soit $i \in [1,n]_{\mathbb{N}}$ fixé.

$$E(X_i) = \sum_{x_j \in [1,n]_{\mathbb{N}}} x_j P_{X_i}(x_j) = \sum_{j=1}^n j P_{X_i}(j) = \sum_{j=1}^n j \frac{1}{n} = \frac{1}{n} \frac{n(n+1)}{2} = \frac{(n+1)}{2}.$$

$$E(S) = E(X_1) + ... + E(X_p) = \frac{p(n+1)}{2}.$$

$$E(X_i^2) = \sum_{x_j \in [1,n]_{\mathbb{N}}} x_j^2 P_{X_i}(x_j) = \sum_{j=1}^n j^2 \frac{1}{n} = \frac{1}{n} \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.$$

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

$$V(S) = \sum_{i=1}^p V(X_i) - \sum_{\substack{i,j=1,\dots,p \\ i \neq j}} \mathrm{cov}(X_i, X_j) = p \frac{n^2-1}{12} - 0,$$

car $\mathrm{cov}(X_i, X_j) = 0$ quand $i \neq j$, les variables aléatoires étant indépendantes. ∎

Exercice 6.2 Même énoncé qu'à l'exercice précédent 6.1, sauf qu'on suppose maintenant les tirages <u>sans</u> remise (donc ici $p \le n$).

21. Donner l'univers et son cardinal.
22. Montrer que la variable X_i suit une loi uniforme à définir.
23. Vérifier que, pour $h, k \in [1,n]_{\mathbb{N}}$ et $i \neq j$, $P((X_i{=}k) \cap (X_j{=}h)) = \begin{cases} 0 & \text{si } k = h, \\ \dfrac{1}{n(n-1)} & \text{si } k \neq h \end{cases}$.

Les variables X_i et X_j sont-elles indépendantes ?

24. Calculer $E(X_i)$ et $V(X_i)$.
25. Calculer $\mathrm{cov}(X_i, X_j)$ et $V(S)$.

Réponse. 21. Ω est l'ensemble des résultats des tirages sans remise, card$\Omega = A_n^p = \frac{n!}{(n-p)!}$.

22. Pour $k \in [1,n]_{\mathbb{N}}$:

$$P(\{\omega : X_i(\omega) = k\} = \frac{A_1^1 A_{p-1}^{n-1}}{A_p^n} = \frac{1 * \frac{(n-1)!}{(n-p)!}}{\frac{n!}{(n-p)!}} = \frac{1}{n} = P_{X_i}(k).$$

23. Pour $h = k$, le tirage étant sans remise, on ne peut pas avoir $X_i = k$ en même temps que $X_j = k$, donc $P((X_i{=}k) \cap (X_j{=}k)) = P(\emptyset) = 0 \neq P(X_i{=}k)P(X_j{=}k)$. Donc les v.a. ne sont pas indépendantes.

Pour $h \neq k$ on a :

$$P((X_i{=}k) \cap (X_j{=}h)) = \frac{A_{p-2}^{n-2}}{A_p^n} = \frac{\frac{(n-2)!}{(n-p)!}}{\frac{n!}{(n-p)!}} = \frac{1}{n(n-1)} \qquad (\neq \frac{1}{n^2}).$$

(On vérifie de nouveau que les v.a. ne sont pas indépendantes.)

24. Même calculs que précédemment :

$$E(X_i) = \frac{n+1}{2}, \qquad E(X_i^2) = \frac{(n+1)(2n+1)}{6}, \qquad V(X_i) = \frac{n^2-1}{12}.$$

25. On a $\mathrm{cov}(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j)$, avec, pour $i \neq j$:

$$E(X_iX_j) = \sum_{h,k=1,\dots,n} hkP((X_i{=}k) \cap (X_j{=}h)) = \frac{1}{n(n-1)} \sum_{\substack{h,k=1,\dots,n \\ h \neq k}} hk = \frac{1}{n(n-1)} \sum_{k=1}^{n} k(\sum_{\substack{h=1,\dots,n \\ h \neq k}} h),$$

avec $\displaystyle\sum_{\substack{h=1,\dots,n \\ h \neq k}} h = (\sum_{h=1,\dots,n} h) - k = \frac{n(n+1)}{2} - k$. D'où :

$$\begin{aligned} E(X_iX_j) &= \frac{1}{n(n-1)} (\frac{n(n+1)}{2} \sum_{k=1}^{n} k - \sum_{k=1}^{n} k^2) = \frac{1}{n(n-1)} (\frac{n(n+1)}{2} \frac{n(n+1)}{2} - \frac{n(n+1)(2n+1)}{6}) \\ &= \frac{n+1}{n-1} \frac{3n^2+3n-4n-2}{12} = \frac{n+1}{n-1} \frac{3n^2-n-2}{12} = \frac{n+1}{n-1} \frac{(n-1)(3n+2)}{12} = \frac{(n+1)(3n+2)}{12}. \end{aligned}$$

D'où, pour $i \neq j$:

$$\begin{aligned} \mathrm{cov}(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j) &= \frac{(n+1)(3n+2)}{12} - \frac{(n+1)^2}{4} \\ &= \frac{(n+1)(3n+2-3n-3)}{12} = -\frac{n+1}{12} \quad (\neq 0). \end{aligned}$$

D'où :

$$V(S) = \sum_{i=1}^{p} V(X_i) + \sum_{\substack{i,j=1,\dots,p \\ i \neq j}} \mathrm{cov}(X_i, X_j) = p\frac{(n+1)(n-1)}{12} - p(p-1)\frac{n+1}{12} = \frac{p(n+1)(n-p)}{12},$$

résultat différent du précédent. ▟

6.2 Epreuve de Bernoulli et loi de Bernoulli

Correspond à un lancer de pile ou face pour une pièce équilibrée ou non.

Définition 6.3 Une épreuve de Bernoulli est une expérience à deux issues (distinctes) dites succès et échec.

Soit a et b les issues d'une épreuve de Bernoulli, a correspondant au succès et b à l'échec, $a \neq b$. Ici $\Omega = \{a, b\}$ est l'univers des issues.

Définition 6.4 Soit $p \in]0, 1[$. La probabilité $P : \Omega \to \mathbb{R}$ définie par :

$$P(a) = p, \qquad P(b) = 1-p \overset{\text{noté}}{=} q. \tag{6.3}$$

est appelée loi de Bernoulli. (Il est immédiat que $P = p\delta_a + q\delta_b$ est une probabilité sur Ω.)

Soit $X : \Omega \to \mathbb{R}$ la v.a.r. définie par, avec $\alpha \neq \beta$:

$$X(a) = \alpha, \qquad X(b) = \beta, \tag{6.4}$$

α (resp. β) étant le "gain" correspondant au résultat a (resp. b).

Définition 6.5 L'ensemble $E = \{\alpha, \beta\}$ est appelé l'ensemble des états.

La loi de Bernoulli s'exprime au travers de la v.a.r. X par :

$$P_X(\alpha) = p \quad (= P(X{=}\alpha)) \qquad \text{et} \qquad P_X(\beta) = q \quad (= P(X{=}\beta)). \tag{6.5}$$

Définition 6.6 Cas particulier $\alpha = 1$ (gain en cas de succès) et $\beta = 0$ (gain en cas d'échec) :

$$P_X(1) = p, \qquad P_X(0) = 1{-}p \overset{\text{noté}}{=} q. \tag{6.6}$$

Ce cas particulier est appelé "l'épreuve de Bernoulli".

6.3 Processus de Bernoulli

On renouvelle la même expérience de Bernoulli r fois, où $r \in \mathbb{N}^*$ (éventuellement infini), chaque expérience donnant pour résultat soit 1 (dit succès) soit 0 (dit échec). L'univers des résultats (des r tirages avec remise) est $\Omega = \{0,1\}^r$ (ou bien $\{0,1\}^{\mathbb{N}^*}$). Pour $i \in [1,r]_{\mathbb{N}}$ (ou bien pour $i \in \mathbb{N}^*$) on note :

$$X_i : \begin{cases} \Omega & \to \{0,1\} \subset \mathbb{R} \\ \vec{\omega} = (\omega_1, \omega_2, ...) & \to X_i(\vec{\omega}) = \omega_i = \begin{cases} 1 & \text{si } \omega_i = 1, \\ 0 & \text{si } \omega_i = 0, \end{cases} \end{cases} \tag{6.7}$$

la v.a.r. donnant le résultat du i-ème tirage (l'opérateur de projection sur la i-ème composante). Ici l'ensemble des états est $E = \{0,1\}$.

Exemple 6.7 $r = 4$: le 4-uplet $\vec{\omega} = (1,0,1,1) \in \Omega$ est une issue (un vecteur à 4 composantes), et $X_2(\vec{\omega}) = 0$, $X_3(\vec{\omega}) = 1$...

On suppose que la suite $(X_i)_{i=1,...,r}$ (ou bien $(X_i)_{i\in\mathbb{N}^*}$) est i.i.d., i.e. est une suite de v.a.r. indépendantes identiquement distribuées, i.e. indépendantes toutes de même loi : il existe $p \in]0,1[$ t.q. pour tout $i = 1, ..., r$ (ou bien $i \in \mathbb{N}^*$) :

$$P_{X_i}(1) = p = P(\{\vec{\omega} : \omega_i = 1\}) \qquad \text{et} \qquad P_{X_i}(0) = 1{-}p \overset{\text{noté}}{=} q = P(\{\vec{\omega} : \omega_i = 0\}), \tag{6.8}$$

avec $p =$ probabilité dite de succès, et $q = 1{-}p =$ probabilité dite d'échecs.

Définition 6.8 Soit $I = [1,r]_{\mathbb{N}}$ ou $I = \mathbb{N}^*$. Le processus $(X_i)_{i\in I}$ est alors appelé processus de Bernoulli de paramètre p.

6.4 Loi binomiale et tirages avec remise

But : compter le nombre de succès d'un processus de r épreuves de Bernoulli d'issues a et b (par exemple $a =$ pile et $b =$ face).

On note π_i la projection sur la i-ème composante :

$$\pi_i : \begin{cases} \mathbb{R}^r & \to \mathbb{R} \\ \vec{x} = (x_1, ..., x_r) & \to \pi_i(\vec{x}) = x_i. \end{cases} \tag{6.9}$$

Sur $\Omega = \{a,b\}^r$ on définit :

$$X_i = 1_{\{a\}} \circ \pi_i, \qquad \text{donc} \qquad X_i(\vec{\omega}) = 1_{\{a\}}(\omega_i) = \begin{Bmatrix} = 1 \text{ si } \omega_i = a, \\ = 0 \text{ si } \omega_i = b, \end{Bmatrix} \tag{6.10}$$

et :

$$S_r : \begin{cases} \Omega & \to \mathbb{N} \\ \vec{\omega} & \mapsto S_r(\vec{\omega}) \overset{\text{déf}}{=} \displaystyle\sum_{i=1}^{r} X_i(\vec{\omega}) = \text{le nombre de succès,} \end{cases} \tag{6.11}$$

on dit aussi que $S_r(\vec{\omega})$ est le "nombre de visites en a" de l'issue $\vec{\omega}$. En particulier $S_r(\vec{\omega}) \in [0,r]_{\mathbb{N}}$.

Exemple 6.9 Pour $r=4$ et $\vec{\omega}=(1,0,1,1)$, on a $S_4(\vec{\omega})=3$. ♣

Exemple 6.10 On lance r fois une même pièce : $a=$ pile et $b=$ face. ♣

Exemple 6.11 Émetteur de particules, avec a la zone cible et b la zone hors cible. Et S_r donne le nombre de particules, parmi r, ayant atteint la cible ♣

Exemple 6.12 Contrôle qualité : on extrait r objets d'une production donnée, avec a si l'objet est bon et b si l'objet est défectueux. Et S_r donne le nombre de bons produits parmi r. ♣

On veut calculer la probabilité $P_{S_r}(k)=P(S_r{=}k)$ d'avoir k succès après r tirages.

Proposition 6.13 et définition. *Si $p\in[0,1]$ est la probabilité d'obtenir a quand on fait une expérience (un tirage), alors la loi de S_r est donnée par, pour $k\in[0,r]_{\mathbb{N}}$:*

$$P_{S_r}(k)=C_r^k p^k(1-p)^{r-k} \quad \overset{\text{noté}}{=} \quad b(k;r,p). \tag{6.12}$$

La loi de S_r est notée :

$$P_{S_r}=b(r,p) \tag{6.13}$$

et appelée loi binomiale de paramètres r,p. Ainsi $b(k;r,p)=b(r,p)(k)$ est la valeur de la loi binomiale pour k succès.

En particulier $P_{S_1}=b(1,p)$ est la loi de Bernoulli.

Autrement dit, si $q=1-p$ (probabilité d'obtenir b pour un tirage), on a :

$$b(k;r,p)=C_r^k p^k q^{r-k}.$$

Ce sont les termes du développement du binôme $(p+q)^r$, d'où le nom de cette loi.

Preuve. La loi de S_r est une loi discrète à valeurs dans $[0,r]_{\mathbb{N}}$, et donc entièrement déterminée si on connaît les $P_{S_r}(\{k\})=^{\text{noté}} P_{S_r}(k)$ pour tout $k\in[0,r]_{\mathbb{N}}$.

C_r^k est le nombre de possibilités pour que dans $\vec{\omega}\in\{a,b\}^r$ on ait k fois a. Et pour chacune de ces possibilités, la probabilité est p^kq^{r-k} (on a k fois a donc $r-k$ fois b). ♣

Exemple 6.14 Cas r lancers d'une pièce équilibrée ($p{=}\frac12$) : on a $b(k;r,\dfrac12)=\dfrac{1}{2^r}\dfrac{r!}{k!(r-k)!}$ En particulier $b(r;r,\dfrac12)=\dfrac{1}{2^r}$, probabilité d'obtenir r piles de suite. ♣

6.4.1 Application aux statistiques : échantillons de taille r avec replacement.

On considère le cas d'une population $S=\{a_1,...,a_n\}$ de taille $|S|=n$. On suppose $S=S_1\cup S_2$ avec $S_1\cap S_2=\emptyset$ (la population est partitionnée en deux sous-populations distinctes S_1 et S_2). On note $n_1=|S_1|$, et on note :

$$p_1=\frac{n_1}{n} \tag{6.14}$$

la proportion de la population qui se trouve dans S_1, i.e., la probabilité pour qu'un tirage équiprobable donne un élément de S_1.

On s'intéresse aux "échantillons de taille r **avec** replacement" (cf. définition 2.68 page 30) : notre univers est $\Omega=S^r$, et $\vec{\omega}=(\omega_1,...,\omega_r)\in\Omega$ est un résultat possible (une suite de r tirages), la valeur ω_i étant l'élément tiré au i-ème tirage, avec donc ω_i soit dans S_1 soit dans S_2.

On définit la variable aléatoire :

$$X_1:\begin{cases}\Omega & \to \mathbb{N}\\ \vec{\omega} & \mapsto X_1(\vec{\omega})=\text{le nombre de composantes de }\vec{\omega}\text{ qui sont dans }S_1,\end{cases} \tag{6.15}$$

i.e. $X_1(\vec{\omega})$ est le nombre de personnes appartenant à S_1 dans l'échantillon $\vec{\omega}$.

De même on définit X_2 relatif à S_2, avec donc :

$$X_1(\vec{\omega})+X_2(\vec{\omega})=r, \tag{6.16}$$

pour tout ω.

Proposition 6.15 *La loi de X_1 est la loi binomiale $b(r,p_1)$, cf. (6.13) :*

$$P_{X_1} = b(r,p_1). \tag{6.17}$$

Autement dit : pour $0 \le k \le r$:

$$P_{X_1}(k) = b(k;r,p_1) \qquad (= C_r^k p_1^k (1-p_1)^{r-k}). \tag{6.18}$$

Idem, la loi de X_2 est $P_{X_2} = b(r,p_2)$ où $p_2 = 1-p_1$ est la proportion de la population qui se trouve dans S_2.

Preuve. Soit $\vec\omega \in S^r$; $\vec\omega = (\omega_i)_{i=1,...r}$ est une suite finie (ordonnée).

L'événement $(X_1{=}k) = \{\vec\omega \in S^r : X_1(\vec\omega) = k\}$ est l'ensemble des suites $\vec\omega$ dont k termes sont dans S_1 : son cardinal est le nombre de combinaison C_r^k, cf. paragraphe 1.4.2. Et la probabilité de tirer un élément $\vec\omega$ dans $(X_1{=}k)$ est $p_1^k(1-p_1)^{r-k}$. Donc $P_{X_1}(k) = C_r^k p_1^k (1-p_1)^{r-k} = b(k;r,p_1)$. ▪

Exercice 6.16 Montrer que les lois X_1 et X_2 ne sont pas indépendantes.

Réponse. C'est évident avec le lien de dépendance (6.16). Calcul : si $k,\ell \in [0,r]_{\mathbb{N}}$ et $k+\ell \ne r$, alors $P(X_1{=}k, X_2{=}\ell) = 0$, alors que $P(X_1{=}k)$ et $P(X_2{=}\ell)$ sont non nuls, cf. (6.18). ▪

6.5 Loi multinomiale

Plutôt que de lancer r fois une pièce à 2 côtés, on lance r fois un objet à ℓ côtés, $\ell \ge 2$. Soit :

$$V = \{a_1, ..., a_\ell\} \tag{6.19}$$

l'ensemble des ℓ valeurs (résultats) possibles. (Le cas $\ell = 2$ vient d'être traité.)

Exemple 6.17 On lance r fois un dés à $\ell{=}6$ faces, et $V = [1,6]_{\mathbb{N}}$. ▪

Soit r tirages <u>avec</u> replacement : l'univers est :

$$\Omega = V^r, \tag{6.20}$$

et un $\vec\omega \in \Omega$ (un événement) est de la forme $\vec\omega = (\omega_1, ..., \omega_r)$, où $\omega_i \in V$ pour tout $i = 1, ..., r$.

Soit p_i la probabilité d'apparition de a_i (après un lancer) pour $i = 1, ..., \ell$. On note :

$$\vec p = (p_1, ..., p_\ell), \qquad p_1{+}...{+}p_\ell = 1. \tag{6.21}$$

Soient les v.a.r., pour $i = 1, ..., \ell$:

$$X_i : \begin{cases} V^r & \to \mathbb{N} \\ \vec\omega & \mapsto X_i(\vec\omega) = \text{le nombre de composantes de } \vec\omega \text{ qui valent } a_i. \end{cases}$$

On dit aussi "le nombre d'éléments de la suite $\vec\omega$" qui valent a_i.

Exemple : pour $\vec\omega = (2,5,1,5,6,3)$ on a $X_4(\vec\omega) = 0$ et $X_5(\vec\omega) = 2$.

En particulier (longueur de la suite $\vec\omega$) :

$$X_1(\vec\omega){+}...{+}X_\ell(\vec\omega) = r. \tag{6.22}$$

On notera :

$$\vec X = (X_1, ..., X_\ell) : \begin{cases} \Omega & \to \mathbb{N}^\ell \\ \vec\omega & \to \vec X(\omega) = (X_1(\vec\omega), ..., X_\ell(\vec\omega)) \overset{\text{noté}}{=} \begin{pmatrix} X_1(\vec\omega) \\ \vdots \\ X_\ell(\vec\omega) \end{pmatrix} \end{cases} \tag{6.23}$$

le "vecteur aléatoire" de composantes les X_i (v.a.r.). On dispose de la notation :

$$\begin{aligned} P(\vec X = (k_1, ..., k_\ell)) &= P((X_1{=}k_1) \cap ... \cap (X_\ell{=}k_\ell)) \\ &\overset{\text{noté}}{=} P_{\vec X}(k_1, ..., k_\ell), \end{aligned} \tag{6.24}$$

probabilité d'obtenir un tirage contenant exactement k_i fois a_i pour $i = 1, ..., \ell$.

Proposition 6.18 et définition. *Pour $\vec{k} = (k_1, ..., k_\ell) \in \mathbb{N}^\ell$, quand $k_1 + ... + k_\ell = r$, la loi de $\vec{X}$ est donnée par :*

$$P_{\vec{X}}(k_1, ..., k_\ell) = \frac{r!}{k_1!...k_\ell!} p_1^{k_1} ... p_\ell^{k_\ell} \quad \overset{\text{noté}}{=} \quad b(\vec{k}; r, \vec{p}). \tag{6.25}$$

La loi de X est notée :

$$P_{\vec{X}} = b(r, \vec{p}), \tag{6.26}$$

appelée la loi multinomiale de paramètres $r, \vec{p}$. Ainsi $P_{\vec{X}}(\vec{k}) = b(r, \vec{p})(\vec{k}) = b(\vec{k}; r, \vec{p})$ est la valeur de la loi pour le résultat $\vec{k}$.

Preuve. On applique (1.28). ∎

Exemple 6.19 On considère le cas d'une population S de taille $|S| = n$, et on s'intéresse aux échantillons de taille r **avec** replacement.

On suppose la population divisée en ℓ sous-populations distinctes S_1,...,S_ℓ (par exemple en "classes d'age"), et on note $p_i = \frac{|S_i|}{|S|} \in [0, 1]$ la proportion de la population qui se trouve dans S_i.

On note $\vec{\omega} = (\omega_1, ..., \omega_n) \in S^r$ un résultat de r tirages. On définit les variables aléatoires, pour $i = 1, ..., \ell$:

$$X_i : \begin{cases} \Omega & \to \mathbb{N} \\ \vec{\omega} & \mapsto X_i(\vec{\omega}) = \text{le nombre de composantes de } \vec{\omega} \text{ qui sont dans } S_i. \end{cases}$$

Soit $\vec{X}(\vec{\omega})$ le vecteur de composantes les $X_i(\vec{\omega})$. Alors la loi de $\vec{X}$ est la loi multinomiale $b(r, \vec{p})$, cf. (6.26), donnée par (6.25). ∎

Exercice 6.20 Loi trinomiale (cas (6.25) avec $\ell = 3$). Montrer lorsque $a_2 = a_3$ (ces résultats sont indistinguables) que la loi trinomiale se réduit à la loi binomiale $b(r, p)$ où $p = p_1$ (et $q = 1-p_1$).

Réponse. Il n'y a que deux issues possibles. On pose $p = p_1$ et $q = p_2+p_3 = 1-p$. On pose $Y(\vec{\omega}) = X_1(\omega)$ = le nombre de composantes de $\vec{\omega}$ qui valent a_1, $P(Y = k_1) = b(k_1; r, p) = \frac{r!}{k_1!(r-k_1)!} p^{k_1} q^{r-k_1}$. Vérifions que :

$$P(Y = k_1) = \sum_{k_2+k_3=r-k_1} P(X_1{=}k_1, (X_2, X_3){=}(k_2, k_3)).$$

Avec (6.25), pour k_1 fixé, on a :

$$\sum_{k_2+k_3=r-k_1} P(X_1{=}k_1, (X_2, X_3){=}(k_2, k_3)) = p_1^{k_1} \frac{r!}{k_1!} \sum_{k_2=0}^{r-k_1} \frac{1}{k_2!(r-k_1-k_2)!} p_2^{k_2} p_3^{r-k_1-k_2}.$$

Il faut donc vérifier :

$$\begin{aligned} \frac{1}{(r-k_1)!}(p_2+p_3)^{r-k_1} &= \sum_{k_2=0}^{r-k_1} \frac{1}{k_2!(r-k_1-k_2)!} p_2^{k_2} p_3^{r-k_1-k_2} \\ &= \frac{1}{(r-k_1)!} \sum_{k_2=0}^{r-k_1} \frac{(r-k_1)!}{k_2!((r-k_1)-k_2)!} p_2^{k_2} p_3^{r-k_1-k_2}, \end{aligned}$$

OK. ∎

6.6 Loi hypergéométrique : tirages sans remise, avec ou sans ordre

6.6.1 Cas de 2 issues

C'est la loi correspondant aux tirages sans remise où on tient ou pas compte de l'ordre, cf. (2.28) et (2.29) : ce sera la même loi dans les deux cas.

Cadre : soit S une population de cardinal $n \geq 2$.

Soit $S = S_1 \cup S_2$ une partition de S. On note :

$$|S| = n, \qquad |S_1| = n_1, \qquad |S_2| = n_2 = n - n_1, \qquad \vec{n} = (n_1, n_2). \tag{6.27}$$

Exemple 6.21 S est un ensemble de pièces, S_1 est l'ensemble des pièces défectueuses.

S est l'alphabet, S_1 est le nombre de voyelles. ∎

On suppose les tirages équiprobables. On note :

$$p_1 = \frac{n_1}{n}, \qquad p_2 = \frac{n_2}{n} = 1 - p_1, \qquad \vec{p} = (p_1, p_2). \tag{6.28}$$

Soit $\Omega \subset S^r$, où Ω est l'univers des tirages de taille $r \in [1, n]_{\mathbb{N}}$ sans remise, ordonnés ou non. Donc un échantillon est un élément $\vec{\omega} = (\omega_1, ..., \omega_r)$ si on tient compte de l'ordre, et un élément $\vec{\omega} = \{\omega_1, ..., \omega_r\}$ si on ne tient pas compte de l'ordre, où $\omega_i \in S_1 \cup S_2$ et $\omega_i \neq \omega_j$ pour tout $i \neq j$.

On définit les variables aléatoires, pour $i = 1, 2$:

$$X_i : \begin{cases} \Omega & \to \mathbb{N} \\ \vec{\omega} & \mapsto X_i(\vec{\omega}) \overset{\text{déf}}{=} \text{le nombre de composantes de } \vec{\omega} \text{ qui sont dans } S_i. \end{cases} \tag{6.29}$$

Exemple 6.22 X_1 compte le nombre de pièces défectueuses dans l'échantillon de taille r, et donc X_2 compte le nombre de pièces correctes.

X_1 compte le nombre de voyelles dans un mot de r lettres.

But : Connaître la probabilité pour que $X_1(\vec{\omega}) = k_1$, pour $k_1 \in [0, n_1]_{\mathbb{N}}$, i.e. on veut connaître $P_{X_1}(k_1)$, i.e. on veut connaître la loi P_{X_1} de X_1. (Remarque : si $k_1 > n_1$ alors $P_{X_1}(k_1) = 0$.)

Exemple 6.23 Connaître la probabilité qu'il y ait k_1 pièces défectueuses dans l'échantillon $\vec{\omega}$.

Connaître la probabilité qu'il y ait k_1 voyelles dans le mot $\vec{\omega}$.

Remarque : on s'attend (intuitivement) à ce que la probabilité « $\vec{\omega}$ contient k_1 fois un élément de S_1 » ne dépend pas du fait qu'on tienne compte, ou pas, de l'ordre d'apparition des k_1 éléments : ce qui compte c'est qu'on en ait obtenu k_1.

Proposition 6.24 et définition. *Soit $k_1 \in [0, n_1]_{\mathbb{N}}$, et soit $k_2 = r - k_1$. On suppose $k_2 \in [0, n_2]_{\mathbb{N}}$. On note $\vec{k} = (k_1, k_2)$. La loi de X_1 est donnée par :*

$$P_{X_1}(k_1) = \frac{C_{n_1}^{k_1} C_{n_2}^{k_2}}{C_n^r} \overset{\text{noté}}{=} H(\vec{k}; \vec{n}, r, \vec{p}), \tag{6.30}$$

et $H(\vec{n}, r, \vec{p})$ est appelée la loi hypergéométrique de paramètres $\vec{n} = (n_1, n_2)$, r et $\vec{p} = (p_1, p_2)$, quand $n = n_1 + n_2$.

Et $P_{X_1}(k_1) = H(\vec{n}, r, \vec{p})(\vec{k}) =^{\text{noté}} H(\vec{k}; \vec{n}, r, \vec{p})$, aussi égal à $P_{X_2}(k_2)$.

Preuve. r tirages sans remise dans une population de taille n.

1- L'ordre dans l'échantillon est pris en compte. Le nombre total d'arrangements dans la population S est A_n^r. Le nombre d'arrangements possibles dans la population S_i est $A_{n_i}^{k_i}$ (car k_i tirages dans une population de taille n_i) . Et il y a $C_r^{k_1}$ combinaisons possibles de distributions de k_1 éléments parmi r. D'où :

$$P(X_1 = k_1) = C_r^{k_1} \frac{A_{n_1}^{k_1} A_{n_2}^{k_2}}{A_n^r} = \frac{r!}{k_1! k_2!} \frac{n_1!}{(n_1 - k_1)!} \frac{n_2!}{(n_2 - k_2)!} \frac{(n-r)!}{n!} = \frac{C_{n_1}^{k_1} C_{n_2}^{k_2}}{C_n^r}.$$

2- L'ordre dans l'échantillon n'est pas pris en compte. Le nombre total de combinaisons est C_n^r. Le nombre de combinaisons possibles pour les éléments de S_i est $C_{n_i}^{k_i}$. D'où :

$$P(X_1 = k_1) = \frac{C_{n_1}^{k_1} C_{n_2}^{k_2}}{C_n^r}.$$

Exemple 6.25 Population $S = \{R_1, R_2, N\} = S_1 \cup S_2$, où $S_1 = \{R_1, R_2\}$ contient les deux boules rouges et $S_2 = \{N\}$ la boule noire. On extrait successivement 2 boules (tirage sans remise), et on s'intéresse à $P(X_1 = 1)$.

1- Cas où on tient compte de l'ordre. L'univers est
$\Omega = \{(R_1, R_2), (R_1, N), (R_2, R_1), (R_2, N), (N, R_1), (N, R_2)\}$ de cardinal $A_3^2 = 6$,
et $P(X_1 = 1) = P(\{(R_1, N), (R_2, N), (N, R_1), (N, R_2)\}) = \frac{4}{6} = \frac{2}{3}$. On vérifie que $\frac{C_2^1 C_1^1}{C_3^2} = \frac{2}{3}$.

2- Cas où on ne tient pas compte de l'ordre. L'univers est
$\Omega = \{\{R_1, R_2\}, \{R_1, N\}, \{R_2, N\}\}$ de cardinal $C_3^2 = 3$,
et $P(X_1 = 1) = P(\{\{R_1, N\}, \{R_2, N\}\}) = \frac{2}{3}$.

Exemple 6.26 Jeu de cartes $S = \{s_1, ..., s_{52}\}$ de 52 cartes, 4 joueurs et 13 cartes distribuées par joueur (= une main). On veut connaître la probabilité d'avoir 0, 1, 2, 3 ou 4 as dans une main. Ici S = les 52 cartes, $n = |S| = 52$, S_1 = les 4 as, $n_1 = 4$, $\vec{n} = (4, 48)$. Une main est notée $\vec{\omega} = (\omega_1, ..., \omega_{13})$ (13 cartes). (L'univers Ω des mains possibles a pour cardinal $|\Omega| = C_{52}^{13}$.) Donc, pour $k_1 \in [0,4]$, on a $P(X_1 = k_1) = H(\vec{n}, 13, \vec{k}) = \frac{C_4^k C_{48}^{13-k}}{C_{52}^{13}} = \frac{4!48!13!39!}{k!(4-k)!(13-k)!(35+k)!52!}$. Donc $P(X_1 = 0) \simeq 0.30$, $P(X_1 = 1) \simeq 0.44$, $P(X_1 = 2) \simeq 0.21$, $P(X_1 = 3) \simeq 0.04$, $P(X_1 = 4) \simeq 0.003$.

Même résultat si on prend en compte l'ordre des cartes dans une main. ▪

Proposition 6.27 *(Convergence de la loi hypergéométrique vers la loi binomiale.)*

Quand on fait tendre $n \to \infty$ (la taille de la population est infiniment grande), si le rapport :

$$p_1 = \frac{n_1}{n} = constant \ \in]0, 1[\tag{6.31}$$

(le pourcentage de pièces défectueuse ne dépend pas du nombre de pièces fabriquées), alors :

$$H(\vec{n}, r, \vec{p}) \underset{n\to\infty}{\longrightarrow} b(r, p_1). \tag{6.32}$$

Et le résultat est conservé quand on remplace $\frac{n_1}{n} = p_1$ constant par : $\exists p \in]0,1[$ t.q. $\lim_{n\to\infty} \frac{n_1}{n} = p_1$.

Autrement dit, notant $p_{1,n} \overset{\text{déf}}{=} \frac{p_1}{n}$ et $p_{2,n} = 1 - p_{1,n}$, si $p_{1,n} \longrightarrow_{n\to\infty} p_1$, alors, pour $k_1 \in [0, r]_{\mathbb{N}}$ et $k_2 = r - k_1$ avec $k_1 \leq n_1$ et $k_2 \leq n_2$, on a :

$$\frac{C_{p_{1,n}}^{k_1} C_{p_{2,n}}^{k_2}}{C_n^r} \underset{n\to\infty}{\longrightarrow} C_r^{k_1} p_1^{k_1} p_2^{k_2}. \tag{6.33}$$

Et dans ce cas, on peut remplacer le calcul (6.30) (coûteux) de la loi hypergéométrique par le calcul de la loi binomiale.

Preuve. Ici, comme $0 < p_1 < 1$ on a $n_1, n_2 \to \infty$ quand $n \to \infty$. Fixons r et k_1 (qui eux ne tendent pas vers l'infini), puis $k_2 = r - k_1$:

$$\begin{aligned}
\frac{C_{n_1}^{k_1} C_{n_2}^{k_2}}{C_n^r} &= \frac{n_1!}{k_1!(n_1 - k_1)!} \frac{n_2!}{k_2!(n_2 - k_2)!} \frac{r!(n-r)!}{n!} \\
&= \frac{r!}{k_1! k_2!} \frac{n_1!}{(n_1 - k_1)!} \frac{n_2!}{(n_2 - k_2)!} \frac{(n-r)!}{n!} \\
&\underset{n=\infty}{\sim} \frac{r!}{k_1! k_2!} \frac{n_1^{k_1} n_2^{k_2}}{n^r} = C_r^{k_1} (\frac{n_1}{n})^{k_1} (\frac{n_2}{n})^{k_2}.
\end{aligned}$$

▪

6.6.2 Cas ℓ issues

Plus généralement, soit un entier $\ell \geq 2$ et soit une partition $S = S_1 \cup ... \cup S_\ell$ de S. On note :

$$n = |S|, \qquad n_i = |S_i|, \qquad p_i = \frac{n_i}{n}, \tag{6.34}$$

où donc p_i est la probabilité de présence dans S_i (tirage équiprobable), et $\sum_{i=1}^{\ell} p_i = 1$.

Exemple 6.28 S un ensemble de pièces, $\ell = 3$, S_1 l'ensemble des pièces sans défaut, S_2 l'ensemble des pièces avec défaut mais vendables à prix réduit, S_3 l'ensemble des pièces invendables. ▪

Soit $r \in \mathbb{N}^*$ et soit Ω l'univers des suites de r tirages ordonnés ou non. Donc un élément de Ω est une suite $\vec{\omega} = (\omega_1, ..., \omega_r)$.

On définit le vecteur aléatoire (la variable aléatoire à valeurs vectorielles) :

$$\vec{X} : \begin{cases} \Omega & \to \mathbb{N}^\ell \\ \vec{\omega} & \mapsto \vec{X}(\vec{\omega}) = \begin{pmatrix} X_1(\vec{\omega}) = \text{le nombre de composantes de } \vec{\omega} \text{ qui sont dans } S_1 \\ \vdots \\ X_\ell(\vec{\omega}) = \text{le nombre de composantes de } \vec{\omega} \text{ qui sont dans } S_\ell \end{pmatrix} \end{cases}. \tag{6.35}$$

Proposition 6.29 *Soit $(k_1,...k_\ell) \in \mathbb{N}^\ell$ t.q. $k_1+...+k_\ell = r$ et $k_i \le n_i$ pour tout $i=1,...,\ell$. La loi de $\vec X$ est donnée par :*

$$P(\vec X = (k_1,...,k_\ell)) = \frac{C_{n_1}^{k_1}...C_{n_\ell}^{k_\ell}}{C_n^r}, \tag{6.36}$$

et cette loi est appelée loi hypergéométrique $H(\vec n, r, \vec p)$ (ou loi polygéométrique), où on a noté $\vec n = (n_1,...,n_\ell)$ et $\vec p = (p_1,...,p_\ell)$, ainsi que $n = n_1+...+n_\ell$.

Preuve. Même démarche que la démonstration de la proposition 6.24. ∎

Exemple 6.30 Soit S un ensemble de boules de k couleurs différentes, $k \ge 2$. On pose $S = S_1 \cup ... \cup S_k$, où les sous-populations S_i sont disjointes 2 à 2. On pose $n = |S|$, $n_i = |S_i|$ et $p_i = \frac{n_i}{n}$. Pour un tirage sans remise de m boules, on pose $P_{m_1,...,m_k}$ la probabilité de tirer m_1 boules de S_1,... , m_k boules de S_k (avec donc $m = m_1+...+m_k$ si on veut une probabilité non nulle). On a :

$$P_{m_1,...,m_k} = \frac{C_{n_1}^{m_1}...C_{n_k}^{m_k}}{C_n^m}. \tag{6.37}$$

∎

6.7 Loi géométrique (nombre d'échecs successifs, ou temps d'attente)

On considère n épreuves de Bernoulli, $n \ge 1$, de résultats a ou b.

On pourra interpréter a comme un succès et b comme un échec.

Soit $p = P(a)$ et $q = 1-p = P(b)$. Soit X la v.a.r. qui donne le premier tirage où on obtient a (on obtient le premier succès) :

$$X : \begin{cases} [a,b]^n \to \mathbb{N}, \\ \vec\omega = (\omega_1,...,\omega_n) \mapsto X(\vec\omega) = \begin{cases} i \text{ où } \omega_i \text{ est le premier } a \text{ si } i \text{ existe}, \\ n{+}1 \text{ si } \vec\omega = (b,...,b), \end{cases} \end{cases} \tag{6.38}$$

Donc, pour $k \in [1,n]_{\mathbb{N}}$:

$$X(\vec\omega) = k \iff \forall j = 1,...,k{-}1,\ \omega_j = b,\ \text{ et } \omega_k = a. \tag{6.39}$$

Donc :

$$P_X(1) = p, \quad P_X(2) = qp, \quad P_X(3) = q^2p, ..., \quad P_X(k) = q^{k-1}p, \quad P_X(n{+}1) = q^n. \tag{6.40}$$

Définition 6.31 La loi géométrique de paramètre p est la loi de probabilité P définie sur $\mathbb{N}^*$ (cas d'une infinité d'épreuves de Bernoulli) par, pour $p \in]0,1[$, notant $q = p-1$:

$$P(\{k\}) = q^{k-1}p \qquad (= (1-p)^{k-1}p). \tag{6.41}$$

(La suite $(u_k)_{k\in\mathbb{N}} = (P(\{k\}))_{k\in\mathbb{N}}$ est une suite géométrique de raison q et de premier terme p.) Autrement dit :

$$P = \sum_{k=1}^{\infty} q^{k-1}p\,\delta_k, \tag{6.42}$$

où δ_k est la masse de Dirac en k. (On vérifie que $P(\mathbb{N}) = \sum_{k=1}^\infty q^{k-1}p = p\sum_{k=0}^\infty q^k = p\frac{1}{1-q} = 1$.)

Proposition 6.32 *Pour la loi géométrique, quand $0<p<1$ et $q=1-p$, on a :*

$$P(X > k) = q^k, \tag{6.43}$$

i.e. la probabilité d'au moins k échecs successifs est q^k ; et donc $P(X \le k) = 1-q^k$ (probabilité d'avoir au moins un succès après k tirages).

Et le premier succès arrive de façon presque certaine après un nombre fini k d'épreuves (mais on peut attendre longtemps...). Autrement dit $P(X{=}\infty) = 0$.

Preuve. $P(X>k) = P(X \in [k{+}1,\infty]) = q^kp + q^{k+1}p + ... = q^kp\sum_{k=0}^\infty q^k = q^kp\frac{1}{1-q} = q^k$. (Et on vérifie que $P(X \le k) = p\sum_{j=0}^{k-1} q^j = p\frac{1-q^k}{1-q} = 1-q^k$).

Et $P(X=k{+}1) = pq^k \underset{k\to\infty}{\longrightarrow} 0$ donne $\lim\limits_{k\to\infty} P(X{=}k) = 0$, soit $P(X{=}\infty) = 0$ avec (2.9).

(Ou encore, pour tout $\varepsilon > 0$ on a $P(X=k) < \varepsilon$ dès que k est assez grand.) ∎

6.8 Loi normale (gaussienne) $\mathcal{N}(m,\sigma^2)$

Définition 6.33 Soit $m \in \mathbb{R}$ et soit $\sigma > 0$. Soit $p : \mathbb{R} \to \mathbb{R}$ donnée par :

$$p(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2\sigma^2}}. \tag{6.44}$$

La probabilité P de densité p est appelée “loi de gauss de moyenne m et d'écart type σ”, et notée :

$$P = \mathcal{N}(m,\sigma^2), \tag{6.45}$$

encore appelée “loi normale de moyenne m et d'écart type σ”. (On verra plus loin qu'effectivement m = la valeur moyenne, et σ = l'écart type).

Quand $m = 0$, la loi $P = \mathcal{N}(0,\sigma^2)$ est dite centrée.

Quand $m = 0$ et $\sigma = 1$, la loi $P = \mathcal{N}(0,1)$ est dite “loi normale standard” : la densité de probabilité est alors donnée par la gaussienne centrée réduite :

$$dP = p(t)\,dt \qquad \text{où} \qquad p(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}. \tag{6.46}$$

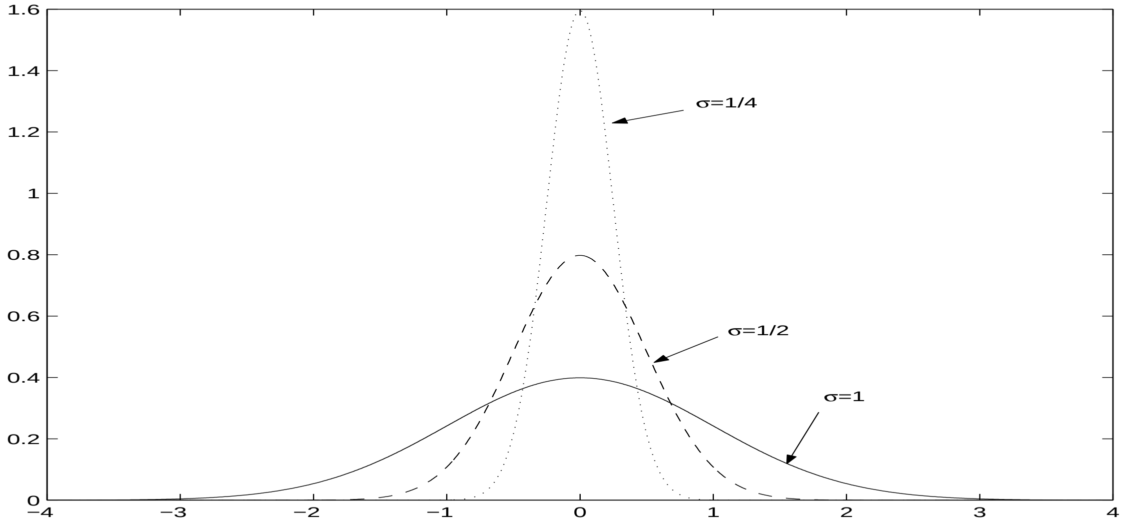


FIGURE 6.1 – Densités de probabilités gaussienne centrée $N(0,\sigma^2)$ pour $\sigma = 1$, $\sigma = \frac{1}{2}$ et $\sigma = \frac{1}{4}$. Et $P([-\alpha\sigma,\alpha\sigma])$ = aire sous la courbe entre les abscisses $t = -\alpha\sigma$ et $t = \alpha\sigma$. Voir tables de loi de la gaussienne.

La loi $P = \mathcal{N}(m,\sigma^2)$ est donc donnée par :

$$P([a,b]) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(t-m)^2}{2\sigma^2}}\,dt. \tag{6.47}$$

On vérifie que $\int_{\mathbb{R}} p(t)\,dt = 1$, cf. (1.33), et P est bien une probabilité.

On montre que $P \underset{\sigma\to 0}{\longrightarrow} \delta_m$ la masse de Dirac en m, voir cours de distribution.

Notation : on pose pour $m \in \mathbb{R}$ et $\sigma = 0$:

$$\mathcal{N}(m,0) = \delta_m. \tag{6.48}$$

Remarque 6.34 Si X est une v.a.r. on notera $P_X = P \circ X^{-1}$ la loi de X. Et dire que P_X suit la loi $\mathcal{N}(m,0)$ signifiera $P_X = \delta_m$, donc la valeur m est la valeur certaine pour X. Autrement dit $P_X = \mathcal{N}(m,0)$ ssi X est la v.a.r. constante $X = m1_{\mathbb{R}}$ (fonction déterministe). ▪

Proposition 6.35 *Soit $X : \Omega{=}\mathbb{R} \to \mathbb{R}$ une v.a.r. de loi $\mathcal{N}(m,\sigma^2)$, où $m \in \mathbb{R}$ et $\sigma > 0$. Alors $Y = \frac{1}{\sigma}(X-m)$ est une v.a.r. de loi $\mathcal{N}(0,1)$ (loi normale standard = loi de densitée gaussienne centrée réduite).*

Et plus généralement, si $P_X = \mathcal{N}(m,\sigma^2)$, où $m \in \mathbb{R}$ et $\sigma > 0$, alors toute translatée-dilatée $Z = \alpha + \beta X$ est de loi gaussienne, pour $\alpha, \beta = 0$.

Preuve. $Y^{-1}([c,d]) = \{\omega : \frac{X(\omega)-m}{\sigma} \in [c,d]\} = X^{-1}([m{+}\sigma c, m{+}\sigma d])$.

Donc $P_Y([c,d]) = P(Y^{-1}([c,d])) = P(X^{-1}([m{+}\sigma c, m{+}\sigma d])) = \int_{x=m+\sigma c}^{m+\sigma d} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx = \int_{y=c}^{d} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y^2}{2}}\,dy$.

Démonstration similaire dans le cas général quand $\beta \neq 0$. Et si $\beta = 0$, alors $Z = \alpha$ est une v.a.r. constante (fonction déterministe) de loi $\mathcal{N}(\alpha, 0) = \delta_\alpha$. ∎

6.9 Loi exponentielle (durée de vie d'une particule radioactive)

Définition 6.36 Soit $\lambda > 0$. La probabilité sur $\mathbb{R}_+$ de densité $p : \mathbb{R}_+ \to \mathbb{R}$ donnée par :

$$p(t) = \lambda e^{-\lambda t}, \tag{6.49}$$

est appelée loi exponentielle de paramètre λ. C'est donc la probabilité donnée sur $\Omega = \mathbb{R}_+$ par :

$$P([a,b]) = \int_a^b \lambda e^{-\lambda t}\,dt \qquad (= e^{-\lambda a} - e^{-\lambda b}). \tag{6.50}$$

Sa fonction de répartition est (immédiat), pour $t \geq 0$:

$$(P([0,t]) =) \qquad F(t) = 1 - e^{-\lambda t} \tag{6.51}$$

Cette loi est une très bonne modélisation de la désintégration d'une particule radioactive au cours du temps : on note $X : t \in \mathbb{R}_+ \to X(t)$ la (quantité de) radioactivité d'une particule au temps t. On note $\tau = \frac{1}{\lambda}$ appelé "temps caractéristique" (dimension d'un temps, et ainsi $P([a,b])$ est sans-dimension : c'est un pourcentage). Alors :

$$F_X(T) = P_X([0,T]) = \int_0^T \frac{1}{\tau} e^{-\frac{t}{\tau}}\,dt = 1 - e^{-\frac{T}{\tau}} \tag{6.52}$$

est la probabilité de désintégration de la particule dans l'intervalle $[0,T]$ (avec F_X la fonction de répartition de P_X). En particulier $P_X(\mathbb{R}_+) = 1$, et la particule a perdu, de manière certaine, toute sa radioactivité au bout d'un temps infini.

Et :

$$G_X(t) \overset{\text{déf}}{=} P(]t,\infty[) = 1 - F_X(t) = e^{-\frac{t}{\tau}} \tag{6.53}$$

est la fonction décroissante (de $1 = 100\%$ vers 0) qui mesure la quantité de radioactivité restante à l'instant t, avec pour unité de mesure : radioactivité $= 1$ à l'instant $t = 0$.

La demie durée de vie de radioactivité est le temps $t_{1/2}$ tel que :

$$P_X([0,t_{1/2}]) = \frac{1}{2}. \tag{6.54}$$

Comme $P_X([0,t_{1/2}]) = 1 - e^{-\frac{t_{1/2}}{\tau}}$, on a $t_{1/2} = -\tau \ln(\frac{1}{2})$, soit :

$$t_{1/2} = \tau \ln(2) \simeq 0.69\,\tau, \tag{6.55}$$

d'où la signification du temps caractéristique $\tau = \frac{1}{\lambda} = \frac{t_{1/2}}{\ln(2)} \simeq 1.44\,t_{1/2}$.

Exercice 6.37 Soit $X : \mathbb{R}_+ \to \mathbb{R}_+$ v.a.r t.q. $P(X > s) > 0$ pour tout $s > 0$. Montrer que X a la "propriété de non vieillissement" :

$$\forall s,t \in \mathbb{R}_+^*, \quad P(X > t{+}s \,|\, X > t) = P(X > s) \quad \Longleftrightarrow \quad X \text{ suit une loi exponentielle.} \tag{6.56}$$

Réponse. $\Leftarrow$. Hypothèse $p_X(t) = \lambda e^{-\lambda t}$. On a $P(X > t{+}s \,|\, X > t) = \frac{P((X>t+s))\cap(X>t))}{P(X>t)} = \frac{P(X>t+s)}{P(X>t)} = \frac{1-F(t+s)}{1-F(t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda(t)}} = e^{-\lambda s}$.

$\Rightarrow$. Hypothèse $P(X > t{+}s \,|\, X > t) = P(X > s)$, pour tout $s, t > 0$. Donc $P((X > t{+}s)) \cap (X > t)) = P(X > t)P(X > s)$, soit $P(X > t{+}s)) = P(X > t)P(X > s)$, soit $G(t{+}s) = G(t)G(s)$ où $G(s) = 1 - F(s)$. Avec F est croissante (fonction de répartition), donc G décroissante. Avec hypothèse $P(X > s) > 0$ pour tout $s > 0$, donc G est non nulle. Donc il existe $\lambda \in \mathbb{R}_+^*$ t.q. $G(s) = e^{-\lambda s}$. Donc $F(s) = 1 - e^{-\lambda s}$. Donc P a pour densité $F'(s) = \lambda e^{-\lambda s}$. ∎

Lien avec la loi géométrique : soit X une v.a.r. telle que P_X suit la loi géométrique de paramètre $p \in]0,1[$, et on pose $q = 1-p$, et soit Y une v.a.r. telle que P_Y suive la loi exponentielle de paramètre λ. Donc, cf. (6.43) et (6.53) :

$$P(X > n) = q^n = e^{n \log q} = e^{-n(-\log q)}, \qquad \text{et} \qquad P(Y > n) = e^{-n\lambda}.$$

Si :

$$\lambda = -\log q,$$

alors la loi de probabilité géométrique discrète P_X peut servir à approcher la loi de probabilité exponentielle continue P_Y, au sens où $P(X > n) = P(Y > n)$ pour tout entier n.

Remarque 6.38 Le temps caractéristique τ est en fait l'espérance (le centre de gravité) de la loi exponentielle : par intégration par parties :

$$E = \int_0^\infty t\,p(t)\,dt = \frac{1}{\tau}\int_0^\infty t\,e^{-\frac{t}{\tau}}\,dt = +\int_0^\infty e^{-\frac{t}{\tau}}\,dt - [t\,e^{-\frac{t}{\tau}}]_0^\infty = [-\tau e^{-\frac{t}{\tau}}]_0^\infty - 0 = \tau.$$

Et le moment d'ordre 2 est $= \tau^2$: par intégration par parties :

$$\int_0^\infty t^2\,p(t)\,dt = \tau^2.$$

Donc $\tau = \frac{1}{\lambda}$ "est vraiment" une grandeur caractéristique de la loi exponentielle. ▪

6.10 Loi de Poisson (ou loi des événements rares)

(Siméon Denis Poisson, 1781–1840.)

Définition 6.39 Cas $\Omega = \mathbb{N}$ et tribu $\mathcal{P}(\mathbb{N})$. Soit $\lambda > 0$. La loi de Poisson de paramètre λ est la loi de probabilité P définie pour $k \in \mathbb{N}$ par :

$$P(\{k\}) = \frac{\lambda^k}{k!}\,e^{-\lambda} \overset{\text{noté}}{=} p(k;\lambda). \tag{6.57}$$

(On vérifie que $P(\mathbb{N}) = \sum_{\mathbb{N}} P(\{k\}) = e^{-\lambda}\sum_{\mathbb{N}} \frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$.)

La loi de Poisson est appelée la loi des événements rares, voir proposition 6.40 et remarque 6.41.

Proposition 6.40 *Soient n épreuves de Bernoulli identiques d'issues a ou b. Soit $\lambda > 0$ fixé. On a, pour $k \in \mathbb{N}$:*

$$\lim_{n\to\infty} b(k;n,\frac{\lambda}{n}) = e^{-\lambda}\frac{\lambda^k}{k!}. \tag{6.58}$$

En particulier (aucun événement ne se produit pour cette loi) :

$$\lim_{n\to\infty} b(0;n,\frac{\lambda}{n}) = e^{-\lambda}. \tag{6.59}$$

(Appeler convergence étroite : la mesure P_n converge vers la mesure P.)

Preuve. Soit λ fixé. Notons $p_n = \frac{\lambda}{n}$. On a $b(k;n,p_n) = \frac{n!}{k!(n-k)!}p_n^k(1-p_n)^{n-k}$.

Pour $k = 0$ on a $b(0;n,p_n) = (1-p_n)^n$ (que des échecs), soit $b(0;n,p_n) = (1-\frac{\lambda}{n})^n \longrightarrow_{n\to\infty} e^{-\lambda}$.

Pour $k \in \mathbb{N}^*$ on a $\frac{n!}{(n-k)!}p_n^k = n(n-1)...(n-k+1)(\frac{\lambda}{n})^k \sim \lambda^k$ quand $n \to \infty$.

Et $(1-p_n)^{n-k} = e^{(n-k)\ln(1-\frac{\lambda}{n})} = e^{(n-k)(-\frac{\lambda}{n}+o(\frac{1}{n}))} = e^{-\lambda+k\frac{\lambda}{n}+o(\frac{1}{n}))} \sim e^{-\lambda}$ quand $n \to \infty$. ▪

Remarque 6.41 - Interprétation - On considère n tirages pour n "très grand", et on s'intéresse au cas d'une loi de Bernoulli $b(n,p)$ pour $p \ll 1$ (succès "très rare"). Posant $\lambda = pn$, la probabilité $b(k;n,\frac{\lambda}{n})$ d'obtenir k fois un succès est "bien approximée" par la valeur $e^{-\lambda}\frac{\lambda^k}{k!}$ quand $k \ll n$ (l'équation (6.58) suppose k fixé et $n \to \infty$) : cela permet d'éviter le calcul de $b(k;n,p) = C_n^k p^k q^{n-k}$ difficile à faire (essayer avec $B(2;1000,\frac{1}{500})$. Et même pour $k = 0$, le calcul de $b(0;n,p) = q^n$ est difficile à faire (essayer avec $n = 1000$ et $p = \frac{1}{500}$). ▪

Exemple 6.42 Soit $n \in \mathbb{N}^*$, $n \gg 1$, et la partition $\bigcup_{i=1}^n [\frac{i-1}{n}, \frac{i}{n}[$ de $[0,1[$ en n intervalles, chaque intervalle correspondant à un capteur de particules.

Soit $m \in \mathbb{N}^*$ un nombre fixé (petit) de particules.

Soit p_n la probabilité que l'intervalle $[\frac{i-1}{n}, \frac{i}{n}[$ contienne au moins un point, probabilité supposée indépendante de i. Soit $\alpha > 0$ fixé, on pose :

$$p_n = \frac{\alpha}{n},$$

On s'intéresse au cas $p_n \ll 1$, modélisé par $p_n \to 0$ quand $n \to \infty$.

Soit X la v.a.r. qui compte le nombre d'intervalles occupés par les m particules : on suppose que $P_X(k) = b(k; n, p_n)$ (probabilité pour que m points distribués au hasard soient dans k intervalles). En particulier :

$$P_X(0) = b(0; n, p_n) = (1-p_n)^n \qquad (\underset{\infty}{\sim} e^{-\alpha})$$

est la probabilité “toutes les particules sont à l'extérieur de $[0,1[$” (aucune particule n'a été détectée dans $[0,1[$). Cette probabilité tend vers la limite finie $e^{-\alpha}$ constante, probabilité pour que les particules à détecter ne parviennent pas aux capteurs.

Donc $f(\alpha) = 1 - e^{-\alpha}$ est la probabilité d'avoir au moins un point $[0,1[$.

Et $b(k; n, p_n) \simeq e^{-\alpha}\frac{\alpha^k}{k!} = p(k; \alpha)$ quand $n \to \infty$, cf. (6.58).

Exercice 6.43 Soit une population de $n = 500$ personnes. On souhaite connaître connaître la probabilité que zéro, une, deux,..., personnes soient nées le 14 juillet. Calculer cette probabilité.

Réponse. Ici $b(k; n, p) = b(k; 500, \frac{1}{365})$ est la probabilité pour que k personnes soient nées le 14 juillet, en supposant l'équiprobabilité des naissances ($p = \frac{1}{365}$ est la probabilité de naître le jour j).

Le calcul exacte par exemple de $b(2; 500, \frac{1}{365}) = C_{500}^2(\frac{1}{365})^2(\frac{364}{365})^{498}$ ne peut pas être fait sur une petite calculatrice de manière évidente.

Un résultat approché peut être calculé par la loi de Poisson à l'aide de (6.58) : ici $\lambda = np = \frac{500}{365}$.

Ici $k = 0, 1, 2, 3, 4 \ll n = 500$, et $\lambda = pn = \frac{500}{365} \ll 500 = n$ rentrent dans le cadre de l'approximation : $P_X(k) \simeq e^{-\lambda}\frac{\lambda^k}{k!}$. Et donc $P_X(0) \simeq 25\%$, $P_X(1) \simeq 35\%$, $P_X(2) \simeq 24\%$, $P_X(3) \simeq 11\%$, $P_X(4) \simeq 4\%$, $P_X(5) \simeq 1\%$...

(Pour 333 personnes $\simeq$ population d'élèves de l'ISIMA : $P_X(0) \simeq 40\%$, $P_X(1) \simeq 37\%$, $P_X(2) \simeq 17\%$, $P_X(3) \simeq 5\%$, $P_X(4) \simeq 1\%$...)

Exercice 6.44 On considère un médicament. Soit p la probabilité (le risque individuel) qu'une personne ait un effet secondaire. On suppose $p \ll 1$.

On prend un groupe de n personnes, $n \gg 1$.

1- Soit X la v.a.r. qui compte le nombre de personnnes qui présente l'effet secondaire.

Donner la loi de X, puis $P_0 = P(X{=}0)$ et $P_1 = P(X{=}1)$ et $P(X{\geq}2)$. Calculer P_0 pour $p = \frac{1}{10000}$ et $n = 10000$.

2- On ne connaît pas p.

On applique la règle 1 : pour $P_0 \geq 5\%$ il est raisonnable de penser observer des effets secondaires.

* Quel est le risque individuel p ?

* Que vaut p pour $n = 10000$?

* Si ce médicament est donné à 10^6 personnes, combien de personnes risquent d'avoir un effet secondaire ?

3- On applique la règle 2 : pour $P_0 \geq 90\%$ il est raisonnable de penser observer des effets secondaires. Mêmes questions.

Réponse. 1- La loi de X est la loi binomiale $b(n, p)$. Ainsi $P(X{=}k) = C_n^k p^k (1-p)^{n-k}$, et $P(X{=}0) = P_0 = (1-p)^n = (1 - \frac{pn}{n}) \simeq e^{-np}$ quand $n \gg 1$ (on a posé $\lambda = np$ dans (6.59)), probabilité de ne pas observer d'effets secondaire dans le groupe de n personnes.

Pour $p = \frac{1}{10000}$ et $n = 10000$ on obtient $P_0 = e^{-1} \simeq 0.37$, probabilité qu'aucune personne n'ait d'effet secondaire. Et $P_1 \simeq e^{-np}(np) = 0.37$. D'où $P(X{\geq}2)1 - 0.37 - 0.37 = 0.26$.

2- * On doit avoir $e^{-np} \geq \frac{5}{100} = \frac{1}{20}$. Donc $-np \geq \ln\frac{1}{20} = -\ln(20)$. Donc $p \leq \frac{\ln(20)}{n} \simeq \frac{3}{n}$.

* Pour $n = 10000$ on obtient $p \leq \frac{3}{10000} = 0.03\%$.

* Avec ce $p = 0.03\%$ (risque individuel), si on donne ce médicament à 10^6 personnes, on aura $10^6 \times \frac{3}{10000} = 300$ personnes qui risquent d'avoir un effet secondaire. C'est beaucoup.

3- * On doit avoir $e^{-np} \geq \frac{90}{100}$. Donc $-np \geq -\ln\frac{90}{100}$. Donc $p \leq \frac{\ln 0.9}{n} \simeq \frac{0.1}{n}$.

* Pour $n = 10000$ on obtient $p \leq \frac{1}{100000} = 0.001\%$.

* Avec ce $p = 0.001\%$ (risque individuel), si on donne ce médicament à 10^6 personnes, on aura $10^6 \times \frac{1}{100000} = 10$ personnes qui risquent d'avoir un effet secondaire.

6.11 Loi γ de la fonction factorielle Γ

Cette loi sera utilisée dans le paragraphe suivant : loi du ki-2.

Pour un entier $n \geq 1$, on note :

$$\Gamma(n+1) = n! \qquad (= \prod_{k=1}^{n} k = 1 \times 2 \times ... \times (n-1) \times n). \tag{6.60}$$

Et par convention on pose $\Gamma(1) = 0! = 1$. Une propriété immédiate est pour $n \geq 0$:

$$\Gamma(n+1) = n\Gamma(n), \tag{6.61}$$

ainsi que, pour tout $n > 0$ entier :

$$\Gamma(n) = \int_0^\infty t^{n-1} e^{-t}\, dt.$$

En effet, par intégration par parties, $\int_0^\infty t^n e^{-t}\, dt = \int_0^\infty n t^{n-1} e^{-t}\, dt + [t^n e^{-t}]_0^\infty$ quand $n \geq 1$, et on retrouve $\Gamma(n+1) = n\Gamma(n)$, avec, pour $n = 1$, $\int_0^\infty t^0 e^{-t}\, dt = 1$.

D'où la généralisation de la définition aux réels positifs :

Définition 6.45 On note $\Gamma : \mathbb{R}_+^* \to \mathbb{R}$ la fonction définie par :

$$\Gamma(z) \stackrel{\text{déf}}{=} \int_0^\infty t^{z-1} e^{-t}\, dt \qquad (= \int_0^\infty e^{(z-1)\log(t)} e^{-t}\, dt). \tag{6.62}$$

Et on généralise cette définition à tout complexe z de partie réelle > 0.

On vérifie par intégration par parties que, pour tout $z > 0$:

$$\Gamma(z+1) = z\Gamma(z). \tag{6.63}$$

En particulier, on trouve, pour $n \geq 1$:

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}, \qquad \Gamma(n+\frac{1}{2}) = (n-\frac{1}{2})(n-\frac{3}{2})...\frac{3}{2}\frac{1}{2}\sqrt{\pi}, \tag{6.64}$$

car $\int_0^\infty \frac{1}{\sqrt{t}} e^{-t}\, dt = \int_0^\infty \frac{1}{y} e^{-y^2} 2y\, dy = 2\int_0^\infty e^{-y^2}\, dy = \int_{-\infty}^\infty e^{-y^2}\, dy = \sqrt{\pi}$.

Définition 6.46 Soit $\alpha > 0$ et $z \in \mathbb{C}$ de partie réelle > 0. La loi $\gamma_{\alpha,z}$ dite “loi gamma d'indice α et z” est la loi de densité $\gamma_{\alpha,z} : \mathbb{R} \to \mathbb{R}_+$ définie par :

$$\gamma_{\alpha,z}(t) = \begin{cases} \dfrac{\alpha^z}{\Gamma(z)} t^{z-1} e^{-\alpha t} & \text{si } t > 0, \\ 0 \quad \text{si } t \leq 0. \end{cases} \tag{6.65}$$

Exercice 6.47 Vérifier que $\int_{\mathbb{R}_+} \gamma_{\alpha,z}(t)\, dt = 1$ pour $\alpha > 0$.

Réponse. $\displaystyle\int_0^\infty \gamma_{\alpha,z}(t)\, dt = \frac{\alpha^z}{\Gamma(z)} \int_{t=0}^\infty t^{z-1} e^{-\alpha t}\, dt = \frac{\alpha^z}{\Gamma(z)} \int_{y=0}^\infty (\frac{y}{\alpha})^{z-1} e^{-y} \frac{dy}{\alpha} = 1.$ ▪

En particulier :

$$\gamma_{1,z}(t) = \frac{1}{\Gamma(z)} t^{z-1} e^{-t} 1_{\mathbb{R}_+}(t). \tag{6.66}$$

Proposition 6.48 *Pour tout $t \geq 0$:*

$$(\gamma_{\alpha,y} * \gamma_{\alpha,z})(t) = \gamma_{\alpha,y+z}(t). \tag{6.67}$$

Preuve. Les fonctions étant nulles pour $t \le 0$, on a d'une part :

$$\begin{aligned}
(\gamma_{\alpha,y} * \gamma_{\alpha,z})(t) &= \int_{\tau=-\infty}^{\infty} \gamma_{\alpha,y}(\tau)\gamma_{\alpha,z}(t-\tau)\,d\tau = \int_{\tau=0}^{t} \gamma_{\alpha,y}(\tau)\gamma_{\alpha,z}(t-\tau)\,d\tau \\
&= \int_{\tau=0}^{t} \frac{\alpha^{y+z}}{\Gamma(y)\Gamma(z)} \tau^{y-1}(t-\tau)^{z-1} e^{-\alpha\tau-\alpha(t-\tau)}\,d\tau \\
&= \frac{\alpha^{y+z}}{\Gamma(y)\Gamma(z)} e^{-\alpha t} \int_{\tau=0}^{t} \tau^{y-1}(t-\tau)^{z-1}\,d\tau \\
&\overset{\tau=tu}{=} \frac{\alpha^{y+z}}{\Gamma(y)\Gamma(z)} e^{-\alpha t}\, t^{y+z-1} \int_{u=0}^{1} u^{y-1}(1-u)^{z-1}\,du \\
&= \gamma_{\alpha,y+z}(t) \frac{\Gamma(y+z)}{\Gamma(y)\Gamma(z)} \int_{u=0}^{1} u^{y-1}(1-u)^{z-1}\,du.
\end{aligned}$$

D'autre part avec Fubini, les fonctions étant positives :

$$\begin{aligned}
\int_{t=-\infty}^{\infty} (\gamma_{\alpha,y} * \gamma_{\alpha,z})(t)\,dt &= \int_{\tau=-\infty}^{\infty} \int_{t=-\infty}^{\infty} \gamma_{\alpha,y}(\tau)\gamma_{\alpha,z}(t-\tau)\,dt d\tau \\
&= (\int_{\tau=-\infty}^{\infty} \gamma_{\alpha,y}(\tau)\,d\tau)(\int_{u=-\infty}^{\infty} \gamma_{\alpha,z}(u)\,du) = 1 \times 1 = 1.
\end{aligned}$$

Comme $\int_{t=-\infty}^{\infty} \gamma_{\alpha,y+z}(t)\,dt = 1$, on en déduit $\int_{u=0}^{1} u^{y-1}(1-u)^{z-1}\,du = \frac{\Gamma(y)\Gamma(z)}{\Gamma(y+z)}$, d'où (6.67). ▪

6.12 Loi de X^2, loi du χ^2 (loi du ki-2 = loi du chi-2 = loi de Pearson à n degrés de liberté)

[Chi-square distribution.]

Soit $X_1, ..., X_n$ sont n v.a.r. sur un même espace probabilisé $(\Omega, \mathcal{T}, P)$. On note $\vec{X} = (X_1, ..., X_n)$ (un vecteur aléatoire sur $(\Omega, \mathcal{T}, P)$), et on note :

$$S^2 = ||\vec{X}||^2 \overset{\text{déf}}{=} \sum_{i=1}^{n} X_i^2. \tag{6.68}$$

Ainsi $S^2 : \Omega \to \mathbb{R}$ est la fonction (la v.a.r.) définie sur $(\Omega, \mathcal{T}, P)$ par, pour tout $\omega \in \Omega$:

$$S^2(\omega) = \sum_{i=1}^{n} X_i(\omega)^2 \qquad (= ||\vec{X}||^2(\omega) \overset{\text{déf}}{=} ||\vec{X}(\omega)||^2_{\mathbb{R}^n} = \sum_{i=1}^{n} X_i^2(\omega)). \tag{6.69}$$

Remarque 6.49 S^2 correspond à une énergie (ou encore à la somme des carrées des erreurs = dispersion des résultats), et peut être dérivée (voir également la méthode des moindres carrés). Alors que pour la somme des valeurs absolues, $\sum_{i=1}^{n} |X_i|$, la dérivation (et la recherche du minimum de l'erreur) n'est pas possible en $\vec{X} = \vec{0}$ (alors que le minimum est justement en $\vec{X} = \vec{0}$). ▪

Définition 6.50 Lorsque toutes les v.a.r. X_i, $i = 1, ..., n$, sont indépendantes et de même loi normale centrée réduite (toutes de densité $p(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$), on appelle fonction ki-2 la fonction :

$$S^2 \overset{\text{noté}}{=} \chi^2(n) \qquad (= ||\vec{X}||^2). \tag{6.70}$$

Et on appelle loi du χ^2 (loi du "ki deux") à n degrés de liberté : la loi de la v.a.r. $\chi^2(n)$:

$$P_{\chi^2(n)} = \text{loi du ki-2}. \tag{6.71}$$

Proposition 6.51 *Cas n=1. Soit X une v.a.r. de loi P_X normale centrée réduite. La loi de $X^2 =^{\text{noté}} \chi^2(1)$ est la loi $P_{X^2} = P_{\chi^2(1)}$ de densité $p1_{\mathbb{R}_+}$ où, pour $t > 0$:*

$$p(t) = \gamma_{\frac12,\frac12}(t) \qquad (= \frac{1}{\sqrt{2\pi}} t^{-\frac12} e^{-\frac{t}{2}} = \frac{1}{\sqrt{2\pi t e^t}}). \tag{6.72}$$

cf. (6.65) avec $z = \frac12$ et $\alpha = \frac12$.

Et si X est une v.a.r. de loi P_X normale de moyenne m et écart type σ, alors la loi de $(X-m)^2 =^{\text{noté}} \chi^2(1,\sigma)$ est la loi $P_{(X-m)^2} = P_{\chi^2(1,\sigma)}$ de densité $p1_{\mathbb{R}_+}$ où, pour $t > 0$:

$$p(t) = \gamma_{\frac{1}{2\sigma^2},\frac{1}{2}}(t) \qquad (= \frac{1}{\sigma\sqrt{2\pi}}\, t^{-\frac{1}{2}} e^{-\frac{t}{2\sigma^2}}), \tag{6.73}$$

cf. (6.65) avec $z = \frac{1}{2}$ et $\alpha = \frac{1}{2\sigma^2}$.

Preuve. On a (avec $t = y^2$ quand $t \geq 0$) :

$$\begin{aligned} P((X-m)^2 \leq t) &= P(|X-m| \leq \sqrt{t}) = P(X-m \in [-\sqrt{t},\sqrt{t}]) = P_{X-m}([-\sqrt{t},\sqrt{t}]) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\tau=-\sqrt{t}}^{\sqrt{t}} e^{-\frac{\tau^2}{2\sigma^2}}\, d\tau = \frac{2}{\sigma\sqrt{2\pi}} \int_{\tau=0}^{\sqrt{t}} e^{-\frac{\tau^2}{2\sigma^2}}\, d\tau \\ &= \frac{2}{\sigma\sqrt{2\pi}} \int_{y=0}^{t} e^{-\frac{y}{2\sigma^2}} \frac{dy}{2\sqrt{y}} = \frac{1}{\sigma\sqrt{2\pi}} \int_{y=0}^{t} y^{-\frac{1}{2}} e^{-\frac{y}{2\sigma^2}}\, dy = \int_{y=0}^{t} p(y)\, dy. \end{aligned}$$

∎

Proposition 6.52 *Soient $X_1,...,X_n$ n lois indépendantes normales centrées réduites (moyenne nulle et écart type $\sigma = 1$). La loi de $||\vec{X}||^2_{\mathbb{R}^n} =^{\text{noté}} \chi^2(n)$ est la loi $P_{||\vec{X}||^2_{\mathbb{R}^n}} = P_{\chi^2(n)}$ de densité $p1_{\mathbb{R}_+}$ où, pour $t > 0$:*

$$p(t) = \gamma_{\frac{1}{2},\frac{n}{2}}(t) \qquad (= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\, t^{\frac{n}{2}-1} e^{-\frac{t}{2}}). \tag{6.74}$$

En particulier, pour n pair, notant $n = 2m$, on a $p(t) = \frac{1}{2^m(m-1)!}\, t^{m-1} e^{-\frac{t}{2}}$.

Soient $X_1,...,X_n$ n lois indépendantes normales de moyenne m_i et toutes de même écart type σ. La loi de $\sum_{i=1}^n (X_i - m_i)^2$ est la loi $P_{\chi^2(n,\sigma)}$ de densité $p1_{\mathbb{R}_+}$ où, pour $t > 0$:

$$p(t) = \gamma_{\frac{1}{2\sigma^2},\frac{n}{2}}(t). \tag{6.75}$$

Preuve. On verra, cf. (8.7), que, lorsque les v.a.r. X et Y sont indépendantes, pour des loi de densités, $dP_X(x) = \rho_X(x)\,dx$ et $dP_Y(x) = \rho_Y(x)\,dx$, la loi P_{X+Y} est la loi de densité $dP_{X+Y}(x) = (\rho_X * \rho_Y)(x)\,dx$. Et on a, pour tout $t \geq 0$:

$$(\gamma_{\alpha,y} * \gamma_{\alpha,z})(t) = \gamma_{\alpha,y+z}(t),$$

cf. (6.67). D'où quand X_1 et X_2 sont deux lois normales centrées indépendantes de même variance σ^2 :

$$dP_{X_1^2+X_2^2}(t) = (\gamma_{\frac{1}{2\sigma^2},\frac{1}{2}} * \gamma_{\frac{1}{2\sigma^2},\frac{1}{2}})(t)\,dt = \gamma_{\frac{1}{2\sigma^2},1}(t)\,dt,$$

d'où (6.74). Puis la formule par récurrence. ∎

6.13 Loi de Cauchy

C'est la loi de densité définie sur $\mathbb{R}$ par :

$$dP(x) = p(x) = \frac{1}{\pi}\,\frac{1}{1+x^2}. \tag{6.76}$$

(On vérifie que $\int_{\mathbb{R}} p(x)\,dx = \frac{1}{\pi}[\arctan]_{-\infty}^{\infty} = 1$.)

C'est une loi qui n'admet pas d'espérance ($\int_0^\infty x\,\frac{1}{1+x^2}\,dx = \infty$), contrairement aux lois présentées avant.

Exemple 6.53 Un phare face à la mer. Soit :

$$q(\theta) = \frac{1}{\pi} 1_{]-\frac{\pi}{2},\frac{\pi}{2}[}(\theta)$$

l'émission (uniforme) de lumière suivant l'angle θ, l'angle $\theta = 0$ correspondant à l'axe perpendiculaire à la côte (faire un dessin). Soit une droite parallèle à la côte située "dans la mer" à une distance 1. Soit $x = X(\theta)$ le point d'impact de la lumière sur cette droite : $\tan(\theta) = \frac{x}{1} = x$.

La fonction de répartition est $F(x) = P_X(]-\infty,x[) = P(X \le x) = P(\{\theta : X(\theta) \le x\})$, i.e. :

$$F(x) = \int_{\theta=-\infty}^{\arctan x} q(\theta)\,d\theta = \frac{1}{\pi}\arctan(x) + \frac{1}{2}.$$

Et $F'(x) = \frac{1}{\pi}\frac{1}{1+x^2} = p(x)$: la probabilité d'atteindre le point x est une probabilité continue de densité la loi de Cauchy. ▪

Exercice 6.54 Soit $d\mu(x) = \frac{1}{\pi}dx$ la densité de probabilité uniforme sur $]-\frac{\pi}{2},\frac{\pi}{2}[$. Montrer que la loi de Cauchy est la densité de la loi image $\nu = \mu \circ \arctan$.

Réponse. On prend un peu d'avance, voir paragraphe 9.1.3. Il s'agit de montrer, cf. (9.13), que $\nu(g) = \mu(g\circ\tan) = \frac{1}{\pi}\int_{\mathbb{R}} g(y)\,\frac{dy}{1+y^2}$, pour une fonction mesurable g. On a :

$$\mu(g\circ\tan) = \int_{x\in\mathbb{R}} g(\tan(x))\,d\mu(x) = \int_{x\in-\frac{\pi}{2}}^{\frac{\pi}{2}} g(\tan(x))\frac{1}{\pi}dx = \int_{y=-\infty}^{\infty} g(y)\frac{1}{\pi}\frac{dy}{1+y^2}, = \int_{y=-\infty}^{\infty} g(y)\,dP(y),$$

à l'aide du changement de variable $y = \tan x$ qui donne $dy = (1+\tan^2 x)\,dx$. ▪

7 Processus stochastiques

7.1 Définitions

7.1.1 Définition d'un processus

Soit I un ensemble (d'indices) donné (par exemple $[0,n]_{\mathbb{N}}$, $\mathbb{N}^*$, $[a,b]$, $\mathbb{R}_+$,...).
Soit E un ensemble muni d'une tribu $\mathcal{T}_E$.

Définition 7.1 Une famille $(X_i)_{i\in I}$ de va.r. $X_i : \Omega \to E$ toutes définies sur un même espace probabilisé $(\Omega,\mathcal{T}_\Omega,P)$ est appelé un processus stochastique.

Si $I = [1,n]_{\mathbb{N}}$ ou $\mathbb{N}^*$ (ou plus généralement un ensemble au plus dénombrable), alors $(X_i)_{i\in I}$ est un processus discret (ou à temps discret).

Si $I = [0,T]$ ou $\mathbb{R}_+$, alors $(X_t)_{t\in I}$ est un processus continu (ou à temps continu).

Définition 7.2 $E \supset \mathrm{Im}X_i$ est appelé l'ensemble des états (fini ou non).

Exemple 7.3 Processus de Bernoulli (processus discret) $(X_i)_{i\in\mathbb{N}^*}$: l'univers est $\{a,b\}^{\mathbb{N}^*}$, et $X_i(\vec{\omega}) = \omega_i$ pour $\vec{\omega} = (\omega_1,\omega_2,...)$. Ici $X_i = \pi_i$ la projection sur la i-ème coordonnée. ▪

Exemple 7.4 Cas continu. $I = [0,T]$ (ou $I = \mathbb{R}_+$). Soit $\{c_1\}$ l'ensemble constitué d'un capteur (une cible). Pour $t \in [0,T]$, soit S_t le nombre de particules qui sont arrivées sur le capteur c_1 dans l'intervalle $[0,t]$ (idem pour compter les voitures passant un péage donné, le nombre de clients passant par une caisse...). Ainsi $(S_t)_{t\in[0,T]}$ est un processus continu à valeurs discrètes, les S_t étant définies sur $\mathcal{F}([0,T];\mathbb{N})$ (plus précisément sur le sous-espace des fonctions croissantes nulles en 0).

Donc avec δ_t la masse de Dirac :

$$S_t(\omega) = \delta_t(\omega), \qquad \text{i.e.} \qquad S_t = \delta_t. \tag{7.1}$$

Si on dispose de plusieurs capteurs c_i, $i = 1,...,m$, on dispose des processus $(S_{c_i,t})_{t\in[0,T]}$, où $S_{c_i,t} : \Omega = \mathcal{F}([0,T],\mathbb{N})$:

$$S_{c_i,t}(\omega) = \omega(t) = \text{nb de particules captées par } c_i \text{ dans } [0,t]. \tag{7.2}$$

Et pour une suite strictement croissante $(t_j)_{j=1,\dots,n}$, les suites $(S_{c_i,t_j})_{j\in[1,n]_{\mathbb{N}}}$ sont des processus discrets.

▪

7.1.2 Processus i.i.d.

Définition 7.5 Un processus discret (X_i) est dit i.i.d. = "independant identically distributed random sequence" ssi les X_i sont indépendantes et de même loi ($P_{X_i} = P_{X_1}$ pour tout i).

Exemple 7.6 Les X_i du processus de Bernoulli forme un processus i.i.d.. ▪

7.1.3 Trajectoires

Soit un processus $(X_i)_{i\in I}$, où $X_i : \Omega \to E$.

Définition 7.7 À $\omega \in \Omega$ fixé, la fonction :

$$X_\omega : \begin{cases} I & \to E \\ i & \to X_\omega(i) \stackrel{\text{déf}}{=} X_i(\omega) \qquad (= X(i,\omega)), \end{cases} \tag{7.3}$$

est appelée la trajectoire de ω. Et son image $\mathrm{Im}X_\omega = X_\omega(I)$ est également appelée la trajectoire de ω (attention au contexte).

Exemple 7.8 Processus discret, voir § suivant. ▟

Exemple 7.9 Processus continu de l'exemple 7.4 : $\omega : t \to \omega(t)$ est la trajectoire de ω au cours du temps. Le graphe de ω est la trajectoire géométrique. ▟

Exemple 7.10 Le processus de Bernoulli a été décrit au paragraphe 6.3 page 61. ▟

7.1.4 Application : marche aléatoire (ou promenade aléatoire)

Soit $\vec{Z} = (Z_n)_{n\in\mathbb{N}}$ un processus de Bernoulli (une suite vérifiant (6.7)).
L'univers est $\Omega = \{a,b\}^{\mathbb{N}^*}$.

Définition 7.11 Une marche aléatoire est un processus $\vec{X} = (X_n)_{n\in\mathbb{N}^*}$ à valeurs dans $\{-1,1\}$ défini par, pour tout n :

$$X_n = 2Z_n - 1, \qquad X_n(\vec{\omega}) = \begin{cases} 1 & \text{si } \omega_n = a, \\ -1 & \text{si } \omega_n = b, \end{cases} \tag{7.4}$$

on avance (si pile) ou recule (si face) de 1 au temps n. L'ensemble des états est $E = \{0,1\}$.

Définition 7.12 Si la marche aléatoire (X_n) est i.i.d. avec, pour tout n :

$$\begin{cases} P_{X_n}(1) = p & (= P(\{\vec{\omega} : X_n(\vec{\omega}) = \omega_n = 1\}), \\ P_{X_n}(-1) = 1-p & (= P(\{\vec{\omega} : X_n(\vec{\omega}) = \omega_n = -1\}). \end{cases} \tag{7.5}$$

où $p \in]0,1[$, alors la marche aléatoire (X_n) est appelée marche aléatoire de paramètre p.
Si $p = q = \frac{1}{2}$, la marche est dite symétrique, ou appelée promenade aléatoire.

On pose :

$$S_r = \sum_{i=1}^{r} X_r. \tag{7.6}$$

Exemple 7.13 Avec $r = 5$ et $\vec{\omega} = (a,b,b,a,b)$ (résultat d'une suite de 5 tirages) : $\vec{X}(\vec{\omega}) = (1,-1,-1,1,-1)$ et donc $S_r(\vec{\omega}) = S_5(\vec{\omega}) = -1$. ▟

Proposition 7.14 *Soit $(X_n)_{n\in\mathbb{N}^*}$ une marche aléatoire de paramètre p. Alors, $b(r,p)$ étant la loi binomiale, cf. paragraphe 6.3, on a, si $k \in [-r,r]_{\mathbb{N}}$:*

$$P_{S_r}(k) = \begin{cases} = b(\dfrac{r+k}{2}; r,p) = C_r^{\frac{r+k}{2}} p^{\frac{r+k}{2}} (1-p)^{\frac{r-k}{2}}, & \textit{si } r+k \textit{ pair}, \\ = 0, \quad \textit{sinon}. \end{cases} \tag{7.7}$$

Preuve. Notons $T_r = \sum_{i=1}^r Z_i$ la somme (7.6) correspondant à la suite de Bernoulli Z_i. Avec S_r donné par (7.6), on a $S_r = \sum_{i=1}^r X_i = \sum_{i=1}^r (2Z_i - 1) = 2Z_r - r$.
D'où $P(S_r = k) = P(2T_r - r = k) = P(T_r = \frac{k+r}{2})$, cf. (6.12).
Donc si $k+r$ est impair alors $P(S_r = k) = 0$, et si $k+r$ est pair alors $P(S_r = k) = b(\frac{r+k}{2}; r,p)$. ▟

7.1.5 Probabilités de transition

Soit un processus $(X_i)_{i\in I}$, où $X_i : \Omega \to E$. Soit $\mathcal{T}_E$ une tribu de E.

Définition 7.15 Dans le cas $I \subset \mathbb{R}$, comme I est alors ordonné, si $t \in I$, et si $\Delta t > 0$ est tel que $t + \Delta t \in I$, alors, pour $B_1, B_2 \in \mathcal{T}_E$, le réel (probabilité conditionnelle) :

$$P_{|X_t\in B_1}(X_{t+\Delta t} \in B_2) \qquad (= P(X_{t+\Delta t} \in B_2 \mid X_t \in B_1)) \tag{7.8}$$

est appelée la probabilité de transition de B_1 à B_2 entre t et $t + \Delta t$ (probabilité d'être dans B_2 à l'instant $t + \Delta t$ sachant qu'on était dans B_1 à l'instant t).

7.2 Processus discrets et chaines de Markov à temps discrets

7.2.1 Définition

Soit $(t_n)_{n\in\mathbb{N}}$ une suite strictement croissante de réels. Soit $(X_{t_n})_{n\in\mathbb{N}}$ un processus discret, les $X_{t_n} : \Omega \to E$ étant à valeurs dans un ensemble d'états discret $E = \{x_0, x_1, ..., x_n, ...\}$ et vérifiant, pour tout t_n et tout x_i :

$$P(X_{t_n}{=}x_i) \neq 0, \qquad \text{soit} \qquad P(\{\vec{\omega} : X_n(\vec{\omega}) = x_i)\}) \neq 0. \tag{7.9}$$

(Toutes les fonctions X_n sont surjectives et $P(X_n^{-1}(x_i)) \neq 0$ pour tout n et tout i.)

Définition 7.16 Le processus $(X_{t_n})_{n\in\mathbb{N}}$ possède la propriété de Markov ssi pour tout $n \in \mathbb{N}^*$:

$$P_{|(X_{t_0}=x_{i_0},...,X_{t_{n-1}}=x_{i_{n-1}})}(X_{t_n}{=}x_{i_n}) = P_{|(X_{t_{n-1}}=x_{i_{n-1}})}(X_{t_n}{=}x_{i_n}), \tag{7.10}$$

soit $P(X_{t_n}{=}x_{i_n}|X_{t_0}{=}x_{i_0}, ..., X_{t_{n-1}}{=}x_{i_{n-1}}) = P(X_{t_n}{=}x_{i_n}|X_{t_{n-1}}{=}x_{i_{n-1}})$.

Et $(X_{t_n})_{\mathbb{N}}$ est alors appelée une chaine de Markov.

Interprétation en génétique : $t_n = n$ correspond à la n-ième génération, et (7.10) indique que les propriétés de la n-ième génération (génération enfants) ne dépend que de la génération $n-1$ (génération parents) et pas des générations précédentes (génération grands-parents et plus).

Remarque 7.17 Cas des X_{t_i} indépendantes : alors $P_{|(X_{t_{n-1}}=x_{i_{n-1}})}(X_{t_n}{=}x_{i_n}) = P(X_{t_n}{=}x_{i_n})$: on dispose également d'une chaîne de Markov, mais ce cas est exclu en général. ■

Définition 7.18 La chaîne de Markov est dite homogène ssi, pour tout $n \geq 1$ et pour tout i, j :

$$P_{|(X_{t_{n-1}}=x_i)}(X_{t_n}{=}x_j) = P_{|X_{t_0}=x_i}(X_{t_1}{=}x_j), \tag{7.11}$$

soit $P(X_{t_n}{=}x_j|X_{t_{n-1}}{=}x_i) = P(X_{t_1}{=}x_j|X_{t_0}{=}x_i)$.

Interprétation en génétique : les propriétés sont conservées de génération en génération.

Remarque 7.19 Dans le cas d'une chaîne de Markov homogène, on appelle loi de la chaîne de Markov issue de l'état i, i.e. la probabilité :

$$P_i = P_{|X_{t_0}=x_i}. \tag{7.12}$$

■

Définition 7.20 La matrice de transition d'une chaîne de Markov homogène est la matrice $[p_{ij}]$ définie pour $i, j \in \mathbb{N}$ par :

$$p_{ij} = P_{|(X_{t_0}=x_i)}(X_{t_1}{=}x_j) \overset{\text{noté}}{=} p(x_i, x_j). \tag{7.13}$$

Proposition 7.21 *Pour une chaîne de Markov homogène, on a, pour tout n :*

$$P((X_{t_0}{=}x_0) \cap ... \cap (X_{t_n}{=}x_n)) = P(X_{t_0}{=}x_0)\, p_{01}\, p_{12} \,...\, p_{n-1,n}, \tag{7.14}$$

où seules interviennent la loi de X_{t_0} et la matrice de transition $[p_{ij}]$.

Preuve. Dans le cas d'une chaine de Markov la règle de multiplication (3.13) page 39 s'énonce simplement :

$$P(X_{t_0}{=}x_0, ..., X_{t_n}{=}x_n) = P(X_{t_0}{=}x_0)\, P_{|(X_{t_0}=x_0)}(X_{t_1}{=}x_1) \,...\, P_{|(X_{t_{n-1}}=x_{n-1})}(X_{t_n}{=}x_n).$$

Et (7.11) donne (7.14). ■

7.2.2 Matrice stochastique et matrice de transition

Définition 7.22 Soit I un ensemble au plus dénombrable. Une matrice stochastique sur I est un tableau matriciel $[p_{ij}]_{\substack{i\in I\\ j\in I}}$ tel que :

$$\forall i,j \in I, \quad p_{ij} \geq 0, \qquad \text{et} \qquad \forall i \in I, \quad \sum_{j\in I} p_{ij} = 1. \tag{7.15}$$

(La somme sur chaque ligne vaut 1.) $P_{X_{t_1}|(X_{t_0}=x_i)}$.)

Proposition 7.23 *La matrice de transition $[p_{ij}]$ d'une chaîne de Markov homogène est une matrice stochastique : pour tout $i \in \mathbb{N}$:*

$$\sum_{j\in I} p_{ij} = P_{|(X_{t_0}=x_i)}(\Omega) = 1. \tag{7.16}$$

Preuve. À i fixé, $\bigcup_{j\in I}\{X_{t_1}{=}x_j\}$ est une partition de Ω, et $P_{|(X_{t_0}=x_i)}$ est une probabilité. Donc $\sum_{j\in I} P_{|(X_{t_0}=x_i)}(X_{t_1}{=}x_j) = 1$. ▪

Exemple 7.24 Cas particulier X_{t_1} et X_{t_2} indépendantes de même loi (cas exclu en général dans l'étude de Markov) : $p_{ij} = P(X_{t_1}{=}x_j) = p_{kj}$ pour tout i,k,j : les colonnes de $[p_{ij}]$ sont toutes égales. ▪

Exemple 7.25 Marche aléatoire $(S_n)_{\mathbb{N}}$, cf. (7.5) : soit X_n v.a.r. t.q. $P_{X_n}(1) = p$ et $P_{X_n}(-1) = q = 1{-}p$, pour $n \in \mathbb{N}$. Soit $S_n = S_0 + \sum_{i=1}^n X_n = S_{n-1} + X_n$, où S_0 est un entier donné. Hypothèse de Markov homogène pour S_n, pour $n \geq 1$:

$$P_{|S_n=i}(S_{n+1}{=}j) = P_{|S_1=i}(S_2{=}j) = p_{ij} = \left\{ \begin{array}{r} p \text{ si } j = i{+}1 \\ q \text{ si } j = i{-}1 \\ 0 \text{ sinon} \end{array} \right\}.$$

▪

Exemple 7.26 Chaîne à deux états a et b (ou on-off, ou pile-face).

Intervalle de temps $[0,t_1[\cup[t_1,t_2[\cup ... \cup [t_n,t_{n+1}[$ (par exemple t_i = la i-ème seconde), et ligne téléphonique :

$$\left\{ \begin{array}{l} X_{t_i} = 0 \ \text{ si ligne libre au temps } t_i, \\ X_{t_i} = 1 \ \text{ si ligne occupée au temps } t_i. \end{array} \right. \tag{7.17}$$

Ici $\{x_1,x_2\} = \{0,1\}$ est l'ensemble des états de la ligne.

Soit p_{01} la probabilité pour que la ligne soit "libre à t_i et occupée à t_{i+1}" (probabilité indépendante de i), et soit p_{10} la probabilité pour que la ligne soit "occupée à t_i et libre à t_{i+1}" :

$$\left\{ \begin{array}{ll} P_{|(X_{t_n}=0)}(X_{t_{n+1}}{=}1) = p_{01} & (= P(X_{t_{n+1}}{=}1|X_{t_n}{=}0)), \\ P_{|(X_{t_n}=1)}(X_{t_{n+1}}{=}0) = p_{10} & (= P(X_{t_{n+1}}{=}0|X_{t_n}{=}1)). \end{array} \right. \tag{7.18}$$

(7.16) donne :

$$\left\{ \begin{array}{ll} P_{|(X_{t_n}=0)}(X_{t_{n+1}}{=}0) = 1{-}p_{01} = p_{00} & (= P(X_{t_{n+1}}{=}0|X_{t_n}{=}0)), \\ P_{|(X_{t_n}=1)}P(X_{t_{n+1}}{=}1) = 1{-}p_{10} = p_{11} & (= P(X_{t_{n+1}}{=}1|X_{t_n}{=}1)). \end{array} \right. \tag{7.19}$$

La matrice de transition est donc $\begin{pmatrix} 1{-}p_{01} & p_{01} \\ p_{10} & 1{-}p_{10} \end{pmatrix}$.

N.B. : attention : la somme sur les lignes donnent 1, pas la somme sur les colonnes en général : prendre $p_{01} = 1/2$ et $p_{10} = 9/10$. ▪

Exercice 7.27 Dans l'exemple 7.26 précédent, supposant que $P_{X_n}(0)$ est indépendant de n, montrer que :

$$P_{X_{t_n}}(0) = \frac{p_{10}}{p_{01}+p_{10}} \qquad \text{et} \qquad P_{X_{t_n}}(1) = \frac{1-p_{10}}{p_{01}+p_{10}},$$

probabilités que la ligne soit libre ou occupée à t_n.

Réponse. Posons $\alpha = P(X_{t_n}{=}0)$ et $\beta = P(X_{t_n}{=}1)$ pour tout n (avec donc $\alpha+\beta = 1$) (processus i.i.d).

On a $P(X_{t_{n+1}}{=}0) = P(X_{t_{n+1}}{=}0|X_{t_n}{=}0)P(X_{t_n}{=}0) + P(X_{t_{n+1}}{=}0|X_{t_n}{=}1)P(X_{t_n}{=}1)$ (Bayes) et donc $\alpha = (1{-}p_{01})\alpha + p_{10}\beta$, soit $p_{01}\alpha = p_{10}\beta$. Ayant $\alpha+\beta = 1$ on obtient $p_{01}\alpha = p_{10}(1{-}\alpha)$, d'où $\alpha = \frac{p_{10}}{p_{01}+p_{10}}$. ▪

Exemple 7.28 Chaîne à 3 états (file d'attente simplifiée). On reprend l'exemple 7.26 auquel on ajoute l'état d'attente noté état 2 : si la ligne est occupée à l'instant t_i et qu'un appel arrive, il est mis en attente (état 2) (et si un second appel arrive, il est perdu). Ici $\{x_1, x_2, x_3\} = \{0, 1, 2\}$ est l'ensemble des états de la ligne.

On suppose qu'un seul appel peut être détecter–libérer par intervalle de temps, ce qui donne :

* état initial libre :
$P(X_{t_{n+1}}{=}1|X_{t_n}{=}0) = p = p_{01}$, $P(X_{t_{n+1}}{=}2|X_{t_n}{=}0) = 0 = p_{02}$, d'où
$P(X_{t_{n+1}}{=}0|X_{t_n}{=}0) = 1{-}p = p_{00}$,

* état initial en attente :
$P(X_{t_{n+1}}{=}0|X_{t_n}{=}2) = 0 = p_{20}$, $P(X_{t_{n+1}}{=}1|X_{t_n}{=}2) = q = p_{21}$, d'où
$P(X_{t_{n+1}}{=}2|X_{t_n}{=}2) = 1{-}q = p_{22}$,

* état initial occupé :
$p_{10} = P(X_{t_{n+1}}{=}0|X_{t_n}{=}1) = q(1{-}p)$ (deux possibilités : libération de la ligne, ou pas de nouvel appel, événements supposés indépendants),
$p_{12} = P(X_{t_{n+1}}{=}2|X_{t_n}{=}1) = p(1{-}q)$ (deux possibilités : nouvel appel, ou pas de libération de ligne, événements supposés indépendants), d'où
$p_{11} = P(X_{t_{n+1}}{=}1|X_{t_n}{=}1) = 1 - p_{10} - p_{12}$.

La matrice de transition est $\begin{pmatrix} 1{-}p & p & 0 \\ q(1{-}p) & 1{-}q(1{-}p){-}p(1{-}q) & p(1{-}q) \\ 0 & q & 1{-}q \end{pmatrix}$. (Attention aux notations : ici $p + q \neq 1$ en général.) ▲

7.2.3 Modèle génétique simplifié (Wright et Fisher)

Modèle génétique simplifié (modèle de Wright et Fisher). On suppose que les caractéristiques d'une future génération ne dépend que des caractéristiques de la génération présente.

Remarque 7.29 Rappels rapides de génétique.

Chromosome (46 pour l'être humain) : structure constituée de molécules d'ADN. Les chromosomes sont regroupés par paires (23 pour l'être humain, dont 22 paires commune pour l'homme et la femme, et une paire XX chez la femme et XY chez l'homme).

Caryotype (caractéristique d'une espèce) = nombre de paires de chromosomes d'une espèce, classées par ordre de taille décroissante.

Gène : sous-structure sur un chromosome donné (une séquence d'ADN) qui renferme le patrimoine héréditaire. L'ensemble des gènes constitue le génotype. ▲

On suppose ici que les gènes ne sont que de deux types notés a et b.

Soit m le nombre de paires de gènes, donc de type $\{a, a\}$, $\{a, b\}$ ou $\{b, b\}$, et $n = 2m$ est le nombre de gènes présents pour une population donnée.

Soit E_N l'éprouvette contenant les paires de gènes de la génération N.

On note :

$$X_N = \text{nombre de gènes de type } a \text{ dans } E_N.$$

Passage E_N à E_{N+1} :

1- on extrait au hasard une paire de gènes de l'éprouvette E_N (parents), on fait une copie d'un des gènes choisi au hasard, et on remet la paire de gènes dans l'éprouvette E_N (qui contient donc toujours m paires de gènes).

2- On recommence une seconde fois la manipulation 1-..

3- On construit une paire de gènes à l'aide des deux copies obtenues en 1- et 2- : on a obtenu une paire de gènes "enfant", et on met cette paire dans l'éprouvette E_{N+1} (enfants).

4- On recommence m fois la démarche 1-, 2-, 3- : l'éprouvette E_{N+1} contient m paires de gènes.

Comme X_n ne s'intéresse pas aux paires de gènes mais aux gènes individuels, les hypothèses équivalent à : on fait n tirages (avec remise) indépendants d'un gène dans E_N, donc :

$$P_{|(X_N=i)}(X_{N+1}{=}j) = b(j; n, \frac{i}{n}) \overset{\text{noté}}{=} p_{ij} \qquad (= C_n^j(\frac{i}{n})^j(1 - \frac{i}{n})^{n-j})),$$

car n est le nombre de gènes dans E_n et, quand $X_N{=}i$, $\frac{i}{n}$ est la probabilité de présence du gène a dans E_N.

Donc p_{ij} est indépendant de N : on est dans le cas d'une chaîne de Markov homogène. La matrice de transition est la matrice $[p_{ij}]_{\substack{j\in\mathbb{N}^* \\ i\in\mathbb{N}^*}}$. Par exemple, dans le cas équiprobable $i = \frac{n}{2}$, $p = \frac{1}{2}$, et $p_{ij} = (\frac{1}{2})^n C_n^j$.

7.3 Problème de ruine du joueur

7.3.1 Position du problème

Soit $k \in \mathbb{N}^*$ le capital initial du joueur, $k \geq 1$ (sinon le jeu ne peut pas commencer).

Soit $N \in \mathbb{N}^*$ la somme d'argent totale en jeu, avec $N \geq k+1$, sinon l'adversaire n'a pas d'argent et le jeu ne peut pas commencer.

À chaque étape on ajoute 1 si le joueur gagne et on retire 1 si le joueur perd. La probabilité de gain est notée p.

Le jeu s'arrête quand le joueur on son adversaire n'a plus d'argent

7.3.2 Modélisation

Soit l'univers $\Omega = \{-1,1\}^{\mathbb{N}^*}$. Une issue est une suite $\vec{\omega} = (\omega_n)_{n\in\mathbb{N}^*}$ où $\omega_n =^{\text{noté}} X_n(\vec{\omega}) \in \{-1,1\}$ donne le résultat de la n-ème partie. Ainsi $(X_n)_{\mathbb{N}^*}$ est une marche aléatoire, cf. (7.4). L'hypothèse est : $(X_n)_{\mathbb{N}^*}$ est une marche aléatoire de paramètre $p \in]0,1[$, cf. (7.5), avec $q = 1-p$:

$$\begin{cases} P_{X_n}(1) = p & (= P(\{\vec{\omega} : X_n(\vec{\omega}) = \omega_n = 1\}), \\ P_{X_n}(-1) = q & (= P(\{\vec{\omega} : X_n(\vec{\omega}) = \omega_n = -1\}). \end{cases} \tag{7.20}$$

Soit $S_n : \Omega \to \mathbb{R}$ la v.a.r. argent du joueur à l'étape $n \in \mathbb{N}$:

$$S_n = S_{n-1} + X_n = S_0 + \sum_{i=1}^{n} X_i, \qquad \text{avec} \qquad S_0 = k, \tag{7.21}$$

donc, pour une issue $\vec{\omega} \in \Omega$, on a $S_n(\vec{\omega}) = S_{n-1}(\vec{\omega}) + \omega_n$, avec $S_0(\vec{\omega}) = k$.

Pour une issue $\vec{\omega}$ donnée, la suite $(S_n(\vec{\omega}))_{\mathbb{N}}$ donne l'argent du joueur après le n-ème tirage.

S'il existe $n \in \mathbb{N}$ t.q. $S_n = 0$, on dit que le joueur est ruiné (et que l'adversaire a gagné), et s'il existe $n \in \mathbb{N}$ t.q. $S_n = N$, on dit que le joueur a gagné (et que l'adversaire est ruiné)

Définition 7.30 Le "temps d'arrêt" d'une issue $\vec{\omega}$ est par définition "l'entier éventuellement infini" :

$$\begin{aligned} t_{k,N}(\vec{\omega}) &= \inf\{n \in \mathbb{N}^*,\ k + \omega_1 + ... + \omega_n = 0 \text{ ou } N\} \\ &= \inf\{n \in \mathbb{N}^*,\ S_n(\vec{\omega}) = 0 \text{ ou } N,\ S_0(\vec{\omega}) = k\} \overset{\text{noté}}{=} n_{\text{fin}}(\vec{\omega}). \end{aligned} \tag{7.22}$$

(Vaut $+\infty$ pour $k = 2$, $N = 4$ et la série alternée $\vec{\omega} = ((-1)^n)_{\mathbb{N}^*}$ par exemple).

Donc le jeu s'arrête dès que l'un des joueurs est ruiné : c'est le cas où $n_{\text{fin}}(\vec{\omega}) \in \mathbb{N}$ (est fini). Et, pour $\vec{\omega} \in \Omega$ donné, on ne considèrera que la suite $(S_n(\vec{\omega}))_{n=0,...,n_{\text{fin}}(\vec{\omega})}$.

Le processus (la suite de v.a.r. aléatoire) qui nous intéresse est la suite $(S_n)_{\mathbb{N}}$.

L'ensemble des états du processus $(S_n)_{\mathbb{N}}$ est $E = \{x_0, x_1, ..., x_N\} = \{0, 1, ..., N\}$ (les sommes d'argent possibles pour le joueur), cf. définition 7.2.

Hypothèse : $(S_n)_{\mathbb{N}}$ est un processus de Markov homogène :

$$\begin{aligned} P(S_{n+1}{=}x_{n+1} \mid S_n{=}x_n, ..., S_1{=}x_1) &= P(S_{n+1}{=}x_{n+1} \mid S_n{=}x_n) \\ &= P(S_1{=}x_{n+1} \mid S_0{=}x_n) = p(x_n, x_{n+1}). \end{aligned} \tag{7.23}$$

La matrice de transition est

$$\begin{pmatrix} 1 & 0 & ... & & & \\ q & 0 & p & 0 & ... & \\ 0 & q & 0 & p & 0 & ... \\ \vdots & & \ddots & & & \\ & & ... & q & 0 & p \\ & & & ... & 0 & 1 \end{pmatrix} \tag{7.24}$$

la première ligne correspond à $P_{|(S_0=0)}$ (le joueur n'a pas d'argent) : donc $P_{|(S_0=0)}(S_1 = 0) = 1$.

La deuxième ligne correspond à $P_{|(S_0=1)}$ (le joueur a 1 au début), et donc $p(1,0) = P_{|(S_0=1)}(S_1{=}0) = q$ (perte), $p(1,1) = P_{|(S_0=1)}(S_1{=}1) = 0$ (impossible), $p(1,2) = P_{|(S_0=1)}(S_1{=}2) = p$ (gain), $P_{|(S_0=1)}(S_1 > 2) = 0$ (impossible)...

De manière générique, pour $1 \le i \le N{-}1$ (correspond à la ligne i, la "première" ligne étant la ligne 0) :

$$p(i,j) = P_{|(S_0=i)}(S_1{=}j) = q\delta_{i,j+1} + p\delta_{i,j-1} \qquad (= P_{|(S_n=i)}(S_{n+1}{=}j)), \tag{7.25}$$

perte quand $j{+}1 = i$ (donc $j = i{-}1$), gain quand $j{-}1 = i$ (donc $j = i{+}1$), impossible sinon.

Et la dernière ligne quand $S_0 = N$: le jeu ne peut pas commencer, donc $P_{|(S_0=N)}(S_1 = N) = 1$.

7.3.3 Probabilité de ruine du joueur

Soit N fixé (somme totale en jeu). Soit A_k l'événement "ruine du joueur possédant le capital initial k" :

$$A_k = \{\vec\omega : n_{\text{fin}}(\vec\omega) < \infty\} = \{\vec\omega : t_{k,N}(\vec\omega) < \infty\}, \tag{7.26}$$

cf. (7.22). On note :

$$\begin{aligned} Q_k &\stackrel{\text{déf}}{=} P(A_k) = \text{probabilité de ruine du joueur} \\ &= P(\{\vec\omega : \exists n \in \mathbb{N}^*,\ n_{\text{fin}}(\vec\omega) = n\}. \end{aligned} \tag{7.27}$$

On ajoute les hypothèses :

$$\begin{cases} Q_0 = 1 & (= P_{|(S_0=0)}(S_1 = 0)), \\ Q_N = 0, & (= 1 - P_{|(S_0=N)}(S_1 = N)), \end{cases} \tag{7.28}$$

i.e. cas du joueur qui commence alors qu'il est déjà ruiné (perte certaine : le joueur a perdu avant de commencer), et cas de l'adversaire ruiné (victoire certaine : le joueur a gagné avant de commencer).

Proposition 7.31 *Pour $k \in \mathbb{N}^*$:*

$$Q_k = qQ_{k-1} + p\,Q_{k+1}, \tag{7.29}$$

équation aux différences d'ordre 2 résoluble avec (7.28).

Preuve. Soit $n \in \mathbb{N}$ fixé. On a $(S_n{=}0) = \{\vec\omega : n_{\text{fin}}(\vec\omega) = n\}$. Donc $Q_k = P(\bigcup_{\mathbb{N}}(S_n{=}0)|S_0{=}k) = \sum_{i=1}^{\infty} P(S_n{=}0|S_0{=}k)$.

La formule de Bayes (3.19) page 40 donne, pour $k \ge 1$:

$$\begin{aligned} P(S_n{=}0|S_0{=}k) &= \sum_{j=k\pm1} P(S_n{=}0|((S_0{=}k) \cap (S_1{=}j))\ P(S_1{=}j|S_0{=}k) \\ &= P(S_n{=}0|(S_1{=}k{-}1) \cap (S_0{=}k))\ P_{X_n}(-1) + P(S_n{=}0|(S_1{=}k{+}1) \cap (S_0{=}k))\ P_{X_n}(1) \end{aligned} \tag{7.30}$$

Et $(S_1{=}k{-}1) \subset (S_0{=}k)$, donc $(S_1{=}k{-}1) \cap (S_0{=}k) = (S_1{=}k{-}1)$, idem avec $k{+}1$, d'où :

$$P(S_n{=}0|S_0{=}k) = P(S_n{=}0|(S_1{=}k{-}1))\ q + P(S_n{=}0|(S_1{=}k{+}1))\ p,$$

d'où (7.29). ∎

Proposition 7.32 *(Probabilités de ruine des joueurs.)*

Si $p = q = \frac12$ (jeu équilibré), alors :

$$Q_k = 1 - \frac{k}{N}, \tag{7.31}$$

et si $\rho = \dfrac{q}{p} \ne 1$, alors :

$$Q_k = 1 - \frac{1-\rho^k}{1-\rho^N} \qquad (= \frac{\rho^k - \rho^N}{1-\rho^N}). \tag{7.32}$$

Et on vérifie que si P_k est la probabilité de victoire du joueur qui a un capital initial k (probabilité de ruine de l'adversaire), on a :

$$P_k + Q_k = 1, \tag{7.33}$$

et donc tôt au tard le joueur va gagner ou perdre : la probabilité d'atteindre un bord est $P_k + Q_k$ et vaut 1. Ou encore le temps d'arrêt $t_{k,N}(\vec\omega)$ est presque surement fini.

Ou encore, comme P_k est la probabilité de ruine de l'adversaire, tôt ou tard l'un des deux joueurs va être ruiné.

Preuve. Rappel : solution $(Q_k)_{k\in\mathbb{N}}$ pour l'équation (7.29) avec conditions initiales (7.28) :

* Cas $q\neq p$: deux suites solutions indépendantes de (7.29) sont données par la suite constante (1) et la suite $((\frac{q}{p})^k=\rho^k)_{k\in\mathbb{N}}$ (puisque $q(\frac{q}{p})^{k-1}+p(\frac{q}{p})^{k+1}=p(\frac{q}{p})^k+q(\frac{q}{p})^k=(\frac{q}{p})^k$), et la forme générale des solutions est donnée par la suite $(Q_k)_{k\in\mathbb{N}}$ vérifiant $Q_k=\alpha\,1+\beta\,\rho^k$, avec α et β données par les conditions initiales. Ici les conditions initiales sont données par (7.28), donc $Q_N=0=\alpha\,1+\beta\,\rho^N$ donne $\alpha=-\beta\rho^N$, et $Q_0=1=\alpha+\beta$ donne $\beta(1-\rho^N)=1$. D'où $Q_k=-\frac{\rho^N}{1-\rho^N}+\frac{1}{1-\rho^N}\rho^k$, d'où (7.32) ;

* Cas $q=p=\frac12$: deux solutions indépendantes de (7.29) sont données par la suite constante (1) et la suite $(k)_{k\in\mathbb{N}}$ (puisque $\frac12(k-1)+\frac12(k+1)=k$), et la forme générale des solutions est donnée par la suite $(Q_k)_{k\in\mathbb{N}}$ vérifiant $Q_k=\alpha\,1+\beta\,k$, avec α et β données par les conditions initiales. Ici (7.28) donne $Q_0=1=\alpha$, et $Q_N=0=\alpha+N\beta$ donne $\beta=-\frac1N$, d'où (7.31).

Et P_k est la probabilité de ruine de l'adversaire qui a un capital initial de $N-k$ et les probabilités respectives q et p de gagner $+1$ ou -1 à chaque étape, d'où si $p=q$ on a $P_k=1-\frac{N-k}{N}=\frac{k}{N}$, et si $p\neq q$ on a $P_k=\frac{(\frac1\rho)^{N-k}-(\frac1\rho)^N}{1-(\frac1\rho)^N}=\frac{\rho^k-1}{\rho^N-1}$. Donc $P_k+Q_k=1$, i.e. (7.33).

(Autre démonstration : ici $p=q=1$: écrire (7.29) comme $(p+q)Q_k=q\,Q_{k-1}+p\,Q_{k+1}$, donc $p(Q_{k+1}-Q_k)=q(Q_k-Q_{k-1})$, donc $(Q_{k+1}-Q_k)=\frac{q}{p}(Q_k-Q_{k-1})$, suite géométrique, donc $(Q_{k+1}-Q_k)=(\frac{q}{p})^k(Q_1-Q_0)$. Donc $Q_N-Q_0=\sum_{k=0}^{N-1}(\frac{q}{p})^k(Q_1-Q_0)$. Et les conditions initiales donnent $Q_N-Q_0=-1$...) ∎

Proposition 7.33 *Cas limite $N\to\infty$ (le joueur adversaire est infiniment riche = cas du casino). On a alors, pour k fixé :*

$$\begin{cases} p\leq q \;\Rightarrow\; Q_k=1 \quad \text{(ruine certaine)},\\ p>q \;\Rightarrow\; Q_k=(\dfrac{q}{p})^k. \end{cases}\tag{7.34}$$

Preuve. On a $\dfrac{k}{N}\underset{N\to\infty}{\longrightarrow}0$, et donc si $p=q$ on a $Q_k\underset{N\to\infty}{\longrightarrow}1$.

Si $p>q$ alors $\rho=\frac{q}{p}<1$ et $\rho^N\underset{N\to\infty}{\longrightarrow}0$, et donc $Q_k\underset{N\to\infty}{\longrightarrow}\rho^k$.

Si $p<q$ alors $\rho=\frac{q}{p}>1$ et $Q_k\simeq\frac{\rho^N}{\rho^N}=1$ au voisinage de $N=\infty$. ∎

7.3.4 Espérance de gain du joueur

k et N sont fixés.

Corollaire 7.34 *Le gain du joueur à la fin de la partie est une v.a.r. G qui prend la valeur $N-k$ avec la probabilité $P_k=1-Q_k$ (gain de $N-k$), et la valeur $-k$ avec la probabilité Q_k (perte de k) :*

$$P(G{=}-k)=Q_k,\qquad P(G{=}N-k)=P_k.\tag{7.35}$$

Preuve. C'est la probabilité de ruine du joueur ou de son adversaire. ∎

Corollaire 7.35 *L'espérance de gain du joueur est :*

$$\begin{aligned} E_G=&\;-kQ_k+(N-k)P_k\\ =&\;NP_k-k, \end{aligned}\tag{7.36}$$

Et si on note $E_G=E_G(p)$ l'espérance de gain pour la probabilité p, alors E_G est une fonction croissante de p, avec alors $E_G(\frac12)=0$ (l'espérance de gain est nulle quand le jeu est équilibré).

Preuve. Le joueur perd k avec probabilité Q_k et gagne $N-k$ avec probabilité $P_k=1-Q_k$, d'où $E_G=Q_k(-k)+P_k(N-k)=-k(1-P_k)+(N-k)P_k$, d'où (7.36).

Si $p=q=\frac12$ alors (7.31) donne $E(G)=0$.

Sinon, comme E_G varie comme P_k, étudions $P_k : p \to P_k(p)$ sur $]0,1[$ (on exclu les cas triviaux $p=0$ et 1). On a, cf. (7.32) :

$$P_k = \frac{(\frac{q}{p})^k - 1}{(\frac{q}{p})^N - 1} = \frac{x(p)-1}{x(p)^{1+\beta}-1} \overset{\text{noté}}{=} f(x(p)), \qquad \text{où} \qquad \beta = \frac{N}{k} - 1 > 0,$$

et où :

$$f(y) = \frac{y-1}{y^{1+\beta}-1} \qquad \text{et} \qquad x(p) = (\frac{1-p}{p})^k \geq 0.$$

Montrons que f et x sont décroissantes ($f' \leq 0$ et $x' \leq 0$), et on aura bien P_k croissante (car $P_k'(p) = f'(x(p))x'(p) \geq 0$). On a pour $y \in \mathbb{R}_+$:

$$f'(y) = \frac{g(y)}{(y^{1+\beta}-1)^2},$$

où :

$$g(y) = y^{1+\beta} - 1 - (y-1)(1+\beta)y^\beta = -\beta y^{1+\beta} + (1+\beta)y^\beta - 1.$$

Donc :

$$g'(y) = -\beta(1+\beta)y^\beta + \beta(1+\beta)y^{\beta-1} = \beta(1+\beta)y^{\beta-1}(1-y).$$

Donc g' est positive sur $[0,1[$ et négative sur $]1,\infty[$, avec $g(1)=0$, donc g est négative sur $\mathbb{R}_+$, donc f' est négative sur $\mathbb{R}_+$, donc f est décroissante sur $\mathbb{R}_+$. On a $p \in]0,1[$ donc $\frac{1}{p} > 1$ et donc :

$$x'(p) = -\frac{k}{p^2}(\frac{1}{p} - 1)^{k-1} \leq 0.$$

Donc P_k est croissante comme fonction de p. Donc E_G est croissante comme fonction de p. ▪

7.3.5 Durée moyenne de la partie

Soit N fixé. Pour $k \in [1, N-1]$, on dispose de la v.a.r. $t_k \overset{\text{déf}}{=} t_{k,N} : \Omega \to [0,\infty]$ donnant le temps d'arrêt, cf. (7.22). Notons $T_k = E(t_k)$ la durée moyenne de la partie pour le capital initial $k \in \mathbb{N}^*$ (l'espérance de la v.a.r. t_k).

Proposition 7.36 *La durée moyenne T_k de la partie vérifie l'équation aux différences :*

$$T_k = qT_{k-1} + pT_{k+1} + 1, \qquad T_0 = T_N = 0. \tag{7.37}$$

D'où :

$$\begin{cases} T_k = k(N-k), & \textit{si} \quad p = q = \dfrac{1}{2}, \\ T_k = \dfrac{1}{1-2p}(k - N\dfrac{1-(\frac{q}{p})^k}{1-(\frac{q}{p})^N}), & \textit{sinon.} \end{cases} \tag{7.38}$$

Preuve. On prend de l'avance : voir le § 10.4 pour les espérances de v.a.r..

Si $k=0$ alors $t_k = 0$, donc $T_k = 0$: la partie ne commence pas. Idem pour $k = N$.

On a :

$$\begin{aligned} P(t_{k-1}{=}i) &= P(\{\vec\omega : n_{\text{fin}}(\vec\omega) = i\}) \\ &= P(\{\vec\omega : \inf\{m : k{-}1{+}\omega_1{+}...{+}\omega_m = 0 \text{ ou } N\} = i\}) \\ &= P(\{\vec\omega : \inf\{m : k{-}1{+}X_1(\vec\omega){+}...{+}X_m(\vec\omega) = 0 \text{ ou } N\} = i\}), \end{aligned} \tag{7.39}$$

Et, cf. (10.7) :

$$E(t_{k-1}) = \sum_{i \geq k-1} i\, P(t_{k-1}{=}i) \tag{7.40}$$

Et, les v.a.r. X_i étant indépendantes :

$$
\begin{aligned}
&P(t_k{=}j\ ,\ X_1{=}{-1})\\
&= P(\{\vec\omega :\ X_1(\vec\omega) = -1 \text{ et } \inf\{n : k{-}1{+}X_2(\vec\omega){+}...{+}X_n(\vec\omega){=}0 \text{ ou } N\} = j\})\\
&= P(\{\vec\omega :\ X_1(\vec\omega) = -1\})\ P(\{\vec\omega : \inf\{n : k{-}1{+}X_2(\vec\omega){+}...{+}X_n(\vec\omega) = 0 \text{ ou } N\} = j\})\\
&= q\ P(\{\vec\omega : \inf\{n : k{-}1{+}X_1(\vec\omega){+}...{+}X_{n-1}(\vec\omega) = 0 \text{ ou } N\} = j\})\\
&= q\ P(\{\vec\omega : 1 + \inf\{n{-}1 : k{-}1{+}X_1(\vec\omega){+}...{+}X_{n-1}(\vec\omega) = 0 \text{ ou } N\} = j\})\\
&= q\ P(\{\vec\omega : \inf\{n{-}1 : k{-}1{+}X_1(\vec\omega){+}...{+}X_{n-1}(\vec\omega) = 0 \text{ ou } N\} = j{-}1\})\\
&= q\ P(\{\vec\omega : \inf\{m : k{-}1{+}X_1(\vec\omega){+}...{+}X_m(\vec\omega) = 0 \text{ ou } N\} = j{-}1\})\\
&= q\ P(t_{k-1}{=}j{-}1).
\end{aligned}
$$

D'où, cf. (10.3) et (13.8) :

$$
\begin{aligned}
E(t_k|X_1{=}{-1}) = P_{(X_1=-1)}(t_k) &= \sum_{j\geq k} j\ P_{(X_1=-1)}(t_k{=}j)\\
&= \sum_{j\geq k} j\ \frac{P(t_k{=}j\ ,\ X_1{=}{-1})}{P(X_1{=}{-1})} = \sum_{j\geq k} j\ P(t_{k-1}{=}j{-}1)\\
&= \sum_{i\geq k-1} (i{+}1)\ P(t_{k-1}{=}i) = E(t_{k-1}) + 1.
\end{aligned}
$$

De même $E(t_k|X_1{=}{+1}) = E(t_{k+1}) + 1$. D'où, cf. (13.7) :

$$
\begin{aligned}
E(t_k)\ &= E(t_k|X_1{=}{-1})\ P(X_1{=}{-1}) + E(t_k|X_1{=}{+1})\ P(X_1{=}{+1})\\
&= q(E(t_{k-1}) + 1) + p(E(t_{k+1}) + 1)
\end{aligned}
\tag{7.41}
$$

D'où (7.37).

Cas $p = \frac12 = q$: une solution particulière est $-k^2$. En effet $-q(k{-}1)^2 - p(k{+}1)^2 + 1 = -k^2$. D'où la solution générale $T_k = \alpha + \beta k - k^2$. Puis $T_0 = 0$ donne $\alpha = 0$, et $T_N = 0$ donne $\beta = N$.

Cas $p \neq \frac12$ (donc $q - p = 1 - 2p \neq 0$) : une solution particulière est $\frac{-k}{1-2p}$. En effet $-q\frac{k-1}{1-2p} - p\frac{k+1}{1-2p} + 1 = \frac{-k}{1-2p}$. D'où la solution générale $T_k = \alpha + \beta(\frac{q}{p})^k + \frac{-k}{1-2p}$. Puis $T_0 = 0$ donne $\alpha = -\beta$, et $T_N = 0$ donne $\beta((\frac{q}{p})^N - 1) = \frac{N}{1-2p}$. ▪

7.4 Processus continus de comptage

Soit $T > 0$ et soit l'univers $\Omega = \mathcal{F}([0,T];\mathbb{R})$, l'ensemble des fonctions $\omega : t \in [0,T] \to \omega(t) \in \mathbb{R}$. Soit P une probabilité sur $(\mathbb{R}, \mathcal{T}_\mathbb{R})$.

Soit $(X_t)_{t\in[0,T]}$ un processus stochastique continu (famille de v.a.r. $\Omega \to \mathbb{R}$).

Exemple $X_t = \delta_t$, où donc $X_t(\omega) = \omega(t)$, cf. (7.1), qui donnera un processus de comptage.

On suppose $X_0 = 0$.

7.4.1 Processus à accroissements indépendants

Définition 7.37 Pour un processus $(X_t)_{t\in[0,T]}$ avec $X_0 = 0$, et pour $h > 0$, la v.a.r. $X_{t+h} - X_t$ est appelée l'accroissement du processus sur $]t, t{+}h]$.

(Et $\frac{X_{t+h}-X_t}{h}$ est la v.a.r. taux d'accroissement du processus.)

Soit $E = \{x_1, ..., x_n, ...\} = (x_i)_{i\in\mathbb{N}^*}$ l'ensemble des états (ensemble discret de réels) du processus, où donc $\mathrm{Im}X_t \subset E$ pour tout t : pour tout $t \in [0,T]$ et tout $\omega \in \Omega$ il existe $i \in \mathbb{N}$ t.q. $X_t(\omega) = x_i$.

Définition 7.38 Le processus $(X_t)_{t\geq 0}$ est à accroissements indépendants (processus sans mémoire) ssi, pour tout $t, h, k > 0$, les v.a.r. d'accroissement $X_{t+h} - X_t$ et $X_t - X_{t-k}$ sont indépendantes (accroissements indépendants). Autrement dit, pour tout $t, h, k > 0$ et tout $m, n \in \mathbb{N}^*$:

$$P_{|X_t - X_{t-k}=x_m}(X_{t+h}{-}X_t{=}x_n) = P(X_{t+h}{-}X_t = x_n). \tag{7.42}$$

Exemple 7.39 Marche aléatoire. ▪

7.4.2 Hypothèse de Markov en temps continu, espace d'états discret

L'hypothèse de Markov "en temps continu", est, pour tout $t,h,k>0$ et tout $x_i,x_j,x_m\in E$:

$$P_{|(X_t=x_i,\,X_{t-k}=x_m)}(X_{t+h}{=}x_j) = P_{|(X_t=x_i)}(X_{t+h}{=}x_j), \tag{7.43}$$

soit $P(X_{t+h}{=}x_j|X_t{=}x_i,\ X_{t-k}{=}x_m) = P(X_{t+h}{=}x_j|X_t{=}x_i)$. Donc ne dépend pas des temps $<t$.

Et le processus est homogène ssi, pour tout $t,h>0$ et tout $x_i,x_j\in E$:

$$P_{|(X_t=x_i)}(X_{t+h}{=}x_j) = P_{|(X_0=x_i)}(X_h{=}x_j), \tag{7.44}$$

soit $P(X_{t+h}{=}x_j|X_t{=}x_i) = P(X_h{=}x_j|X_0{=}x_i)$.

On note alors, pour tout i,j,t :

$$p_{ij}(t) = P_{|(X_0=x_i)}(X_t{=}x_j) \qquad (= P(X_t{=}x_j|X_0{=}x_i)). \tag{7.45}$$

Remarque 7.40 Soit (X_t) un processus de Markov en temps continu homogène, cf. (7.44). Discrétisons $[0,T]$ en intervalles égaux : soit $n\in\mathbb{N}^*$, soit $h=\frac{T}{n}$ et soit $t_i=ih$ pour $i=0,...,n$. Et soit $(X_{t_i})_{i\in[0,n]_\mathbb{N}}$ le processus discret défini à partir du processus continu homogène ci-dessus. On a alors :

$$P_{|(X_{t_k}=x_i)}(X_{t_{k+1}}{=}x_j) = P_{|(X_0=x_i)}(X_h{=}x_j) = p_{ij}(h) \overset{\text{noté}}{=} p_{ij}, \tag{7.46}$$

et on retrouve un processus de Markov homogène à temps discret avec sa matrice de transition $[p_{ij}]$, cf. (7.13). ▪

7.4.3 Fonction et processus de comptage

Définition 7.41 Une fonction de comptage est une fonction $N:\mathbb{R}_+\to\mathbb{N}$ croissante, en escalier (à valeurs discrètes), continue à droite (donc $N(t_+)=N(t)$ pour tout t), avec $N(0)=0$, et telle que les sauts valent 1 (on monte une marche à la fois).

Donc N est une fonction de comptage ssi il existe $n\in\mathbb{N}^*$, il existe $0=t_0<t_1<...<t_n=T$ (donc $[0,T]=\bigcup_{k=1}^n[t_k-t_{k-1}]$), il existe $n_0\in\mathbb{N}\cup\{-1\}$ t.q. :

$$N(t) = \sum_{k=1}^{n-1} n_k 1_{[t_k,t_{k+1}[}, \qquad n_k = n_0{+}k. \tag{7.47}$$

Exemple 7.42 $N(t)$ est le nombre de particules qui sont détectées par un capteur donné pendant l'intervalle de temps $[0,t]$. Ou le nombre d'appels téléphoniques à un standard dans cet intervalle. Ou le nombre de clients à un guichet dans cet intervalle... ▪

Définition 7.43 Soit $(X_t)_{t\in\mathbb{R}}$ un processus continu d'ensemble des états $E=\mathbb{N}$. C'est un processus (stochastique) de comptage ssi pour tout $\omega\in\Omega$ la trajectoire de ω est une fonction de comptage. Donc, pour tout $\omega\in\Omega$:

$$\tilde X_\omega : \left\{\begin{array}{ll} \mathbb{R}_+ & \to \mathbb{N} \\ t & \to \tilde X_\omega(t) \overset{\text{déf}}{=} X_t(\omega) \end{array}\right\} \quad \text{est une fonction de comptage.} \tag{7.48}$$

En particulier, pour tout $\omega\in\Omega$:

$$\tilde X_\omega(0) = 0 = X_0(\omega), \tag{7.49}$$

et la fonction X_0 est la fonction nulle.

7.4.4 Processus de comptage stationnaire

Définition 7.44 Un processus de comptage $(X_t)_{t\in\mathbb{R}^*}$ est stationnaire (ou à homogénéité temporelle) ssi, pour tout $t,h>0$, les lois de probabilités des v.a.r. $X_{t+h}-X_t$ et X_h-X_0 $(=X_h)$ sont les mêmes : pour tout $t,h>0$:

$$P_{X_{t+h}-X_t} = P_{X_h-X_0} \qquad (= P_{X_h}). \tag{7.50}$$

7.4.5 Processus de Poisson

Définition 7.45 Un processus de comptage $(X_t)_{t\in\mathbb{R}^*}$ est dits à événements rares ssi :

$$\lim_{h\to 0+} P(X_h \geq 1) = 0 \qquad \text{et} \qquad \lim_{h\to 0+} \frac{P(X_h > 1)}{P(X_h = 1)} = 0. \tag{7.51}$$

Définition 7.46 Un processus de comptage $(X_t)_{t\in\mathbb{R}^*}$ est un processus de Poisson (simple) ssi il est :

(i) à accroissements indépendants,
(ii) stationnaire, et
(iii) à événements rares.

Proposition 7.47 *Dans la définition précédente, il est équivalent de remplacer (iii) par :*
(iii)' la loi de la variable aléatoire X_t est la loi de Poisson de paramètre λt :

$$\exists \lambda > 0 \ t.q.\ \forall t > 0,\ \forall k \in \mathbb{N},\ P_{X_t}(k) = \frac{\lambda^k t^k}{k!} e^{-\lambda t}. \tag{7.52}$$

Preuve. Supposons (iii)' et montrons (iii). En particulier $P_{X_t}(0) = e^{-\lambda t}$, et donc $\longrightarrow_{t\to 0} 1$. Donc $P(X_t > 0) = 1 - P_{X_t}(0) \longrightarrow_{t\to 0} 0$, i.e. $(7.51)_1$.

Puis $P_{X_t}(0) = e^{-\lambda t} = 1 - \lambda t + o(t)$ au voisinage de t=0 avec $P_{X_t}(1) = \lambda t e^{-\lambda t} = \lambda t(1 - \lambda t + o(t))$ au voisinage de t=0, d'où $P(X_t > 1) = 1 - P_{X_t}(0) - P_{X_t}(1) = \lambda^2 t^2 + o(t^2)$. D'où $\lim_{h\to 0+} \dfrac{P(X_h > 1)}{P(X_h = 1)} \sim h$ au voisinage de 0.

Réciproquement supposons (i)+(ii)+(iii) et montrons (iii)'.

Posons $f_k(t) = P_{X_t}(k)$ $(= P(X_t{=}k))$. Il s'agit de montrer que $f_k(t) = \frac{\lambda^k t^k}{k!} e^{-\lambda t}$. On a, avec les hypothèses (i) et (ii) :

$$\begin{aligned}
f_k(t+h) = P(X_{t+h}{=}k) &= \sum_{i=0}^{k} P((X_{t+h}{=}k) \cap (X_t{=}i)) = \sum_{i=0}^{k} P((X_t{=}i) \cap (X_{t+h}{=}k)) \\
&= \sum_{i=0}^{k} P((X_t{=}i) \cap (X_{t+h}{-}X_t{=}k{-}i)) = \sum_{i=0}^{k} P((X_t{-}X_0{=}i) \cap (X_{t+h}{-}X_t{=}k{-}i)) \\
&= \sum_{i=0}^{k} P(X_t{-}X_0{=}i) P(X_h{-}X_0{=}k{-}i) = \sum_{i=0}^{k} P(X_t{=}i) P(X_h{=}k{-}i) \\
&= \sum_{i=0}^{k} f_i(t) f_{k-i}(h).
\end{aligned} \tag{7.53}$$

En particulier pour $k = 0$ (un seul terme avec $i = 0$) on a $f_0(t+h) = f_0(t) f_0(h)$. Donc il existe $a \in \mathbb{R}$ t.q. $f_0(t) = e^{at}$, voir exercice suivant 7.48.

Et comme f est définie sur $\mathbb{R}_+$ à valeurs dans $[0,1]$ et est non nulle, on a $a < 0$. On pose $\lambda = -a > 0$.

Montrons par récurrence que $f_k(t) = \frac{\lambda^k t^k}{k!} e^{-\lambda t}$. C'est vrai pour $k = 0$, et supposons-le jusqu'à k. Montrons que f_{k+1} vérifie l'équation différentielle :

$$f'_{k+1}(t) + \lambda f_{k+1}(t) = \lambda f_k(t), \qquad \text{i.e.} \qquad \lim_{h\to 0} \frac{f_{k+1}(t+h) - f_{k+1}(t)}{h} = \lambda(f_k(t) - f_{k+1}(t)). \tag{7.54}$$

Avec (7.53) on a :

$$\begin{aligned}
f_{k+1}(t+h) &= \sum_{i=0}^{k+1} f_i(t) f_{k+1-i}(h) \\
&= \sum_{i=0}^{k-1} f_i(t) f_{k+1-i}(h) + f_k(t) f_1(h) + f_{k+1}(t) f_0(h).
\end{aligned}$$

Et $(7.51)_2$ donne, pour $m \geq 2$, $f_m(h) = P(X_h{=}m) = o(P(X_h{=}1)) = o(f_1(h))$, d'où :

$$f_{k+1}(t+h) = f_k(t) f_1(h) + f_{k+1}(t) f_0(h) + o(f_1(h)).$$

Et $f_0(h) = P_{X_h}(0) = 1 - P(X_h{\ge}1) = 1 - P(X_h{=}1) + o(P(X_h{=}1)) = 1 - f_1(h) + o(f_1(h))$. D'où :

$$f_{k+1}(t+h) = f_k(t)f_1(h) + f_{k+1}(t)(1 - f_1(h) + o(f_1(h))) + o(f_1(h)).$$

Donc :

$$\frac{f_{k+1}(t+h) - f_{k+1}(t)}{h} = \frac{f_1(h)}{h}(f_k(t) - f_{k+1}(t) + o(1)] \underset{h\to 0}{\longrightarrow} (f_k(t) - f_{k+1}(t)) \lim_{h\to 0} \frac{f_1(h)}{h}.$$

Avec $f_1(h) = P(X_h{=}1) = 1 - P(X_h{=}1) - P(X_h{>}1) = 1 - f_0(h) + f_1(h)o(1)$, voir ci-dessus, donc $\lim_{h\to 0} \frac{1 - f_0(h)}{h} = \lim_{h\to 0} \frac{f_1(h)(1 + o(1))}{h}$. Comme $f_0(h) = e^{-\lambda h}$, on a $\lim_{h\to 0} \frac{f_0(h) - 1}{h} = f_0'(0) = -\lambda$, d'où $\lim_{h\to 0} \frac{f_1(h)}{h} = \lambda$. D'où (7.54), qu'on résout par la méthode de variation de la constante, voir exercice suivant 7.49, d'où (iii)'. ▪

Exercice 7.48 Montrer : une fonction continue strictement positive $f : \mathbb{R} \to \mathbb{R}$ vérifie $f(0) = 1$ et $f(x + y) = f(x)f(y)$, pour tout $x, y \in \mathbb{R}$, ssi il existe $a \in \mathbb{R}$ tel que $f(x) = e^{ax}$.

Réponse. $\Leftarrow$. Si $a = 0$ alors $f = 1_\mathbb{R}$ vérifie bien $f(x+y) = f(x)f(y)$, et si $a \neq 0$ on a bien $e^{a(x+y)} = e^{ax}e^{ay}$.

$\Rightarrow$. Supposons $f > 0$ sur $\mathbb{R}$. On pose $g(x) = \log(f(x))$. Montrons que g est linéaire. Pour $n \in \mathbb{N}^*$ on a $g(nx) = \log(f(x)^n) = n\log(f(x)) = ng(x)$, et en particulier on a $g(0) = 0g(0) = 0$. Et avec $x = \frac{1}{n}$ on a $\frac{1}{n}g(1) = g(\frac{1}{n})$, soit $g(\frac{1}{n}) = \frac{1}{n}g(1)$ donc pour tout $m \in \mathbb{N}^*$ on a $g(\frac{m}{n}) = \frac{m}{n}g(1)$. Donc pour tout $r \in \mathbb{Q}$ on a $g(r) = rg(1)$. Et g est continue car f et log le sont, donc $g(x) = xg(1)$ pour tout $x \in \mathbb{R}$. Donc g est linéaire, donc de la forme $g(x) = ax$. Et tout $a \in \mathbb{R}^*$ convient. ▪

Exercice 7.49 Résoudre l'équation différentielle $u'(t) + \lambda u(t) = \lambda \frac{\lambda^k t^k}{k!} e^{-\lambda t}$ sur $\mathbb{R}$.

Réponse. 1- Solution homogène $u_h(t) = e^{-\lambda t}$: on a bien $u_h'(t) + \lambda u_h(t) = 0$.

2- Solution particulière : $u_p(t)$ cherchée sous la forme $u_p(t) = c(t)e^{-\lambda t}$. Donc $u_p'(t) + \lambda u_p(t) = \lambda \frac{\lambda^k t^k}{k!} e^{-\lambda t}$ donne $c'(t) = \lambda \frac{\lambda^k t^k}{k!}$. D'où (à une constante près) $c(t) = \frac{\lambda^{k+1}}{k!} \int_0^t \tau^k \, d\tau = \frac{\lambda^{k+1}\tau^{k+1}}{(k+1)!}$. D'où $u_p(t) = \frac{\lambda^{k+1}\tau^{k+1}}{(k+1)!} e^{-\lambda t}$ est une solution particulière.

3- Solution générale : $u(t) = cu_h(t) + u_p(t)$ où $c \in \mathbb{R}$ (fixée par la donnée d'une condition initiale). ▪

8 Loi de sommes et de produits

8.1 Loi d'une somme : convolution discrète si indépendance

8.1.1 Loi d'une somme : cas discret

On rappelle que, pour $f, g : \mathbb{Z} \to \mathbb{R}$, le produit de convolution discret $f * g$ est défini pour $z \in \mathbb{Z}$ par, quand il a un sens :

$$(f * g)(z) \overset{\text{déf}}{=} \sum_{k\in\mathbb{Z}} f(k)g(z{-}k) = \sum_{\substack{k,\ell\in\mathbb{Z} \\ k+\ell=z}} f(k)g(\ell). \tag{8.1}$$

Le produit de convolution est trivialement commutatif (quand il est défini).

Par exemple $(f * g)(0) = \sum_{k\in\mathbb{Z}} f(k)g(-k)$

Proposition 8.1 *Soient X et Y deux v.a.r. à valeurs dans $\mathbb{Z}$ définies sur un même espace probabilisé $(\Omega, \mathcal{T}, P)$, de lois respectives P_X et P_Y. Alors $X{+}Y$ est une v.a.r. à valeurs dans $\mathbb{Z}$ sur l'espace probabilisé $(\Omega, \mathcal{T}, P)$ de loi P_{X+Y} donnée par, pour $z \in \mathbb{Z}$:*

$$P_{X+Y}(z) = \sum_{k\in\mathbb{Z}} P((X{=}k) \cap (Y{=}z{-}k)). \tag{8.2}$$

Si de plus X et Y sont indépendantes alors :

$$P_{X+Y} = P_X * P_Y, \qquad \text{quand } X \text{ et } Y \text{ indépendantes,} \tag{8.3}$$

i.e. pour tout $z \in \mathbb{Z}$:

$$P_{X+Y}(z) = \sum_{k\in\mathbb{Z}} P_X(k)P_Y(z{-}k) \qquad (= (P_X * P_Y)(z)). \tag{8.4}$$

Preuve. Soit $z \in \mathbb{Z}$. Soit l'événement $A_z = \{\omega \in \Omega : (X{+}Y)(\omega) = z\}$. Soit $P(A_z) = P_{X+Y}(z)$ sa probabilité. On a $A_z = \{\omega \in \Omega : \exists k \in \mathbb{Z},\ X(\omega) = k,\ Y(\omega) = z{-}k\}$, soit $A_z = \bigcup_{k\in\mathbb{Z}}((X{=}k) \cap (Y{=}z{-}k))$. C'est une union d'ensembles disjoints, d'où $P(A_z) = \sum_{k\in\mathbb{Z}} P((X{=}k) \cap (Y{=}z{-}k))$.

Et quand les v.a.r. sont indépendantes, on obtient (8.4). ▪

Exemple 8.2 On lance deux fois un dé, X donne la valeur du premier lancer et Y celle du second lancer. X et Y étant supposées indépendantes, on a $P_{X+Y}(3) = P_X(1)P_Y(2) + P_X(2)P_Y(1) = \frac{2}{36}$.

Détails : ici $\Omega = [1,6]_{\mathbb{N}}^2$, $P_X(k) = P(\{\vec{\omega} = (\omega_1, \omega_2) : \omega_1 = k\})$... ▪

8.1.2 Loi d'une somme : cas continu

Pour $f, g : \mathbb{R} \to \mathbb{R}$, le produit de convolution est défini formellement par :

$$(f * g)(x) = \int_{t\in\mathbb{R}} f(t)g(x{-}t)\,dt. \tag{8.5}$$

Soit $X, Y : \Omega \to \mathbb{R}$ deux v.a.r. de loi de probabilité de densité ρ_X et ρ_Y. Ainsi :

$$P_X(]a,b[) = \int_a^b \rho_X(t)\,dt \qquad (= P(X \in]a,b[)),$$

où $\rho_X : \mathbb{R} \to \mathbb{R}$ est positive et intégrable telle que $\int_{\mathbb{R}} \rho_X(t)\,dt = 1$. De même pour Y.

Proposition 8.3 *Si X et Y sont indépendantes, la loi P_{X+Y} de $X{+}Y$ est la loi de densité :*

$$\rho_{X+Y} = \rho_X * \rho_Y, \qquad \textit{quand } X \textit{ et } Y \textit{ indépendantes} \tag{8.6}$$

(convolution des densités), soit, pour tout $t \in \mathbb{R}$:

$$\rho_{X+Y}(t) = (\rho_X * \rho_Y)(t) = \int_{\tau\in\mathbb{R}} \rho_X(\tau)\rho_Y(t{-}\tau)\,d\tau. \tag{8.7}$$

En particulier :

$$P_{X+Y}(]a,b[) = \int_{t=a}^b \rho_{X+Y}(t)\,dt = \int_{t=a}^b \int_{\tau\in\mathbb{R}} \rho_X(\tau)\rho_Y(t{-}\tau)\,d\tau\,dt. \tag{8.8}$$

Preuve. On a $P_{X+Y}(]a,b[) = P(X+Y \in [a,b])$, et :

$$\begin{aligned}(X+Y \in [a,b]) &= \{\omega \in \Omega : x = X(\omega) \in \mathbb{R},\ y = Y(\omega) \in \mathbb{R},\ x+y \in [a,b]\} \\ &= \{\omega \in \Omega : x = X(\omega) \in \mathbb{R},\ y = Y(\omega) \in [a{-}x, b-x]\} \\ &= \{(X,Y) \in D,\ D = \{(x,y) : x \in \mathbb{R},\ y \in [a{-}x, b-x]\},\end{aligned}$$

D étant la "bande oblique" de $\mathbb{R}^2$ limité par les droites $y = a{-}x$ et $y = b{-}x$, faire un dessin. D'où, les v.a.r. X et Y étant indépendantes et avec Fubini (les densités sont positives) :

$$\begin{aligned}P_{X+Y}(]a,b[) &= P_{X,Y}(D) = (P_X \otimes P_Y)(x \in \mathbb{R},\ y \in [a-x, b-x]) \\ &= \int_{x=-\infty}^{\infty} \int_{y=a-x}^{b-x} \rho_X(x)\rho_Y(y)\,dy\,dx = \int_{x=-\infty}^{\infty} \int_{t=a}^{b} \rho_X(x)\rho_Y(t-x)\,dt\,dx \\ &= \int_{t=a}^{b} \int_{x=-\infty}^{\infty} \rho_X(x)\rho_Y(t-x)\,dx\,dt = \int_{t=a}^{b} (\rho_X * \rho_Y)(t)\,dt.\end{aligned}$$

▪

Remarque 8.4 De manière générique, étant donnée une mesure μ, le produit de convolution de deux fonctions f et g relativement à μ est formellement défini par $(f * g)(x) = \int_{t\in\mathbb{R}} f(t)g(x{-}t)\,d\mu$, soit :

$$(f * g)(x) = \mu(f\,\tau_x\check{g}) \tag{8.9}$$

mesure de la fonction produit de f par $\tau_x\check{g}$ où, pour une fonction h générique, $\check{h}(t) =^{\text{déf}} h(-t)$ (symétrie par rapport à l'axe vertical) et $\tau_x h(t) =^{\text{déf}} h(t{-}x)$ (translation).

Ainsi $(\tau_x\check{g})(t) = \check{g}(t-x) = g(x-t)$.

Ici la mesure μ était soit la mesure de comptage $\mu = \sum_{k\in\mathbb{Z}} \delta_k$ pour le cas discret, soit la mesure de Lebesgue $\mu = dx$ pour le cas continu. ▪

8.1.3 Convolution de deux mesures

On renvoie à un cours de théorie de la mesure pour les détails, ou bien au cours de distribution (§ convolution de distributions), les mesures étant des distributions particulières (celles d'ordre 0).

Soit la fonction somme :
$$S(x,y) = x+y. \tag{8.10}$$

Définition 8.5 Si μ et ν sont deux mesures sur une tribu $\mathcal{T}$ de $\mathbb{R}$, leur convolution $\mu*\nu$ est formellement l'image de la mesure produit $\mu\otimes\nu$ par S, cf. (4.9) :
$$\mu*\nu = (\mu\otimes\nu)\circ S^{-1}, \tag{8.11}$$
soit, pour tout ensemble $B\in\mathcal{T}$:
$$(\mu*\nu)(B) \overset{\text{déf}}{=} (\mu\otimes\nu)(\{(x,y)\in\mathbb{R}^2 :\ x+y\in B\}). \tag{8.12}$$

Définition alternative sur les fonctions (en passant par les fonctions en escalier, voir remarque suivante 8.7) :

Définition 8.6 Si μ et ν sont deux mesures, leurs convolution $\mu*\nu$ est formellement définie par, si f est une fonction mesurable :
$$(\mu*\nu)(f) = (\mu\otimes\nu)(f\circ S), \tag{8.13}$$
encore noté :
$$\int_{\mathbb{R}} f(x)\,d(\mu*\nu)(x) = \iint_{\mathbb{R}^2} f(x{+}y)\,d\mu(x)d\nu(y). \tag{8.14}$$

Remarque 8.7 Pour passer de la définition 8.5 à la définition 8.6 :

On pose $(\mu*\nu)(1_B) =^{\text{déf}} (\mu*\nu)(B)$ où B est un intervalle quelconque de $\mathbb{R}$ (ou de manière plus générale un borélien). Et ainsi on obtient avec (8.11) :
$$(\mu*\nu)(1_B) = (\mu\otimes\nu)(S^{-1}(B)) = (\mu\otimes\nu)(1_B\circ S),$$
où 1_B la fonction indicatrice d'un borélien B de $\mathbb{R}$, car $1_B\circ S = 1_{S^{-1}(B)}$, cf. (1.10). Puis à partir des fonctions indicatrices, on passe aux fonctions en escalier puis aux fonctions mesurables. ▪

Exemple 8.8 On retrouve le calcul des distributions : δ_0 (masse de Dirac en 0 donnée par $\delta_0(g) = g(0)$ pour g continue en 0) est élément neutre pour le produit de convolution : $\delta_0*\mu = \mu*\delta_0 = \mu$. En effet, avec l'abus de notation usuel cf. cours de distribution,
$$(\delta_0*\mu)(f) = (\delta_0\otimes\mu)(f\circ S) = \langle\delta_0(x)\mu(y), f(x{+}y)\rangle_{dxdy} = \langle\mu(y), \langle\delta_0(x), f(x{+}y)\rangle_{dx}\rangle_{dy},$$
par Fubini qui est toujours vrai au sens des distributions (donc ici pour toute fonction f C^∞ à support compact). Donc $(\delta_0*\mu)(f) = \langle\mu(y), f(y)\rangle_{dy} = \mu(f)$, pour toute fonction f C^∞ à support compact. ▪

Corollaire 8.9 *Si X et Y sont deux v.a.r. de lois P_X et P_Y, alors, toujours avec S la fonction somme donnée en (8.10) :*
$$P_{X+Y} = P_{(X,Y)}\circ S^{-1}, \tag{8.15}$$
i.e. P_{X+Y} est la mesure image de $P_{(X,Y)}$ par S.

Si X et Y sont deux v.a.r. indépendantes de lois P_X et P_Y, alors :
$$P_{X+Y} = P_X * P_Y, \tag{8.16}$$
i.e., dans ce cas, la loi de $X{+}Y$ est la convolée des deux lois.

Preuve. On a $S\circ(X,Y) = X+Y$ définie de Ω dans $\mathbb{R}$. Et donc $P_{X+Y} = P_{S\circ(X,Y)} = P_{(X,Y)}\circ S^{-1}$.

Et quand X et Y sont indépendantes, on applique (5.15) avec $\varphi(x,y) = S(x,y) = x+y$. ▪

Exercice 8.10 Retrouver le résultat de la sommation des mesures discrètes.

Réponse. On a $P_{X+Y}(\{z\}) = P_{(X,Y)}(S^{-1}(\{z\}))$ avec $S^{-1}(\{z\}) = \{(x,y) : x{+}y = z\} = E_z$ (droite du plan $\mathbb{R}^2$). D'où $P_{X+Y}(\{z\}) = (P_X\otimes P_Y)(E_z) = \sum_{\substack{x,y\in\mathbb{R}^2\\ x+y=z}} P_X(x)P_Y(y)$ quand X et Y sont indépendantes. ▪

Exercice 8.11 Retrouver le résultat de la sommation des mesures continues.

Réponse. On a $P_{X+Y}(\{[a,b]\}) = P_{(X,Y)}(S^{-1}(\{[a,b]\}))$ avec $S^{-1}(\{[a,b]\}) = \{(x,y) : x{+}y\in[a,b]\} = E_{ab}$ (bande oblique du plan $\mathbb{R}^2$). Donc, quand X et Y sont indépendantes : $P_{X+Y}(\{[a,b]\}) = (P_X\otimes P_Y)(E_{ab}) = \iint_{\substack{(x,y)\in\mathbb{R}^2\\ x+y\in[a,b]}} \rho_X(x)\rho_Y(y)\,dxdy = \int_{x=a}^{b}\int_{z\in\mathbb{R}} \rho_X(x)\rho_Y(z{-}x)\,dzdx$, et $P_{X+Y}(\{[a,b]\})$ est la mesure de densité $\rho_{X+Y}(z) = \int_{z\in\mathbb{R}} \rho_X(x)\rho_Y(z{-}x)\,dz = (\rho_X*\rho_Y)(z)$. ▪

8.2 Loi d'un produit

Soient $X, Y : \Omega \to \mathbb{R}$ deux v.a.r. sur un même espace probabilisé $(\Omega, \mathcal{T}, P)$ de lois respectives P_X et P_Y. La fonction produit $XY : \omega \in \Omega \to X(\omega)Y(\omega) \in \mathbb{R}$ est une v.a.r. (car le produit de deux fonctions mesurables est une fonction mesurable) de loi P_{XY} définie par $P_{XY}(I) = P((XY)^{-1}(I))$ pour tout $I \in \mathcal{T}_\mathbb{R}$.

8.2.1 Loi d'un produit : cas discret

Ici $\Omega = \{\omega_1, ..., \omega_n, ...\}$ ensemble fini ou dénombrable.

On note $X(\Omega) = \{x_i : i \in I_x\}$ et $Y(\Omega) = \{y_j \in I_y\}$, où I_x, I_y sont des ensembles finis ou dénombrables.

Quitte à renuméroter et à étendre, on prend $I_x = \mathbb{Z} = I_y$ en imposant éventuellement des probabilités nulles aux points non définis précédemment.

On note, pour $k \in \mathbb{Z}$:

$$p_k = P_X(x_k), \qquad q_k = P_Y(y_k), \tag{8.17}$$

et donc $P_X = \sum_{k\in\mathbb{Z}} p_k \delta_{x_k}$ et $P_Y = \sum_{\ell\in\mathbb{Z}} q_\ell \delta_{y_\ell}$.

Exemple 8.12 $\Omega = [1,6]_\mathbb{N}$, $X(i) = i =^{\text{noté}} x_i$ et $Y(i) = 7 - i =^{\text{noté}} y_i$, pour tout $i \in \Omega$.

On a $X(\Omega) = \{x_1, ..., x_6\}$ et $Y(\Omega) = \{y_1, ..., y_6\}$.

Et on se donne des points x_k et y_k pour $k \in \mathbb{Z} - [1,6]_\mathbb{N}$ auxquels on attribue une probabilité nulle.

Et $XY : \Omega \to \mathbb{R}$ est donnée, pour $i \in \Omega$, par $(XY)(i) = i(7-i) \in (XY)(\Omega) = \{6, 10, 12\}$. ▪

Proposition 8.13 *La loi P_{XY} est la loi discrète donnée par, pour $z \in \mathbb{R}$:*

$$P_{XY}(z) = P(XY = z) = \sum_{\substack{k,\ell\in\mathbb{Z}^2\\ x_k y_\ell = z}} P((X{=}x_k) \cap (Y{=}y_\ell)). \tag{8.18}$$

Si de plus X et Y sont indépendantes alors :

$$P_{XY}(z) = \sum_{\substack{k,\ell\in\mathbb{Z}\\ x_k y_\ell = z}} p_k q_\ell = \sum_{k,\ell\in\mathbb{Z}} p_k q_\ell \delta_{z, x_k y_\ell}, \tag{8.19}$$

où $\delta_{x_k y_\ell, z}$ est le symbole de Kronecker (=1 si $x_k y_\ell{=}z$, et =0 sinon). Autrement dit :

$$P_{XY}(z) = (P_X \otimes P_Y)(H_z),$$

où $H_z = \{(x,y) \in \mathbb{R}^2 : xy = z\}$ est l'hyperbole "$xy = z$" dans $\mathbb{R}^2$ (pour $z \neq 0$).

Avec $F = (XY)(\Omega) = \{z \in \mathbb{R} : \exists (k,\ell) \in \mathbb{Z}^2,\ z = x_k y_\ell\}$, on a :

$$P_{XY} = \sum_{z\in F} r_z \delta_z, \qquad \text{où} \qquad r_z = \sum_{\substack{k,\ell\in\mathbb{Z}^2\\ x_k y_\ell = z}} p_k q_\ell.$$

(On vérifie que :

$$\sum_{z\in F} P_{XY}(z) = \sum_{k\in\mathbb{Z}} P_X(x_k) \sum_{\ell\in\mathbb{Z}} P_Y(y_\ell) = 1. \tag{8.20}$$

Autrement dit, si on fait la somme sur toutes les hyperboles, dont les hyperboles dégénérées (les axes), on fait la somme sur $\mathbb{R}^2$, et la probabilité $(P_X \otimes P_Y)(\mathbb{R}^2) = 1$.)

Preuve. Soit $z \in \mathbb{Z}$. On s'intéresse à l'ensemble $A_z = \{\omega \in \Omega : (XY)(\omega) = z\}$ et à sa valeur $P(A_z) = P_{XY}(z)$. On a :

$$A_z = \{\omega : \exists k, \ell \in \mathbb{Z},\ X(\omega) = x_k,\ Y(\omega) = y_\ell \text{ et } x_k y_\ell = z\} = \bigcup_{\substack{k,\ell\in\mathbb{Z}\\ x_k y_\ell = z}} ((X{=}x_k) \cap (Y{=}y_\ell))$$

union d'ensembles disjoints. D'où :

$$P_{XY}(z) = P(A_z) = \sum_{\substack{k,\ell\in\mathbb{Z}\\ x_k y_\ell = z}} P((X{=}x_k) \cap (Y{=}y_\ell)).$$

En particulier, pour X et Y indépendantes on obtient (8.19).

Et $\sum_{z\in F} P_{XY}(z) = \sum_{k,\ell\in\mathbb{Z}} \sum_{z\in F} \delta_{x_k y_\ell, z} p_k q_\ell = \sum_{k,\ell\in\mathbb{Z}} p_k q_\ell = \sum_{k\in\mathbb{Z}} p_k \sum_{\ell\in\mathbb{Z}} q_\ell = 1.$ ▪

Exemple 8.14 On lance une dé équilibré deux fois. X donne la valeur $X(i) = x_i$ au premier lancer si le dé donne i, et de même et $Y(i) = y_i$ pour le second lancer. Ici $P_X(x_i) = p_i = P_Y(y_i) = q_i$ (dé équilibré et lancers indépendants). Comme X et Y sont indépendantes, on a $P_{XY}(z) = \sum_{\substack{k,\ell\in[1,6]^2\\ x_k y_\ell = z}} p_k q_\ell$ non nul pour $z \in F$, où $F = \bigcup_{i,j=1,\dots,6}\{x_i y_j\}$.

Par exemple avec $X(i) = i = Y(i)$ et $p_i = q_i = \frac{1}{6}$, on a $P_{XY}(6) = P_X(1)P_Y(6) + P_X(2)P_Y(3) + P_X(3)P_Y(2) + P_X(6)P_Y(1) = \frac{4}{36}$. ▪

8.2.2 Loi d'un produit : cas continu

Soit $\psi : \mathbb{R}^2 \to \mathbb{R}$:

$$\psi(x,y) = xy,$$

la fonction produit. Si $c \in \mathbb{R}^*$, l'ensemble $\psi^{-1}(\{c\}) \subset \mathbb{R}^2$ est constitué des deux branches de l'hyperbole "$y = \frac{c}{x}$", branches dans le 1er et le 3ème quadrant si $c > 0$, et branches dans le 2nd et 4ème quadrant si $c < 0$. Et pour $c = 0$, l'ensemble $\psi^{-1}(\{c\}) \subset \mathbb{R}^2$ est constitué des axes x=0 et y=0.

Corollaire 8.15 *Toujours avec X et Y indépendantes, on déduit : pour tout fonction f mesurable telle que $f \circ \psi$ est $P_X \otimes P_Y$ intégrable, on a :*

$$P_{XY}(f) = (P_X \otimes P_Y)(f \circ \psi), \tag{8.21}$$

encore noté :

$$\int_{z\in\mathbb{R}} f(z)\, dP_{XY}(z) = \iint_{(x,y)\in\mathbb{R}^2} f(xy)\, dP_X(x) dP_Y(y). \tag{8.22}$$

En particulier en considérant la fonction tronquée $f1_{[a,b]}$:

$$\int_{z=a}^{b} f(z)\, dP_{XY}(z) = \iint_{\{x,y\,:\,xy\in[a,b]\}} f(xy)\, dP_X(x) dP_Y(y). \tag{8.23}$$

(Dessin de $\{(x,y) : xy \in [a,b]\}$; union d'hyperboles.)

En particulier pour P_X et P_Y probabilités de densités ρ_X et ρ_Y :

$$\int_{z=a}^{b} f(z)\, dP_{XY}(z) = \iint_{x,y\,:\,xy\in[a,b]} f(xy)\rho_X(x)\rho_Y(y)\, dxdy. \tag{8.24}$$

Le domaine d'intégration du membre de droite est $\Delta = \{(x,y) \in \mathbb{R}^2 :\ xy \in [a,b]\}$: c'est l'union des hyperboles du plan comprises entre les hyperboles $y = \frac{a}{x}$ et $y = \frac{b}{x}$, faire un dessin.

Preuve. On applique (5.15) avec $\varphi(x,y) = \psi(x,y) = xy$. ▪

Corollaire 8.16 *Quand X et Y sont indépendantes et de loi de densité ρ_X et ρ_Y, la loi de P_{XY} est la loi de densité ρ_{XY} donnée par, pour $z \neq 0$:*

$$\rho_{XY}(z) = \int_{x\in\mathbb{R}} \rho_X(x)\rho_Y(\frac{z}{x})\, |x|\, dx. \tag{8.25}$$

Preuve. Soit $z > 0$. Soit $\varepsilon < z$ et soient $a = z - \varepsilon > 0$ et $b = z + \varepsilon > 0$ (donc a et b sont positifs).

Dans (8.24) on prend $f = 1_{[a,b]}$ on fait le changement de variables $\chi : \begin{pmatrix} x \\ y \end{pmatrix} \to \chi(x,y) = \begin{pmatrix} x \\ z = xy \end{pmatrix}$. La matrice jacobienne de χ est $\begin{pmatrix} 1 & 0 \\ y & x \end{pmatrix}$ de déterminant $= x$. Et le domaine Δ (union d'hyperboles) est transformé par χ en le domaine (bande) $D = \mathbb{R} \times [a,b] - \{0\} \times [a,b]$, où $\{0\} \times [a,b]$ est négligeable pour la mesure de Lebesgue. D'où avec (8.24) :

$$\begin{aligned}
\int_{z=a}^{b} dP_{XY}(z) &= \iint_{(x,z)\in\mathbb{R}\times[a,b]} \rho_X(x)\rho_Y(\frac{z}{x})\, |x|\, dxdz \\
&= \int_{z\in[a,b]} (\int_{x\in\mathbb{R}} \rho_X(x)\rho_Y(\frac{z}{x})\, |x|\, dx)\, dz,
\end{aligned}$$

l'application du théorème de Fubini étant légitime pour les fonction positives, voir cours d'intégration. Même démarche quand $z < 0$. ▪

9 Proposition de la mesure image

Ce paragraphe reprend le paragraphe 4.2.3. On insiste car la proposition de la mesure image est souvent mal comprise, alors qu'elle est essentielle à la compréhension de l'espérance entre autres.

9.1 Changement de variables dans les intégrales : proposition de la mesure image

9.1.1 Rappel : mesure et intégration d'une fonction (au sens de Legesgue)

Soit $(E, \mathcal{T}_E, \mu)$ un espace mesuré. Si $A \subset E$ vérifie $A \in \mathcal{T}_E$ (i.e. A mesurable), la réel positif $\mu(A)$ a un sens dans $\overline{\mathbb{R}_+} = [0, \infty]$.

Notation. Soit $\tilde{\mu}$ définie sur les fonctions indicatrices 1_A pour $A \in \mathcal{T}_E$ par :

$$\tilde{\mu}(1_A) \stackrel{\text{déf}}{=} \mu(A), \qquad \tilde{\mu}(1_A) \stackrel{\text{noté}}{=} \int 1_A \, d\mu \stackrel{\text{noté}}{=} \int_A d\mu. \tag{9.1}$$

Interprétation pour la mesure de Lebesgue : "l'aire sous la courbe 1_A est l'aire du rectangle de base A et de hauteur 1".

Puis on impose à $\tilde{\mu}$ d'être "linéaire" sur l'ensemble des fonctions indicatrices : pour tout $n \in \mathbb{N}^*$ et tout $c_1, ..., c_n \in \mathbb{R}$:

$$\tilde{\mu}(\sum_{i=1}^n c_i 1_{A_i}) \stackrel{\text{déf}}{=} \sum_{i=1}^n c_i \tilde{\mu}(1_{A_i}) \qquad (= \sum_{i=1}^n c_i \mu(A_i)). \tag{9.2}$$

Puis on impose à $\tilde{\mu}$ de passer à la limite sous le signe $\tilde{\mu}$ pour les fonctions en escalier positives : pour tout $(c_i)_{\mathbb{N}^*}$ suite dans $\mathbb{R}_+$:

$$\tilde{\mu}(\lim_{n\to\infty} \sum_{i=1}^n c_{in} 1_{A_{in}}) = \lim_{n\to\infty} \sum_{i=1}^n c_{in} \tilde{\mu}(1_{A_{in}}) \quad \in [0, \infty]. \tag{9.3}$$

Une fonction mesurable positive $f : E \to \mathbb{R}$ étant limite simple de fonctions en escalier, on a ainsi défini $\tilde{\mu}(f)$ pour toute fonction f mesurable positive.

Puis on dit qu'une fonction positive est μ-intégrable ssi (elle est mesurable et) $\tilde{\mu}(f) < \infty$ (i.e. si $\tilde{\mu}(f) \in [0, \infty[$).

Et on dit qu'une fonction f est μ-intégrable ssi f_+ et f_- sont μ-intégrables, où f_+ et f_- sont les fonctions positives définies par $f_+ = \sup(f, 0)$ et $f_- = \sup(-f, 0)$ (avec donc $f = f_+ - f_-$). Et on note :

$$\tilde{\mu}(f) = \int f \, d\mu. \tag{9.4}$$

Puis, on note (abusivement), le contexte levant les ambiguïtés, $\tilde{\mu} = \mu$:

$$\tilde{\mu}(f) \stackrel{\text{noté}}{=} \mu(f) \qquad (= \int f \, d\mu). \tag{9.5}$$

9.1.2 Changement de variables dans les intégrales

Soit :

$$\varphi : \left\{ \begin{array}{ll} \Omega{=}[a, b] & \to V{=}[c, d] \\ \omega & \mapsto x = \varphi(\omega) \stackrel{\text{noté}}{=} x(\omega) \end{array} \right\}$$

un difféomorphisme (fonction bijective C^1 d'inverse C^1 = "un changement de variables").

La formule générique de changement de variables dans les intégrales est, lorsque lorsque $g : V \to \mathbb{R}$ est intégrable sur $\varphi(\Omega)$:

$$\int_{x\in\varphi(\Omega)} g(x) \, dx = \int_{\omega\in\Omega} g(x(\omega)) \, |x'(\omega)| \, d\omega. \tag{9.6}$$

(la valeur absolue pour tenir compte du cas φ décroissante, les notations $\Omega = [a, b]$ et $\varphi(\Omega) =$

$[c,d]=V$ imposant implicitement $a<b$ et $c<d$). Soit donc ici :

$$\int_{x=c}^{d} g(x)\,dx = \int_{\omega=a}^{b} g(x(\omega))\,|x'(\omega)|\,d\omega. \tag{9.7}$$

En vue de la proposition de la mesure image, on pose :

$$d\mu(\omega) = |\varphi'(\omega)|\,d\omega \qquad (= |x'(\omega)|\,d\omega), \tag{9.8}$$

et (9.6) s'écrit alors aussi :

$$\int_V g(x)\,dx = \int_{\varphi^{-1}(V)} (g\circ\varphi)(\omega)\ \ d\mu(\omega). \tag{9.9}$$

De même (ici φ est inversible) :

$$\int_{\omega\in\Omega} f(\omega)\,d\omega = \int_{x\in\varphi(\Omega)} (f\circ\varphi^{-1})(x)\ \ d\nu(x) \qquad \text{où} \qquad d\nu(x) = |(\varphi^{-1})'(x)|\,dx. \tag{9.10}$$

Remarque 9.1 Rappel. Ici bien sûr dx et $d\omega$ sont les mesures de Lebesgue sur $\mathbb{R}$: si μ_ℓ est la mesure de Lebesgue, on commence par écrire (9.6) sous la forme $\mu_\ell(1_{\varphi(\Omega)}g) = \mu_\ell(1_\Omega(g\circ\varphi)|\varphi'|)$, soit :

$$\int_{x\in\varphi(\Omega)} g(x)\,d\mu_\ell(x) = \int_{\omega\in\Omega} g(\varphi(\omega))\,|\varphi'(\omega)|\,d\mu_\ell(\omega).$$

Et l'usage fait que, pour la mesure de Lebesgue, on oublie d'écrire μ_ℓ, et ainsi on écrit $d\mu_\ell(x)=dx$ et $d\mu_\ell(\omega)=d\omega$, d'où l'écriture (9.6).

9.1.3 Proposition de la mesure image sur les fonctions

Soit $(\Omega,\mathcal{T}_\Omega)$ et $(V,\mathcal{T}_V)$ deux ensembles mesurables.
Soit $\varphi:\Omega\to V$ une fonction mesurable (non nécessairement inversible).
Soit $\mu:\mathcal{T}_\Omega\to\mathbb{R}_+$ une mesure sur Ω.
Soit $\nu:\mathcal{T}_V\to\mathbb{R}_+$ la mesure image de $\mu:\mathcal{T}_\Omega\to\mathbb{R}_+$ par $\varphi:\Omega\to V$:

$$\nu = \mu\circ\varphi^{-1}, \tag{9.11}$$

cf. (4.9), où donc $\nu(B)=\mu(\varphi^{-1}(B)$ pour B ensemble mesurable.

En particulier $\nu(V)=\mu(\varphi^{-1}(\mathbb{R}))=\mu(\Omega)=1$, ainsi que $\nu(V)=\mu(\varphi^{-1}(\varphi(\Omega)))=1$.

Le passage aux mesures de fonctions est basé sur $\mu(1_A) \stackrel{\text{déf}}{=} \mu(A)$, cf. (9.1), et sur (1.10) qu'on rappelle :

$$1_{f^{-1}(B)} = 1_B\circ f. \tag{9.12}$$

Proposition 9.2 *Si $g:V\to\mathbb{R}_+$ est mesurable positive on a,*

$$\nu(g) = \mu(g\circ\varphi), \tag{9.13}$$

noté encore :

$$\int_V g(x)\,d\nu(x) = \int_\Omega g(\varphi(\omega))\ d\mu(\omega). \tag{9.14}$$

(À comparer avec (9.9) relative à la mesure de Lebesgue et aux fonctions inversibles.)

(9.13) reste vraie si $g\in L^1(V,\nu)$ (i.e. $\int_V |g(x)|\,d\nu<\infty$).

(9.13) reste vraie si g est bornée dans le cas ν mesure bornée.

Et pour tout $B\in\mathcal{T}_V$ (la fonction 1_Bg étant alors mesurable) :

$$\int_B g(x)\,d\nu(x) = \int_{\varphi^{-1}(B)} (g\circ\varphi)(\omega)\,d\mu(\omega). \tag{9.15}$$

Et on retrouve (4.11) (page 51) si $g=1_B$.

Preuve. La mesure de la fonction indicatrice de $B \in \mathcal{T}_V$ est définie par $\nu(1_B) =^{\text{déf}} \nu(B)$, cf. (9.1). En particulier (4.11) peut être réécrite avec (9.12) :

$$\int_{x\in V} 1_B(x)\,d\nu(x) = \int_{\omega\in\Omega} 1_{\varphi^{-1}(B)}(\omega)\,d\mu(\omega) = \int_{\omega\in\Omega} (1_B\circ\varphi)(\omega)\,d\mu(\omega), \tag{9.16}$$

ce qui donne (9.15) dans le cas $g = 1_B$. D'où, par linéarité de l'intégrale :

$$\int_V \sum_{i=1}^N c_i 1_{B_i}(x)\ d\nu(x) = \int_\Omega ((\sum_{i=1}^N c_i 1_{B_i})\circ\varphi)(\omega)\ d\mu(\omega), \tag{9.17}$$

pour tout $N \in \mathbb{N}$, $c_i \in \mathbb{R}_+$ et $B_i \in \mathcal{T}_V$. D'où (9.15) dans le cas g fonction en escalier.

Et les fonctions $g : V \to \mathbb{R}$ mesurables positives étant limites de combinaisons linéaires de fonctions en escalier positives, on obtient (9.15) dans $\overline{\mathbb{R}_+}$ pour les fonctions mesurables positives.

Puis si $|g|$ est ν-intégrable, alors $g = g_+ - g_-$ avec g_+ et g_- positives, et avec g_+ et g_- satisfaisant (9.15), et $\nu(g) = \nu(g_+) - \nu(g_-)$.

Puis si g est bornée et ν est bornée, alors $|g|$ est ν-intégrable (de valeur $\leq ||g||_\infty \nu(V)$), donc g est ν-intégrable. ■

Exemple 9.3 Si $\varphi : \Omega \to V$ est un difféomorphisme (est C^1 inversible d'inverse C^1), et si μ est une mesure de densité ρ_μ, i.e. $d\mu(\omega) = \rho_\mu(\omega)d\omega$, on obtient :

$$d\nu(y) = (\rho_\mu\circ\varphi^{-1})(y)\ |(\varphi^{-1})'(y)|\ dy, \tag{9.18}$$

i.e. ν est une mesure de densité $(\rho_\mu\circ\varphi^{-1})(y)\ |(\varphi^{-1})'(y)|$. En particulier, si μ est la mesure de Lebesgue, avec donc $\rho_\mu = 1_\Omega$, on retrouve $d\nu(y) = |(\varphi^{-1})'(y)|\ dy$, cf. (9.10). ■

Exemple 9.4 Soit μ une mesure finie. Soit $a < b$, soit $y_0 \in \mathbb{R}$ fixé, et soit $\varphi : \Omega \to V = \mathbb{R}$ la fonction constante :

$$\varphi = y_0 1_\Omega, \tag{9.19}$$

Ici $\varphi(\Omega) = \{y_0\}$ (singleton). Donc, pour $B \in \mathcal{T}_\mathbb{R}$, on a $\varphi^{-1}(B) = \left\{\begin{array}{l}\Omega \ \text{ si } y_0 \in B\\ \emptyset \ \text{ si } y_0 \notin B\end{array}\right\}$. Donc :

$$\nu(B) = \mu(\varphi^{-1}(B)) = \left\{\begin{array}{ll}\mu(\Omega) & \text{si } y_0 \in B\\ 0 & \text{si } y_0 \notin B\end{array}\right\}, \quad \text{donc} \quad \nu = \mu(\Omega)\delta_{y_0}, \tag{9.20}$$

et ν est une mesure ponctuelle. Et $(g\circ\varphi)(\omega) = g(y_0)$, soit $g\circ\varphi = g(y_0)1_\Omega$, donc :

$$\nu(g) = \mu(\Omega)\delta_{y_0}(g) = \mu(\Omega)g(y_0), \tag{9.21}$$

et $\mu(g\circ\varphi) = \mu(g(y_0)1_\Omega) = g(y_0)\mu(1_\Omega, = \nu(g)$ comme il se doit. ■

Remarque 9.5 Si $f = g\circ\varphi$ est une fonction, $f\circ\varphi^{-1}$ n'est pas une fonction quand φ n'est pas bijective. Dans ce cas (9.14) est bien défini, mais par contre $\int_\Omega f\ d\mu = \int_V (f\circ\varphi^{-1})\,d\nu$ ne l'est pas : la proposition de la mesure image ne marche "que dans un sens" quand φ n'est pas bijective. ■

Exemple 9.6 Voir exercice 6.54. ■

9.2 Application aux probabilités

9.2.1 Probabilité image

Soit $\mathbb{R}$ muni de sa tribu borélienne $\mathcal{T}_\mathbb{R}$. Soit P une probabilité sur un espace mesurable $(\Omega, \mathcal{T}_\Omega)$.

Soit $X : \Omega \to V \subset \mathbb{R}$ une v.a.r. (une fonction mesurable), où $V = X(\Omega)$. On dispose de la mesure image :

$$P_X = P\circ X^{-1} \tag{9.22}$$

Proposition 9.7 *P_X est une mesure de probabilité.*

Preuve. On a $P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$. ▪

Et (9.13) donne :

$$P_X(g) = P(g \circ X), \tag{9.23}$$

soit :

$$\int_{x\in V} g(x)\,dP_X(x) = \int_{\omega\in\Omega} g(X(\omega))\,dP(\omega). \tag{9.24}$$

Et quitte à remplacer g par $1_B g$ où $B \in \mathcal{T}_\mathbb{R}$ (la fonction g tronquée), avec (9.12) :

$$\int_{x\in B} g(x)\,dP_X(x) = \int_{\omega\in X^{-1}(B)} g(X(\omega))\,dP(\omega). \tag{9.25}$$

En particulier, avec $g = id$ l'identité :

$$\int_{x\in V} x\,dP_X(x) = \int_{\omega\in\Omega} X(\omega)\,dP(\omega) = P(X) \quad (\overset{\text{noté}}{=} E(X) = \text{espérance de } X). \tag{9.26}$$

9.2.2 Cas discret

Exemple : $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$, et P la mesure atomique :

$$P = \sum_{i=1}^n q_i \delta_{\omega_i}, \tag{9.27}$$

où les q_i sont positifs et $\sum_{i=1}^n q_i = 1$.

Soit $X : \Omega \to \mathbb{R}$ une fonction donnée. Soit $V = X(\Omega) = \{x_1, ..., x_m\}$ l'image (avec donc $m \le n$).

La mesure image P_X est donnée par $P_X(B) = P(\varphi^{-1}(B))$ pour tout $B \subset \mathbb{R}$. C'est donc la mesure :

$$P_X = \sum_{i=1}^m p_i \delta_{x_i}, \qquad \text{où} \qquad p_i = \sum_{\{j\in\mathbb{N}\,:\,\omega_j\in X^{-1}\{x_i\}\}} q_j.$$

Et pour toute fonction $g : \mathbb{R} \to \mathbb{R}$, on a :

$$P_X(g) = \sum_{i=1}^m p_i\, g(x_i) = \sum_{i=1}^n q_i\, g(X(\omega_i)). \tag{9.28}$$

En particulier $E(X) = P(X) = \sum_{i=1}^n q_i\, X(\omega_i) = P_X(id) = \sum_{i=1}^m p_i\, x_i$ est l'espérance.

Exemple 9.8 Soit X la fonction constante $X \equiv 36$ (et n'est donc pas inversible). Donc $m = 1$, $V = X(\Omega) = \{36\} = \{x_1\}$, et donc $p_1 = \sum_{j=1}^n q_j = 1$.

Et $P_X(id) = p_1\, g(36) = 1 * 36 = 36 = E(X) = P(X)$ (attendu). ▪

9.2.3 Cas continu

Exemple : $\Omega = \mathbb{R}$, et P la mesure continue de densité q :

$$dP(\omega) = q(\omega)\,d\omega, \tag{9.29}$$

où $\int_\mathbb{R} q(\omega)\,d\omega = 1$.

Soit $X : \Omega \to \mathbb{R}$ une fonction dérivable donnée. Ici $V = X(\Omega) \subset \mathbb{R}$.

On a $P_X(g) = \mu_\ell(g \circ X)$.

Exemple 9.9 Soit X la fonction constante $X \equiv 36$ (et n'est donc pas inversible). Donc $V = X(\Omega) = \{36\} = \{x_1\}$, et donc $p_1 = \sum_{j=1}^n q_j = 1$.

Et $P_X(g) = p_1\, g(36) = g(36)$, donc $P_X = \delta_{36}$ est une mesure discrète.

Et $P_X(id) = 36 = P(X) = \int_\Omega 36\, q(\omega)\,d\omega$ est l'espérance. ▪

Exemple 9.10 X est un difféomorphisme $\Omega \to V$. P_X est une mesure de densité "$dP_X = p(x)\,dx$" :

$$P_X(g) = \int_{x\in V} g(x)\,p(x)\,dx = \int_{\omega\in X^{-1}(V)} (g\circ X)(\omega)\,q(\omega)\,d\omega, \tag{9.30}$$

à comparer à (9.28), avec :

$$p(x) = q(X^{-1}(x))\,|(X^{-1})'(x)|, \tag{9.31}$$

cf. (9.10). ▪

10 Espérance, écart type, variance, moments d'une v.a.r.

Soit un espace probabilisé $(\Omega, \mathcal{T}, P)$. Soit une v.a.r. (une fonction mesurable) $X : \omega \in \Omega \rightarrow X(\omega) =^{\text{noté}} x \in \mathbb{R}$, et soit $P_X = P \circ X^{-1}$ sa loi de probabilité.

10.1 Moment d'ordre 0 : la masse unité

$P_X(\mathbb{R}) = P(\Omega) = 1$ est appelée la masse de X, et on note :

$$\int_\Omega dP = \int_\mathbb{R} dP_X = 1 = 100\% = \frac{100}{100}. \tag{10.1}$$

(Toute la masse est dans Ω.) Donc, respectivement dans les cas discret et continu :

$$\sum_{i \in I} p_i = 1 \qquad \text{et} \qquad \int_\mathbb{R} p(x)\, dx = 1. \tag{10.2}$$

10.2 Moment d'ordre 1 : l'espérance

(Espérance = [Expectation].)

10.2.1 Définition

Dictionnaire Larousse : espérance = sentiment qui porte à considérer ce que l'on désire comme réalisable ; confiance, certitude.

Ici l'espérance aura le sens suivant : si on recommence "le même jeu" (la même expérience) de nombreuses fois, quel est la moyenne des gains ?

En mécanique, l'espérance est le centre de gravité : si on doit fixer un objet à l'aide d'un seul point pour qu'il reste en équilibre, ce point est le centre de gravité.

Définition 10.1 L'espérance $\overline{X}$ d'une v.a.r. X est le réel $P(X)$ = mesure de la fonction X par la mesure P :

$$\overline{X} \overset{\text{déf}}{=} P(X) \overset{\text{noté}}{=} E(X) \overset{\text{noté}}{=} \int_{\omega \in \Omega} X(\omega)\, dP(\omega) \overset{\text{noté}}{=} \int_\Omega X\, dP, \tag{10.3}$$

lorsque $\int X\, dP$ existe dans $\mathbb{R}$. Egalement noté $E_P(X)$ s'il y a un doute sur la probabilité.

Faire un dessin : l'aire sous la fonction constante $\overline{X} 1_\Omega$ est égale à l'aire sous la fonction X.

Donc :

$$\text{cas } X \text{ v.a.r. discrète,} \qquad \overline{X} = E(X) = \sum_\Omega X(\omega)\, q_\omega, \tag{10.4}$$

où $q_\omega = P(\{\omega\})$ est la "fréquence" d'obtention de l'issue ω. Et :

$$\text{cas } X \text{ v.a.r. continue,} \qquad \overline{X} = E(X) = \int_\Omega X(\omega)\, q(\omega) d\omega, \tag{10.5}$$

P étant la mesure de densité $dP(\omega) = q(\omega)d\omega$.

Remarque 10.2 Voir annexe 15. En particulier, comme $P(\Omega) = 1$, cette valeur $E(X)$ est le centre de gravité de X. Rappel : le centre de gravité x_g relativement à une mesure μ est le réel vérifiant $x_g \Big(\int_\Omega d\mu(\omega) \Big) = \int_\Omega X(\omega)\, d\mu(\omega)$, soit $x_g = \int_\Omega X(\omega)\, d\tilde{\mu}(\omega)$ où $d\tilde{\mu}_\omega = \frac{d\mu(\omega)}{\int_\Omega d\mu(\omega)}$ est une mesure de probabilité : $\int_\Omega d\tilde{\mu}(\omega) = 1$. ▟

Exemple 10.3 On lance 2 fois une pièce équilibrée. Ici $\frac{1}{2} = P(\text{pile}) = P(\text{face})$ est la "fréquence moyenne" d'obtention de l'issue pile ou de face. Supposons qu'on gagne 1 si pile et 10 si face, soit $X(\text{pile}) = 1$ et $X(\text{face}) = 10$. En moyenne on gagne $\overline{X} = X(\text{pile})P(\text{pile}) + X(\text{face})P(\text{face}) = 1 * \frac{1}{2} + 10 * \frac{1}{2} = 5,5$ (si on fait une infinité de lancers, en moyenne on gagne $5,5$).

On lance 2 fois une pièce déséquilibrée avec $P(\text{pile}) = p$ et $P(\text{face}) = q$. On gagne $X(\text{pile}) = x_p$ si pile et $X(\text{face}) = x_q$ si face. En moyenne on gagne $\overline{X} = X(\text{pile})P(\text{pile}) + X(\text{face})P(\text{face}) = x_p * p + x_q * q$. ▟

Exercice 10.4 Un élève obtient 12 sur 20 en matière m_1, 15 sur 20 en en matière m_2, et 4 sur 10 en matière m_3. La matière m_1 a pour coefficient 1, m_2 a pour coefficient 2, m_3 a pour coefficient 3. Donner la moyenne de l'élève.

Réponse. Valeur pour chaque matière : $X(m_1)=\frac{12}{20}$, $X(m_2)=\frac{15}{20}$, $X(m_1)=\frac{4}{10}=\frac{8}{20}$.
Coefficients des matières : $c(m_1)=1$, $c(m_2)=2$, $c(m_3)=3$.
Somme des coefficients : $C=c(m_2)+c(m_3)+c(m_3)=6$.
Poids relatifs des matières : $P(m_k)=\frac{c(m_k)}{C}=\frac{k}{6}$.
$\overline{X}=\sum_{k=1}^3 X(m_k)P(m_k)=\frac{12}{20}*\frac{1}{6}+\frac{15}{20}*\frac{2}{6}+\frac{4}{10}\frac{3}{6}=\frac{1}{20}\frac{12+15*2+8*3}{6}=\frac{11}{20}$, donc 11 sur 20. ▪

Exercice 10.5 A la fin d'une année scolaire, on fait la moyenne des n notes affectées des coefficients c_i. Un élève à la note $\frac{N_i}{20}$ à la i-ème matière. Calculer sa moyenne sur 20. Une erreur s'est produite sur la j-ème note : au lieu d'avoir eu N_j, il a eu N_j+e_j. De combien de points cela modifie-t-il sa moyenne ? Application numérique : $\sum c_i=50$ avec $e_j=2$.

Réponse. Soit $C=\sum_{i=1}^n c_i$.
Moyenne initiale $M_1=\sum_{i=1}^n \frac{N_i}{20}\frac{c_i}{C}=\frac{1}{20}\frac{1}{C}\sum_{i=1}^n N_ic_i$.
Moyenne corrigée $M_2=\frac{1}{20}\frac{1}{C}\sum_{i=1}^n (N_i+e_i)c_i$, où e_i est l'erreur pour la i-ème matière.
$M_2-M_1=\frac{1}{20}\frac{1}{C}\sum_{i=1}^n c_ie_i$. Si $e_i=0$ pour tout $i\neq j$, alors $M_2-M_1=\frac{1}{20}\frac{c_j}{C}e_j$.
A.N. : $M_2-M_1=\frac{1}{20}\frac{1}{50}2\simeq\frac{1}{20}\frac{1}{25}$, donc 0,04 point sur 20 de plus sur la moyenne générale. ▪

Exemple 10.6 Si $X=1_A$ avec A ensemble mesurable (donc X est la v.a.r. "on-off' $X(\omega)=1$ si $\omega\in A$ et $X(\omega)=0$ si $\omega\notin A$), alors $\overline{X}=\int 1_A\,dP=P(A)=^{\text{noté}}E(A)$. ▪

10.2.2 Calcul à l'aide de la mesure image : le moment d'ordre 1

On a $P(g\circ X)=P_X(g)$, cf. (9.23), donc avec $g=id=$ l'identité ($id(x)=x$).

Corollaire 10.7 *Soit $X:\Omega\to\mathbb{R}$. À l'aide de la mesure image $P_X=P\circ X^{-1}$, on a également :*

$$\overline{X}=E(X)=\int_{x\in\mathbb{R}} x\,dP_X(x)=P_X(id),\tag{10.6}$$

Ainsi $P(X)=E(X)=\int x\ dP_X(x)$ est le moment d'ordre 1 par rapport 0 (l'origine des x) relativement à la mesure P_X (voir annexe § 15.3).

Donc :

$$P_X=\sum_{i\in I} p_{x_i}\delta_{x_i}\text{ proba discrète}\qquad\Longrightarrow\qquad\overline{X}=\sum_{i\in I}x_i\,p_i,\tag{10.7}$$

et :

$$dP_X=p(x)\,dx\text{ proba continue}\qquad\Longrightarrow\qquad\overline{X}=\int_{x\in\mathbb{R}}x\,p(x)\,dx.\tag{10.8}$$

Exemple 10.8 On lance 2 fois une pièce. Soit $p=P(\text{pile})$ et $q=P(\text{face})$. Supposons qu'on gagne $X(\text{pile})=x_p$ si pile et $X(\text{face})=x_q$ si face. En moyenne on gagne :

$$\overline{X}=x_pP_X(x_p)+x_qP_X(x_q)=x_p*p+x_q*q.\tag{10.9}$$

C'est le barycentre des points x_1 et x_2 affectés des coordonnées barycentriques p_1 et p_2. ▪

Exemple 10.9 Si $X=1_A$ avec A ensemble mesurable. Ici $X(\Omega)=1_A(\Omega)=\{0,1\}=\{x_0,x_1\}$ ensemble (des états) discret des valeurs de 1_A. Ici P_X est une probabilité discrète avec $P_X(1)=P(1_A^{-1}(1))=P(A)=^{\text{noté}}p_1$ et $P_X(0)=P(1_A^{-1}(0))=P(A^C)=^{\text{noté}}p_0$. Donc $\overline{X}=0*p_0+1*p_1=p_1=P(A)$. ▪

Exemple 10.10 Jeux type "roue de la fortune", où $X(\omega)=$ le gain écrit sur le "secteur angulaire ω". Et $E(X)$ donne l'espérance de "gain moyen". ▪

Exemple 10.11 Lancers d'un dé équilibré (à 6 faces) : on gagne 1 si résultat pair et -1 si résultat impair. En moyenne, on gagne $E=3\frac{1}{6}+3\frac{-1}{6}=0$. ▪

Exemple 10.12 Lancers d'un dé équilibré : on gagne k si le dé donne k. En moyenne, on gagne $E=\sum_{k=1}^6 k\frac{1}{6}=1\frac{1}{6}+2\frac{1}{6}+...+6\frac{1}{6}=\frac{21}{6}=\frac{7}{2}=3,5$. ▪

Exercice 10.13 On lance un dé, et on attribue les valeurs 1,3,5,7,9,11 aux éventualités 1,2,3,4,5,6, et on suppose le dé équilibré. Calculer l'espérance.

Réponse. Ici $X(i) = 2i - 1 = x_i$ et donc $E(X) = \sum_i x_i \frac{1}{6} = \frac{1}{6}(2\sum_{i=1}^6 i - 6) = \frac{1}{6}(42 - 6) = 6$.

On vérifie que $\sum_i (x_i - 6)\frac{1}{6} = 0$, i.e. que 6 est le centre de gravité. ▪

Exemple 10.14 On considère n familles f_i et on note n_i le nombre d'enfants de la famille i. Le nombre moyen d'enfants par famille est $\frac{\sum_i n_i}{n} = \sum_i n_i \frac{1}{n}$.

Ici implicitement X est la v.a.r. qui à une famille f_i associe le nombre d'enfants $X(f_i)=n_i$ et $p_i = \frac{1}{n}$ la probabilité de choisir la famille f_i. Et $E(X) = \sum_i X(f_i)p_i = \sum_i n_i \frac{1}{n}$. ▪

Exercice 10.15 Soit $\varphi : \mathbb{R}^2 \to \mathbb{R}$ une v.a.r., et soit $X, Y : \Omega \to \mathbb{R}$ deux v.a.r.. On note $\varphi(X,Y) = \varphi \circ (X,Y)$ la v.a.r. sur Ω donnée par $\varphi(X,Y)(\omega) = \varphi(X(\omega), Y(\omega))$. Si X et Y sont indépendantes, montrer $E(\varphi(X,Y)) = E(f(X)) = E(g(Y))$, où $f(x) = E(\varphi(x,Y))$ et $g(y) = E(\varphi(X,y))$.

Réponse. X et Y étant indépendantes, on a (5.15) :

$$\int_{z\in\mathbb{R}} z\,dP_{\varphi(X,Y)}(z) = \iint_{(x,y)\in\mathbb{R}^2} \varphi(x,y)\,dP_X(x)dP_Y(y) = \int_y (\int_x \varphi(x,y)\,dP_X(x))dP_Y(y). \tag{10.10}$$

▪

10.2.3 V.a.r. centrée

Définition 10.16 La v.a.r. X est dite centrée ssi $E(X) = 0$, i.e. ssi $\overline{X} = 0$.

Exemple 10.17 Dans l'exercice 10.13, la v.a.r. $Y = X - 6\,1_\Omega$ est centrée : $\overline{Y} = 0$ (vérification immédiate), ou on applique la proposition suivante (linéarité de E) : $\overline{Y} = \overline{X} - 6\,\overline{1_\Omega} = 6 - 6 = 0$. ▪

10.2.4 Premières propriétés

Proposition 10.18 *L'espérance définie par (10.3) est une application linéaire sur l'espace vectoriel des fonctions $X : \Omega \to \mathbb{R}$ Riemann intégrables (au sens de Riemann t.q. $\int X(\omega)\,dP(\omega) < \infty$), et sur l'ensemble des fonctions Lebesgue intégrables (mesurables au sens de Lebesgue et t.q. $\int |X(\omega)|\,dP(\omega) < \infty$).*

Preuve. La linéarité $E(X + \lambda Y) = E(X) + \lambda E(Y)$ est immédiate avec (10.3) puisque P est une mesure. ▪

Proposition 10.19 *1- Toute v.a.r. constante $X = c1_\Omega$, où $c \in \mathbb{R}$, a une espérance qui vaut $E(X) = c$, i.e. $\overline{X} = c$. Autrement dit, si la seule valeur possible est c, on est sûr d'obtenir c : $E(c1_\Omega) = cE(1_\Omega) = cP(1_\Omega) = cP(\Omega) = c$, cf. (9.1) (une fonction constante est une fonction déterministe : une seule issue).*

2- Pour toute v.a.r. X qui admet une espérance $\overline{X} = E(X)$ finie, on a :

$$\overline{X} = E(X) = E(\overline{X}1_\Omega) \overset{\text{noté}}{=} E(\overline{X}). \tag{10.11}$$

Et donc la v.a.r. $X - \overline{X}1_\Omega =^{\text{noté}} X - \overline{X}$ est centrée (centre de gravité en 0) :

$$E(X - \overline{X}1_\Omega) = 0, \qquad \text{et} \qquad E(X - \overline{X}1_\Omega) \overset{\text{noté}}{=} E(X - \overline{X}). \tag{10.12}$$

Preuve. 1- Soit $X = c1_\Omega$ (fonction constante sur Ω). On a $E(X) = \int_\Omega X(\omega)\,dP(\omega) = \int_\Omega c\,dP(\omega) = c\int_\Omega dP(\omega) = c$: donc $E(X) = c$ est "la valeur de X".

2- Par linéarité de E on a $E(X - \overline{X}1_\Omega) = E(X) - \overline{X}E(1_\Omega) = E(X) - \overline{X} = 0$. ▪

Remarque 10.20 N.B. : attention aux notations, voir § 1.3.6 page 9. Ici $\overline{X}$ est à la fois :

(i) la notation d'un réel dans l'égalité $E(X) = \overline{X}$, et

(ii) la notation de la fonction constante $\overline{X}1_\Omega =^{\text{noté}} \overline{X}$.

Ainsi la notation usuelle $E(\overline{X})$ est la notation abusive de $E(\overline{X}1_\Omega)$ (espérance de la fonction constante $\overline{X}1_\Omega$), de même que la notation usuelle $E(X - \overline{X})$ est la notation abusive de $E(X - \overline{X}1_\Omega)$ (espérance de la fonction $X - \overline{X}1_\Omega$). ▪

10.2.5 Généralisation

$P(g \circ X) = P_X(g)$ donne :

Corollaire 10.21 *Si $X \in L^1(\Omega, P)$ et si $g \in L^1(\mathbb{R}; P_X)$ (i.e. si $\int_\Omega |X(\omega)|\, dP(\omega) < \infty$ et si $\int_\mathbb{R} |g(x)|\, dP_X(x) < \infty$), alors on a :*

$$E(g \circ X) = P_X(g), \qquad \text{soit} \qquad E(g \circ X) = \int_{x \in \mathbb{R}} g(x)\, dP_X(x) \overset{\text{noté}}{=} \int g\, dP_X, \tag{10.13}$$

mesure de la fonction g par P_X. (En particulier $g = id$ (l'identité) redonne l'espérance de X.)

Donc dans le cas P_X discret :

$$E(f \circ X) = \sum_{j \in j} f(x_j) p_{x_j}, \tag{10.14}$$

et dans le cas P_X continu :

$$E(f \circ X) = \int_{x \in \mathbb{R}} f(x)\, p(x)\, dx. \tag{10.15}$$

Exemple 10.22 Dans le cas discret, on retrouve le résultat :

$$E(f \circ X) = \sum_{x_j \in X(\Omega)} f(x_j) P_X(x_j). \tag{10.16}$$

En effet, on sommant par "paquets" (les paquets $X^{-1}(x_j) = \{\omega : X(\omega) = x_j\}$ pour les $x_j \in X(\Omega)$) :

$$\begin{aligned} E(f \circ X) &= \sum_{\omega_i \in \Omega} f(X(\omega_i)) P(\omega_i) = \sum_{x_j \in X(\Omega)} \sum_{\omega_i \in X^{-1}(x_j)} f(x_j) P(\omega_i) \\ &= \sum_{x_j \in X(\Omega)} f(x_j) \Big(\sum_{\omega_i \in X^{-1}(x_j)} P(\omega_i) \Big) = \sum_{x_j \in X(\Omega)} f(x_j)\, P(\{X{=}x_j\}) = \sum_{x_j \in X(\Omega)} f(x_j) P_X(x_j) \end{aligned}$$

Ce calcul est licite dès que la sommation par paquets l'est, ce dont on est sûr si la suite $(f(x_i))_\mathbb{N}$ est P_X-sommable, i.e. si $\sum_i |f(x_i)| < \infty$. ▲

Exemple 10.23 La fonction $f \circ \overline{X} 1_\Omega$ est la fonction constante $f(\overline{X} 1_\Omega) = f(\overline{X}) 1_\Omega$ (puisque $f(\overline{X} 1_\Omega(\omega)) = f(\overline{X})$ pour tout $\omega \in \Omega$). Donc :

$$P(f \circ \overline{X} 1_\Omega) = E(f \circ \overline{X} 1_\Omega) = E(f(\overline{X}) 1_\Omega) = \int_\Omega f(\overline{X})\, dP = f(\overline{X}) \int_\Omega dP = f(\overline{X}). \tag{10.17}$$

Et, comme $P_{f \circ \overline{X} 1_\Omega} = P_{f(\overline{X}) 1_\Omega} = P \circ (f(\overline{X}) 1_\Omega)^{-1} = \delta_{f(\overline{X})}$, cf. (9.19) et (9.20), c'est bien égal à :

$$P_{f \circ \overline{X} 1_\Omega}(id) = \delta_{f(\overline{X})}(id) = id(f(\overline{X})) = f(\overline{X}), \tag{10.18}$$

cf. (9.21). ▲

10.2.6 Espérance d'un produit : cas de v.a.r. indépendantes

L'espérance (l'intégrale) étant linéaire, l'espérance d'une somme est la somme des espérances :

$$E(X + Y) = E(X) + E(Y). \tag{10.19}$$

Pour le produit : on se donne deux v.a.r. X et Y sur un même espace probabilisé $(\Omega, \mathcal{T}, P)$. La fonction produit $XY : \omega \in \Omega \to (XY)(\omega) = X(\omega) Y(\omega)$ est une v.a.r. (c'est une fonction mesurable).

Proposition 10.24 *Soit X et Y deux v.a.r. indépendantes. On a :*

$$E(XY) = E(X) E(Y) \qquad \textit{quand indépendance.} \tag{10.20}$$

(C'est évidemment faux si X et Y ne sont pas indépendantes : prendre $Y = X \neq 0$ avec $E(X) = 0$ et $E(X^2) > 0$.)

Preuve. Quand X et Y sont indépendantes, on dispose de (5.15) page 56 qui donne, avec $f = id$:

$$E(XY) = \int_z z\, dP_{XY}(z) = \iint_{x,y} xy\, dP_X(x) dP_Y(y) = \int_x x\, dP_X(x) \int_y y\, dP_Y(y) = E(X)E(Y).$$

Exercice 10.25 Réécrire la démonstration dans les cas de probabilités discrètes et continues.

Réponse. Cas discret et v.a.r. indépendantes, avec (8.19) :

$$P_{XY}(\{z\}) = \sum_{\substack{x_i,y_j\\ x_iy_j=z}} P_X(x_i)P_Y(y_j) = \sum_{x_i,y_j} P_X(x_i)P_Y(y_j)\delta_{x_iy_i}(\{z\}).$$

D'où, notant $F = (XY)(\Omega) = \{z \in \mathbb{R} : \exists (i,j) \in \mathbb{Z}^2,\ z = x_iy_j,\ x_i \in X(\Omega), y_i \in Y(\Omega)\}$, et avec $P(g \circ XY) = P_{XY}(g)$ (cf. (9.23)) :

$$\begin{aligned} P(XY) &= \sum_{z\in F} z\, P_{XY}(id) = \sum_{z\in F} z \sum_{x_i,y_j} P_X(x_i)P_Y(y_j)\delta_{z,x_iy_i} = \sum_{x_i,y_j} P_X(x_i)P_Y(y_j) \sum_{z\in F} z\, \delta_{z,x_iy_i} \\ &= \sum_{x_i,y_j} x_iy_j\, P_X(x_i)P_Y(y_j) = (\sum_{x_i} x_i\, dP_X(x_i))(\sum_{y_j} y_j\, dP_Y(y_j)). \end{aligned}$$

Cas continu et v.a.r. indépendantes de lois continues avec avec $\varphi(x,y) = xy$ et (5.14), et avec (8.24), :

$$P(XY) = P_{XY}(id) = (P_X \otimes P_Y)(id \circ \varphi) = \int_{(x,y)\in\mathbb{R}^2} xy\, \rho_X(x)\rho_Y(y)\, dxdy = \int_{x\in\mathbb{R}} x\rho_X(x)\, dx \int_{y\in\mathbb{R}} y\rho_Y(y)\, dy.$$

10.3 Exemples des lois usuelles

10.3.1 Alternative simple et loi binomiale

Alternative simple : $P(X{=}1) = p$ et $P(X{=}0) = 1{-}p$: on obtient $E(X) = 1 \times p + 0 \times (1{-}p)$:

$$E(X) = p. \tag{10.21}$$

C'est la position du centre de gravité d'une tige horizontale sans masse de longueur 1 à laquelle on a mis à droite ("en 1") le poids p et à gauche ("en 0") le poids $1{-}p$. En particulier, si $p = \frac{1}{2}$ le centre de gravité est au milieu, et si $p = 1$ le centre de gravité est à droite (à l'endroit de la seule masse présente).

Même espérance pour $X = 1_F$ (fonction caractéristique de l'ensemble $F \subset \Omega$) lorsque $P(F) = p$ (et $P(F^C) = 1{-}p$).

Espérance de la loi binomiale $b(n,p)$. Ici $P(X{=}k) = b(k;n,p) = P_X(k)$: on obtient avec $q = p{-}1$:

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k\, b(k;n,p) = \sum_{k=1}^{n} k\, C_n^k p^k q^{n-k} = \sum_{k=1}^{n} k\, \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!((n{-}1)-(k{-}1)!} p^{k-1} q^{n-k} \\ &= np \sum_{m=0}^{n-1} \frac{(n{-}1)!}{m!(n-1-m)!} p^m q)^{n-1-m} = np \sum_{m=0}^{n-1} C_{n-1}^m p^m q^{n-1-m} = np(p+q)^{n-1}, \end{aligned}$$

et donc :

$$E(X) = np \qquad \text{(espérance de la loi } b(n,p)). \tag{10.22}$$

10.3.2 Loi géométrique

Exercice 10.26 Montrer que $\displaystyle\sum_{k=1}^{\infty} kp^k = \frac{p}{(1-p)^2}$ pour $|p|<1$.

Réponse. $S(p)=\sum_{k=0}^{\infty}p^k$ est une série entière de rayon de convergence $R=1$. On peut donc dériver sous le signe $\sum$ pour $|p|<1$. On obtient $S'(p)=\sum_{k=1}^{\infty}kp^{k-1}=\frac{1}{p}\sum_{k=1}^{\infty}kp^k$. Et comme $\sum_{k=0}^{\infty}p^k=\frac{1}{1-p}$, on a aussi $S'(p)=\frac{1}{(1-p)^2}$. Donc $\frac{1}{p}\sum_{k=1}^{\infty}kp^k=\frac{1}{(1-p)^2}$. ▪

Ici $P(X{=}k)=(1-p)p^k$. D'où $E(X)=\sum_{k=0}^{\infty}k(1-p)p^k=(1-p)\sum_{k=0}^{\infty}kp^k=(1-p)\sum_{k=1}^{\infty}kp^k$. Donc :

$$E(X)=\frac{p}{1-p}. \tag{10.23}$$

10.3.3 Loi gaussienne

Soit X la v.a.r. gaussienne de loi $\mathcal{N}(m,\sigma^2)$, i.e. P_X de densité $p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$ sur $\mathbb{R}$.

$$\begin{aligned}E(X)&=\int_{\mathbb{R}}x\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx\\&=\frac{1}{\sigma\sqrt{2\pi}}\int_{\mathbb{R}}(x-m)e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx+m\int_{\mathbb{R}}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx\\&=\frac{1}{\sigma\sqrt{2\pi}}[-\sigma^2e^{-\frac{(x-m)^2}{2\sigma^2}}]_{-\infty}^{\infty}+m.\end{aligned} \tag{10.24}$$

Donc :

$$E(X)=m. \tag{10.25}$$

Exercice 10.27 Soit X la v.a.r. gaussienne de loi $\mathcal{N}(m,\sigma^2)$. Montrer que, pour $\lambda\in\mathbb{R}$:

$$E(e^{\lambda X})=e^{\lambda m+\frac{\lambda^2\sigma^2}{2}}. \tag{10.26}$$

Réponse.

$$\begin{aligned}E(e^{\lambda X})&=\int_{\omega}e^{\lambda X(\omega)}dP(\omega)=\int_{x}e^{\lambda x}dP_X(x)=\int_{x}e^{\lambda x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx=e^{\lambda m}\int_{y}e^{\lambda y}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{y^2}{2\sigma^2}}\,dy\\&=e^{\lambda m}\int_{y}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\sigma^2\lambda)^2}{2\sigma^2}+\frac{\sigma^2\lambda^2}{2}}\,dy=e^{\lambda m+\frac{\lambda^2\sigma^2}{2}}\int_{z}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{z^2}{2\sigma^2}}\,dz,\end{aligned}$$

la deuxième égalité par définition de la mesure image. ▪

Exercice 10.28 Montrer que si pour tout $\lambda\in\mathbb{R}$ on a $E(e^{\lambda X})=e^{\lambda m+\frac{\lambda^2\sigma^2}{2}}$, alors X est de loi $\mathcal{N}(m,\sigma^2)$. ▪

10.3.4 Loi de Poisson

Loi de Poisson $P_X(k)=e^{-\lambda}\frac{\lambda^k}{k!}$ pour $k\in\mathbb{N}$ et $\lambda>0$:

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

10.3.5 Exercices

Exercice 10.29 Un berger trie un groupe de n chèvres dont m ne sont pas vaccinées, $0<m<n$. Les chèvres rentrent une à une dans un sas. Le berger repère, à chaque passage d'une chèvre, si elle est vaccinée ou non. On note X la v.a. indiquant le rang d'apparition de la dernière chèvre non vaccinée.

1. Donner l'univers Ω, son cardinal, et l'image $X(\Omega)$.
2. Calculer $P(X{=}k)$ pour tout $k\in[m,n]_{\mathbb{N}}$. Vérifier que $\sum_k P(X{=}k)=1$.
3. Montrer que $E(X)=\frac{(n+1)m}{m+1}$.

Réponse. 1. Ω est l'ensemble des n-uplets ordonnés $\vec{\omega} = (\omega_1, ..., \omega_n)$ où ω_i est l'une des chèvres. $|\Omega| = n!$. Et $X(\Omega) \subset \{m, m+1, ..., n\}$ (avec $X(\vec{\omega}) = m$ si $\vec{\omega}$ a ses m premières composantes = les chèvres non vaccinées, et $X(\vec{\omega}) = n$ si $\vec{\omega}$ a sa dernière composante = une chèvre non vaccinée).

2. Déterminons le cardinal de $(X{=}k)$: la m-ième (la dernière) est passée en k-ième position, soit m possibilités ; les $m{-}1$ chèvres restantes non vaccinées sont passées parmi les $k{-}1$ premières, soit $A_{k-1}^{m-1} = \frac{(k-1)!}{(k-m)!}$ possibilités ; et les $n{-}m$ chèvres vaccinées sont mises dans les $n{-}m$ positions restantes, soit $(n{-}m)!$ possibilités. D'où $mA_{k-1}^{m-1}(n{-}m)!$ est le nombre de n-uplets possibles. D'où :

$$P(X{=}k) = \frac{mA_{k-1}^{m-1}(n{-}m)!}{n!} = \frac{m(k-1)!(n{-}m)!}{(k-m)!n!} = \frac{(k-1)!m!(n{-}m)!}{(k-m)!(m-1)!n!} = \frac{C_{k-1}^{m-1}}{C_n^m}. \tag{10.27}$$

Et $\displaystyle\sum_{k=m}^{n} C_{k-1}^{m-1} = C_n^m$, cf. (1.27), donc $\sum_k P(X{=}k) = 1$.

3.

$$E(X) = \sum_{k=m}^{n} kP(X{=}k) = \frac{1}{C_n^m}\sum_{k=m}^{n} kC_{k-1}^{m-1} = \frac{1}{C_n^m}\sum_{k=m}^{n}\frac{k!}{(m-1)!(k-m)!} = \frac{m}{C_n^m}\sum_{k=m}^{n}\frac{k!}{m!(k-m)!}$$
$$= \frac{m}{C_n^m}\sum_{k=m}^{n} C_k^m = \frac{m}{C_n^m}C_{n+1}^{m+1} = m\frac{m!(n-m)!}{n!}\frac{(n+1)!}{(m+1)!(n-m)!} = \frac{m(n+1)}{(m+1)}.$$

Exercice 10.30 Deux archers s'affrontent au tir à l'arc. La cible est constituée de quatres zones A, B, C et D, qui rapportent respectivement 10, 5, 1 et 0 point(s). Le jeu consiste à tirer trois flèches et le score final est la somme des trois résultats (les tirs sont indépendants). Le jeu est répété un grand nombre de fois.

Le premier archet, M. Flute, tire dans A 6 fois sur 10, dans B 1 fois sur 10, dans C 1 fois sur 10. Le deuxième archet, M. Saxo, tire dans A 5 fois sur 10, dans B 3 fois sur 10, dans C 1 fois sur 10.

1- Quelle est l'espérance du score de M. Flute pour le tir d'une flèche ?

2- Quelle est l'espérance du score final obtenu (3 flèches) par M. Flute ?

3- Mêmes questions pour M. Saxo.

4- M. Flute et M. Saxo tirent chacun une flèche. Remplir le tableau ci-dessous :

Flute \ Saxo	10	5	1	0	Total
10					
5					
1					
0					
Total					

dans lequel la case intersection de la ligne i et de la colonne j donne la probabilité pour que M. Flute obtienne i points et M. Saxo obtienne j points (on précisera l'hypothèse qu'il est nécessaire de faire pour justifier cette probabilité). La "colonne Total" doit donner, à la ligne i la probabilité pour que M. Flute obtienne i points, indépendamment du résultat de M. Saxo ; et la "ligne Total" doit donner, à la colonne j la probabilité pour que M. Saxo obtienne i points, indépendamment du résultat de M. Flute.

5- Quelle est la probabilité que M. Flute ait un score identique à celui de M. Saxo après un seul tir de flèche ?

6- Quelle est la probabilité que M. Flute ait un score strictement plus élevé que M. Saxo après un seul tir de flèche ?

7- Quelle est la probabilité que M. Flute ait un score identique ou plus élevé que M. Saxo après un seul tir de flèche ?

8- Quelle est la probabilité que M. Flute ait un score identique ou plus élevé que celui de M. Saxo, conditionnellement au fait que les deux archers n'ont pas tiré hors zone ?

Réponse. 1- Soit P_F la loi de probabilité de M. Flute : $P_F(A) = \frac{6}{10}$, $P_F(B) = \frac{1}{10}$, $P_F(C) = \frac{1}{10}$, $P_F(D) = \frac{2}{10}$. Donc $E(F) = 10P_F(A) + 5P_F(B) + 1P_F(C) + 0P_F(D) = 10\frac{6}{10} + 5\frac{1}{10} + 1\frac{1}{10} + 0 = \frac{66}{10}$ (espérance de résultat $= 6,6$ points).

2- Espérance finale (3 tirs indépendants) : $3E(F) = \frac{198}{10}$ (espérance de résultat $= 19,8$ points).

3- $P_S(A) = \frac{5}{10}$, $P_S(B) = \frac{3}{10}$, $P_S(C) = \frac{1}{10}$, $P_S(D) = \frac{1}{10}$. Donc $E(S) = 10\frac{5}{10} + 5\frac{3}{10} + 1\frac{1}{10} + 0 = \frac{66}{10}$ et $3E(S) = \frac{198}{10}$. Même résultat : les deux tireurs peuvent espérer un même score.

4- Résultat rapide : Le tableau donne le produit des probabilités en supposant que les tirs sont indépendants : soit $P((F{=}x)\cap(S{=}y)) = P(F{=}x)P(S{=}y)$:

Flute \ Saxo	10	5	1	0	Total
10	$\frac{30}{100}$	$\frac{18}{100}$	$\frac{6}{100}$	$\frac{6}{100}$	$\frac{60}{100}$
5	$\frac{5}{100}$	$\frac{3}{100}$	$\frac{1}{100}$	$\frac{1}{100}$	$\frac{10}{100}$
1	$\frac{5}{100}$	$\frac{3}{100}$	$\frac{1}{100}$	$\frac{1}{100}$	$\frac{10}{100}$
0	$\frac{10}{100}$	$\frac{6}{100}$	$\frac{2}{100}$	$\frac{2}{100}$	$\frac{20}{100}$
Total	$\frac{50}{100}$	$\frac{30}{100}$	$\frac{10}{100}$	$\frac{10}{100}$	$\frac{100}{100}$

Et $P(F=x) = P(F{=}x\cap S{=}10) + P(F{=}x\cap S{=}5) + P(F{=}x\cap S{=}1) + P(F{=}x\cap S{=}0)$ donne la somme sur une ligne.

Résultat détaillé : univers $\Omega = \{A,B,C,D\}^2 = \{(\omega_F,\omega_S) \in \{A,B,C,D\}^2\}$ ensemble des couples donnant un résultat de Flute et Saxo. La probabilité sur Ω est $Q((F,S){=}(x,y)) = P(F{=}x)P(S{=}y) = P_F(x)P_S(y)$ donne la valeur dans le tableau au croisement de la ligne x et de la colonne y.

5- Résultat rapide : Diagonale : $P(F{=}S) = P((F{=}A)\cap(S{=}A)) + P((F{=}B)\cap(S{=}B)) + P((F{=}C)\cap(S{=}C)) + P((F{=}D)\cap(S{=}D)) = \frac{30+3+1+2}{100} = \frac{36}{100}$ est la probabilité que les tireurs aient le même score.

Résultat détaillé : l'ensemble $(F > S)$ est l'ensemble $\{(\omega_F,\omega_S) : \omega_F > \omega_S$. D'où $Q(F > S)$ = somme des termes de la matrice supérieure stricte.

6- $P(F > S) = P((F{=}10)\cap(S{=}5\cup S{=}1\cup S{=}0)) + ...$ = sur-diagonale stricte = $\frac{18+6+6+1+1+1}{100} = \frac{33}{100}$ pour un score strictement plus élevé de M. Flute. (Sous-diagonale stricte : $\frac{5+5+10+3+6+2}{100} = \frac{31}{100}$ pour un score strictement plus élevé de M. Saxo.)

7- $P(F \geq S) = \frac{36}{100} + \frac{33}{100} = \frac{69}{100}$.

8- $P(F\in[1,10]\cup S\in[1,10]) = \frac{30+18+6+5+3+1+5+3+1}{100} = \frac{72}{100}$.

$P(F\geq S|(F\in[1,10]\cup S\in[1,10])) = \frac{P((F\geq S)\cap(F\in[1,10]\cup S\in[1,10]))}{P(F\in[1,10]\cup S\in[1,10])} = \frac{\frac{30+18+6+3+1+1}{100}}{\frac{72}{100}} = \frac{59}{72}$. ▪

10.4 Variance et écart type

[Variance and standard deviation.]

10.4.1 Définitions

Définition 10.31 Soit X une v.a.r. sur Ω et soit $E(X) = \overline{X} \in \mathbb{R}$ son espérance (on suppose qu'elle existe). Sa variance $\mathrm{Var}(X)$ est (quand elle existe) :

$$\mathrm{Var}(X) = P((X{-}\overline{X}1_\Omega)^2) \overset{\text{noté}}{=} E((X{-}\overline{X}1_\Omega)^2) \overset{\text{noté}}{=} E((X{-}\overline{X})^2), \tag{10.28}$$

soit :

$$\mathrm{Var}(X) = \int_{\omega\in\Omega} (X(\omega){-}\overline{X})^2\, dP(\omega) = ||X{-}\overline{X}1_\Omega||^2_{L^2(\Omega,P)} \tag{10.29}$$

C'est "l'énergie" de la v.a.r. centrée $X{-}\overline{X}1_\Omega =^{\text{noté}} X - \overline{X}$ relativement à la mesure P.

Définition 10.32 L'écart type est :

$$\sigma(X) = \sqrt{\mathrm{Var}(X)} = ||X{-}\overline{X}||_{L^2(\Omega,P)}, \tag{10.30}$$

homogène à une "norme"=longueur, contrairement à la variance qui est homogène au "carré d'une norme" = surface = énergie.

Donc :

$$\mathrm{Var}(X) = \sigma(X)^2. \tag{10.31}$$

Définition 10.33 Si $\sigma(X) = 1$ alors la v.a.r. X est dite réduite.

Définition 10.34 Si $E(X) = 0$ et si $\sigma(X) = 1$ alors la v.a.r. X est dite centrée réduite.

Proposition 10.35 $\mathrm{Var}(X) = 0 \iff X$ *est une fonction constante (presque surement). (On est dans le cadre "déterministe".)*

Preuve. $\mathrm{Var}(c1_\Omega) = P(0) = 0$: la variance s'annule sur les fonctions constantes. Réciproquement, si $\mathrm{Var}(X) = 0$, alors (10.29) donne : $X(\omega){-}\overline{X} = 0$ P-presque surement, donc X est une fonction constante P-presque surement (i.e. pour P-presque tout ω). ▪

10.4.2 Avec la loi image

Proposition 10.36 *Soit X une v.a.r. qui admet une espérance et une variance, et $P_X = P \circ X^{-1}$ la loi de probabilité de la v.a.r. X.*

Alors la variance est le moment d'inertie de la mesure P_X par rapport à $\overline{X}$ (où P_X est considérée comme distribution de masse) :

$$\mathrm{Var}(X) = \int_{x\in\mathbb{R}} (x-\overline{X})^2 \, dP_X(x), \tag{10.32}$$

i.e. moment d'ordre 2 de l'identité id par rapport à $\overline{X}$ pour la mesure P_X, cf. (15.16).

On dit que la variance donne une idée de la "dispersion" autour de la valeur moyenne $\overline{X}$.

Preuve. On applique (9.24) : $\int_{x\in\mathbb{R}} g(x)\, dP_X = \int_{\omega\in X^{-1}(\mathbb{R})} (g(X(\omega))\, dP$, proposition de la mesure image. Soit $g(x) \stackrel{\text{déf}}{=} (x-\overline{X})^2$: on a donc posé $g = (id - \overline{X}1_{\mathbb{R}})^2$. On a $g(X(\omega)) = (X(\omega) - \overline{X})^2$, d'où $\int_{x\in\mathbb{R}} (x-\overline{X})^2 \, dP_X(x) = \int_{\omega\in\Omega} (X(\omega)-\overline{X})^2 \, dP(\omega)$, d'où (10.32) grâce à (10.29). ▪

Donc :

1- dans le cas discret $P_X = \sum_i p_i \delta_{x_i}$:

$$\mathrm{Var}(X) = \sum_{i\in I} (x_i - \overline{X})^2 p_{x_i} = \sigma^2(X), \tag{10.33}$$

2- dans le cas continu $dP_X = p(x)\, dx$:

$$\mathrm{Var}(X) = \int_{x\in\mathbb{R}} (x - \overline{X})^2 \, p(x)\, dx = \sigma^2(X). \tag{10.34}$$

10.4.3 Propriétés

Soit :

$$L^2(\Omega, P) = \{X : \Omega \to \mathbb{R} \text{ t.q. } P(X) < \infty\} = \{X : \Omega \to \mathbb{R} \text{ t.q. } \int_{\omega\in\Omega} X(\omega)^2 \, dP(\omega) < \infty\} \tag{10.35}$$

l'ensemble des fonctions (mesurables) de carré P-intégrable (fonctions d'énergie finie). C'est un sous-espace vectoriel de $(\mathcal{F}(\Omega;\mathbb{R}), +, .)$ (immédiat par linéarité de l'intégrale).

Proposition 10.37 *La forme bilinéaire $(\cdot,\cdot)_P : (X,Y) \to (X,Y)_P$ définie par :*

$$(X,Y)_P \stackrel{\text{déf}}{=} E(XY) = \int_{\omega\in\Omega} X(\omega) Y\omega)\, dP(\omega). \tag{10.36}$$

est un produit scalaire dans $L^2(\Omega, P)$.

Preuve. $(\cdot,\cdot)_P : (X,Y) \to (X,Y)_P$ est trivialement bilinéaire et symétrique. et $\int X^2(\omega)\, dP = 0$ implique X^2 nulle P-presque surement, donc X nulle P-presque surement, donc X est nulle au sens L^2, voir cours d'intégration. ▪

On note :

$$||X||_P = \sqrt{(X,X)_P} = ||X||_{L^2(\Omega),P} = ||X||_2 = (\int_\Omega |X|^2 \, dP)^{\frac12} = (P(X^2))^{\frac12} = (E(X^2))^{\frac12} \tag{10.37}$$

la norme associée.

Proposition 10.38 *Soit $X \in L^2(\Omega, P)$, i.e. soit X une v.a.r. telle que X^2 admette une espérance (X est dans l'espace des fonctions de carré intégrable pour la mesure P). Alors :*

1. *On a $L^2(\Omega, P) \subset L^1(\Omega, P)$: si $X \in L^2(\Omega, P)$ alors $X \in L^1(\Omega, P)$. (Reste vraie si on remplace P par une mesure finie.) Et :*

$$|E(X)| \le E(|X|) \le \sqrt{E(X^2)}. \tag{10.38}$$

2. *X admet une variance et on a :*

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 \qquad (= \overline{X^2} - \big(\overline{X}\big)^2). \tag{10.39}$$

(Réciproquement, si $\mathrm{Var}(X)$ existe, ce qui nécessite l'existence de $\overline{X}$, alors $E(X^2)$ existe, i.e. $X \in L^2(\Omega, P)$.)

3. *Si $\sigma(X) = 0$, alors X est une fonction constante : la fonction constante $\overline{X}1_\Omega$:*

$$\mathrm{Var}(X) = 0 \quad \Longleftrightarrow \quad X = \overline{X}1_\Omega. \tag{10.40}$$

Dans ce cas X est une v.a.r. bien déterminée (on rentre dans le cas déterministe : il n'y a plus de "dispersion" autour de la valeur moyenne).
(Réciproque immédiate : si X est une fonction constante alors $\mathrm{Var}(X) = 0$, cf. (10.29).)

4. *Si $\sigma(X) \neq 0$, alors la v.a.r. $\frac{X-\overline{X}}{\sigma(X)}$ est centrée réduite. En particulier :*

$$\mathrm{Var}(\frac{X - \overline{X}}{\sigma(X)}) = 1. \tag{10.41}$$

5. *Si $b \in \mathbb{R}$, alors la v.a.r. $X{+}b1_\Omega \stackrel{\text{noté}}{=} X{+}b$ admet une variance et :*

$$\mathrm{Var}(X + b) = \mathrm{Var}X, \tag{10.42}$$

et la variance (qui est une "mesure de la dispersion") ne dépend pas de b. En particulier :

$$\mathrm{Var}(X - \overline{X}) = \mathrm{Var}X. \tag{10.43}$$

6. *Si $\alpha \in \mathbb{R}$, alors la v.a.r. αX admet une variance et :*

$$\mathrm{Var}(\alpha X) = \alpha^2 \mathrm{Var}(X), \tag{10.44}$$

ou encore $\sigma(\alpha X) = |\alpha|\, \sigma(X)$. D'où si $\alpha, b \in \mathbb{R}$:

$$\mathrm{Var}(\alpha X + b) = \alpha^2 \mathrm{Var}(X). \tag{10.45}$$

7. *Pour tout $a \in \mathbb{R}$:*

$$||X - a||_2^2 = ||X - \overline{X}||_2^2 + (a - \overline{X})^2, \tag{10.46}$$

(propriété des moments d'inertie) et en particulier, pour tout $a \in \mathbb{R}$:

$$||X - a||_2^2 \geq ||X - \overline{X}||_2^2, \tag{10.47}$$

i.e. le moment d'inertie autour d'un point a est minimal si on prend a le centre d'inertie = le centre de gravité.

Preuve.

1. C'est Cauchy–Schwarz dans $L^2(\Omega, P)$: $|(1_\Omega, |X|)_P| \leq ||1_\Omega||_P ||X||_P$, soit $|\int_\Omega |X(\omega)|\, dP| \leq (1)^{\frac12}\, (\int_\Omega X(\omega)^2\, dP)^{\frac12}$. Idem en remplaçant $|X|$ par X.
2. Comme $(X - \overline{X})^2 = X^2 - 2\overline{X}X + \overline{X}^2$, chaque terme du membre de droite admettant une espérance, la linéarité de E donne l'existence de $\mathrm{Var}(X) = E(X^2) - 2\overline{X}^2 + \overline{X}^2 = E(X^2) - \overline{X}^2$.
3. Si $\sigma(X) = 0$, alors $||X{-}\overline{X}1_\Omega||_P^2 = 0$, donc $X - \overline{X}1_\Omega = 0$ presque surement.
4. Sinon, avec la linéarité de E, $\mathrm{Var}(\frac{X-\overline{X}}{\sigma(X)}) = E(\frac{(X-\overline{X})^2}{\sigma^2(X)}) = \frac{E((X-\overline{X})^2)}{\sigma^2(X)} = \frac{\mathrm{Var}(X)}{\sigma^2(X)} = 1$.
5. $(X{+}b) - E(X{+}b) = X + b - E(X) - b = X - \overline{X}$, d'où le résultat.
(Sans abus de notation : $(X{+}b1_\Omega) - E(X{+}b1_\Omega)1_\Omega = X + b1_\Omega - E(X)1_\Omega - E(b1_\Omega)1_\Omega = X + b1_\Omega - E(X)1_\Omega - b1_\Omega = X - E(X)1_\Omega$.)
6. $E((\alpha X)^2 - (E(\alpha X))^2) = E(\alpha^2 X^2 - (\alpha E(X))^2) = E(\alpha^2 X^2 - \alpha^2 E(X)^2) = \alpha^2 E(X^2) - \alpha^2 E(X)^2$ par linéarité de E.
7. $(X(\omega){-}a)^2 = ((X(\omega){-}\overline{X}) + (\overline{X}{-}a))^2 = (X(\omega){-}\overline{X})^2 + (\overline{X}{-}a)^2 + 2(X(\omega){-}\overline{X})(\overline{X}{-}a)$, et donc $E((X{-}a)^2) = E((X{-}\overline{X})^2) + (\overline{X}{-}a)^2 E(1_\Omega) + 2(\overline{X}{-}a)E(X{-}\overline{X}) = E((X{-}\overline{X})^2) + (\overline{X}{-}a)^2 + 0$.

▪

10.4.4 Exemples

Exemple 10.39 Loi de Bernoulli pour $X : \{a,b\} \to \{0,1\}$ avec $X(a) = 1$ (succès) et $X(b) = 0$ (échec), et avec $P(X{=}1) = P_X(1) = p$, $P(X{=}0) = P_X(0) = 1{-}p =^{\text{noté}} q$. On a $E(X) = X(a)P_X(1) + X(b)P_X(0) = 1 * p + 0 * q = p$ (déjà vu). D'où :

$$\begin{aligned}\mathrm{Var}(X) &= E((X{-}p)^2) = (1{-}p)^2 P_X(1) + (0{-}p)^2 P_X(0) = (1{-}p)^2 p + p^2(1{-}p)\\ &= p(1{-}p) = pq.\end{aligned} \tag{10.48}$$

En particulier $\mathrm{Var}(X) \le \frac14$ car $p(1{-}p) \le \frac14$ (cf. tableau de variations de $f(p) = p(1{-}p)$).

Exercice 10.40 Loi de Bernoulli généralisée : $X(a) = \alpha$, $X(b) = \beta$, $P(X{=}\alpha) = P_X(\alpha) = p$, $P(X{=}\beta) = P_X(\beta) = 1{-}p =^{\text{noté}} q$. Calculer $E(X)$ et $\mathrm{Var}(X)$.

Réponse. On a $E(X) = \alpha P_X(\alpha) + \beta P_X(\beta) = \alpha p + \beta q = \beta + (1-\alpha)p$.
On a $E(X^2) = \alpha^2 P_X(\alpha) + \beta^2 P_X(\beta) = \alpha^2 p + \beta^2 q$. D'où :

$$\begin{aligned}\mathrm{Var}(X) &= \alpha^2 p + \beta^2 q - (\alpha p + \beta q)^2 = \alpha^2(p-p^2) + \beta^2(q-q^2) - 2\alpha\beta qp\\ &= pq(\alpha-\beta)^2.\end{aligned} \tag{10.49}$$

En particulier $\mathrm{Var}(X) \le \frac14 (a-b)^2$.

Exemple 10.41 Loi gaussienne de moyenne m et d'écart type σ, cf. (6.44), i.e. loi P_X sur $\mathbb{R}$ de densité $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$. Son espérance est m, cf. (10.24).

Et $(x{-}m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} = (x{-}m)((x{-}m)e^{-\frac{(x-m)^2}{2\sigma^2}})$ "$= uv'$" donne par intégration par parties :

$$\mathrm{Var}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} (x{-}m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx = \frac{1}{\sigma\sqrt{2\pi}}([-\sigma^2(x{-}m)e^{-\frac{(x-m)^2}{2\sigma^2}}]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \sigma^2 e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx).$$

Donc $\mathrm{Var}(X) = 0 + \sigma^2 = \sigma^2$ et :

$$\sigma(X) = \sigma. \tag{10.50}$$

Quand $m = 0$ la loi gaussienne est centrée, et quand $\sigma = 1$ la loi gaussienne est réduite. Et elle est donc centrée et réduite quand $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Exercice 10.42 Soit un troupeau de n chèvres. Les chèvres peuvent indifféremment restées dehors, noté A_0, ou aller dans un des deux abris A_1 et A_2, chaque abris étant assez grand pour abriter toutes les chèvres. Soit Y la v.a. donnant le nombre de chèvres dans l'abris A_1 et Z la v.a. donnant le nombre d'abris vides.

11- Donner la loi $P(Y = k)$ de Y, et en particulier $P(Y = 0)$ et $P(Y = n)$.
12- Donner l'espérance et la variance de Y.
13- Donner l'univers sur lequel agit Y et donner $\mathrm{Im}(Y)$.
21- Que vaut l'événement $(Y{=}0) \cap (Z{=}0)$?
22- Que vaut $P((Y{=}0) \cap (Z{=}2))$?
23- En déduire que $P((Y{=}0) \cap (Z{=}1)) = \frac{2^n-1}{3^n}$
24- Les lois de Y et Z sont-elles indépendantes ?
25- Donner l'univers sur lequel agit Z et $\mathrm{Im}(Z)$.
31- Calculer $P((Y{=}k) \cap (Z{=}2))$ pour $k = 1, ..., n$.
32- Vérifier que $P((Y{=}k) \cap (Z{=}1)) = \frac{C_n^k}{3^n}$ pour $k = 1, ..., n$.
33- En déduire que $P((Y{=}k) \cap (Z{=}0))$.
41- Donner la loi de Z : commencer par donner $P(Z = 2)$ puis $P(Z = 1)$.
42- Calculer l'espérance de Z.

Réponse.

11- La probabilité qu'une chèvre soit dans A_1 est $p = \frac13$, et $q = 1 - p = \frac23$ est la probabilité qu'une chèvre aille ailleurs (dans A_0 ou A_2). La loi de Y est donc la loi binomiale $B(n,p) = B(n, \frac13)$. Ainsi $P(Y = k) = B(k; n, p) = C_n^k p^k (1-p)^{n-k} = C_n^k \frac{2^{n-k}}{3^n}$, pour $k \in [0,n]_{\mathbb{N}}$. En particulier $P(Y = 0) = \frac{2^n}{3^n}$ (chaque chèvre à la probabilité $\frac{2}{n}$ d'aller ailleurs que dans A_1), et $P(Y = n) = \frac{1}{3^n}$ (chaque chèvre à la probabilité $\frac13$ d'aller dans A_1).

12- Donc $E(Y) = np = \frac{n}{3}$ et $V(Y) = npq = \frac{2n}{9}$, cf. loi binomiale.

13- Soit $C = \{c_1, ..., c_n\}$ l'ensemble des chèvres. L'univers de Y est suggéré par la loi binomiale. on fait n lancers, le i-ème correspondant au choix de c_i, d'issue $a = A_1$ et $b =$ ailleurs. Donc l'univers est

$\Omega = \{a,b\}^n$. Et pour $\vec{\omega} = (\omega_1, ..., \omega_n) \in \{a,b\}^n$, on a $Y(\vec{\omega})$ = le nombre de composante $\omega_i = A$, soit $Y = \sum_{i=1}^n 1_{\{a\}}(\omega_i)$.

Et $Y(\Omega) = [0,n]_{\mathbb{N}}$.

21- Z=0 indique qu'aucun abris n'est vide : en particulier A_1 n'est pas vide ; et Y=0 indique que A_1 est vide : l'événement $(Y{=}0) \cap (Z{=}0)$ est impossible.

22- Z=2 indique toutes les chèvres sont dehors, et Y=0 indique que A_1 est vide, cette dernière information étant inutile. Donc $P((Y{=}0) \cap (Z{=}2)) = P(Z{=}2)$ est la probabilité que toutes les chèvres sont dehors, chaque chèvre ayant la probabilité $\frac{1}{3}$ d'être dehors. Donc $P((Y{=}0) \cap (Z{=}2)) = P(Z{=}2) = \frac{1}{3^n}$.

23- $P(Y=0) = P((Y{=}0) \cap (Z{=}0)) + P((Y{=}0) \cap (Z{=}1)) + P((Y{=}0) \cap (Z{=}2))$, avec $P(Y=0) = (\frac{2}{3})^n$, car chaque chèvre à la probabilité $\frac{2}{3}$ de ne pas aller dans A_1. Donc $\frac{2^n}{3^n} = 0 + P((Y{=}0) \cap (Z{=}1)) + \frac{1}{3^n}$, donc $P((Y{=}0) \cap (Z{=}1)) = \frac{2^n-1}{3^n}$.

24- $P((Y{=}0) \cap (Z{=}2)) = P(Z{=}2) = \frac{1}{3^n}$ est différent de $P((Y{=}0))P((Z{=}2))$ car $P(Y{=}0) \neq 1$: les v.a. Y et Z ne sont pas indépendantes.

25- Z a même ensemble de définition que Y (sinon $(Y=0) \cap (Z=0)$ n'a pas de sens). Mais on a un problème avec l'univers de Y défini précédemment.

Changeons l'univers de Y en posant $\Omega = (\{A_0, A_1, A_2\})^n$ l'ensemble des issues possibles. Et on pose $Y : \Omega \to [0,n]_{\mathbb{N}}$ définie par $Y(\vec{\omega}) = \sum_{i=1}^n 1_{\{A_1\}}(\omega_i)$ (compte le nombre de chèvres dans A_1).

Et on pose $Z : \Omega \to \{0,1,2\}$ définie par :

$Z(\vec{\omega}) = 2$ ssi $\forall i = 1, ..., n$ on a $\omega_i = A_0$ (les abris sont vides),

$Z(\vec{\omega}) = 0$ ssi $\exists i,j \in [1,n]_{\mathbb{N}}$ t.q. $\omega_i = A_1$ et $\omega_j = A_2$ (aucun abris n'est vide),

$Z(\vec{\omega}) = 1$ sinon (il existe ω_i t.q. $\omega_i \neq A_0$ et notant A_{k_i}, avec $k_i = 1$ ou 2, l'abris dans lequel se trouve ω_i, alors pour tout $j \neq i$ on a $\omega_j \neq A_\ell$ où $\ell = 3 - k_i$).

31- $Z = 2$ indique que toutes les chèvres sont dehors, donc aucune dans A_1, donc pour $k \geq 1$ on a $P((Y{=}k) \cap (Z{=}2)) = 0$.

32- Soit $k \in [1,n]_{\mathbb{N}}$. $P((Y{=}k) \cap (Z{=}1))$ impose donc k chèvres dans A_1 parmi n, soit C_n^k issues possibles, les autres étant dans A_0, soit une issue possible. Et le nombre total d'issues est le nombre de fonctions $C \to \{A_1, A_2, A_3\}$, soit 3^n. Donc $P((Y{=}k) \cap (Z{=}1)) = \frac{C_n^k}{3^n}$ pour $k = 1, ..., n$.

33- Donc $P((Y{=}k) \cap (Z{=}0)) = P(Y{=}k) - P((Y{=}k) \cap (Z{=}1)) - P((Y{=}k) \cap (Z{=}2)) = C_n^k \frac{2^{n-k}-1}{3^n}$.

41- $P(Z{=}2) = \sum_{k=0}^n P((Y{=}k) \cap (Z{=}2)) = P((Y{=}0) \cap (Z{=}2)) + \sum_{k=1}^n P((Y{=}k) \cap (Z{=}2)) = \frac{1}{3^n} + 0$.

$P(Z{=}1) = \sum_{k=0}^n P((Y{=}k) \cap (Z{=}1)) = P((Y{=}0) \cap (Z{=}1)) + \sum_{k=1}^n P((Y{=}k) \cap (Z{=}1)) = \frac{2^n-1}{3^n} + \sum_{k=1}^n \frac{C_n^k}{3^n}$, avec $\sum_{k=1}^n \frac{C_n^k}{3^n} = \frac{1}{3^n}(-C_n^0 + \sum_{k=0}^n C_n^k) = \frac{1}{3^n}(-1 + (1+1)^n) = \frac{2^n-1}{3^n}$. Donc $P(Z{=}1) = \frac{2^n-1}{3^n} + \frac{2^n-1}{3^n} = \frac{2^{n+1}-2}{3^n}$.

$P(Z{=}0) = 1 - P(Z{=}1) - P(Z{=}2) = \frac{3^n - 2^{n+1} - 1}{3^n}$.

42- $E(Z) = 0P(Z{=}0) + 1P(Z{=}1) + 2P(Z{=}2) = 0 + \frac{2^{n+1}-2}{3^n} + \frac{2}{3^n} = \frac{2^{n+1}}{3^n} = 2(\frac{2}{3})^n$. ▪

Exercice 10.43 Soit n articles $a_1, ..., a_n$, avec $n \geq 2$, de prix respectifs $\alpha_1, ..., \alpha_n$. Un client achète aléatoirement k articles ($0 \leq k \leq n$). Soit S le prix payé par ce client. But : calculer l'espérance et la variance de S.

Pour ce : On note Y la v.a. qui donne le nombre d'articles achetés par le client (donc entre 0 et n). On suppose que la loi de probabilité de la v.a. Y est la loi uniforme. On note X_i la v.a. qui vaut 1 si le client achète a_i et 0 sinon.

10- Donner l'univers Ω sur lequel sont définis Y et X, ainsi que $|\Omega|$.

11- Donner $P(Y{=}k)$, puis déterminer P.

12- Calculer $P(X_i{=}1|Y{=}k)$.

13- En déduire que $P(X_i{=}1) = \dfrac{1}{2}$.

14- Calculer $E(X_i)$ et $V(X_i)$.

21- Exprimer S en fonction des X_i et des α_i.

22- Calculer $E(S)$ en fonction des α_i.

3- Soit $(i,j) \in [1,n]_{\mathbb{N}}^2$ avec $i \neq j$.

31- Calculer $P((X_i{=}1) \cap (X_j{=}1)|Y{=}k)$ pour $k = 0, ..., n$.

32- En déduire que $P((X_i{=}1) \cap (X_j{=}1)) = \dfrac{1}{3}$. Les v.a. X_i et X_j sont-elles indépendantes ?

33- Calculer la covariance $\mathrm{cov}(X_i, X_j)$ et le coefficient de corrélation.

34- En déduire $V(S) = \dfrac{1}{12}(\sum\limits_{i=1}^n \alpha_i)^2 + \dfrac{1}{6}\sum\limits_{i=1}^n \alpha_i^2$.

Réponse. 10- Soit $E = \{a_1, ..., a_n\}$ l'ensemble des articles.

L'ensemble des issues est $\Omega = \mathcal{P}(E)$ l'ensemble des sous-ensembles de E (articles achetés par un client), de cardinal 2^n.

Y est Z sont définies sur l'espace probabilisé $(\Omega,\mathcal{P}(\Omega),P)$, où P est à définir grâce à l'hypothèse P_Y uniforme.

$Y:\Omega\to[0,n]_{\mathbb{N}}$ avec, pour $F\subset E$, $Y(F)=|F|$ cardinal de F.

$X_i:\Omega\to\{0,1\}$ avec, pour $F\subset E$, $X_i(F)=\delta_{a_i}(F)=1_F(a_i)$.

11- Pour $k=0,...,n$ (avec $k=0$ si le client n'achète rien), on a $P_Y(k)=\dfrac{1}{n+1}$ (loi uniforme).

Et $P_Y=P\circ Y^{-1}$, donc $P_Y(k)=P(\{F\subset\Omega:|F|=k\})=\frac{1}{n+1}$.

On suppose l'équiprobabilité du choix d'un même nombre d'articles, i.e. $P(F)=P(G)$ quand $|F|=|G|$.

Et il y a C_n^k sous-ensembles de cardinal $|F|=k$, parmi 2^n sous-ensembles de Ω. Donc la probabilité d'un choix de k éléments donnés est $\dfrac{1}{n+1}\dfrac{C_n^k}{2^n}=P(F)$ pour un $F\subset\Omega$ tel que $|F|=k$. C'est la loi P de probabilité cherchée.

12- $((X_i{=}1|Y{=}k))$ est l'ensemble des sous-ensembles de taille k qui contiennent a_i. Pour $k\geq1$, il y a C_{n-1}^{k-1} tels sous-ensembles parmi les C_n^k sous-ensembles de taille k (puisque a_i est fixé), donc en proportion il y a $P((X_i{=}1|Y{=}k))=\dfrac{C_{n-1}^{k-1}}{C_n^k}=\dfrac{(n-1)!k!(n-k)!}{(k-1)!(n-k)!n!}=\dfrac{k}{n}$ (hypothèse d'équiprobabilité du choix d'un même nombre d'articles).

Et pour $k=0$, comme $(Y{=}0)$ signifie que le client n'achète rien, $(X_i{=}1)$ est incompatible avec $(Y{=}0)$, donc $P(X_i{=}1|Y{=}0)=\dfrac{P((X_i{=}1)\cap(Y{=}0))}{P(Y{=}k)}=0$. Donc la formule $=\frac{k}{n}$ est vrai pour tout $k=0,...,n$.

12'- Calcul en se servant de la définition d'une probabilité conditionnelle : pour $1\leq k\leq n$: $P(X_i{=}1|Y{=}k)=\dfrac{(P((X_i{=}1)\cap(Y{=}k))}{P(Y{=}k)}=\dfrac{\frac{1}{n+1}\frac{C_{n-1}^{k-1}}{2^n}}{\frac{1}{n+1}\frac{C_n^k}{2^n}}=\dfrac{k}{n}$.

13- $P(X_i{=}1)=\displaystyle\sum_{k=0}^n P(X_i{=}1|Y{=}k)P(Y{=}k)=\sum_{k=0}^n\frac{k}{n}\frac{1}{n+1}=\frac{1}{n(n+1)}\frac{n(n+1)}{2}=\frac{1}{2}$.

14- $E(X_i)=0P(X_i{=}0)+1P(X_i{=}1)=0+\dfrac{1}{2}=\dfrac{1}{2}$ (loi de Bernoulli).

$E(X_i^2)=0^2P(X_i{=}0)+1^2P(X_i{=}1)=\dfrac{1}{2}$.

$V(X_i)=E(X_i^2)-E(X_i)^2=\dfrac{1}{2}-(\dfrac{1}{2})^2=\dfrac{1}{4}$.

21- Pour $F\subset\Omega$: $S(F)=\displaystyle\sum_{i=1}^n X_i(F)\alpha_i$.

22- $E(S)=\displaystyle\sum_{i=1}^n E(X_i)\alpha_i=\frac{1}{2}\sum_{i=1}^n\alpha_i$.

31- On a $P((X_i{=}1)\cap(X_j{=}1)|Y{=}0)=0=P((X_i{=}1)\cap(X_j{=}1)|Y{=}1)=0$ (car on achète strictement moins de 2 articles).

Soit $k\in[2,n]_{\mathbb{N}}$ et $i\neq j$. On a $P((X_i{=}1)\cap(X_j{=}1)|Y{=}k)=P(\{F:a_i,a_j\in F\})=\dfrac{C_{n-2}^{k-2}}{C_n^k}=\dfrac{(n-2)!k!(n-k)!}{(k-2)!(n-k)!n!}=\dfrac{k(k-1)}{n(n-1)}$.

32- $P((X_i{=}1)\cap(X_j{=}1))=0+\displaystyle\sum_{k=2}^n P((X_i{=}1)\cap(X_j{=}1)|Y{=}k)P(Y{=}k)=\frac{1}{n(n-1)(n+1)}\sum_{k=2}^n k(k-1)=\frac{1}{n(n-1)(n+1)}\sum_{k=1}^n k(k-1)=\frac{1}{n(n-1)(n+1)}(\frac{n(n+1)(2n+1)}{6}-\frac{n(n+1)}{2})=\frac{1}{n-1}\frac{2n+1-3}{6}=\frac{1}{3}$.

Comme $\frac{1}{3}\neq\frac{1}{2}\frac{1}{2}$, les v.a. ne sont pas indépendantes.

33- $E(X_iX_j)=0+1P((X_i{=}1)\cap(X_j{=}1))=\frac{1}{3}$.

D'où $\mathrm{cov}(X_i,X_j)=E(X_iX_j)-E(X_i)E(X_j)=\frac{1}{3}-\frac{1}{4}=\frac{1}{12}$ (et les v.a. ne sont effectivement pas indépendantes).

D'où $\rho(X_i,X_j)=\dfrac{\mathrm{cov}(X_i,X_j)}{\sigma(X_i)\sigma(X_j)}=\dfrac{\frac{1}{12}}{\frac{1}{4}}=\dfrac{1}{3}$.

34- $V(S)=V(\displaystyle\sum_{i=1}^n X_i(\omega)\alpha_i)=\sum_{i=1}^n\alpha_i^2V(X_i)+\sum_{1\leq i<j\leq n}\alpha_i\alpha_j\mathrm{cov}(X_i,X_j)=\frac{1}{4}\sum_{i=1}^n\alpha_i^2+\frac{1}{12}2\sum_{1\leq i<j\leq n}\alpha_i\alpha_j$.

Comme $(\displaystyle\sum_{i=1}^n\alpha_i)^2=\sum_{i=1}^n\alpha_i^2+2\sum_{1\leq i<j\leq n}\alpha_i\alpha_j$, on a $V(S)=(\dfrac{1}{4}-\dfrac{1}{12})\displaystyle\sum_{i=1}^n\alpha_i^2+\frac{1}{12}(\sum_{i=1}^n\alpha_i)^2$, avec $\frac{1}{4}-\frac{1}{12}=\frac{1}{6}$. ▪

10.5 Moments d'ordre p

10.5.1 Définition des moments d'ordre p

On généralise les définitions précédentes :

Définition 10.44 Si $p \in \mathbb{N}$, le moment d'ordre P de la v.a.r. X est :

$$E(X^p) \stackrel{\text{déf}}{=} P(X^p) \qquad (= \int_\Omega X^p(\omega)\,dP(\omega)). \tag{10.51}$$

On a donc également (proposition de la mesure image) :

$$E(X^p) = P_X(x^p) \qquad (= \int_\mathbb{R} x^p\,dP_X(x)). \tag{10.52}$$

10.5.2 Inégalité de Jensen

Définition 10.45 Une fonction $\varphi : \mathbb{R} \to \mathbb{R}$ est convexe ssi, pour tout $t \in [0,1]$ et tous $x_1, x_2 \in \mathbb{R}$:

$$\varphi(tx_1 + (1-t)x_2) \le t\varphi(x_1) + (1-t)\varphi(x_2). \tag{10.53}$$

Autrement dit, entre les points $(x_1, \varphi(x_1))$ et $(x_2, \varphi(x_2))$, le graphe de la fonction φ est sous le segment de droite joignant ces points. Faire un dessin.

Lemme 10.46 *Si $f : \mathbb{R} \to \mathbb{R}$ est affine, alors :*

$$E(f \circ X) = f(E(X)). \tag{10.54}$$

Preuve. f est de la forme $f(x) = ax + b$. D'où :

$$\int_\Omega f(X(\omega))\,dP(\omega) = \int_\Omega (aX(\omega)+b)\,dP(\omega) = a\int_\Omega X(\omega)\,dP(\omega) + b = aE(X) + b = f(E(X)).$$

▪

Proposition 10.47 *Si $\varphi : \mathbb{R} \to \mathbb{R}$ est convexe, si X et $\varphi \circ X =^{noté} \varphi(X)$ a une espérance, alors $\varphi(E(X)) \le E(\varphi(X))$:*

$$\varphi \text{ convexe} \quad \Longrightarrow \quad \varphi(E(X)) \le E(\varphi(X)). \tag{10.55}$$

En particulier, $\varphi : x \to |x|^p$ étant convexe pour $p \ge 1$:

$$\forall p \in \mathbb{N}^*, \quad |E(X)|^p \le E(|X|^p) \tag{10.56}$$

Preuve. Soit Λ_φ l'ensemble des fonctions affines inférieures à φ. Donc pour tout $x \in \mathbb{R}$ on a $\varphi(x) \ge f(x)$ pour tout $f \in \Lambda_\varphi$, et donc, pour tout $f \in \Lambda_\varphi$:

$$\int_\Omega \varphi(X(\omega))\,dP(\omega) \ge \int_\Omega f(X(\omega))\,dP(\omega) = f(E(X)),$$

cf. (10.54). D'où $E(\varphi \circ X) \ge \sup_{f\in\Lambda_\varphi}(f(E(X))) = \varphi(E(X))$. ▪

10.6 Intervalle de confiance

10.6.1 Définition et exemple d'intervalle de confiance

But : estimer la "précision" par rapport à la moyenne $\overline{X}$. On se servira de l'écart type qui donne une idée de la "dispersion" par rapport à la valeur moyenne.

Soit $\gamma \in]0,100]$, et $\gamma\% = \frac{\gamma}{100}$ un pourcentage. Soit un réel Δ_e qui vérifie :

$$P(X \in [\overline{X}-\Delta_e, \overline{X}+\Delta_e]) \ge \frac{\gamma}{100} \qquad (= \gamma\%). \tag{10.57}$$

Définition 10.48 Le plus petit des réels Δ_e est l'écart à la moyenne à $\gamma\%$.
Et l'intervalle $[\overline{X}-\Delta_e, \overline{X}+\Delta_e]$ est alors l'intervalle de confiance à $\gamma\%$.

(Si $\Delta_e = 0$ et $P_X(\overline{X}) \ge \gamma\% > 0$, alors P_X ne peut pas être une mesure diffuse.)

Exemple 10.49 Si $\gamma = 95$, alors l'intervalle $[\overline{X}-\Delta_e, \overline{X}+\Delta_e]$ est un intervalle de confiance à 95% : la probabilité pour que l'éventualité ω ait sa valeur $X(\omega)$ dans l'intervalle $[\overline{X}-\Delta_e, \overline{X}+\Delta_e]$ est de 95%. ▪

Exemple 10.50 Cas P_X la loi normale, loi de densité $p(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$ de moyenne m et écart-type σ. On veut calculer un réel Δ_e tel que :

$$\int_{x=m-\Delta_e}^{m+\Delta_e} \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx \geq 95\%,$$

i.e., avec $y = \frac{(x-m)}{\sigma}$ tel que

$$\sigma\int_{y=-\frac{\Delta_e}{\sigma}}^{\frac{\Delta_e}{\sigma}} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\,dy \geq 95\%,$$

i.e tel que, la fonction $e^{-\frac{y^2}{2}}$ étant paire, le reste vérifie :

$$\int_{y=\frac{\Delta_e}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\,dy \leq \frac{1}{2\sigma}5\% = \frac{1}{\sigma}\,2.5\%.$$

La valeur minimale de Δ_e est donnée dans les tables de lois normales, tables qui donnent les valeurs de la fonction $z \to \int_{-\infty}^{z} e^{-\frac{y^2}{2}}\,dy$.

Par exemple, les tables donnent $\int_{-\infty}^{1.96} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\,dy \simeq 0.975$, et donc $\int_{y=1.96}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\,dy \simeq 1 - 0.975 = 0.025 = 2.5\%$; l'intervalle de confiance de 5% pour la loi normale centrée réduite est $\simeq [-1.96, 1.96]$ (intervalle dans lequel on trouve ne moyenne 95 tirs sur 100 quand la loi de probabilité du tir est la loi normale centrée réduite). ▪

10.6.2 Formule de Bienaymé–Tchebytcheff et intervalle de confiance

Tchebytcheff s'écrit aussi Chebichev, Chebyshev... (traduction du cyrillique).

En fonction de σ il s'agit de connaître la probabilité d'être à plus d'une certaine distance de l'espérance (= le centre de gravité).

Lemme 10.51 *(Inégalité de Markov). Si X est une v.a.r. (fonction mesurable) qui admet une espérance alors, pour tout $a > 0$:*

$$P(X{\geq}a) \;\leq\; \frac{P(X)}{a} \qquad (= \frac{E(X)}{a}). \tag{10.58}$$

Plus généralement, si f est une fonction mesurable croissante et $f \circ X$ admet une espérance, alors pour tout $a \in \mathbb{R}$:

$$P(f \circ X) \;\geq\; f(a)\,P(X{\geq}a), \tag{10.59}$$

ce résultat étant trivial pour les a t.q. $f(a) \leq 0$. En particulier si $f(a) > 0$:

$$P(X{\geq}a) \;\leq\; \frac{E(f(X))}{f(a)}. \tag{10.60}$$

Preuve. Pour f croissante, on a $f(x) \geq f(a)$ pour tout $x \geq a$, et donc $f(X(\omega)) \geq f(a)$ pour tout ω t.q. $X(\omega) \geq a$, i.e. pour tout ω t.q. $\omega \in (X \geq a) =^{\text{déf}} X^{-1}([a,\infty[)$, d'où :

$$P(f \circ X) = \int_{\omega\in\Omega} f(X(\omega))\,dP(\omega) \geq \int_{\omega\in(X\geq a)} f(a)\,dP(\omega) = f(a)\,P(X \geq a),$$

d'où (10.59), d'où (10.58) avec $f = id$. ▪

Corollaire 10.52 *(Formule de Bienaymé–Tchebycheff.) Soit $X \in L^2(\Omega, P)$ une v.a.r. qui admet donc une espérance $\overline{X} = E(X)$ et une variance $\sigma(X)^2 = E((X - \overline{X})^2)$. On a, pour tout $\tau > 0$:*

$$P(|X| \geq \tau) \;\leq\; \frac{E(X^2)}{\tau^2}, \tag{10.61}$$

et ainsi il est peu probable (probabilité inférieure à 1%) qu'une v.a.r. X soit en valeur absolue 10 fois supérieure à sa norme $\sqrt{E(X^2)}$ (prendre $\tau = 10\sqrt{E(X^2)}$).

Quitte à considérer la v.a.r. $X-\overline{X}$, on a également :

$$P(|X-\overline{X}|\geq\tau)\ \leq\ \frac{\sigma(X)^2}{\tau^2},\tag{10.62}$$

et ainsi il est peu probable (probabilité inférieure à 1%) qu'une v.a.r. X soit éloignée de sa moyenne de plus de plus de 10 fois son écart type $\sigma(X)$.

(N.B. : quand $\tau\leq\sigma(X)$ c'est évident puisqu'une probabilité est toujours inférieure à 1. Dans ce cas la formule ne donne aucune information.)

Preuve. C'est (10.60) avec $I=[0,\infty[$ et $f(x)=x^2$. ∎

Corollaire 10.53

$$P(|X-\overline{X}|<\tau\sigma(X))\ \geq\ 1-\frac{1}{\tau^2}.\tag{10.63}$$

Preuve. on a $P(|X-\overline{X}|<\tau\sigma(X))+P(|X-\overline{X}|\geq\tau\sigma(X))=1$, d'où (10.63) avec (10.62). ∎

(10.63) peut servir à grossièrement établir des intervalles de confiance :

$$P(X\in]\overline{X}-\tau\sigma,\overline{X}+\tau\sigma[)\ \geq\ 1-\frac{1}{\tau^2},\tag{10.64}$$

soit :

$$P_X(]\overline{X}-\tau\sigma,\overline{X}+\tau\sigma[)\ \geq\ 1-\frac{1}{\tau^2}.\tag{10.65}$$

Remarque 10.54 La formule de Bienaymé–Tchebytcheff est très grossière : elle est valable quelle que soit la probabilité, et sa démonstration est basée sur $1_{|X|\geq a}\leq\frac{X^2}{a^2}$ pour $a>0$ (i.e. sur $1\leq\frac{x^2}{a^2}$ quand $x\geq a>0$).

Si on connaît la loi de probabilité de X, on ne se sert pas la formule de Bienaymé–Tchebytcheff : on fait un calcul direct de $P(|X-\overline{X}|\geq\tau\sigma)$. ∎

Exemple 10.55 Avec la loi normale réduite centrée (de densité $p(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$) : la formule de Tchebytcheff et $\frac{1}{\tau^2}=5\%$, i.e. avec $\tau=\sqrt{\frac{100}{5}}=\sqrt{20}\simeq4.7$, donne $P_X([\overline{X}-\tau,\overline{X}+\tau])\geq95\%$.

Ici $\tau\simeq4.7$ donne une estimation très grossière de Δ_e, voir exemple 10.50) où $\tau\simeq1.96$ convient aussi. ∎

10.7 Fonction génératrice d'une v.a.r. discrète

Définition 10.56 Si $(a_n)_{\mathbb{N}}$ est une suite de réels, alors la fonction :

$$G:s\to G(s)=\sum_{n=0}^{\infty}a_ns^n\tag{10.66}$$

est appelée la fonction génératrice de la suite $(a_n)_{\mathbb{N}}$.

La fonction génératrice permet en particulier de transformer une série (somme discrète sur $\mathbb{N}$) en une série entière (fonction C^∞ sur $]-R,R[$ où R est le rayon de convergence).

Exemple 10.57 Si $a_n=1$ pour tout n, alors $G(s)=\sum_{n=0}^\infty s^n$ est la série entière de Neumann définie pour $s\in]-1,1[$ (de rayon de convergence $R=1$) et $G(s)=\frac{1}{1-s}$.

Si $a_n=\frac{1}{n!}$, alors $G(s)=e^s$ (rayon de convergence infini).

Si $a_k=C_n^k$ pour $0\leq k\leq n$ et $a_k=0$ pour $k>n$, alors $G(s)=(1+s)^n$. ∎

En particulier si $X:\Omega\to\mathbb{R}$ est une v.a.r. à valeurs discrètes x_n, i.e. $X(\Omega)$ est une suite $(x_n)_{\mathbb{N}}$ (i.e. $\forall\omega\in\Omega$, $\exists n\in\mathbb{N}$, $X(\omega)=x_n$), notant P_X la loi de X, alors :

$$G(s)=\sum_{n=0}^{\infty}p_ns^n,\quad\text{où}\quad p_n=P(X{=}x_n)=P_X(x_n),\tag{10.67}$$

est la fonction génératrice de la suite $(p_n)_{\mathbb{N}}$. Comme $\sum_{\mathbb{N}}p_n=1$, le rayon de la série entière G est $R\geq1$, et on a $G(1)=\sum_{n=0}^\infty p_n=1$.

Proposition 10.58 *Cas particulier $X(\Omega)=\mathbb{N}$ et $x_n = n$ pour tout $\mathbb{N}$. Soit R le rayon de convergence de la série entière G (fonction génératrice). Si $1\in]-R,R[$, alors $G'(1)$ existe et :*

$$G'(1)=E(X). \tag{10.68}$$

De plus $G''(1)$ existe et :

$$G''(1)=E(X(X-1))=E(X^2)-E(X) \qquad \text{et} \qquad \mathrm{Var}(X)=G''(1)+G'(1)-(G'(1))^2. \tag{10.69}$$

Plus généralement :

$$G^{(k)}(1)=E(X(X-1)...(X-k+1)). \tag{10.70}$$

Preuve. Pour $s\in]-R,R[$ la dérivée k-ième est obtenue par dérivation sous Σ. Donc :

$$G'(s)=\sum_{n=1}^{\infty} np_nt^{n-1}, \qquad G''(s)=\sum_{n=2}^{\infty} n(n-1)p_ns^{n-2}=\sum_{n=2}^{\infty} n^2p_ns^{n-2}-\sum_{n=2}^{\infty} np_ns^{n-2}.$$

Donc si $1\in]-R,R[$ alors $G'(1)=\sum_{n=1}^\infty np_n = E(X)$ (car $x_n=n$ pour tout n), d'où (10.68), et $G''(1)=E(X(X-1))=\sum_{n=2}^\infty n^2p_n-\sum_{n=2}^\infty np_n=\sum_{n=1}^\infty n^2p_n-\sum_{n=1}^\infty np_n$, d'où (10.69). Idem pour (10.70). ∎

Exercice 10.59 4. Loi uniforme sur $[1,n]_{\mathbb{N}}$. Montrer que $G(s)=\frac{1}{n}\frac{s(1-s^n)}{(1-s)}$.
2. Loi binomiale $B(n,p)$. Montrer que $G(s)=(sp+1-p)^n$.
3. Loi géométrique. Montrer que $G(s)=\dfrac{ps}{1-(1-p)s}$.
4- Loi de Poisson. Montrer que $G(s)=e^{\lambda(s-1)}$.
Puis donner les espérances et variances.

Réponse. 1. $G(s)=\sum_{k=1}^n\frac{1}{n}s^k=\frac{s}{n}\sum_{k=0}^{n-1}s^k=\frac{s}{n}\frac{1-s^n}{1-s}=\frac{1}{n}\frac{s-s^{n+1}}{1-s}$.
$G'(s)=\sum_{k=1}^n\frac{1}{n}ks^{k-1}=\frac{1}{n}\sum_{m=0}^{n-1}(m+1)s^m$. D'où $G'(1)=\frac{1}{n}\sum_{k=1}^n k=\frac{1}{n}\frac{n(n+1)}{2}=\frac{n+1}{2}=E(X)$.
$G''(s)=\frac{1}{n}\sum_{m=1}^{n-1}m(m+1)s^{m-1}$. D'où $G''(1)=\frac{1}{n}\sum_{m=1}^{n-1}m(m+1)=\frac{1}{n}(\sum_{m=1}^{n-1}m^2+\sum_{m=1}^{n-1}m)=\frac{1}{n}(\frac{(n-1)n(2n-1)}{6}+\frac{n(n-1)}{2})=\frac{n^2-1}{3}$. D'où $\mathrm{Var}(X)=\frac{n^2-1}{12}$.
2. $G(s)=\sum_{k=0}^n s^kC_n^kp^k(1-p)^{n-k}=((sp)+(1-p))^n=(sp+1-p)^n$.
D'où $G'(s)=np(sp+1-p)^{n-1}$ et $G''(s)=n(n-1)p^2(sp+1-p)^{n-2}$ (pour $n\geq 2$).
D'où $E(X)=G'(1)=np$, $\mathrm{Var}(X)=G''(1)+G'(1)-(G'(1))^2=n(n-1)p^2+np-(np)^2=np(1-p)$.
3. $G(s)=\sum_{k=0}^\infty s^kq^{k-1}p=ps\sum_{k=0}^n(sq)^{k-1}=ps\frac{1}{1-qs}$.
D'où $G'(s)=\frac{p(1-qs)+psq}{(1-qs)^2}=\frac{p}{(1-qs)^2}=p(1-qs)^{-2}$ et $G''(s)=2pq(1-qs)^{-3}$.
D'où $E(X)=G'(1)=\frac{p}{p^2}=\frac{1}{p}$, $\mathrm{Var}(X)=G''(1)+G'(1)-(G'(1))^2=\frac{2pq}{p^3}+\frac{1}{p}-\frac{1}{p^2}=\frac{2p(1-p)+p^2-p}{p^3}=\frac{1-p}{p^2}$.
4. $G(s)=\sum_{k=0}^\infty s^ke^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^\infty\frac{(\lambda s)^k}{k!}=e^{-\lambda}e^{\lambda s}$.
D'où $G'(s)=\lambda e^{\lambda(s-1)}$ et $G''(s)=\lambda^2e^{\lambda(s-1)}$
D'où $E(X)=G'(1)=\lambda$, $\mathrm{Var}(X)=G''(1)+G'(1)-(G'(1))^2=\lambda^2+\lambda-\lambda^2=\lambda$. ∎

Proposition 10.60 *Soit $s>0$. La fonction $s^X=e^{\log(s)X}:\Omega\to\mathbb{R}$ étant définie par $(s^X)(\omega)=s^{X(\omega)}=e^{\log(s)X(\omega)}$, on a :*

$$G(s)=E(s^X) \tag{10.71}$$

Si X et Y sont deux v.a.r. indépendantes de même loi (v.a.r. i.i.d.) à valeurs dans $\mathbb{N}$, alors s^X et s^Y sont indépendantes et, pour $s\neq 0$:

$$G_{X+Y}(s)=G_X(s)G_Y(s). \tag{10.72}$$

Preuve. Sachant $X(\omega)=k\in\mathbb{N}$ (v.a.r. à valeurs entières), on a $E(s^X)=^{\text{déf}}P(s^X)=\sum_{k\in\mathbb{N}}s^kP(X{=}k)=G(s)$.
$P((s^X{=}x)\cap(s^Y{=}y))=P((e^{X\log(s)}{=}x)\cap(e^{Y\log(s)}{=}y))=P((X{=}\frac{\log(x)}{\log(s)})\cap(Y{=}\frac{\log(x)}{\log(s)}))=P(X{=}\frac{\log(x)}{\log(s)})P(Y{=}\frac{\log(x)}{\log(s)})$, car X et Y sont indépendantes, donc $=P(s^X{=}x)P(s^Y{=}y)$, et s^X et s^Y sont indépendantes.
D'où $G_{X+Y}(s)=E(s^{X+Y})=E(s^Xs^Y)=E(s^X)E(s^Y)$ car s^X et s^Y sont indépendantes. ∎

11 Vecteur aléatoire, suite aléatoire, fonction aléatoire

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé.
On reprend des notions déjà vues en partie, avec un vocabulaire différent.

11.1 Vecteur aléatoire dans $\mathbb{R}^n$, loi conjointe

Soit $\mathbb{R}^n = \mathbb{R} \times ... \times \mathbb{R}$ muni de sa base canonique $(\vec{E}_1, ..., \vec{E}_n)$. On note $\pi_i : \mathbb{R}^n \to \mathbb{R}$ la projection sur la i-ème composante, cf. (2.34) :

$$\vec{x} = \sum_i x_i \vec{E}_i \quad \Longrightarrow \quad \pi_i(\vec{x}) = x_i. \tag{11.1}$$

Soit $\vec{X} : \Omega \to \mathbb{R}^n$ une fonction à valeurs vectorielles. :

$$\vec{X} = (X_1, ..., X_n) \overset{\text{noté}}{=} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} : \begin{cases} \Omega & \to \mathbb{R}^n = \mathbb{R} \times ... \times \mathbb{R} \\ \omega & \mapsto \vec{X}(\omega) = (X_1(\omega), ..., X_n(\omega)) \overset{\text{noté}}{=} \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_2(\omega) \end{pmatrix}, \end{cases} \tag{11.2}$$

la notation en "vecteur colonne" par utilisation implicite de la base canonique de $\mathbb{R}^n$.

Définition 11.1 Lorsque les X_i sont des v.a.r. (fonctions P-mesurables), $\vec{X} = (X_1, ..., X_n)$ est appelé un vecteur aléatoire dans $\mathbb{R}^n$ (un couple dans $\mathbb{R}^2$) de variables aléatoires réelles.

Définition 11.2 Les composantes $X_i = \pi_i \circ \vec{X}$ de $\vec{X}$ sont appelées les v.a.r. marginales de $\vec{X}$.

Notons $\mathcal{T}_{\mathbb{R}^n}$ la tribu borélienne de $\mathbb{R}^2$.

Définition 11.3 La fonction $P_{\vec{X}} : \mathcal{T}_{\mathbb{R}^n} \to [0, 1]$, appelée loi de probabilité de $\vec{X}$, définie par :

$$P_{\vec{X}} \overset{\text{déf}}{=} P \circ \vec{X}^{-1}, \tag{11.3}$$

est également appelée la loi conjointe des X_i. Donc, pour tout ensemble $D \in \mathcal{T}_{\mathbb{R}^n}$:

$$P_{\vec{X}}(D) \overset{\text{déf}}{=} P(\vec{X} \in D), \tag{11.4}$$

soit $P_{\vec{X}}(D) = P(\vec{X}^{-1}(D)) = P(\{\omega \in \Omega : (X_1(\omega), ..., X_n(\omega)) \in D\})$.

Autrement dit, c'est la loi définie par, pour tous les intervalles $I_1, ..., I_n$ de $\mathbb{R}$:

$$\begin{aligned} P_{\vec{X}}(I_1 \times ... \times I_n) &= P(\vec{X} \in I_1 \times ... \times I_n) = P(\vec{X}^{-1}(I_1 \times ... \times I_n)) \\ &= P(\{\omega \in \Omega : \vec{X}(\omega) \in I_1 \times I_2 ...\} \\ &= P(\{\omega \in \Omega : X_1(\omega) \in I_1,\ X_2(\omega) \in I_2,\ ...\} \\ &= P(\{\omega \in \Omega : \omega \in X_1^{-1}(I_1) \cap X_2^{-1}(I_2) \cap ...\} \\ &= P(X_1 \in I_1, X_2 \in I_2, ...) = P_{(X_1, X_2, ...)}(I_1 \times I_2 ...) \\ &\overset{\text{noté}}{=} P_{(X_1, X_2, ...)}(I_1, I_2, ...), \end{aligned} \tag{11.5}$$

avec les notations déjà vu, cf. (5.5) et (5.6).

11.2 Lois marginales

Les lois marginales de mesures de probabilités $P : \mathcal{T}_{\mathbb{R}^n} \to [0, 1]$ (où il n'est pas question de v.a.r.) ont été définies à la définition 2.81 page 33 : ainsi :

$$P_1(]a, b[) = P(]a, b] \times \mathbb{R} \times ...) \qquad (= (P \circ \pi_1^{-1})(]a, b])). \tag{11.6}$$

Soit $\vec{X} = (X_1, ..., X_n)$ un vecteur aléatoire sur un espace probabilisé $(\Omega, \mathcal{T}, P)$, et soit $P_{\vec{X}} = P_{(X_1, ..., X_n)}$ sa loi, cf. (11.4).

Définition 11.4 Les lois $P_{X_i} = P \circ X_i^{-1}$ des v.a.r. marginales $X_i : \Omega \to \mathbb{R}$ de $\vec{X}$ sont appelées les lois marginales de $\vec{X}$.

Ayant $(X_1 \in I) = (X_1 \in I_1) \cap (X_2 \in \mathbb{R}) \cap (X_3 \in \mathbb{R}) \cap ...$:

Proposition 11.5 *Les relations entre les lois marginales P_{X_i} et $P_{\vec{X}}$ sont données par, pour $i = 1, ..., n$:*

$$P_{X_i} = P \circ X_i^{-1}, \tag{11.7}$$

soit, pour tout intervalle I de $\mathbb{R}$, pour $i = 1, 2$:

$$P_{X_i}(I) = P(X_i^{-1}(I)) \qquad (= P(\{\omega \in \Omega : X_i(\omega) \in I\})), \tag{11.8}$$

Donc, pour tout intervalle I de $\mathbb{R}$, par exemple :

$$P_{X_1}(I) = P_{\vec{X}}(I \times \mathbb{R}^{n-1}) \qquad \text{soit} \qquad P(X_1 \in I) = P(\vec{X} \in I \times \mathbb{R}^{n-1}). \tag{11.9}$$

Preuve. $P_{\vec{X}}(I \times \mathbb{R}^{n-1}) = P(\vec{X} \in I \times \mathbb{R}^{n-1}) = P(X_1 \in I, X_2 \in \mathbb{R}, X_3 \in \mathbb{R}, ...) = P(X_1 \in I) = P_{X_1}(I)$. Idem pour X_i pour $i = 2, ..., n$. ▟

11.3 Exemple : couple de v.a.r. (ou vecteur aléatoire dans $\mathbb{R}^2$)

11.3.1 Cas discret : matrice de la loi

Dans $\mathbb{R}^n = \mathbb{R}^2$, $\vec{X} = (X_1, X_2)$. Soit $\Omega = \{x_i,\ i \in I\}$ où I est ensemble fini ou dénombrable. On note :

$$p_{ij} = P_{\vec{X}}(x_i, y_j) \qquad \text{où donc} \qquad \sum_{i,j \in I} p_{ij} = 1, \tag{11.10}$$

soit (loi discrète) :

$$P_{\vec{X}} = \sum_{i,j \in I} p_{ij}\, \delta_{(x_i, y_j)}. \tag{11.11}$$

Définition 11.6 La matrice $[p_{ij}]_{\substack{i \in I \\ j \in I}}$ est appelée matrice de la loi $P_{\vec{X}}$.

(Ce n'est pas une matrice stochastique : à i fixé (une ligne) on n'a pas $\sum_j p_{ij} = 1$.)

On donne souvent p_{ij} sous forme matriciel (tableau), pour $\vec{X} = (X, Y)$:

$X \backslash Y$	y_1	y_2	$\cdots$
x_1	p_{11}	p_{12}	
x_2	p_{21}	p_{22}	
x_3	p_{31}	p_{32}	
$\vdots$			$\ddots$

où I a été redéfini en $\{1, ..., n, ...\}$.

Exemple 11.7 Voir exercice 10.30. ▟

Définition 11.8 Notations. Pour tout $i, j \in I$, on note $p_{i\cdot}$ et $p_{\cdot j}$ les réels définis par :

$$p_{i\cdot} \overset{\text{déf}}{=} P_{\vec{X}}(\{i\} \times \mathbb{R}) = P_{X_1}(i) \qquad \text{et} \qquad p_{\cdot j} \overset{\text{déf}}{=} P_{\vec{X}}(\mathbb{R} \times \{j\}) = P_{X_2}(j). \tag{11.12}$$

On a immédiatement :

$$p_{i\cdot} = \sum_{j \in I} p_{ij} \qquad \text{et} \qquad p_{\cdot j} = \sum_{i \in I} p_{ij}, \tag{11.13}$$

soit relativement à la matrice $[p_{ij}]$:

$$p_{i\cdot} = \text{somme sur la ligne } i \qquad \text{et} \qquad p_{\cdot j} = \text{somme sur la colonne } j. \tag{11.14}$$

Et on a trivialement :

$$\sum_{i \in I} p_{i\cdot} = \sum_{j \in I} p_{\cdot j} = 1 \qquad (= \sum_{i,j \in I} p_{ij}). \tag{11.15}$$

Exemple 11.9 On lance deux pièces. Soit $\vec{X} : \Omega \to \{a,b\} \times \{c,d\}$, avec donc $\vec{X}(\omega) = (X_1(\omega), X_2(\omega)) \in \{a,b\} \times \{c,d\}$, avec donc $X_1 : \Omega \to \{a,b\}$ et $X_2 : \Omega \to \{c,d\}$. Pour des lancers indépendants et un tirage équiprobable, la loi de $P_{\vec{X}}$ est donnée par le tableau :

$X_1 \backslash X_2$	c	d
a	$\frac{1}{4}$	$\frac{1}{4}$
b	$\frac{1}{4}$	$\frac{1}{4}$

Donc $P_{\vec{x}} = \frac{1}{4}\delta_{(a,c)} + \frac{1}{4}\delta_{(a,d)} + \frac{1}{4}\delta_{(b,c)} + \frac{1}{4}\delta_{(b,d)}$, ou encore, pour tout $(x,y) \in \{a,b\} \times \{c,d\}$:

$$P_{\vec{X}}(x,y) = P(X_1{=}x\,,\; X_2{=}y) = \frac{1}{4}. \tag{11.16}$$

Et $p_{i\cdot} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} = p_{\cdot j}$.

Exemple 11.10 Soit $\vec{Y} : \Omega \to \{a,b\} \times \{c,d\}$. On définit sa loi à l'aide du tableau :

$Y_1 \backslash Y_2$	c	d
a	0	α
b	β	0

où $\alpha, \beta \geq 0$ et $\alpha + \beta = 1$. Donc $P_{\vec{Y}} = \alpha\delta_{(a,d)} + \beta\delta_{(b,c)}$, soit :

$$P_{\vec{Y}}(a,c) = 0 = P_{\vec{Y}}(b,d), \qquad P_{\vec{Y}}(a,d) = \alpha, \qquad P_{\vec{Y}}(b,c) = \beta. \tag{11.17}$$

Et $p_{1\cdot} = \alpha = p_{\cdot 2}$, et $p_{2\cdot} = \beta = p_{\cdot 1}$.

11.3.2 Cas continu

La loi de $\vec{X}$ est continue quand elle est de la forme, pour $D \in \mathcal{T}_{\mathbb{R}}^2$:

$$P_{\vec{X}}(D) = \iint_D p(x,y)\,dxdy,$$

où $p : \mathbb{R}^2 \to \mathbb{R}$ est une fonction positive dans $\mathbb{R}^2$ telle que $\displaystyle\iint_{\mathbb{R}^2} p(x,y)\,dxdy = 1$.

Exemple 11.11 Soit A un borélien et soit $p(x,y) = ce^{xy}1_A(x,y)$ où $c \in \mathbb{R}$ est une constante telle que $\iint_A p(x,y)\,dxdy = 1$.

On notera :

$$p_1(x) = p_{X_1}(x) = \int_{y\in\mathbb{R}} p(x,y)\,dy, \qquad p_2(y) = p_{X_2}(y) = \int_{x\in\mathbb{R}} p(x,y)\,dx. \tag{11.18}$$

11.3.3 Fonction de répartition

Définition 11.12 On appelle fonction de répartition F du couple (X,Y), ou fonction de répartition conjointe, la fonction $F : \mathbb{R}^2 \to [0,1]$ donnée par :

$$\begin{aligned} F(x,y) &= P((X \in]-\infty,x]) \cap (Y \in]-\infty,y])) = P(X \leq x \text{ et } Y \leq y) \\ &= P_{\vec{X}}(]-\infty,x]\times]-\infty,y]). \end{aligned} \tag{11.19}$$

11.4 Remarque

Proposition 11.13 *1- Si on connaît la loi $P_{\vec{X}}$ on en déduit les lois marginales.*

2- La réciproque est fausse : on peut avoir des lois marginales identiques qui appartiennent à des vecteurs aléatoires différents.

3- N.B. : on peut avoir $P_{X_1}, ..., P_{X_n}$ des lois densités, alors que $P_{\vec{X}}$ n'est pas une loi de densité.

Preuve. 1- Si $P_{\vec{X}}$ existe, alors les lois marginales sont données par (11.8).

2- La réciproque est fausse : reprenons les exemples 11.9 et 11.10 avec $\alpha = \beta = \frac{1}{2}$: on a $P_{\vec{X}} \neq P_{\vec{Y}}$ (par exemple $P_{\vec{X}}(a,c) = \frac{1}{4}$ et $P_{\vec{Y}}(a,c) = 0$).

Et $P_{X_1} = P_{Y_1}$ et $P_{X_2} = P_{Y_2}$ sont les lois équiprobables. En effet, on a $P_{X_1}(a) = P_{\vec{X}}(\{a\},\mathbb{R}) = P_{\vec{X}}(a,c) + P_{\vec{X}}(a,d) = \frac{1}{2}$ (somme sur la ligne a), et $P_{X_1}(b) = \frac{1}{2}$ (somme sur la ligne b). De même $P_{X_2}(c) = \frac{1}{2} = P_{X_2}(d)$ (somme sur la colonne adéquat).

Et $P_{Y_1}(a) = \frac{1}{2} = P_{Y_1}(b)$, et $P_{Y_2}(c) = \frac{1}{2} = P_{Y_2}(d)$.

Donc les P_{X_i} et P_{Y_i} sont les lois équiprobables, mais $P_{\vec{X}} \neq P_{\vec{Y}}$.

3- Soit $\Delta = \{(x,x) : x \in \mathbb{R}\}$ la diagonale de $\mathbb{R}^2$. Soit $P : \mathbb{R}^2 \to \mathbb{R}_+$ la probabilité de densité linéique notée $\delta_\Delta p(x,y)dxdy$ avec δ_Δ la mesure de Dirac le long de Δ : si $D \subset \mathbb{R}^2$, $P(D) = \int_{x\in\mathbb{R}:(x,x)\in D} p(x,x)\,dx$. P n'est pas une loi de densité ("volumique") dans $\mathbb{R}^2$ (c'est une "densité linéique").

Soit $\vec{X}$ la fonction identité, $\vec{X}(x,y) = (x,y)$, de composantes X_1 et X_2 données par $X_1(x,y) = x$ et $X_2(x,y) = y$. On a donc $P_{\vec{X}}([a,b]\times[c,d]) = P(\vec{X} \in [a,b]\times[c,d]) = P([a,b]\times[c,d])$.

Et on a $P_{X_1}([a,b]) = (P \circ X_1^{-1})([a,b]) = P(X_1 \in [a,b]\}) = P(\vec{X} \in [a,b]\times\mathbb{R}) = \int_a^b p(x,x)\,dx$, et donc P_{X_1} est la loi de densité $p_1(x) = p(x,x)$. De même pour X_2. ▟

11.5 Lois marginales indépendantes

11.5.1 Définition

Définition 11.14 Le vecteur aléatoire $\vec{X} = (X_1, ..., X_n)$ définit des v.a.r. indépendantes ssi :

$$P_{\vec{X}} = P_{X_1} \otimes ... \otimes P_{X_n}, \tag{11.20}$$

i.e. ssi pour tous les intervalles I_1, ..., I_n de $\mathbb{R}$:

$$P_{\vec{X}}(I_1, ..., I_n) = P_{X_1}(I_1)...P_{X_n}(I_n). \tag{11.21}$$

Proposition 11.15 *Dans le cas d'indépendance, les lois marginales déterminent de manière unique la loi de la variable vectorielle.*

Preuve. La loi $P_{\vec{X}}$ est définie de manière unique par (11.20). ▟

Exemple 11.16 Cas discret et indépendance : $P_{\vec{X}} = \sum_{ij} p_{ij}\delta_{x_i,y_j}$ avec nécessairement $p_{ij} = P_{X_1}(i)P_{X_2}(j)$ (car indépendance). ▟

Exemple 11.17 Retour sur la proposition 11.13 et le contre-exemple de la démonstration : $\vec{X}$ définit des v.a.r. indépendantes, mais pas $\vec{Y}$.

En effet, on doit avoir $(P_{Z_1} \otimes P_{Z_2})(z_1,z_2) = P_{Z_1}(z_1)P_{Z_2}(z_2) = \frac{1}{4}$ pour tout z_1, z_2, vérifié pour $\vec{Z} = \vec{X}$, et faux quand $\vec{Z} = \vec{Y}$. ▟

Exemple 11.18 Les lois X_1, X_2 relatives à l'échantillon dans une population, avec (6.16), définissent un vecteur aléatoire $\vec{X} = (X_1, X_2)$ où lois marginales ne sont pas indépendantes puisque $X_1{+}X_2{=}n$ (lien entre X_1 et X_2). Voir les calculs plus loin exemple 11.21. ▟

11.5.2 Cas discret et exemples

Proposition 11.19 *Dans le cas discret* $P_{\vec{X}} = \sum_{i,j\in I} p_{ij}\delta_{(x_i,y_j)}$, *cf. (11.11), on a :*

$$\begin{cases} P_{X_1} = \displaystyle\sum_{i\in I} p_{i\cdot}\,\delta_{x_i}, & p_{i\cdot} = \displaystyle\sum_j p_{ij}, \\ P_{X_2} = \displaystyle\sum_{j\in I} p_{\cdot j}\,\delta_{y_j}, & p_{\cdot j} = \displaystyle\sum_i p_{ij}. \end{cases} \tag{11.22}$$

Et les variables aléatoires X_1 *et* X_2 *sont indépendantes ssi* $P_{\vec{X}}(x_i,y_j) = P_{X_1}(x_i)P_{X_2}(y_j)$:

$$X_1 \text{ et } X_2 \text{ indépendantes} \iff p_{ij} = p_{i\cdot}p_{\cdot j} \quad \forall i,j. \tag{11.23}$$

Et on a alors :

$$\begin{cases} E(X_1) = \sum_i x_i p_{i\cdot} & (= P(X_1) = \int X_1\, dP = \int x\, dP_{X_1}(x)), \\ E(X_2) = \sum_j y_j p_{\cdot j} & (= P(X_2) = \int X_2\, dP = \int y\, dP_{X_2}(y)). \end{cases} \tag{11.24}$$

Preuve. (11.9) donne :

$$\begin{cases} P_{X_1}(x_i) = P(X_1{=}x_i) = P(X_1{=}x_i, X_2 \in \mathbb{R}) = p_{i\cdot}, \\ P_{X_2}(y_j) = P(X_2{=}y_j) = P(X_1{\in}\mathbb{R}, X_2{=}y_j) = p_{\cdot j}, \end{cases} \tag{11.25}$$

i.e. (11.22). D'où (11.24). ▪

Exemple 11.20 On fait deux lancers d'une pièce équilibrée. L'espace des résultats est $\Omega = \{a,b\}^2 = \{(a,a),(a,b),(b,a),(b,b)\}$ et $P(\vec\omega) = \frac14$ pour tout $\vec\omega = (\omega_1,\omega_2) \in \Omega$.

Soit $X_1 : \Omega \to \mathbb{R}$ v.a.r. relative au premier lancer, qui donne la valeur α si on a obtenu a et la valeur β si on a obtenu b. Donc ici $X_1(a,a) = X_1(a,b) = \alpha$ et $X_1(b,a) = X_1(b,b) = \beta$.

Soit $X_2 : \Omega \to \mathbb{R}$ v.a.r. relative au second lancer, qui donne la valeur γ si on a obtenu a et la valeur δ si on a obtenu b. Donc ici $X_2(a,a) = X_2(b,a) = \gamma$ et $X_2(a,b) = X_2(b,b) = \delta$.

La pièce étant équilibrée, on a $P_{X_1}(\alpha) = P_{X_1}(\beta) = \frac12 = P_{X_2}(\gamma) = P_{X_2}(\delta)$.

Soit $\vec X = (X_1,X_2) : \vec X(a,a) = \begin{pmatrix}\alpha\\ \gamma\end{pmatrix}$, $\vec X(a,b) = \begin{pmatrix}\alpha\\ \delta\end{pmatrix}$, $\vec X(b,a) = \begin{pmatrix}\beta\\ \gamma\end{pmatrix}$ et $\vec X(b,b) = \begin{pmatrix}\beta\\ \delta\end{pmatrix}$.

Avec les notations précédentes $x_1 = \alpha$, $x_2 = \beta$, $y_1 = \gamma$, $y_2 = \delta$, et $I = \{1,2\}$.

Et $p_{11} = P_{\vec X}(x_1,y_1) = P_{\vec X}(\alpha,\gamma) = P((X_1{=}\alpha) \cap (X_2{=}\gamma)) = P(\{(a,a),(a,b)\} \cap \{(a,a),(b,a)\}) = P(\{a,a\}) = \frac14$.

Et $p_{21} = P_{\vec X}(x_2,y_1) = P(\{(b,a),(b,b)\} \cap \{(a,a),(b,a)\}) = P(b,a) = \frac14$, et de même $p_{12} = p_{22} = \frac14$. La matrice stochastique est donc la matrice $[p_{ij}]_{\substack{i=1,2\\ j=1,2}} = \frac14 \begin{pmatrix}1&1\\1&1\end{pmatrix}$.

D'où, en sommant sur les lignes, $p_{1.} = \frac12$ et $p_{2.} = \frac12$ donnent l'équiprobabilité pour le 1er lancer, et en sommant sur les colonnes, $p_{.1} = \frac12$ et $p_{.2} = \frac12$ donnent l'équiprobabilité pour le 2nd lancer.

Ici les v.a.r. X_1 et X_2 sont indépendantes : on vérifie que $p_{ij} = p_{i.}p_{.j}$ $(= \frac14)$.

N.B. : ici les événements sont indépendants, et cela **ne** donne **pas** une matrice diagonale : une matrice stochastique n'est pas une matrice de covariance, voir plus loin. ▪

Exemple 11.21 Voir proposition 6.15 page 63. Soit une population $S = \{a_1,...,a_n\}$ et soit un échantillon de taille r avec replacement. Donc ici $\Omega = S^r$. Et soit $\vec X = \begin{pmatrix}X_1\\X_2\end{pmatrix} : S^r \to \mathbb{N}^2$ le vecteur aléatoire dont la première composante $X_1(\vec\omega)$ donne le nombre de personnes de S_1 présentes dans l'échantillon $\vec\omega$, et la seconde composante $X_2(\vec\omega)$ le nombre personnes de S_2. (On a $X_1(\vec\omega) + X_2(\vec\omega) = r$ et donc X_1 et X_2 sont liées).

Soient $k_1, k_2 \in \mathbb{N}$ deux entiers. Calculons :

$$P_{\vec X}(k_1,k_2) \stackrel{\text{déf}}{=} P(\vec X = (k_1,k_2)) = P(\vec X^{-1}(k_1,k_2)).$$

Si $k_1 + k_2 \neq r$ alors $P_X(k_1,k_2) = 0$ (par hypothèses les échantillons contiennent exactement r individus). Supposons $k_1 + k_2 = r$. On a :

$$\vec X^{-1}(k_1,k_2) \stackrel{\text{déf}}{=} \vec X^{-1}(\{k_1\} \times \{k_2\}) = \{\vec\omega \in S^r : X_1(\vec\omega) = k_1,\ X_2(\vec\omega) = k_1\} = X_1^{-1}(k_1)$$

puisque $k_1 + k_2 = r$ et que l'information sur X_2 est inutile : ici $\{X_1{=}k_1\} \cap \{X_2{=}k_2\} = \{X_1{=}k_1\}$ et donc $P(\{X_1{=}k_1\} \cap \{X_2{=}k_2\}) = P(\{X_1{=}k_1\})$. La loi de X est donc "celle de X_1" au sens, quand $k_2 = r - k_1$:

$$P_{\vec X}(k_1,k_2) = P_{X_1}(k_1) = b(k_1;r,p) = C_r^{k_1} p^{k_1}(1{-}p)^{k_2} = \frac{r!}{k_1!k_2!} p^{k_1} q^{k_2},$$

où $p = \frac{|S_1|}{|S|}$ et $q = 1{-}p$. En particulier, si $P(X_2{=}k_2) \neq 1$, alors $P(\{X_1{=}k_1\} \cap \{X_2{=}k_2\}) \neq P_{X_1}(k_1)P_{X_2}(k_2)$: les lois ne sont pas indépendantes, ce qu'on savait déjà avec la relation "$X_2(\vec\omega) = r - X_1(\vec\omega)$".

(Calcul direct : $P_{X_1}(k_1)P_{X_2}(k_2) = \frac{r!}{k_1!k_2!} p^{k_1}(1{-}p)^{k_2} \frac{r!}{k_2!k_1!} q^{k_2}(1{-}q)^{k_1}$ avec $q = 1{-}p$, et donc $P_{X_1}(k_1)P_{X_2}(k_2) = (\frac{r!}{k_1!k_2!} p^{k_1} q^{k_2})^2 \neq P_X(k_1,k_2) = P_{X_1}(k_1)$ en général.) ▪

11.5.3 Cas continu

Ici $P_{\vec{X}} : \mathcal{T}_{\mathbb{R}^2} \to [0,1]$ (avec $\mathcal{T}_{\mathbb{R}^2}$ la tribu borélienne de $\mathbb{R}^2$) est définie par une densité $p(x,y)$: pour tout D domaine (mesurable) de $\mathbb{R}^2$:

$$P_{\vec{X}}(D) = \iint_D p(x,y)\,dxdy. \tag{11.26}$$

Comme :

$$P_{X_1}(A) = P_{\vec{X}}(A \times \mathbb{R}) = \iint_{A\times\mathbb{R}} p(x,y)\,dxdy = \int_A \Big(\int_{y\in\mathbb{R}} p(x,y)\,dy\Big)\,dx, \tag{11.27}$$

la première loi marginale P_{X_1} est la loi de densité p_1 :

$$P_{X_1}(A) = \int_A p_{X_1}(x)\,dx \qquad \text{où} \qquad p_{X_1}(x) = \int_{y\in\mathbb{R}} p(x,y)\,dy. \tag{11.28}$$

Et la seconde loi marginale est la loi P_{X_2} de densité $p_2(y) = \int_{x\in\mathbb{R}} p(x,y)\,dx$.

Si p est une fonction à variables séparées, i.e. de la forme $p(x,y) = q_1(x)q_2(y)$, alors les lois de X_1 et X_2 sont indépendantes. En effet, dans ce cas on a bien (11.20) :

$$P_{\vec{X}}(A\times B) = \iint_{A\times B} p(x,y)\,dxdy = \int_A q_1(x)\,dx \int_B q_2(y)\,dy = P_{X_1}(A)P_{X_2}(B), \tag{11.29}$$

car $\int_{y\in\mathbb{R}} p(x,y)\,dy = \int_{y\in\mathbb{R}} q_1(x)q_2(y)\,dy = q_1(x)$ est bien la densité de la première loi marginale, cf. (11.28). Idem pour q_2.

11.6 Fonction aléatoire

Dans le cas des fonctions $\Omega \to E$ où E est un espace de fonctions, on parle de fonction aléatoire :

Définition 11.22 Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé et E un espace de fonctions muni d'une topologie et de la tribu engendrée. Une fonction mesurable :

$$X : \begin{cases} \Omega & \to E \\ \omega & \mapsto X(\omega) \end{cases}$$

est appelée une fonction aléatoire.

C'est le cas des fonctions (mesurables) dépendant d'un paramètre :

Exemple 11.23 Soit $X_t : x \in [a,b] \to X_t(x) \in \mathbb{R}$ une fonction dépendant du paramètre $t \in [0,T]$.

Soit $\Omega = [0,T]$ muni de la tribu borélienne, et soit $X : t \in \Omega \to X_t \in E$ où $E = \mathcal{F}_M([a,b];\mathbb{R})$ est l'ensemble des fonctions mesurables. ▟

12 Espérance d'un vecteur aléatoire réel, covariance

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé.

12.1 Espérance

Soit $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ un vecteur aléatoire réel (fonction $\Omega \to \mathbb{R}^n$ mesurable à valeurs vectorielles).

On note $\vec{x} = \vec{X}(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ l'image d'un point $\omega \in \Omega$ par $\vec{X}$.

Définition 12.1 L'espérance de $\vec{X}$ (quand elle existe) est, avec $E(X_i) = P(X_i)$:

$$E(\vec{X}) \stackrel{\text{déf}}{=} \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} \stackrel{\text{noté}}{=} \overline{\vec{X}} = \begin{pmatrix} \overline{X}_1 \\ \vdots \\ \overline{X}_n \end{pmatrix} \stackrel{\text{noté}}{=} P(\vec{X}), \tag{12.1}$$

ou bien avec d'autres notations :

$$E(\vec{X}) = \int_\Omega \vec{X}(\omega)\, dP(\omega) = \begin{pmatrix} \int_\Omega X_1(\omega)\, dP(\omega) \\ \vdots \\ \int_\Omega X_n(\omega)\, dP(\omega) \end{pmatrix}. \tag{12.2}$$

Proposition 12.2 *Soit $\vec{X} = (X_1, ..., X_n)$ un vecteur aléatoire. On a :*

$$E(\vec{X}) = \begin{pmatrix} \int_\mathbb{R} x_1\, dP_{X_1}(x_1) \\ \vdots \\ \int_\mathbb{R} x_n\, dP_{X_n}(x_n) \end{pmatrix}. \tag{12.3}$$

Si de plus les v.a.r. X_i sont indépendantes, alors $P_{\vec{X}} = P_{X_1} \otimes ... \otimes P_{X_n}$ et :

$$E(\vec{X}) = \int_{\mathbb{R}^n} \vec{x}\, dP_{\vec{X}}.$$

C'est alors le centre de gravité $E(\vec{X}) = \int_{\mathbb{R}^n} \vec{x}\, d\mu(\vec{x})$ de la répartition de masse $\mu = P_{\vec{X}}$.

Preuve. Chaque v.a.r. X_i admet une loi de probabilité $P_{X_i} = P \circ X_i^{-1} : \mathcal{T} \to [0,1]$, et on applique la proposition de la mesure image.

Supposons de plus les X_i indépendantes. Alors pour la 1ère composante de $\vec{x}$, ayant $P_{\vec{X}} = P_{X_1} \otimes ... \otimes P_{X_n}$:

$$\begin{aligned} \int_{\mathbb{R}^n} x_1\, dP_{\vec{X}}(x_1) &= \int_{\mathbb{R}^n} x_1\, dP_{X_1}(x_1)...dP_{X_n}(x_n) = \int_\mathbb{R} x_1\, dP_{X_1}(x_1) \int_\mathbb{R} dP_{X_2}(x_2)... \int_\mathbb{R} dP_{X_n}(x_n) \\ &= \int_\mathbb{R} x_1\, dP_{X_1}(x_1), \end{aligned}$$

et avec (12.3) on obtient $E(\vec{X}) = \begin{pmatrix} \int_{\mathbb{R}^n} x_1\, dP_{\vec{X}} \\ \vdots \\ \int_{\mathbb{R}^n} x_n\, dP_{\vec{X}} \end{pmatrix} = \int_{\mathbb{R}^n} \vec{x}\, dP_{\vec{X}}$. ▪

Soit $M = [M_{ij}]_{\substack{i=1,...,m \\ j=1,...,n}}$ une matrice aléatoire réelle sur $(\Omega, \mathcal{T}, P)$ (une fonction $\omega \to M(\omega)$ mesurable dans $\mathbb{R}^{mn}$).

Définition 12.3 L'espérance de la fonction M est :

$$E(M) = [E(M_{ij})]_{\substack{i=1,...,m \\ j=1,...,n}} = \int_\Omega M(\omega)\, dP = [\int_\Omega M_{ij}(\omega)\, dP]_{\substack{i=1,...,m \\ j=1,...,n}}. \tag{12.4}$$

Remarque 12.4 En dimension finie, une matrice $m * n$ peut être considérée comme un vecteur de $\mathbb{R}^{mn}$, l'écriture usuelle consistant à mettre toutes les colonnes de la matrice les unes à la suite des autres pour former un "vecteur colonne". ▪

Remarque 12.5 On peut rendre la définition de $E(\vec{X})$ intrinsèque, i.e. ici indépendante du choix de la base $(\vec{e}_i)$ sur laquelle $\vec{X} = X_1\vec{e}_1 + ... + X_n\vec{e}_n$ est représenté. Pour ce on utilise la démarche usuelle : si F est un espace vectoriel de dimension n (précédemment $F = \mathbb{R}^n$), on considère son dual $F^* = L(F;\mathbb{R})$, ensemble des formes linéaires sur F, et pour une fonction $\vec{X} : \Omega \to F$ donnée et $\ell \in F^*$ on regarde $\ell(\vec{X}) =^{\text{déf}} \ell \circ \vec{X} : \Omega \to \mathbb{R}$, donc donnée par $\ell(\vec{X})(\omega) = \ell(\vec{X}(\omega))$ (précédemment on avait considéré les $\ell = \pi_i$ définis par $\pi_i(\vec{x}) = x_i$). Comme $\ell(\vec{X})$ est une v.a.r., son espérance a un sens déjà défini.

On définit alors le vecteur $E(\vec{X}) \in F$ par $\ell(E(\vec{X})) = E(\ell(\vec{X}))$ pour tout $\ell \in F^*$, car on sait qu'un vecteur $\vec{v}$ est entièrement déterminé par l'ensemble des valeurs $(\ell(\vec{v}))_{\ell \in F^*}$. Il suffit d'ailleurs de se contenter de n formes linéaires de base dans F^*, et la définition précédente (12.1) à été donnée dans le cas $F = \mathbb{R}^n$ muni de sa base canonique et F^* de la base duale (π_i). ▪

Lemme 12.6 *Si $\vec{X}$ est un vecteur aléatoire dans $\mathbb{R}^n$ alors, pour tout $\vec{y} \in \mathbb{R}^n$:*

$$E(\vec{y}^T.\vec{X}) = \vec{y}^T.E(\vec{X}). \tag{12.5}$$

(Permet de se ramener aux v.a.r. dans $\mathbb{R}$.)

Preuve. $\vec{y}^T.\vec{X} = \sum_{=1}^n y_i X_i$ et E linéaire donne $E(\vec{y}^T.\vec{X}) = \sum_{=1}^n y_i E(X_i) = \vec{y}^T.E(\vec{X})$. ▪

12.2 Covariance de deux v.a.r.

12.2.1 Définition

On considère deux v.a.r. X et Y sur un même espace probabilisé $(\Omega, \mathcal{T}, P)$, et on considère le vecteur aléatoire $\vec{X} = (X, Y)$ (le couple de v.a.r.).

Définition 12.7 Au couple (X, Y) on associe le réel :

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) \qquad (= \int XY\, dP - \int X\, dP \int Y\, dP), \tag{12.6}$$

appelé la covariance de X et de Y.

En particulier :

$$\text{cov}(X, X) = \text{Var}(X). \tag{12.7}$$

Pour une probabilité discrète $P = \sum_{i\in\mathbb{N}^*} q_{\omega_i} \delta_{\omega_i}$, on a :

$$\text{cov}(X, Y) = \sum_{i\in\mathbb{N}^*} X(\omega_i)Y(\omega_i)q_{\omega_i} - (\sum_{i\in\mathbb{N}^*} X(\omega_i)q_{\omega_i})(\sum_{i\in\mathbb{N}^*} Y(\omega_i)q_{\omega_i}), \tag{12.8}$$

et $\text{cov}(X, X) = \text{Var}(X) = \sum_{i\in\mathbb{N}^*} X(\omega_i)^2 q_{\omega_i} - (\sum_{i\in\mathbb{N}^*} X(\omega_i)q_{\omega_i})^2$.

Pour une probabilité continue donnée par $dP(\omega) = q(\omega)d\omega$, on a :

$$\text{cov}(X, Y) = \int_\Omega X(\omega)Y(\omega)\, q(\omega)d\omega - \int_\Omega X(\omega)\, q(\omega)d\omega \int_\Omega Y(\omega)\, q(\omega)d\omega), \tag{12.9}$$

et $\text{cov}(X, X) = \text{Var}(X) = \int_\Omega X(\omega)^1\, q(\omega)d\omega - (\int_\Omega X(\omega)\, q(\omega)d\omega)^2$.

Exemple 12.8 Soit un dé équilibré à 3 faces de valeurs 1, 2 et 3, soit X et Y les résultats respectifs de deux lancers successifs. Ici X et Y sont deux v.a.r. indépendantes à valeurs dans $[1, 3]_\mathbb{N}$ de même loi de probabilité $P_X = P_Y = \sum_{i=1}^3 \frac{1}{3}\delta_i$. On a $E(X) = E(Y) = \sum_{i=1}^3 i\,\frac{1}{3} = \frac{6}{3} = 2$.

On a $XY : \omega \in \{1, 2, 3\} \to \{r \in \mathbb{R} : \exists i \in [1, 3]_\mathbb{N} \text{ t.q. } r = X(i)Y(i)\} = \{1, 2, 3, 4, 6, 9\}$, avec la probabilité $q_1 = \frac{1}{9}$, $q_2 = \frac{2}{9}$, $q_3 = \frac{2}{9}$, $q_4 = \frac{1}{9}$, $q_6 = \frac{2}{9}$, $q_9 = \frac{1}{9}$ (faire le produit matriciel $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} . \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$ et compter le nombre de 1, de 2...) D'où $E(XY) = 1\frac{1}{9} + 2\frac{2}{9} + 3\frac{2}{9} + 4\frac{1}{9} + 6\frac{2}{9} + 9\frac{1}{9} = 4$. D'où $\text{cov}(X, Y) = 4 - 2^2 = 0$. On verra que si X et Y sont indépendantes (ce qui est le cas ici), on a nécessairement $\text{cov}(X, Y) = 0$. ▪

Exercice 12.9 Voir § 6.4.1. Population $S = \{a_1, a_2, a_3\}$ de taille 3. Tirages équiprobables avec replacement et échantillons de taille r=2. Sous populations $S_1 = \{a_1\}$ et $S_2 = \{a_2, a_3\}$. Pour un échantillon (résultat de deux tirages avec remise) $\vec{\omega} = (\omega_1, \omega_2) \in \Omega$, soit $X_1(\vec{\omega})$ le nombre de composantes qui sont dans S_1, et soit $X_2(\vec{\omega}) = 2 - X_1(\vec{\omega})$ le nombre de composantes qui sont dans S_2. Calculer $\text{cov}(X_1, X_2)$. (Ici X_1 et X_2 ne sont pas indépendantes : $X_1 + X_2 = 2$.)

Réponse. La probabilité d'obtenir un élément de S_1 lors d'un tirage est $p_1 = \frac{|S_1|}{|S|} = \frac{1}{3}$ (équiprobabilité), et la probabilité d'obtenir un élément de S_2 lors d'un tirage est $p_2 = \frac{|S_2|}{|S|} = \frac{2}{3}$ (équiprobabilité). On a $P_{X_i} = b(|S|, p_i)$, cf. (6.17), donc $E(X_i) = |S|p_i$, cf. (10.22), d'où $E(X_1) = 3\frac{1}{3} = 1$ et $P_{X_2} = 3\frac{2}{3} = 2$.

On a $X_2 = 2 - X_1$, donc $X_1X_2 = 2\,X_1 - X_1$ avec $X_1(\vec{\omega}) \in \{0, 1, 2\}$, donc $(X_1X_2)(\vec{\omega}) \in \{0, 1\}$, et :
$P_{X_1X_2}(0) = P(\{(a_1, a_1), (a_2, a_2), (a_2, a_3), (a_3, a_2), (a_3, a_3)\} = \frac{5}{9}$,
$P_{X_1X_2}(1) = P(\{(a_1, a_2), (a_1, a_3), (a_2, a_1), (a_3, a_1)\} = \frac{4}{9}$.

Donc $E(X_1X_2) = 0\,\frac{5}{9} + 1\,\frac{4}{9} = \frac{4}{9}$.

Donc $\text{cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = \frac{4}{9} - 1 * 2 = -\frac{14}{9}$. ▪

12.2.2 Bilinéarité de la covariance (semi-produit scalaire)

Proposition 12.10 *La covariance* cov : $(X,Y) \to \mathrm{cov}(X,Y)$ *est une forme bilinéaire symétrique positive sur* $L^2(\Omega,P)$ *: c'est un semi-produit scalaire. Et donc* cov *vérifie l'inégalité de Cauchy–Schwarz* $|\mathrm{cov}(X,Y)| \le \sqrt{\mathrm{cov}(X,X)}\sqrt{\mathrm{cov}(Y,Y)}$*, soit :*

$$|\mathrm{cov}(X,Y)| \le \sigma(X)\sigma(Y). \tag{12.10}$$

Et c'est une produit scalaire sur le sous-espace de $L^2(\Omega,P)$ *des v.a.r. de moyenne nulle.*

Preuve. Si $X,Y \in L^2(\Omega,P)$ alors $XY \in L^1(\Omega,P)$ et $X,Y \in L^1(\Omega,P)$ (car P mesure finie implique $L^2(\Omega,P) \subset L^1(\Omega,P)$). Donc $\mathrm{cov}(X,Y)$ est bien défini sur $L^2(\Omega,P)$.

La symétrie $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$ et la bilinéarité sont immédiates avec (12.6).

Et une forme bilinéaire symétrique positive vérifie l'inégalité de Cauchy–Schwarz.

Puis $\mathrm{cov}(X,X) = \mathrm{Var}(X) \ge 0$. Et $\mathrm{Var}(X) = 0$ ssi $X = \overline{X}$ (presque surement). Et le sous-espace de $L^2(\Omega,P)$ des v.a.r. de moyenne nulle est un sous-espace vectoriel. ▪

Définition 12.11 La matrice 2×2 symétrique positive $K = \begin{pmatrix} \mathrm{cov}(X,X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & \mathrm{cov}(Y,Y) \end{pmatrix}$ est la matrice de covariance du couple (X,Y).

Plus généralement, étant donné un vecteur aléatoire $\vec{X} = (X_1,...,X_n)$, la matrice $n \times n$ définie par $K = [\mathrm{cov}(X_i,X_j)]_{\substack{i=1,\dots,n \\ j=1,\dots,n}}$ est la matrice de covariance du vecteur aléatoire $\vec{X}$.

12.2.3 Propriétés

Proposition 12.12 *(12.6) est équivalent à (définition alternative), dans* $L^2(\Omega,P)$ *:*

$$\mathrm{cov}(X,Y) = E((X-\overline{X})(Y-\overline{Y})) \qquad (= P((X-\overline{X})(Y-\overline{Y}))), \tag{12.11}$$

ce qui s'écrit aussi :

$$\mathrm{cov}(X,Y) = \int_\Omega (X(\omega)-\overline{X})(Y(\omega)-\overline{Y})\,dP(\omega) = (X-\overline{X},Y-\overline{Y})_{L^2(\Omega,P)}. \tag{12.12}$$

Et on a, pour tout $a,b \in \mathbb{R}$ *:*

$$\mathrm{cov}(X-a,Y-b) = \mathrm{cov}(X,Y). \tag{12.13}$$

Preuve. $E((X-\overline{X})(Y-\overline{Y})) = E(XY) - \overline{X}E(Y) - \overline{Y}E(X) + \overline{X}\,\overline{Y} = E(XY) - \overline{X}\,\overline{Y} = \mathrm{cov}(X,Y)$ par linéarité de E.

Puis $X-a-E(X-a) = X-\overline{X}$, par linéarité de E, d'où $\mathrm{cov}(X-a,Y-b) = E((X-\overline{X})(Y-\overline{Y})) = \mathrm{cov}(X,Y)$, i.e. (12.13). ▪

Donc, cas discret :

$$\mathrm{cov}(X,Y) = \sum_{i\in\mathbb{N}^*} (X(\omega_i) - \overline{X})(Y(\omega_i) - \overline{Y})\,q_{\omega_i} = \sum_{i\in\mathbb{N}^*} (x_i - \overline{X})(y_i - \overline{Y})\,q_{\omega_i}, \tag{12.14}$$

et dans le cas continu :

$$\mathrm{cov}(X,Y) = \int_\omega (X(\omega) - \overline{X})(Y(\omega) - \overline{Y})\,q(\omega)\,d\omega. \tag{12.15}$$

Proposition 12.13 *On a pour* $X,Y \in L^2(\Omega,P)$ *:*

$$X \text{ et } Y \text{ indépendantes} \quad \Longrightarrow \quad \mathrm{cov}(X,Y) = 0. \tag{12.16}$$

(Orthogonalité relative au semi-produit scalaire cov*.) La réciproque est fausse.*

En particulier, dans le sous-espace vectoriel de $L^2(\Omega,P)$ *des fonctions de moyenne nulle,* X *et* Y *indépendantes implique* $X \perp Y$ *au sens* $L^2(\Omega,P)$*.*

Preuve. Si les v.a.r. sont indépendantes, on a $E(XY) = E(X)E(Y)$, cf. (10.20), d'où (12.16). Et prenant $X = Y = 1_A$ on a $\mathrm{cov}(X,Y) = 0$ alors que X et Y ne sont pas indépendantes pour A t.q. $P(A) \ne 0,1$, car alors $P((1_A \in \{0\}) \cap (1_A \in \{0\})) \ne P(1_A \in \{0\})P(1_A \in \{0\})$, cf. (4.8) et (5.7). Voir également exercice 12.15. ▪

Exemple 12.14 Si les X_k indépendantes ont pour loi $dP_{X_k} = \frac{1}{\sigma_k\sqrt{2\pi}}e^{-\frac{(x-m_k)^2}{2\sigma_k^2}}\,dx$ (gaussiennes cf. (12.11)), alors $\mathrm{cov}(X_1,X_2)=0$, donc $E(X_1X_2)=E(X_1)E(X_2)$, cf. (12.6), donc :

$$E(X_1X_2) = (\int_{\mathbb{R}} x\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma_1^2}}\,dx)(\int_{\mathbb{R}} y\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{y^2}{2\sigma_2^2}}\,dy) - 0 = 0.$$

(Les intégrants sont des fonctions impaires.) ▪

Exercice 12.15 Soit X et Y deux v.a.r. gaussiennes. Donner un exemple où $\mathrm{cov}(X,Y)=0$ bien que X et Y sont dépendantes.

(En revanche : le vecteur aléatoire $\vec{X}=(X,Y)$ est gaussien $\Longleftrightarrow$ $\mathrm{cov}(X,Y)=0$, car, par définition d'un vecteur gaussien, le vecteur (X,Y) est gaussien ssi $\alpha X+\beta Y$ gaussien pour tout α,β.)

Réponse. Soit $X = id : \omega\in\mathbb{R}\to x = X(\omega)=\omega\in\mathbb{R}$. On suppose que la loi de X est la loi $\mathcal{N}(0,1)$, soit loi de densité $p_X(x)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$ (ici $P_X = P\circ X^{-1} = P$). Soit $a>0$ (qu'on va choisir ultérieurement), et soit Y la v.a.r. donnée par $Y = \left\{\begin{matrix} +X \text{ si } |X|\le a \\ -X \text{ si } |X|>a \end{matrix}\right\}$, donc $Y(\omega) = \left\{\begin{matrix} +X(\omega) \text{ si } X(\omega)\in[-a,a] \\ -X(\omega) \text{ sinon} \end{matrix}\right\}$.

Montrons que la loi de Y est la loi $\mathcal{N}(0,1)$. Pour $b>a$ on a On a $P(Y<-b) = P((-X)<-b) = P(X>b) = P(X<-b)$ par symétrie de la gaussienne centrée. Et donc $P(Y\in[a,b]) = P(-X\in[a,b]) = P(X\in[-b,-a]) = P(X\in[a,b])$ par symétrie de la gaussienne centrée. Puis pour $b\in[0,a]$ on a $P(Y\in[-a,b]) = P(X\in[-a,b])$. Donc Y et X ont la même fonction de répartition : $P_Y = P_X = \mathcal{N}(0,1)$.

On a $XY = X^2$ sur $X^{-1}([-a,a])$ et $XY=-X^2$ sur $X^{-1}(\mathbb{R}-[-a,a])$. Donc $E(XY) = P(XY) = \int_{\omega\in[-a,a]} X^2(\omega)\,dP(\omega) - \int_{\omega\in\mathbb{R}-[-a,a]} X^2(\omega)\,dP(\omega) = \int_{x\in[-a,a]} x^2 dP_X(x) - \int_{x\in\mathbb{R}-[-a,a]} x^2\,dP_X(x)$, cf. loi image avec $X = id$.

On choisi a t.q. $\int_{x\in[-a,a]} x^2dP_X(x) = \int_{x\in\mathbb{R}-[-a,a]} x^2\,dP_X(x)$: donc $E(XY)=0$. Et comme $E(X) = P_X(id) = 0$ (gaussienne centrée), on a $\mathrm{cov}(X,Y)=0$.

Montrons que X et Y ne sont pas indépendantes. $P(X\in[-a,a]\cap(Y\in[a,a]) = P(X\in[-a,a]) \ne P(X\in[-a,a])^2 = P(X\in[-a,a])P(Y\in[a,a])$ (car $a>0$).

Ici les v.a.r. sont non corrélées mais sont dépendantes.

(Ici $X+Y = 2X1_{|X|\le a}$ n'est pas gaussien : le vecteur $\vec{X}=(X,Y)$ n'est pas gaussien.) ▪

Exercice 12.16 Montrer que :

$$\mathrm{cov}(aX+b,cY+d) = ac\,\mathrm{cov}(X,Y). \tag{12.17}$$

Réponse. $E(aX+b1_\Omega) = aE(X)+b$ par linéarité de E, Donc $aX+b-E(aX+b) = a(X-\overline{X})$. Idem pour Y. Donc $(aX+b-E(aX+b))(cY+d-E(cY+d)) = ac(X-\overline{X})(Y-\overline{Y})$. Ayant (12.11) et E linéaire on en déduit le résultat. ▪

12.2.4 Coefficient de corrélation

Définition 12.17 Soit X et Y deux v.a.r. définies sur un même espace probabilisé.

Quand $\sigma(X)\ne 0$ et $\sigma(Y)\ne 0$ (les v.a.r. ne sont pas constantes), le coefficient de corrélation des deux v.a.r. X et Y est le réel "adimensionnel" :

$$\rho_{XY} \stackrel{\text{déf}}{=} \frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)}. \tag{12.18}$$

Comme $|\mathrm{cov}(X,Y)|\le\sigma(X)\sigma(Y)$, cf. Cauchy–Schwarz (12.10), on a toujours :

$$|\rho_{XY}|\le 1. \tag{12.19}$$

Et avec $\mathrm{cov}(X,Y) = \mathrm{cov}(X-\overline{X},Y-\overline{Y})$ et la bilinéarité de cov, on a également :

$$\rho_{XY} = E(\frac{X-\overline{X}}{\sigma(X)}\frac{Y-\overline{Y}}{\sigma(Y)}) = (\frac{X-\overline{X}}{\sigma(X)},\frac{Y-\overline{Y}}{\sigma(Y)})_{L^2(\Omega,P)}, \tag{12.20}$$

i.e. c'est le produit scalaire des v.a.r. centrées réduites (appelées unitaires) dans l'espace $L^2(\Omega,P)$ (ici on suppose donc $X,Y\in L^2(\Omega,\mathcal{T},P)$).

Proposition 12.18 *Quand $K_{\vec X}$ est définie positive et quand les v.a.r. ne sont pas constantes :*

$$\rho_{XY} = \pm 1 \quad \Longleftrightarrow \quad \exists (a,b,c) \in \mathbb{R}^* \times \mathbb{R}^* \times \mathbb{R} \quad : \quad aX + bY = c \text{ presque surement}, \tag{12.21}$$

i.e. X et Y sont dépendantes de manière affine.

Preuve. $\Rightarrow$. On a $\mathrm{cov}(X,Y) = (X-\overline{X}, Y-\overline{Y})_{L^2(\Omega,P)}$, cf. (12.12), donc (Cauchy–Schwarz) $\mathrm{cov}(X,Y) \le \sigma(X)\sigma(Y)$ avec égalité ssi $X-\overline{X}$ est parallèle à $Y-\overline{Y}$, i.e. ssi $\exists a,b \in \mathbb{R}$ t.q. $a(X-\overline{X}) + b(Y-\overline{Y}) = 0$, avec l'un des deux non nuls. Donc ici les deux non nuls (si $a=0$ alors $b \neq 0$, donc $Y = \overline{Y}$ et $\sigma(Y) = 0$, cas exclu, idem si $b=0$).

$\Leftarrow$. Si $aX + bY = c$, avec a et b non nuls, alors $Y = \frac{1}{b}(c - aX)$, donc $(X-\overline{X})(Y-\overline{Y}) = \frac{1}{b}(X-\overline{X})(c-aX-c+a\overline{X}) = \frac{a}{b}(X-\overline{X})^2$, d'où $E((X-\overline{X})(Y-\overline{Y})) = \frac{a}{b}E((X-\overline{X})^2)$, et $\mathrm{Var}(Y) = \mathrm{Var}(\frac{1}{b}(c-aX)) = (\frac{a}{b})^2\mathrm{Var}(X)$, d'où $\rho_{XY} = \pm 1$. ∎

12.2.5 Approximation linéaire d'une v.a.r par une autre

Soit X et Y deux v.a.r. sur un même espace probabilisé. But : trouver la meilleure approximation (au sens des moindres carrés) de Y par la fonction $aX + b$:

$$\text{trouver } a,b \in \mathbb{R} \text{ qui réalisent le minimum de } ||Y - (aX+b)||_P^2 = E\big((Y - (aX+b))^2\big). \tag{12.22}$$

Proposition 12.19 *Le minimum est réalisé pour :*

$$a = \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(X)}, \qquad b = E(Y) - aE(X). \tag{12.23}$$

Preuve. $(Y-(aX+b))^2 = Y^2 - 2aXY - 2bY + a^2X^2 + 2abX + b^2$, d'où $E\big((Y-(aX+b))^2\big) = E(Y^2) - 2aE(XY) - 2bE(Y) + a^2E(X^2) + 2abE(X) + b^2 =^{\text{noté}} f(a,b)$, donc $\frac{\partial f}{\partial a}(a,b) = -2E(XY) + 2aE(X^2) + 2bE(X)$ et $\frac{\partial f}{\partial b}(a,b) = -2E(Y) + 2aE(X) + 2b$. On veut $\frac{\partial f}{\partial a}(a,b) = \frac{\partial f}{\partial b}(a,b) = 0$, donc $b = E(Y) - aE(X)$ puis $-2E(XY) + 2aE(X^2) + 2E(Y)E(X) - 2aE(X)^2 = 0$. ∎

12.2.6 Variance d'une somme

Proposition 12.20 *On a, pour des v.a.r. X et Y sur un même espace probabilisé $(\Omega, \mathcal{T}, P)$:*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{cov}(X,Y). \tag{12.24}$$

Et de manière plus générale, pour n v.a.r. X_i :

$$\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \sum_{\substack{i,j=1 \\ i<j}}^n \mathrm{cov}(X_i, X_j). \tag{12.25}$$

En particulier, dès que les v.a.r. X_i sont indépendantes (covariances croisées nulles), la variance de la somme est égale à la somme des variances.

Preuve. $\mathrm{cov}(X+Y, X+Y) = \mathrm{cov}(X,X) + 2\mathrm{cov}(X,Y) + \mathrm{cov}(Y,Y)$ car cov est une forme bilinéaire symétrique. D'où (12.24). De même pour (12.25).

Calcul direct : on a $((X+Y) - E(X+Y))^2 = ((X-\overline{X}) + (Y-\overline{Y}))^2$. Donc :

$$\mathrm{Var}(X+Y) = \int_\omega (X(\omega)-\overline{X})^2\,dP + \int_\omega (Y(\omega)-\overline{Y})^2\,dP + 2\int_\omega (X(\omega)-\overline{X})(Y(\omega)-\overline{Y})\,dP.$$

Et $\mathrm{Var}(\sum_{i=1}^{n-1} X_i + X_n) = \mathrm{Var}(\sum_{i=1}^{n-1} X_i) + \mathrm{Var}(X_n) + 2\mathrm{cov}(\sum_{i=1}^{n-1} X_i, X_n)$, d'où, avec la bilinéarité et la symétrie de cov, et par récurrence :

$$\mathrm{Var}(\sum_{i=1}^{n-1} X_i + X_n) = \sum_{i=1}^{n-1} \mathrm{Var}(X_i) + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n-1} \mathrm{cov}(X_i, X_j) + \mathrm{Var}(X_n) + 2\sum_{i=1}^{n-1} \mathrm{cov}(X_i, X_n),$$

soit (12.25). ∎

12.3 Matrice de covariance

Soit $X_1, ..., X_n : \Omega \to \mathbb{R}$ n v.a.r.. Soit $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} : \omega \in \Omega \to \vec{X}(\omega) = \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix} \in \mathbb{R}^n$ le vecteur aléatoire associé.

Définition 12.21 La matrice de covariance (ou de dispersion) du vecteur aléatoire $\vec{X}$ est la matrice $n \times n$:

$$K_{\vec{X}} = [\mathrm{cov}(X_i, X_j)] = \begin{pmatrix} \mathrm{cov}(X_1, X_1) & ... & \mathrm{cov}(X_1, X_n) \\ \vdots & & \vdots \\ \mathrm{cov}(X_n, X_1) & ... & \mathrm{cov}(X_n, X_n) \end{pmatrix} \overset{\text{noté}}{=} [K_{\vec{X},ij}]. \tag{12.26}$$

Exemple 12.22 Suite de l'exemple 12.8 : $K_{X,Y} = \begin{pmatrix} V & 0 \\ 0 & V \end{pmatrix}$, où $V = \mathrm{Var}(X) = \mathrm{Var}(Y)$.

Exemple 12.23 Suite de l'exemple 12.9 : $K_{X,Y} = \begin{pmatrix} V_1 & -\frac{14}{9} \\ -\frac{14}{9} & V_2 \end{pmatrix}$, où $V_i = \mathrm{Var}(X_i)$.

Si $\vec{y} \in \mathbb{R}^n$, on note $\vec{y}^T.\vec{X}$ la v.a.r. $\omega \to \vec{y}^T.\vec{X}(\omega) = \sum_{j=1}^n y_j X_j(\omega)$, et on note $\vec{y}^T.K_{\vec{X}}.\vec{y}$ la v.a.r. $\omega \to (\vec{y}^T.K_{\vec{X}}.\vec{y})(\omega) = \vec{y}^T.\vec{X}(\omega).\vec{y} = \sum_{i=1}^n y_i K_{\vec{X},ij}(\omega) y_j$.

Lemme 12.24 *Pour tout $\vec{y} \in \mathbb{R}^n$ on a :*

$$\vec{y}^T.K_{\vec{X}}.\vec{y} = \mathrm{Var}(\vec{y}^T.\vec{X}). \tag{12.27}$$

(Permet de se ramener aux v.a.r..)

Preuve. $\mathrm{Var}(\vec{y}^T.\vec{X}) = \mathrm{Var}(\sum_{i=1}^n y_i X_i) = \mathrm{cov}(\sum_{i=1}^n y_i X_i, \sum_{j=1}^n y_j X_j) = \sum_{i,j=1}^n y_i \mathrm{cov}(X_i, X_j) y_j$ par bilinéarité de cov.

Proposition 12.25 *Une matrice de covariance est symétrique positive.*

Preuve. C'est la symétrie et la positivité de cov, cf. proposition 12.10.

Proposition 12.26 *Si les composantes $X_i \in L^2(\Omega, P)$ de $\vec{X}$ sont indépendantes alors la matrice de covariance est diagonale :*

$$\text{les } X_i \text{ indépendantes} \quad \Longrightarrow \quad K_{\vec{X}} = \mathrm{diag}(\mathrm{Var}(X_i)). \tag{12.28}$$

(La réciproque est fausse, voir proposition 12.13.)

Preuve. Cf. (12.16) : $\mathrm{cov}(X_i, X_j) = 0$ dans ce cas.

Soit $[\vec{Y}] = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ la notation générique de la matrice (colonne) représentant $\vec{Y}$ dans la base canonique de $\mathbb{R}^n$, et $[\vec{Y}]^T$ sa matrice (ligne) transposée. Ainsi on dispose de la matrice $n \times n$ donnée par $\vec{Y}.\vec{Y}^T = [Y_i Y_j]_{i,j=1,...,n}$ dont la trace vaut $\mathrm{Tr}(\vec{Y}.\vec{Y}^T) = \sum_{i=1}^n Y_i^2 = ||\vec{Y}||^2$.

Proposition 12.27 *Pour $\vec{X} \in L^2(\Omega, P)^n$ on a :*

$$\begin{aligned} K_{\vec{X}} &= E([\vec{X} - \overline{\vec{X}}].[\vec{X} - \overline{\vec{X}}]^T) = E(\vec{X}.\vec{X}^T) - E(\vec{X})E(\vec{X})^T \\ &= E([(X_i - \overline{X}_i)(X_j - \overline{X}_j)]_{ij}) = E([X_i X_j]_{ij}) - [\overline{X}_i \overline{X}_j]_{ij}, \end{aligned} \tag{12.29}$$

Et la matrice de covariance $K_{\vec{X}}$ est symétrique et positive (elle est donc diagonalisable et toutes ses valeurs propres sont positives ou nulles). Et on a :

$$\mathrm{Tr}(K_{\vec{X}}) = E(||\vec{X} - E(\vec{X})||^2). \tag{12.30}$$

Et si M est une matrice $m * n$ (en particulier si M est une matrice $n * n$ de changement de base) :

$$K_{M.\vec{X}} = M.K_{\vec{X}}.M^T. \tag{12.31}$$

Preuve. Soit $\vec{Y} = \vec{X} - E(\vec{X})$, on a $Y_iY_j = (X_i - \overline{X}_i)(X_j - \overline{X}_j)$, donc $E(Y_iY_j) = \mathrm{cov}(X_i, X_j)$, d'où $K_{\vec{X}} = E([\vec{X} - \vec{\overline{X}}].[\vec{X} - \vec{\overline{X}}]^T) = E([(X_i - \overline{X}_i)(X_j - \overline{X}_j)]_{ij})$, d'où (12.30).

Puis $[\vec{X} - E(\vec{X})].[\vec{X} - E(\vec{X})]^T = \vec{X}.\vec{X}^T - \vec{X}.E(\vec{X})^T - E(\vec{X}).\vec{X}^T + E(\vec{X}).E(\vec{X})^T$, avec $E(\vec{X}.E(\vec{X})^T) = E([X_i\overline{X}_j]_{ij}) = [E(X_i\overline{X}_j)]_{ij} = [\overline{X}_jE(X_i)]_{ij} = [\overline{X}_j\overline{X}_i)]_{ij} = E(\vec{X}).E(\vec{X}]^T = E(E(\vec{X}).\vec{X}^T)$. D'où $E([\vec{X} - E(\vec{X})].[\vec{X} - E(\vec{X})]^T) = E(\vec{X}.\vec{X}^T) - E(\vec{X})E(\vec{X})^T = K_{\vec{X}}$. ▪

Remarque 12.28 Autre introduction de la matrice de covariance : comme matrice d'une forme quadratique dans une base orthonormée (lien usuel forme bilinéaire symétrique $b : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ et forme quadratique associée $q : \mathbb{R} \to \mathbb{R}$ définie par $q(\vec{x}) = b(\vec{x}, \vec{x})$) :

On note $(\cdot,\cdot)_{\mathbb{R}^d}$ le produit scalaire euclidien de $\mathbb{R}^d$. Soit $\vec{X} : \Omega \to \mathbb{R}^d$ un vecteur aléatoire. Si $\vec{y} \in \mathbb{R}^d$ est un vecteur, alors $(\vec{X}, \vec{y})_{\mathbb{R}^d} : \Omega \to \mathbb{R}$ est la fonction (v.a.r.) définie par :

$$(\vec{X}, \vec{y})_{\mathbb{R}^d}(\omega) = (\vec{X}(\omega), \vec{y})_{\mathbb{R}^d}.$$

(Connaître $\vec{X}$ c'est connaître tous les $(\vec{X}(\omega), \vec{y})_{\mathbb{R}^d}$.) Soit $\Psi_{\vec{X}} : \mathbb{R}^d \to \mathbb{R}$ définie par :

$$\Psi_{\vec{X}}(\vec{y}) = \mathrm{Var}((\vec{X}, \vec{y})_{\mathbb{R}^d}) \qquad (= \mathrm{Var}(\sum_{i=1}^d y_iX_i)).$$

(Cette fonction de $\vec{y}$ est indépendante du choix de la base de $\mathbb{R}^d$, elle ne dépend que du choix du produit scalaire, ici le produit scalaire euclidien).

Soit une base orthonormée de $\mathbb{R}^d$ (relativement au produit scalaire euclidien), dans laquelle $\vec{X} = (X_1, ..., X_d)$ (fonction) et $\vec{y} = (y_1, ..., y_d)$ (vecteur). Comme $E((\vec{X}, \vec{y})_{\mathbb{R}^d}) = E(\sum_i y_iX_i) = \sum_i y_iE(X_i)$, on a :

$$\begin{aligned}\Psi_{\vec{X}}(\vec{y}) &= E([\sum_{i=1}^d y_i(X_i - \overline{X}_i)]^2) = E(\sum_{i,j=1}^d y_iy_j(X_i - \overline{X}_i)(X_j - \overline{X}_j)) = \sum_{i,j=1}^d y_iy_j\mathrm{cov}(X_i, X_j)\\ &= \vec{y}^T.K.\vec{y}.\end{aligned}$$

Ainsi $\Psi_{\vec{X}}$ est une forme quadratique et $K_{\vec{X}}$ est la matrice de $\Psi_{\vec{X}}$ dans la b.o.n. choisie. ▪

12.4 Exemple : problème d'échantillonnage

N.B. : comme souvent, le problème de compréhension réside dans les notations.

Cadre initial. Population $S = \{a_1, ..., a_n\}$. Tirage équiprobable : probabilité $P = \sum_{i=1}^n \frac{1}{n}\delta_{a_i}$, i.e. la probabilité d'obtention de a_i est $P(a_i) = \frac{1}{n}$ pour tout i.

Soit $f : S \to \mathbb{R}$ la v.a.r. qui donne la valeur d'une caractéristique de a_i, par exemple $f(a_i)$ est la taille de a_i. La valeur moyenne (espérance) de f et sa variance sont :

$$\left\{\begin{aligned} &E(f) \overset{\text{déf}}{=} P(f) = \frac{1}{n}\sum_{i=1}^n f(a_i),\\ &\mathrm{Var}(f) = E((f - m1_S)^2) = \frac{1}{n}\sum_{i=1}^n (f(a_i) - m)^2 = \sigma(f)^2.\end{aligned}\right. \tag{12.32}$$

On souhaite estimer cette moyenne et cette variance à l'aide d'un échantillon de taille $r < n$.

Cadre des échantillons. Tirage de taille r avec replacement.
L'univers des échantillons est S^r de cardinal $|S^r| = n^r$.

Pour un échantillon $\vec{\omega} = (\omega_1, ..., \omega_r) \in S^r$ (considéré comme une population), avec hypothèse d'équiprobabilité d'obtention des ω_i, la probabilité est $P_{r,\vec{\omega}} = \sum_{i=1}^r \frac{1}{r}\delta_{\omega_i}$, i.e. la probabilité d'obtention de ω_i est $P_{r,\vec{\omega}}(\omega_i) = \frac{1}{r}$ pour tout $i = 1, ..., r$. L'espérance ("valeur moyenne") et la variance de

f sur cet échantillon $\vec{\omega}$ sont :

$$\begin{cases} E_{r,\vec{\omega}}(f) \stackrel{\text{déf}}{=} P_{r,\vec{\omega}}(f) = \dfrac{1}{r}\displaystyle\sum_{j=1}^{r} f(\omega_j) \stackrel{\text{noté}}{=} E_{r,f}(\vec{\omega}), \\ \mathrm{Var}_{r,\vec{\omega}}(f) = \dfrac{1}{r}\displaystyle\sum_{j=1}^{r}(f(\omega_j) - E_{r,\vec{\omega}}(f))^2. \end{cases} \tag{12.33}$$

On a ainsi en particulier défini la v.a.r. $X =^{\text{noté}} E_{r,f}$:

$$X = E_{r,f} : \begin{cases} S^r & \to \mathbb{R} \\ \vec{\omega} & \to E_{r,f}(\vec{\omega}) \stackrel{\text{déf}}{=} E_{r,\vec{\omega}}(f) = \dfrac{1}{r}\displaystyle\sum_{j=1}^{r} f(\omega_j), \end{cases} \tag{12.34}$$

qui a un échantillon de taille r associe la valeur moyenne sur cet échantillon. (Exemple : $E_{r,f}(\vec{\omega}) =$ taille moyenne de l'échantillon.)

On suppose que les échantillons sont équiprobables : probabilité $Q = \displaystyle\sum_{i=1}^{n^r} \frac{1}{n^r}\delta_{\vec{\omega}}$, où on a noté $\vec{\omega} = (\omega_1, ..., \omega_r) \in S^r$ un échantillon (où donc $\omega_i \in S$ est le résultat du i-ème tirage), i.e. $Q(\vec{\omega}) = \dfrac{1}{n^r}$ pour tout $\vec{\omega} \in S^r$.

Proposition 12.29 *L'espérance de la v.a.r. $X = E_{r,f}$ est :*

$$(Q(E_{r,f}) =) \qquad E(E_{r,f}) = E(f) \qquad (= P(f)), \tag{12.35}$$

i.e. la moyenne de "la moyenne sur chaque échantillon" est égale à la moyenne sur la population.

Preuve. Il y a n^r échantillons possibles. on peut ordonner l'ensemble des échantillons possibles. Notons $\vec{\nu}(j) = (\nu_1(j), ..., \nu_r(j)) \in S^r$ le j-ème, $j = 1, ..., n^r$. Pour $i \in [1, r]_{\mathbb{N}}$ notons $\nu_i : j \in [1, n^r]_{\mathbb{N}} \to \nu_i(j) \in S$. Les ν_i sont des v.a.r. indépendantes et toutes distribuées suivant la loi uniforme sur S, à savoir $P_{\nu_i} = P \circ \nu_i^{-1} = \sum_{\ell=1}^{n} \frac{1}{n}\delta_{a_\ell}$, i.e. $P_{\nu_i}(a_k) = P(\nu_i^{-1}(a_k)) = P(\{j \in [1, n^r]_{\mathbb{N}} : \nu_i(j) = a_k\}) = \frac{1}{n}$).

Les tirages d'échantillons étant supposés équiprobables, la moyenne sur l'ensemble des $n^r = |S^r|$ échantillons distincts possibles est donc :

$$Q(E_{r,f}) = \sum_{\vec{\omega}\in S^r} \frac{1}{n^r} E_{r,f}(\vec{\omega}) = \sum_{j=1}^{n^r} \frac{1}{n^r} E_{r,f}(\vec{\nu}(j)) = \sum_{j=1}^{n^r} \frac{1}{n^r}\frac{1}{r}\sum_{i=1}^{r} f(\nu_i(j)) = \frac{1}{r}\sum_{i=1}^{r}(\frac{1}{n^r}\sum_{j=1}^{n^r} f(\nu_i(j))).$$

Et, pour $j = 1, ..., n^r$, $\nu_i(j)$ prend la valeur a_k un nombre n^{r-1} fois, pour $k = 1, ..., n$. D'où, pour $i = 1, ..., r$:

$$\frac{1}{n^r}\sum_{j=1}^{n^r} f(\nu_i(j)) = \frac{1}{n}\sum_{k=1}^{n} f(a_k) = P(f). \tag{12.36}$$

D'où $E(E_{r,f}) =^{\text{déf}} Q(E_{r,f}) = \frac{1}{r}\sum_{i=1}^{r} P(f) = P(f) =^{\text{noté}} E(f)$. ∎

Proposition 12.30 *Notons $m = E(f) = E(E_{r,f})$. L'écart type $\sigma(E_{r,f}) = (E((E_{r,f} - m)^2)^{\frac{1}{2}} = ||E_{r,f} - m||_{L^2(\Omega,P)}$ (l'erreur quadratique moyenne) est :*

$$\sigma(E_{r,f}) = \frac{\sigma(f)}{\sqrt{r}}, \tag{12.37}$$

où $\sigma(f)$ est l'écart type de f, cf. (12.32). Et cette erreur est d'autant plus faible que l'échantillon est grand, la "qualité de l'approximation" étant en $\frac{1}{\sqrt{r}}$.

Preuve. Quitte à changer f en $f-m$, prenons $m=0$. Et :

$$E(E_{r,f}^2) = \sum_{j=1}^{n^r} \frac{1}{n^r} E_{r,f}^2(\vec{\nu}(j)) = \sum_{j=1}^{n^r} \frac{1}{n^r} \Big(\frac{1}{r}\sum_{i=1}^{r} f(\nu_i(j))\Big)\Big(\frac{1}{r}\sum_{k=1}^{r} f(\nu_k(j))\Big)$$

Avec, les v.a.r. ν_i étant indépendantes, il en est de même des v.a.r. $f(\nu_i)$, cf. (5.10), d'où :

$$0 = \mathrm{cov}(f(\nu_i), f(\nu_k)) = E((f\circ\nu_i)(f\circ\nu_k)) = \sum_{j=1}^{n^r} \frac{1}{n^r} f(\nu_i(j)) f(\nu_k(j)) \quad \text{si } i \neq j. \tag{12.38}$$

D'où :

$$E(E_{r,f}^2) = \frac{1}{r^2}\sum_{j=1}^{n^r}\frac{1}{n^r}\Big(\sum_{i=1}^{r} f(\nu_i(j))^2\Big) = \frac{1}{r^2}\sum_{i=1}^{r}\sum_{j=1}^{n^r}\frac{1}{n^r} f(\nu_i(j))^2 = \frac{1}{r^2}\sum_{i=1}^{r}\sum_{i=1}^{n}\frac{1}{n} f(a_i)^2 = \frac{1}{r}\mathrm{Var}(f),$$

l'avant dernière égalité comme pour la preuve précédente. ∎

12.5 Méthode des moindres carrés : notations probabilistes

Soit $X \in L^2(\Omega, P)$ muni du produit scalaire $(f,g)_{L^2} = \int_\Omega f(x)g(x)\,dP(x) = P(fg)$.

(Si $\Omega = [0,1]$ et $dP = dx$ la mesure de Lebesgue, on retrouve méthode des moindres carrées usuelle.)

Soit $V_h \subset L^2(\Omega, P)$ un sous-espace vectoriel de dimension finie n (par exemple $V_h = P_1 =$ les fonctions continues affines par morceau sur un maillage donné). Ce sous-espace est fermé (car de dimension finie), donc pour un $X \in L^2(\Omega,P)$ donné, il existe une unique fonction $X_h \in V_h$ telle que :

$$||X - X_h||_{L^2} = \min_{Y_h\in V_h} ||X - Y_h||_{L^2}.$$

(On dit que X_h réalise le minimum.) C'est le théorème de projection orthogonal, l'opérateur π_h : $X \to X_h$ étant l'opérateur de projection orthogonal sur V_h relativement au produit scalaire $(\cdot,\cdot)_{L^2}$:

$$\pi_h : \begin{cases} L^2(\Omega,P) & \to V_h \\ \qquad X & \to X_h = \pi_h(X) \qquad \text{t.q.} \quad (X-X_h, Y_h)_{L^2} = 0, \quad \forall Y_h \in V_h, \end{cases} \tag{12.39}$$

soit $\int_\Omega (X-X_h) Y_h\,dP = 0$, pour tout $Y_h \in V_h$. Autrement dit, avec la notation E de l'espérance, X_h est caractérisé par :

$$(P((X-X_h)Y_h) =) \qquad E((X-X_h)Y_h) = 0, \quad \forall Y_h \in V_h.$$

Supposons qu'on connaisse un ensemble de générateurs de V_h, i.e. une famille $(\varphi_1, ..., \varphi_m)$ de v.a.r. telle que $V_h = \mathrm{Vect}\{\varphi_1, ..., \varphi_m\}$. La famille $(\varphi_1, ..., \varphi_m)$ n'est pas nécessairement libre : les φ_i sont par exemple donnée par l'expérience "passée". (Si on en extrait une famille libre, alors cette famille forme une base de V_h.)

On va supposer que V_h contient les fonctions constantes.

Calcul : matrice de covariance

La notation $\overline{X} = E(X)$ représente le réel "valeur moyenne" ainsi que la fonction constante $\overline{X} 1_\Omega$.

Proposition 12.31 *On suppose que la famille $(1_\Omega, \varphi_1, ..., \varphi_m)$ est génératrice de V_h (il n'est pas exclu que l'une des φ_i soit une fonction constante : la famille n'est pas supposée libre). Pour X une v.a.r., on a :*

$$X_h = \pi_h(X) \quad \Longrightarrow \quad E(X_h) = E(X), \tag{12.40}$$

et posant $X_h = \sum_{j=1}^m a_j \varphi_j$ on a :

$$X_h - E(X_h) = \sum_{j=1}^{m} a_j(\varphi_j - E(\varphi_j)), \tag{12.41}$$

où les a_j sont une solution du système linéaire :

$$\sum_{j=1}^{m} [\mathrm{cov}(\varphi_i, \varphi_j)].a_j = \mathrm{cov}(X, \varphi_i), \tag{12.42}$$

i.e. les a_j sont une solution du système linéaire :

$$K_{\vec\varphi}.\vec a=\vec b \qquad \text{où} \qquad K_{\vec\varphi}=\mathrm{cov}[(\varphi_i,\varphi_j)], \quad \vec a=\begin{pmatrix}a_1\\ \vdots\\ a_m\end{pmatrix}, \quad \vec b=\begin{pmatrix}\mathrm{cov}(X,\varphi_1)\\ \vdots\\ \mathrm{cov}(X,\varphi_n)\end{pmatrix}. \tag{12.43}$$

Et dans ce cas "le carré de l'erreur" commise en approchant X par X_h est appelée "variance résiduelle" (l'erreur au carrée) et vaut :

$$||X-X_h||^2_{L^2}=\mathrm{Var}(X)-\mathrm{Var}(X_h), \tag{12.44}$$

soit :

$$E((X-X_h)^2)=\mathrm{Var}(X)-(\vec a,\vec b)_{\mathbb{R}^{m+1}}. \tag{12.45}$$

Si de plus les φ_i forment une famille libre, alors $K_{\vec\varphi}$ est inversible et $\vec a=K_{\vec\varphi}^{-1}.\vec b$.

Preuve. $1_\Omega\in V_h$, et $(X_h,1_\Omega)_{L^2}=(X,1_\Omega)_{L^2}$ s'écrit aussi $E(X_h)=E(X)$, i.e. (12.40).

La famille $(1_\Omega,\varphi_1,...,\varphi_m)$ étant génératrice dans V_h, il en est de même de la famille $(1_\Omega,\varphi_1-\overline\varphi_1 1_\Omega,...,\varphi_m-\overline\varphi_m 1_\Omega)=^{\text{noté}}(1_\Omega,\varphi_1-\overline\varphi_1,...,\varphi_m-\overline\varphi_m)$ (trivial). Donc il existe des a_i t.q. $X_h=a_0 1_\Omega+\sum_{j=1}^m a_j(\varphi_j-\overline\varphi_j)$, et donc $\overline X_h=a_0$ $(=\overline X)$.

Puis $\varphi_i-\overline\varphi_i\in V_h$ donne les m équations $(X_h,\varphi_i-\overline\varphi_i)_{L^2}=(X,\varphi_i-\overline\varphi_i)_{L^2}$, pour tout $i=1,...,m$, soit :

$$0+\sum_{j=1}^m a_j(\varphi_j-\overline\varphi_j,\varphi_i-\overline\varphi_i)_{L^2}=(X,\varphi_i-\overline\varphi_i)_{L^2}=(X-\overline X,\varphi_i-\overline\varphi_i)_{L^2}, \tag{12.46}$$

i.e. (12.42) qui se réécrit matriciellement sous la forme (12.43).

Puis l'erreur au sens des moindres carrés est $||X-X_h||_{L^2}=\sqrt{E((X-X_h)^2)}$, avec :

$$||X-X_h||^2_{L^2}=||X-\overline X-X_h-\overline X_h||^2_{L^2}=||X-\overline X||^2_{L^2}-||X_h-\overline X_h||^2_{L^2},$$

car $\overline X=\overline X_h$ et $||X-\overline X-X_h-\overline X_h||^2_{L^2}=||X-\overline X||^2_{L^2}-2(X-\overline X,X_h-\overline X_h)_{L^2}+||X_h-\overline X_h||^2_{L^2}$ avec $2(X-\overline X,X_h-\overline X_h)_{L^2}=2(X_h-\overline X_h,X_h-\overline X_h)_{L^2}=2||X_h-\overline X_h||^2_{L^2}$, puisque $(X,X_h)_{L^2}=(X_h,X_h)_{L^2}$, cf. (12.39) avec $Y_h=X_h$. D'où (12.44). Et :

$$\mathrm{Var}(X_h)=||X_h-\overline X_h||^2_{L^2}=(\sum_{i=1}^m a_i(\varphi_i-\overline\varphi_i),\sum_{j=1}^m a_j(\varphi_j-\overline\varphi_j))_{L^2}=\sum_{i,j=1}^m a_i a_j\mathrm{cov}(\varphi_i,\varphi_j)=\vec a^T.K_{\vec\varphi}.\vec a.$$

d'où (12.45) puisque $\vec b=K_{\vec\varphi}.\vec a$.

Et si les v.a.r. φ_i forment une famille libre, alors la matrice des covariances est inversible. ▪

Remarque 12.32 Quand $(1_\Omega,\varphi_1,...,\varphi_m)$ est une famille libre, alors $K_{\vec\varphi}=(\varphi_i-\overline\varphi_i,\varphi_j-\overline\varphi_j)_{L^2}$ est une matrice de masse, donc positive, et donc $\mathrm{Var}(X_h)=\vec a^T.K_{\vec\varphi}.\vec a=\vec a^T.\vec b\geq 0$, et (12.44) donne :

$$\frac{||X-X_h||^2_{L^2}}{\mathrm{Var}(X)}\leq 1, \tag{12.47}$$

rapport adimensionnel de l'erreur au carré sur la variance, rapport qui est d'autant plus petit que X est "proche" de l'espace V_h, et qui s'annule si X appartient à V_h. ▪

Corollaire 12.33 *Dans le cas où V_h est l'espace vectoriel* $\mathrm{Vect}\{1_\Omega,\varphi_1\}$ *avec φ_1 non constante ($\dim V_h=2$ et cas $m=1$ de la proposition précédente correspond au problème de "meilleure approximation par une fonction affine" quand φ_1 est affine), la meilleure approximation de X (au sens des moindres carrés) est la v.a.r. $X_h=\alpha\varphi_1+\beta$ donnée par :*

$$X_h=\frac{\mathrm{cov}(X,\varphi_1)}{\mathrm{Var}(\varphi_1)}(\varphi_1-E(\varphi_1))+E(X), \tag{12.48}$$

et l'erreur commise est donnée par (au carré) :

$$E((X-X_h)^2)=\mathrm{Var}(X)(1-\rho^2_{X,X_h}), \tag{12.49}$$

où $\rho_{X,Y}=\frac{\mathrm{cov}(X,Y)}{\sigma(X)\sigma(Y)}$ est le coefficient de corrélation, cf. (12.18).

Preuve. Avec (12.46), on a $a_1\mathrm{Var}(\varphi_1) = \mathrm{cov}(X,\varphi_1)$, d'où $a_1 = \frac{\mathrm{cov}(X,\varphi_1)}{\mathrm{Var}(\varphi_1)}$, et (12.41) donne (12.48).

Et $\vec{a}^T.\vec{b} = \vec{a}^T.K_{\vec{\varphi}}.\vec{a} = a_1^2\mathrm{Var}(\varphi_1) = \frac{\mathrm{cov}(X,\varphi_1)^2}{\mathrm{Var}(\varphi_1)}$.

Comme $X_h = a_0 + a_1\varphi_1$ on a $\mathrm{Var}(X_h) = a_1^2\mathrm{Var}(\varphi_1)$ et $\mathrm{cov}(X,X_h) = a_1\mathrm{cov}(X,\varphi_1)$, et donc $\rho_{X,X_h} = \frac{a_1\mathrm{cov}(X,\varphi_1)}{a_1\sigma(X)\sigma(\varphi_1)} = \frac{\mathrm{cov}(X,\varphi_1)}{\sigma(X)\sigma(\varphi_1)}$.

Donc $\mathrm{Var}(X)\rho_{X,X_h}^2 = \frac{\mathrm{cov}(X,\varphi_1)^2}{\mathrm{Var}(\varphi_1)} = \vec{a}^T.\vec{b}$. D'où (12.45) donne (12.49). ▪

13 Espérance conditionnelle de Y sachant X

13.1 Loi conditionnelle $P_{|X\in I}$

Soit $(\Omega,\mathcal{T},P)$ un espace probabilisé, et soit X une v.a.r. sur Ω.

Si $I \in \mathcal{T}_\mathbb{R}$ est tel que $P(X{\in}I) \neq 0$, on rappelle que $P_{|X\in I}$ est définie par, cf. (3.3), si $B \in \mathcal{T}$:

$$\begin{aligned} P(B|X{\in}I) = P_{|X\in I}(B) &= \frac{P(B\cap(X{\in}I))}{P(X{\in}I)} = \frac{P(B\cap X^{-1}(I))}{P(X^{-1}(I))} \\ &\overset{\text{noté}}{=} \int_{\omega\in B} dP_{|X\in I}(\omega) \overset{\text{noté}}{=} \int_{\omega\in B\cap(X\in I)} \frac{dP(\omega)}{P(X\in I)}. \end{aligned} \tag{13.1}$$

Dans le cas discret $\Omega = \{\omega_k : k\in K\}$ où K est au plus dénombrable, on a donc :

$$P(B|X{\in}I) = P_{|X\in I}(B) = \sum_{\omega_k\in B} P_{|X\in I}(\omega_k) = \sum_{\omega_k\in B\cap(X\in I)} \frac{P(\omega_k)}{P(X\in I)}. \tag{13.2}$$

Donc, si $B \in \mathcal{T}$, toujours quand $P(X{\in}I) \neq 0$:

$$P(1_B|X{\in}I) \overset{\text{déf}}{=} P_{|X\in I}(1_B) = \int_{\omega\in\Omega} 1_B(\omega)dP_{|X\in I}(\omega) = \frac{1}{P(X{\in}I)} \int_{\omega\in(X\in I)} 1_B(\omega)dP(\omega),$$

et donc, par linéarité et passage à la limite, pour Y v.a.r. définie sur $(\Omega,\mathcal{T},P)$:

$$P(Y|X{\in}I) \overset{\text{déf}}{=} P_{|X\in I}(Y) = \int_{\omega\in\Omega} Y(\omega)dP_{|X\in I}(\omega) = \frac{1}{P(X{\in}I)} \int_{\omega\in X^{-1}(I)} Y(\omega)dP(\omega). \tag{13.3}$$

Dans le cas discret $\Omega = \{\omega_k : k\in K\}$ où K est au plus dénombrable, toujours quand $P(X{\in}I) \neq 0$:

$$P(Y|X{\in}I) = P_{|X=I}(Y) = \sum_{k\in K} Y(\omega_k)P_{|X=I}(\omega_k) = \frac{1}{P(X{\in}I)} \sum_{\omega_k\in X^{-1}(I)} Y(\omega_k)P(\omega_k). \tag{13.4}$$

13.2 Espérance conditionnelle $E(Y|X{\in}I) = P_{|X\in I}(Y)$

Définition 13.1 Si Y est une v.a.r. sur Ω, si $P(X{\in}I) \neq 0$, l'espérance conditionnelle de Y sachant $X \in I$ est l'espérance de Y relativement à la loi $P_{|X\in I}$:

$$E(Y|X{\in}I) \overset{\text{déf}}{=} P_{|X\in I}(Y) = P(Y|X{\in}I) \overset{\text{noté}}{=} \int_\Omega Y(\omega)dP_{|X\in I}(\omega), \tag{13.5}$$

cf. (13.3). En particulier dans le cas discret $E(Y|X{\in}I)$ est donné par (13.4).

Cas discret : soit $\mathcal{X}_{\neq 0}$ le sous-ensemble de $\mathbb{R}$ défini par :

$$\mathcal{X}_{\neq 0} \overset{\text{déf}}{=} \{x\in X(\Omega) : P_X(x)\neq 0\} \qquad (= \{x\in X(\Omega) : P(X{=}x)\neq 0\}). \tag{13.6}$$

Proposition 13.2 *(Probabilités totales). Pour une probabilité discrète sur $\Omega = \{\omega_k : k\in K\}$ où K est au plus dénombrable :*

$$P(Y) = \sum_{k\in\mathcal{X}_{\neq 0}} P_{|(X=x_k)}(Y)P(X{=}x_k). \tag{13.7}$$

Preuve.

$$P(B) = \sum_{k\in K} P(B\cap(X{=}x_k)) = \sum_{k\in\mathcal{X}_{\neq 0}} P(B\cap(X{=}x_k)) = \sum_{k\in\mathcal{X}_{\neq 0}} P_{|X=x_k}(B)P(X{=}x_k),$$

la deuxième égalité car si $k\notin\mathcal{X}_{\neq 0}$ alors $P(X{=}x_k)=0$ et $B\cap(X{=}x_k)\subset(X{=}x_k)$.

Donc $P(1_B) = \sum_{k\in\mathcal{X}_{\neq 0}} P_{|X=x_k}(1_B)P(X{=}x_k)$, d'où (13.7). ▪

Exemple 13.3 Soit $\Omega=\{a_1,...,a_n\}$ et P une probabilité sur la tribu discrète de Ω.

Soit $X:\Omega\to\{x_1,...,x_k\}$ (une v.a.r. discrète), $k\geq 1$, t.q. $P(X{=}x_i))\neq 0$ pour tout i.

Soit $Y:\Omega\to\{y_1,...,y_m\}$ (une v.a.r. discrète), $m\geq 1$.

Alors, pour $i\in[1,n]_{\mathbb{N}}$:

$$E(Y|X{=}x_i) = P_{|X=x_i}(Y) = \sum_{j=1}^m y_j\, P_{|(X=x_i)}(Y{=}y_j) \qquad (= \sum_{j=1}^m y_j\, \frac{P((Y{=}y_j)\cap(X{=}x_i))}{P(X{=}x_i)}). \quad (13.8)$$

▪

13.3 Loi conditionnelle image $(P_{|X\in I})_Y = P_{Y|X\in I} = P_{|X\in I}\circ Y^{-1}$

La loi image d'une loi de probabilité a été définie en (4.5).

Définition 13.4 Si X et Y sont deux v.a.r. sur Ω, si $I\in\mathcal{T}_{\mathbb{R}}$, si $P(X{\in}I)\neq 0$, alors la loi image de la loi $P_{|X\in I}$ par Y est notée $P_{Y|X\in I}$:

$$(P_{|X\in I})_Y \overset{\text{déf}}{=} P_{|X\in I}\circ Y^{-1} \overset{\text{noté}}{=} P_{Y|X\in I}. \quad (13.9)$$

Définition 13.5 Pour $J\in\mathcal{T}_{\mathbb{R}}$ on note :

$$P_{Y|X\in I}(J) \overset{\text{noté}}{=} P_{Y|X}(J|I) \qquad (= P_{|X\in I}(Y{\in}J) = P(Y{\in}J|X{\in}I)). \quad (13.10)$$

Donc : pour tout $J\in\mathcal{T}_{\mathbb{R}}$:

$$P_{Y|X}(J|I) = \frac{P((Y{\in}J)\cap(X{\in}I))}{P(X{\in}I)}. \quad (13.11)$$

Et, avec la notation intégrale :

$$P_{Y|X}(J|I) = \int_{\omega\in(Y\in J)} dP_{|X\in I}(\omega) = \frac{1}{P(X{\in}I)}\int_{(X\in I)\cap(Y\in J)} dP. \quad (13.12)$$

En particulier on a bien $P_{Y|X\in I}(\mathbb{R})=1$, car $(Y{\in}\mathbb{R})=\Omega$.

Et la proposition de la mesure image donne, pour toute v.a.r. $g:\mathbb{R}\to\mathbb{R}$:

$$P_{Y|X\in I}(g) = \int_{y\in\mathbb{R}} g(y)dP_{Y|X\in I}(y) = \int_{\omega\in\Omega} g(Y(\omega))dP_{|X\in I}(\omega). \quad (13.13)$$

Corollaire 13.6 *L'espérance conditionnelle de Y sachant $X\in I$ vaut, quand $P(X{\in}I)\neq 0$:*

$$E(Y|X{\in}I) = P_{|X\in I}(Y) = P_{Y|X\in I}(id) = \int_{\mathbb{R}} y\, dP_{Y|X\in I}(y). \quad (13.14)$$

Preuve. C'est la proposition de la loi image (13.13) appliquée à (13.5). ▪

Exercice 13.7 Soit $s\in\mathbb{N}^*$. Une urne contient initialement s boules rouges, s boules vertes et $2s$ boules bleues, les boules étant indiscernables au touché. On effectue n tirages successifs selon le protocole suivant :

- si la boule tirée est rouge, on ne la remet pas, mais on la remplace par une boule bleue.
- si la boule tirée est bleue, on ne la remet pas, mais on la remplace par une boule rouge.
- si la boule tirée est verte, on la remet.

On note X_k la v.a.r. qui donne le nombre de boules rouges dans l'urne après le k-ème tirage.

On note R_k l'événement "la k-ième boule tirée est rouge", V_k l'événement "la k-ième boule tirée est verte", B_k l'événement "la k-ième boule tirée est bleue".

1- Déterminer la loi de X_1. Calculer son espérance et sa variance.

2- On suppose $n \geq 2s$. Justifier que $X_n(\Omega) \subset \{0, 1, ..., 3s\}$.

3- Soit $k \in [1, 3s-1]_{\mathbb{N}}$.

31- Quelle est la composition de l'urne quand $(X_n = k)$ vient de se réaliser ? Justifier que les 3 couleurs sont présentes avant d'effectuer le n+1-ième tirage.

32- Justifier $P_{|(X_n=k)}(X_{n+1}=h) = P((X_{n+1}=h)|(X_n=k)) = \begin{cases} = \dfrac{k}{4s} \text{ si } h = k-1, \\ = \dfrac{s}{4s} \text{ si } h = k, \\ = \dfrac{3s-k}{4s} \text{ si } h = k+1, \\ = 0 \text{ sinon }, \end{cases} =$ $P_{X_{n+1}|X_n}(h|k)$ qui donne la loi de probabilité conditionnelle $P_{|(X_n=k)}$.

33- En déduire $E(X_{n+1}|(X_n=k)) = (1 - \dfrac{1}{2s})\ k + \dfrac{3}{4}$. Vérifier que l'égalité est encore vérifiée pour $k = 0$ et $k = 3s$.

34- En déduire $E(X_{n+1}) = (1 - \frac{1}{2s})\ E(X_n) + \frac{3}{4}$.

41- Calculer $E(X_n)$ en fonction de n, s et $E(2s)$, pour $n \geq 2s$.

42- En déduire $\lim_{n\to\infty} E(X_n)$.

Réponse. 1- $P(X_1 = i) = 0$ pour tout $i \neq s, s+1, s-1$ puisqu'après le premier tirage il y a soit $s+1$, soit s, soit $s-1$ boule rouge.

La probabilité de tirer une boule rouge ou une boule verte est $\frac{s}{4s} = \frac{1}{4}$ et la probabilité de tirer une boule bleue est $\frac{2s}{4s} = \frac{1}{2}$. Donc $P(X_1 = s-1) = \frac{1}{4}$ (on a tiré une boule rouge), $P(X_1 = s) = \frac{1}{4}$ (on a tiré une boule verte), $P(X_1 = s+1) = \frac{1}{2}$ (on a tiré une boule bleue).

Donc $E(X_1) = (s-1)\frac{1}{4} + s\frac{1}{4} + (s+1)\frac{1}{2} = s + \frac{1}{4}$.

Et $V(X_1) = E((X_1 - s - \frac{1}{4})^2) = (-\frac{5}{4})^2\frac{1}{4} + (-\frac{1}{4})^2\frac{1}{4} + (\frac{3}{4})^2\frac{1}{2} = \frac{1}{16}\frac{25+1+18}{4} = \frac{11}{16}$.

2- $n \geq 2s$. Si une suite de tirages donne des boules rouges et aucune boule bleue, X_k peut prendre toutes les valeurs dans $[0, s]$. Et on ne peut pas avoir de valeurs de $X_k < 0$. Si une suite de tirages donne des boules bleues et aucune boule rouge, X_k peut prendre toutes les valeurs dans $[2, 3n]$. Si pour un k donné $X_k = 3n$, alors il n'y a plus de boule bleue, et donc X ne peut que décroître en fonction de k. Donc $X_n(\Omega) \subset [0, 3s]$.

31- $k \in [1, 3s-1]_{\mathbb{N}}$. Quel que soit n, au n-ième tirage l'urne contient s boules vertes. Et pour $(X_n = k)$ il y a k boules rouges, donc $3s-k$ boules bleues. Et comme $k \geq 1$, et $k \leq 3s-1$, les trois couleurs sont présentes.

32- $P((X_{n+1}=k-1)|(X_n=k))$: on a tiré une boule rouge alors qu'il y avait k boules rouges sur $4s$: probabilité de tirage $= \frac{k}{4s}$.

$P((X_{n+1}=k)|(X_n=k))$: on a tiré une boule verte alors qu'il y avait s boules vertes sur $4s$: probabilité de tirage $= \frac{s}{4s}$.

$P((X_{n+1}=k+1)|(X_n=k))$: on a tiré une boule bleue alors qu'il y avait $3s-k$ boules bleues sur $4s$: probabilité de tirage $= \frac{3s-k}{4s}$.

Autres cas impossibles.

33- $E(X_{n+1}|(X_n=k)) = (k-1)\dfrac{k}{4s} + k\dfrac{s}{4s} + (k+1)\dfrac{3s-k}{4s} = \dfrac{k^2 - k + ks + 3ks + 3s - k^2 - k}{4s}$.

41- Soit $n \geq 2s$. La suite $(E(X_n))_{n \geq 2s}$ est une suite arithmético-géométrique de point fixe λ donné par $\lambda = (1 - \frac{1}{2s})\ \lambda + \frac{3}{4}$, soit $\lambda = \frac{3s}{2}$. Donc la suite $(E(X_n) - \lambda)_{n \geq 2s}$ est une suite géométrique de raison $(1 - \frac{1}{2s})$, de premier terme $E(X_{2s})$. D'où pour $n \geq 2s$ on a $E(X_n) = (1 - \frac{1}{2s})^n(E(X_{2s}) - \frac{3s}{2}) + \frac{3s}{2}$.

42- Donc $\lim_{n\to\infty} E(X_n) = \frac{3s}{2}$. ▪

13.4 La v.a.r. espérance $E(Y|X)$: cas discret

13.4.1 La v.a.r. espérance $E(Y|X)$ sur l'espace probabilisé $(\mathbb{R}, \mathcal{T}_{\mathbb{R}}, P_X)$

On suppose ici que $X(\Omega)$ est discret (et donc P_X est une probabilité discrète).

On reprend l'ensemble $\mathcal{X}_{\neq 0} = \{x \in X(\Omega) : P_X(x) \neq 0\}$ défini en (13.6).

Définition 13.8 Soit X et Y deux v.a.r : $\Omega \to \mathbb{R}$. La fonction $E(Y|X) : \mathcal{X}_{\neq 0} \to \mathbb{R}$ définie sur $\mathcal{X}_{\neq 0}$, cf. (13.6), par :

$$E(Y|X)(x) \overset{\text{déf}}{=} E(Y|X{=}x) \qquad (= P_{|X=x}(Y)) \tag{13.15}$$

est appelée espérance conditionnelle de Y sachant X (ou par rapport à X).

Donc $E(Y|X)(x)$ est l'espérance de Y pour la probabilité $P_{|X=x}$, soit :

$$E(Y|X)(x) \tag{13.16}$$

, donc :

Proposition 13.9 *L'espérance de la fonction $E(Y|X)$ dans l'espace probabilisé image $(\mathbb{R}, \mathcal{T}_\mathbb{R}, P_X)$ vérifie :*

$$E(E(Y|X)) = E(Y). \tag{13.17}$$

Preuve. La fonction $E(Y|X)$ est définie sur $(\mathcal{X}_{\neq 0}, \mathcal{T}_\mathbb{R}, P_X)$, donc :

$$E(E(Y|X)) = P_X(E(Y|X)) = \sum_{x\in\mathcal{X}_{\neq 0}} E(Y|X)(x)P_X(x) = \sum_{x\in\mathcal{X}_{\neq 0}} E(Y|X{=}x)P(X{=}x) = E(Y),$$

cf. (13.7). ∎

13.4.2 Notations

Pour $X : \Omega \to \mathbb{R}$ et $Y : \Omega \to \mathbb{R}$ des v.a.r., on note $X(\Omega) = (x_i)_{i=1,\dots,n}$ et $Y(\Omega) = (y_j)_{j=1,\dots,m}$ (avec éventuellement n ou $m = \infty$).

On reprend la notation (11.10). Ainsi :

$$P(Y{=}y_j|X{=}x_i) = \frac{P((Y{=}y_j)\cap(X{=}x_i))}{P(X{=}x_i)} = \frac{P_{\vec{X}}(x_i,y_j)}{P_X(x_i)} = \frac{p_{ij}}{p_{i\cdot}} = P_{|X=x_i}(Y{=}y_j) = P_{Y|X=x_i}(y_j).$$

Et donc $P_{Y|X=x_i} = P_{Y|X}(\cdot|x_i)$ est une loi discrète donnée par :

$$P_{Y|X=x_i} = \sum_j \frac{p_{ij}}{p_{i\cdot}}\delta_{y_j} \qquad (= \sum_j P_{Y|X=x_i}(y_j)\delta_{y_j}). \tag{13.18}$$

Et on note :

$$P_{Y|X}(y_j|x_i) \overset{\text{déf}}{=} P_{Y|X=x_i}(y_j) \qquad (= \frac{p_{ij}}{p_{i\cdot}}). \tag{13.19}$$

Et l'espérance conditionnelle de Y sachant $X = x_i$ est le réel :

$$E(Y|X)(x_i) = E(Y|X{=}x_i) = P_{|X=x_i}(Y) = P_{Y|X=x_i}(id) = \sum_j y_j\,\frac{p_{ij}}{p_{i\cdot}}. \tag{13.20}$$

Exemple 13.10 (Loi binomiale.) Si $X = n$ est le nombre de tirages et $Y{=}k$ le nombre de succès, on a $P(Y{=}k|X{=}n) = b(k;n,p) = P_{|X=n}(Y{=}k)$, voir (6.12), et $E(Y|X{=}n) = P_{|X=n}(Y) = np$, cf. (10.22). ∎

Remarque 13.11 Dans le cas où les lois marginales sont indépendantes, on trouve bien sûr $P_{Y|X=x_i} = P_Y$. En effet, notant $P(Y{=}y_j) = p_{\cdot j} = P_Y(y_j)$, on a dans ce cas $p_{ij} = p_{i\cdot}p_{\cdot j}$ et donc $P_{Y|X=x_i}(y_j) = \dfrac{p_{i\cdot}p_{\cdot j}}{p_{i\cdot}} = p_{\cdot j} = P_Y(y_j)$: la loi conditionnelle est donc inconditionnelle (cas des v.a.r. indépendantes). ∎

13.5 Loi conditionnelle : cas continu

Soit P une probabilité continue sur Ω de densité $dP = p(\omega)d\omega$. Soit X une v.a.r. sur Ω, et la probabilité image $P_X = P\circ X^{-1}$ est également supposée être une probabilité de densité :

$$dP_X = p_X(x)\,dx. \tag{13.21}$$

Donc dans ce cas $P_X(x) = P(X{=}x) = 0$, et on n'a pas l'espoir de définir brutalement $E(Y|X{=}x) = P_{|X=x}(Y)$, puisque $\frac{1}{P(X=x)} = \frac{1}{0}$.

On va servir de la loi $P_{\vec{X}} = P_{(X,Y)}$ du couple $\vec{X} = (X,Y)$, dont on suppose que c'est une loi continue :

$$dP_{\vec{X}}(x,y) = p_{\vec{X}}(x,y)\,dxdy. \tag{13.22}$$

et $P_{\vec{X}}(I,J) =^{\text{déf}} P_{\vec{X}}(I\times J) = P(X{\in}I, Y{\in}J) = \int_{(x,y)\in I\times J} p_{\vec{X}}(x,y)\,dxdy$.

13.5.1 Définition de la loi conditionnelle $P_{Y|X=x}$

On a (13.11) ici sous la forme, toujours pour $P_X(I) \neq 0$:

$$\begin{aligned} P_{Y|X\in I}(J) &= \frac{P_{\vec{X}}(I,J)}{P_X(I)} = \frac{\int_{y\in J}\int_{x\in I} p_{\vec{X}}(x,y)\,dxdy}{\int_{x\in I} p_X(x)\,dx} \overset{\text{noté}}{=} P_{Y|X}(J|I) \\ &= \int_{y\in J} \widetilde{p}_{Y|X\in I}(y)\,dy, \qquad \text{où} \qquad \widetilde{p}_{Y|X\in I}(y) = \frac{\int_{x\in I} p_{\vec{X}}(x,y)\,dx}{\int_{x\in I} p_X(x)\,dx}, \end{aligned} \tag{13.23}$$

et donc $P_{Y|X\in I}$ est la loi continue de densité $\widetilde{p}_{Y|X\in I}$. (On a pu appliquer le théorème de Fubini car les densités sont des fonctions positives.)

Pour définir $p_{Y|X=x}$, on remarque que pour $I = [x, x+h[$ et $h << 1$ on a :

$$\widetilde{p}_{Y|X\in[x,x+h[}(y) = \frac{\int_x^{x+h} p_{\vec{X}}(t,y)\,dt}{\int_x^{x+h} p_X(t)\,dt} \simeq \frac{h\,p_{\vec{X}}(x,y)}{h\,p_X(x)} \overset{\text{noté}}{=} p_{Y|X=x}(y).$$

D'où la définition (ce n'est pas un résultat, juste une définition) :

Définition 13.12 Quand $p_X(x) \neq 0$ (la densité p_x est non nulle en x), on définit la loi $P_{Y|X=x}$ comme étant la loi de densité $p_{Y|X=x}$ définie par :

$$p_{Y|X=x}(y) = \frac{p_{\vec{X}}(x,y)}{p_X(x)} = \frac{p_{X,Y}(x,y)}{p_X(x)} \overset{\text{noté}}{=} p_{Y|X}(y|x), \tag{13.24}$$

soit donc $dP_{Y|X=x}(y) = p_{Y|X=x}(y)\,dy$. Donc, pour tout $J \in \mathcal{T}_{\mathbb{R}}$, quand $p_X(x) \neq 0$:

$$P_{Y|X=x}(J) = \int_{y\in J} p_{Y|X=x}(y)\,dy = \frac{1}{p_X(x)} \int_{y\in J} p_{\vec{X}}(x,y)\,dy. \tag{13.25}$$

(Comme $\int_{y\in\mathbb{R}} p(x,y)\,dy = p_X(x)$, on vérifie en particulier que $P_{Y|X=x}(\mathbb{R}) = 1$.)

Remarque 13.13 Dans le cas où $p_{\vec{X}}(x,y) = p_X(x)p_Y(y)$, cas des lois marginales sont indépendantes, on trouve bien sûr $\widetilde{p}_{Y|X\in I} = p_Y = p_{Y|X\in I}$. ▪

13.5.2 Définition de l'espérance conditionnelle $E(Y|X=x)$

Définition 13.14 Y étant une v.a.r. définie sur $(\Omega, \mathcal{T}_\Omega, P)$, l'espérance conditionnelle $E(Y|X=x)$ est l'espérance associée à la probabilité de densité $p_{Y|X=x}$ définie sur $\mathbb{R}$ en (13.24) :

$$E(Y|X=x) = P_{Y|X=x}(id) = \int_{y\in\mathbb{R}} y\,p_{Y|X=x}(y)\,dy, \tag{13.26}$$

cf. (13.25).

(Voir (13.14) : ici on se sert essentiellement du fait que P_X est une mesure de densité, cf. (13.24).)

13.5.3 La fonction espérance conditionnelle $E(Y|X)$

Définition 13.15 C'est la fonction $\mathbb{R} \to \mathbb{R}$ donnée par :

$$E(Y|X)(x) = E(Y|X=x). \tag{13.27}$$

Proposition 13.16 *L'espérance de la v.a.r. $E(Y|X)$ relativement à la probabilité P_X vérifie :*

$$E(E(Y|X)) = E(Y). \tag{13.28}$$

Preuve. Par définition, cf. (10.4), on a, $E(Y|X)$ étant une v.a.r. sur $(\mathbb{R}, \mathcal{T}_{\mathbb{R}}, P_X)$:

$$\begin{aligned} E(E(Y|X)) = P_X(E(Y|X)) &= \int_{x\in\mathbb{R}} E(Y|X)(x)\,p_X(x)\,dx = \int_x \int_{y\in\mathbb{R}} y\,p_{Y|X=x}(y)\,dy\,p_X(x)\,dx \\ &= \int_{y\in\mathbb{R}} y \int_{x\in\mathbb{R}} \frac{p_{\vec{X}}(x,y)}{p_X(x)}\,p_X(x)\,dxdy = \int_{y\in\mathbb{R}} y\,(\int_{x\in\mathbb{R}} p_{\vec{X}}(x,y)\,dx)dy \\ &= \int_{y\in\mathbb{R}} y\,p_Y(y)\,dy = E(Y), \end{aligned}$$

puisque $\int_{x\in\mathbb{R}} p_{\vec{X}}(x,y)\,dx = p_Y(y)$. ▪

13.6 * Loi conditionnelle : cas général

Une difficulté essentielle de la notion de v.a.r. $E(Y|X)$ est : quand $X, Y : \Omega \to \mathbb{R}$ sont deux v.a.r. sur Ω, alors $E(Y|X) : X(\Omega) \subset \mathbb{R} \to \mathbb{R}$ est une v.a.r. sur $\mathbb{R}$ (sur l'espace image $X(\Omega)$). Et donc, pour la définition de $E(Y|X)$ à l'aide de X et Y (définies sur Ω) on va devoir se ramener à l'image réciproque $X^{-1}(\mathbb{R}) \subset \Omega$, plus précisément travailler avec la tribu engendrée par X qui correspond à la contrainte "sachant X". Ce cadre général permet de traiter indifféremment les cas discrets et continus. Ce § peut être omis en première lecture.

13.6.1 Pull-back

Définition 13.17 Soit $X : \Omega \to \mathbb{R}$ une v.a.r., soit $g : \mathbb{R} \to \mathbb{R}$ une fonction définie sur $X(\Omega)$. La fonction $f = g \circ X : \Omega \to \mathbb{R}$ est appelée le pull-back de g par X :

$$f(\omega) \stackrel{\text{déf}}{=} g(x), \quad \forall \omega \in X^{-1}(x) \tag{13.29}$$

Le pull-back, comme son nom l'indique, est la fonction qui ramène à l'espace initial Ω.

13.6.2 Retour sur le cas discret

Cas discret : le pull-back de $g = E(Y|X)$ par X est défini pour les $\omega \in X^{-1}(x)$ quand $x \in \mathcal{X}_{\neq 0}$ par :

$$f(\omega) = E(Y|X)(x), \quad \forall \omega \in X^{-1}(x), \tag{13.30}$$

avec donc :

$$f(\omega) = E(Y|X)(X(\omega)), \quad \forall \omega \in X^{-1}(x), \tag{13.31}$$

soit :

$$f(\omega) = \frac{1}{P(X{=}x)} \int_{(X=x)} Y \, dP, \qquad \forall \omega \in X^{-1}(x). \tag{13.32}$$

Et la proposition de la mesure image P_X de P par X donne, cf. (9.24) :

$$E(E(Y|X)) = \int_{\mathbb{R}} E(Y|X)(x) \, dP_X(x) = \int_{\Omega} f(\omega) \, dP(\omega) = E(f). \tag{13.33}$$

Le calcul de l'espérance de $E(Y|X)$ dans $\mathbb{R}$ a été ramené a un calcul dans Ω.

Ici : $E(E(Y|X))$ est calculé dans l'espace probabilisé $(\mathbb{R}, \mathcal{T}_{\mathbb{R}}, P_X)$, et $E(f)$ est calculé dans l'espace probabilisé $(\Omega, \mathcal{T}, P)$.

Avec (13.32), cas discret, on voit que la mesure $E(f) = P(f)$ ne dépend de X qu'au travers des ensembles $(X{=}x) = X^{-1}(x)$:

si X et $\widetilde{X}$ sont deux v.a.r. engendrant la même tribu $\mathcal{T}(X) = \mathcal{T}(\widetilde{X})$, donc avec les mêmes ensembles $A_i = X^{-1}(x_i) = \widetilde{X}^{-1}(x'_i)$ quand $X(\Omega) = (x_i)_{i=1,\dots,n}$ et $\widetilde{X}(\Omega) = (x'_i)_{i=1,\dots,n}$, alors si $f = E(Y|X) \circ X$ et $f' = E(Y|\widetilde{X}) \circ \widetilde{X}$, on a, pour tout $\omega \in A_i$:

$$f(\omega) = \frac{1}{P(A_i)} \int_{\omega \in A_i} Y(\omega) \, dP(\omega) = \widetilde{f}(\omega),$$

i.e. $f = \widetilde{f}$(les pull-back sont égaux). Autrement dit les v.a.r. X et $\widetilde{X}$ sont indistinguables lors du calcul de $E(f)$: on construit donc naturellement une classe d'équivalence.

13.6.3 Espérance $E(Y|\mathcal{A}) : \Omega \to \mathbb{R}$

Définition 13.18 Soit $A \in \mathcal{T}$ tel que $0 < P(A) < 1$, et soit $\mathcal{T}(A) = \{\emptyset, A, A^C, \Omega\}$ la tribu engendré par A. On appelle espérance conditionnelle de Y sachant A la v.a.r. $\mathcal{A}$-mesurable définie par :

$$E(Y|\mathcal{T}(A))(\omega) = E(Y|A)1_A(\omega) + E(Y|A^C)1_{A^C}(\omega). \tag{13.34}$$

Et si $X = 1_A$, la tribu engendrée par X étant $\mathcal{T}(A)$ on note également :

$$E(Y|\mathcal{T}(1_A)) = E(Y|\mathcal{T}(A)). \tag{13.35}$$

On a donc pour cette tribu simple, si $\omega \in A$ alors :

$$E(Y|\mathcal{T}(A))(\omega) = E(Y|A)(\omega) = \frac{1}{P(A)} \int_{t\in A} Y(t)\,dP(t), \tag{13.36}$$

et si $\omega \in A^C$ alors :

$$E(Y|\mathcal{A})(\omega) = E(Y|A^C)(\omega) = \frac{1}{P(A^C)} \int_{t\in A^C} Y(t)\,dP(t). \tag{13.37}$$

.

On généralise cette définition aux sous-tribus quelconques :

Proposition 13.19 et définition. *Soit $Y : \Omega \to \mathbb{R}$ une v.a.r. (fonction $\mathcal{T}$-mesurable) qui est P-intégrable (ou positive), i.e. $Y \in L^1(\Omega, \mathcal{T}, P)$ (ou $Y \geq 0$).*

Soit $\mathcal{A}$ une sous-tribu de $\mathcal{T}$.

Alors il existe unique une fonction $Z : \Omega \to \mathbb{R}$ qui est $\mathcal{A}$-mesurable et P-intégrable (ou positive) sur la tribu $\mathcal{A}$ telle que :

$$\forall A \in \mathcal{A}, \quad \int_A Y\,dP = \int_A Z\,dP. \tag{13.38}$$

L'égalité (13.38) définie la relation d'équivalence : $Z \sim Y$ ssi, $\forall A \in \mathcal{A}$, $P(Y1_A) = P(Z1_A)$.

Et un représentant de la classe de Y est noté :

$$E(Y|\mathcal{A}), \tag{13.39}$$

et est appelé espérance conditionnelle de Y par rapport à $\mathcal{A}$.

Preuve. 0- Visualisation du problème. Cas $L^2(\Omega, \mathcal{T}, P)$: on se restreint aux v.a.r. dans $L^2(\Omega, \mathcal{T}, P)$: alors (13.38) s'écrit : pour toute fonction $U \in L^2(\Omega, \mathcal{A}, P)$ (une telle fonction est $\mathcal{A}$-mesurable et donc limite de combinaisons linéaires de fonctions indicatrices) on a :

$$\int_\Omega (Y - Z)U\,dP = 0. \tag{13.40}$$

Donc $Y - Z \in L^2(\Omega, \mathcal{A}, P)^\perp$ (dans l'orthogonal à $L^2(\Omega, \mathcal{A}, P)$ dans $L^2(\Omega, \mathcal{T}, P)$. Comme $Z \in L^2(\Omega, \mathcal{A}, P)$, on a : Z est la projection orthogonale de Y sur $L^2(\Omega, \mathcal{A}, P)$. A condition que la projection existe, i.e. que $L^2(\Omega, \mathcal{A}, P)$ soit fermé dans $L^2(\Omega, \mathcal{T}, P)$, ce qui est le cas par "stabilité de la tribu $\mathcal{A}$".

Démonstration de la proposition.

1- Unicité : si Z et Z' vérifient (13.38), alors l'ensemble $A = \{Z < Z'\}$ étant dans la tribu $\mathcal{A}$ on a :

$$\int_{\{Z<Z'\}} Y\,dP = \int_{\{Z<Z'\}} Z\,dP = \int_{\{Z<Z'\}} Z'\,dP,$$

et donc $\int_{\{Z-Z'<0\}}(Z - Z')\,dP = 0$, et donc $(Z - Z')1_{\{Z-Z'<0\}} = 0$ p.s., donc $P(Z - Z' < 0) = 0$ pour la probabilité restreinte à $\mathcal{A}$. De même $P(Z' < Z) = 0$. Donc $Z = Z'$ p.s. sur la tribu $\mathcal{A}$.

2- Existence. Par hypothèse $Y \in L^1(\Omega, \mathcal{T}, P)$. Quitte à poser $Y = Y_+ - Y_-$, supposons $Y \geq 0$. Soit $Y_n = \inf(Y, n)$. Alors on a $Y_n \in L^2$ puisque borné (on a $\int Y_n^2\,dP \leq \int n\,Y_n\,dP \leq n \int Y\,dP$.)

Posons Z_n le Z trouvé au cas 0- relativement à Y_n. La suite (Y_n) est croissante vers Y et $\int_A Y_n\,dP = \int_A Z_n\,dP$ pour tout n et tout $A \in \mathcal{A}$, qui donne $\int_A Z_{n+1} - Z_n\,dP \geq 0$ pour tout n et tout A. Donc la suite (Z_n) est croissante. De plus elle est positive car $Y_n \geq 0$ et donc $0 = \int_{Z_n<0} Y_n\,dP = \int_{Z_n<0} Z_n\,dP$ et donc $Z_n 1_{Z_n<0} = 0$. Et le théorème de convergence monotone indique que la suite croissante et positive Z_n est convergente vers une v.a.r. Z avec $\lim \int_A Z_n\,dP = \int_A \lim Z_n\,dP = \int_A Z\,dP$. ▟

13.6.4 Espérance $E(Y|X)$

Soit $\mathcal{A} = \mathcal{T}(X)$ la tribu engendrée par X : on vient de définir $E(Y|\mathcal{T}(X))$. Plaçons-nous dans l'espace $\mathbb{R} \supset \mathrm{Im}(X)$ sur lequel on dispose de la loi P_X image de P par X.

Définition 13.20 On appelle espérance conditionnelle de Y sachant X (ou relativement à X) un élément de la P_X-classe d'équivalence des fonctions $z : \mathbb{R} \to \mathbb{R}$ qui sont $\mathcal{T}_\mathbb{R}$ mesurables telles que leur pull-back $f = z \circ X$ vérifient $z \circ X \in$ classe de $E(Y|\mathcal{T}(X))$. Et on note $E(Y|X)$ un tel élément z.

Avec (13.38) et $z = E(Y|X)$ on a donc, pour tout $A \in \mathcal{T}(X)$:

$$\int_A (z \circ X)(\omega)\, dP(\omega) = \int_A Y(\omega)\, dP(\omega), \tag{13.41}$$

soit de manière équivalente : pour tout $I \in \mathcal{T}_\mathbb{R}$:

$$\int_I z(x)\, dP_X(x) = \int_{X^{-1}(I)} Y(\omega)\, dP(\omega), \tag{13.42}$$

soit, comme $1_{X^{-1}(I)} = 1_I \circ X$, pour tout $I \in \mathcal{T}_\mathbb{R}$:

$$\int_\mathbb{R} z(x) 1_I(x)\, dP_X(x) = \int_\Omega Y(\omega)(1_I \circ X)(\omega)\, dP(\omega) = \int_{x\in I, y\in\mathbb{R}} y\, dP_{X,Y}(x,y). \tag{13.43}$$

Exemple 13.21 Dans le cas discret, la tribu $\mathcal{T}(X)$ est engendrée par les ensembles discrets $A_i = X^{-1}(x_i)$ pour les $x_i \in X(\Omega)$ (tribu dite atomique, les atomes étant les $A_i = X^{-1}(x_i)$ pour $x_i \in \mathcal{X}_{\neq 0}$). Un élément $Z \in E(Y|\mathcal{T}(X))$ vérifie, pour tout i, pour $z = E(Y|X)$:

$$\int_{X^{-1}(x_i)} Y(\omega)\, dP(\omega) = \int_{X^{-1}(x_i)} Z(\omega)\, dP(\omega) = \int_{x_i} z(x)\, dP_X,$$

avec $P_X = \sum_i P(A_i)\delta_{x_i}$. On retrouve bien, pour tout i :

$$z(x_i) = \frac{1}{P(A_i)} \int_{A_i} Y(\omega)\, dP(\omega) = E(Y|X)(x_i).$$

▪

13.6.5 Loi conditionnelle $P_{Y|X=x}$

Définition 13.22 On dit que Y admet une loi conditionnelle relativement à X (ou sachant X) si pour tout $x \in \mathbb{R}$ il existe une loi de probabilité ν_x sur $\mathbb{R}$ telle que : pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$ mesurable et bornée, de pull-back par Y l'application $\varphi \circ Y$, l'application :

$$z : x \to \int_{y\in\mathbb{R}} \varphi(y)\, d\nu_x(y)$$

est dans la classe de $E(\varphi \circ Y|X)$.

Donc (13.42) donne, pour tout $I \in \mathcal{T}_\mathbb{R}$:

$$\int_{x\in I} \int_{y\in\mathbb{R}} \varphi(y)\, d\nu_x(y)\, dP_X(x) = \int\int_{x\in I, y\in\mathbb{R}} \varphi(y)\, dP_{X,Y}(x,y). \tag{13.44}$$

Exemple 13.23 Cas discret. Vérifions que $\nu_{x_i} = \sum_j \frac{p_{ij}}{p_{i.}} \delta_{y_j} = P_{Y|X=x_i}$ convient. D'une part, comme $P_X = \sum_i p_{i.}\delta_{x_i}$, on a :

$$\int_{x_i} \int_{y\in\mathbb{R}} \varphi(y)\, d\nu_x(y)\, dP_X(x) = p_{i.} \int_{y\in\mathbb{R}} \varphi(y)\, d\nu_{x_i}(y) = p_{i.} \sum_j \frac{p_{ij}}{p_{i.}} \varphi(y_j) = \sum_j p_{ij}\varphi(y_j),$$

et d'autre part on a :

$$\int\int_{x=x_i, y\in\mathbb{R}} \varphi(y)\, dP_{X,Y}(x,y) = \sum_j \varphi(y_j)\, p_{ij},$$

et on a bien égalité. ▪

Exemple 13.24 Cas continu. On vérifie immédiatement que la loi ν_x de densité $\mu_x(y) = \frac{p_{\vec{X}}(x,y)}{p_X(x)} = p_{Y|X=x}$ convient, i.e. vérifie (13.44). ▪

14 Lois des grands nombres

Soit $(\Omega, \mathcal{T}, P)$ un espace probabilisé.

14.1 Convergence en probabilité (convergence stochastique)

Ce paragraphe est de fait un paragraphe sur la mesure et l'intégration, avec pour mesure une mesure de probabilité. Voir polycopié intégral de Lebesgue.

Définition 14.1 Une suite $(Y_n)_{n\in\mathbb{N}^*}$ de v.a.r. est dite converger en probabilité (ou converger stochastiquement) vers la v.a.r. Y ssi :

$$\forall \varepsilon > 0, \quad P(|Y_n - Y| \geq \varepsilon) \underset{n\to\infty}{\longrightarrow} 0, \tag{14.1}$$

i.e., avec la notation intégrale, ssi :

$$\forall \varepsilon > 0, \quad \int_{\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}} dP(\omega) \underset{n\to\infty}{\longrightarrow} 0. \tag{14.2}$$

Et on note :

$$Y_n \underset{n\to\infty}{\overset{P}{\longrightarrow}} Y, \qquad \text{ou encore} \qquad Y_n \overset{P}{\longrightarrow} Y, \tag{14.3}$$

la dernière notation quand il n'y a pas d'ambiguïté.

Exemple 14.2 Soit $(X_i)_{i\in\mathbb{N}^*}$ un processus de Bernoulli de paramètre p, cf. (6.7). Soit $Y_n = \frac{1}{n}\sum_{i=1}^n X_i$ (valeur moyenne des résultats des n premiers tirages). On verra que $Y_n \overset{P}{\longrightarrow} p$ (loi faible des grands nombres). ▪

Remarque 14.3 La convergence en probabilité est importante en statistique : Y étant une v.a.r. inconnue qu'on souhaite estimer à l'aide de v.a.r. Y_n moyenne des résultats d'une "même" expérience, la probabilité $P(|Y_n - Y| > \varepsilon)$ représente la probabilité de se tromper à ε près quand on affirme que Y_n est une approximation de Y. Voir la loi faible des grands nombres. ▪

Proposition 14.4 *Soit $m \in \mathbb{N}^*$. Si $Y_n \to Y$ dans $L^m(\Omega, P)$ (convergence dans L^m), alors $Y_n \overset{P}{\longrightarrow} Y$ (convergence en probabilité).*

En particulier la convergence en moyenne (cas m=1) et la convergence quadratique (cas m=2) impliquent la convergence en probabilité.

Preuve. Hypothèse : $P(|Y_n - Y|^m) = ||Y_n - Y||_m^m = \int_\Omega |Y_n(\omega) - Y(\omega)|^m \, dP(\omega) \underset{n\to\infty}{\longrightarrow} 0$ pour $m \geq 1$. D'où, pour $m \geq 1$ et $\varepsilon > 0$:

$$P(|Y_n - Y| \geq \varepsilon) = P(|Y_n - Y|^m \geq \varepsilon^m) \;\leq\; P(|Y_n - Y|^m \geq 0) \;=\; P(|Y_n - Y|^m) \underset{n\to\infty}{\longrightarrow} 0.$$

▪

Exercice 14.5 Donner un exemple où la réciproque est fausse.

Réponse. Cas $m = 1$ et $L^1([0,1], dx)$, $P = dx$ mesure de Lebesgue sur $[0,1]$; soit $(Y_n)_{\mathbb{N}^*}$ la suite de fonctions $Y_n = n 1_{[0,\frac{1}{n}]}$. Cette suite converge en probabilité vers la fonction nulle, car, avec $\varepsilon > 0$ on a $P(|Y_n - 0| \geq \varepsilon) = P(\{x \in [0,1] : Y_n(x) \geq \varepsilon\}) = P([0, \frac{1}{n}]) = \frac{1}{n}$ dès que n est assez grand.

Mais $\int_0^1 |Y_n(x) - 0| \, dx = n \int_0^{\frac{1}{n}} dx = 1 \not\to 0$: la suite ne converge pas dans $L^1([0,1], dx)$ vers la fonction nulle. (La suite (Y_n) converge à l'extérieur de $L^1([0,1])$ vers la masse de Dirac.) ▪

Exercice 14.6 Montrer que la convergence P-p.p. entraine la convergence en probabilité (alors que la convergence P-p.p. n'entraine pas la convergence en norme L^1).

Réponse. Le lemme de Fatou (cours d'intégration) indique, pour une suite (f_n) de fonctions positives mesurables :

$$\mu(\liminf_{n\to\infty} f_n) \le \liminf_{n\to\infty}(\mu(f_n)). \tag{14.4}$$

Soit $(Y_n)_\mathbb{N}$ une suite de v.a.r. qui converge P-p.p. vers Y. Soit $\varepsilon > 0$. Soit $A_n = (|Y_n - Y| < \varepsilon)$. On pose $f_n = 1_{A_n}$. Comme $P(A_n) = P(1_{A_n})$, On obtient :

$$P\Big(\liminf_{n\to\infty}(|Y_n - Y| < \varepsilon)\Big) \le \liminf_{n\to\infty}\Big(P(|Y_n - Y| < \varepsilon)\Big). \tag{14.5}$$

Comme $(Y_n)_\mathbb{N}$ converge p.p. vers Y, il existe B t.q. $P(B) = 0$ et pour tout $\omega \in B^C$ on a $\lim_{n\to\infty} |Y_n(\omega) - Y(\omega)| = 0$, donc il existe $N \in \mathbb{N}$ t.q. pour tout $n \ge N$ on a $|Y_n(\omega) - Y(\omega)| < \varepsilon$. Donc $P(|Y_n - Y| < \varepsilon) = P(B^C) = 1$. Donc $\liminf_{n\to\infty}\Big(P(|Y_n - Y| < \varepsilon)\Big) = 1$. Donc $\liminf_{n\to\infty}\Big(P(|Y_n - Y| \ge \varepsilon)\Big) = 0$. ▪

Exercice 14.7 Montrer que la réciproque est fausse : la convergence en probabilité n'entraine pas la convergence P-p.p. D'ailleurs on peut avoir la convergence en probabilité tout en ayant la divergence en tout point.

Réponse. Soit $\Omega = [0, 1[$, soit $dP(\omega) = d\omega$ (mesure de Lebesgue). Soit $n \in \mathbb{N}^*$ et soit k l'unique entier tel que $2^k \le n < 2^{k+1}$ (soit $k \le \log_2(n) < k+1$ et k est la partie entière de $\log_2(n)$). Puis soit $j = n - 2^k \in [0, 2^k[$. Posons :

$$Y_n(t) = 1_{[\frac{j}{2^k}, \frac{j+1}{2^k}[}(t).$$

Alors $Y_n \to 0$ en probabilité, car pour les $0 < \varepsilon < 1$ on a

$P(|Y_n - 0| > \varepsilon) = P(|Y_n| > \varepsilon) = \int_{\frac{j}{2^k}}^{\frac{j+1}{2^k}} d\omega = \frac{1}{2^k} = \frac{2}{2^{k+1}} \le \frac{2}{n}$.

Par contre la suite (Y_n) ne converge en aucun point de $[0, 1]$. En effet, si elle converge au point t, ayant $Y_n(t) = 1$ ou 0, on a $\lim_{n\to\infty} Y_n(t) = 1$ ou 0, et elle est nécessairement stationnaire : à partir d'un certain rang N, on a, soit pour tout $n > N$ $Y_n(t) = 1$, soit pour tout $n > N$ $Y_n(t) = 0$. Or :

1- si $Y_n(t) = 1$, alors $Y_{n+1}(t) = 0$. En effet, pour $n < 2^k$, on a soit $n+1 < 2^k$ et les intervalles $[\frac{n-2^k}{2^k}, \frac{n+1-2^k}{2^k}[$ et $[\frac{n+1-2^k}{2^k}, \frac{n+2-2^k}{2^k}[$ sont disjoints, soit $n+1 = 2^k$ et les intervalles $[\frac{n-2^k}{2^k}, \frac{n+1-2^k}{2^k}[$ et $[0, \frac{1}{2^{k+1}}[$ sont disjoints.

2- Si $Y_n(t) = 0$, on a $t \notin [\frac{j}{2^k}, \frac{j+1}{2^k}[= [\frac{n-2^k}{2^k}, \frac{n+1-2^k}{2^k}[$. Ayant $t \in [0, 1[$, soit l'unique entier $a \in \{0, 1, ..., 2^{k+1}-1\}$ tel que $t \in [\frac{a}{2^{k+1}}, \frac{a+1}{2^{k+1}}[$. Posons $b = a + 2^{k+1} \in \mathbb{N}$, avec donc $t \in [\frac{b-2^{k+1}}{2^k}, \frac{b+1-2^{k+1}}{2^k}[$: on a $Y_b(t) = 1$, avec $b > n$.

Donc la suite $(Y_n(t))$ n'est pas stationnaire : il n'y a convergence ponctuelle en aucun point. ▪

Exercice 14.8 Montrer : si $Y_n \underset{n\to\infty}{\longrightarrow} Y$ en probabilité, alors il existe une sous-suite extraite $(Y_{n_k})_{k\in\mathbb{N}^*}$ telle que $Y_{n_k} \underset{k\to\infty}{\longrightarrow} Y$ P-p.p.

Réponse. Soit $k \in \mathbb{N}$ fixé. On a $P(|Y_n - Y| > \frac{1}{k}) \underset{n\to\infty}{\longrightarrow} 0$, i.e. :

$\forall \varepsilon > 0$, $\exists N_{k,\varepsilon} \in \mathbb{N}$, $\forall n \ge N_{k,\varepsilon}$, $P(|Y_n - Y| > \frac{1}{k}) < \varepsilon$.

En particulier c'est vrai avec $\varepsilon = \frac{1}{2^k}$. Soit $n_k = n(k) = N_{k,\varepsilon}$. Alors la suite $(Y_{n_k})_{k_i n\mathbb{N}^*}$ est telle que la série $\sum_{k=1}^{\infty} P(|Y_{n_k} - Y| > \frac{1}{k})$ est convergente (car majorée par $\sum_k \frac{1}{2^k} < \infty$).

Soit alors $Z = \sum_{k=1}^{\infty} |Y_{n_k} - Y|$ (v.a.r.). On a $P(Z) = P(\sum_{k=1}^{\infty} |Y_{n_k} - Y|) = \sum_{k=1}^{\infty} P(|Y_{n_k} - Y| > \frac{1}{k}) < \infty$ (passage à la limite sous la mesure à l'aide du théorème de convergence monotone de Beppo-Lévi, ici pour la suite $(Z_n)_{\mathbb{N}^*}$ où $Z_n = \sum_{k=1}^{n} |Y_{n_k} - Y|$, la suite $(Z_n)_{\mathbb{N}^*}$ étant une suite croissante de fonctions positives, qui donne $P(\lim_\infty(Z_n)) = \lim_\infty(P(Z_n))$, soit $P(\sum_{k=1}^{\infty} |Y_{n_k} - Y|) = \lim_\infty(\sum_{k=1}^{n} P(|Y_{n_k} - Y|))$).

Donc $Z \in L^1(\Omega, P)$ (on a $Z \ge 0$ et $P(|Z|) = P(Z) < \infty$). Donc pour presque tout $\omega \in \Omega$ on a $Z(\omega) < \infty$ (sinon on prend A non négligeable tel que $Z = \infty$ sur A et $P(Z) = \infty$, absurde).

Donc $|Y_{n_k} - Y|(\omega) \underset{k\to\infty}{\longrightarrow} 0$ pour presque tout ω. ▪

Exercice 14.9 (Suite de l'exercice 14.5.) Montrer que dans L^1 la réciproque de la proposition 14.4 est "presque vraie" sous l'hypothèse de domination :

Si $Y_n \to Y$ en probabilité et si $\exists Z \in L^1(\Omega, P)$, $\forall n \in \mathbb{N}^*$, $|Y_n| < Z$, alors on peut extraire une sous-suite qui converge vers Y dans L^1.

Réponse. Résultat de l'exercice 14.8 : comme $Y_n \to Y$ en probabilité, on peut extraire une sous-suite (Y_{n_k}) qui converge p.p. vers Y.

On applique alors le théorème de convergence dominée : la sous-suite (Y_{n_k}) étant de plus dominée, elle converge vers Y dans L^1 : $P(|Y_{n_k} - Y|) \underset{k\to\infty}{\longrightarrow} 0$. ▪

Exercice 14.10 Soit $(Y_n)_{\mathbb{N}^*}$ une suite de v.a.r. à deux états 0 et 1 : $\mathrm{Im}(Y_n) = \{0,1\}$pour tout n. On suppose les Y_n indépendantes et chaque Y_n suit une loi de Bernoulli de paramètre p_n. Montrer pour la convergence en probabilité :

$$Y_n \xrightarrow{P} 0 \qquad \Longleftrightarrow \qquad p_n \underset{n\to\infty}{\longrightarrow} 0, \tag{14.6}$$

et pour la convergence presque surement (P-presque partout) :

$$Y_n \underset{n\to\infty}{\longrightarrow} 0 \quad \text{p.s.} \qquad \Longleftrightarrow \qquad \sum_{n=1}^{\infty} p_n < \infty. \tag{14.7}$$

(La différence entre les deux types de convergence est flagrante.)

Réponse. $P(|Y_n - 0| \geq \varepsilon) = P(|Y_n| \geq \varepsilon) = P(|Y_n| = 1) = p_n$, d'où (14.6).

Pour (14.7) : pour ω fixé, $Y_n(\omega) \in \{0,1\}$ et $Y_n(\omega) \longrightarrow_{n\to\infty} 0$ implique que $Y_n = 0$ à partir d'un certain n (la suite (Y_n) est stationnaire à partir d'un certain n). Donc, avec (3.40) :

$$\begin{aligned}\{\omega : Y_n(\omega) \underset{n\to\infty}{\longrightarrow} 0\} &= \{\omega : Y_n(\omega) = 0 \text{ à partir d'un certain } n\} \\ &= \bigcup_{n\in\mathbb{N}^*} \Big(\bigcap_{k\geq n} \{\omega : Y_k(\omega) = 0\}\Big) = \liminf A_n = (\limsup A_n^C)^C,\end{aligned}$$

où $A_k = \{\omega : Y_k(\omega) = 0\}$, avec $P(A_k) = 1 - p_k$ et $P(A_c^C) = p_k$.

D'où (14.7) $\Leftarrow$: le lemme de Borel–Cantelli (théorème 3.60) indique que, cf. (3.43) : si $\sum_n P(A_n^C) < \infty$, alors $P(\limsup A_n^C) = 0$. Donc $B = \limsup A_n^C$ est P-négligeable, avec $B = (\{\omega : Y_n(\omega) \longrightarrow_{n\to\infty} 0\})^C$: l'ensemble des points ω t.q. $Y_n(\omega)$ ne converge pas vers 0 est négligeable.

Et (14.7) $\Rightarrow$: on applique le lemme de Borel–Cantelli (théorème 3.60), avec (3.44) : les A_n étant indépendants, si $\sum_n P(A_n^C) = \infty$ alors $P(\limsup A_n^C) = 1$. Donc $P((\limsup A_n^C)^C) = 0 = P(\liminf A_n)$. Donc $C = \{\omega : Y_n(\omega) \longrightarrow_{n\to\infty} 0\}$ est P-négligeable : l'ensemble des points ω t.q. $Y_n(\omega)$ converge vers 0 est négligeable. Donc si $Y_n \longrightarrow_{n\to\infty} 0$ presque surement, alors $\sum_n P(A_n^C) < \infty$. ▪

14.2 Loi faible des grands nombres et théorème de Bernoulli

La convergence en probabilité a été définie en (14.1).

Théorème 14.11 Loi faible des grands nombres. *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r. indépendantes dans $L^2(\Omega, P)$, les X_n ayant toutes la même moyenne et le même écart type :*

$$\forall n \in \mathbb{N}^*, \qquad E(X_n) = m, \qquad \sigma(X_n) = \sigma. \tag{14.8}$$

Pour $n \in \mathbb{N}^*$, soit :

$$Y_n = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + \ldots + X_n}{n}, \tag{14.9}$$

la moyenne partielle. Alors :

$$E(Y_n) = m, \tag{14.10}$$

et :

$$\sigma(Y_n) = \frac{\sigma}{\sqrt{n}}. \tag{14.11}$$

D'où :

$$||Y_n - m||^2_{L^2(\Omega,P)} \underset{n\to\infty}{\longrightarrow} 0 \qquad \text{et} \qquad Y_n \underset{n\to\infty}{\overset{P}{\longrightarrow}} m \tag{14.12}$$

Ainsi, Y_n converge en probabilité vers la v.a.r. constante $m1_\Omega$, car $\sigma(Y_n) = \frac{\sigma}{\sqrt{n}} \longrightarrow_{n\to\infty} 0$ (donc Y_n est de "moins en moins aléatoire").

N.B. : ainsi l'écart type de la moyenne de n termes indépendants n'est pas la moyenne des écarts types (n'est pas constant) : il décroît comme $\frac{1}{\sqrt{n}}$. Dans ce cas de v.a.r. indépendantes, les valeurs positives et négatives des $X_k - \overline{X_k}$, dont les valeurs absolues au carré sont mesurées par σ^2 (mesure de la dispersion), se "compensent partiellement" dans Y_n, pour n "grand".

Preuve. $E(Y_n) = \dfrac{E(X_1)+...+E(X_n)}{n} = \dfrac{nm}{n} = m$ (linéarité de E). D'où :

$$\begin{aligned}\mathrm{Var}(Y_n) &= \int (Y_n - m)^2\, dP = \int (\frac{X_1-m}{n} + ... + \frac{X_n-m}{n})^2\, dP \\ &= \sum_{i=1}^n \mathrm{Var}(\frac{X_i-m}{n}) + 2\sum_{i<j} \mathrm{cov}(\frac{X_i-m}{n}, \frac{X_j-m}{n}).\end{aligned}$$

Les v.a.r. X_i étant indépendantes, il en est de même des $\frac{X_i-m}{n}$, cf. (5.11), et donc les covariances sont nulles. Et(10.45) donne, pour $i \in \mathbb{N}^*$:

$$\mathrm{Var}(\frac{X_i-m}{n}) = \frac{1}{n^2}\mathrm{Var}(X_i) = \frac{1}{n^2}\sigma^2.$$

D'où $\mathrm{Var}(Y_n) = \sum_{i=1}^n \frac{1}{n^2}\sigma^2 = \frac{1}{n}\sigma^2$, d'où (14.11). D'où $\mathrm{Var}(Y_n) \underset{n\to\infty}{\longrightarrow} 0$, soit $||Y_n - m1_\Omega||^2_{L^2(\Omega,P)} \underset{n\to\infty}{\longrightarrow} 0$, i.e. $(14.12)_1$, donc $Y_n \underset{n\to\infty}{\longrightarrow} m1_\Omega$ dans $L^2(\Omega, P)$. D'où (Y_n) converge en probabilité vers $m1_\Omega$ cf. proposition 14.4. ∎

Théorème 14.12 (Bernoulli). *(Corollaire.) Soit $(X_n)_{\mathbb{N}^*}$ une suite de v.a.r. indépendantes de même loi de Bernoulli de paramètre p $(= (E(X_n))$. Alors, pour $\varepsilon > 0$:*

$$P(|Y_n - p| \ge \varepsilon) \le \frac{1}{4n\varepsilon^2}. \tag{14.13}$$

Preuve. La formule de Bienaymé–Tchébytcheff (10.61) donne, avec (14.10) et (14.11) :

$$P(|Y_n - p| \ge \varepsilon) \le \frac{E(|Y_n - p|^2)}{\varepsilon^2} = \frac{\sigma(Y_n)^2}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Et pour la loi de Bernoulli, cf. (10.48), $\sigma^2 = pq = p(1-p) \le \frac{1}{4}$. ∎

Exemple 14.13 Lancers d'une pièce de côté a = pile et b = face. Soit g la fonction gain donnée par $g(a) = \alpha$ (gain si a) et $g(b) = \beta$ (gain si b). On note $P_g(\alpha) = p$ et $P_g(\beta) = 1-p$. Gain moyen $m = E(g) = \alpha P_g(\alpha) + \beta P_g(\beta) = \alpha p + \beta(1-p)$.

Lancers de cette pièce une infinité de fois. Soit $\Omega = \{a,b\}^{\mathbb{N}^*}$. Soit X_n la v.a.r. qui donne le gain du tirage n : $X_n : \vec{\omega} = (\omega_1, \omega_2, ...) \in \Omega \to X_n(\vec{\omega}) = f(\omega_n) \in \{\alpha, \beta\}$, avec $P_{X_n}(\alpha) = P_g(\alpha) = p$ et $P_{X_n}(\beta) = P_g(\beta) = 1-p$. Alors $(Y_n)_{\mathbb{N}^*} = (\frac{\sum_{k=1}^n X_k}{n})_{\mathbb{N}^*}$ tend vers la v.a.r. constante $m1_\Omega$, où $m = E(X_1) = \alpha P_g(\alpha) + \beta P_g(\beta)$: c'est le gain qu'on peut espérer après de très nombreux tirages. Par exemple si a = pile donne $\alpha = 10$ et b = face donne $\beta = 0$ alors $m = 10 * p$, gain moyen qu'on peut espérer après n lancers, pour n "grand". Et pour $\varepsilon = 10^{-2}$, l'erreur en probabilité sera $P(|Y_n - m| \ge \varepsilon) \le \frac{1}{10}$ pour $n \ge \frac{10^5}{4} = 25000$, cf. (14.13) (on rappelle que l'inégalité de Bienaymé–Tchébytcheff est grossière). ∎

14.3 Loi forte des grands nombres

La loi forte des grands nombres est une loi de convergence "en moyenne quadratique" et presque sûrement (presque partout au sens de la mesure P).

Théorème 14.14 *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r. de carré P-intégrable (suite dans $L^2(\Omega, P)$). On suppose :*

1- $\exists m \in \mathbb{R}$, $\forall i \in \mathbb{N}^*$, $E(X_i) = m$ (toutes les v.a.r. ont même espérance),

2- $\exists C > 0$, $\forall i \in \mathbb{N}^*$, $\mathrm{Var}(X_i) \le C$ (les variances sont uniformément bornées), et

3- $\mathrm{cov}(X_i, X_j) = 0$ pour tout i, j (en particulier vrai si les X_i sont indépendantes).

Alors les moyennes partielles $Y_n = \frac{\sum_{i=1}^n X_i}{n}$, comme en (14.9), vérifient $E(Y_n) = m$ et convergent à la fois en moyenne quadratique et au sens presque sûrement (convergence ponctuelle) vers la fonction constante $m1_\Omega$:

$$E(Y_n) = m, \qquad \int |Y_n - m|^2\, dP \underset{n\to\infty}{\longrightarrow} 0, \qquad \text{et} \qquad Y_n \underset{n\to\infty}{\longrightarrow} m \quad P\text{-p.s.}.$$

N.B. : en particulier les hypothèses sont vérifiées quand les X_i sont indépendantes, et ont toutes même espérance (finie) et même variance (finie) (hypothèses de la loi faible des grands nombres).

Preuve. Supposons $m=0$, ce qui revient à considéré X_i-m au lieu de X_i, ce qui ne modifie pas les hypothèses, cf. (12.13), mais allège les écritures.

Avec (12.25), on a, avec l'hypothèse $\mathrm{cov}(X_i,X_j)=0$ pour tout $i\neq j$:

$$\mathrm{Var}(\frac{X_1}{n}+...+\frac{X_n}{n})=\sum_{i=1}^n \mathrm{Var}(\frac{X_i}{n})+0=\sum_{i=1}^n \frac{1}{n^2}\,\mathrm{Var}(X_i),$$

et donc, avec l'hypothèse 2- :

$$0\le \mathrm{Var}(Y_n)\le \sum_{i=1}^n \frac{1}{n^2}\,C\le \frac{C}{n}\underset{n\to\infty}{\longrightarrow} 0.$$

D'où la convergence en moyenne quadratique vers $m=0$.

Montrons $Y_n^2\underset{n\to\infty}{\longrightarrow} 0$ P-p.p., ce qui montrera que $Y_n\underset{n\to\infty}{\longrightarrow} 0$ P-p.p..

Considérons la suite extraite $(Z_k)_{k\in\mathbb{N}^*}=(Y_{k^2})_{k\in\mathbb{N}^*}$. On a :

$$\sum_{k=1}^\infty \mathrm{Var}(Z_k)=\sum_{k=1}^\infty \mathrm{Var}(Y_{k^2})\le \sum_{k=1}^\infty \frac{C}{k^2}=C\frac{\pi^2}{6}<\infty,$$

et donc la série $\sum_{k=1}^\infty \mathrm{Var}(Z_k)$ converge dans $\mathbb{R}$.

Et $E(Z_n)=0$ pour tout n car $E(X_k)=0$ pour tout k, donc $\mathrm{Var}(Z_k)=E(Z_k^2)=P(Z_k^2)=\int_\Omega Z_k^2(\omega)\,dP$. D'où, l'inversion des signes $\sum$ et $\int$ étant vérifiée pour les suites croissantes de fonctions positives (théorème de convergence monotone de Beppo-Lévi), ici avec la suite des fonctions $S_N=\sum_{k=1}^N Z_k^2$, on a :

$$\lim_{N\to\infty}(\int_\Omega S_N\,dP)=\int_\Omega(\lim_{N\to\infty} S_N)\,dP.$$

On obtient :

$$\infty>C\frac{\pi^2}{6}\ge\sum_{k=1}^\infty \mathrm{Var}(Z_k)=\sum_{k=1}^\infty(\int_\Omega Z_k(\omega)^2\,dP)=\int_\Omega(\sum_{k=1}^\infty Z_k(\omega)^2)\,dP,$$

D'où la série de fonctions $\sum_{k=1}^\infty Z_k^2$ est bornée P-p.p., voir lemme suivant 14.15. D'où $Z_k^2\longrightarrow 0$ P-p.p. sur Ω, d'où $Z_k\longrightarrow 0$ P-p.p. sur Ω.

Retour à (Y_n). Soit n donné. Et soit $\ell_n=\max\{k\in\mathbb{N}^*:k^2\le n\}$. On a :

$$Y_n-Z_{\ell_n}=Y_n-Y_{\ell_n^2}=\sum_{i=1}^n\frac{X_i}{n}-\sum_{i=1}^{\ell_n^2}\frac{X_i}{\ell_n^2}=\underbrace{(\sum_{i=1}^n\frac{X_i}{n}-\sum_{i=1}^{\ell_n^2}\frac{X_i}{n})}_{T_n}+\underbrace{(\sum_{i=1}^{\ell_n^2}\frac{X_i}{n}-\sum_{i=1}^{\ell_n^2}\frac{X_i}{\ell_n^2})}_{U_n}.$$

Pour le deuxième terme U_n de la somme on a :

$$|U_n|=|\sum_{i=1}^{\ell_n^2}\frac{X_i}{n}-\sum_{i=1}^{\ell_n^2}\frac{X_i}{\ell_n^2}|=|\sum_{i=1}^{\ell_n^2}\frac{X_i}{\ell_n^2}\frac{\ell_n^2}{n}-\sum_{i=1}^{\ell_n^2}\frac{X_i}{\ell_n^2}|\le|\sum_{i=1}^{\ell_n^2}\frac{X_i}{\ell_n^2}|\,(\frac{n-\ell_n^2}{n})=|Z_{\ell_n}|\,\frac{n-\ell_n^2}{n}.$$

Comme la suite extraite $(Z_k)_{k\in\mathbb{N}^*}$ converge P-p.p. vers 0, on déduit $U_n\longrightarrow_{n\to\infty} 0$ P-p.p..

Pour le premier terme $T_n=\displaystyle\sum_{\ell_n^2+1}^n\frac{X_i}{n}$ de la somme, de moyenne nulle, on a, avec l'hypothèse 3- :

$$\mathrm{Var}(T_n)=\frac{1}{n^2}\sum_{i=\ell_n^2+1}^n \mathrm{Var}(X_i)\le\frac{C}{n^2}(n-\ell_n^2)\le C(\frac{2}{n^{\frac32}}+\frac{1}{n^2})=O(n^{-3/2}),$$

car $(\ell_n+1)^2>n$ donne $\ell_n^2+2\ell_n+1>n$, donc $n-\ell_n^2<2\ell_n+1\le 2\sqrt{n}+1$.

Donc la série $\sum_{n=1}^\infty \mathrm{Var}(T_n)$ converge dans $\mathbb{R}$, avec (Fubini pour les fonctions positives) :

$$\sum_{n=1}^\infty \mathrm{Var}(T_n) = \sum_{n=1}^\infty \int_\Omega T_n^2\,dP = \int_\Omega (\sum_{n=1}^\infty T_n^2)\,dP.$$

Donc la fonction $\sum_{n=1}^\infty T_n^2$ est finie P-p.p., cf. lemme 14.15, i.e. $\sum_{n=1}^\infty T_n^2(\omega) < \infty$ pour P-p.p. ω.

D'où $T_n^2 \longrightarrow 0$ P-p.p. sur Ω, d'où $T_n \longrightarrow_{n\to\infty} 0$ P-p.p. sur Ω.

Donc, $|Y_n - Y_{\ell_n^2}| \le |T_n| + |U_n| \underset{n\to\infty}{\longrightarrow} 0$ P-p.p.. Avec $Y_{\ell_n^2} \underset{n\to\infty}{\longrightarrow} Z_{\ell_n}$ P-p.p.. Donc $Y_n \underset{n\to\infty}{\longrightarrow} 0$ P-p.p.. ▪

Lemme 14.15 *Toute fonction intégrable est finie presque partout : si $(\Omega, \mathcal{T}_\Omega, \mu)$ est un espace mesuré, si $f \in \mathcal{F}(\Omega;\mathbb{R})$ est mesurable et μ-intégrable (i.e. $\mu(|f|) < \infty$) alors f est finie μ-presque partout :*

$$\mu(\{f = \infty\}) = 0, \tag{14.14}$$

i.e., si $Z = \{\omega : f(\omega) = \infty\}$, alors $\mu(Z) = 0$, i.e., pour tout $\omega \in \Omega - Z$ on a $f(\omega) < \infty$.

Preuve. Supposons $f \ge 0$. Soit $n \in \mathbb{N}^*$ et $A_n = \{f \ge n\}$ $(= \{\omega \in \Omega : f(\omega) \ge n\} = f^{-1}([n,\infty[))$.

On a $1_{A_n} \le \frac{1}{n} f$, car vrai si $\omega \in A_n$ et vrai si $\omega \in A_n^C$ (on a supposé $f \ge 0$).

Donc $\mu(A_n) \le \frac{1}{n}\mu(f) \underset{n\to\infty}{\longrightarrow} 0$. Comme (A_n) est une suite décroissante et $\mu(A_1) \le \mu(f) < \infty$, on a $\lim_{n\to\infty}(\mu(A_n)) = \mu(\lim_{n\to\infty} A_n)$, donc $\mu(\lim_{n\to\infty} A_n) = 0$, donc $\mu(\{f = \infty\}) = 0$.

Et si $f \not\ge 0$, on pose $f_+ = \sup(0, f) \ge 0$, d'où $\mu(\{f_+ = \infty\}) = 0$. Idem on pose $f_- = \sup(0, -f) \ge 0$, d'où $\mu(\{f_- = \infty\}) = 0$. Et $f = f_+ - f_-$, d'où $\mu(\{f = \infty\}) = 0$. ▪

Remarque 14.16 On garde la convergence p.p. sans supposer l'existence de la constante C (hypothèse 2.), à condition de supposer que les v.a.r. sont indépendantes (toujours toutes de moyenne m). Voir par exemple Métivier [8]. ▪

Remarque 14.17 Soit $(X_i)_{i=1,\dots,n}$ une suite de v.a.r. indépendantes toutes la même loi, loi qu'on ne connaît pas. On peut cependant estimer la valeur moyenne $m = E(X_i)$ à l'aide de "l'algorithme stochastique" suivant, calculer à l'issue du n+1-ième tirage :

$$Y_{n+1} = Y_n + \frac{1}{n+1}(X_{n+1} - Y_n)$$

(puisque $(n+1)Y_{n+1} = \sum_{i=1}^{n+1} X_i$ et $(n+1)Y_n + X_{n+1} - Y_n = nY_n + X_{n+1} = \sum_{i=1}^{n+1} X_i$).

On a : (Y_n) tend vers la fonction constante $= m$ (loi forte des grands nombres). ▪

14.4 Théorème de la limite centrale (convergence vers la loi de Gauss)

Théorème Central Limite (TCL).

14.4.1 Convergence étroite et convergence en loi

(La convergence étroite est la convergence des distributions, voir cours de distributions, les mesures étant les distributions d'ordre 0.)

Soit U un ouvert dans $\mathbb{R}$ (même démarche dans $\mathbb{R}^n$), et soit $C_c^0(U;\mathbb{R})$ l'ensemble des fonctions continues à support compact dans U.

Définition 14.18 On se place dans un espace mesurable $(U, \mathcal{T})$. Soit μ une mesure bornée, et soit $(\mu_n)_\mathbb{N}$ une suite de mesures bornées. La suite $(\mu_n)_\mathbb{N}$ converge étroitement vers la mesure μ ssi :

$$\forall f \in C_c^0(U;\mathbb{R}), \quad \mu_n(f) \underset{n\to\infty}{\longrightarrow} \mu(f), \qquad \text{noté} \qquad \mu_n \underset{n\to\infty}{\rightharpoonup} \mu, \tag{14.15}$$

appelée également convergence simple (convergence en chaque point f), ou convergence faible.

Autre notation : par définition, $(\mu_n)_\mathbb{N}$ converge étroitement vers la mesure μ ssi :

$$\forall f \in C_c^0(U;\mathbb{R}), \quad \lim_{n\to\infty} \int_U f\,d\mu_n = \int_U f\,d\mu \qquad (= \int_U f \lim_{n\to\infty} d\mu_n), \tag{14.16}$$

i.e. on peut passer à la limite sous le signe $\int$.

Exemple 14.19 La probabilité de loi normale $P_n = \mathcal{N}(0,\sigma_n^2)$ converge étroitement vers $P = \mathcal{N}(0,\sigma^2)$ quand la suite $(\sigma_n)_{\mathbb{N}}$ converge vers σ (appliquer le théorème de convergence dominée). ▟

Définition 14.20 Quand la suite $(\mu_n{=}P_{X_n})_{\mathbb{N}}$ est une suite de lois de v.a.r.X_n et que la limite $\mu{=}P_X$ est la loi d'une v.a.r. X, on dit que (X_n) converge en loi vers la loi de X :

$$P_{X_n} \underset{n\to\infty}{\rightharpoonup} P_X \quad \overset{\text{déf}}{\Longleftrightarrow} \quad (X_n) \text{ converge en loi vers la loi de } X. \tag{14.17}$$

Donc (X_n) converge en loi vers la loi de X ssi :

$$\forall f \in C_c^0(U;\mathbb{R}), \quad P_{X_n}(f) \underset{n\to\infty}{\longrightarrow} P_X(f), \tag{14.18}$$

soit :

$$\forall f \in C_c^0(U;\mathbb{R}), \quad \lim_{n\to\infty} \int_{x\in U} f(x)\, dP_{X_n}(x) = \int_{x\in U} f(x)\, dP_X(x). \tag{14.19}$$

14.4.2 Théorème de la limite centrale (convergence vers la loi de Gauss)

Soit $(\Omega,\mathcal{T},P)$ un espace probabilisé. On se restreint au cas simple $L^2(\Omega,P)$ (fonction de carré P-intégrable).

Théorème 14.21 *(Cas simple.) Soit $(X_k)_{k\in\mathbb{N}^*} \in L^2(\Omega,P)^{\mathbb{N}^*}$ une suite de v.a.r. i.i.d. (indépendantes identiquement distribuées), i.e. les X_k sont de carré intégrable, sont indépendantes et ont toutes la même loi $P_{X_k} = P_{X_1}$ pour tout $k \in \mathbb{N}^*$.*

On note $m = E(X_1) = E(X_k)$ et $\sigma = \sigma(X_1) = \sigma(X_k)$, valeurs indépendantes de k.

On note $S_n = \sum_{i=1}^n X_i$ la somme, $Y_n = \frac{S_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$ la moyenne, comme en (14.9), et :

$$Z_n = \frac{Y_n - m}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - nm}{\sqrt{n}\,\sigma}. \tag{14.20}$$

la v.a.r. centrée réduite associée. Alors :

$$P_{Z_n} \underset{n\to\infty}{\rightharpoonup} \mathcal{N}(1,0) \qquad (\textit{convergence en loi}), \tag{14.21}$$

i.e. la loi converge étroitement vers la loi de Gauss centrée réduite $\mathcal{N}(1,0)$ de densité $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

Ainsi $P_{Y_n} \sim \mathcal{N}(m,\frac{\sigma^2}{n})$ et $P_{S_n} \sim \mathcal{N}(nm, n\sigma^2)$ quand $n \to \infty$.

La démonstration est donnée au paragraphe suivant 14.5.

14.5 Démonstration du théorème de la limite centrale

L'ingrédient essentielle est la transformée de Fourier des mesures (transformée de Fourier des distributions d'ordre 0) que les probabilistes appellent fonction caractéristique.

Pourquoi passer par la transformée de Fourier ?

Parce que la transformée de Fourier de toute mesure de probabilité est une mesure de densité, voir proposition suivante 14.33, ce qui simplifie.

Et que la gaussienne centrée réduite est conservée par Fourier (à une constante multiplicative près). Résultat essentiel pour la suite : c'est "le point central" du "théorème de la limite centrale".

Les résultats :

1- l'exponentielle $\xi \to e^{ix\xi}$ transforme "la somme $\xi_1 + ... + \xi_n$ en le produit $e^{ix\xi_1}...e^{ix\xi_n}$", et

2- par Fourier un produit de convolution se transforme en produit simple,

permettent de démontrer le théorème central limite : convergence vers le centre = la gaussienne centrée réduite.

On rappelle que, dans le cas de v.a.r. indépendantes, la loi $P_{X_1+...+X_n}$ est la loi image de $P_{X_1} \otimes ... \otimes P_{X_n}$ par l'application somme $\varphi(x_1,...,x_n) = x_1 + ... + x_n$, et qu'alors $P_{X_1+...+X_n} = P_{X_1} * ... * P_{X_n}$ (a un sens car les mesures P_{X_i} sont bornées). Et donc que pour toute fonction f intégrable on a :

$$\int_{\mathbb{R}} f(y)\,(dP_{X_1} * ... * P_{X_n})(y) = \int_{\mathbb{R}^n} f(x_1 + ... + x_n)\, dP_{X_1}(x_1)...dP_{X_n}(x_n).$$

14.5.1 Transformée de Fourier d'une fonction

Soit $f \in L^1(\mathbb{R};\mathbb{C}) =^{\text{noté}} L^1(\mathbb{R})$. On prend ici comme définition : la transformée de Fourier de f est la fonction $\mathcal{F}(f) \overset{\text{noté}}{=} \widehat{f} : \mathbb{R} \to \mathbb{C}$ donnée par :

$$\mathcal{F}(f)(\xi) \overset{\text{noté}}{=} \widehat{f}(\xi) \overset{\text{déf}}{=} \int_{x\in\mathbb{R}} e^{-i\xi x} f(x)\,dx \overset{\text{noté}}{=} \langle e^{-i\xi\cdot}, f\rangle \quad \in \mathbb{C}. \tag{14.22}$$

(Intégrale qui dépend du paramètre $\xi \in \mathbb{R}$.)

(N.B. : Même si $f : \mathbb{R} \to \mathbb{R}$ est à valeurs réelles, sa transformée de Fourier $\widehat{f} : \mathbb{R} \to \mathbb{C}$ est à valeurs complexes puisque $e^{-iy} = \cos(y) - i\sin(y)$: on ne peut pas se contenter de travailler dans $\mathbb{R}$.)

Exemple 14.22 Exemple fondamental de la gaussienne centrée réduite :

$$p_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad\Longrightarrow\quad \mathcal{F}(p_{0,1})(\xi) = e^{-\frac{\xi^2}{2}} \overset{\text{noté}}{=} \widehat{p_{0,1}}(\xi), \tag{14.23}$$

voir plus loin (14.32) : la gaussienne centrée réduite est conservé à la constante $\sqrt{2\pi}$ près (la masse n'est pas conservée : la mesure de probabilité da masse 1 est transformée en une mesure finie de masse $\sqrt{2\pi}$). Cette gaussienne réduite est le "point central" du théorème central limite. ▪

Remarque 14.23 Si on prend la définition $\mathcal{F}_e(f)(\nu) = \int_{t\in\mathbb{R}} e^{-2i\pi\nu t} f(t)\,dt$ de la transformée de Fourier (celle du traitement du signal), la gaussienne conservée par $\mathcal{F}_e$ est $t \to e^{-\pi t^2}$. Et les modifications à apporter pour la démonstration du théorème central limite sont simples. ▪

Proposition 14.24 *Si $f \in L^1(\mathbb{R};\mathbb{C})$ alors $\widehat{f}$ est bornée et continue.*
Et la transformée de Fourier $\mathcal{F} : L^1(\mathbb{R};\mathbb{C}) \to L^\infty(\mathbb{R};\mathbb{C})$ est linéaire.

Preuve. $|e^{i\theta}| = 1$ donne $|\widehat{f}(\xi)| \le \int_{x\in\mathbb{R}} |f(x)|\,dx = ||f||_{L^1} < \infty$ pour tout $\xi \in \mathbb{R}$: $\widehat{f}$ est bornée.

L'intégrant $(x,\xi) \to e^{-i\xi x} f(x)$ est continu en ξ et borné par $|f(x)|$ (indépendamment de ξ) avec $f \in L^1(\mathbb{R})$, donc $\widehat{f}$ est continue (théorème de convergence dominée).

La linéarité est immédiate. ▪

On rappelle que $xf =^{\text{déf}} \mathit{id}f$ où id est la fonction identité et x le nom imposé de la variable : $xf =^{\text{déf}} \mathit{id}f : x \to (xf)(x) =^{\text{déf}} (\mathit{id}f)(x) = xf(x)$.

Lemme 14.25 *1- Si $f \in L^1$ et de plus $xf \in L^1$ alors $\widehat{f} \in C^1$ et, pour tout $\xi \in \mathbb{R}$:*

$$(\widehat{f})'(\xi) = -i\widehat{(xf)}(\xi). \tag{14.24}$$

2- Si $f \in L^1$ vérifie $f' \in L^1$ alors, pour tout $\xi \in \mathbb{R}$:

$$\widehat{f'}(\xi) = i\xi\widehat{f}(\xi). \tag{14.25}$$

3- Si $f \in L^1$, si $m \in \mathbb{R}$, si $\tau_m f$ est la translaté de m, i.e. $\tau_m f(x) = f(x-m)$, alors :

$$\widehat{\tau_m f}(\xi) = e^{-i\xi m}\widehat{f}(\xi). \tag{14.26}$$

Preuve. (14.24) est obtenue par dérivation de (14.22) sous $\int$ (théorème de convergence dominée car $xf \in L^1$).

(14.25) est obtenue par intégration par parties :

$$\int_{x\in\mathbb{R}} e^{-i\xi x} f'(x)\,dx = -\int_{x\in\mathbb{R}} (-i\xi)\,e^{-i\xi x} f(x)\,dx + [e^{-i\xi x} f(x)]_{-\infty}^{\infty},$$

le terme de bord s'annulant car f <u>et</u> f' dans L^1 donnent f s'annule en $\pm\infty$ (voir transformées de Fourier, cours de distributions).

Puis $\widehat{\tau_m f}(\xi) = \int_{x\in\mathbb{R}} e^{-i\xi x} f(x-m)\,dx = \int_{y\in\mathbb{R}} f(y) e^{-i\xi(y+m)}\,dy$, d'où (14.26). ▪

14.5.2 Transformée de Fourier inverse

La transformée de Fourier inverse est donnée par, pour $h \in L^1(\mathbb{R};\mathbb{C})$ telle que $\widehat{h} \in L^1(\mathbb{R};\mathbb{C})$:

$$\mathcal{F}^{-1}(h)(x) \overset{\text{déf}}{=} \frac{1}{2\pi} \int_{\xi \in \mathbb{R}} e^{+i\xi x} h(\xi)\, d\xi \overset{\text{noté}}{=} \langle \frac{e^{+ix\cdot}}{2\pi}, h \rangle_{d\xi} \quad \in \mathbb{C}, \tag{14.27}$$

cf. cours de distribution.

14.5.3 Transformée de Fourier d'une gaussienne : une gaussienne

On rappelle qu'une gaussienne est une fonction g de type, pour $a > 0$ et $c \neq 0$:

$$g(x) = c\, e^{-ax^2}. \tag{14.28}$$

Proposition 14.26 *La transformée de Fourier d'une gaussienne est une gaussienne :*

$$\widehat{e^{-ax^2}}(\xi) = \sqrt{\frac{\pi}{a}}\, e^{-\frac{\xi^2}{4a}}, \qquad \widehat{e^{-\frac{i^2}{4a}}}(\xi) = 2\sqrt{a\pi}\, e^{-a\xi^2}, \qquad \widehat{\widehat{e^{-ax^2}}}(\xi) = 2\pi e^{-a\xi^2}. \tag{14.29}$$

En particulier la gaussienne $g_{\frac{1}{2}}(x) = e^{-\frac{x^2}{2}}$ est conservée par Fourier (c'est un point central) au facteur multiplicatif $\sqrt{2\pi}$ près :

$$e^{-\frac{x^2}{2}} \quad \overset{\mathcal{F}}{\Longrightarrow} \quad \widehat{e^{-\frac{x^2}{2}}}(\xi) = \sqrt{2\pi}\, e^{-\frac{\xi^2}{2}}. \tag{14.30}$$

Et pour, pour $\sigma > 0$:

$$e^{-\frac{x^2}{2\sigma^2}} \quad \overset{\mathcal{F}}{\Longrightarrow} \quad \widehat{e^{-\frac{x^2}{2\sigma^2}}}(\xi) = \sqrt{2\pi}\,\sigma\, e^{-\frac{\sigma^2\xi^2}{2}}. \tag{14.31}$$

Et plus généralement, pour $m \in \mathbb{R}$ et $\sigma > 0$:

$$e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \overset{\mathcal{F}}{\Longrightarrow} \quad \widehat{e^{-\frac{(x-m)^2}{2\sigma^2}}}(\xi) = e^{-im\xi}\sqrt{2\pi}\,\sigma\, e^{-\frac{\sigma^2\xi^2}{2}}. \tag{14.32}$$

Preuve. Soit $g(x) = e^{-ax^2}$ (la transformée de Fourier est linéaire : on prend $c = 1$ pour alléger). On dérive : $g'(x) = -2ax\, e^{-ax^2}$, soit, pour tout $x \in \mathbb{R}$:

$$g'(x) + 2ax\, g(x) = 0. \tag{14.33}$$

D'où par Fourier (qui est une transformation linéaire) : $\widehat{(g')}(\xi) + 2a\widehat{(xg)}(\xi) = 0$, d'où, avec (14.24) et (14.25), $i\,\xi\,\widehat{g}(\xi) + 2a\frac{1}{-i}\,(\widehat{g})'(\xi) = 0$, soit, pour tout $\xi \in \mathbb{R}$:

$$(\widehat{g})'(\xi) + \frac{1}{2a}\,\xi\,\widehat{g}(\xi) = 0, \tag{14.34}$$

équation similaire à (14.33), où a été remplacé par $\frac{1}{4a}$. D'où $\widehat{g}(\xi) = ke^{-\frac{\xi^2}{4a}}$ est solution de (14.34) pour toute constante k (existence et unicité à l'aide du théorème de Cauchy–Lipschitz). Et $k = \widehat{g}(0) = \int_{x\in\mathbb{R}} g(x)e^{-i0x}\,dx = \int_{x\in\mathbb{R}} e^{-ax^2}\,dx = \sqrt{\frac{\pi}{a}}$, cf. (1.33). D'où (14.29). D'où (14.30) avec $a = \frac{1}{2}$.

Puis avec $a = \frac{1}{2\sigma^2}$ on a $\sqrt{\frac{\pi}{a}} = \sqrt{2\pi}\,\sigma$ et $\frac{1}{4a} = \frac{\sigma^2}{2}$, d'où (14.31).

Puis (14.26) donne (14.32). ▪

14.5.4 Transformée de Fourier de mesures bornées et notations

On renvoie au cours de distributions et aux transformées de Fourier de distributions tempérées, les mesures étant des distributions d'ordre 0 (donc d'ordre fini, donc tempérées).

On note $\mathcal{D}(\mathbb{R})$ l'ensemble des fonctions $C^\infty(\mathbb{R})$ à support compact.

On rappelle : si $f : \mathbb{R} \to \mathbb{C}$ est dans $L^1(\mathbb{R})$, on note $T_f : \mathcal{D}(\mathbb{R}) \to \mathbb{C}$, dite distribution régulière associée à f, la distribution définie par, pour tout $\varphi \in \mathcal{D}(\mathbb{R})$:

$$T_f(\varphi) \overset{\text{déf}}{=} \int_{\mathbb{R}} f(x)\varphi(x)\,dx \overset{\text{noté}}{=} \langle T_f, \varphi\rangle \overset{\text{noté}}{=} \langle f, \varphi\rangle, \tag{14.35}$$

la notation $\langle .,. \rangle$ du crochet de dualité étant une notation usuelle de la linéarité. Attention à la notation abusive $T_f = f$: elle est à l'origine de beaucoup d'incompréhensions. Le contexte doit levé les ambiguïtés.

Insistons : la difficulté essentielle est dans les notations, essentiellement la notation
$T_f =^{\text{noté}} f$ "au sens des distributions",
identifiant une distribution régulière $T_f : \mathcal{D}(\mathbb{R}) \to \mathbb{R}$ (qui n'est pas une fonction $\mathbb{R} \to \mathbb{R}$) avec $f : \mathbb{R} \to \mathbb{R}$ (qui est une fonction $\mathbb{R} \to \mathbb{R}$).

Ainsi la fonction caractéristique d'une (mesure de) probabilité P sera la fonction régulière Φ_X correspondant à la distribution régulière $\widehat{P} = T_{\Phi_X}$ (transformée de Fourier de P), et on notera abusivement $T_{\Phi_X} = \Phi_X$, bien que T_{Φ_X} soit une mesure alors que Φ_X est la densité de la mesure...

On note $\mathcal{S}$ l'ensemble (de Schwartz) des fonctions $C^\infty(\mathbb{R})$ à décroissance rapide ("plus rapide que toute fraction rationnelle", et on dit abusivement "fonction à décroissance exponentielle"). On a besoin de $\mathcal{S}$ car une gaussienne $x \to e^{-ax^2}$ n'est pas dans $\mathcal{D}(\mathbb{R})$ (n'est pas à support compact), mais est bien dans $\mathcal{S}$ (elle est C^∞ à décroissance rapide).

En particulier si $\varphi \in \mathcal{S}$ alors $\varphi \in L^p(\mathbb{R})$ pour tout $p \in [1, \infty]$.

Et les distributions tempérées sont les distributions éléments de $\mathcal{S}'$ le dual de S (voir cours de distributions).

Proposition 14.27 *Si $\varphi \in \mathcal{S}$ (est à décroissance rapide) alors $\widehat{\varphi} \in \mathcal{S}$ (est à décroissance rapide).*

Preuve. Voir cours de distributions. ▪

Cette proposition permet de poser :

Définition 14.28 Si $T \in \mathcal{S}'$ (est une distribution tempérée sur $\mathbb{R}$), sa transformée de Fourier est la distribution $\widehat{T} \in \mathcal{S}'$ définie par, pour tout $\varphi \in \mathcal{S}$:

$$\widehat{T}(\varphi) \overset{\text{déf}}{=} T(\widehat{\varphi}), \tag{14.36}$$

ce qui s'écrit également avec le crochet de dualité, pour tout $\varphi \in \mathcal{S}$:

$$\langle \widehat{T}, \varphi\rangle \overset{\text{déf}}{=} \langle T, \widehat{\varphi}\rangle. \tag{14.37}$$

C'est en particulier le cas si $T = \mu$ est une mesure.

On vérifie que $\widehat{T}$ définie par (14.36) est une distribution tempérée, voir cours de distributions.

Exemple 14.29 Si $\mu = \delta_0$ (mesure de Dirac en 0) alors, pour tout $\varphi \in \mathcal{S}$:

$$\langle \widehat{\delta_0}, \varphi\rangle \overset{\text{déf}}{=} \langle \delta_0, \widehat{\varphi}\rangle = \widehat{\varphi}(0) = \int_{\mathbb{R}} \varphi(x)\,dx = \int_{\mathbb{R}} \varphi(x) 1_{\mathbb{R}}(x)\,dx = \langle T_{1_{\mathbb{R}}}, \varphi\rangle. \tag{14.38}$$

À retenir :
1- δ_0 n'est ni une fonction ni une distribution régulière,
2- la transformée de Fourier $\widehat{\delta_0}$ de δ_0 est une distribution régulière :

$$\widehat{\delta_0} = T_{1_{\mathbb{R}}} \quad \in \mathcal{D}'(\mathbb{R}). \tag{14.39}$$

Donc $\widehat{\delta_0}$ est la mesure de densité $1_{\mathbb{R}}$ définie sur $\mathcal{S}$, cf. l'intégrale dans (14.38). Et on note abusivement :

$$\widehat{\delta_0} \overset{\text{noté}}{=} 1_{\mathbb{R}} \qquad \text{(fonction constante)}, \tag{14.40}$$

cf. (14.35), et on dit abusivement que $\widehat{\delta_0}$ est une fonction, alors que c'est la distribution $T_{1_{\mathbb{R}}}$ mesure de densité donnée par $d\widehat{\delta_0}(x) = 1_{\mathbb{R}}(x)\,dx = dx$ mesure de Lebesgue (définie sur $\mathcal{S}$). ▪

Exemple 14.30 Si $\mu = \delta_a$ (mesure de Dirac en a) on a, pour tout $\varphi \in \mathcal{S}$:

$$\langle \widehat{\delta_a}, \varphi \rangle \overset{\text{déf}}{=} \langle \delta_a, \widehat{\varphi} \rangle = \widehat{\varphi}(a) = \int_{\mathbb{R}} \varphi(x) e^{-iax} \, dx \overset{\text{noté}}{=} \langle e^{-ia\cdot}, \varphi \rangle \overset{\text{noté}}{=} \langle e^{-iax}, \varphi(x) \rangle_{dx}, \tag{14.41}$$

et donc :

$$\widehat{\delta_a} = T_{e^{-ia\cdot}} \tag{14.42}$$

est la mesure de densité $x \to e^{-iax}$, cf. l'intégrale dans (14.41) : $d\widehat{\delta_a}(x) = e^{-iax} dx$ (définie sur $\mathcal{S}$). Et on note abusivement :

$$\widehat{\delta_a}(x) \overset{\text{noté}}{=} e^{-iax}, \tag{14.43}$$

au sens (14.41). ▪

Proposition 14.31 *Si $f \in L^1(\mathbb{R}, \mathbb{C})$ alors pour tout $\varphi \in \mathcal{S}$:*

$$\int_{\mathbb{R}} f(x) \widehat{\varphi}(x) \, dx = \int_{\mathbb{R}} \widehat{f}(\xi) \varphi(\xi) \, d\xi, \tag{14.44}$$

soit, pour les distributions régulières associées, pour tout $\varphi \in \mathcal{S}$:

$$\langle \widehat{T_f}, \varphi \rangle = \langle T_{\widehat{f}}, \varphi \rangle, \qquad \text{i.e.} \qquad \widehat{T_f} = T_{\widehat{f}}. \tag{14.45}$$

Et on note abusivement $T_f = f$ et $\widehat{T_f} = \widehat{f}$, i.e. on note abusivement (14.44) comme :

$$\langle \widehat{f}, \varphi \rangle = \langle f, \widehat{\varphi} \rangle. \tag{14.46}$$

Preuve. Comme $f \in L^1$ on a $\widehat{f} \in L^\infty$ et toutes les quantités suivantes ont un sens : T_f est une distribution tempérée (cours de distribution) et pour $\varphi \in \mathcal{S}$:

$$\begin{aligned}
\langle \widehat{T_f}, \varphi \rangle \overset{\text{déf}}{=} \langle T_f, \widehat{\varphi} \rangle \overset{\text{déf}}{=} & \int_\xi f(\xi) \, \widehat{\varphi}(\xi) \, d\xi = \int_\xi f(\xi) \int_x \varphi(x) \, e^{-ix\xi} \, dx \, d\xi \\
= & \int_x \Big(\int_\xi f(\xi) \, e^{-ix\xi} \, d\xi \Big) \, \varphi(x) \, dx = \int_x \widehat{f}(x) \, \varphi(x) \, dx \\
= & \langle T_{\widehat{f}}, \varphi \rangle \overset{\text{noté}}{=} \langle \widehat{f}, \varphi \rangle,
\end{aligned}$$

à l'aide du théorème de Fubini : en effet la "fonction intégrant" $g(x,\xi) = \varphi(x) f(\xi) \, e^{-ix\xi}$ est dans $L^1(\mathbb{R}^2)$ car $|g(x,\xi)| \leq |\varphi(x)||f(\xi)|$ (fonction à variables séparées) et φ et f sont dans $L^1(\mathbb{R})$. ▪

Exemple 14.32 Soit $\mu = P$ mesure de probabilité de densité gaussienne $p(x) = \sqrt{\frac{a}{\pi}} e^{-ax^2}$, cf. (1.33), $a > 0$, i.e. :

$$P(f) = \sqrt{\frac{a}{\pi}} \int_{x \in \mathbb{R}} f(x) \, e^{-ax^2} \, dx \qquad (\overset{\text{noté}}{=} \langle P, f \rangle), \tag{14.47}$$

pour $f \in \mathcal{S}$. Alors $\widehat{\mu} = \widehat{P}$ est la mesure de densité gaussienne $\widehat{p}(\xi) = \sqrt{\frac{a}{\pi}} \sqrt{\frac{\pi}{a}} e^{-\frac{\xi^2}{4a}} = e^{-\frac{\xi^2}{4a}}$, cf. (14.29) : pour $h \in \mathcal{S}$:

$$\widehat{P}(h) = \int_{x \in \mathbb{R}} h(\xi) \, e^{-\frac{\xi^2}{4a}} \, d\xi \qquad (\overset{\text{noté}}{=} \langle \widehat{P}, h \rangle). \tag{14.48}$$

Cependant, avec (1.33) :

$$P(\mathbb{R}) = 1, \qquad \text{alors que} \qquad \widehat{P}(\mathbb{R}) = \int_{\mathbb{R}} \widehat{p}(\xi) \, d\xi = 2\sqrt{a\pi}. \tag{14.49}$$

Donc $\widehat{p}$ n'est pas une densité de probabilité pour $a \neq \frac{1}{4\pi}$: la mesure $\widehat{P}$ n'est pas une probabilité : c'est juste une mesure bornée. ▪

14.5.5 La transformée de Fourier d'une probabilité est une mesure de densité

Proposition 14.33 *Soit μ une mesure bornée sur $\mathbb{R}$ (c'est en particulier le cas d'une probabilité). Alors la fonction $p : \mathbb{R} \to \mathbb{R}$ définie par :*

$$p(\xi) \stackrel{\text{déf}}{=} \mu(e^{-i\xi\cdot}) \stackrel{\text{noté}}{=} \int_{x\in\mathbb{R}} e^{-ix\xi}\, d\mu(x) \qquad (= \langle \mu_x, e^{-ix\xi}\rangle_{dx}). \tag{14.50}$$

est une fonction bornée, continue, et même uniformément continue.

Et $\widehat{\mu}$ est la mesure de densité p (la distribution régulière de fonction p) :

$$\widehat{\mu} = T_p, \qquad \text{soit} \qquad d\widehat{\mu}(\xi) = p(\xi)d\xi, \tag{14.51}$$

soit, pour toute fonction $\varphi \in \mathcal{S}$:

$$\widehat{\mu}(\varphi) = \int_{\xi\in\mathbb{R}} \varphi(\xi)\, p(\xi)\, d\xi \qquad (= \langle \widehat{\mu}, \varphi\rangle). \tag{14.52}$$

Et on note abusivement ("identification" de la distribution régulière et de sa fonction associée) :

$$\widehat{\mu}(\xi) \stackrel{\text{noté}}{=} p(\xi) \qquad (\text{au sens } d\widehat{\mu}(\xi) = p(\xi)\, d\xi \text{ des distributions}). \tag{14.53}$$

Preuve. À ξ fixé, la fonction $f : x \to f(x) = e^{-ix\xi}$ est mesurable et bornée (par 1), donc μ étant bornée, $\mu(f)$ est bien définie. Donc $p(\xi)$ est bien défini pour tout ξ. (De plus $|p(\xi)| \leq \int_{x\in\mathbb{R}} |e^{-ix\xi}|\, d\mu(x) = \mu(\mathbb{R}) < \infty$, et donc p est une fonction bornée).

Relativement à la mesure μ, $p(\xi)$ est une "intégrale" dépendant du paramètre ξ d'intégrant $e^{-ix\xi}$ qui est continu en ξ et est borné indépendamment de ξ par $1_\mathbb{R}$ qui est μ-intégrable (car μ est une mesure bornée). Donc p est continu (théorème de convergence dominée).

On a $|1 - e^{ia}|^2 = (1 - e^{ia})(1 - e^{-ia}) = 2 - 2\cos a$. La fonction $\beta(a) = 2 - 2\cos a - a^2$ étant négative sur $\mathbb{R}$ (faire le tableau de variation), on obtient $|1 - e^{ia}| \leq \min(\sqrt{2}, |a|)$. Donc $|e^{-ix\xi} - e^{-iy\xi}| = |e^{-ix\xi}(1 - e^{i\xi(x-y)})| \leq \min(\sqrt{2}, |\xi(x-y)|)$, et donc, pour $|x - y| \leq \eta$:

$$|p(x) - p(y)| \leq |\int_{\xi\in\mathbb{R}} \frac{1}{\sqrt{2\pi}}(e^{-ix\xi} - e^{-iy\xi})\, d\mu(\xi)| \leq \frac{1}{\sqrt{2\pi}} \int_{\xi\in\mathbb{R}} \min(\sqrt{2}, |\xi|\eta)\, d\mu(\xi).$$

L'intégrant $(\xi, \eta) \to \min(\sqrt{2}, |\xi|\eta)$ est borné par $\sqrt{2}$ indépendamment de η, donc le théorème de convergence dominée donne $\displaystyle\int_{\xi\in\mathbb{R}} \min(\sqrt{2}, |\xi|\eta)\, d\mu(\xi) \underset{\eta\to 0}{\longrightarrow} \int_{\xi\in\mathbb{R}} \min(\sqrt{2}, 0)\, d\mu(\xi) = 0$. Donc, pour tout $\varepsilon > 0$, il existe $\eta > 0$ t.q. $|x-y| \leq \eta$ entraine $|p(x)-p(y)| \leq \varepsilon$: la fonction p est uniformément continue.

Notation : quand y est la variable, et T une distribution tempérée, pour $\psi \in \mathcal{S}$ on note $T(\psi) = \langle T, \psi\rangle =^{\text{noté}} \langle T(y), \psi(y)\rangle_{dy}$ (pour s'y retrouver dans les noms des variables). Ainsi, pour $\varphi \in \mathcal{D}(\mathbb{R})$:

$$\begin{aligned}\langle \widehat{\mu}(\xi), \varphi(\xi)\rangle_{d\xi} &\stackrel{\text{déf}}{=} \langle \mu(x), \widehat{\varphi}(x)\rangle_{dx} = \int_{x\in\mathbb{R}} \widehat{\varphi}(x)\, d\mu(x) = \int_{x\in\mathbb{R}} (\int_{\xi\in\mathbb{R}} \varphi(\xi) e^{-ix\xi}\, d\xi)\, d\mu(x) \\ &= \int_{\xi\in\mathbb{R}} (\int_{x\in\mathbb{R}} e^{-ix\xi}\, d\mu(x))\varphi(\xi)\, d\xi = \int_{\xi\in\mathbb{R}} p(\xi)\varphi(\xi)\, d\xi = \langle T_p, \varphi\rangle,\end{aligned}$$

car on peut appliquer Fubini : l'intégrant est $g : (x, \xi) \to \frac{1}{\sqrt{2\pi}}\varphi(x)e^{-ix\xi}$ relativement à la mesure $dx \otimes \mu$, avec $|g(x, \xi)| \leq \frac{1}{\sqrt{2\pi}}|\varphi(x)|\,|1_\mathbb{R}(\xi)|$ qui est bien dans $L^1(\mathbb{R}^2, dx \otimes \mu)$ car $\varphi \in \mathcal{D}(\mathbb{R})$ et μ mesure bornée. D'où $\widehat{\mu} = T_p$, i.e. (14.51). ∎

Exemple 14.34 Si μ est une mesure de probabilité discrète $\mu = \displaystyle\sum_{k=1}^n p_k \delta_{x_k}$, avec (14.42) on a :

$$\widehat{\mu}(\xi) = \sum_{k=1}^n p_k\, e^{-i\xi x_k} \quad \text{au sens } d\widehat{\mu}(\xi) = \textstyle\sum_{k=1}^n p_k\, e^{-i\xi x_k}\, d\xi \text{ des distributions.} \tag{14.54}$$

On retrouve $\widehat{\mu}(f) = \displaystyle\int_{\xi\in\mathbb{R}} f(\xi) \sum_{k=1}^n p_k\, e^{-i\xi x_k}\, d\xi = \sum_{k=1}^n p_k \widehat{f}(x_k) = \mu(\widehat{f})$ (définition de $\widehat{\mu}$). ∎

Exercice 14.35 Soit $(\mu_n)_{\mathbb{N}^*}$ et μ des distributions tempérées. Montrer que :

$$(\mu_n)_{\mathbb{N}^*} \text{ converge étroitement vers } \mu \Longleftrightarrow \widehat{\mu_n} \text{ converge étroitement vers } \widehat{\mu}. \tag{14.55}$$

En déduire le théorème de Lévy : pour $(P_n)_{\mathbb{N}^*}$ et P des mesures de probabilités.

$$(P_n)_{\mathbb{N}^*} \text{ converge étroitement vers } P \Longleftrightarrow \widehat{P_n} \text{ converge simplement vers } \widehat{P}, \tag{14.56}$$

au sens, quand $d\widehat{P_n}(\xi) = p_n(\xi)d\xi$ et $d\widehat{P}(\xi) = p(\xi)d\xi$, on a $p_n \underset{n\to\infty}{\longrightarrow} p$ (convergence simple).

Réponse. (14.55) : $\Rightarrow$: $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ est un isomorphisme, cf. cours de distributions. Donc $\langle \mu_n, \widehat{\varphi}\rangle \underset{n\to\infty}{\longrightarrow} \langle \mu, \widehat{\varphi}\rangle$ (hypothèse de convergence étroite) donne $\langle \widehat{\mu}_n, \varphi\rangle \underset{n\to\infty}{\longrightarrow} \langle \widehat{\mu}, \varphi\rangle$, pour tout $\varphi \in \mathcal{S}$, donc $(\widehat{\mu}_n)_{\mathbb{N}^*}$ converge étroitement vers $\widehat{\mu}$.

$\Leftarrow$: $\langle \widehat{\widehat{T}}, \varphi\rangle = \langle \widehat{T}, \widehat{\varphi}\rangle = \langle T, \widehat{\widehat{\varphi}}\rangle = \langle T, \check{\varphi}\rangle$ où $\check{\varphi}(x) = \varphi(-x)$ pour $\varphi \in \mathcal{S}$: cf. cours de distribution (et transformée de Fourier inverse : $\widehat{\widehat{\varphi}} = \check{\varphi}$). Et on applique $\Rightarrow$.

(14.56) : $\Rightarrow$: $(P_n)_{\mathbb{N}^*}$ et P mesures bornées, donc admettent une transformée de Fourier mesure de densité continue, cf. (14.52), $d\widehat{P_n}(\xi) = p_n(\xi)\, d\xi$ et $d\widehat{P}(\xi) = p(\xi)\, d\xi$, donc hypothèse $\int_{\mathbb{R}} \varphi(\xi)(p_n(\xi) - p(\xi))\, d\xi \underset{n\to\infty}{\longrightarrow} 0$ pour tout $\varphi \in \mathcal{S}$. Donc : $p_n \underset{n\to\infty}{\longrightarrow} p$ p.p. pour la mesure de Lebesgue. Comme p et p_n sont continues, la convergence est ponctuelle.

$\Leftarrow$: on utilise l'inverse de Fourier pour les distributions. ∎

14.5.6 La fonction caractéristique Φ_X

Quand la mesure bornée est la loi de probabilité P_X d'une v.a.r. X, la transformée de Fourier $\widehat{P_X}$ est une mesure de densité, cf. proposition 14.33, notée Φ_X, donnée par, pour ξ fixé dans $\mathbb{R}$, cf. (14.50) et (14.51) :

$$\Phi_X(\xi) = P_X(e^{-i\xi\cdot}) = \int_{x\in\mathbb{R}} e^{-ix\xi} dP_X(x). \tag{14.57}$$

Définition 14.36 La densité $\Phi_X : \mathbb{R} \to \mathbb{R}$ est appelée fonction caractéristique de P_X :

$$\widehat{P_X} = T_{\Phi_X} \overset{\text{noté}}{=} \Phi_X, \qquad \text{soit} \qquad d\widehat{P_X}(\xi) = \Phi_X(\xi)\, d\xi. \tag{14.58}$$

Ainsi, pour tout fonction $\varphi \in \mathcal{D}(\mathbb{R})$:

$$\widehat{P_X}(\varphi) = \int_{\xi\in\mathbb{R}} \varphi(\xi)\, \Phi_X(\xi)\, d\xi. \tag{14.59}$$

Remarque 14.37 Définition alternative, cf. (14.57) :

$$\Phi_X(\xi) = P(e^{-i\xi X}) \qquad (= E(e^{-i\xi X}) = \int_{\omega\in\Omega} e^{-iX(\omega)\xi} dP(\omega)). \tag{14.60}$$

∎

Exemple 14.38 Si P_X est la probabilité de densité gaussienne $g_{m,\sigma}(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}$, alors $\widehat{P_X}$ est la mesure finie de densité Φ_X donnée par :

$$\Phi_X(\xi) = e^{-im\xi} e^{-\frac{\sigma^2\xi^2}{2}}, \tag{14.61}$$

cf. (14.32). ∎

Proposition 14.39 *Φ_X caractérise la loi P_X : si les v.a.r. X et Y ont la même fonction caractéristique, soit $\Phi_X = \Phi_Y$, alors $P_X = P_Y$ (les v.a.r. X et Y ont la même loi).*

Preuve. $d\widehat{P_X}(\xi) = \Phi_X(\xi)d\xi = \Phi_Y(\xi)d\xi = d\widehat{P_Y}(\xi)$, donc P_X et P_Y ont la même transformée de Fourier, donc elles sont égales (Fourier inverse). ∎

14.5.7 Développement limité de Φ_X

Lemme 14.40 *Soit X une v.a.r. qui admet un moment d'ordre $k \in \mathbb{N}$. Alors sa fonction caractéristique Φ_X est dérivable k fois sur $\mathbb{R}$ et :*

$$\Phi_X^{(k)}(0) = (-i)^k P(X^k) \qquad (= (-i)^k E(X^k)). \tag{14.62}$$

En particulier $\Phi_X(0) = 1$.

Preuve. Pour $k = 0$ (14.60) donne $\Phi_X(0) = \int_\Omega dP = 1$. Puis on applique le théorème de convergence dominée à (14.60). Pour $k = 1$ on obtient $\Phi_X^{(k)}(\xi) = \int_{\omega\in\Omega}(-iX(\omega))e^{-iX(\omega)\xi}dP(\omega)$ (car $|(-iX(\omega))e^{-iX(\omega)\xi}| \le |X(\omega)|$ avec $X \in L^1$), donc pour $\xi = 0$ on obtient $\Phi_X^{(k)}(0) = (-i)\int_{\omega\in\Omega} X(\omega)dP(\omega) = (-i)P(X)$. Puis récurrence sur k. ∎

Proposition 14.41 *Pour toute v.a.r. X de carré P-intégrable (qui admet donc une moyenne et une variance) on a, $\Phi_X \in C^2(\mathbb{R},\mathbb{C})$, et au voisinage de $\xi = 0$:*

$$\Phi_X(\xi) = 1 - iE(X)\xi - \frac{1}{2}E(X^2)\xi^2 + o(\xi^2), \tag{14.63}$$

et donc :

$$\log(\Phi_X(\xi)) = -iE(X)\xi - \frac{1}{2}\sigma^2(X)\xi^2 + o(\xi^2). \tag{14.64}$$

Preuve. (14.63) grâce à (14.62).

Et $\log(1-z) = -(z+\frac{z^2}{2}+o(z))$ au voisinage de $z = 0$ (par intégration de $\frac{1}{1-z} = 1+z+z^2+...$). D'où $\log(1-(iE(X)\xi + \frac{1}{2}E(X^2)\xi^2 + o(\xi^2)) = -(iE(X)\xi + \frac{1}{2}E(X^2)\xi^2 + \frac{1}{2}(iE(X)\xi)^2 + o(\xi^2)) = -iE(X)\xi - \frac{1}{2}(E(X^2) - E(X)^2)\xi^2 + o(\xi^2)$ comme annoncé. ∎

14.5.8 Fonction caractéristique d'une somme de v.a.r. indépendantes

Proposition 14.42 *Si $X_1, ..., X_n$ sont n v.a.r. indépendantes, alors pour tout ξ :*

$$\Phi_{X_1+...+X_n}(\xi) = \prod_{k=1}^n \Phi_{X_k}(\xi). \tag{14.65}$$

Preuve. On a $e^{-i(X_1(\omega)+...+X_n(\omega))\xi} = \prod_{k=1}^n e^{-iX_k(\omega)\xi}$.

Et, les v.a.r. étant indépendantes, on a $P_{X_1+...+X_n} = P_{X_1} * ... * dP_{X_n}$, et donc :

$$P_{X_1+...+X_n}(f) = \int ... \int_{\mathbb{R}^n} f(x_1 + ... + x_n)dP_{X_1}(x_1)...dP_{X_n}(x_n),$$

cf. convolution de probabilités (ou convolution de distributions). D'où le résultat avec $f(x) = e^{-ix\xi}$ qui donne $f(x_1 + ... + x_n) = \prod_{i=1}^n f(x_i)$.

∎

14.5.9 Démonstration du théorème de la limite centrale

Preuve. Il s'agit de démontrer la convergence (14.21). On va s'intéresser à P_{Z_n} au travers de sa transformée de Fourier $\widehat{P_{Z_n}} = T_{\Phi_{Z_n}}$ qui est plus simple (mesure de densité).

Quitte à remplacer X_i par $X_i - m$, on suppose que m=0. On a donc $Z_n = \frac{\sqrt{n}Y_n}{\sigma}$.

La fonction caractéristique de la v.a.r. $\frac{\sqrt{n}Y_n}{\sigma} = \sum_k \frac{X_k}{\sigma\sqrt{n}}$ est donnée par, sachant $P_{X_k} = P_{X_1}$ pour tout k, cf. (14.65) et (14.66) (dans le lemme suivant) :

$$\Phi_{Z_n}(\xi) = \prod_{k=1}^n \Phi_{\frac{X_k}{\sigma\sqrt{n}}}(\xi) = (\Phi_{X_1}(\frac{\xi}{\sigma\sqrt{n}}))^n.$$

Avec ici $E(X_1) = 0$ et donc $E(X_1^2) = \sigma^2$, on a, cf. (14.64) pour ξ fixé et au voisinage de $n = \infty$:

$$\log(\Phi_{X_1}(\frac{\xi}{\sigma\sqrt{n}})) = -\frac{1}{2}\frac{\xi^2}{n} + o(\frac{\xi^2}{n}) = \frac{\xi^2}{n}(-\frac{1}{2} + o(1)),$$

d'où, pour ξ fixé et au voisinage de $n = \infty$:

$$\Phi_{Z_n}(\xi) = \exp(\log(\Phi_{Z_n}(\xi))) = \exp(n\log(\Phi_{X_1}(\frac{\xi}{\sigma\sqrt{n}}))) = e^{\xi^2(-\frac{1}{2}+o(1))}.$$

D'où, pour ξ fixé et au voisinage de $n = \infty$:

$$\Phi_{Z_n}(\xi) \sim e^{-\frac{1}{2}\xi^2},$$

Donc la mesure P_{Z_n} est transformé par Fourier en $\widehat{P_{Z_n}}$ mesure de densité Φ_{Z_n} vérifiant l'équivalence $\Phi_{Z_n}(\xi) \sim_\infty e^{-\frac{1}{2}\sigma^2\xi^2}$. D'où le résultat par Fourier inverse, cf. (14.27). ▪

Lemme 14.43 *Si $X : \Omega \to \mathbb{R}$ est une v.a.r., si $a > 0$, si $h : \mathbb{R} \to \mathbb{R}$ est positive ou bien est P_X-intégrable (i.e. $h \in L^1(\mathbb{R}, P_X)$, i.e. $\int_{\mathbb{R}} |h(x)|\, dP_X < \infty$), alors :*

$$h_a(x) \stackrel{\text{déf}}{=} h(ax) \quad \Longrightarrow \quad P_{aX}(h) = P_X(h_a). \tag{14.66}$$

En particulier, si P_X est une mesure de densité f alors $\frac{1}{a}P_{aX}$ est une mesure de densité $f_{\frac{1}{a}}$:

$$dP_X = f(x)\,dx \quad \Longrightarrow \quad dP_{aX} = \frac{1}{a}f(\frac{x}{a})\,dx. \tag{14.67}$$

Ou encore, si P_{aX} est une mesure de densité g, alors P_X est une mesure de densité ag_a.

Preuve. Avec la proposition de la mesure image, pour $a > 0$:

$$P_{aX}(h) = \int_y h(y)\,dP_{aX}(y) = \int_\omega h((aX)(\omega))\,dP(\omega) = \int_\omega h_a(X(\omega))\,dP(\omega) = \int_x h_a(x)\,dP_X(x)$$

D'où (14.66). D'où dans le cas d'une mesure P de densité f :

$$P_X(h_a) = \int_x h(ax)f(x)\,dx = \int_y h(y)f(\frac{y}{a})\,\frac{dy}{a} = P_{aX}(h).$$

Et $P_X(h) = P_{aX}(h_{\frac{1}{a}}) = \int_y h(\frac{y}{a})g(y)\,dy = \int_x h(x)g(ax)\,a\,dx$. ▪

15 Annexe : les moments

C'est un rappel d'intégration appliqué à la mécanique.

15.1 Notations

On se place dans le cas d'un sous-ensemble $B \subset \mathbb{R}^n$. Soit $\mathcal{T}$ une tribu sur B, et soit μ une mesure sur $(B, \mathcal{T})$. Donc $(B, \mathcal{T}, \mu)$ est un ensemble mesuré.

On considèrera les cas où :

1- μ est une mesure discrète définie finie sur $B = \{\vec{x}_i : i \in I\}$ (un ensemble de points) où I est un ensemble fini ou dénombrable, i.e. :

$$\mu = \sum_{i\in I} \rho_i \delta_{\vec{x}_i}, \tag{15.1}$$

dite mesure atomique de poids $\rho_i > 0$ aux points $\vec{x}_i$.

2- μ est une mesure continue finie sur $B \subset \mathbb{R}^n$, i.e. mesure de densité $\rho > 0$ auquel cas on écrit

$$d\mu = \rho(\vec{x})\,dx_1...dx_n \stackrel{\text{noté}}{=} \rho(\vec{x})\,dx, \tag{15.2}$$

où $dx_1...dx_n =^{\text{noté}} dx$ est la mesure de Lebesgue (dans $\mathbb{R}^n$).

15.2 Moment d'ordre 0 : la masse

Définition 15.1 Pour $A \subset \mathcal{T}$ (A sous-ensemble mesurable de B) :

$$\text{masse de } A \stackrel{\text{déf}}{=} \mu(A) \stackrel{\text{noté}}{=} \int_A d\mu \quad \in \overline{\mathbb{R}_+}. \tag{15.3}$$

Donc, respectivement dans le cas discret et dans le cas continu :

$$\mu(A) = \sum_{i \in I} \rho_i \qquad \text{et} \qquad \mu(A) = \int_A \rho(\vec{x})\, dx. \tag{15.4}$$

Définition 15.2 Soit $f : B \to \mathbb{R}$ une fonction mesurable positive. Quand il existe, le moment d'ordre 0 de f (relativement à la mesure μ) est "la mesure de f" et est appelé "la masse de f" :

$$\text{masse de } f \stackrel{\text{déf}}{=} \mu(f) \stackrel{\text{noté}}{=} \int_B f\, d\mu, \tag{15.5}$$

encore appelée la masse de f relativement à la mesure μ. (On retrouve $\mu(1_B) = \mu(B)$.)

Donc, respectivement dans le cas discret et dans le cas continu :

$$\mu(f) = \sum_{i \in I} f(\vec{x}_i)\, \rho_i \qquad \text{et} \qquad \mu(f) = \int_B f(\vec{x})\, \rho(\vec{x})\, dx. \tag{15.6}$$

15.3 Moment d'ordre 1 et le centre de gravité

Définition 15.3 Quand il existe, le moment d'ordre 1 de la mesure μ par rapport à un point $\vec{x}_0$ est le vecteur :

$$\vec{\mu}_{1,\vec{x}_0}(B) \stackrel{\text{déf}}{=} \mu((\vec{x}-\vec{x}_0)) \stackrel{\text{noté}}{=} \int_B (\vec{x}-\vec{x}_0)\, d\mu(\vec{x}). \tag{15.7}$$

(On a noté $(\vec{x}-\vec{x}_0)$ la fonction $\vec{f} : \vec{x} \to \vec{f}(\vec{x}) = \vec{x}-\vec{x}_0$).)

Donc, dans le cas discret, avec $\vec{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{in} \end{pmatrix}$, $\vec{\mu}_{1,\vec{x}_0}(B) \in \mathbb{R}^n$ est le vecteur :

$$\vec{\mu}_{1,\vec{x}_0}(B) = \sum_{i \in I} (\vec{x}_i - \vec{x}_0)\, \rho_i = \begin{pmatrix} \sum_{i \in I} (x_{i1} - x_{i0})\, \rho_i \\ \vdots \\ \sum_{i \in I} (x_{in} - x_{in})\, \rho_i \end{pmatrix} \tag{15.8}$$

et dans le cas continu, avec $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\vec{\mu}_{1,\vec{x}_0}(B) \in \mathbb{R}^n$ est le vecteur :

$$\vec{\mu}_{1,\vec{x}_0}(B) = \int_B (\vec{x}-\vec{x}_0)\, \rho(\vec{x})\, dx = \begin{pmatrix} \int_B (x_1 - x_{01})\, \rho(\vec{x})\, dx \\ \vdots \\ \int_B (x_n - x_{0n})\, \rho(\vec{x})\, dx \end{pmatrix}. \tag{15.9}$$

On a donc immédiatement, quand les moments existent :

$$\vec{\mu}_{1,\vec{x}_0}(B) = \vec{\mu}_{1,0}(B) - \mu(B)\, \vec{x}_0, \tag{15.10}$$

encore noté :

$$\int_B (\vec{x}-\vec{x}_0)\, d\mu(\vec{x}) = \int_B \vec{x}\, d\mu(\vec{x}) - \vec{x}_0 \int_B d\mu(\vec{x}). \tag{15.11}$$

Définition 15.4 Dans le cas $\mu(B) < \infty$ (masse finie) et $\mu(B) \neq 0$ (masse non nulle), le centre de gravité de B (ou encore point d'équilibre, ou encore centre d'inertie) est le point $\vec{x}_G$ tel que :

$$\vec{\mu}_{1,\vec{x}_G}(B) = 0, \tag{15.12}$$

i.e. $\vec{x}_G$ est le point donné par :

$$\vec{\mu}_{1,0}(\vec{x}) = \vec{x}_G\, \mu(B), \qquad \text{i.e.} \qquad \vec{x}_G \int_B d\mu = \int_B \vec{x}\, d\mu. \tag{15.13}$$

Autrement dit :

$$\vec{x}_G = \frac{\int_B \vec{x}\, d\mu}{\int_B d\mu} = \int_B \vec{x}\, dP \qquad \text{où} \qquad dP = \frac{1}{\mu(B)} d\mu, \tag{15.14}$$

où P est une mesure de probabilité (mesure de masse $P(B) = 1 = \frac{100}{100} = 100\%$).

Respectivement dans le cas discret et dans le cas continu, le centre de gravité x_G est donc le point tel que :

$$\begin{cases} \vec{x}_G = \sum_{i \in I} \vec{x}_i \, \frac{\rho_i}{\mu(B)} = \sum_{i \in I} \vec{x}_i p_i, & p_i = \frac{\rho_i}{\sum_I \rho_i}, \\ \vec{x}_G = \int_B \vec{x}\, \frac{\rho(\vec{x})}{\mu(B)}\, dx = \int_B \vec{x}\, p(\vec{x})\, dx, & p(\vec{x}) = \frac{\rho(\vec{x})}{\int_B \rho(\vec{x})\, dx}, \end{cases} \tag{15.15}$$

où p_i et $p(\vec{x})$ sont les masses et densités de masse adimensionnées (masse totale considérée comme étant la masse unité par changement d'unité de mesure).

Remarque 15.5 Si remplace μ par $c\mu$ où c est une constante > 0, la position du centre de gravité n'est pas changée, cf. (15.13) où on considère $c\,d\mu$ au lieu de $d\mu$. Ou encore $dP = \frac{d(c\mu)}{(c\mu)(B)} = \frac{d\mu}{\mu(B)}$.

En d'autres termes, un changement d'unité de mesure (par exemple passage des kilogrammes aux "pounds" anglais) ne modifie pas la position du centre de gravité.

Donc, quitte à faire un changement d'unité de mesure, pour calculer le centre de gravité on peut remplacer μ par $\frac{\mu}{m}$ où $m = \mu(B)$, ce qu'exprime $dP = \frac{1}{\mu(B)} d\mu$.

Cela revient à considérer des objets de masse unité (interprétation probabilistique), et dans ce cadre, le centre de gravité est appelé l'espérance et noté $\vec{x}_g = \overline{\vec{X}}$.

Et dans le cas discret, lorsque $I = 1, ..., n{+}1$ (donc $n{+}1$ points dans $\mathbb{R}^n$) en considérant l'objet comme ayant une masse unité, on a $\sum_{i=1}^{n+1} \rho_i \vec{x}_i = \vec{x}_G$ avec $\sum_{i=1}^{n+1} \rho_i = 1$: le centre de gravité est le barycentre des x_i de coordonnées barycentriques les masses ρ_i. ▲

Exemple 15.6 Demi-cercle supérieur Γ de centre 0 et de rayon $R : x^2 + y^2 = R^2$, $y \geq 0$. On suppose qu'un point quelconque de ce demi-cercle peut-être atteint de manière équiprobable, i.e. avec une densité de probabilité $p = p_0$ constante.

L'altitude moyenne des tirs est $y_m = \int_\Gamma y\, p(y)\, d\Gamma = p_0 \int_\Gamma y\, d\Gamma$. Ici $\vec{r} = \begin{pmatrix} x = R\cos\theta \\ y = R\sin\theta \end{pmatrix}$ est la position d'un point du cercle, avec $\theta \in [0, \pi]$, d'où $d\vec{r} = \begin{pmatrix} -R\sin\theta \\ R\cos\theta \end{pmatrix} d\theta$, d'où l'élément de longueur $d\Gamma = ||d\vec{r}|| = R\, d\theta$; d'où $y_m = p_0 \int_{\theta=0}^{\pi} (R\sin\theta)(R\, d\theta) = 2p_0 R^2$.

D'autre part, comme p_0 est tel que $\int_\Gamma p_0\, d\Gamma = 1$, on a $\int_{\theta=0}^{\pi} p_0\, R\, d\theta = 1 = p_0 \pi R$, et donc $p_0 = \frac{1}{\pi R}$. D'où $y_m = \frac{2}{\pi} R$. C'est l'altitude moyenne des tirs, ou encore c'est la position verticale du centre de gravité d'une demie roue. La position moyenne horizontale est bien sûr $x_m = 0$ (exercice). ▲

15.4 Moment d'ordre 2, la variance, l'écart type

Définition 15.7 Quand il existe, le moment d'ordre 2 de μ par rapport à $\vec{x}_0$, encore appelé moment d'inertie de μ par rapport à $\vec{x}_0$, est :

$$\mu_{2,\vec{x}_0}(B) = \mu(||\vec{x} - \vec{x}_0||^2) \overset{\text{noté}}{=} \int_B ||\vec{x} - \vec{x}_0||^2 \, d\mu(x). \tag{15.16}$$

(On a noté $||\vec{x} - \vec{x}_0||^2$ la fonction $\vec{f} : \vec{x} \to \vec{f}(\vec{x}) = ||\vec{x} - \vec{x}_0||^2$.)

Donc, respectivement dans le cas discret et dans le cas continu :

$$\mu_{2,\vec{x}_0}(B) = \sum_{i \in I} ||\vec{x}_i - \vec{x}_0||^2 \, \rho_i, \qquad \text{et} \qquad \mu_{2,\vec{x}_0}(B) = \int_B ||\vec{x} - \vec{x}_0||^2 \, \rho(\vec{x})\, dx. \tag{15.17}$$

Définition 15.8 Quand $\vec{x}_0 = \vec{x}_G$ (centre de gravité) et $\mu = P$ est une mesure de probabilité, le moment d'ordre 2 par rapport à $\vec{x}_G$ est appelé variance $\text{Var}(B)$, et sa racine carrée $\sigma = \sqrt{\text{Var}(B)}$ est appelé l'écart type :

$$\text{Var}(B) = \sigma^2(B) \overset{\text{déf}}{=} \mu_{2,\vec{x}_G} = \int ||\vec{x} - \vec{x}_G||^2 \, d\mu(x).$$

Proposition 15.9 *Quand la moyenne et le moment d'ordre 2 existent dans* $\mathbb{R}$*, pour tout* $\vec{x}_0 \in B$ *on a :*

$$\mu(||\vec{x}-\vec{x}_0||^2) = \mu(||\vec{x}-\vec{x}_G||^2) + ||\vec{x}_G-\vec{x}_0||^2\mu(B), \tag{15.18}$$

Et donc le moment d'inertie relativement à $\vec{x}_0$ *est minimum quand* $\vec{x}_0 = \vec{x}_G$*.*

D'où le nom de "centre d'inertie" donné au centre de gravité.

Preuve. $\vec{x}-\vec{x}_0 = \vec{x}-\vec{x}_G + \vec{x}_G-\vec{x}_0$, d'où $||\vec{x}-\vec{x}_0||^2 = ||\vec{x}-\vec{x}_G||^2 + ||\vec{x}_G-\vec{x}_0||^2 + 2(\vec{x}-\vec{x}_G, \vec{x}_G-\vec{x}_0)_{\mathbb{R}^n}$. D'où $\mu(||\vec{x}-\vec{x}_0||^2) = \mu(||\vec{x}-\vec{x}_G||^2) + ||\vec{x}_G-\vec{x}_0||^2\mu(1_B) + 2(\vec{x}_G-\vec{x}_0, \mu(\vec{x}-\vec{x}_G)_{\mathbb{R}^n}$, avec $\mu(\vec{x}-\vec{x}_G) = 0$ par définition de $\vec{x}_G$. D'où (15.18).

Et (15.18) indique que $\mu(||\vec{x}-\vec{x}_0||^2) > \mu(||\vec{x}-\vec{x}_G||^2)$ pour tout $\vec{x}_0 \neq \vec{x}_G$. ▟

Exemple 15.10 Cas discret : Soit une tige "sans masse" sur laquelle on dispose de masses ponctuelles ρ_i aux points $x_i = i$ pour $i = 1,2,3,4,5,6$. Soit $f = 1$.

La masse de l'objet obtenu est $\mu_0(1) = \rho_1+...+\rho_6$ somme des masses ponctuelles. En particulier si toutes les masses sont égales une même masse ponctuelle $m = \rho_i$, on a $\mu_0(1) = 6m = M$ la masse totale.

Son moment d'ordre 1 par rapport à x_0 est $m_{1,x_0}(1) = (1-x_0)\rho_1 + (2-x_0)\rho_2 + ... + (6-x_0)\rho_6$. Ce moment d'ordre 1 est nul pour $x_0 \stackrel{\text{noté}}{=} x_G$ tel que $\rho_1 + 2\rho_2 + ...6\rho_6 = x_G M$. En particulier si toutes les masses sont égales à la masse ponctuelle m, on a $21m = x_0 6m$, i.e. $x_0 \stackrel{\text{noté}}{=} x_G = \frac{21}{6} = \frac{7}{2} = 3,5$: c'est la position du centre de gravité. Si on met au point x_G un support sous la tige horizontale, la tige reste en équilibre.

Son moment d'ordre 2 par rapport à x_0 est $m_{2,x_0}(1) = (1-x_0)^2\rho_1 + ... + (6-x_0)^2\rho_6$. Et si on veut faire tourner la tige sur elle même, ce sera plus facile de le faire si on la tient en x_G : c'est là où le moment d'inertie est le plus faible. ▟

15.5 Moment d'ordre m

Définition 15.11 Quand il existe, le moment d'ordre $m > 1$ de la mesure μ par rapport à $\vec{x}_0$ est :

$$\mu_{m,\vec{x}_0}(B) \stackrel{\text{déf}}{=} \mu(||\vec{x}-\vec{x}_0||^m). \tag{15.19}$$

Donc, respectivement dans le cas discret et dans le cas continu :

$$\mu_{m,\vec{x}_0}(B) = \sum_i ||\vec{x}_i-\vec{x}_0||^m \rho_i, \qquad \text{et} \qquad \mu_{m,\vec{x}_0}(B) = \int_a^b ||\vec{x}-\vec{x}_0||^m \rho(\vec{x})\,dx.$$

Remarque 15.12 Le moment $\mu_{3,x_G}(B)$ d'ordre 3 permet de savoir si les masses sont disposées "symétriquement par rapport au centre de gravité". Si cette disposition est symétrique on a immédiatement $\mu_{3,x_G}(B) = 0$.

Exemple : 2 masses $m_1 = 1$ au point 1 et $m_2 = 1$ au point 2 ont pour centre de gravité le point $x_G = \frac{1}{2}(1*1+2*1) = \frac{3}{2}$ et pour moment $\mu_{3,x_G}(1) = 0$: la disposition des masses est "symétrique" par rapport au centre de gravité.

Exemple : 2 masses $m_1 = 1$ au point 1 et $m_2 = 2$ au point 2 ont pour centre de gravité le point $x_G = \frac{1}{3}(1*1+2*2) = \frac{5}{3}$ et pour moment $\mu_{3,x_G}(1) = (\frac{-2}{3})^3*1+(\frac{1}{3})^3*2 = \frac{-4}{9} \neq 0$: la disposition des masses n'est pas "symétrique" par rapport au centre de gravité. ▟

Remarque 15.13 Le moment d'ordre 4 permet de mesurer "l'aplatissement". S'il est grand la distribution est étalée, s'il est petit elle est concentré. On général on regarde $\gamma = \dfrac{\mu_{4,x_G}}{\mu_{2,x_G}}$ qu'on compare à la distribution de Gauss (loi normale). On a $\gamma = 3$ pour la distribution de Gauss, et pour $\gamma > 3$, on est plus étalé, et pour $\gamma < 3$ on est plus concentré. ▟

Références

[1] Bertoin Jean : *Probabilités : cours de licence de mathématiques appliquées*. Université Paris VI, 2000/2001, http ://www.proba.jussieu.fr/supports.php

[2] Degrave C., Degrave D. : *Probabilités – Statistiques*. HEC, Options Générale économie, Précs de mathématiques, Bréal, juin 1990.

[3] Feller William : *An Introduction to Probability Theory and its Applications.* Third Edition 1968, Revised Printing (1970 ?), John Wiley & Sons.

[4] Jolion Jean-Michel : *Probabilités et statistiques.* `http://rfv.insa-lyon.fr/~jolion/`

[5] Leray Philippe :
`http://asi.insa-rouen.fr/enseignement/siteUV/rna/Cours2/01-ProbaCond.pdf`

[6] Mazel Claude : cours de l'ISIMA 2007.

[7] Méléard Sylvie : *Aléatoire.* Cours de l'école Polytechnique, promotion X2009,
`http://catalogue.polytechnique.fr/cours.php?id=2392`

[8] Metivier Michel : *Probabilités.* Cours de l'École Polytechnique 1983.

[9] Metivier M., Neveu J. : *Probabilités.* Cours de l'École Polytechnique 1983.

[10] Nedzela Michel : *Modèles probabilistes d'aide à la décision.* Presses de l'Université du Québec, 1987.

[11] Phan Thérèse : *Probabilités.* Cours de l'École Centrale de Paris 1984.

[12] Réau J.-P., Chauvat G. : *Probabilités et statistiques pour les sciences économiques et sociales, résumés de cours, exercices et problèmes corrigés.* Armand Colin, Flash U, Deug sciences eco, 1988.

[13] Rudin Walter : *Analyse réelle et complexe.* Masson, 1995.

[14] Schwartz L. : *Méthodes mathématiques pour les sciences physiques.* Hermann, collection enseignement des sciences, 1965.

[15] Suquet Charles : *Introduction au calcul des probabilités (Deug=L2).* Université des Sciences et Technologies de Lille, 2006, `http://math.univ-lille1.fr/~suquet/`

[16] Suquet Charles : *Cours IPE (Intégration et Probabilités Élémentaires).* Université des Sciences et Technologies de Lille, 2007, `http://math.univ-lille1.fr/~suquet/`

[17] http ://www.bacamaths.net/.

[18] *La distribution gaussienne : une histoire de cloches.*
`http://elementaire.web.lal.in2p3.fr/documents/numero2/ANALYSE2.pdf`

[19] Müller Didier : *Probabilités.* `http://www.apprendre-en-ligne.net/MADIMU2/PROBA/`