XXII Probabilités sur un univers fini

5 juin 2016

La théorie des probabilités cherche à modéliser des phénomènes faisant intervenir le hasard. Puisqu'il s'agit d'une modélisation, il conviendra, pour chaque définition que nous allons donner, d'une part d'apprendre sa définition mathématique et d'autre part de comprendre en quoi cette définition modélise un phénomène aléatoire.

1. Événements, probabilités

1.1. Expérience aléatoire et univers

a. Introduction

On parlera d'expérience aléatoire pour modéliser un processus dont le résultat est incertain. Exemple : tirage au sort d'une boule dans une urne, tirage à pile ou face avec une pièce de monnaie, lancer d'un ou plusieurs dés, choix au hasard d'une personne dans la population française, tirage de trois cartes à jouer au hasard dans un paquet, etc.

On appellera généralement *univers des possibles* ou *univers*, l'ensemble des résultats possibles d'une expérience aléatoire. Cet univers dépend bien évidemment de la modélisation choisie : par exemple pour un tirage à pile ou face, on peut modéliser l'univers comme étant l'ensemble $\{ \text{pile}, \text{face} \}$ ou comme $\{ \text{pile}, \text{face}, \text{tranche} \}$.

Il arrive parfois qu'on s'intéresse à une expérience aléatoire donnant plusieurs résultats. Par exemple, si on tire une carte à jouer dans un jeu, on peut s'intéresser à la couleur de la carte, auquel cas on considérera l'univers $\Omega_1 = \{ \heartsuit, \diamondsuit, \spadesuit, \clubsuit \}$ ou à sa valeur, auquel cas on s'intéressera à l'univers $\Omega_2 = \{ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \text{Valet}, \text{Dame}, \text{Roi} \}$. Si on s'intéresse aux deux simultanément, on prendra plutôt comme univers $\Omega = \Omega_1 \times \Omega_2$.

De manière générale, on prendra pour univers un ensemble nous permettant de représenter simultanément tous les résultats qui nous intéresseront.

Sur l'exemple précédent, on peut s'intéresser à l'*événement* qui consiste à tirer une carte rouge et de valeur trois ou quatre. Cet événement est modélisé comme une partie de Ω, en l'espèce la partie $\{ (\heartsuit, 3), (\heartsuit, 4), (\diamondsuit, 3), (\diamondsuit, 4) \}$.

Dans toute la suite de ce chapitre, on utilisera souvent des expériences imagées : tirages de dés, de boules dans une urne *etc.* On adoptera les conventions suivantes, sauf mention du contraire :

- les dés sont équilibrés, à six faces ;
- les urnes sont opaques, les boules sont indiscernables au toucher et, si une urne contient n boules, ces dernières sont numérotées de 1 à n ;
- les jeux de cartes sont parfaitement mélangés.

b. Univers, événements

Définition 1.1.1 (Univers)**.**
On appelle *univers* un ensemble non vide Ω.

Cette année, on se limitera au cas où Ω est fini.

Dans ce cas [1], on appelle *événement* toute partie de l'univers, c'est-à-dire tout élément de $\mathscr{P}(\Omega)$ et on appelle *événement élémentaire* ou *éventualité* les événements singletons, c'est-à-dire de la forme $\{ \omega \}$, pour $\omega \in \Omega$ (selon le contexte, le terme événement élémentaire peut parfois désigner les éléments de Ω et non les singletons).

Un événement est dit *impossible* s'il désigne la partie vide ($\varnothing$) et *certain* s'il désigne la partie pleine (Ω).

Étant donnés deux événements A et B, on définit

— l'*événement contraire de* A : $\Omega \setminus A$, noté $\bar{A}$.
— l'événement «A et B» (conjonction de A et B) : $A \cap B$.
— l'événement «A ou B» (disjonction de A et B) : $A \cup B$.

On dit que A et B sont incompatibles si $A \cap B = \varnothing$, autrement dit si leur conjonction est impossible.

On dit que des événements $A_1, \ldots, A_n$ sont *mutuellement incompatibles* si leur conjonction est impossible. On dit qu'ils sont *deux à deux incompatibles* si pour tout i et j, $i \neq j$ implique A_i et A_j incompatibles. Dans ce dernier cas, on dit que leur union $A_1 \cup \ldots \cup A_n$ est une *union disjointe.*

Remarque 1.1.2.
L'incompatibilité deux à deux de n événements (avec $n \geqslant 2$) implique l'incompatibilité mutuelle mais la réciproque est fausse. Considérons par exemple $\Omega = \{ 1, 2, 3, 4, 5, 6 \}$ l'univers des résultats d'un tirage d'un dé à six faces. Alors les trois événements «le résultat est pair», «le résultat est divisible par 3» et «le résultat est un nombre premier» ne sont pas deux à deux incompatibles mais sont mutuellement incompatibles.

Remarque 1.1.3.
Dans le cas où Ω n'est pas fini ni dénombrable, modéliser les événements par les éléments de $\mathscr{P}(\Omega)$ pose des problèmes techniques. Pour les résoudre, on impose aux événements d'être des élément d'un sous-ensemble $\mathscr{T}$ de $\mathscr{P}(\Omega)$, cet ensemble $\mathscr{T}$ devant former ce qu'on appelle une *tribu*. Dans le cas fini ou dénombrable, $\mathscr{P}(\Omega)$ est une tribu.

Exemple 1.1.4.
Pour modéliser les tirages successifs, sans remise, de deux boules dans une urne contenant n boules numérotées de 1 à n, on peut utiliser l'univers $[\![1, n]\!]^2$.

Ici, l'événement $\{(i, k)\}$ modélise « on tire d'abord la boule i, puis la boule k » Les événements du type $\{(i, i)\}$ n'ont pas d'interprétation dans notre modèle. Ce n'est pas grave : on leur attribuera plus tard une probabilité nulle.

c. Système complet d'événements

Définition 1.1.5.
On dit qu'une famille $(A_i)_{i \in I}$ d'événements dans un univers Ω est un *système complet d'événements* si ces événements sont deux à deux incompatibles et que leur union (disjointe) est certaine : $\Omega = \bigcup_{i \in I} A_i$.

Exemple 1.1.6.
Si A est un événement, $\left\{A, \overline{A}\right\}$ est un système complet d'événements.

1. Dans le cas infini, la définition est un peu plus subtile. La cas dénombrable sera traité en seconde année.

Exemple 1.1.7.
Pour l'exemple du tirage d'une carte donné plus haut, la famille (A_1, A_2, A_3), où A_1 est l'événement «la carte tirée est rouge», A_2 l'événement «la carte tirée est un sept noir» et A_3 l'événement «la carte tirée est noire mais n'est pas un sept» constitue un système complet d'événements.

Remarque 1.1.8.
La notion de système complet d'événements est très proche de celle de partition. Les différences sont les suivantes :

1. un système complet d'événements est une famille de parties de Ω alors qu'une partition est un ensemble de parties de Ω ;
2. la notion de système complet d'événements ne s'utilise qu'en probabilités ;
3. rien dans la définition de système complet d'événements n'impose aux A_i d'être non vides.

Proposition 1.1.9.
Soit $(A_i)_{i \in I}$ un système complet d'événements sur un univers Ω, soit B un événement. Alors, B peut s'écrire comme l'union suivante :

$$B = \bigcup_{i \in I} B \cap A_i$$

et cette union est une union disjointe (les $(B \cap A_i)_{i \in I}$ sont deux à deux incompatibles).

Démonstration.
Montrons tout d'abord que B est inclus dans cette réunion. Soit $b \in B$. On a $b \in \Omega$ et la réunion des A_i pour $i \in I$ est égale à Ω, donc il existe un $i_0 \in I$ vérifiant $b \in A_{i_0}$. On a alors $b \in B \cap A_{i_0}$. On a donc

$$b \in \bigcup_{i \in I} B \cap A_i$$

Ce qui montre cette première inclusion.

L'inclusion réciproque est immédiate : pour tout $i \in I$, on a $B \cap A_i \subset B$, donc

$$\bigcup_{i \in I} B \cap A_i \subset B$$

On a donc l'égalité voulue.

On peut aussi montrer cela par calcul sur les ensembles : par la relation de De Morgan,

$$B = B \cap \Omega = B \cap \bigcup_{i \in I} A_i = \bigcup_{i \in I} B \cap A_i.$$

Le fait que l'union soit disjointe est immédiat : soit $(i, j) \in [\![1, n]\!]^2$ vérifiant $i \neq j$. Alors $(B \cap A_i) \cap (B \cap A_j) \subset A_i \cap A_j = \varnothing$. □

1.2. Espaces probabilisés finis

a. Définition

À tout événement, on veut associer sa *probabilité*, qui modélise la « probabilité de réalisation » de cet événement lors de la réalisation de cette expérience aléatoire. Mais que veut dire cette phrase ? On peut le voir comme la fréquence de la réalisation de cet événement au bout d'un « grand » nombre de répétitions « indépendantes » de cette expérience. La loi des grands nombres (sa version faible dans le cas dénombrable sera vue en seconde année) justifie la consistance de la définition suivante avec son interprétation concrète.

Définition 1.2.1.
Une *probabilité* (ou *mesure de probabilité*) sur un univers fini Ω est une application P de $\mathscr{P}(\Omega)$ dans $\mathbb{R}$ vérifiant les trois propriétés suivantes :

1. Pour tout événement A, $0 \leqslant P(A) \leqslant 1$.
2. $P(\Omega) = 1$.
3. Pour tout couple (A, B) d'événements, si A et B sont incompatibles alors $P(A \cup B) = P(A) + P(B)$.

Un *espace probabilisé fini* est un couple (Ω, P), où Ω est un univers fini et P une probabilité.

Pour tout événement A d'un espace probabilisé (Ω, P), la valeur $P(A)$ est appelée probabilité de l'événement A.

On dit qu'un événement A est *presque sûr* (ou *quasi certain*) si $P(A) = 1$ et est *négligeable* (ou *quasi impossible*) si $P(A) = 0$.

Remarque 1.2.2. 1. Dans le cas où l'univers est infini, il faut adapter un peu la définition : l'ensemble de départ de P est alors la tribu des événements et on donne au troisième axiome une forme plus générale. Comme on ne verra cette année que le cas fini, on s'autorisera à omettre la précision « fini » quand on parlera d'espaces probabilisés.

2. L'intérêt de la notion de quasi certitude ou quasi impossibilité n'est pas évidente quand il s'agit de probabilités sur un univers fini. Donnons un exemple intuitif dans le cas d'un univers infini : si on tire un réel au hasard dans $[0, 1]$, il n'est pas impossible d'obtenir exactement le réel $1/3$ mais la probabilité de cette événement est nulle (la probabilité d'obtenir un réel situé dans un intervalle donné est proportionnelle au diamètre de cet intervalle). C'est donc un événement quasi-impossible mais non impossible.

3. Il est important de retenir que la notion de quasi impossibilité ne veut pas dire «probabilité faible». Un physicien dirait qu'un événement de probabilité 10^{-100} est impossible (le nombre d'atomes dans l'univers est de l'ordre de 10^{80}) mais pour un mathématicien, un tel événement n'est même pas un événement quasi impossible.

Définition 1.2.3.
Un prédicat défini sur Ω vrai sur un événement de probabilité 1 sera dit « presque-sûr ».

Exemple 1.2.4.
Reprenons l'exemple 1.1.4 : on tire successivement, sans remise, deux boules dans une urne contenant $n \geqslant 2$ boules numérotées de 1 à n. Il est pratique de considérer comme univers $\Omega = [\![1, n]\!]^2$. On prend alors une probabilité P telle que l'événement $\{ (i, i) \mid 1 \leqslant i \leqslant n \}$ est négligeable.

Sur cet univers, presque-sûrement on aura « $i \neq k$ » pour $1 \leqslant i, k \leqslant n$.

b. Probabilité uniforme

Définition 1.2.5 (probabilité uniforme)**.**
Soit Ω un univers fini. L'application

$$P: \begin{array}{rcl} \mathscr{P}(\Omega) & \to & \mathbb{R} \\ A & \mapsto & \frac{\operatorname{Card} A}{\operatorname{Card} \Omega} \end{array}$$

est une probabilité, appelée *probabilité uniforme.*

Démonstration.
Remarquons que Ω est un ensemble fini, donc ses parties sont finies également donc ont des cardinaux entiers. Comme de plus $\Omega \neq \varnothing$, son cardinal est non nul, donc on peut diviser par $\operatorname{Card} \Omega$. P est donc bien définie sur $\mathscr{P}(\Omega)$.

Pour montrer que P est une probabilité, il suffit de vérifier les trois propriétés de la définition :

1. Soit $A \in \mathscr{P}(\Omega)$, alors $0 \leqslant \operatorname{Card} A \leqslant \mathscr{P}(\Omega)$ donc $0 \leqslant P(A) \leqslant 1$.
2. On a bien $P(\Omega) = 1$.
3. Soit (A, B) deux événements incompatibles. Alors $A \cap B = \varnothing$ donc $\operatorname{Card}(A \cup B) = \operatorname{Card} A + \operatorname{Card} B$, donc $P(A \cup B) = P(A) + P(B)$.

□

Remarque 1.2.6.
Cette probabilité modélise souvent l'expression « au hasard », prise dans son acceptation courante (tirer une boule au hasard dans une urne, etc.).

Exemple 1.2.7.
Continuons l'exemple 1.2.4 : la probabilité que l'on considérera sur Ω n'est pas uniforme, mais sa restriction à $\{ (i, k) \mid 1 \leqslant i, k \leqslant n \text{ et } i \neq k \}$ le sera.

c. Propriétés élémentaires

Proposition 1.2.8.
Soit (Ω, P) un espace probabilisé. Soit A et B deux événements. Alors on a

1. $P(\varnothing) = 0$;
2. $A \subset B \Rightarrow P(A) \leqslant P(B)$;
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
4. $P(A) = 1 - P(\bar{A})$.

Démonstration.
1. $\varnothing = \varnothing \sqcup \varnothing$ donc $P(\varnothing) = P(\varnothing) + P(\varnothing) = 0$, d'où le résultat.
2. Supposons $A \subset B$. Alors $B = A \cup (B \setminus A)$. De plus, A et $B \setminus A$ sont incompatibles, donc $P(B) = P(A) + P(B \setminus A) \geqslant P(A)$.
3. On a $A \cup B = A \cup (B \setminus A \cap B)$ et $B \setminus A \cap B$ et A sont incompatibles, donc $P(A \cup B) = P(A) + P(B \setminus (A \cap B))$. Or $B = (B \setminus (A \cap B)) \cup (A \cap B)$ et $B \setminus (A \cap B)$ et $A \cap B$ sont incompatibles, donc $P(B) = P(B \setminus (A \cap B)) + P(A \cap B)$. On en déduit le résultat.
4. $\Omega = A \cup \bar{A}$ et A et $\bar{A}$ sont incompatibles, donc $1 = P(\Omega) = P(A) + P(\bar{A})$ d'où le résultat.

□

Remarque 1.2.9.
La formule donnée en 3 se généralise en la formule du *crible de Poincaré*, comme pour le cardinal : pour des événements $A_1, \dots, A_n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^k \sum_{1 \leqslant i_1 < \dots < i_k \leqslant n} P(A_{i_1} \cap \dots \cap A_{i_k}).$$

Proposition 1.2.10.
Soit (Ω, P) un espace probabilisé. Soit $n \in \mathbb{N}$ et $A_1, \dots, A_n$ des événements *deux à deux incompatibles.* Alors la probabilité de leur union (appelée union disjointe) est la somme de leurs probabilités :

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

Démonstration.
Pour $k \in [\![0, n]\!]$, notons $E(k)$ la proposition

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i)$$

Montrons $\forall k \in [\![0, n]\!] \quad E(k)$ par récurrence :

— On a $P\left(\bigcup_{i=1}^{0} A_i\right) = P(\varnothing) = 0 = \sum_{i}^{0} P(A_i)$.

— Montrons $\forall k \in [\![0, n-1]\!] \quad P(k) \Rightarrow P(k+1)$. Soit $k \in [\![0, n-1]\!]$ vérifiant $P(k)$. Alors, comme les événements A_i pour $i \in [\![1, n]\!]$ sont deux à deux incompatibles, on a $\forall i \in 1, k \quad A_i \cap A_{k+1} = \varnothing$. Donc

$$\left(\bigcup_{i=1}^{k} A_i\right) \cap A_{k+1} = \varnothing$$

On a donc

$$P\left(\left(\bigcup_{i=1}^{k} A_i\right) \cup A_{k+1}\right) = P\left(\bigcup_{i=1}^{k} A_i\right) + P(A_{k+1})$$

Or on a $E(k)$, donc

$$P\left(\bigcup_{i=1}^{k+1} A_i\right) = \left(\sum_{i=1}^{k}\right) + P(A_{k+1})$$

On a donc $E(k+1)$.

On a donc $\forall k \in [\![0, n]\!] \quad E(k)$. □

Proposition 1.2.11 (Formule des probabilités totales, première forme)**.**
Soit (Ω, P) un espace probabilisé, $n \in \mathbb{N}^*$, $(A_i)_{i\in[\![1,n]\!]}$ un système complet d'événements et B un événement. Alors

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

Démonstration.
D'après la proposition 1.1.9, on a

$$B = \bigcup_{i=1}^{n} A_i \cap B.$$

et cette union est une union disjointe. Donc d'après la proposition 1.2.10, on a le résultat. □

d. Détermination par les images des singletons

Dans cette partie, on considère un univers fini $\Omega = \{\omega_1, \ldots, \omega_n\}$ de cardinal $n \in \mathbb{N}^*$.

Proposition 1.2.12.
Soit P une probabilité sur Ω. Alors pour tout événement A, on a

$$P(A) = \sum_{\omega \in A} P(\{\, \omega \,\}). \tag{1}$$

Démonstration.
Il suffit de remarquer que pour tout événement A, A est l'union disjointe des $\{\, \omega \,\}$ pour $\omega \in A$ et d'utiliser la proposition 1.2.10. □

Corollaire 1.2.13.
En particulier, des réels $p_1, \ldots, p_n$ étant donnés, il existe au plus une probabilité sur Ω vérifiant $\forall i \in [\![1, n]\!] \quad P(\{\, \omega_i \,\}) = p_i$.

Démonstration.
En effet, considérons deux probabilités P_1 et P_2 vérifiant cette condition. Alors, d'après la proposition, pour tout événement A, $P_1(A) = P_2(A)$. □

Remarque 1.2.14.
Remarquons que pour qu'une probabilité vérifiant cette condition existe, il est nécessaire que les p_i soit tous positifs ou nuls (car ce sont des probabilités) et que leur somme soit égale à 1 (car d'après l'égalité (1) c'est la probabilité de Ω). La proposition suivante montre que ces deux conditions sont suffisantes.

Proposition 1.2.15.
Soit $p_1, \ldots, p_n$, n réels positifs ou nuls de somme égale à 1. Alors il existe une (unique) fonction P de probabilité sur Ω telle que pour tout $i \in [\![1, n]\!]$, on a $P(\{\, \omega_i \,\}) = p_i$.

Démonstration.
On a déjà vu l'unicité sous réserve d'existence. Montrons l'existence. Notons P l'application de $\mathscr{P}(\Omega)$ dans $\mathbb{R}$ qui à tout événement A associe

$$P(A) = \sum_{i\in[\![1,n]\!],\omega_i\in A} p_i$$

Il est clair que pour tout $i \in [\![1, n]\!]$, on a $P(\{\, \omega_i \,\}) = p_i$. Montrons que P est une probabilité sur Ω.

1. Soit A un événement. Les p_i pour $i \in [\![1, n]\!]$ étant positifs ou nuls, on a pour tout A, $P(A) \geqslant 0$. De plus, $P(A)$ est inférieur ou égal à la somme des p_i, qui vaut 1, donc $0 \leqslant P(A) \leqslant 1$.
2. On a $P(\Omega) = \sum_{i=1}^{n} p_i = 1$.
3. Pour tout couple (A, B) d'événements incompatibles, comme $\{\, i \in [\![1, n]\!] \mid \omega_i \in A \cup B \,\}$ et la réunion disjointe de $\{\, i \in [\![1, n]\!] \mid \omega_i \in A \,\}$ et

$\{ i \in [\![1, n]\!] \mid \omega_i \in B \}$, on a

$$\begin{aligned} P(A \cup B) &= \sum_{i \in [\![1,n]\!], \omega_i \in A \cup B} p_i \\ &= \sum_{i \in [\![1,n]\!], \omega_i \in A} p_i + \sum_{i \in [\![1,n]\!], \omega_i \in B} p_i \\ &= P(A) + P(B). \end{aligned}$$

Ainsi, P est donc bien une probabilité. □

Exemple 1.2.16.
Nous pouvons maintenant définir des mesures de probabilités de la manière suivante : « la probabilité P est définie sur $[\![0, 5]\!]$ par

k	0	1	2	3	4	5
$P(\{k\})$	1/4	0	1/2	1/12	1/12	1/12

».

Exemple 1.2.17.
Étant donné des points $A_1, \dots, A_n$ dans un ensemble (fini) Ω, on définit la (mesure de) probabilité empirique par rapport à $A_1, \dots, A_n$ par

$$\forall x \in \Omega,\ P_n(\{x\}) = \frac{1}{n} \operatorname{Card} \{ i \in [\![1, n]\!] \mid A_i = x \}.$$

1.3. Probabilités conditionnelles

a. Définition

Le résultat d'une expérience aléatoire dépend parfois du résultat d'une autre. Par exemple, en prenant pour univers l'ensemble des jours de l'année 2016 muni de la probabilité uniforme P, si on appelle A l'événement «j'attrape un rhume aujourd'hui» et B l'événement «il fait un temps froid et humide aujourd'hui», on aimerait pouvoir exprimer que lorsque B est réalisé, A a une plus grande probabilité d'être réalisé.

Pour cela, on peut restreindre notre univers Ω à l'ensemble des jours où B est réalisé et regarder quelle est la probabilité de l'événement «attraper un rhume» dans cet univers restreint. On appellera cette probabilité la probabilité de A sachant B.

Dans l'univers Ω, l'événement A est l'ensemble des jours où j'attrape un rhume, B est l'ensemble des jours froids et humides. Notre univers restreint est B et l'événement «j'attrape un rhume» dans cet univers est l'ensemble $A \cap B$. Sa probabilité, si on le munit de la probabilité uniforme est $\operatorname{Card}(A \cap B) / \operatorname{Card} B = P(A \cap B)/P(B)$.

Cet exemple nous conduit à la définition suivante.

Définition 1.3.1 (Probabilité conditionnelle)**.**
Soit (Ω, P) un espace probabilisé et A et B deux événements, avec B de probabilité non nulle. On appelle probabilité de A sachant B, et on note $P_B(A)$ ou $P(A|B)$, le réel $\dfrac{P(A \cap B)}{P(B)}$.

Remarque 1.3.2.
Même si l'exemple donné concernait une probabilité uniforme, la définition donnée s'applique à toute probabilité.

Proposition 1.3.3.
Sous les hypothèses de la définition ci-dessus, l'application $P_B : \mathscr{P}(\Omega) \to \mathbb{R}$ est une probabilité sur Ω.

Démonstration.
En effet, elle est bien définie car $P(B) > 0$. Elle est à valeurs positives ou nulles et pour tout $A \in \mathscr{P}(\Omega)$, on a $A \cap B \subset B$ donc $P(A \cap B) \leqslant P(B)$ donc $P_B(A) \leqslant 1$. De plus $P_B(\Omega) = P(B)/P(B) = 1$. Enfin, pour tout couple (A, C) d'événements incompatibles, $A \cap B$ et $C \cap B$ sont incompatibles, donc

$$\begin{aligned} P_B(A \cup C) &= \frac{P((A \cap B) \cup (C \cap B))}{P(B)} \\ &= \frac{P(A \cap B) + P(C \cap B)}{P(B)} \\ &= P_B(A) + P_B(C) \end{aligned}$$

□

Remarque 1.3.4.
Dans les exercices de probabilités, l'un des points délicats (et donc intéressant) est souvent de traduire correctement l'énoncé en termes de probabilités conditionnelles. En effet, les énoncés ne sont pas toujours donnés de manière mathématisée et vous avez alors un (petit) travail de modélisation à effectuer. On pourra commencer par s'entraîner sur les exercices 1.3.11 et 1.3.12.

Exercice 1.3.5.
Dans une urne, on place deux boules blanches et une boule noire. On effectue un premier tirage dans l'urne, dans laquelle on remet la boule tirée en y rajoutant une boule de même couleur. On effectue un second tirage dans l'urne.

Modéliser (*i.e.* traduire l'énoncé mathématiquement, ici en termes de probabilités conditionnelles).

b. Probabilités composées, probabilités totales

Proposition 1.3.6 (Formule des probabilités composées)**.**
Soit A et B deux événements, avec $P(A) > 0$. Alors

$$P(A \cap B) = P(A) \times P(B|A).$$

Démonstration.
C'est une simple réécriture de la définition. □

Remarque 1.3.7.
On généralise cela au cas de plusieurs événements. Soit par exemple A, B, C et D quatre événements tels que $P(A \cap B \cap C \cap D) \neq 0$. Alors

$$P(A \cap B \cap C \cap D) = P(A) \times P(B|A) \\ \times P(C|A \cap B) \times P(D|A \cap B \cap C).$$

Exercice 1.3.8.
Soit $(A_1, \dots, A_n)$ des événements dont la probabilité de l'intersection est non nulle. Exprimer $P\left(\bigcap_{i=1}^{n} A_i\right)$ à l'aide de la formule des probabilités composées.

Proposition 1.3.9 (Formule des probabilités totales, deuxième forme)**.**
Soit $n \in \mathbb{N}^*$, $(A_i)_{i \in [\![1,n]\!]}$ un système complet d'événements de probabilités toutes non nulles et B un événement. Alors

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

Démonstration.
C'est une conséquence immédiate de la première forme (proposition 1.2.11) et de la formule des probabilités composées. □

Remarque 1.3.10.
On adopte souvent la convention suivante, fort utile dans l'utilisation de la formule des probabilités totales : si $P(A_i) = 0$, on pose $P(B|A_i) \times P(A_i) = 0$. Ainsi, la formule est valide pour tout système complet d'événements, ce qui peut éviter certaines contorsions particulièrement douloureuses ...

Attention : cette convention n'est pas au programme de la filière MP (mais elle l'est en PSI). Si vous voulez l'utiliser, rappelez la *clairement* avant. Ou mieux : revenez à la formule utilisant les intersections.

Exercice 1.3.11.
On considère une urne contenant quatre boules blanches et trois boules noires. On tire successivement et sans remise trois boules.

Calculer la probabilité de tirer exactement deux boules noires.

Note : il peut être intéressant de dessiner un arbre des possibilités pour raisonner. Mais un tel arbre n'est en *aucun cas* une justification.

Exercice 1.3.12.
On effectue $N \geqslant 2$ tirages successifs dans une urne, contenant initialement une boule blanche et une boule noire.

Chaque fois que l'on tire une boule blanche, on la remet et on rajoute une boule blanche supplémentaire dans l'urne.

Chaque fois que l'on tire une boule noire, on la remet dans l'urne.

On suppose que cette expérience est modélisée par un espace probabilisé fini (Ω, P). Quelle est la probabilité p_N d'obtenir la première boule noire au N^e tirage ?

Remarque 1.3.13. — Un des points attendus est de justifier rigoureusement l'utilisation d'une certaine formule ...

— L'exercice précédent est souvent modifié en « si l'on tire une boule noire, on s'arrête ». Cela change-t-il la modélisation ?
— L'année prochaine, vous pourrez modéliser cela en considérant une suite infinie de tirages. Comme nous ne considérons que des espaces probabilisés finis, nous nous limitons à N tirages successifs ... avec N quelconque.
— Que vaut $\displaystyle\sum_{n=1}^{+\infty} p_n$? Pouvez-vous l'interpréter, au moins intuitivement ?

c. Formule de Bayes

Exercice 1.3.14 (fondamental)**.**
On effectue un test de dépistage d'une maladie. Le test rend un résultat binaire : positif ou négatif. La probabilité que le test rende un résultat positif pour une personne si cette personne a contracté la maladie est appelé sensibilité du test et est notée p_1. La probabilité que le test rende un résultat négatif pour une personne qui n'a pas contracté cette maladie est appelée spécificité et est notée p_2.

Une personne prise au hasard dans la population française effectue le test et celui-ci rend un résultat positif. Quelle est la probabilité que cette personne soit malade ?

On donne : $p_1 = 0,99$, $p_2 = 0,98$, population française : $N = 6 \times 10^7$ personnes, nombre de personnes ayant contracté la maladie dans la population française : $m = 10^3$.

Proposition 1.3.15 (Formule de Bayes, cas particulier)**.**
Soit A et B deux événements tels que $P(A) > 0$ et $P(B) > 0$. Alors

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

Démonstration.
Il suffit de remplacer $P(A|B)$ et $P(B|A)$ par leurs définitions pour constater le résultat. □

Remarque 1.3.16.
Le théorème de Bayes dit, que pour deux événements A et B de probabilité non nulle, la probabilité d'avoir l'événement A, sachant qu'on a observé B (appelée probabilité *a posteriori*) est la probabilité d'avoir B sachant A multipliée par le rapport de la probabilité d'avoir A (probabilité *a priori*) et de la probabilité d'avoir B.

Proposition 1.3.17 (Formule de Bayes, cas général)**.**
Soit $(A_i)_{i\in[\![1,n]\!]}$ un système complet d'événements de probabilités toutes non nulles et B un événement de probabilité non nulle. Alors pour tout $j \in [\![1, n]\!]$, on a

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

Démonstration.
Soit $j \in [\![1, n]\!]$. Alors en utilisant la version précédente du théorème de Bayes, on obtient

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{P(B)}$$

Or d'après le théorème des probabilités totales, on a

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

□

Remarque 1.3.18.
Il est essentiel d'avoir compris ce que dit cette formule et d'être capable de la redémontrer rapidement.

Exercice 1.3.19.
Monsieur C. vient au lycée à pied, à cheval ou en voiture avec des probabilités respectives 9/100, 9/10 et 1/100. Quand il vient à pied, il met des chaussures de sport avec probabilité 9/10 ; à cheval, avec probabilité 5/10 et en voiture avec probabilité 1/10. Aujourd'hui, vous constatez qu'il a mis des chaussures de sport. Quelle est la probabilité qu'il soit venu à cheval ?

Remarque 1.3.20.
Remarquez l'abus de langage de l'exercice ci-dessus : en réalité, si on se place aujourd'hui et que Monsieur C. est déjà au lycée, ou bien il est venu à cheval ou bien il est venu par un autre moyen et parler de probabilité n'a plus de sens. En réalité, ce que signifie la formulation, c'est : « si on se place dans l'univers de tous les jours possibles, quelle est la probabilité que M. C. soit venu à cheval sachant qu'il a mis des chaussures de sport ». Cet abus de langage est typique de nombreux problèmes de probabilités et modéliser correctement le problème n'est pas toujours chose facile (voir le problème des trois portes).

1.4. Événements indépendants

a. Couple d'événements indépendants

Définition 1.4.1.
Soit A et B deux événements. On dit que A et B sont indépendants si et seulement si $P(A \cap B) = P(A) \times P(B)$.

Remarque 1.4.2.
1. Si $P(B) > 0$, alors cette condition est équivalente à $P(A|B) = P(A)$. Autrement dit, de façon informelle, savoir B ne modifie pas la probabilité de A.
2. Si $P(B) > 0$ et $P(\bar{B}) > 0$, elle est également équivalente à $P(A|B) = P(A|\bar{B})$.
3. Il arrive que l'on démontre l'indépendance de deux événements mais le plus souvent, il s'agit d'une hypothèse de modélisation du problème considéré.
4. Dans tout exercice de probabilités, il est primordiale de bien repérer dans l'énoncé les hypothèses d'indépendance.

Exercice 1.4.3.
Quels sont les événements indépendants d'eux-mêmes ?

Exemple 1.4.4.
Si on lance deux fois un dé à six faces et qu'on note A et B les événements «obtenir un 6» respectivement au premier et deuxième tirage, on aura tendance à modéliser le problème en disant que les deux événements sont indépendants ce qui correspond à l'intuition physique : le fait qu'on obtienne un 6 au deuxième tirage ne dépend pas du fait qu'on a obtenu un 6 au tirage précédent, le dé n'ayant pas de «mémoire» de ce qui s'est passé. Attention : même si le dé est pipé, il est raisonnable de considérer que les deux événements sont indépendants.

Si au lieu de considérer deux lancers d'un même dé, on considère plutôt le lancer simultané d'un dé rouge et d'un dé vert, il est encore raisonnable de penser que les deux événements sont indépendants, même si les dés sont pipés et même s'ils sont pipés de deux façons différentes. À moins par exemple que les dés soient aimantés, auquel cas les résultats des deux dés pourraient être reliés.

Exercice 1.4.5.
Le jeu de la roulette russe à deux joueurs consiste à placer une unique balle dans un revolver à 6 coups puis à faire tourner le barillet de façon aléatoire. Chacun à son tour, chaque joueur pointe le revolver sur sa propre tempe avant d'actionner la détente. La partie s'arrête dès le chien percute la cartouche (le perdant est celui qui tenait le revolver).

On note A l'événement «le premier joueur perd dès son premier essai», B l'événement «le deuxième joueur perd dès son premier essai».

Modéliser le problème et calculer $P(A)$ et $P(B)$. Peut-on raisonnablement penser que A et B sont indépendants ?

Variante : on fait tourner de nouveau le barillet à chaque tour. Calculer $P(A)$ et $P(B)$.

NB : Bien qu'il soit rare de développer une addiction à ce jeu, y jouer est fortement déconseillé.

Exercice 1.4.6.
On considère une urne contenant 10 boules noires et 10 boules blanches. On tire successivement deux boules, sans remise. On note A (resp. B) l'événement «la première (resp. seconde) est blanche».

Modéliser ce problème.

A et B sont-elles indépendantes ? Calculer $P(A)$ et $P(B)$. Que vaut $P(A|B)$? Que vaut $P(B|A)$?

Mêmes questions si on tire maintenant simultanément deux boules, l'une de la main gauche, l'autre de la main droite et qu'on note A (resp. B) l'événement «la boule tirée par la main gauche (resp. droite) est blanche».

b. Famille finie d'événements mutuellement indépendants

Définition 1.4.7.
Soit n un entier. On dit que des événements A_1, $\dots$, A_n sont *mutuellement indépendants* si pour toute partie I de $[\![1,n]\!]$, on a

$$P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i).$$

On dit qu'ils sont *deux à deux indépendants* si et seulement si pour tout $(i,j) \in [\![1,n]\!]^2$, avec $i \neq j$, A_i et A_j sont indépendants.

Remarque 1.4.8.
1. L'ordre des éléments n'a aucune importance.
2. L'indépendance mutuelle entraîne l'indépendance deux à deux.
3. La réciproque est fausse. Considérer par exemple deux tirages d'un dé et les événements «le premier tirage donne un nombre pair», «le second tirage donne un nombre pair» et «la somme des deux nombres obtenus est paire».
4. Il ne suffit pas de vérifier que la probabilité de l'intersection des A_i est égale au produit des $P(A_i)$ pour l'ensemble de tous les indices mais bien de le vérifier pour tous les ensembles d'indices possibles. Considérer par exemple l'univers $\Omega = [\![1,8]\!]$ des résultats possibles d'un dé équilibré à 8 faces, et les événements $A_1 = \{1,2,3,4\}$, $A_2 = \{1,2,3,4\}$ et $A_3 = \{1,5,6,7\}$.
5. Le plus souvent, l'indépendance mutuelle est une hypothèse faite lors de la modélisation du problème. Le problème, c'est que dans de nombreux énoncés, cette hypothèse n'est écrite nulle part : c'est le mathématicien qui analyse le problème[2] qui doit se poser la question de l'indépendance des événements.
6. Lorsqu'il s'agit de montrer l'indépendance mutuelle de plusieurs événements, il faut vérifier autant de conditions que de sous-ensembles de $[\![1,n]\!]$, soit 2^n, dont $n+1$ sont trivialement vérifiées (celles pour lesquelles I possède 0 ou 1 élément).

Proposition 1.4.9.
Remplacer, dans une famille d'événements mutuellement indépendants, certains événements par leurs contraires donne une nouvelle famille d'événements mutuellement indépendants. En d'autres termes, soit n un entier et A_1, $\dots$, A_n n événements indépendants. On se donne n événements B_1, $\dots$, B_n tels que pour tout $i \in [\![1,n]\!]$, B_i est égal à A_i ou à $\bar{A_i}$. Alors la famille $B_1, \dots, B_n$ est une famille d'événements indépendants.

Démonstration.
Il suffit de montrer le cas où $B_i = A_i$ pour $i = 1, \dots, n-1$ et où $B_n = \bar{A_n}$. En effet, on peut alors en déduire que si, dans une famille d'événements mutuellement indépendants, on change l'un des événements en son contraire, on obtient de nouveau une famille d'événements mutuellement indépendants. Par une récurrence immédiate, il vient alors que si on change p événements en leurs contraires, on obtient de nouveau une famille d'événements mutuellement indépendants.

Posons donc $B_i = A_i$ pour $i \in [\![1,n-1]\!]$ et $B_n = \bar{A_n}$.

Soit alors I une partie de $[\![1,n]\!]$, montrons qu'on a

$$P\left(\bigcap_{i\in I} B_i\right) = \prod_{i\in I} P(B_i)$$

Si I ne contient pas n, c'est évident.

Si I contient n, alors posons $J = I \setminus \{n\}$. On a succes-

2. Donc vous en particulier !

sivement :

$$\begin{aligned}
P\left(\bigcap_{i\in I} B_i\right) &= P\left(\bar{A_n} \cap \bigcap_{i\in J} A_i\right)\\
&= P\left((\Omega \setminus A_n) \cap \bigcap_{i\in J} A_i\right)\\
&= P\left(\bigcap_{i\in J} A_i \setminus \left(A_n \cap \bigcap_{i\in J} A_i\right)\right)\\
&= P\left(\bigcap_{i\in J} A_i\right) - P\left(\bigcap_{i\in I} A_i\right)\\
&= \prod_{i\in J} P(A_i) - \prod_{i\in I} P(A_i)\\
&= (1-P(A_n))\prod_{i\in J} P(A_i)\\
&= P(B_n)\prod_{i\in J} P(B_i)\\
&= \prod_{i\in I} P(B_i).
\end{aligned}$$

□

2. Variables aléatoires

2.1. Définitions

Définition 2.1.1.
Une *variable aléatoire* (v.a.) X est une application définie sur l'univers Ω à valeurs dans un ensemble E. Lorsque $E \subset \mathbb{R}$, la variable aléatoire est dite réelle. On appelle parfois *univers image* l'image directe de Ω par X.

Exemple 2.1.2.
Intuitivement, X représente une valeur associée à une expérience aléatoire : si l'on prend l'exemple du cas d'une personne jouant au loto, la valeur X, exprimée en euros, de son gain au loto lors du tirage qui aura lieu à une certaine date peut être modélisée par une variable aléatoire à valeur réelle (l'univers Ω étant l'ensemble des tirages de loto possible).

Remarque 2.1.3.
Une variable aléatoire modélise donc un « objet aléatoire ». Si l'on considère une matrice aléatoire, on manipulera donc des variables aléatoires à valeurs matricielles, tandis que si l'on considère des triangles aléatoires on manipulera des variables aléatoires à valeurs dans l'ensemble des triangles du plan.

Par exemple, si $X_1, \dots, X_n$ sont des variables aléatoires à valeurs dans un ensemble A, si P_n est la (mesure de) probabilité empirique par rapport à $X_1, \dots, X_n$, alors P_n est une variable aléatoire à valeurs dans l'ensemble des (mesures de) probabilités sur A.

Remarque 2.1.4.
En toute généralité, la définition de variable aléatoire est plus subtile. On se place ici dans un cadre très simplifié (univers fini).

Définition 2.1.5.
Soit X une variable aléatoire à valeurs dans un ensemble E.

Pour toute partie A de E, on note $\{ X \in A \}$ ou $(X \in A)$, (voire $[X \in A]$) l'événement $X^{-1}(A)$.

Si $E \subset \mathbb{R}$, et $x \in \mathbb{R}$, X est dite *réelle* et on note $(X = x)$, $(X \leqslant x)$, $(X < x)$, $(X \geqslant x)$, $(X > x)$ respectivement les événements $X^{-1}(\{ x \})$, $X^{-1}(]-\infty, x])$, ...

On note $P(X \in A)$, $P(X = x)$, $P(X \leqslant x)$, ... les probabilités de ces événements.

Exemple 2.1.6.
Pour reprendre l'exemple précédent, $(X \geqslant 1000)$ représente l'événement «le gain du joueur au loto est supérieur ou égal à mille euros» et $P(X \geqslant 1000)$ représente la probabilité de cet événement.

Proposition 2.1.7.
Soit X une variable aléatoire à valeurs dans un ensemble fini E. Alors $([X = x])_{x\in E}$ est un système complet d'événements.

Démonstration.
Soit $\omega \in \Omega$, alors $\omega \in [X = X(\omega)]$ donc $\cup_{x\in E}[X = x] = \Omega$.
Soit $(x, y) \in E^2$, avec $x \neq y$. Si $\omega \in [X = x] \cap [X = y]$, alors $X(\omega) = x = y$, ce qui est impossible, donc $[X = x] \cap [X = y] = \varnothing$. □

Remarque 2.1.8.
C'est souvent ce type de système complet d'événements que l'on utilisera.

2.2. Loi d'une variable aléatoire

Définition 2.2.1.
Soit X une variable aléatoire à valeurs dans un ensemble E.

On appelle *loi de la variable aléatoire* X la loi de probabilité $P_X : \mathscr{P}(E) \to \mathbb{R}$ qui à tout élément x de $X(\Omega)$ associe la probabilité de l'événement $X = x$:

$$\forall x \in X(\Omega) \quad P_X(\{x\}) = P(X = x).$$

Proposition 2.2.2.
Soit X une variable aléatoire définie sur un espace probabilisé fini (Ω, P). Alors $X(\Omega)$ est fini et P_X est une probabilité sur $X(\Omega)$.

Démonstration.
Par la caractérisation d'une loi par l'image de ses singletons, il suffit de remarquer que

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

□

Corollaire 2.2.3.
Si $A \subset X(\Omega)$, $P_X(A) = P(X \in A)$.

Remarque 2.2.4.
On peut donc définir une variable aléatoire en précisant la probabilité que cette v.a. soit égale à chaque élément de son image. Par exemple, on peut dire « la v.a. X à valeurs dans $[\![0, 5]\!]$ dont la loi est déterminée par

k	0	1	2	3	4	5
$P(X = k)$	1/4	0	1/2	1/12	1/12	1/12

».

Remarque 2.2.5.
Pour déterminer la loi d'une variable aléatoire X, on procédera *systématiquement* de la manière suivante :
• on détermine l'image de X ;
• pour chaque $k \in X(\Omega)$, on calcule $P(X = k)$;
• si on obtient une loi connue, on la nomme.

Exercice 2.2.6.
Soit X à valeurs dans $\{-1; 0; 1\}$ dont la loi est déterminée par

$$P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}.$$

Comment s'appelle la loi de X ? Quelle est la loi de $-X$? A-t-on $X = -X$?

Remarque 2.2.7.
La définition ci-dessus pose problème dans le cas où X est une variable aléatoire réelle continue (hors-programme mais vu en terminale) : la probabilité d'avoir $P(X = x)$ ne donne aucune information puisque pour tout x fixé, $P(X = x) = 0$.

C'est pourquoi on trouve parfois une autre définition de P_X : P_X est alors l'application de $\mathscr{P}(E)$ dans $\mathbb{R}$ qui, à une partie A de E, associe $P(X \in A)$. Dans ce cas, P_X a la propriété d'être une probabilité sur E (lorsque E est fini) et est déterminée de façon unique par les valeurs des $P(X = x)$ pour $x \in E$ puisqu'on a $P(X \in A) = \sum_{x \in A} P(X = x)$.

Cette deuxième définition pose également des problèmes dans le cas des variables continues mais ils sont plus facilement réparables (on ne peut plus définir P_X sur $\mathscr{P}(E)$ mais seulement sur une partie).

Définition 2.2.8.
Soit E et F deux ensembles. Soit X une variable aléatoire à valeurs dans un ensemble E. Soit $f : E \to F$. L'application $f \circ X : \Omega \to F$ est une variable aléatoire à valeurs dans F appelée *image de X par f* et parfois notée $f(X)$.

Remarque 2.2.9.
Le terme «image d'une variable aléatoire» vient

du fait que pour tout $\omega \in \Omega$, $f(X)(\omega)$ est l'image de $X(\omega)$ par f.

Exemple 2.2.10.
Si X est une v.a. réelle, on pourra considérer les v.a. réelles X^2, $|X|$ *etc.*

Proposition 2.2.11.
La loi associée à la variable aléatoire $f(X)$ introduite ci-dessus est

$$P_{f(X)} : \begin{array}{rcl} f(X(\Omega)) & \to & \mathbb{R} \\ y & \mapsto & P(X \in f^{-1}(\{ y \})) \end{array}.$$

Démonstration.
Il suffit de remarquer

$$\begin{aligned}(X \in f^{-1}(\{ y \})) &= \left\{ \omega \in \Omega \;\middle|\; X(\omega) \in f^{-1}(\{ y \}) \right\} \\ &= \{ \omega \in \Omega \mid f(X(\omega)) \in \{ y \} \} \\ &= \{ \omega \in \Omega \mid f(X)(\omega) \in \{ y \} \} \\ &= \left\{ \omega \in \Omega \;\middle|\; \omega \in (f(X))^{-1}(\{ y \}) \right\} \\ &= (f(X) = y).\end{aligned}$$

□

Exemple 2.2.12.
Soit X un v.a. à valeurs dans $\{-1, 0, 1\}$ telle que $P(X = -1) = P(X = 0) = P(X = 1) = \dfrac{1}{3}$. Déterminer les lois de X^2 et de X^3.

Définition 2.2.13 (Fonction de répartition)**.**
Soit X une variable aléatoire réelle. On appelle fonction de répartition de X la fonction

$$F_X : \begin{array}{rcl} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & P(X \leqslant x) \end{array}.$$

Exemple 2.2.14.
S'il faut retenir une chose, c'est le dessin de la fonction de répartition, figure 1.

Proposition 2.2.15.
La fonction de répartition d'une variable réelle X sur un univers Ω *fini* est une fonction en escalier. Plus précisément, Ω étant fini, $X(\Omega)$ est fini et s'écrit $\{ x_1, \ldots, x_n \}$ avec $n \geqslant 1$, où $x_1 < \ldots < x_n$. Alors F_X est constante sur les intervalles $]-\infty, x_1[$, $[x_1, x_2[$, $[x_2, x_3[$, $\ldots$, $[x_{n-1}, x_n[$, $[x_n, +\infty[$, prend pour valeur 0 sur le premier de ces intervalles, 1 sur le dernier et pour tout $i \in [\![1, n-1]\!]$ prend la valeur $\displaystyle\sum_{k=1}^{i} P(X = x_k)$ sur $[x_i, x_{i+1}[$.

Démonstration.
Il suffit de constater que pour tout $i \in [\![1, n-1]\!]$, et tout $t \in [x_i, x_{i+1}[$, l'événement $(X \leqslant t)$ n'est autre que l'union des événements deux à deux disjoints $(X = x_k)$ pour $k \leqslant i$, d'où

$$F_X(t) = \sum_{k=1}^{i} P(X = x_k).$$

De même on a $F_X(t) = 0$ pour $t \in]-\infty, x_1[$ et $F_X(t) = \displaystyle\sum_{k=1}^{n} P(X = x_k) = 1$ pour $t \in [x_n, +\infty[$. On en déduit que F_X est en escalier. □

Proposition 2.2.16.
La fonction de répartition d'une v.a. réelle est croissante et continue à droite. Elle a pour limite 0 en $-\infty$ et 1 en $+\infty$.

Remarque 2.2.17.
Cette propriété est vraie même pour les v.a. réelles définies sur un univers infini (programme de seconde année).

Démonstration.
C'est un corollaire immédiat de ce qui précède. □

Proposition 2.2.18.
La fonction de répartition caractérise la loi d'une variable X, au sens où pour tout réel t $P(X = t) > 0$ si et seulement si F_X n'est pas continue en t, et $P(X = t) = F_X(t) - \lim\limits_{t-} F_X$.

Démonstration.
Là encore, cela découle de ce qui précède. □

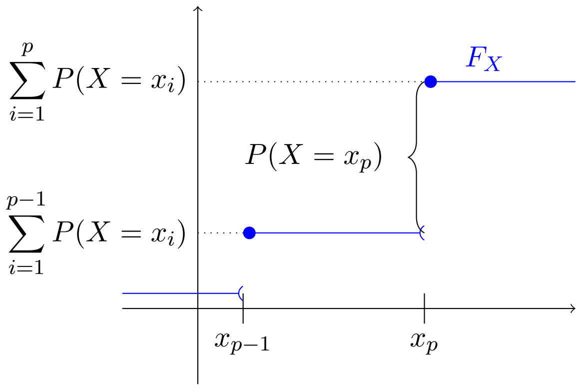

FIGURE 1 – Illustration de la fonction de répartition d'une variable aléatoire X à valeurs dans $\{x_1, \dots, x_n\}$ avec $x_1 < \dots < x_n$.

2.3. Loi usuelles

Dans cette partie, on étendra automatiquement les définitions données en commettant l'abus de notation suivant. Soit X une variable aléatoire X à valeurs dans E, $A \subset E$ tel que $\forall x \in E \backslash A, P(X=x)=0$ et $\forall a \in A, P(X=a)>0$ (on dit que A est le support de la loi de X). Formellement, $X(\Omega)=E$, mais on étendra les définitions suivantes comme si $X(\Omega)=A$.

Exemple 2.3.1.
Si $X : \Omega \to \{1,2,3,4\}$, avec $P(X=1)=P(X=2)=P(X=3)=\dfrac{1}{3}$ et $P(X=4)=0$, on s'autorisera à dire que X suit une loi uniforme sur $\{1,2,3\}$.

a. Loi uniforme

Définition 2.3.2.
Soit E un ensemble fini, non vide. On dit qu'une variable aléatoire réelle X suit *la loi uniforme* sur E et on note $X \hookrightarrow \mathscr{U}(E)$ (voire $X \equiv \mathscr{U}(E)$) si $X(\Omega)=E$ et P_X est la probabilité uniforme sur E (autrement dit, pour tout $x \in E$, $P(X=x)=1/\#E$).

En particulier pour tout couple (a,b) d'entiers relatifs avec $a \leqslant b$, on a $X \hookrightarrow \mathscr{U}(\llbracket a,b \rrbracket)$ si et seulement si

$$\forall x \in \llbracket a,b \rrbracket \quad P(X=x)=\frac{1}{b-a+1}.$$

Exemple 2.3.3. — La variable aléatoire modélisant le nombre obtenu par tirage d'un dé équilibré à 6 faces suit la loi uniforme sur $\llbracket 1,6 \rrbracket$.
— Pour modéliser le numéro d'une boule tirée dans une urne contenant n boules numérotées de 1 à n on prendra une variable aléatoire suivant la loi uniforme sur $\llbracket 1,n \rrbracket$.

Exercice 2.3.4.
Un joueur possède six dés : le dé n° 1 a 4 faces, le n° 2 en a 6, le n° 3 8, le n° 4 10, le n° 5 12 et le n° 6 20. Il lance un dé à 6 faces, obtient i, lance le dé n° i et note X le résultat obtenu.

Quelle est la loi de X ?

b. Loi de Bernoulli

Définition 2.3.5.
Soit $p \in [0,1]$. On dit qu'une variable aléatoire X est une variable de Bernoulli (ou suit la loi de Bernoulli) de paramètre p et on note $X \hookrightarrow \mathscr{B}(p)$ si X est à valeurs dans $\{0,1\}$ et $P(X=1)=p$.

Exemple 2.3.6. — Modélisons le tirage d'une pièce à pile ou face par la variable X valant 0 si l'on obtient pile et 1 si l'on obtient face. Si la pièce est supposée équilibrée, on supposera que X suit la loi uniforme sur $\{0,1\}$; que ce soit le cas ou non, il s'agit d'une variable de Bernoulli de paramètre la probabilité d'obtenir face.
— Si X est à valeurs dans $\{0,1\}$, alors $X \hookrightarrow \mathscr{B}(P(X=1))$.
— De manière générale, notons A un événement sur un espace probabilisé Ω. Alors χ_A,

la fonction indicatrice de A, définie par

$$\chi_A : \begin{cases} \Omega & \longrightarrow & \{ 0, 1 \} \\ \omega & \longmapsto & \begin{cases} 1 \text{ si } \omega \in A, \\ 0 \text{ si } \omega \notin A. \end{cases} \end{cases}$$

est une variable de Bernoulli de paramètre $P(A)$.

Très souvent, on s'intéressera à un événement A lors d'une expérience aléatoire et on dira que l'expérience est un succès si A est réalisé et un échec si A ne l'est pas. χ_A est alors la variable de Bernoulli prenant la valeur 1 en cas de succès et 0 en cas d'échec.

Exercice 2.3.7.
Montrer que toute variable aléatoire réelle définie sur un espace probabilisé fini peut s'écrire comme une combinaison linéaire de variables aléatoires suivant des lois de Bernoulli.

c. Loi binomiale

Définition 2.3.8.
Soit $n \in \mathbb{N}$ et $p \in [0, 1]$. On dit qu'une variable aléatoire X suit la loi binomiale de paramètres n et p et on note $X \hookrightarrow \mathscr{B}(n, p)$ si X est à valeurs dans $[\![0, n]\!]$ et pour tout $k \in [\![0, n]\!]$, on a

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

Exemple 2.3.9.
- Considérons n expériences aléatoires mutuellement indépendantes toutes de probabilité de succès p. On modélisera le nombre de succès par une variable binomiale de paramètres n et p. On verra dans la proposition 2.5.20 une justification mathématique à cette modélisation.
- En particulier considérons une urne opaque contenant des B boules blanches et N boules noires indiscernables au toucher ($B \in \mathbb{N}^*$, $N \in \mathbb{N}^*$), dans laquelle on tire n fois une boule, *avec* remise. Alors on modélisera le nombre de fois où l'on tire une boule noire par une variable aléatoire suivant la loi binomiale de paramètres n et $N/(B + N)$.

Exercice 2.3.10.
On considère un joueur jouant au jeu suivant :
- Il mise un euro. Cette mise est définitivement perdue.
- Il lance quatre pièces de monnaie.
- Si exactement trois des pièces tombent sur pile, il perçoit un euro. Si les quatre pièces tombent sur pile, il perçoit dix euros.

On note X le gain du joueur (mise incluse, le gain peut donc être négatif). Donner la loi de X.

2.4. Couples de variables aléatoires

Dans cette partie, on considère Ω un espace probabilisé fini et deux variables aléatoires X et Y.

On écrira $X(\Omega)$ sous la forme $\{ x_1, \dots, x_N \}$ et $Y(\Omega)$ sous la forme $\{ y_1, \dots, y_P \}$ où N et P sont des entiers.

Remarque 2.4.1.
On peut considérer, en commettant un léger abus de notation, que (X, Y) est une variable aléatoire, à valeurs dans $\{ x_1, \dots, x_N \} \times \{ y_1, \dots, y_P \}$.

Proposition 2.4.2.
La famille $((X = x_i) \cap (Y = y_j))_{(i,j) \in [\![1,N]\!] \times [\![1,P]\!]}$ est un système complet d'événements.

Démonstration.
Soit $(i, j) \in [\![1, N]\!] \times [\![1, P]\!]$, il suffit de remarquer que $[(X, Y) = (x_i, y_j)] = [X = x_i] \cap [Y = y_j]$ et d'utiliser la proposition 2.1.7. □

Remarque 2.4.3.
On notera souvent l'événement $[X = x] \cap [Y = y]$ par $[X = x, Y = y]$.

Définition 2.4.4 (Loi conjointe)**.**
On appelle *loi conjointe de X et Y* la loi du couple (X, Y), soit la loi $P_{X,Y} : \mathscr{P}(X(\Omega) \times Y(\Omega)) \to \mathbb{R}$ vérifiant $\forall (x, y) \in X(\Omega) \times Y(\Omega)$, $P_{X,Y}(\{(x, y)\}) = P((X = x) \cap (Y = y)) = P(X = x, Y = y)$.

Exercice 2.4.5.
On tire un dé équilibré à six faces et on lance une pièce de monnaie équilibrée. On note X la variable de Bernoulli associée à l'événement «obtenir pile» et Y la valeur tirée sur le dé.

Donner la loi conjointe de X et Y.

Exercice 2.4.6.
On tire deux dés équilibrés à quatre faces, un vert et un rouge. On appelle X la valeur obtenue sur le dé vert, Y la valeur obtenue sur le dé rouge et Z la somme des deux. Donner la loi conjointe de X et Y puis de X et Z.

Définition 2.4.7.
On appelle *première (resp. seconde) loi marginale du couple* (X, Y) la loi de X (resp. de Y).

Proposition 2.4.8.
Les lois marginales du couple (X, Y) sont déterminées de façon unique par la loi du couple $P_{X,Y}$. Plus précisément, on a pour tout $x \in X(\Omega)$

$$\begin{aligned} P(X = x) = P_X(\{x\}) &= \sum_{y \in Y(\Omega)} P_{X,Y}(\{(x, y)\}) \\ &= \sum_{y \in Y(\Omega)} P(X = x, Y = y). \end{aligned}$$

Symétriquement, pour tout $y \in Y(\Omega)$, on a

$$\begin{aligned} P(Y = y) = P_Y(\{y\}) &= \sum_{x \in X(\Omega)} P_{X,Y}(\{(x, y)\}) \\ &= \sum_{x \in X(\Omega)} P(X = x, Y = y). \end{aligned}$$

En revanche, les lois marginales ne suffisent pas à déterminer la loi du couple.

Démonstration.
Pour le premier point, considérons $x \in X$ et remarquons alors qu'on a

$$(X = x) = \bigcup_{y \in Y(\Omega)} (X = x) \cap (Y = y).$$

Or cette union est une union disjointe, d'où l'égalité

$$P(X = x) = \sum_{y \in Y(\Omega)} P((X = x) \cap (Y = y)).$$

D'où le premier point. Autrement dit : la famille des $(Y = y)$ pour $y \in Y(\Omega)$ constitue un système complet d'événements et le résultat se déduit immédiatement de la formule des probabilités totales.

Pour le second point, il suffit de montrer que deux couples peuvent avoir des lois différentes bien qu'ayant les mêmes lois marginales. On peut par exemple considérer, sur l'univers $\Omega = \{ 1, 2 \}$ muni de la probabilité uniforme, les variables de Bernoulli X et Y définies par $X(1) = 1$, $X(2) = 0$, $Y(1) = 0$ et $Y(2) = 1$. Alors les variables X et Y ont même loi, les deux couples de variables aléatoires (X, X) et (X, Y) ont donc mêmes lois marginales. Cependant, ils ont des lois conjointes différentes puisque $P_{X,X}(1, 1) = \frac{1}{2}$ et $P_{X,Y}(1, 1) = 0$. □

Exemple 2.4.9.
On lance deux dés à 6 faces et l'on note X et Y le résultat de chacun. Alors (après avoir modélisé cela), les lois jointes de (X, X) et de (X, Y) sont différentes alors que (X, X) et (X, Y) ont les mêmes lois marginales.

Définition 2.4.10 (Loi conditionnelle)**.**
Soit X et Y deux variables aléatoires respectivement à valeurs dans des ensembles E et F. Soit $x \in X(\Omega)$ vérifiant $P(x) \neq 0$. Alors on appelle loi conditionnelle de Y sachant $(X = x)$ la loi de la variable Y sur l'univers Ω muni de la probabilité $P_{X=x}$. C'est donc la fonction $P_{Y|X=x} : \mathscr{P}(Y(\Omega)) \to \mathbb{R}$ vérifiant :

$$\forall y \in Y(\Omega),\ P_{Y|X=x}(y) = P(Y = y|X = x).$$

Remarque 2.4.11.
On définit de même, pour $y \in Y(\Omega)$ vérifiant $P(y) \neq 0$, la loi conditionnelle de X sachant $(Y = y)$.

Remarque 2.4.12.
La loi de Y conditionnellement à $X = x$ ne dépend que de la loi jointe de X et de Y :

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{\sum_{z \in Y(\Omega)} P(X = x, Y = z)}.$$

On peut généraliser de la même manière ces notions sur les couples de variables aléatoires et définir la loi conjointe d'un n-uplet de variables aléatoires, les n lois marginales de ce n-uplet ainsi que la loi conditionnelle d'une variable, par exemple X_n, sachant $X_1 = x_1$, … et $X_{n-1} = x_{n-1}$.

2.5. Variables aléatoires indépendantes

Dans ces parties, sauf mention expresse du contraire, X et Y sont des variables aléatoires à valeurs respectivement dans des ensembles E et F.

Définition 2.5.1.
On dit que X et Y sont indépendantes si pour tout $x \in E$ et tout $y \in F$ on a

$$P([X = x] \cap [Y = y]) = P(X = x)P(Y = y). \quad (2)$$

On dira que deux variables aléatoires sont *indépendantes et identiquement distribuées* (i.i.d.) si elles sont indépendantes et de même loi.

Remarque 2.5.2.
On écrira souvent cela comme

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

Deux v.a. sont donc indépendantes si et seulement si leur loi jointe se factorise en le produit de leurs lois.

Exemple 2.5.3.
Si l'on tire deux dés ou si l'on effectue deux tirages avec remise dans une urne, on modélisera les deux résultats comme des v.a. indépendantes.

Exercice 2.5.4.
On effectue deux tirages sans remise dans une urne contenant n boules. Modéliser. Les numéros tirés sont-ils indépendants ?

Proposition 2.5.5.
X et Y sont indépendantes si et seulement si pour tout $A \subset E$ et tout $B \subset F$, on a

$$P((X, Y) \in A \times B) = P(X \in A)P(Y \in B). \quad (3)$$

Démonstration.
Pour le sens indirect, c'est-à-dire pour montrer l'égalité (2), sous l'hypothèse que l'égalité (3) est vérifiée pour tout A et tout B, il suffit de choisir $A = \{ x \}$ et $B = \{ y \}$.

Montrons le sens direct : supposons donc qu'on a (2) pour tout $(x, y) \in E \times F$ et montrons que pour tout $(A, B) \in \mathscr{P}(E) \times \mathscr{P}(F)$, on a (3). Quitte à remplacer A par $A \cap X(\Omega)$ et B par $B \cap Y(\Omega)$, on peut supposer que A et B sont des ensembles finis. On a alors

$$\begin{aligned}
[X \in A] &= \bigcup_{x \in A} [X = x] \\
[Y \in B] &= \bigcup_{y \in B} [Y = y] \\
[(X, Y) \in A \times B] &= \bigcup_{(x,y) \in A \times B} [X = x] \cap [Y = y]
\end{aligned}$$

Or ces trois unions sont des unions disjointes, donc :

$$\begin{aligned}
P((X, Y) \in A \times B) &= P\left(\bigcup_{(x,y) \in A \times B} [X = x] \cap [Y = y] \right) \\
&= \sum_{(x,y) \in A \times B} P([X = x] \cap [Y = y]) \\
&= \sum_{(x,y) \in A \times B} P(X = x) \times P(Y = y) \\
&= \left(\sum_{x \in A} P(X = x) \right) \times \left(\sum_{y \in B} P(Y = y) \right) \\
&= P\left(\bigcup_{x \in A} [X = x] \right) \times P\left(\bigcup_{y \in B} [Y = y] \right) \\
&= P(X \in A) \times P(Y \in B).
\end{aligned}$$

□

Exercice 2.5.6.
Soit X, Y indépendantes de loi uniforme sur $[\![1, n]\!]$. Calculer $P(X \leqslant Y)$.

Proposition 2.5.7.
Soit E, E', F et F' quatre ensembles. Soit $f : E \to E'$ et $g : F \to F'$. Soit alors X et Y deux variables aléatoires indépendantes. Alors $f(X)$ et $g(Y)$ sont des variables aléatoires indépendantes.

Démonstration.
Soit $A' \subset E'$ et $B' \subset F'$. D'après la proposition qui précède, il suffit de montrer

$$P\big((f(X), g(Y)) \in A' \times B'\big) = P(f(X) \in A') \times P(f(Y) \in B').$$

Or on a

$$\big[(f(X), g(Y)) \in A' \times B'\big] = \big[f(X) \in A'\big] \cap \big[g(Y) \in B'\big].$$

Et d'autre part, on a

$$\begin{aligned}\big[f(X) \in A'\big] &= \left\{ \omega \in \Omega \;\middle|\; f(X(\omega)) \in A' \right\} \\ &= \left\{ \omega \in \Omega \;\middle|\; X(\omega) \in f^{-1}(A') \right\} \\ &= \big[X \in f^{-1}(A')\big]\end{aligned}$$

et de la même façon, on obtient :

$$\big[g(Y) \in B'\big] = \big[Y \in g^{-1}(B')\big].$$

On en déduit

$$\begin{aligned}&\big[(f(X), g(Y)) \in A' \times B'\big] \\ &= \big[f(X) \in A'\big] \cap \big[g(Y) \in B'\big] \\ &= \big[X \in f^{-1}(A')\big] \cap \big[Y \in g^{-1}(B')\big] \\ &= \big[(X, Y) \in f^{-1}(A') \times g^{-1}(B')\big].\end{aligned}$$

Puis :

$$\begin{aligned}&P\big((f(X), g(Y)) \in A' \times B'\big) \\ =&P\big((X, Y) \in f^{-1}(A') \times g^{-1}(B')\big) \\ =&P\big(X \in f^{-1}(A')\big) \times P\big(Y \in f^{-1}(B')\big) \\ =&P\big(f(X) \in A'\big) \times P\big(f(Y) \in B'\big),\end{aligned}$$

qui est ce qu'on voulait démontrer. □

Définition 2.5.8.
Soit n un entier et X_1, ..., X_n des variables aléatoires à valeurs dans des ensembles respectifs E_1, ..., E_n. On dit que les variables X_1, ..., X_n sont mutuellement indépendantes si pour tous x_1, ..., x_n appartenant respectivement à X_1, ..., X_n, on a

$$P\left(\bigcap_{i=1}^{n} (X_i = x_i)\right) = \prod_{i=1}^{n} P(X_i = x_i). \qquad (4)$$

On dira que des variables aléatoires sont *indépendantes et identiquement distribuées* (i.i.d.) si elles sont mutuellement indépendantes et de même loi.

Remarque 2.5.9.
On montrera bientôt que des variables aléatoires mutuellement indépendantes le sont deux à deux.

Comme le montre l'exemple 2.5.10, des v.a. indépendantes deux à deux ne le sont pas forcément mutuellement.

Exemple 2.5.10.
Soit X, Y i.i.d. de loi de Rademacher : $P(X = 1) = P(X = -1) = \dfrac{1}{2}$, soit $Z = XY$. Montrer que X, Y et Z sont indépendantes deux à deux, mais pas mutuellement.

Proposition 2.5.11.
Les variables aléatoires X_1, ..., X_n à valeurs dans des ensembles respectifs E_1, ..., E_n sont mutuellement indépendantes si et seulement pour tous sous ensembles respectifs A_1, ..., A_n de E_1, ..., E_n, on a

$$P\left(\bigcap_{i=1}^{n} (X_i \in A_i)\right) = \prod_{i=1}^{n} P(X_i \in A_i). \qquad (5)$$

Démonstration.
La démonstration est similaire à celle du cas de deux variables.

Pour le sens indirect, c'est-à-dire pour montrer l'égalité (4), sous l'hypothèse que l'égalité (5) est vérifiée pour tous ensembles A_1, ..., A_n, il suffit de choisir $A_i = \{ x_i \}$ pour $i = 1, \dots n$.

Montrons le sens direct : supposons donc qu'on a (4) pour tous $x_1, \dots, x_n$ appartenant respectivement à E_1, ..., E_n et montrons que pour tous sous ensembles respectifs A_1, ... A_n de E_1, ..., E_n, on a (5). Quitte à remplacer chaque A_i par $A_i \cap X_i(\Omega)$, on peut supposer que les A_i sont des ensembles finis. On a alors, si $1 \leqslant i \leqslant n$,

$$(X_i \in A_i) = \bigcup_{x_i \in A_i} (X_i = x_i).$$

Ainsi,

$$\begin{aligned}P\left(\bigcap_{i=1}^{n} (X_i \in A_i)\right) =& P\left(\bigcap_{i=1}^{n} \bigcup_{x_i \in A_i} (X_i = x_i)\right) \\ =& P\left(\bigcup_{(x_1, \dots, x_n) \in A_1 \times \dots \times A_n} \bigcap_{i=1}^{n} (X_i = x_i)\right).\end{aligned}$$

Or, cette réunion est disjointe, donc

$$P\left(\bigcap_{i=1}^{n} (X_i \in A_i)\right) = \sum_{(x_1, \dots, x_n) \in A_1 \times \dots \times A_n} P\left(\bigcap_{i=1}^{n} (X_i = x_i)\right).$$

Par indépendance mutuelle,

$$\begin{aligned} P\left(\bigcap_{i=1}^{n}(X_i \in A_i)\right) &= \sum_{(x_1,\dots,x_n)\in A_1\times\dots\times A_n} \prod_{i=1}^{n} P(X_i = x_i) \\ &= \prod_{i=1}^{n} \sum_{x_i \in A_i} P(X_i = x_i) \\ &= \prod_{i=1}^{n} P(X_i \in A_i), \end{aligned}$$

d'où le résultat. □

Remarque 2.5.12.
Notez la différence entre la définition d'événements mutuellement indépendants et celle de variables mutuellement indépendantes : dans ce dernier cas, la définition ne demande pas de regarder pour tous les sous-ensembles de $[\![1, n]\!]$, pour une bonne raison : le résultat suivant assure que si l'égalité (4) est vérifiée, alors elle est vraie aussi si l'on effectue le produit et l'intersection seulement pour un sous-ensemble de $[\![1, n]\!]$.

Proposition 2.5.13.
Toute sous-famille d'une famille de variables aléatoires mutuellement indépendantes est constituée de variables mutuellement indépendantes : soit $X_1, \dots, X_n$ des variables aléatoires mutuellement indépendantes et $I \subset [\![1, n]\!]$; alors les X_i pour $i \in I$ sont mutuellement indépendantes.

Démonstration.
D'après la proposition 2.5.11, il suffit de montrer que pour toute famille d'événements A_i pour $i \in I$, on a

$$P\left(\bigcap_{i\in I}(X_i \in A_i)\right) = \prod_{i\in I} P(X_i \in A_i).$$

Considérons donc une telle famille quelconque, et posons $A_i = E_i$ pour $i \in [\![1, n]\!] \setminus I$. Alors, pour tout $i \in [\![1, n]\!] \setminus I$, l'événement $X_i \in A_i$ est certain, c'est-à-dire est égal à Ω. Donc

$$\begin{aligned} P\left(\bigcap_{i\in I}(X_i \in A_i)\right) &= P\left(\bigcap_{i=1}^{n}(X_i \in A_i)\right) \\ &= \prod_{i=1}^{n} P(X_i \in A_i) \\ &= \prod_{i\in I} P(X_i \in A_i). \end{aligned}$$

□

Corollaire 2.5.14 (Lemme des coalitions)**.**
Soit $(X_1, \dots, X_n)$ des variables aléatoires mutuellement indépendantes, I et J deux sous-ensembles disjoints de $\{1, \dots, n\}$.

Alors les variables aléatoires $(X_i)_{i\in I}$ et $(X_j)_{j\in J}$ sont indépendantes.

Démonstration.
Direct, par indépendance mutuelle de la famille $(X_k)_{k\in I\cup J}$. □

Remarque 2.5.15.
Ce lemme se généralise directement à l'indépendance mutuelle de m « paquets » de variables aléatoires pris sur m sous-ensembles disjoints deux à deux de $\{1, \dots, n\}$.

Proposition 2.5.16.
Si $X_1, \dots, X_n$ sont des variables aléatoires mutuellement indépendantes et $f_1, \dots, f_n$ sont des fonctions définies sur $X_1(\Omega), \dots, X_n(\Omega)$, alors $f_1(X_1), \dots, f_n(X_n)$ sont mutuellement indépendantes.

Démonstration.
Exactement comme pour deux v.a. □

Exemple 2.5.17.
Si X, Y, Z, T sont quatre v.a. réelles mutuellement indépendantes, alors $(X, Y), Z, T$ sont aussi mutuellement indépendantes, tout comme $X + Y, \mathrm{e}^Z$ et T^2.

Proposition 2.5.18.
Soit $n \in \mathbb{N}$, $A_1, \dots, A_n$ n événements et $X_1, \dots, X_n$ les variables de Bernoulli respectivement associées à ces événements. Alors les événements $A_1, \dots, A_n$ sont indépendants si et seulement si les variables aléatoires $X_1, \dots, X_n$ sont indépendantes.

Démonstration.
Supposons que les événements $A_1, \dots, A_n$ sont indépendants. Soit alors $x_1, \dots, x_n$ n éléments de $\{0, 1\}$. Notons B_i l'événement $X_i = x_i$ pour $i \in [\![1, n]\!]$. Alors pour tout i, on a $B_i = A_i$ ou $B_i = \bar{A}_i$. Donc les événements B_1,

..., B_n sont mutuellement indépendants. On en déduit successivement :

$$\begin{aligned} P\left(\bigcap_{i=1}^{n}(X_i = x_i)\right) &= P\left(\bigcap_{i=1}^{n} B_i\right) \\ &= \prod_{i}^{n} P(B_i) \\ &= \prod_{i}^{n} P(X_i = x_i). \end{aligned}$$

Ainsi, les variables aléatoires X_1, ..., X_n sont mutuellement indépendantes.

Réciproquement, supposons que les variables aléatoires X_1,...,X_n sont mutuellement indépendantes et montrons que les événements $A_1, \dots, A_n$ le sont aussi. Soit $I \subset [\![1, n]\!]$. Alors les X_i pour $i \in I$ sont mutuellement indépendantes, en particulier, on a

$$P\left(\bigcap_{i\in I}(X_i = 1)\right) = \prod_{i\in I} P(X_i = 1)$$

Or pour tout $i \in I$, l'événement $(X_i = 1)$ n'est autre que A_i. On en déduit donc le résultat. $\square$

Remarque 2.5.19.
Si $X_1, \dots, X_n$ sont des variables aléatoires suivant des loi de Bernoulli, alors

$$\sum_{i=1}^{n} X_i = \operatorname{Card}\{\, i \in [\![1, n]\!] \mid X_i = 1 \,\}.$$

On remarquera, sans s'émouvoir, que $\{\, i \in [\![1, n]\!] \mid X_i = 1 \,\}$ est une variable aléatoire à valeurs dans $\mathscr{P}([\![1, n]\!])$.

Proposition 2.5.20.
Soit $p \in [0, 1]$ et $n \in \mathbb{N}$. Soit X_1, ...,X_n n variables aléatoires suivant toutes la loi de Bernoulli de paramètre p et mutuellement indépendantes. Alors la variable X à valeurs dans $[\![0, n]\!]$ définie par

$$X = X_1 + \ldots + X_n$$

suit la loi binomiale de paramètres n et p.

Démonstration.
On a, pour tout $i \in [\![1, n]\!]$, $X_i(\Omega) = \{\, 0, 1 \,\}$, donc $X(\Omega) \subset [\![0, n]\!]$.

Posons $E = \{\, 0, 1 \,\}^n$, et pour tout élément $(x_1, \dots, x_n)$ de E, on note $A_{(x_1,\dots,x_n)}$ l'événement $\bigcap_{i=1}^{n}(X_i = x_i)$. La famille $(A_x)_{x\in E}$ est un système complet d'événements. Donc on a, pour tout $k \in [\![0, n]\!]$

$$\begin{aligned} (X = k) &= \bigcup_{x\in E} (X = k) \cap A_x \\ &= \bigcup_{\substack{(x_1,\dots,x_n)\in E \\ x_1+\cdots+x_n=k}} A_{(x_1,\dots,x_n)}. \end{aligned}$$

Or les A_x forment un système complet d'événements. On a donc

$$\begin{aligned} P(X = k) &= \sum_{(x_1,\dots,x_n)\in E} P(A_{(x_1,\dots,x_n)} \cap [X = k]) \\ &= \sum_{\substack{(x_1,\dots,x_n)\in E \\ x_1+\cdots+x_n=k}} P(A_{(x_1,\dots,x_n)}). \end{aligned}$$

Or, les X_i étant mutuellement indépendantes, pour tout $(x_1, \dots, x_n) \in E$, on a

$$P(A_{(x_1,\dots,x_n)}) = \prod_{i=1}^{n} P(X_i = x_i).$$

Or pour tout $i \in [\![1, n]\!]$, X_i est une variable de Bernoulli de paramètre p, donc si $x_i = 1$, $P(X_i = x_i)$ vaut p et si $x_i = 0$, $P(X_i = x_i)$ vaut $(1 - p)$. Donc si k est le nombre de membres du n-uplet $(x_1, \dots, x_n)$ valant 1, on a

$$P(A_{(x_1,\dots,x_n)}) = p^k(1 - p)^{n-k}$$

D'où

$$\begin{aligned} P(X = k) &= \sum_{\substack{(x_1,\dots,x_n)\in E \\ x_1+\cdots+x_n=k}} p^k(1 - p)^{n-k} \\ &= \binom{n}{k} p^k(1 - p)^{n-k}. \end{aligned}$$

On peut aussi le montrer par récurrence sur n. En notant $S_n = \sum_{k=1}^{n} X_k$, alors naturellement $S_1 \hookrightarrow \mathscr{B}(1, p)$.

Supposons que $S_{n-1} \hookrightarrow \mathscr{B}(n - 1, p)$. Alors X_n est indépendante de $(X_1, \dots, X_{n-1})$ donc de S_{n-1}. Il ne reste plus qu'à montrer que si $Y \hookrightarrow \mathscr{B}(n - 1, p)$ est indépendante de X_n, alors $S = X_n + Y \hookrightarrow \mathscr{B}(n, p)$.

Comme $X_1(\Omega) = \{0, 1\}$ et $Y(\Omega) = [\![0, n - 1]\!]$, alors $S(\Omega) \subset [\![0, n]\!]$. Soit $k \in [\![0, n]\!]$, calculons $P(S = k)$.

— Si $k = 0$, $P(S = 0) = P([X_n = 0] \cap [Y = 0])$ et donc par indépendance de X_n et de Y, $P(S = 0) = P(X_n = 0)P(S = 0) = (1-p)\cdot(1-p)^{n-1} = (1-p)^n$.
— De même, si $k = n$, $P(S = n) = P([X_n = 1] \cap [Y = n - 1])$ et donc $P(S = n) = P(X_n = 1)P(S = n - 1) = p \cdot p^{n-1} = p^n$.
— Si $0 < k < n$, alors $([X_n = 0], [X_n = 1])$ est un système complet d'événements, donc par la formule

des probabilités totales

$$\begin{aligned} P(S=k) &= P([S=k]\cap[X_n=0]) \\ &\quad + P([S=k]\cap[X_n=1]) \\ &= P([Y=k]\cap[X_n=0]) \\ &\quad + P([Y=k-1]\cap[X_n=1]) \\ \text{ind.} &= P(Y=k)P(X_n=0) \\ &\quad + P(Y=k-1)P(X_n=1) \\ &= \binom{n-1}{k} p^k(1-p)^{n-1-k}\cdot(1-p) \\ &\quad + \binom{n-1}{k-1} p^{k-1}(1-p)^{n-1-(k-1)}\cdot p \\ &= p^k(1-p)^{n-k}\left[\binom{n-1}{k}+\binom{n-1}{k-1}\right] \\ &= p^k(1-p)^{n-k}\binom{n}{k} \end{aligned}$$

par la formule du triangle de Pascal, ce qui permet de conclure.

□

Remarque 2.5.21.
Dans le dernier calcul, on pouvait aussi écrire

$$\begin{aligned} P(S=k) &= P(S=k|X=0)P(X=0) \\ &\quad + P(S=k|X=1)P(X=1) \\ &= P(Y=k|X=0)P(X=0) \\ &\quad + P(Y=k-1|X=1)P(X=1) \end{aligned}$$

puis de voir, par indépendance de X et de Y, que $P(Y=k|X=0)=P(Y=k)$ ainsi que $P(Y=k-1|X=1)=P(Y=k-1)$.

Nous n'avons fait que démontrer ceci.

2.6. Espérance

Définition 2.6.1.
Soit X une variable aléatoire réelle. On appelle espérance de X et on note $\mathrm{E}(X)$ la somme

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

On dit qu'une variable aléatoire X est *centrée* si son espérance est nulle.

Remarque 2.6.2.
1. L'espérance est la moyenne des valeurs prises par X, pondérées par leurs probabilités.
2. L'espérance de X ne dépend que de la loi de X, donc deux variables ayant même loi ont même espérance.

Exercice 2.6.3.
On lance deux dés et on note X la somme des deux résultats. Calculer l'espérance de X.

Proposition 2.6.4.
Soit X une variable aléatoire réelle. Alors on a

$$\mathrm{E}(X) = \sum_{\omega\in\Omega} P(\{\,\omega\,\})X(\omega).$$

Démonstration.
Il suffit de remarquer que pour tout $x\in X(\Omega)$, on a $\{\,X=x\,\}=\{\,\omega\in\Omega \mid X(\omega)=x\,\}$, donc

$$P(X=x) = \sum_{\substack{\omega\in\Omega\\ X(\omega)=x}} P(\{\,\omega\,\})$$

On a alors successivement :

$$\begin{aligned} \mathrm{E}(X) &= \sum_{x\in X(\Omega)}\left(\sum_{\substack{\omega\in\Omega\\ X(\omega)=x}} P(\{\,\omega\,\})\right)x \\ &= \sum_{x\in X(\Omega)}\left(\sum_{\substack{\omega\in\Omega\\ X(\omega)=x}} P(\{\,\omega\,\})X(\omega)\right) \\ &= \sum_{\omega\in\Omega} P(\{\,\omega\,\})X(\omega) \end{aligned}$$

□

Proposition 2.6.5.
L'espérance est linéaire et positive, donc croissante. Plus précisément, soit X et Y deux variables aléatoires réelles et α et β deux réels.

Alors

$$\mathrm{E}(\alpha X+\beta Y) = \alpha\mathrm{E}(X)+\beta\mathrm{E}(Y)$$

De plus si X est presque sûrement à valeurs positives ($P(X\geqslant 0)=1$) alors $\mathrm{E}(X)\geqslant 0$.

Enfin, si l'événement $X\leqslant Y$ est presque sûr, alors $\mathrm{E}(X)\leqslant\mathrm{E}(Y)$.

Démonstration.
Les deux premiers points se déduisent immédiatement de la proposition 2.6.4 (pour le second on utilise que pour $x < 0$, on a $P(X = x) \leqslant P(X < 0) = 1 - P(X \geqslant 0) = 0$).

Le troisième se déduit du fait que $\mathrm{E}(Y) = \mathrm{E}(Y - X) + \mathrm{E}(X)$ et que $P(Y - X \leqslant 0) = P(X \leqslant Y) = 1$. □

Remarque 2.6.6.
La propriété précédente s'étend naturellement, par récurrence, à toute combinaison linéaire (finie !) de variables aléatoires.

Proposition 2.6.7.
Une variable aléatoire positive d'espérance nulle est nulle presque-sûrement.

Démonstration.
On a une somme de termes positifs qui est nulle. □

Proposition 2.6.8 (Espérance des lois usuelles.)**.**
Soit $C \in \mathbb{R}$, $p \in [0,1]$, $n \in \mathbb{N}$, $a, b \in \mathbb{Z}$.

1. L'espérance d'une variable aléatoire constante de valeur C est égale à C.
2. L'espérance d'une variable aléatoire suivant la loi de Bernoulli de paramètre p est p.
3. L'espérance d'une variable aléatoire suivant la loi binomiale de paramètres n et p vaut np.
4. L'espérance d'une variable aléatoire suivant la loi uniforme sur $[\![a, b]\!]$, avec $a < b$, vaut $\dfrac{a+b}{2}$.

Démonstration.
Les deux premiers points se déduisent directement de la définition. Pour le troisième, on sait que l'espérance d'une variable aléatoire ne dépend que de sa loi. Étant donné une variable X sur un espace probabilisé Ω suivant la loi binomiale de paramètres n et p où $n \in \mathbb{N}$ et $p \in [0,1]$, on construit une variable aléatoire Y, sur un autre espace de probabilité Ω', de façon à ce que d'une part Y suive la loi binomiale de paramètres n et p et d'autre part Y s'écrive comme somme de de variables $Y_1, \dots, Y_n$ de Bernoulli indépendantes, toutes de paramètre p. On a alors

$$\begin{aligned}
\mathrm{E}(X) &= \mathrm{E}(Y) \\
&= \mathrm{E}\left(\sum_{i=1}^{n} Y_i\right) \\
&= \sum_{i=1}^{n} \mathrm{E}(Y_i) \\
&= \sum_{i=1}^{n} p \\
&= np.
\end{aligned}$$

Voici comment construire Ω' et Y. On pose $\Omega' = \{\, 0, 1 \,\}^n$ et pour tout $\omega = (\omega_1, \dots, \omega_n) \in \Omega'$, on pose $p_\omega = \prod_{i=1}^{n} \alpha(\omega_i)$, où $\alpha(0) = 1 - p$ et $\alpha(1) = p$.

On a

$$\begin{aligned}
\sum_{\omega \in \Omega'} p_\omega &= \sum_{(\omega_1, \dots, \omega_n) \in \{\, 0,1 \,\}^n} \prod_{i=1}^{n} \alpha(\omega_i) \\
&= \prod_{i=1}^{n} \sum_{\omega_i \in \{\, 0,1 \,\}} \alpha(\omega_i) \\
&= \prod_{i=1}^{n} (1 - p + p).
\end{aligned}$$

Ainsi, il existe une probabilité sur Ω' vérifiant pour tout $\omega \in \Omega'$, $P(\{\, \omega \,\}) = p_\omega$.

On définit alors, pour $i \in I$, Y_i la variable aléatoire de Bernoulli vérifiant $Y_i(\omega_1, \dots, \omega_n) = \omega_i$. Alors les Y_i, pour $i = 1, \dots, n$ sont des variables aléatoires mutuellement indépendantes. En effet, soit x_i pour $i = 1, \dots, n$ des éléments de $\{\, 0, 1 \,\}$.

Soit $i \in [\![1, n]\!]$, on a alors

$$\begin{aligned}
P(Y_i = x_i) &= \sum_{\substack{(\omega_1, \dots, \omega_n) \in \{\, 0,1 \,\}^n \\ \omega_i = x_i}} \prod_{j=1}^{n} \alpha(\omega_j) \\
&= \alpha(x_i) \prod_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\omega_j \in \{\, 0,1 \,\}} \alpha(\omega_j) \\
&= \alpha(x_i) \prod_{\substack{j=1 \\ j \neq i}}^{n} (1 - p + p) \\
&= \alpha(x_i).
\end{aligned}$$

Par ailleurs, on a

$$\begin{aligned}
P\left(\bigcap_{i=1}^{n} \{\, Y_i = x_i \,\}\right) &= P(\{\, (x_1, \dots, x_n) \,\}) \\
&= \prod_{i=1}^{n} \alpha(x_i),
\end{aligned}$$

donc
$$P\left(\bigcap_{i=1}^{n}\{ Y_i = x_i \}\right) = \prod_{i=1}^{n} P(Y_i = x_i).$$

Concernant le dernier point, il suffit de voir que si $X \hookrightarrow \mathscr{U}(\llbracket a, b\rrbracket)$, alors $X - a \hookrightarrow \mathscr{U}(\llbracket 0, b-a\rrbracket)$. Soit donc $n \in \mathbb{N}$ et $Y \hookrightarrow \mathscr{U}(\llbracket 0, n\rrbracket)$, alors

$$\mathrm{E}Y = \sum_{k=0}^{n} kP(Y = k) = \frac{1}{n+1}\sum_{k=0}^{n} k = \frac{n}{2},$$

ce qui permet de conclure par $\mathrm{E}X = \mathrm{E}[X-a] + a = \dfrac{b-a}{2} + a = \dfrac{a+b}{2}$. □

Proposition 2.6.9 (Formule de transfert)**.**
Soit X une variable aléatoire à valeur dans E et $f : X(\Omega) \to \mathbb{R}$. Alors

$$\mathrm{E}[f(X)] = \sum_{x \in X(\Omega)} P(X = x) f(x)$$

Démonstration.
On a déjà noté que pour tout $x \in X(\Omega)$, on a

$$P(X = x) = \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} P(\{ \omega \}).$$

En posant $s = \displaystyle\sum_{x \in X(\Omega)} P(X = x) f(x)$, on a alors successivement :

$$\begin{aligned}
s &= \sum_{x \in X(\Omega)} \left(\sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} P(\{ \omega \}) \right) f(x) \\
&= \sum_{x \in X(\Omega)} \left(\sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} P(\{ \omega \}) f(X(\omega)) \right) \\
&= \sum_{\omega \in \Omega} P(\{ \omega \}) f(X)(\omega) \\
&= \mathrm{E}[f(X)].
\end{aligned}$$

□

Remarque 2.6.10.
La formule de transfert montre que $\mathrm{E}[f(X)]$ dépend juste de la loi de X et de f. Pour calculer cette espérance, nul besoin d'obtenir la loi de $f(X)$!

Exercice 2.6.11.
Si $X \hookrightarrow \mathscr{U}(\llbracket 0, n\rrbracket)$, calculer $\mathrm{E}[X^2]$.

Proposition 2.6.12 (Espérance de variables indépendantes)**.**
Soit X et Y deux variables à valeurs réelles indépendantes. Alors $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$.

Démonstration.
On a, par la formule de transfert (appliquée à (X, Y) et à $(x, y) \mapsto xy$),

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

□

Remarque 2.6.13.
La réciproque est bien entendu fausse. Considérez par exemple la variable $X \hookrightarrow \mathscr{U}(-1, 0, 1)$ et $Y = X^2$.

Remarque 2.6.14.
Cela se généralise : si $X_1, \dots, X_n$ sont des v.a. réelles mutuellement indépendantes, alors $\mathrm{E}[X_1 \dots X_n] = \mathrm{E}[X_1] \dots \mathrm{E}[X_n]$.

Proposition 2.6.15 (Inégalité de Markov)**.**
Soit X une variable aléatoire réelle à *valeurs positives* presque-sûrement. Soit $t \in \mathbb{R}_+^*$. On a :

$$P(X \geqslant t) \leq \frac{\mathrm{E}(X)}{t}.$$

Démonstration.

On a

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{\omega\in\Omega} P(\{\,\omega\,\})X(\omega) \\
&= \sum_{\substack{\omega\in\Omega\\ X(\omega)\geqslant t}} P(\{\,\omega\,\})X(\omega) + \sum_{\substack{\omega\in\Omega\\ X(\omega)<t}} P(\{\,\omega\,\})X(\omega) \\
&\geqslant \sum_{\substack{\omega\in\Omega\\ X(\omega)\geqslant t}} P(\{\,\omega\,\})X(\omega) \\
&\geqslant \sum_{\substack{\omega\in\Omega\\ X(\omega)\geqslant t}} P(\{\,\omega\,\})t \\
&\geqslant tP(\{\,\omega\in\Omega \mid X(\omega)\geqslant t\,\}) \\
&\geqslant tP(X\geqslant t)
\end{aligned}$$

Sinon, il suffit de voir que $t\mathbf{1}_{X\geqslant t}$ est une variable aléatoire inférieure à X p.s, puis de considérer son espérance ... □

Remarque 2.6.16.
C'est une inégalité fondamentale en probabilité, vous la retrouverez de nombreuses fois : retenez la bien !

Exemple 2.6.17.
n étudiants ont travaillé au mois de juillet dernier. En moyenne, leur travail leur a rapporté 750 euros chacun. Donner un majorant de la proportion d'étudiants ayant gagné au moins 1000 euros.

2.7. Variance, écart type et covariance

Définition 2.7.1 (Moments)**.**
Soit X une variable aléatoire réelle. Soit $r \in \mathbb{N}^*$.

On appelle moment d'ordre r de X et on note $m_r(X)$ l'espérance de X^r :

$$m_r(X) = \mathrm{E}(X^r).$$

On appelle moment centré d'ordre r de X et on note $\mu_r(X)$ le moment d'ordre r de $X - \mathrm{E}(X)$:

$$\mu_r(X) = \mathrm{E}((X - \mathrm{E}(X))^r).$$

On appelle variance de X et on note $V(X)$ son moment centré d'ordre 2 :

$$V(X) = \mathrm{E}\Big((X - \mathrm{E}(X))^2\Big).$$

On appelle écart-type de X et on note $\sigma(X)$ la racine carrée de la variance :

$$\sigma(X) = \sqrt{V(X)}.$$

On dit que X est une variable aléatoire *réduite* si sa variance est 1.

Remarque 2.7.2.
1. Les moments et les moments centrés ne dépendent que de la loi de X.
2. Intuitivement la variance mesure la « moyenne » du carré des écarts à la « moyenne ». C'est donc une mesure de la dispersion des valeurs autour de la valeur moyenne prise par la variable aléatoire X. Si la variable X est exprimée dans une unité u, la variance sera exprimée dans l'unité u^2 (et l'écart-type dans l'unité u).

Proposition 2.7.3 (Formule de König-Huygens)**.**
Soit X une variable aléatoire réelle. Alors la variance est la différence entre l'espérance de son carré et le carré de son espérance :

$$V(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2.$$

Démonstration.
Il suffit d'utiliser la définition, la linéarité de l'espérance et le résultat sur l'espérance d'une variable constante pour obtenir successivement :

$$\begin{aligned}
V(X) &= \mathrm{E}\big((X-\mathrm{E}(X))^2\big) \\
&= \mathrm{E}\big(X^2 - 2\mathrm{E}(X)X + \mathrm{E}(X)^2\big) \\
&= \mathrm{E}\big(X^2\big) - 2\mathrm{E}(X)\mathrm{E}(X) + \mathrm{E}(X)^2 \\
&= \mathrm{E}\big(X^2\big) - \mathrm{E}(X)^2.
\end{aligned}$$

□

Proposition 2.7.4.
Soit X une variable aléatoire réelle, et a et b deux réels.

Alors

$$\begin{aligned}
\mathrm{E}(aX+b) &= a\mathrm{E}(X) + b, \\
V(aX+b) &= a^2V(X), \\
\sigma(aX+b) &= |a|\sigma(X).
\end{aligned}$$

En particulier, si $\sigma(X) \neq 0$, la variable Y définie par

$$Y = \frac{X - \mathrm{E}(X)}{\sigma(X)}$$

est une variable centrée réduite.

Démonstration.
On connaît déjà le premier point. Il suffit de remplacer par les définitions et de calculer pour conclure pour les autres. □

Proposition 2.7.5 (Variance des lois usuelles)**.**
Soit $p \in [0,1]$ et $n \in \mathbb{N}^*$.

1. Toute variable aléatoire constante p.s. est de variance nulle.
2. Toute variable aléatoire réelle suivant la loi de Bernoulli de paramètre p a pour variance $p(1-p)$.
3. Toute variable aléatoire réelle suivant la loi binomiale de paramètres p et n a pour variance $np(1-p)$.
4. Toute variable aléatoire rélle suivant la loi uniforme sur $[\![1, n]\!]$, a pour variance $\dfrac{n^2-1}{12}$.

Démonstration.
1. Direct.
2. Considérons X une variable aléatoire réelle suivant la loi de Bernoulli de paramètre p. Alors $V(X) = E(X^2) - E(X)^2$. or $P(X^2 = 1) = P(X = 1) = p$ et donc $E(X^2) = p$. donc $V(X) = p - p^2 = p(1-p)$.
3. Soit X une variable aléatoire réelle suivant la loi binomiale de paramètres p et n. Alors on peut construire une variable aléatoire réelle Y de même loi que X, donc de même variance, de façon que Y soit la somme de n variables de Bernoulli $Y_1, \dots, Y_n$ de paramètre p mutuellement indépendantes.

   On verra plus loin que *dans le cas particulier de variables deux à deux indépendantes*, la variance de la somme est égale à la somme des variances. on a donc ici :
   $$V(Y) = \sum_{i=1}^{n} V(Y_i) = np(1-p)$$
4. Faisons le calcul dans le cas général $X \hookrightarrow \mathscr{U}([\![a, b]\!])$, avec $a < b$. $X - a \hookrightarrow \mathscr{U}([\![0, b-a]\!])$ et $V(X-a) = V(X)$. Soit $Y \hookrightarrow \mathscr{U}([\![0, n]\!])$, avec $n \in \mathbb{N}$, calculons $V(Y)$. Par la formule de König-Huygens, $V(Y) = \mathrm{E}\left[Y^2\right] - \mathrm{E}\left[Y\right]^2 = \mathrm{E}\left[Y^2\right] - \dfrac{n^2}{4}$. De plus, par la formule de transfert,
   $$\begin{aligned}\mathrm{E}\left[Y^2\right] &= \sum_{k=0}^{n} k^2 P(Y = k)\\ &= \frac{1}{n+1}\sum_{k=0}^{n} k^2\\ &= \frac{n(2n+1)}{6}.\end{aligned}$$
   Ainsi,
   $$V(Y) = \frac{2n(2n+1)}{12} - \frac{3n^2}{12} = \frac{n(n+2)}{12}.$$
   Donc $V(X) = \dfrac{(b-a)(b-a+2)}{12}$ et l'on retrouve bien l'énoncé avec $b = n$ et $a = 1$.

□

Théorème 2.7.6 (Inégalité de Bienaymé-Tchebychev)**.**
Soit X une variable aléatoire réelle, $V(X)$ sa variance, soit $\varepsilon > 0$. Alors

$$P(|X - \mathrm{E}X| \geqslant \varepsilon) \leqslant \frac{V(X)}{\varepsilon^2}.$$

Démonstration.
Il suffit de constater que $(X - \mathrm{E}X)^2$ est une v.a. positive, donc d'après l'inégalité de Markov :

$$\begin{aligned}P(|X - \mathrm{E}X| \geqslant \varepsilon) &= P\left((X - \mathrm{E}X)^2 \geqslant \varepsilon^2\right)\\ &\leqslant \frac{\mathrm{E}\left[(X - \mathrm{E}X)^2\right]}{\varepsilon^2}\\ &\leqslant \frac{V(X)}{\varepsilon^2}.\end{aligned}$$

□

Remarque 2.7.7.
Cette inégalité contrôle la déviation d'une variable aléatoire par rapport à sa moyenne. C'est un résultat fondamental !

Corollaire 2.7.8.
Une variable aléatoire de variance nulle est constante presque-sûrement.

Démonstration.
Pour tout $\mathrm{E} > 0$, par l'inégalité de Bienaymé-Tchebychev on a $P(|X - \mathrm{E}X| \geqslant \varepsilon) = 0$, donc s'il existe $x \in \mathbb{R}$, avec $x \neq \mathrm{E}X$ et $P(X = x) \neq 0$, alors il suffit de prendre $\varepsilon = \frac{1}{2}|x - \mathrm{E}X|$ pour obtenir une contradiction. Ainsi, $P(X = \mathrm{E}X) = 1$.

On pouvait aussi simplement remarquer que $(X - \mathrm{E}[X])^2$ est positive, donc que si $V(X) = 0$ alors p.s. $(X - \mathrm{E}[X])^2 = 0$, donc p.s. $X = \mathrm{E}[X]$. □

Définition 2.7.9 (Covariance)**.**
Soit X et Y deux variables aléatoires réelles. On appelle covariance de X et Y et on note $\mathrm{Cov}(X, Y)$ le réel

$$\mathrm{Cov}(X, Y) = \mathrm{E}((X - \mathrm{E}X)(Y - \mathrm{E}Y)).$$

Si $\mathrm{Cov}(X, Y) = 0$, on dit que X et Y sont décorrélées.

Remarque 2.7.10.
$\mathrm{Cov}(X, X)$ n'est autre que $V(X)$ et $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

Remarque 2.7.11.
Si $\sigma(X) = 0$, nous avons vu que $X = \mathrm{E}[X]$ p.s, donc $\mathrm{Cov}(X, Y) = 0$.

Si $\sigma(X)$ et $\sigma(Y)$ ne sont pas nuls, on définit le *coefficient de corrélation* de X et de Y par

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

L'inégalité de Cauchy-Schwarz (que nous verrons bientôt) permet de montrer deux résultats intéressants :

- $\rho(X, Y) \in [-1, 1]$;
- si $|\rho(X, Y)| = 1$, alors il existe $a, b, c \in \mathbb{R}$ tels que $aX + bY = c$ p.s, *i.e.* X et Y sont p.s. en relation affine.

On montre aussi facilement que, si l'on résout le *problème de régression affine de Y sur X par moindres carrés*, *i.e.* si l'on minimise en $a, b \in \mathbb{R}$ la quantité :

$$\mathrm{E}\left[(Y - aX - b)^2\right],$$

alors le coefficient a optimal (c'est la pente de la droite de régression de Y sur X) est

$$a = \frac{\mathrm{Cov}(X, Y)}{V(X)} = \rho(X, Y) \times \frac{\sigma(Y)}{\sigma(X)}.$$

Remarquons que, dans ce cas, effectuer la régression de Y sur X n'est pas la même chose que d'effectuer la régression de X sur Y !

Si X et Y sont centrées-réduites (en pratique, il convient souvent de centrer-réduire ses données avant d'effectuer un traitement statistique dessus), on s'apperçoit que

$$a = \rho(X, Y).$$

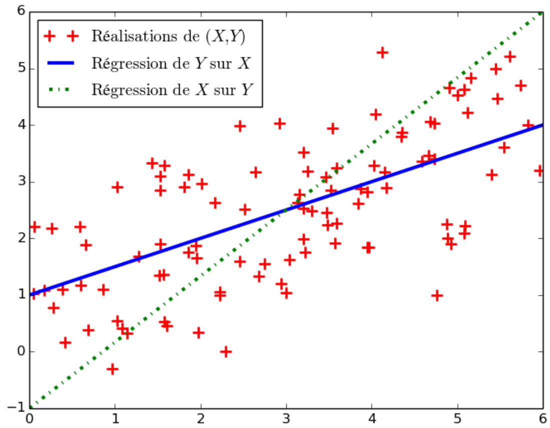

FIGURE 2 – Exemple de régression affine par moindres carrés, ici $\rho(X, Y) = \sqrt{\frac{3}{7}} \approx 0,65$.

Exercice 2.7.12.
Soit (Ω, P) un espace probabilisé fini, X une variable aléatoire réelle définie sur Ω et f une fonction réelle croissante.

Montrer que $\mathrm{Cov}(X, f(X)) \geqslant 0$.

Proposition 2.7.13.
Soit X et Y deux variables aléatoires réelles. Alors on a

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

Démonstration.
Comme pour la formule de König-Huygens :

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

□

Remarque 2.7.14.
Par la formule de transfert,

$$\mathrm{E}\left[XY\right] = \sum_{x\in X(\Omega),\ y\in Y(\Omega)} xyP(X=x, Y=y).$$

Corollaire 2.7.15.
Deux variables aléatoires indépendantes ont une covariance nulle.

Remarque 2.7.16.
Comme montré par l'exemple $X \hookrightarrow \mathscr{U}(-1,0,1)$ et $Y = X^2$, deux variables peuvent être non corrélées mais sans être indépendantes.

Proposition 2.7.17.
Soit $n \in \mathbb{N}$ et $X_1, \dots, X_n$ n variables aléatoires réelles. Alors on a

$$V(\sum_{i=1}^n X_i) = \sum_{i=1}^n V(X_i) + 2\sum_{1\leqslant i<j\leqslant n} \mathrm{Cov}(X_i, X_j).$$

Démonstration.
Posons, pour $i = 1, \dots, n$, $Y_i = X_i - \mathrm{E}(X_i)$ et $X = \sum_{i=1}^n X_i$.

On a :

$$\begin{aligned}V(X) &= \mathrm{E}\left(\left(\sum_{i=1}^n Y_i\right)^2\right)\\ &= \mathrm{E}\left(\sum_{i=1}^n Y_i^2 + 2\sum_{1\leqslant i<j\leqslant n} Y_iY_j\right)\\ &= \sum_{i=1}^n \mathrm{E}\big(Y_i^2\big) + 2\sum_{1\leqslant i<j\leqslant n} \mathrm{E}(Y_iY_j)\\ &= \sum_{i=1}^n V(X_i) + 2\sum_{1\leqslant i<j\leqslant n} \mathrm{Cov}(X_i, X_j).\end{aligned}$$

□

Corollaire 2.7.18.
Dans le cas particulier où les variables $X_1, \dots, X_n$ sont deux à deux décorrélées, la variance de leur somme est égale à la somme de leurs variances.

Remarque 2.7.19.
On déduit de ce résultat la variance de la loi binomiale, comme on l'a vu plus haut.